

Genetic Evidence of Paleolithic Colonization and Neolithic Expansion of Modern Humans on the Tibetan Plateau

Xuebin Qi,^{†,1} Chaoying Cui,^{†,2} Yi Peng,^{†,1,3} Xiaoming Zhang,^{†,1,3} Zhaohui Yang,^{†,1,3} Hua Zhong,^{‡,1} Hui Zhang,¹ Kun Xiang,^{1,3} Xiangyu Cao,^{1,3} Yi Wang,^{1,3} Ouzhuluobu,² Basang,⁴ Ciwangsangbu,⁴ Bianba,² Gonggalanzi,² Tianyi Wu,⁵ Hua Chen,⁶ Hong Shi,^{*,1} and Bing Su^{*,1}

¹State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

²High Altitude Medical Research Center, School of Medicine, Tibetan University, Lhasa, China

³University of Chinese Academy of Sciences, Beijing, China

⁴People's Hospital of Dangxiong County, Dangxiong, China

⁵National Key Laboratory of High Altitude Medicine, High Altitude Medical Research Institute, Xining, China

⁶Department of Epidemiology and Biostatistics, Harvard School of Public Health

[‡]Present address: Department of Pathology & Immunology, Baylor College of Medicine, Houston, TX

[†]These authors contributed equally to this work.

*Corresponding authors: E-mail: sub@mail.kiz.ac.cn, shihong@mail.kiz.ac.cn.

Associate editor: Sarah Tishkoff

Abstract

Tibetans live on the highest plateau in the world, their current population size is approximately 5 million, and most of them live at an altitude exceeding 3,500 m. Therefore, the Tibetan Plateau is a remarkable area for cultural and biological studies of human population history. However, the chronological profile of the Tibetan Plateau's colonization remains an unsolved question of human prehistory. To reconstruct the prehistoric colonization and demographic history of modern humans on the Tibetan Plateau, we systematically sampled 6,109 Tibetan individuals from 41 geographic populations across the entire region of the Tibetan Plateau and analyzed the phylogeographic patterns of both paternal ($n = 2,354$) and maternal ($n = 6,109$) lineages as well as genome-wide single nucleotide polymorphism markers ($n = 50$) in Tibetan populations. We found that there have been two distinct, major prehistoric migrations of modern humans into the Tibetan Plateau. The first migration was marked by ancient Tibetan genetic signatures dated to approximately 30,000 years ago, indicating that the initial peopling of the Tibetan Plateau by modern humans occurred during the Upper Paleolithic rather than Neolithic. We also found evidences for relatively young (only 7–10 thousand years old) shared Y chromosome and mitochondrial DNA haplotypes between Tibetans and Han Chinese, suggesting a second wave of migration during the early Neolithic. Collectively, the genetic data indicate that Tibetans have been adapted to a high altitude environment since initial colonization of the Tibetan Plateau in the early Upper Paleolithic, before the last glacial maximum, followed by a rapid population expansion that coincided with the establishment of farming and yak pastoralism on the Plateau in the early Neolithic.

Key words: Tibetan Plateau, demographic history, human migration, last glacial maximum.

Introduction

The Tibetan Plateau, with an average elevation of more than 4,000 m, represents one of the most extreme environments for human settlement due to severe hypoxia at high-altitude. Humans have had to develop physiological adaptations to cope with the environmental extremes at high altitude. At present, approximately 5 million indigenous Tibetans live on the Plateau and are well adapted to this high-altitude hypoxic environment. Studies have shown that Tibetans acquired an improved performance in utilizing the reduced oxygen as a result of natural selection (Wu 2001; Wu and Kayser 2006; Simonson et al. 2010; Bigham et al. 2010; Yi et al. 2010; Beall 2011; Peng, Yang, et al. 2011; Wang, Zhang, et al. 2011; Xu et al.

2011). However, the chronology of precisely when and how the ancestors of modern Tibetans first settled on the Tibetan Plateau and began permanently residing at high altitude has been under substantial debates (Aldenderfer and Zhang 2004; Brantingham and Gao 2006; Brantingham et al. 2007; Yuan et al. 2007; Shi et al. 2008; Zhao et al. 2009; Aldenderfer 2011). Despite the controversy, the timing of this colonization is crucial for revealing the signatures of natural selection in human populations implicated in the early peopling of the Tibetan Plateau, as well as for delineating the mechanisms and history of genetic adaptation to high-altitude hypoxia.

The exact timing of the peopling of the Tibetan Plateau has been controversial. Archaeological investigations from the Upper Paleolithic (40–10 thousand years ago [ka]) and

Neolithic (10–4 ka) sites throughout the area suggest that the earliest human foraging may have begun approximately 40–30 ka (Zhang and Li 2002; Aldenderfer and Zhang 2004; Brantingham and Gao 2006; Madsen et al. 2006; Brantingham et al. 2007; Yuan et al. 2007; Gao et al. 2008; Hou et al. 2010; Aldenderfer 2011). The initial inhabitants were likely the seasonal hunter–gatherer groups who lived in the relatively low altitude areas (<3,000 m) of the Tibetan Plateau, as permanent occupation at high altitude (>3,000 m) did not seem to begin until the advent of farming and pastoral economies approximately 8.2–6 ka (Zhang and Li 2002; Aldenderfer and Zhang 2004; Brantingham and Gao 2006; Madsen et al. 2006; Brantingham et al. 2007; Yuan et al. 2007; Gao et al. 2008; Hou et al. 2010; Aldenderfer 2011). Current archaeological data suggest that an ancient initial occupation of the Plateau, followed by multiple migrations at different times and from different places may have created a complex, mosaic population history (Aldenderfer 2011). However, the archaeological inference of population history has mainly relied on the limited cultural artifacts, and the chronology of prehistoric colonization of the Tibetan Plateau is still unclear.

Genetic analyses of human populations are also informative for tracing population history. For example, by using a Bayesian Skyline method and 425 whole mitochondrial DNA (mtDNA) sequence, Gignoux et al. (2011) accurately reconstructed demographic fluctuations for European, sub-Saharan African, and Southeastern Asian populations over the past 50,000 years. Analyses of the phylogeographic patterns of both paternal (Y chromosome) and maternal (mtDNA) lineages have identified both ancient and young genetic components in contemporary Tibetan populations (Torroni et al. 1994; Su et al. 2000; Shi et al. 2008; Zhao et al. 2009; Qin et al. 2010). This suggests that modern Tibetan populations may have been genetically formed from more than one ancestral population that ventured into the Plateau during prehistory. In addition, genome-wide scans of autosomal single nucleotide polymorphisms (SNPs) have identified endothelial PAS domain protein 1 (*EPAS1*, also called hypoxia-inducible factor 2α , *HIF-2 α*) as a strong candidate gene for conferring high altitude adaptation to Tibetans (Bigham et al. 2010; Simonson et al. 2010; Yi et al. 2010; Peng, Yang, et al. 2011; Wang, Zhang, et al. 2011; Xu et al. 2011). The onset of natural selection on the *EPAS1* gene can be considered as the beginning of exposure to high-altitude hypoxia. Continuous selection across generations would eventually lead to genetic adaptation to the hypoxic stress. Therefore, the onset of natural selection on the *EPAS1* gene can also be roughly considered the beginning of human settlement on the Tibetan Plateau. But two of these studies have estimated a contradictory time for the onset of natural selection on the *EPAS1* gene in Tibetans (Yi et al. 2010; Peng, Yang, et al. 2011). For example, Yi et al. (2010) estimated that the onset of natural selection on the *EPAS1* gene started approximately 2,750 years ago since the divergence of Han Chinese and Tibetan populations, while Peng, Yang, et al. (2011) estimated the onset of selection to be 18 ka, a timeline close to the end of last glacial maximum (LGM, 22–18 ka) on the Plateau, by re-sequencing the full sequence of the *EPAS1*

gene (94 kb). These genetic findings were generally based on a small sample of Tibetans and hence called for a detailed, systematic genetic analysis of Tibetan populations.

To reconstruct the chronological profile of prehistoric colonization and the demographic history of modern humans on the Tibetan Plateau, we systematically sampled 6,109 Tibetan individuals from 41 geographic populations across the entire region of the Tibetan Plateau, and conducted phylogeographic analyses using paternal, maternal, and genome-wide autosomal SNP markers. From these data, we found evidence of two major migrations of modern humans into the Tibetan Plateau, with the permanent occupation likely occurring in the early Upper Paleolithic before the LGM, and a recent migration and population expansion beginning in the early Neolithic, coincident with the emergence of farming and yak pastoralism on the Plateau.

Results

Y-Chromosomal Diversity in Tibetan Populations

Based on the current phylogeny of human Y-chromosomes, we determined the Y-chromosomal haplogroups of 2,354 Tibetan males from 41 populations (fig. 1 and supplementary fig. S1, Supplementary Material online). The majority of paternal lineages in Tibetans (87.80%) were East Asian specific lineages (Shi et al. 2005, 2008). Among them, D-M174 occurred at the highest frequency (54.33%, the majority of its sublineages were D3*-P99), followed by O-M175 (33.47%, the majority of its sublineages were O3a3c1-M117). Another relatively rare lineage in Tibetans was N-M231 (5.65%), the majority of its sublineages were N1*-LLY22G (5.06%) (supplementary figs. S1 and S2A, Supplementary Material online). N-M231 was also an ancient lineage widespread in Eurasia (Zhong et al. 2011). Besides these three major haplogroups, Tibetans also harbored several other rare lineages such as C3-M130, R-M207, Q, T, K-M, F2, and J occurring at very low frequencies (0.13–2.04%, supplementary fig. S1, Supplementary Material online). In general, within the Tibetan populations, the distribution of the Y-chromosomal haplogroups was relatively homogenous and no obvious substructure was observed (supplementary fig. S1, Supplementary Material online).

D-M174 represents an important lineage in mainland East Asia, as it was prevalent in Tibetans (54.33%) and Japanese (34.7%) (Hammer et al. 2006), but was rare or absent in most other Eurasian populations, including Han Chinese (3.9%) (Su et al. 1999, 2000; Wells et al. 2001). D-M174 is a subbranch of an ancient Y-chromosomal macrohaplogroup DE which was defined by M1 or “YAP” marker dating to 65 ka (Karafet et al. 2008). There was a deep divergence (>37 ka) of D-M174 between Tibetans (D*-M174, D1-M15, and D3*-P99) and Japanese (D2-M55) (Shi et al. 2008). In mainland East Asia, D-M174 carriers were mostly D*-M174 and D1-M15 (Shi et al. 2008). The distribution of the D-M174 sublineages in Tibetans was shown in supplementary figure S1 (Supplementary Material online). D3a-P47, a Tibetan-specific subhaplogroup under D-M174, was the most dominant haplogroup and represents 55.8% of the D-M174 lineages in Tibetans.

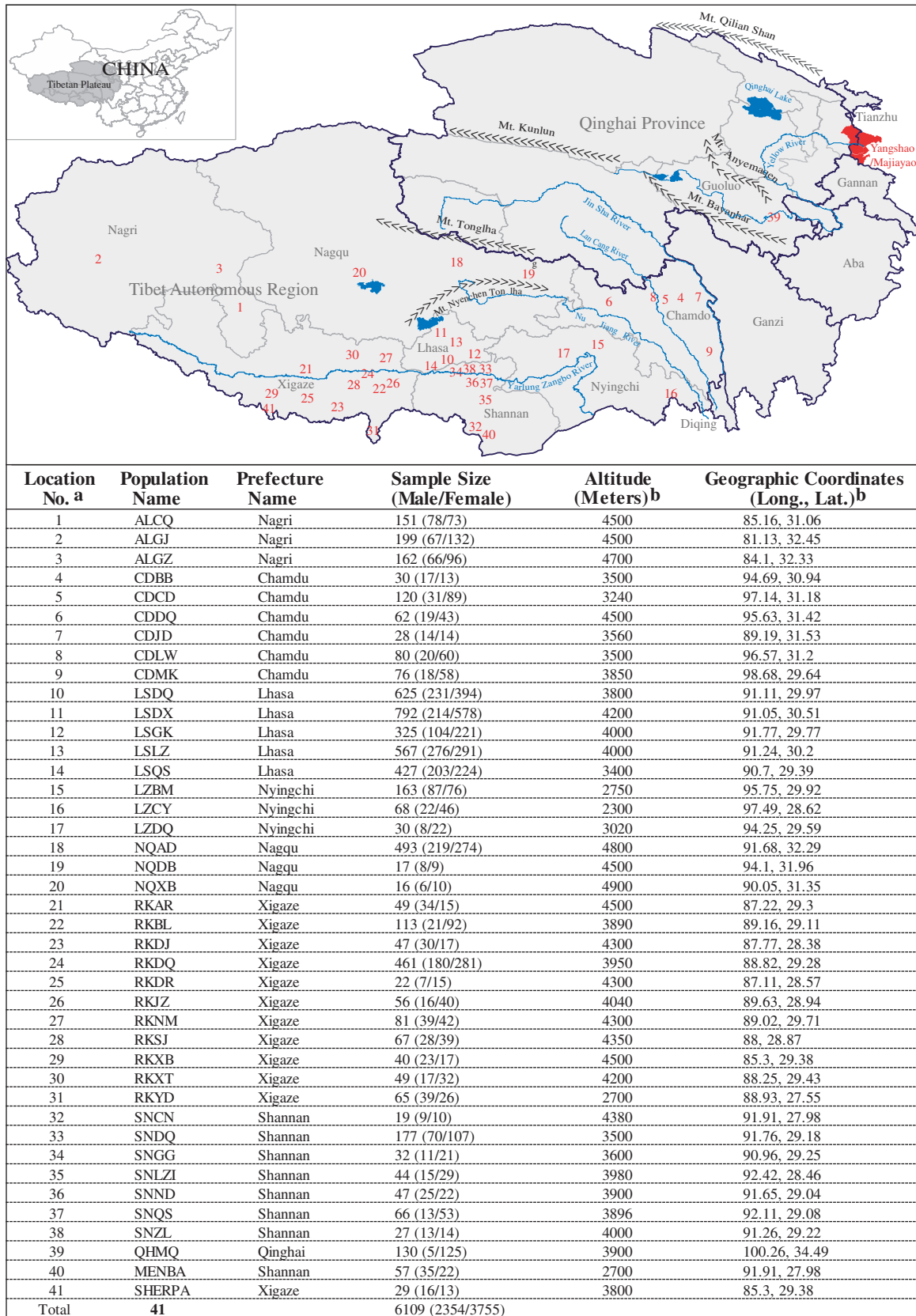


FIG. 1. A map showing the geographic area modern Tibetans inhabit today and the sampling locations for this study. The red area marked with Yangshao/Majiyao is the Neolithic sites (8,000 years ago) neighboring the Tibetan Plateau. ^aThe location no. corresponds to the locations on the map. ^bThe altitude and geographic coordinates data for a population in a particular county were roughly taken from locations where the County Administration Centers were located.

Downloaded from https://academic.oup.com/mbe/article/30/8/1761/1015497 by guest on 20 August 2022

The coalescence age of the Tibetan D-M174 lineage was dated to 32.5 ± 11.5 ka using the T_d estimator (Zhitovovskiy 2001; Zhitovovskiy et al. 2004) (table 1). Further dating of the D-M174 subhaplogroups (D1-M15, D1a-N1, and D3*-P99) supports the antiquity of the Tibetan Y-chromosomes (18.7–27.9 ka) (table 1). In particular, D3*-P99, a Tibetan-specific subhaplogroup, was estimated to 18.7 ± 6.6 ka. The only exception was D3a-P47, a relatively young lineage dated to 10.1 ± 3.6 ka, a time close to the beginning of the Neolithic period. The relatively young age of D3a-P47 is puzzling and seems inconsistent with the very old age estimates for all other D-M174 branches. Since D3a-P47 was the most dominant D-M174 haplogroup in Tibetans, we further analyzed its detailed haplotype structure using the Y-chromosomal STR data. The result revealed a star-like STR network for D3a-P47 (fig. 2A), that is, one or several tightly-linked core STR haplotypes surrounded by other STR haplotypes only one or two mutation steps away from the core, indicating the molecular signature of recent population expansion.

O3a3c1-M117 was another major Y-chromosomal haplogroup in Tibetans (29.82%), and it also occurred at relatively high frequency in Han Chinese (9.6–16.3%) and many other East and Southeast Asian populations (5.5–16.0%) (supplementary fig. S1 and table S1, Supplementary Material online). Among the four basal lineages of O3-M122 (O3a1, O3a2, O3a3, and O3a4), only O3a3c1-M117 was detected in modern Tibetans, whereas nearly all O3a basal lineages were found in other Asian populations including Han Chinese, Tibeto-Burman, Hmong-Mien, and Mon-Khmer populations (supplementary table S1 and references therein). In addition, the coalescence time of O3a3c1 in Tibetans (18.4 ± 7.4 ka) was much younger than that in Han Chinese (32.2 ± 7.9 ka) (table 1 and supplementary table S2, Supplementary Material online). Collectively, this suggests that O3a3c1-M117 originated outside of Tibet, and was brought to the Tibetan Plateau from other places, consistent with previous studies showing a southern origin and an initial northward migration of O3-M122 about 25–30 ka in eastern Asia (Cai et al. 2011; Shi et al. 2005).

The Y-chromosomal STR network analysis of O3a3c1-M117 in Tibetans also revealed a similar star-like network as seen in D3a-P47 (fig. 2B), confirming the proposed recent population expansion within the Tibetan Plateau. We further constructed the Y-chromosomal STR network of O3a3c1-M117 by including data from all available East Asian populations. As shown in supplementary figure S2B (Supplementary Material online), the network was similar with the one that only includes Tibetans, and the major Y-chromosomal STR haplotypes from different populations mingled together with relatively short distances from each other. To further test the demic diffusion of the Neolithic Han Chinese agriculturalists into the Tibetan Plateau, we examined the divergence time of O3a3c1-M117 between Tibetans and Han Chinese. The estimated divergence time was 11.5 ± 4.8 ka (table 1), consistent with the proposed expansion of the earliest Neolithic agriculturalists. This divergence time was also consistent with the rapid expansion of D3a-P47 (10.1 ± 3.6 ka) (table 1), a lineage

specific to the Tibetan autochthons, during the Neolithic agricultural or agro-pastoral transition.

Mitochondrial DNA Diversity in Tibetans

As Y-chromosomal data only represents the paternal lineage, using genotyping and sequencing, we determined the mtDNA haplogroups for 6,109 individuals from 41 populations across the Tibetan Plateau (fig. 1 and supplementary fig. S3, Supplementary Material online). Similar to the profile of the Y-chromosomal haplogroups, the majority of maternal lineages in Tibetans (M9a, D, A, F, G, M8, M13, B, and M62, 90.99%) are also East Asian founding lineages (Torroni et al. 1992, 1994; Torroni, Schurr, et al. 1993; Yao et al. 2002; Kong, Yao, Sun, et al. 2003; Tanaka et al. 2004; Kong et al. 2006, 2011).

M9a, a lineage initially defined by Kong, Yao, et al. (2003), was the most prevalent maternal lineage in Tibetans with an overall frequency of 22.48% ranging from 12.26% to 29.24% among geographic regions within the Plateau (supplementary fig. S3, Supplementary Material Online). This frequency was similar to that previously reported in two studies with relatively small sample sizes (Zhao et al. 2009; Qin et al. 2010). Previous studies suggested that M9a originated in Southeast Asia and there was an inland post-glacial dispersal in East Asia (Torroni et al. 1994; Tanaka et al. 2004; Soares et al. 2008; Peng, Palanichamy, et al. 2011). The M9a lineage can be further divided into two subhaplogroups, M9a1a (16.30%) and M9a1b1 (4.73%), both of which exhibited a star-like phylogeny and was dated to 7.4 ± 1.4 ka and 9.5 ± 2.5 ka respectively (table 1 and fig. 3A).

Haplogroup D, initially defined by Torroni et al. (1992), was the second most dominant maternal lineage in Tibetans (16.53%) (supplementary fig. S3, Supplementary Material online). This lineage represents one of the main maternal founding lineages in eastern Asia with an overall coalescence age of 38 ka (Torroni et al. 1992; Torroni, Schurr, et al. 1993; Torroni, Sukernik, et al. 1993; Yao et al. 2002; Kong, Yao, et al. 2003; Tanaka et al. 2004; Kong et al. 2006; Zheng et al. 2011; Behar et al. 2012). It was further divided into D4 (76.93%), D5 (18.12%) and D6 (1.68%) in Tibetans (fig. 3B and supplementary fig. S3, Supplementary Material online). These three subhaplogroups were previously defined in Asian populations (Yao et al. 2002; Tanaka et al. 2004). D4 was the major subhaplogroup in Tibetans with an ancient coalescence age of 29.9 ± 5.9 ka (table 1 and fig. 3B). Most D5 lineages in Tibetans belong to D5a2a (71.26%), which had a coalescence age of 40.4 ± 13.8 ka (table 1 and fig. 3B). Haplogroup D reflects the antiquity of Tibetan mtDNAs and a deep divergence from other East Asian populations including Han Chinese (26–27 ka) (table 1 and supplementary fig. S4, Supplementary Material online).

Haplogroup A was another major maternal lineage in Tibetans (14.63%) (supplementary fig. S3, Supplementary Material online). This lineage was initially defined by Torroni et al. (1992). The majority subhaplogroups of haplogroup A in Tibetans were A4 (53.80%) and A11 (44.63%). A4 was a widespread lineage in East Asian populations (Torroni et al. 1992; Torroni, Schurr, et al. 1993; Yao et al. 2002; Kong,

Table 1. Estimated Coalescence Time within Tibetan Populations and Divergence Time from Han Chinese for Y Chromosome and mtDNA Haplogroups in Tibetans.

Haplogroups (%) ^a	Subhaplogroups	Coalescence Time (T_c) within Tibetan Populations ^b		Divergence Time (T_d) from Han Chinese Populations ^c		Climatic Period ^d
		<i>N</i>	$T_c \pm SD$ (ka)	<i>N</i>	$T_d \pm SD$ (ka)	
Y Chromosome (by STR variation)						
<u>D-M174 (54.33%)</u>	D-M174	1,278	32.5 ± 11.5	/	ND	Upper Paleolithic
	D1*-M15	33	20.7 ± 8.9	/	ND	LGP
	D1a-N1	328	27.9 ± 9.9	/	ND	Upper Paleolithic
	D3*-P99	203	18.7 ± 6.6	/	ND	LGP
	<u>D3a-P47</u>	<u>714</u>	<u>10.1 ± 3.6</u>	/	ND	Neolithic
N-M231 (5.30%)	N-M231	133	24.4 ± 7.9	144	22.4 ± 7.4	Upper Paleolithic
	N1*-LLY22G	119	23.1 ± 8.5	130	22.4 ± 7.8	Upper Paleolithic
<u>O3-M175 (33.47%)</u>	O3a-M324	772	23.0 ± 8.9	190	16.7 ± 4.9	Upper Paleolithic
	O3a*-M324	13	31.7 ± 12.8	98	16.7 ± 7.3	Upper Paleolithic
	O3a3c*-M134	61	22.1 ± 7.9	31	13.4 ± 5.7	LGP
	<u>O3a3c1-M117</u>	<u>697</u>	<u>18.4 ± 7.4</u>	<u>47</u>	<u>11.5 ± 4.8</u>	<u>LGP</u>
Mitochondrial DNA (by HVS-I variations)						
<u>A (14.63%)</u>	A11-16093	238	5.6 ± 1.4	1	ND	Neolithic
	<u>A11</u>	<u>399</u>	<u>13.4 ± 6.9</u>	<u>2</u>	<u>16.5 ± 10.4</u>	<u>LGP</u>
	<u>A4</u>	<u>481</u>	<u>15.0 ± 4.6</u>	<u>49</u>	<u>13.7 ± 3.2</u>	<u>LGP</u>
B (3.76%)	B4	157	35.9 ± 13.7	21	ND	Upper Paleolithic
	B4-16299	50	4.3 ± 2.1	0	14.0 ± 13.0	Neolithic
	B5b	60	7.5 ± 2.3	18	10.6 ± 3.6 ^e	Neolithic
<u>D (16.53%)</u>	<u>D4</u>	<u>777</u>	<u>29.9 ± 5.9</u>	<u>84</u>	<u>25.9 ± 6.3</u>	<u>Upper Paleolithic</u>
	<u>D4* (w/o D4j3)</u>	<u>705</u>	<u>25.6 ± 5.1</u>	<u>81</u>	<u>20.4 ± 5.0</u>	<u>LGP</u>
	D4j3	94	8.8 ± 3.5	3	8.1 ± 3.3	Neolithic
	D5a2a	119	40.4 ± 13.8	24	27.0 ± 12.0	Upper Paleolithic
	D6a	89	31.6 ± 10.6	3	23.1 ± 10.4	Upper Paleolithic
<u>F (11.44%)</u>	<u>F1*</u>	<u>319</u>	<u>10.2 ± 2.0</u>	<u>23</u>	<u>9.6 ± 2.0</u>	<u>Neolithic</u>
	F1-16284	97	15.1 ± 8.2	2	ND	LGP
	F1b	96	20.4 ± 12.9	9	ND	LGP
	F1c	132	19.0 ± 8.3	10	17.5 ± 8.1	LGP
G (8.22%)	G*	77	20.2 ± 7.1	5	15.6 ± 6.0	LGP
	G2a-16227	67	31.4 ± 8.3	17	30.0 ± 7.9	Upper Paleolithic
	G2a-16129	67	3.8 ± 2.6	2	ND	Neolithic
	G2b	48	20.2 ± 8.9	2	ND	LGP
	G3*	99	17.3 ± 5.2	10	16.0 ± 4.8	LGP
	G3a1	123	13.8 ± 5.4	0	ND	LGP
M8 (7.71%)	C4a1	149	27.0 ± 11.4	0	ND	Upper Paleolithic
	C4a2'3'4	118	21.4 ± 7.6	3	ND	LGP
	Z*	92	21.8 ± 6.0	34	20.2 ± 4.9	LGP
<u>M9a (22.48%)</u>	<u>M9a1a</u>	<u>979</u>	<u>7.4 ± 1.4</u>	<u>18</u>	<u>23.6 ± 16.6</u>	<u>Neolithic</u>
	M9a1b1	287	9.5 ± 2.5	0	ND	Neolithic
M11	M11a	48	11.5 ± 5.2	1	ND	LGP
M13	M13a1	83	6.4 ± 1.9	0	ND	Neolithic
	M13a2	165	7.6 ± 5.0	0	ND	Neolithic
M61	M61	46	16.3 ± 8.4	0	ND	LGP
M62	M62b	122	27.8 ± 10.2	0	ND	Upper Paleolithic

NOTE.—ND, T_d was not determined for either there is no haplogroup sharing or difficult to define the Han Chinese founding haplotype (see network in [supplementary fig. S4, Supplementary Material](#) online). The underlined haplotypes indicate the major haplotypes in Tibetans.

^aHaplogroup frequency in Tibetan populations.

^b T_c was estimated using the index ρ (Forster et al. 1996), which is defined as the average number of substitutions between all other sequences and the putative founding haplotype within Tibetan population (see network and founding haplotype in [fig. 3](#) and [supplementary fig. S5, Supplementary Material](#) online).

^c T_d was estimated using the index ρ (Forster et al. 1996), which is defined as the average number of substitutions between the sequences of Tibetan population and the closest founder haplotype observed in Han Chinese population, and excluded those haplotypes shared among Tibetans and Han Chinese populations (see Han Chinese founding haplotype in [supplementary fig. S4, Supplementary Material](#) online).

^dHaplogroups was divided into three climatic period groups according to their coalescence ages: Neolithic (10–4 ka), LGP (22–10 ka), and Upper Paleolithic (40–22 ka).

^e T_d for Han Chinese B5b2 from Tibetan B5b (see network in [supplementary fig. S4, Supplementary Material](#) online).

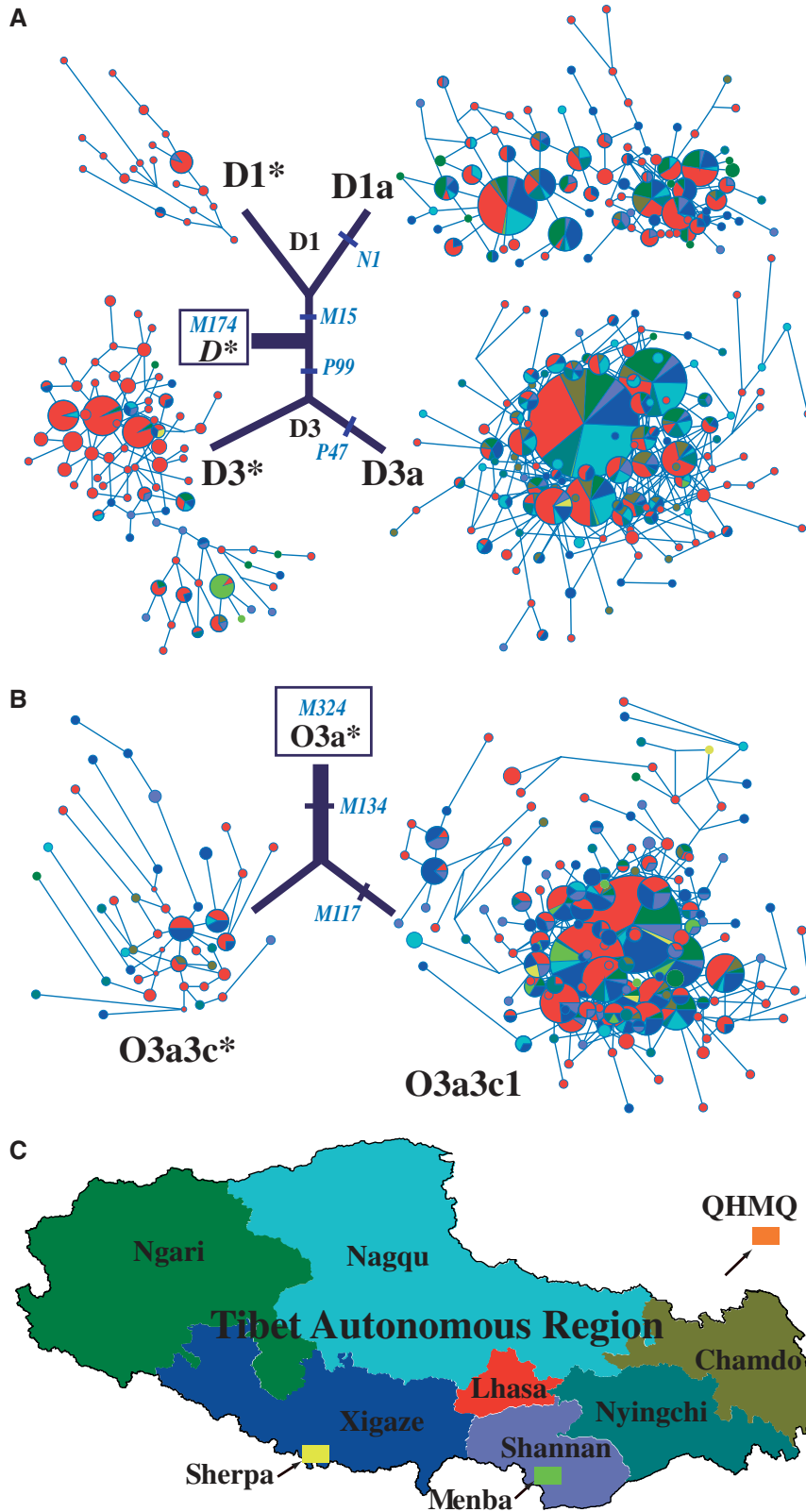


Fig. 2. Median-Joining networks for Y-chromosomal haplogroups D-M174 and O3a-M324. (A) The D-M174 and its subhaplogroups D1*-M15, D1a-N1, D3*-P99, and D3a-P47. (B) The O3a-M324 and its subhaplogroups O3a3c*-M134 and O3a3c1-M117. The diagnostic mutations used to classify the subhaplogroups are labeled on the branches. Each node represents a haplotype and its size is proportional to the haplotype frequency, and the length of a branch is proportional to the mutation steps. (C) A map color-coding the geographic regions of the Tibetan Plateau. The colored areas indicate the geographic origins of the studied populations (merged by prefectures) and the colors correspond to the node colors in the networks.

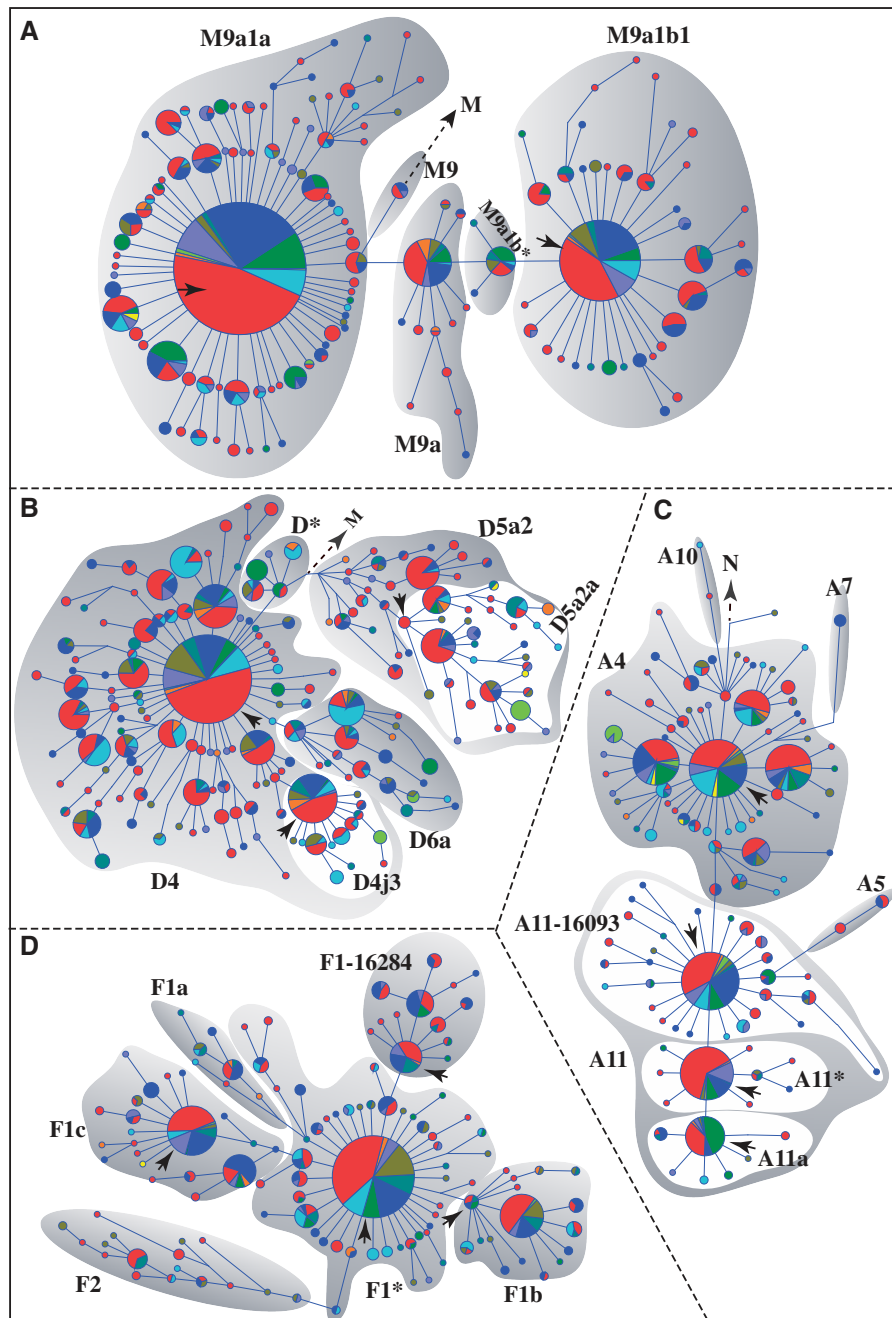


Fig. 3. Median joining networks for mtDNA haplogroups. (A) Haplogroup M9a and its subhaplogroups M9a1a and M9a1b1. (B) Haplogroup D and its subhaplogroups D4, D5, and D6. (C) Haplogroup A and its subhaplogroups A4 and A11. (D) Haplogroup F and its subhaplogroups. To simplify the network, a maximum parsimony calculation was performed to eliminate superfluous links between haplotypes with the default settings. Each node represents a haplotype and its size is proportional to the haplotype frequency, and the length of a branch is proportional to the mutation steps. The node colors correspond to those in figure 2. Those haplotypes marked by an arrow head were defined as the putative founder haplotypes used to calculate coalescence time of a particular haplogroup.

Yao, et al. 2003; Tanaka et al. 2004; Zhao et al. 2009; Qin et al. 2010), and was dated to 15.0 ± 4.6 ka in Tibetans (table 1 and fig. 3C). By contrast, A11 seemed to be a Tibetan-specific lineage (Zhao et al. 2009; Qin et al. 2010), and was dated to 13.4 ± 6.9 ka (table 1 and fig. 3C). In addition to A4 and A11, there were several rare lineages (A5, A7 and A10) with low frequencies in Tibetans (0.03–0.13%).

Haplogroup F was also prevalent in Tibetans with an overall frequency of 11.44% (supplementary fig. S3, Supplementary

Material online). This haplogroup was first defined as haplogroup A by Ballinger et al. (1992), and later renamed as haplogroup F by Torroni et al. (1994). This haplogroup likely originated in Southeast Asia, and shows high diversity in East Asian populations (Ballinger et al. 1992; Torroni et al. 1994; Yao et al. 2002; Kong, Yao, et al. 2003; Tanaka et al. 2004). It can be further divided into F1*, F1-16284, F1a, F1b, F1c and F2 in Tibetans, with coalescence ages ranging from 20–10 ka (table 1 and fig. 3D).

Besides these major haplogroups, the Tibetan populations also harbored several other lineages occurring at a relatively high frequency, including G (8.22%), M8'C/Z (7.71%), M13a (4.06%) and B (3.76%). These haplogroups were also East Asian founding lineages (Ballinger et al. 1992; Torroni et al. 1992, 1994; Torroni, Schurr, et al. 1993; Yao et al. 2002; Kong, Yao, et al. 2003; Tanaka et al. 2004). Like other maternal lineages, most of them were dated to 14–36 ka (table 1 and supplementary fig. S5, Supplementary Material online). In addition, it is worth mentioning that M62b, a Tibetan-specific basal lineage previously dated to 22.3 ka in a relative small data set (Zhao et al. 2009), was now dated to 27.7 ± 10.2 ka in our larger data set ($n = 122$) (table 1 and supplementary fig. S5, Supplementary Material online).

Data from Genome-Wide Autosomal Markers

Additional evidence from our genome-wide autosomal data also supports an ancient occupation of the Tibetan Plateau. We performed genome-wide principal component analysis (PCA) with 171,317 common SNPs in 20 populations, including our previous data of Tibetans ($n = 50$) (Peng, Yang, et al. 2011), and data of other Asian populations from HGDP (Li et al. 2008) and HapMap Phase III (The International HapMap Consortium 2003) databases. The PCAs among representative East Asian populations indicated a genetically distinct group of Tibetans, an implication of relatively long-term genetic isolation between Tibetans and other East Asian populations (fig. 4). A similar pattern of genetic relationships was also seen in PCA plots of Y-chromosomal and mtDNA data (supplementary fig. S6A and B, Supplementary Material online). The

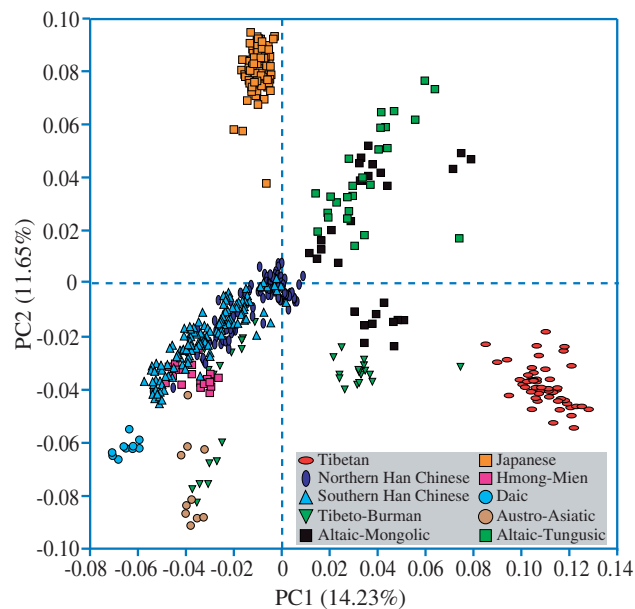


Fig. 4. Genome-wide PCA plots of Tibetans and other Asian populations calculated with 171,317 SNPs. The first two principal components (PC1 and PC2) explain 14.23% and 11.65% of the total genetic variance across the genome, respectively. The Tibetan samples comprise a genetically distinct group. Besides Tibetans, Japanese, and Han Chinese, the other samples were classified and merged by the language group as described in Material and Methods.

relatively closer genetic relationship between Tibetans and Han Chinese populations likely resulted from the admixture of Neolithic Han Chinese agriculturalists expanding into the Tibetan Plateau. Based on the estimation of F_{ST} using the genome-wide SNPs (720,064 SNPs) (Reich et al. 2009; Lukic, et al. 2011), the divergence time between Tibetans and Han Chinese was 12.6 ka, which was much younger than the estimated ages of the Tibetan-specific Y-chromosomal and mtDNA haplogroups, suggesting extensive population interactions between Paleolithic Tibetans and Neolithic Han Chinese agriculturalists.

To avoid the potential influence of linkage disequilibrium (LD) structures on PCA and age estimation, we also generated two pruned SNP sets (100,593 SNPs with R^2 cutoff < 0.5 and 124,964 SNPs with R^2 cutoff < 0.8). The pruned SNP sets showed similar PCA patterns compared with the set of all SNPs (171,317 SNPs) (supplementary fig. S6C and D, Supplementary Material online). Based on the estimation of F_{ST} , the estimated divergence times between Tibetans and Han Chinese from these two pruned SNP sets were 6,650 and 6,796 years ago, respectively, and can be considered as the lower bound of population interactions between Paleolithic Tibetans and Neolithic Han Chinese agriculturalists. We also estimated the divergence time between Tibetans and Han Chinese using a maximum likelihood method proposed by Wang and Nielsen (2012), and the result was similar (6850 ± 150 and 6400 ± 113 years ago for two SNPs sets, respectively).

Genetic Divergence within Tibetan Populations

Given the proposed early colonization and occupation of the Tibetan Plateau by modern humans, a certain degree of genetic divergence among present-day Tibetan populations should be expected, and was indeed observed in the calculated genetic distances among the Tibetan populations (fig. 5A–E). Moreover, we detected a significant correlation between the genetic distances and the geographic distances for both the Y-chromosomal and mtDNA haplogroups data ($r^2 = 0.006–0.261$, $P < 10^{-5}$, Mantel test) (fig. 5A–E and supplementary fig. S7, Supplementary Material online), in line with the proposed ancient occupation of the Tibetan Plateau, suggesting the major migratory pattern within the Plateau fits the isolation-by-distance model. For the Y-chromosomal data, when all paternal haplogroups were considered, no correlation was detected (supplementary fig. S7A, Supplementary Material online). This is likely due to the differential impacts of the two suggested prehistoric migrations on the paternal lineages of modern Tibetan populations, reflected by the two dominant Y-chromosomal haplogroups, D-M174 and O3a3c1-M117. We also tested the correlation between genetic diversity and altitude, and detected no significant relationship (supplementary fig. S8, Supplementary Material online).

To visualize the geographic patterns of gene diversity across the Tibetan Plateau, we constructed contour maps using the unbiased STR gene diversity or haplotype diversity data from Y-chromosomal STRs and mtDNA hyper variable segment I

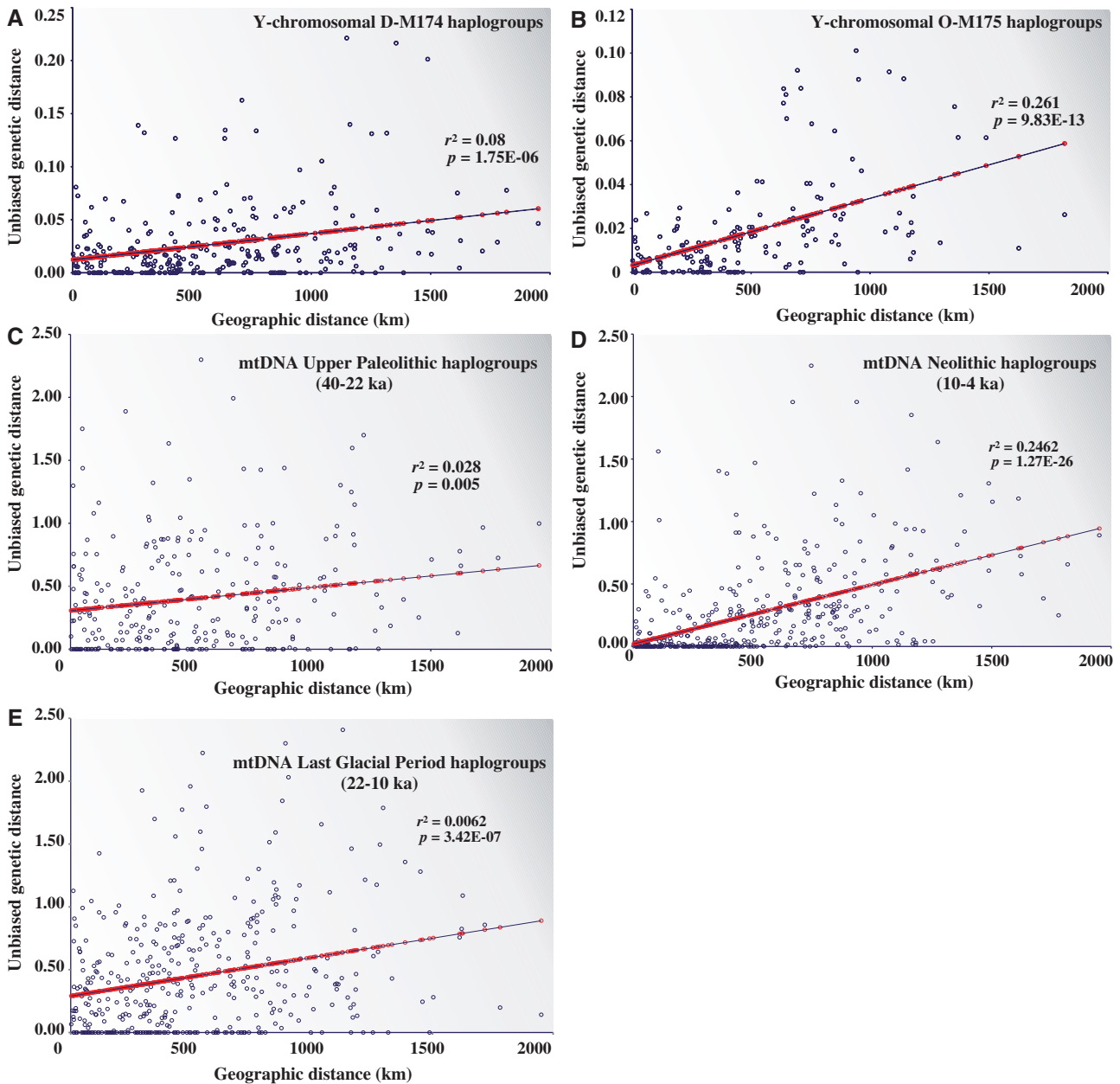


FIG. 5. Correlation between genetic distance and geographic distance across the Tibetan Plateau. Genetic distance was calculated with the Y-chromosomal STR alleles or the mtDNA HVS-I haplotypes (as one locus) for samples within haplogroups. (A) Y-chromosomal D-M174 haplogroup (24 populations). (B) Y-chromosomal O-M175 haplogroup (19 populations). (C) mtDNA haplogroups with TMRCA corresponding to upper Paleolithic (24 populations). (D) mtDNA haplogroups with TMRCA corresponding to Neolithic (29 populations). (E) mtDNA haplogroups with TMRCA corresponding to LGP (29 populations).

(HVS-I) variations. We also constructed contour maps by classifying the mtDNA haplogroups based on their coalescence ages corresponding to different climatic periods, that is, Upper Paleolithic (40–22 ka), last glacial period (LGP, 22–10 ka), and Neolithic (10–4 ka) (fig. 6A–E and supplementary fig. S7C and D, Supplementary Material online). We observed different levels of gene diversity among geographic populations. The Y-chromosomal haplogroup D (fig. 6A), the mtDNA LGP and Upper Paleolithic haplogroups (fig. 6C and E) showed a slightly higher gene diversity in the eastern regions (Chamdo and Nyingchi) and along the contemporary agricultural zones

(Lhasa, Shannan and Xigaze) (fig. 1), but statistically not significant ($P > 0.05$, Mann–Whitney U test). In contrast, the genetic diversity of the Y-chromosomal O-M175 (fig. 6B) and the mtDNA Neolithic (fig. 6D) haplogroups was higher in the eastern regions than in the western regions, though it was only marginally significant for O-M175 ($P = 0.08$, Mann–Whitney U test). This finding was concordant with archaeological evidence that supports the proposed east to west migratory route of the Neolithic population expansion coming out of what is now modern day northwestern China (Wang 1994; Aldenderfer and Zhang 2004).

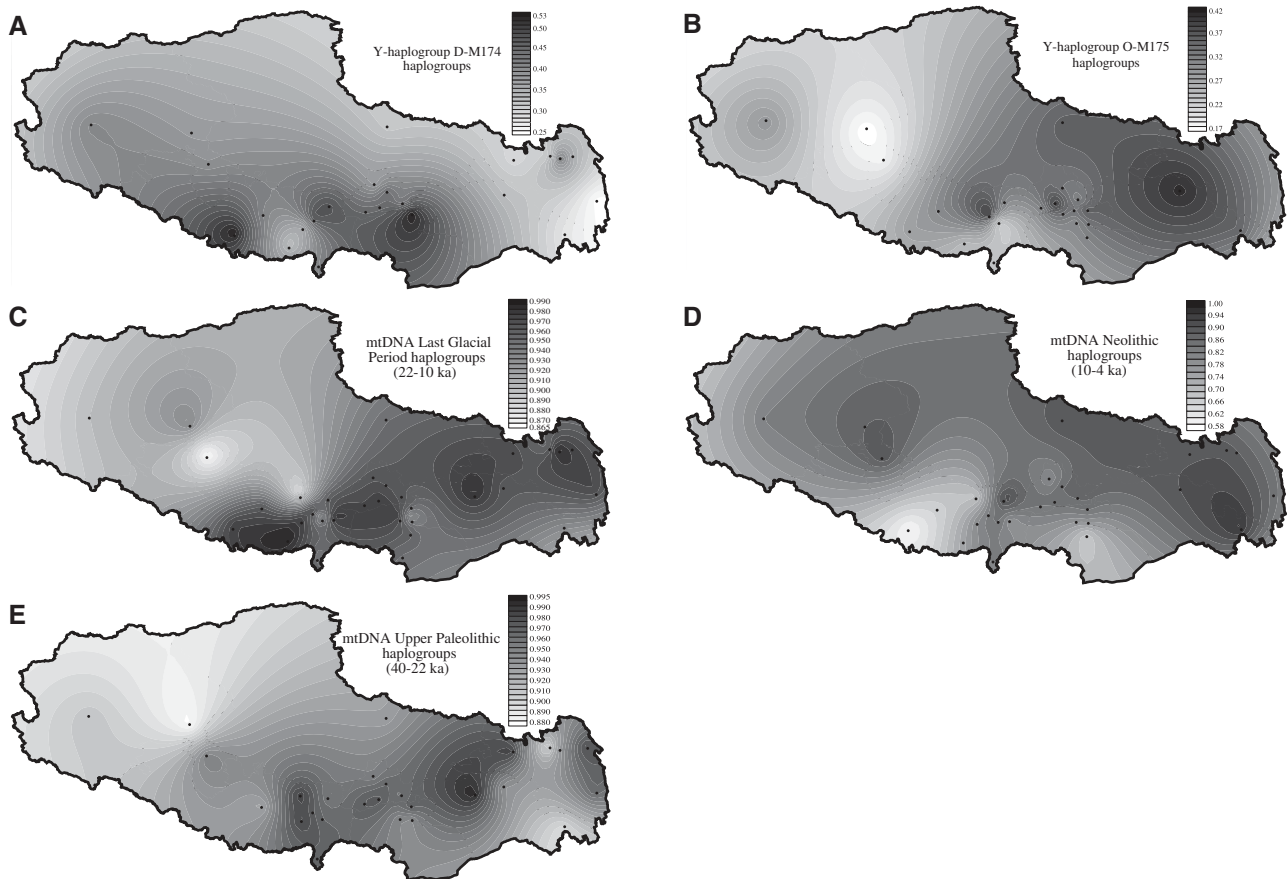


FIG. 6. Geographic patterns of genetic diversity across the Tibetan Plateau. Gene diversity was calculated with the Y-chromosomal STR alleles or the mtDNA HVS-I haplotypes (as one locus) for samples. Refer to the legend of figure 5 for the numbers of populations used in each haplogroup.

Discussion

Origins of Y-Chromosomal and mtDNA Haplogroups in Tibetans

To reconstruct the demographic history and chronology of modern human settlement on the Tibetan Plateau, we systematically screened paternal and maternal lineages in Tibetan populations. We detected two major Y-chromosomal (D-M174 and O-M175) and four major mtDNA haplogroups (M9a, D, A, and F) in Tibetan populations. The majority of both paternal (87.80%) and maternal (90.99%) haplogroups in Tibetans are the founding lineages of East Asians, and many of them are shared between Tibetans and Han Chinese (Torrioni et al. 1992, 1994; Torrioni, Schurr, et al. 1993; Kong, Yao, et al. 2003; Tanaka et al. 2004; Shi et al. 2005, 2008; Kong et al. 2006; Metspalu et al. 2006; Derenko et al. 2007; Tamm et al. 2007; Achilli et al. 2008; Soares et al. 2008; van Oven and Kayser 2009; Zhao et al. 2009; Qin et al. 2010; Kong et al. 2011; Peng, Palanichamy, et al. 2011).

We did not observe Y-chromosomal haplogroups in Tibetans coming from Central Asia and South Asia, even though the Tibetan Plateau is located in the vicinity of these two regions, suggesting that the Himalayas have served as a natural barrier for human migration, as others have previously suggested (Cordaux et al. 2004; Gayden et al. 2007). C3-M130, a major founding haplogroup in

Mongolian population (42.9%) (Zhong et al. 2010), occurred at very low frequency (47/2,354, 1.9%) in Tibetans; the majority of these lineages were found in the Lhasa region (27/47). C3-M130 also occurred at relatively high frequency in Northern Han Chinese (14.8%) and Southern Han Chinese (7.4%); therefore, it is difficult to test whether the Tibetan C3-M130 lineage was introduced by Neolithic Han Chinese or the known Mongolian invasion of Tibet in the 1200s and 1300s. As C3-M130 is very rare in Tibetans, therefore, it appears that the Mongolian invasion did not have a great impact on Tibetans.

For the mitochondrial haplogroups, we detected a low frequency (0.03–0.65%) of other haplogroups in Tibetans (U2, U7, T, M5, M20, M25, M49, M61, M70, R2, R5, and R11) (supplementary figs. S3 and S5, Supplementary Material online). These haplogroups were mainly found in western Eurasian, North Asian and/or Indian populations (Palanichamy et al. 2004; Barnabas et al. 2006; Reddy et al. 2007; Chandrasekar et al. 2009; Dulik et al. 2012; Sharma et al. 2012), but absent in other East Asian populations. The U2 (U2a, U2b and U2e) (0.65%), U7 (0.65%) and T (T1 = 0.41% and T2 = 0.21%) were previously reported in the South Asia and North Asia mtDNA pools (Palanichamy et al. 2004; Dulik et al. 2012). These rare lineages were found to sporadically distribute across the Plateau with a relatively higher incidence in Lhasa, the capital city. Haplogroup U has an origin in western

Eurasia and occurs in most Central Asian populations (Comas et al. 2004), except for U2 and U7 which were first identified in Northeastern and Central Indian tribal populations (Reddy et al. 2007; Chandrasekar et al. 2009; Sharma et al. 2012). Besides U and T, the other rare lineages were initially identified in northeastern Indian (Reddy et al. 2007; Chandrasekar et al. 2009) and central Indian tribal populations (Barnabas et al. 2006; Sharma et al. 2012). These rare maternal lineages in Tibetan populations suggested a possibility of sporadic maternal gene flow from Central and South Asia, and a complex maternal genetic landscape on the Tibetan Plateau (Aldenderfer 2011).

Paleolithic Colonization of Modern Humans on the Tibetan Plateau

The majority of the Y-chromosomal and mtDNA lineages in Tibetan populations were dated to 13–40 ka (table 1), reflecting the antiquity of the Tibetan lineages. In particular, we detected Tibetan specific Y-chromosomal (D3*-P99) and mitochondrial (M62b) lineages that were dated to 19–28 ka, a time range falling into the Upper Paleolithic period. The ancient coalescence ages of such Tibetan-specific lineages serve as strong evidence that the permanent human settlement on the Tibetan Plateau occurred during the Upper Paleolithic about 30 ka.

It is widely accepted that the mountain ranges and higher places on the Plateau were covered by ice sheets during the LGM (Shi et al. 1992; Shi 2004; Shi and Zhao 2009). Therefore, people generally believe that even though modern humans successfully settled on the Plateau in the Upper Paleolithic time, these earlier settlers might not survive the LGM, and the present-day Tibetans are descendants of post-glacial immigrants settled on the Plateau. However, our genetic data provide strong evidence that the earlier settlers could have survived the LGM and contributed to the formation of present-day Tibetan populations.

Current evidence from Paleoclimatic analyses also supports the presence of Tibetan population on the Plateau during the LGM. It was proposed that in the early Upper Paleolithic, approximately 40–30 ka, the Tibetan Plateau was relatively warm and humid with an average annual temperature of 3–4 °C higher than today (Gao et al. 2008; Shi and Zhao 2009; Hou et al. 2010). The lakes on the Plateau were freshwater, with a much higher water level than they have today (Chen and Bowler 1985). In addition, only the high places of the Tibetan Plateau was likely covered by ice sheets during the LGM (Shi et al. 1992; Shi 2004), and ice sheets were never an obstacle to long-term life on the Plateau. The Paleo-environmental context on the Plateau during the early Upper Paleolithic should have been appropriate for the earliest human settlement before the LGM, and the ice-free regions—likely the relatively low altitude areas—could easily have served as the refugia during the LGM.

Evidence from archaeological observations also supports our genetic findings. Although the majority of the archaeological sites dated to the Upper Paleolithic are located at relatively low altitude (<3,000 m) (Aldenderfer and Zhang

2004; Brantingham et al. 2007; Aldenderfer 2011), modern humans probably had occupied the high altitude Plateau during the early Upper Paleolithic. Siling Co., located on the northern Tibetan Plateau with an elevation of 4,600 m, is a Paleolithic site dated to 40–30 ka, which has yielded paleoliths that show technological and typological affinities with the European Middle Paleolithic. The findings at Siling Co. suggest that modern humans entered into the Plateau during the early Upper Paleolithic, in line with the proposed migratory waves of modern humans across the Old World during the Late Pleistocene (Cavalli-Sforza and Feldman 2003; Underhill and Kivisild 2007; Yuan et al. 2007). The other Paleolithic finding came from the hand- and foot-prints and the remnant of a fireplace found on a hot spring travertine dated to 21 ka in the Central Plateau at 4,200 m, supporting the idea that ancestors of modern Tibetans did indeed survive the LGM (22–18 ka) (Zhang and Li 2002).

Given the contexts of the inhospitable high-altitude hypoxic environment of the Plateau, the initial immigrants from the low altitude area must have encountered strong natural selection against hypoxic stress upon their entry into the Plateau. The evolutionary history of genes responsible for adaptation to high-altitude hypoxia in modern Tibetans should then also reflect the time of initial settlement. Recently, several genome-wide studies have suggested *EPAS1* as one of the key genes harboring adaptive sequence changes for high-altitude hypoxia (Beall et al. 2010; Bigham et al. 2010; Simonson et al. 2010; Yi et al. 2010; Peng, Yang, et al. 2011; Xu et al. 2011). The onset of natural selection on the *EPAS1* gene was estimated to be 18 ka (Peng, Yang, et al. 2011), supporting the possibility that continuous human settlement on the Plateau occurred during the Upper Paleolithic following the end of LGM (22–18 ka) around 18 ka on the Tibetan Plateau. Collectively, the genetic, paleoclimatic and archaeological data all support modern humans' permanent settlement of the Tibetan Plateau beginning from the early Upper Paleolithic, and since then the ancestors of modern Tibetan populations have developed genetic adaptations to the high-altitude environment.

Neolithic Expansion of Modern Humans on the Tibetan Plateau

The Tibetans' successful adaptation to the high-altitude is not only reflected by their long history of living on the Plateau, but also by the relatively large population size of over 5 million indigenous Tibetans currently residing there. We identified a molecular signature of recent population expansion during the early Neolithic time in both paternal (Y-chromosomal D3a-P47 and O3a3c1-M117) and maternal (M9a1a and M9a1b1) lineages (10–7 ka) (table 1). The detailed analysis of haplotype sharing and time of divergence between Tibetans and Han Chinese suggests that the Neolithic population expansion on the Plateau was likely caused by the dispersal of the earliest Neolithic Han Chinese agriculturalists originating about 10 ka in what is now northwestern China, reflected by the cultural artifacts of the early Neolithic in China's Gansu and Qinghai provinces, the two neighboring

areas close to the Plateau (Wang 1994; Barton et al. 2009; Bettinger et al. 2010; Yang et al. 2012).

The earliest Neolithic cultures in China are known to have started at the Upper Yellow River basin of northwestern China about 10 ka, with agriculture emerging and expanding since then (Wang 1994; Barton et al. 2009; Bettinger et al. 2010; Yang et al. 2012). Chinese archeologists have proposed that the Di-Qiang people, or Proto-Tibeto-Burman populations, already inhabited the Upper Yellow River area during the early Neolithic (Barton et al. 2009; Bettinger et al. 2010; Yang et al. 2012), and later migrated into the Tibetan Plateau during the Yang-Shao epoch (about 8,000 years ago) bringing agriculture to the Himalayan region (Wang 1994). Additionally, there might have been more than one expansion, or a series of movements, from east to west, rather than a singular major Neolithic migration from Northwestern China into the region (Shelach 2000; Rhode et al. 2007).

We observed rapid population expansion in the majority of Y-chromosomal (D3a-P47 and O3a3c1-M117) and mitochondrial haplogroups (M9a1a, M9a1b1, A4, and A11) (figs. 2 and 3). The result of Bayesian skyline plots indicated a 2- to 4-fold increase in population growth rate during the Neolithic transition around 8,000 years ago in Tibetans (supplementary table S3 and fig. S9, Supplementary Material online). By contrast, there was only a mildly accelerated population growth detected in Han Chinese starting about 18,000 years ago, and the curve for Tibeto-Burman was relatively flat (supplementary fig. S9, Supplementary Material online).

The major mitochondrial haplogroups in Tibetans can be divided into three climatic period groups according to their coalescence ages: Neolithic (10–4 ka), LGP (22–10 ka), and Upper Paleolithic (40–22 ka) (table 1). The BSP analysis suggested a rapid increase in population size starting approximately 6,000 years ago for the Neolithic haplogroups (supplementary table S3 and fig. S10, Supplementary Material online). For the LGP haplogroups, population started to increase approximately 16 ka, whereas the Upper Paleolithic haplogroups showed a relatively flat growth curve.

It has been shown in European populations that the introduction of agricultural or an agro-pastoral economy facilitated a rapid increase in population growth relative to more ancient expansions of hunter-gatherers due to the climate improvement following the end of the LGM (Gignoux et al. 2011; Zheng et al. 2011, 2012). We observed similar scenario in Tibetans showing a rapid population growth during the Neolithic. The remaining question is whether the rapid Neolithic population growth in Tibetans was caused by the expansion of the local Paleolithic autochthons or there were immigrants from outside the Plateau.

We showed that the Y-chromosomal O3a3c1-M117 originated outside of Tibet, and was brought to the Tibetan Plateau by the Neolithic Han Chinese agriculturalists. The Neolithic immigrants were mixed with the earlier Paleolithic autochthons carrying the D-M174 lineage. The introduction of the Neolithic technologies into the Plateau eventually led to the establishment of farming and yak pastoralism on the Plateau, for example, barley cultivation (Dai et al. 2012) and yak domestication (Guo et al. 2006; Wang et al. 2010; Wang,

Yonezawa, et al. 2011), which in turn served to provide a stable livelihood and productive resources needed for the subsequent rapid population growth. Both barley and yak were domesticated locally in the Tibetan Plateau, suggesting that Tibet was a key center of Neolithic cultural change and not simply a recipient of agricultural technology brought by peoples entering from different regions of East Asia.

Considering the prevalence of the paternal lineage O3a3c1-M117 in Tibetans (29.82%), the contribution of the Neolithic Han Chinese agriculturalists to the formation of modern Tibetan populations was large, although the major haplogroup profiles in modern Tibetan populations were probably established prior to the entry of Neolithic agriculturalists into the Plateau. This is also reflected by the relatively old coalescence age of O3a3c1-M117 within Tibetans (18.4 ± 7.4 ka) (table 1), which is older than the estimated divergence time (11.5 ± 4.8 ka) of O3a3c1-M117 between Tibetans and Han Chinese. This is due to a relatively large immigrant population of the Neolithic Han Chinese agriculturalists into the Plateau, resulting in an estimated coalescence time predating the migratory event. Hence, both demic and cultural diffusions might have occurred during the transition of the Neolithic agricultural economy on the Plateau approximately 12.6–6.6 ka, the estimated time range of admixture between Paleolithic Tibetans and Neolithic Han Chinese agriculturalists based on the genome-wide SNP markers.

In summary, genetic data from the Y-chromosome, mtDNA and autosomes all support the hypothesis that modern humans permanently settled on the Tibetan Plateau around the Upper Paleolithic, as early as 30 ka, followed by rapid population expansion and gene flow from outside the region beginning in the early Neolithic, 10–7 ka. Both the Paleolithic migration and Neolithic expansion had a significant impact on the genetic makeup of present-day Tibetan populations. The proposed ancient peopling of the Plateau also suggests a long lasting act of natural selection on Tibetan populations and explains why Tibetans have developed superior adaptation to high-altitude environments, compared with the other high-altitude populations (e.g., Andeans) who have only a relatively short history of inhabitation (Beall 2007). Hence, Tibetans are an ideal population for further research searching for genes and understanding the molecular mechanism of genetic adaptation to high-altitude hypoxia.

Materials and Methods

Samples

We performed a detailed questionnaire for all volunteers about their parents' and grand-parents' origins, and only those volunteers whose parents and grandparents both are Tibetans will be considered as Tibetans and used in this study. Informed written consents were obtained from all volunteers. The sampling was conducted randomly and biologically related individuals were excluded from this study. Two to five milliliters of venous blood samples were collected from 6,109 unrelated volunteers from 41 geographical locations across

the entire region of the Tibetan Plateau, including the Tibetan Autonomous Region and Qinghai Province. Of these samples, a total of 5,893 Tibetan individuals from 38 populations were taken from the Tibetan Autonomous Region, and one Tibetan population (QHMQ, $n = 130$) from Maqin County, Guoluo Tibetan Autonomous Prefecture in Qinghai Province, which is also in the main part of the Tibetan Plateau. The other two populations, Menba ($n = 57$) and Sherpa ($n = 29$), were minority ethnic groups closely related to local Tibetans. The sampling locations and population details were given in [figure 1](#). The protocol of this study was approved by the Institutional Review Board of the Kunming Institute of Zoology, Chinese Academy of Sciences.

Y-Chromosomal Genotyping

The Y-chromosomal haplogroups were classified based on the up-to-date, high-resolution Y-chromosomal phylogenetic tree (Karafet et al. 2008; Underhill et al. 2010) following procedures described in our previous studies (Shi et al. 2005, 2008; Zhong et al. 2011). To evaluate the genetic diversity of the major Y-chromosomal haplogroup in Tibetan populations, eight commonly used Y-chromosomal STR (short tandem repeat) loci (DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393) were genotyped using fluorescence-labeled primers with an ABI 3130XL Genetic Analyzer (Applied Biosystems, USA). The allelic nomenclature of Y-chromosomal STRs was based on the system proposed by Butler et al. (2002).

Mitochondrial DNA Genotyping

The HVS-I of the mitochondrial control region from positions 16024–16466 relative to the revised Cambridge Reference Sequence (rCRS) (Andrews et al. 1999) was amplified with primers L15974 and H16498, following the method described in Yao et al. (2002). The polymerase chain reaction (PCR) products were purified with the resin method, and sequenced with primer L15974 using the BigDye Terminator Cycle Sequencing kit v3.1 (Applied Biosystems, USA) and an ABI 3730XL Genetic Analyzer (Applied Biosystems, USA). Those samples with a 12-bp poly-C stretch were further sequenced in the reverse direction with primer H16498. To further confirm the haplogroup classification, the HVS-II sequence from positions 56-304/412 relative to rCRS were amplified and sequenced for part of the samples ($n = 2,106$) with primers L29 and H408. The mtDNA haplogroups were classified using the HVS-I and/or HVS-II variable sites and at least one additional coding region diagnostic SNP sites according to the up-to-date mtDNA phylogenetic tree (Phylotree.org, build 12, July 12, 2011) (van Oven and Kayser 2009). The coding region diagnostic SNP sites were genotyped by direct sequencing or following a multiplex SNaPshot assay consisting of 21 coding region diagnostic SNP sites for East Asians (Qin et al. 2010).

Y-Chromosomal Data Analysis

Median-joining networks for STR variations of the Y-chromosomal haplogroups were constructed using NETWORK 4.6

(Fluxus Engineering) (Bandelt et al. 1999) with equal weights across all loci. The age of STR variation within each Y-chromosomal haplogroup was estimated following the published method (Zhitovovsky 2001; Zhitovovsky et al. 2004; Sengupta et al. 2006). During the analysis of Y-chromosomal STR alleles, DYS389II was named DYS389b after subtracting DYS389I because the PCR product of DYS389II contains both DYS389II and DYS389I loci. The unbiased STR gene diversity of the Y-chromosomal haplogroup was calculated according to method described by Nei (1987) with GenAlEx (version 6.3) (Peakall and Smouse 2006).

Mitochondrial DNA Data Analysis

The HVS-I and HVS-II sequences were read and checked using DNASTAR v7.1 (DNASTAR Inc.), and aligned with ClustalX (version 2.0) (Larkin et al. 2007). The HVS-I and HVS-II variable sites were extracted with DnaSP (version 5.0) (Librado and Rozas 2009). Median-joining networks were constructed using NETWORK 4.6 (Fluxus Engineering) (Bandelt et al. 1999) based on the HVS-I sequence variations. To simplify the network, a maximum parsimony calculation was performed to eliminate superfluous links between haplotypes with the default settings. All sites with an indel (insertions and deletions) as well as the variable sites at 16181, 16182, and 16183 were removed for network construction and other analyses according to the suggestions by van Oven and Kayser (2009). The coalescence times were estimated with the ρ statistics (Forster et al. 1996) using HVS-I sequence variations from positions 16024–16466, and the standard errors were calculated using Saillard et al.'s method (Saillard et al. 2000). We used an improved mutation rate of one mutation every 16,677 years for the HVS-I sequences (Soares et al. 2009) to estimate the time to the most recent common ancestor (TMRCA) of a haplogroup. The unbiased HVS-I haplotype diversity and its sampling variance was calculated with DnaSP (version 5.0) (Librado and Rozas 2009) following method described by Nei (1987).

Genome-Wide SNP Data Analysis

The genome-wide SNP data were retrieved from three data sets including 508 samples from 20 populations: 1) one Tibetan population ($n = 50$) was obtained from our previous study (Peng, Yang, et al. 2011), which was genotyped on the Affymetrix SNP Array 6.0 platform; 2) 16 populations from the HGDP panel genotyped on the Illumina 650Y platform (Li et al. 2008), which include Cambodian ($n = 10$), Dai ($n = 10$), Hezhen ($n = 8$), Japanese ($n = 28$), Miao ($n = 10$), Naxi ($n = 8$), Oroqen ($n = 9$), She ($n = 10$), Tu ($n = 10$), Tujia ($n = 10$), Xibo ($n = 9$), Han Chinese ($n = 44$), Daur ($n = 9$), Lahu ($n = 8$), Mongola ($n = 10$), Yi ($n = 10$); and 3) the HapMap Phase III samples including Han Chinese from Beijing (CHB) ($n = 84$), Han Chinese from Metropolitan Denver, Colorado (CHD) ($n = 85$), and Japanese from Tokyo (JPT) ($n = 86$) (The International HapMap Consortium 2003). All three data sets were merged and the resulting set of 171,317 common SNPs was used in our analysis. The median of physical spacing between these SNPs are 8,077 bp. The LD between SNPs was

estimated and the mean of pairwise R^2 is 0.570 (SD = 0.270) and median is 0.516.

To investigate the effect of correlation of SNPs on genome-wide PCA and F_{ST} estimation, we also pruned SNPs based on pairwise LD. A sliding window with the size of 30 SNPs is used, and the number of 5 SNPs was used to shift the window at each step. One of a pair of SNPs with $R^2 > R^2$ cutoff is dropped. By repeating the procedures, we can generate a subset of SNPs with the pairwise R^2 lower than R^2 cutoff. We prepared two sets of pruned SNPs, one with R^2 cutoff = 0.5, including 100,593 SNPs; the other with R^2 cutoff = 0.8, including 124,964 SNPs.

We also used a maximum likelihood method proposed by Wang and Nielsen (2012) to estimate the divergence time between Tibetans and Han Chinese. As we did not have resequencing data to estimate the effective population size or the genetic diversity of the Tibetan population, we only used the estimated divergence time of Han Chinese from Tibetans (instead of the opposite) and converted it into years. Similar to the estimation from F_{ST} , we also chose $N_e = 10,000$ for the Han Chinese population, and 25 years for a generation. The 95% confidence interval was estimated by bootstrapping the data for 200 times.

Principal Component Analysis

We examined the genetic structure among Tibetan and other Asian populations with the genome-wide autosomal SNP, Y-chromosomal and mtDNA haplogroup frequency data by using PCA (Patterson et al. 2006). We performed genome-wide PCA with all 171,317 common SNPs and two sets of pruned SNPs to rule out the LD effect on PCA, one with R^2 cutoff = 0.5, including 100,593 SNPs; the other with R^2 cutoff = 0.8, including 124,964 SNPs. The Y-chromosomal and mtDNA PCAs were performed with MVSP 3.13n (Kovach Computing Services, Anglesey, UK). The haplogroup frequency data for other Asian populations were retrieved from the previous studies for Y-chromosome (Hammer et al. 2006; Sengupta et al. 2006; Zhong et al. 2011) and mtDNA (Qian et al. 2001; Yao et al. 2002; Yao and Zhang 2002; Kong, Yao, et al. 2003; Wen, Li, et al. 2004; Wen, Xie, et al. 2004; Wen et al. 2005; Hill et al. 2006, 2007; Derenko et al. 2007; Li et al. 2007). The populations with a sample size of less than 10 individuals were excluded. Ultimately, PCA analysis included 34 Tibetan populations (this study) and 75 other Asian populations for Y-chromosome, and 55 Tibetan populations (41 from this study and 14 from literatures [Zhao et al. 2009; Qin et al. 2010]) and 69 other Asian populations for the mtDNA.

Besides Tibetans, Japanese and Han Chinese, the other samples were classified and merged by the language group. The grouping details for autosomal SNP data are as following:

- Tibetan: Tibetan
- Northern Han Chinese: Han Chinese (HGDP), CHB (HapMap)
- Southern Han Chinese: Han Chinese (HGDP), CHD (HapMap)
- Tibeto-Burman: Tujia, Yi, Naxi, and Lahu

- Altaic-Mongolic: Mongola, Daur, and Tu
- Altaic-Tungusic: Oroqen, Hezhen, and Xibo
- Japanese: JPT (HapMap)
- Hmong-Mien: Miao and She
- Daic: Dai
- Austro-Asiatic: Cambodian

Estimation of Divergence Time between Tibetan and Han Chinese Population

Genome-Wide SNP Data

An ideal choice to estimate population divergence time is to use likelihood methods, modeling the joint allele frequency spectrum of the two populations (Gutenkunst et al. 2009; Reich et al. 2009; Lukic et al. 2011; Chen 2012). But these approaches can only work for re-sequencing data and are unfit for SNP arrays since the SNPs are ascertained with the biased complex procedures. Here, we use a very simple approach based on the estimation of F_{ST} (Cavalli-Sforza 1969): 1) The F_{ST} is estimated with genome-wide SNPs and two sets of pruned SNPs to rule out the LD effect on F_{ST} estimation (100,593 SNPs with R^2 cutoff = 0.5, and 124,964 SNPs with R^2 cutoff = 0.8). According to the Cavalli-Sforza method (Cavalli-Sforza 1969), the divergence time is $\hat{T} = -\log(1 - \hat{F}_{ST})$ and \hat{T} is in units of $2N$ generations. As the effective population sizes for Tibetan and Han Chinese are unknown, a reasonable choice is $N = 10,000$, based on former studies on $4N\mu$ (Voight et al. 2005; Noonan et al. 2006). To further convert generations to years, we used 25 years for a generation; 2) We used a jackknife resampling procedure to get the confidence interval (see Reich et al. 2009 for details), which can account for the correlations among nearby SNP loci. The whole genome was divided into consecutive 5 Mb blocks, and in each resampling, one block was excluded from the analysis.

Mitochondrial DNA Data

For those Tibetan haplogroups that shared between Tibetans and Han Chinese populations, and with a sample size of over 20 individuals, a divergence time between Tibetans and Han Chinese was estimated using the HVS-I sequence variations (16024–16383) using the ρ distance (Forster et al. 1996), which was defined as the average number of substitutions between sequences of Tibetan haplogroups and the closest founder haplotype observed in Han Chinese population (see putative founding haplotype in [supplementary fig. S4, Supplementary Material](#) online), and excluded those haplotypes shared among Tibetans and Han Chinese populations.

Y Chromosomal Data

The divergence time was estimated using the Y-chromosomal STR data using a T_d statistic described elsewhere (Zhivotovsky 2001; Zhivotovsky et al. 2004).

Bayesian Skyline Plot

To reconstruct the demographic changes through time for Tibetan, Tibeto-Burman, and Han Chinese populations, as well as three haplogroups within Tibetans with TMRCA corresponding to different climatic periods, that is, Neolithic (10–4 ka), LGP (22–10 ka) and Upper Paleolithic (40–22 ka)

(table 1), we reconstructed Bayesian Skyline Plot (BSP) (Drummond et al. 2005) in BEAST (version 1.7) (Drummond et al. 2012) with Markov chain Monte Carlo (MCMC) algorithms (Drummond et al. 2002). The BSPs were generated using HVS-I sequences from 16,024 to 16,466 for Tibetan population, and 16,024 to 16,383 for Han Chinese ($n = 973$) (Yao et al. 2002; Wen, Li, et al. 2004) and Tibeto-Burman ($n = 496$) (Wen, Xie, et al. 2004) populations retrieved from the literature. For the Tibetan population, we randomly sampled 1,000 individuals from 6,109 Tibetan individuals for all haplogroups and 500 individuals for each of the haplogroups with TMRCAs corresponding to three climatic periods including Neolithic ($n = 1,925$), LGP ($n = 2,503$) and Upper Paleolithic ($n = 1,461$), and repeated three rounds of independent sampling and BSPs. The best-fit model was selected with Modeltest (version 3.7) (Posada and Buckley 2004). A strict molecular clock with the fixed rate as 1.784×10^{-7} substitutions per site per year (Soares et al. 2009) was applied. The MCMC chain was run for 5×10^7 steps, with sampling of parameters every 5,000 steps, and the initial 5×10^6 steps were discarded as burn-in. In all runs, the effective sample size for parameters of interest was over 200. The BSP was visualized with Tracer 1.5 (<http://tree.bio.ed.ac.uk/software/tracer>), and the female effective population size was plotted on a log scale and assumed a female generation time of 20 years. Population growth rate and time of population growth were calculated from skyline plots using method described elsewhere (Gignoux et al. 2011).

Calculation of Geographic Distance among Tibetan Populations

The geographic coordinates and altitude data for a population in a particular county were roughly taken from locations where the County Administration Centers were located. Map distance was used rather than the real geographic distance, using geographic coordinates to represent the geographic distance among populations. The calculation was performed in GenAEx (version 6.3) (Peakall and Smouse 2006).

Genetic Isolation by Geographic Distance among Populations

The unbiased Nei's genetic distance (Nei 1987) was calculated among populations for all samples or haplogroups with TMRCAs corresponding to three climatic periods (Neolithic, LGP and Upper Paleolithic haplogroups) using GenAEx (version 6.3) (Peakall and Smouse 2006). The Y-chromosomal STR loci or mtDNA haplotype (as one locus) was used to calculate gene diversity and genetic distance. The Mantel-test was performed with GenAEx (version 6.3) (Peakall and Smouse 2006) to test correlation between genetic distance and geographic distance among Tibetan populations. Populations with a sample size of less than 10 individuals, and QHMQ (few individuals were genotyped for the Y chromosomal markers), Menba and Sherpa were removed from this analysis. We used a variety of populations: Y-chromosomal all samples (33 populations), Y-chromosomal haplogroup D-M174 (24

populations), Y-chromosomal haplogroup O-M175 (19 populations), mtDNA all samples (36 populations), and mtDNA haplogroups with TMRCAs corresponding to three climatic periods, that is, Neolithic (29 populations), LGP (29 populations), and Upper Paleolithic (24 populations).

Construction of Contour Map

To visualize the geographic distributions of genetic diversity across the Tibetan Plateau, the unbiased gene diversity (Y-chromosomal STR data) and haplotype diversity (mtDNA HVS-I data) was illustrated geographically using a contour map, constructed using Golden Software Surfer 10.0 (Golden Software Inc., USA) with the Kriging algorithm.

Data Deposit

The GenBank accession numbers for the 6,109 HVS-I and 2,106 HVS-II sequences reported in this paper are JQ630032–JQ636140 and JQ636141–JQ638268, respectively.

Supplementary Material

Supplementary figures S1–S10, tables S1–S3 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

The authors thank the volunteers who donated samples of their blood for this study. They thank all participating staff from High Altitude Medical Research Center, School of Medicine, and Tibetan University for their assistance during sample collection. They thank Prof. Darren Curnoe at the University of New South Wales for critical review of the early version of this manuscript. They also thank Andrew Willden at the Kunming Institute of Zoology for language editing, and Dr Yong Wang for kindly sharing with us their code for estimating population divergence time and helpful discussion. This work was supported by the National 973 Program of China grants 2012CB518202 to T.W. and X.Q., 2011CB512107 to Ou.; the National Natural Science Foundation of China grants 91231203 and 31123005 to B.S., 30870295 to X.Q., and 91131001 to H.S.; State Key Laboratory of Genetic Resources and Evolution grants GREKF11-02 to Ou., GREKF10-02 to C.C., and GREKF13-04 to T.W.; the Ministry of Education of China grant 707051 to Ou.; and the Natural Science Foundation of Yunnan Province grant 2009CD107 and 2010CI044 to H.S.

References

- Achilli A, Perego UA, Bravi CM, Coble MD, Kong Q-P, Woodward SR, Salas A, Torroni A, Bandelt H-J. 2008. The phylogeny of the four Pan-American mtDNA haplogroups: implications for evolutionary and disease studies. *PLoS One* 3:e1764.
- Aldenderfer M. 2011. Peopling the Tibetan Plateau: insights from archaeology. *High Alt Med Biol.* 12:141–147.
- Aldenderfer M, Zhang Y. 2004. The prehistory of the Tibetan Plateau to the seventh century A.D.: perspectives and research from China and the West Since 1950. *J World Prehist.* 18:1–55.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet.* 23:147.

- Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen KH, Wallace DC. 1992. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. *Genetics* 130:139–152.
- Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* 16:37–48.
- Barnabas S, Shouche Y, Suresh CG. 2006. High-resolution mtDNA studies of the Indian population: implications for palaeolithic settlement of the Indian subcontinent. *Ann Hum Genet.* 70:42–58.
- Barton L, Newsome SD, Chen FH, Wang H, Guilderson TP, Bettinger RL. 2009. Agricultural origins and the isotopic identity of domestication in northern China. *Proc Natl Acad Sci U S A.* 106: 5523–5528.
- Beall CM. 2007. Two routes to functional adaptation: Tibetan and Andean high-altitude natives. *Proc Natl Acad Sci U S A.* 104: 8655–8660.
- Beall CM. 2011. Genetic changes in Tibet. *High Alt Med Biol.* 12:101–102.
- Beall CM, Cavalleri GL, Deng L, et al. (29 co-authors). 2010. Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders. *Proc Natl Acad Sci U S A.* 107:11459–11464.
- Behar Doron M, van Oven M, Rosset S, Metspalu M, Loogväli E-L, Silva Nuno M, Kivisild T, Torroni A, Villems R. 2012. A “Copernican” reassessment of the Human mitochondrial DNA tree from its root. *Am J Hum Genet.* 90:675–684.
- Bettinger RL, Barton L, Morgan C. 2010. The origins of food production in north China: a different kind of agricultural revolution. *Evol Anthropol: Issues News Rev.* 19:9–21.
- Bigham A, Bauchet M, Pinto D, et al. (14 co-authors). 2010. Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genet.* 6:e1001116.
- Brantingham PJ, Gao X. 2006. Peopling of the northern Tibetan Plateau. *World Archaeol.* 38:387–414.
- Brantingham PJ, Gao X, Olsen JW, Ma HZ, Rhode D, Zhang HY, Madsen DB. 2007. A short chronology for the peopling of the Tibetan Plateau. In: Madsen DB, Chen F, Gao X, editors. Late quaternary climate change and human adaptation in arid China. Amsterdam (The Netherlands): Elsevier. p. 129–150.
- Butler JM, Schoske R, Vallone PM, Kline MC, Redd AJ, Hammer MF. 2002. A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Sci Int.* 129:10–24.
- Cai X, Qin Z, Wen B, et al. (12 co-authors). 2011. Human migration through bottlenecks from Southeast Asia into East Asia during last glacial maximum revealed by Y chromosomes. *PLoS One* 6: e24282.
- Cavalli-Sforza LL. 1969. Human diversity. In: Proceedings of 12th International Congress Genetics: Tokyo, Japan.
- Cavalli-Sforza LL, Feldman MW. 2003. The application of molecular genetic approaches to the study of human evolution. *Nat Genet.* 33(Suppl):266–275.
- Chandrasekar A, Kumar S, Sreenath J, et al. (21 co-authors). 2009. Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. *PLoS One* 4:e7447.
- Chen H. 2012. The joint allele frequency spectrum of multiple populations: a coalescent theory approach. *Theor Popul Biol.* 81: 179–195.
- Chen K-Z, Bowler JM. 1985. Preliminary study on sedimentary characteristics and evolution of palaeoclimate of Qarhan salt lake in Qaidam Basin. *Scientia Sinica (Series B).* 11:1218–1231 (in Chinese).
- Comas D, Plaza S, Wells RS, Yuldaseva N, Lao O, Calafell F, Bertranpetit J. 2004. Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur J Hum Genet.* 12:495–504.
- Cordaux R, Weiss G, Saha N, Stoneking M. 2004. The northeast Indian passageway: a barrier or corridor for human migrations? *Mol Biol Evol.* 21:1525–1533.
- Dai F, Nevo E, Wu D, Comadran J, Zhou M, Qiu L, Chen Z, Beiles A, Chen G, Zhang G. 2012. Tibet is one of the centers of domestication of cultivated barley. *Proc Natl Acad Sci U S A.* 109:16969–16973.
- Derenko M, Malyarchuk B, Grzybowski T, et al. (12-co-authors). 2007. Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *Am J Hum Genet.* 81:1025–1041.
- Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161: 1307–1320.
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol.* 22:1185–1192.
- Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol.* 29: 1969–1973.
- Dulik MC, Zhadanov SI, Osipova LP, Askapuli A, Gau L, Gokcumen O, Rubinstein S, Schurr TG. 2012. Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. *Am J Hum Genet.* 90:229–246.
- Forster P, Harding R, Torroni A, Bandelt HJ. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet.* 59:935–945.
- Gao X, Zhou Z-Y, Guan Y. 2008. Human cultural remains and adaptation strategies in the Tibetan Plateau margin region in the late Pleistocene. *Quaternary Sci.* 28:969–977 (in Chinese).
- Gayden T, Cadenas AM, Regueiro M, Singh NB, Zhivotovskiy LA, Underhill PA, Cavalli-Sforza LL, Herrera RJ. 2007. The Himalayas as a directional barrier to gene flow. *Am J Hum Genet.* 80:884–894.
- Gignoux CR, Henn BM, Mountain JL. 2011. Rapid, global demographic expansions after the origins of agriculture. *Proc Natl Acad Sci U S A.* 108:6044–6049.
- Guo S, Savolainen P, Su J, Zhang Q, Qi D, Zhou J, Zhong Y, Zhao X, Liu J. 2006. Origin of mitochondrial DNA diversity of domestic yaks. *BMC Evol Biol.* 6:73.
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 5:e1000695.
- Hammer MF, Karafet TM, Park H, Omoto K, Harihara S, Stoneking M, Horai S. 2006. Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. *J Hum Genet.* 51: 47–58.
- Hill C, Soares P, Mormina M, et al. (12 co-authors). 2006. Phylogeography and ethnogenesis of aboriginal Southeast Asians. *Mol Biol Evol.* 23:2480–2491.
- Hill C, Soares P, Mormina M, et al. (11-co-authors). 2007. A mitochondrial stratigraphy for island southeast Asia. *Am J Hum Genet.* 80: 29–43.
- Hou G-L, Xu C-J, Fan Q-S. 2010. Three expansions of prehistoric humans towards northeast margin of Qinghai-Tibet Plateau and environmental change. *Acta Geographica Sinica.* 65:65–71 (in Chinese).
- International HapMap Consortium. 2003. The International HapMap Project. *Nature* 426:789–796.
- Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* 18:830–838.
- Kong QP, Bandelt HJ, Sun C, et al. (12-co-authors). 2006. Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. *Hum Mol Genet.* 15:2076–2086.
- Kong QP, Sun C, Wang HW, et al. (16 co-authors). 2011. Large-scale mtDNA screening reveals a surprising matrilineal complexity in east Asia and its implications to the peopling of the region. *Mol Biol Evol.* 28:513–522.
- Kong QP, Yao YG, Liu M, Shen SP, Chen C, Zhu CL, Palanichamy MG, Zhang YP. 2003. Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. *Hum Genet.* 113: 391–405.

- Kong QP, Yao YG, Sun C, Bandelt HJ, Zhu CL, Zhang YP. 2003. Phylogeny of east Asian mitochondrial DNA lineages inferred from complete sequences. *Am J Hum Genet.* 73:671–676.
- Larkin MA, Blackshields G, Brown NP, et al. (13 co-authors). 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948.
- Li H, Cai X, Winograd-Cort ER, et al. (12 co-authors). 2007. Mitochondrial DNA diversity and population differentiation in southern East Asia. *Am J Phys Anthropol.* 134:481–488.
- Li JZ, Absher DM, Tang H, et al. (11 co-authors). 2008. Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319:1100–1104.
- Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
- Lukic S, Hey J, Chen K. 2011. Non-equilibrium allele frequency spectra via spectral methods. *Theor Popul Biol.* 79:203–219.
- Madsen DB, Haizhou M, Brantingham PJ, Xing G, Rhode D, Haiying Z, Olsen JW. 2006. The late upper Paleolithic occupation of the northern Tibetan Plateau margin. *J Archaeol Sci.* 33:1433–1444.
- Metspalu M, Kivisild T, Bandelt H-J, Richards M, Villems R. 2006. The pioneer settlement of modern humans in Asia. In: Bandelt H-J, Macaulay V, Richards M, editors. *Human mitochondrial DNA and the evolution of Homo sapiens*. Berlin (Germany): Springer-Verlag. p. 181–199.
- Nei M. 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.
- Noonan JP, Coop G, Kudavavalli S, et al. (11 co-authors). 2006. Sequencing and analysis of Neanderthal genomic DNA. *Science* 314:1113–1118.
- Palanichamy MG, Sun C, Agrawal S, Bandelt HJ, Kong QP, Khan F, Wang CY, Chaudhuri TK, Palla V, Zhang YP. 2004. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. *Am J Hum Genet.* 75:966–978.
- Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genet.* 2:e190.
- Peakall ROD, Smouse PE. 2006. genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes.* 6:288–295.
- Peng MS, Palanichamy MG, Yao YG, et al. (18 co-authors). 2011. Inland post-glacial dispersal in East Asia revealed by mitochondrial haplogroup M9a'b. *BMC Biol.* 9:2.
- Peng Y, Yang Z, Zhang H, et al. (15 co-authors). 2011. Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. *Mol Biol Evol.* 28:1075–1081.
- Posada D, Buckley TR. 2004. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Syst Biol.* 53:793–808.
- Qian YP, Chu ZT, Dai Q, Wei CD, Chu JY, Tajima A, Horai S. 2001. Mitochondrial DNA polymorphisms in Yunnan nationalities in China. *J Hum Genet.* 46:211–220.
- Qin Z, Yang Y, Kang L, et al. (14 co-authors). 2010. 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.
- Reddy BM, Langstieh BT, Kumar V, Nagaraja T, Reddy AN, Meka A, Reddy AG, Thangaraj K, Singh L. 2007. Austro-Asiatic tribes of Northeast India provide hitherto missing genetic link between South and Southeast Asia. *PLoS One* 2:e1141.
- Reich D, Thangaraj K, Patterson N, Price AL, Singh L. 2009. Reconstructing Indian population history. *Nature* 461:489–494.
- Rhode D, Haiying Z, Madsen DB, Xing G, Jeffrey Brantingham P, Haizhou M, Olsen JW. 2007. Epipaleolithic/early Neolithic settlements at Qinghai Lake, western China. *J Archaeol Sci.* 34:600–612.
- Saillard J, Forster P, Lynnerup N, Bandelt HJ, Norby S. 2000. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet.* 67:718–726.
- Sengupta S, Zhivotovskiy LA, King R, et al. (15 co-authors). 2006. Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am J Hum Genet.* 78:202–221.
- Sharma G, Tamang R, Chaudhary R, et al. (12 co-authors). 2012. Genetic affinities of the central Indian tribal populations. *PLoS One* 7:e32546.
- Shelach G. 2000. The earliest neolithic cultures of northeast China: recent discoveries and new perspectives on the beginning of agriculture. *J World Prehist.* 14:363–413.
- Shi H, Dong YL, Wen B, Xiao CJ, Underhill PA, Shen PD, Chakraborty R, Jin L, Su B. 2005. Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. *Am J Hum Genet.* 77: 408–419.
- Shi H, Zhong H, Peng Y, et al. (13 co-authors). 2008. Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol.* 6:45.
- Shi Y-F. 2004. The emergence and abandonment of the ice sheet hypothesis over the Qinghai-Xizang Plateau during the Ice Age. *Quaternary Sci.* 24:10–18 (in Chinese).
- Shi Y-F, Zhao J-D. 2009. The special warm-humid climate and environment in China during 40-30 kaBP: Discovery and review. *J Glaciol Geocryol.* 31:1–10 (in Chinese).
- Shi Y-F, Zheng B-X, Li S-J. 1992. Last glaciation and maximum glaciation of the Qinghai-Xizang (Tibet) plateau: a controversy to M. Kuhle's ice sheet hypothesis. *Chinese Geographical Sci.* 2:293–311.
- Simonson TS, Yang Y, Huff CD, et al. (12 co-authors). 2010. Genetic evidence for high-altitude adaptation in Tibet. *Science* 329:72–75.
- Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, Salas A, Oppenheimer S, Macaulay V, Richards MB. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet.* 84:740–759.
- Soares P, Trejaut JA, Loo JH, et al. (14 co-authors). 2008. Climate change and postglacial human dispersals in southeast Asia. *Mol Biol Evol.* 25: 1209–1218.
- Su B, Xiao C, Deka R, et al. (11 co-authors). 2000. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum Genet.* 107:582–590.
- Su B, Xiao J, Underhill P, et al. (21 co-authors). 1999. Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *Am J Hum Genet.* 65:1718–1724.
- Tamm E, Kivisild T, Reidla M, et al. (21 co-authors). 2007. Beringian standstill and spread of Native American founders. *PLoS One* 2:e829.
- Tanaka M, Cabrera VM, Gonzalez AM, et al. (28 co-authors). 2004. Mitochondrial genome variation in eastern Asia and the peopling of Japan. *Genome Res.* 14:1832–1850.
- Torrioni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, Wallace DC. 1994. Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. *Am J Phys Anthropol.* 93:189–199.
- Torrioni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet.* 53:563–590.
- Torrioni A, Schurr TG, Yang CC, et al. (12 co-authors). 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* 130:153–162.
- Torrioni A, Sukernik RI, Schurr TG, Starikorskaya YB, Cabell MF, Crawford MH, Comuzzie AG, Wallace DC. 1993. mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am J Hum Genet.* 53:591–608.
- Underhill PA, Kivisild T. 2007. Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu Rev Genet.* 41:539–564.
- Underhill PA, Myres NM, Rootsi S, et al. (34 co-authors). 2010. Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *Eur J Hum Genet.* 18:479–484.

- van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat.* 30:E386–E394.
- Voight BF, Adams AM, Frisse LA, Qian Y, Hudson RR, Di Rienzo A. 2005. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc Natl Acad Sci U S A.* 102:18508–18513.
- Wang B, Zhang YB, Zhang F, et al. (18 co-authors). 2011. On the origin of Tibetans and their genetic basis in adapting high-altitude environments. *PLoS One* 6:e17002.
- Wang Y, Nielsen R. 2012. Estimating population divergence time and phylogeny from single-nucleotide polymorphisms data with out-group ascertainment bias. *Mol Ecol.* 21:974–986.
- Wang Z, Shen X, Liu B, et al. (11 co-authors). 2010. Phylogeographical analyses of domestic and wild yaks based on mitochondrial DNA: new data and reappraisal. *J Biogeogr.* 37:2332–2344.
- Wang Z, Yonezawa T, Liu B, Ma T, Shen X, Su J, Guo S, Hasegawa M, Liu J. 2011. Domestication relaxed selective constraints on the yak mitochondrial genome. *Mol Biol Evol.* 28:1553–1556.
- Wang ZH. 1994. History of nationalities in China. Beijing: China Social Science Press.
- Wells RS, Yuldasheva N, Ruzibakiev R, et al. (27 co-authors). 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci U S A.* 98:10244–10249.
- Wen B, Li H, Gao S, et al. (18 co-authors). 2005. Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. *Mol Biol Evol.* 22:725–734.
- Wen B, Li H, Lu D, et al. (18 co-authors). 2004. Genetic evidence supports demic diffusion of Han culture. *Nature* 431:302–305.
- Wen B, Xie X, Gao S, et al. (13 co-authors). 2004. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *Am J Hum Genet.* 74:856–865.
- Wu T. 2001. The Qinghai-Tibetan Plateau: how high do Tibetans live? *High Alt Med Biol.* 2:489–499.
- Wu T, Kayser B. 2006. High altitude adaptation in Tibetans. *High Alt Med Biol.* 7:193–208.
- Xu S, Li S, Yang Y, et al. (13 co-authors). 2011. A genome-wide search for signals of high-altitude adaptation in Tibetans. *Mol Biol Evol.* 28:1003–1011.
- Yang X, Wan Z, Perry L, et al. (13 co-authors). 2012. Early millet use in northern China. *Proc Natl Acad Sci U S A.* 109:3726–3730.
- Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP. 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet.* 70:635–651.
- Yao YG, Zhang YP. 2002. Phylogeographic analysis of mtDNA variation in four ethnic populations from Yunnan Province: new data and a reappraisal. *J Hum Genet.* 47:311–318.
- Yi X, Liang Y, Huerta-Sanchez E, et al. (67 co-authors). 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329:75–78.
- Yuan BY, Huang WW, Zhang D. 2007. New evidence for human occupation of the northern Tibetan Plateau, China during the Late Pleistocene. *Chin Sci Bull.* 52:2675–2679.
- Zhang DD, Li SH. 2002. Optical dating of Tibetan human hand- and footprints: An implication for the palaeoenvironment of the last glaciation of the Tibetan Plateau. *Geophys Res Lett.* 29:1072–1074.
- Zhao M, Kong QP, Wang HW, et al. (16 co-authors). 2009. Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. *Proc Natl Acad Sci U S A.* 106:21230–21235.
- Zheng HX, Yan S, Qin ZD, Jin L. 2012. MtDNA analysis of global populations support that major population expansions began before Neolithic Time. *Sci Rep.* 2:745.
- Zheng HX, Yan S, Qin ZD, Wang Y, Tan JZ, Li H, Jin L. 2011. Major population expansion of East Asians began before neolithic time: evidence of mtDNA genomes. *PLoS One* 6:e25835.
- Zhivotovsky LA. 2001. Estimating divergence time with the use of microsatellite genetic distances: impacts of population growth and gene flow. *Mol Biol Evol.* 18:700–709.
- Zhivotovsky LA, Underhill PA, Cinnioglu C, et al. (17 co-authors). 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet.* 74:50–61.
- Zhong H, Shi H, Qi XB, Duan ZY, Tan PP, Jin L, Su B, Ma RZ. 2011. Extended Y-chromosome investigation suggests post-Glacial migrations of modern humans into East Asia via the northern route. *Mol Biol Evol.* 28:717–727.
- Zhong H, Shi H, Qi XB, Xiao CJ, Jin L, Ma RZ, Su B. 2010. Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia. *J Hum Genet.* 55:428–435.