

Bing Su · Chunjie Xiao · Ranjan Deka
Mark T. Seielstad · Daoroong Kangwanpong
Junhua Xiao · Daru Lu · Peter Underhill
Luca Cavalli-Sforza · Ranajit Chakraborty · Li Jin

Y chromosome haplotypes reveal prehistorical migrations to the Himalayas

Received: 27 July 2000 / Accepted: 21 September 2000 / Published online: 11 November 2000

© Springer-Verlag 2000

Abstract By using 19 Y chromosome biallelic markers and 3 Y chromosome microsatellite markers, we analyzed the genetic structure of 31 indigenous Sino-Tibetan speaking populations (607 individuals) currently residing in East, Southeast, and South Asia. Our results showed that a T to C mutation at locus M122 is highly prevalent in almost all of the Sino-Tibetan populations, implying a strong genetic affinity among populations in the same language family. Furthermore, the extremely high frequency of H8, a haplotype derived from M122C, in the Sino-Tibetan speaking populations in the Himalayas including Tibet and northeast India indicated a strong bottleneck effect that occurred during a westward and then southward

migration of the founding population of Tibeto-Burmans. We, therefore, postulate that the ancient people, who lived in the upper-middle Yellow River basin about 10,000 years ago and developed one of the earliest Neolithic cultures in East Asia, were the ancestors of modern Sino-Tibetan populations.

Introduction

As delineating migrations becomes one of the major themes in human evolution studies, Y chromosome markers began to show their power in tracing human prehistory (Jobling and Tyler-Smith 1995; Hammer et al. 1997; Underhill et al. 1996, 1997; Lell et al. 1997; Bianchi et al. 1998; Su et al. 1999, 2000; Karafet et al. 1999; Underhill et al. 2000). Y chromosome haplotypes derived from multiple biallelic markers can be considered as a multiple-allele single-locus system, which occurred in an ordered time series, each of which might mark the footprint of a unique migration event. Since the effective population size of the Y chromosome is much smaller than that of the autosomes, many of the Y chromosome biallelic markers tend to show restricted regional distribution, in other words, population specificity. Therefore, the distribution of multiple Y chromosome haplotypes in human populations is highly informative in the inference of population migration events.

Our recent study on the extant East Asian populations with 19 Y chromosome biallelic markers demonstrated that during the last ice age, about 60,000 years ago, modern humans of African origin first reached the southern part of East Asia, namely, mainland Southeast Asia. Then, a northward diaspora led to the peopling all across East Asia (Su et al. 1999). In that study, ten out of the 19 Y chromosome biallelic markers are East Asian-specific and generally absent in other world populations. Hence, these markers are highly informative in studying the migrations of East Asian populations.

The Himalayas is a plateau located at the western part of East Asia boarded by high mountains (most of them are

B. Su · C. Xiao · R. Chakraborty · L. Jin (✉)
Human Genetics Center, University of Texas-Houston,
6901 Bertner Avenue, Houston, TX 77030, USA
e-mail: ljin@utsph.sph.uth.tmc.edu,
Tel.: +1-713-5009846, Fax: +1-713-5000900

B. Su · J. Xiao · D. Lu · L. Jin
Institute of Genetics, School of Life Sciences,
Fudan University and Morgan-Tan
International Center for Life Sciences, Shanghai, China

B. Su
Kunming Institute of Zoology, Chinese Academy of Sciences,
Kunming, China

C. Xiao
Department of Biology, Yunnan University, Kunming, China

R. Deka
Department of Environmental Health, University of Cincinnati,
Cincinnati, Ohio, USA

M.T. Seielstad
Program for Population Genetics,
Harvard School of Public Health, Boston, Mass., USA

D. Kangwanpong
Department of Biology, Chiang Mai University, Chiang Mai,
Thailand

P. Underhill · L. Cavalli-Sforza
Department of Genetics, Stanford University,
Stanford, Calif., USA

above 3,000 meters), to the west by the Tian-Shan, north by the Kunlun, east by the Hengduan and south by the Himalaya. According to archaeological studies, the Himalayas began to show human activity only in relatively recent time. The earliest human habitation in Tibet was dated back to about 4000–5000 years ago (Cavalli-Sforza and Piazza 1994).

People living in the Himalayan region mostly speak Sino-Tibetan languages, the second largest language family in the world by population size. There are 360 different individual languages in this family spoken in nine Asian countries, including China, Myanmar, Laos, Thailand, Vietnam, Bangladesh, India, Nepal and Bhutan (Martisoff 1991). This family is further divided into two subfamilies, namely Chinese and Tibeto-Burman (Martisoff 1991). Being one of the most ancient written languages in the world, Chinese has the most number of speakers across East Asia, and has diverged into many dialects in the past few thousand years. Because of the cultural significance, Chinese has had extensive influence on other languages during and after the Neolithic expansion of Han Chinese populations. The other subfamily, Tibeto-Burman has highly differentiated individual languages (at least 250), and was further divided into four branches: Baric, Bodic, Burmese-Lolo and Karen (Grimes 1999). People speaking Tibeto-Burman mostly live in the Himalayas, including Tibet, northeast India, west Myanmar, Nepal, Bhutan, Sichuan and Yunnan. However, there are also several other relatively small Tibeto-Burman populations living in Southeast Asia who actually migrated from the plateau area in the middle of the first millennium AD (Martisoff 1991).

The linguistic connection between Tibeto-Burman and Chinese are well established. There are over 300 cognates between Proto-Tibeto-Burman and Old Chinese, grouping them into the same language family (Martisoff 1991). Based on the archaeological findings, the ancestors who spoke Proto-Sino-Tibetan were estimated to live around at least 6000 years ago (Cavalli-Sforza and Piazza 1994; Martisoff 1991). Wang also estimated that Sinitic split away from Tibeto-Burman around 6000 years ago, based on lexical evidence and cladistic methods (Wang 1998). However, the identification of the ancestor Sino-Tibetan population and its relationship with those early Neolithic populations living in East Asia remains a mystery. Besides the Sino-Tibetan family, another five language families are also present in East Asia, which are Altaic, Daic, Hmong-Mien, Austro-Asiatic and Austronesian (Grimes 1999). The Altaic speaking populations are mainly located in northern East Asia, mostly in Siberia. The other four families are spoken in southern China, Southeast Asia and the Pacific Islands.

In order to trace the origin and historic expansion of the Sino-Tibetan populations and their relationship with the other language families, and especially peopling of the Himalayas, a total of 31 indigenous Sino-Tibetan populations were studied by typing 19 Y chromosome biallelic and 3 Y chromosome microsatellite markers in 607 male individuals. As described earlier, 10 of the 19 Y chromo-

some biallelic markers are polymorphic in East Asian populations. It is shown in this study that those markers reveal a series of ancient migration events leading to the inhabitation of the Himalayan plateau.

Materials and methods

A total of 607 male DNA samples were collected, covering all the main language branches in the Sino-Tibetan language family (Table 1, Fig. 4). The Han Chinese populations were sampled from 22 provincial areas across mainland China, and are grouped into 15 populations based on their geographic affinity (Table 1). There are seven Baric populations, with six of them from northeast India and one from northwest Yunnan, a southwest province of China. In the Burmese-Lolo branch, six populations were sampled including one from Hunan (China), and five from Yunnan (China). Two Bodic populations were included with one from southern Tibet and the other from northern Yunnan (Qian et al. 2000). In addition, one Karen population was also sampled from Thailand, representing the Karen branch.

The 19 Y chromosome biallelic markers were chosen based on their population specificity, including M1 (Alu insertion, also called YAP), M3 (C to T substitution), M5 (A to G substitution), M7 (C to G substitution), M9 (C to G substitution), M15 (9 bp insertion), M17 (1 bp deletion), and M45 (G to A substitution), M50 (T to C substitution), M88 (A to G substitution), M89 (C to T substitution), M95 (C to T substitution), M103 (C to T substitution), M110 (T to C substitution), M111 (4 bp deletion), M119 (A to C substitution), M120 (T to C substitution), M122 (T to C substitution), and M134 (1 bp deletion) (Vollrath et al. 1992; Hammer et al. 1997; Underhill et al. 1996, 1997; Su et al. 1999; Kayser et al. 1997). M1 is an ancient polymorphism occurring both in Africans and Asians, but generally absent in other populations (Hammer et al. 1997). M3 is an American Indian specific marker, while M5 is Oceanian specific (Underhill et al. 1996, 1997). M45 has a high frequency in Caucasians, but is relatively rare in East Asian populations (Su et al. 1999). M9 is generally absent in Africans, but frequent in non-African populations (Underhill et al. 1996; Su et al. 1999; Underhill et al. 2000). M122, M95 and M119 define three lineages which are found predominantly in East Asian populations, but are absent in other parts of the world (Su et al. 1999). The haplotype notations (H1–H17) follow our previous study and the phylogenetic network among haplotypes is given in Fig. 1 (Su et al. 1999). An allelic-specific genotyping assay was used to type the 19 Y chromosome biallelic markers as described in our previous report (Su et al. 1999).

Three Y chromosome microsatellites (DYS389, DYS390 and DYS391) were typed in 138 Sino-Tibetan individuals who are haplotype H8. The variances of repeat numbers at each microsatellite locus was calculated and used in the dating of the emergence of Proto-Sino-Tibetan and Proto-Tibeto-Burman populations. The age estimation follows the methods previously described (Su et al. 1999).

The principal component (PC) analysis was conducted based on the haplotype and haplogroup frequency distribution among the Sino-Tibetans. In addition, a total of 22 East Asian populations who speak non-Sino-Tibetan languages were also included in the PC analysis for comparison (Su et al. 1999, 2000).

Results

The haplotypes of 19 Y chromosome biallelic markers of a total of 607 individuals from 31 Sino-Tibetan populations are shown in Table 1. Among them, five populations (Jingpo, Tibetan-Lhasa, Tibetan-Yunnan, Tujia, and Yi) have been previously published (Su et al. 1999; Qian et al. 2000). A total of 17 haplotypes were observed in our pre-

Table 1 Y chromosome haplotype frequency distribution in 31 Sino-Tibetan populations

Population	Sample size	Language	H1	H2	H3	H4	H5	H6	H7	H8	H9	H11	H12	H13	H14
			\	M1 ^a	M15	M89	M9	M122	M7	M134	M119	M95	M111	M120	M45
1. Shandong Han	32	Chinese, Mandarin	9.4			3.1	18.8	28.1		28.1	9.4				3.1
2. Henan Han	28	Chinese, Mandarin	7.1			3.6	25.0	32.1		14.3	10.7				7.1
3. Northern Han	22	Chinese, Mandarin	9.1				22.7	27.3		27.3	9.1				4.5
4. Anhui Han	22	Chinese, Mandarin	13.6				18.2	27.3		18.2	18.2				4.6
5. Zhejiang Han	50	Chinese, Mandarin	12.0				6.0	24.0		26.0	26.0	6.0			
6. Jiangsu Han	55	Chinese, Mandarin	7.3	1.8		1.8	18.2	23.6	3.6	21.8	16.4	3.6			1.8
7. Shanghai Han	30	Chinese, Mandarin	6.7	3.3			16.7	23.3	3.3	16.7	26.7				3.3
8. Hubei Han	18	Chinese, Mandarin	5.6				11.1	27.8	5.6	33.3	16.7				
9. Sichuan Han	14	Chinese, Mandarin		7.1			28.6	7.1	7.1	35.7	7.1	7.1			
10. Jiangxi Han	21	Chinese, Mandarin	4.8	4.8		9.5	19.1	19.1		23.8	14.3	4.8			
11. Hunan Han	15	Chinese, Mandarin					13.3	33.3		26.7	13.3	13.3			
12. Fujian Han	13	Chinese, Mandarin	7.7				7.7	38.5		38.5			7.7		
13. Yunnan Han	27	Chinese, Mandarin	11.1			3.7	3.7	18.5		55.6	3.7	3.7			
14. Guangxi Han	3	Chinese, Mandarin	33.3								66.7				
15. Guangdong Han	15	Chinese, Mandarin						40.0		26.7	20.0		6.7	6.7	
16. Kachari	20	Tibeto-Burman, Baric					15.0			85.0					
17. Jingpo	5	Tibeto-Burman, Baric, Kachinic.								100					
18. Rabha	17	Tibeto-Burman, Baric, Konyak	11.8			5.9	5.9			76.5					
19. Naga	3	Tibeto-Burman, Baric, Konyak						33.3		66.7					
20. Adi	5	Tibeto-Burman, Baric, Mirish								100					
21. Nishi	9	Tibeto-Burman, Baric, Mirish					11.1			88.9					
22. Apatani	5	Tibeto-Burman, Baric, Mirish				20.0				80.0					
23. Tibetan-Lhasa ^b	46	Tibeto-Burman, Bodic, Bodish, Tibetan	8.7	23.9	17.4	4.3		4.3		34.8		2.2		2.2	2.2
24. Tibetan-Yunnan ^b	27	Tibeto-Burman, Bodic, Bodish, Kham		14.8	29.6	3.7	14.8	7.4		29.6					
25. Tujia ^b	10	Tibeto-Burman, Burmese-Lolo, Lolo	10.0				20.0	30.0	10.0		20.0		10.0		
26. Jino	18	Tibeto-Burman, Burmese-Lolo, Lolo	16.7			5.6	38.9	11.1		22.2		5.6			
27. Lahu-Yunnan	13	Tibeto-Burman, Burmese-Lolo, Lolo	15.4			30.8	15.4	15.4		15.4		7.7			
28. Yi ^b	14	Tibeto-Burman, Burmese-Lolo, Lolo			14.3		42.9	21.4		7.1		14.3			
29. Bai	13	Tibeto-Burman, Burmese-Lolo, Lolo	15.4		7.7		15.4	30.8		30.8					
30. Naxi	18	Tibeto-Burman, Burmese-Lolo, Lolo					77.8					16.7			5.5
31. Karen	19	Tibeto-Burman, Karen					36.8	11		52.6					

^aMost recent markers associated with haplotypes^bData published in our previous study (Su et al. 1999; Qian et al. 2000)

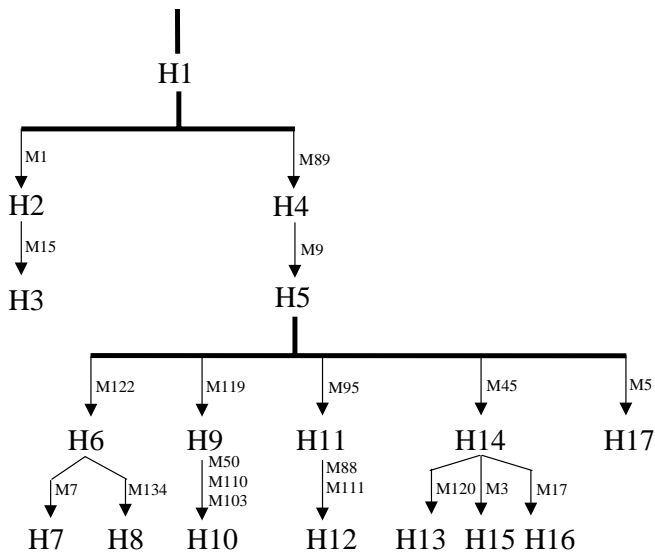


Fig. 1 The phylogenetic network of 17 Y chromosome haplotypes defined by 19 biallelic markers. This is a revised version of that presented by Su et al. (1999) and the most recent markers defining the haplotypes are labeled beside the branches

vious study on the extent East Asian populations and no recurrent mutations were identified (Su et al. 1999). Of the 17 haplotypes, 13 haplotypes are present in the Sino-Tibetan populations. The frequency distribution of the 13 haplotypes is listed in Table 1. It is noteworthy that, sharing a T to C mutation at locus M122, H6, H7 and H8 are the predominant haplotypes in almost all of the Sino-Tibetan populations except Naxi. The M122C allele is East Asian specific and absent in Africans, Europeans and Oceanians, as revealed in our previous study (Su et al. 1999). Recent studies in the extant Siberian and Central Asia populations revealed only sporadic occurrences of M122C and M119C (Lell and Wells, personal communication). H6 is the ancestral haplotype of the M122C alleles, while H7 and H8 are the two derived ones with additional mutations, M7 and M134, respectively (Fig. 1). The average frequency of M122C in Sino-Tibetan populations is 52.2% (28.6–100%), the highest among the six language families in East Asia (Hmong-Mien, 47.6%; Daic, 26.1%; Austronesian, 25.7%; Altaic, 24.5%; and Austroasiatic, 21.8%; detailed data not shown). Given that Hmong-Mien populations had extensive known interactions with the Han Chinese population in the last three millenniums (Wang 1994), we conclude that the prevalence of M122C in Sino-Tibetan populations implies relatively close genetic relationships among these Sino-Tibetan populations and a possible common origin of members in this language family. Interestingly, Naxi is the only exception without the M122C alleles.

The genetic affinity among Sino-Tibetan populations is further reflected in the PC analysis of the extant East Asian populations, in which most of the Sino-Tibetan populations are clustered together (see Fig. 2a). In addition, the PC analysis also showed that Sino-Tibetan popu-

lations are closely related to Altaic populations, the northern East Asian language family, and also with Hmong-Mien, the southern language family. Interestingly, as a predominant haplotype in the Hmong-Mien language family (28.6% on average; Su et al. 1999), H7 also showed up in southern Hans and Tujia with relatively low frequencies (1.8% in southern Han and 10% in Tujia), but was absent in northern Hans and other Sino-Tibetan populations. On the other hand, as compared to the other southern language families, the Hmong-Mien family has a much higher frequency of M122C alleles (47.6%). Hence, this contrast implies strong interactions between the Han and Hmong-Mien peoples that have lasted for several thousand years, as is confirmed in history literature (Wang 1994), although a possible shared ancestry can not be ruled out.

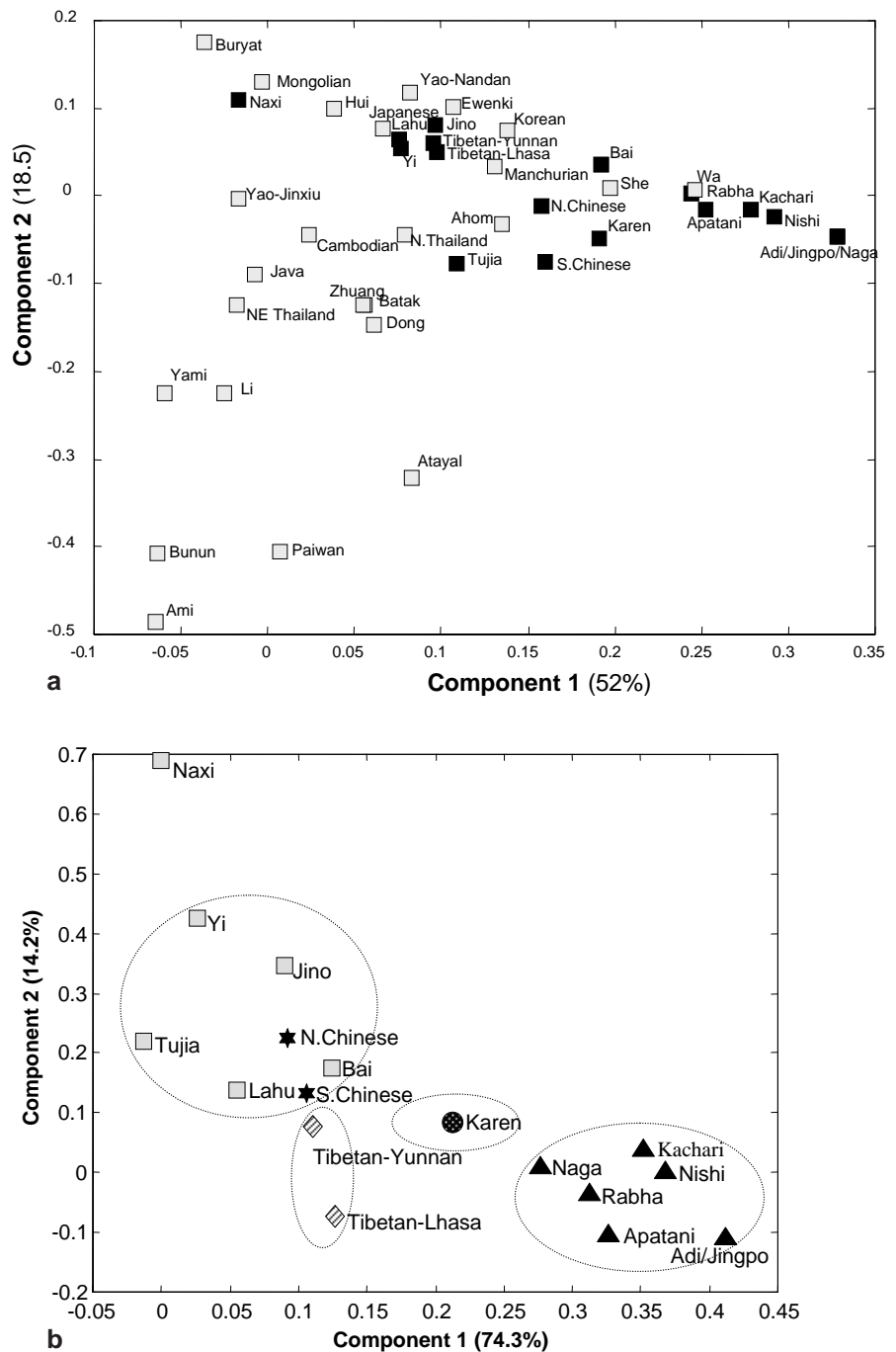
However, despite the prevalence of M122C in most of the Sino-Tibetans, populations under the same language branch tend to share similar patterns of haplotype distribution, but are distinctive across linguistic branches. The genetic differentiation among language branches probably reflects the demographic history of those populations. First of all, the Han Chinese populations, who currently speak Chinese, are highly diversified, especially in southern Hans. This pattern is expected considering the extensive contacts between Han and other ethnic populations after its Neolithic expansion over 5,000 years ago. Besides the M122C haplotypes, H5 (M9G) and H9 (M119C) are also popular in Han populations with medium frequencies. H9 is virtually absent in all the Tibeto-Burman subfamily members except Tujia, indicating that the interaction between Chinese and other southern populations occurred after the divergence of the Chinese and Tibeto-Burmans, and a limited gene flow occurred between them after the divergence.

The Tibeto-Burman subfamily can be further divided into four branches, including Baric, Bodic, Burmese-Lolo and Karen (Grimes 1999). Among them, Burmese-Lolo and Karen show similar patterns of haplotype distribution with a characteristic occurrence of H11 and H12 (defined by a C to T mutation at locus M95) in most of the populations. Since all the populations under these two language branches live in southern China (mostly in Yunnan province) and Southeast Asia, the presence of H11 and H12 might indicate influences from Daic and Austro-Asiatic families, which are neighbors of Burmese-Lolo and Karen speaking populations, and have the highest frequency of H11 and H12 on average (data not shown).

The Bodic branch, the so-called Tibetans, are distinctive from the other three branches with a high frequency of YAP+. The predominant occurrence of both YAP+ and M122C in Tibetan populations indicates a multiple origin of Tibetan Y chromosomes from both East Asia and central Asia/southwest Siberia (Qian et al. 2000).

Surprisingly, the Baric populations have extremely homogeneous Y chromosomes. The M122C alleles are almost fixed in this language branch with only sporadic occurrence of other Y haplotypes. Furthermore, the extremely high frequencies of H8 (M134G, 84.4% on aver-

Fig. 2a, b Principal component (PC) analysis of Sino-Tibetan populations. **a** PC analysis in 43 East Asian populations representing all the six language families spoken in this region. This map accounts for 70.5% of the genetic variation. Considering the small sample size in some of the populations, the haplotype frequencies of the 13 haplotypes were grouped into 8 haplogroups where group 1 = H1 (ancestral for all markers), group 2 = H2+H3 (YAP+), group 3 = H4 (M89C), group 4 = H5 (M9G), group 5 = H6+H7+H8 (M122C), group 6 = H9 (M119C), group 7 = H11+H12 (M95T), and group 8 = H13+H14 (M45A). The *black squares* are Sino-Tibetan populations. The Han Chinese were grouped into two populations, the Northern Han and the Southern Han with the Yangze River as the borderline. Data of 22 populations from the other five language families are from Su et al. (1999, 2000): the Altaic family (Buryat, Mongolian, Ewenki, Hui and Manchurian), the Austro-Asiatic family (Cambodian and Wa), the Daic family (Dong, Li, Zhuang, Ahom, Yao-Jinxiu, N. Thailand and NE. Thailand), the Hmong-Mien family (Yao-Nandan and She) and the Austronesian family (Batak, Java, Ami, Atayal, Paiwan, Bunun and Yami). Korean and Japanese are two isolated languages (Grimes 1999). See Fig. 4 for the geographic locations of all the Sino-Tibetan populations studied. **b** PC analysis in 18 Sino-Tibetan populations based on the 13 haplotype frequency distribution. It accounts for 88.5% of the original genetic variation



age) in all the Baric populations strongly suggests a population bottleneck event probably associated with the branching of Baric from the ancestral Proto-Tibeto-Burman population. In contrast, in Han Chinese, Burmese-Lolo and Karen speaking populations, the frequencies of H6 are usually either higher or equivalent to those of H8, except for Sichuan Han and Yunnan Han, who actually live in the neighboring provinces of Tibet. The homogeneity of Baric Y chromosomes is further reflected by the Y chromosome microsatellite data. Up to 75.5% (37 out of 49) of H8 Baric individuals belong to two mi-

cro-satellite haplotypes (13–24–10 and 13–25–10, see Table 2), and they are only one mutation step away from each other. Interestingly, the H8 frequencies in the two Bodic populations are also quite high, which implies a relatively close genetic relationship between Baric and Bodic, probably due to interaction and/or recent common ancestry.

The genetic divergence of the four language branches in the Tibeto-Burman subfamily were further revealed in the PC analysis (Fig. 2b). Populations in the same language branch tend to cluster together and stay relatively

Table 2 Y chromosome microsatellite haplotype distribution in 138 individuals

Haplotype (389-390-391)	Han Chinese	Tibeto-Burman
12-24-10	7	
12-24-12	1	
12-25-10	6	2
12-25-11	1	
12-26-10	1	
13-22-10	2	
13-23-10	1	1
13-24-10	9	28
13-24-11		2
13-25-09	1	
13-25-10	18	14
13-25-11	2	1
13-26-10	15	2
13-26-11	4	
14-22-10	1	
14-23-10	1	
14-24-11		1
14-25-10	4	3
14-25-11	1	
14-26-10		1
14-26-11	2	
15-21-10	1	
15-24-10	2	
15-25-11	1	
15-26-09	1	
15-26-10	1	
Total	83	55

All the 138 individuals are H8 (M134G) with 83 Han Chinese and 55 Tibeto-Burmans. Haplotypes are defined by combinations of repeat numbers at the three loci, namely DYS389, DYS390 and DYS391

separated from other branches. The northern and southern Han Chinese populations are close to each other, but do not form a separate cluster due to their rather diversified

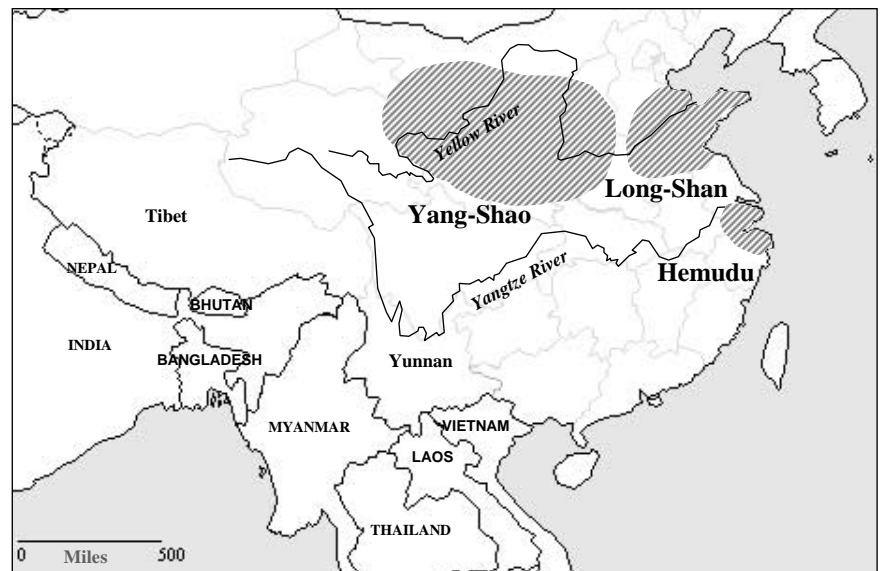
haplotype distribution, resulting from frequent contact with other ethnic populations over the past several millenniums.

Discussion

Our previous study on the extant East Asian populations with 19 Y chromosome biallelic markers showed that around 60,000 years ago, modern humans from Africa arrived in mainland Southeast Asia, and then a northward diaspora led to the peopling of East Asia (Su et al. 1999). According to the archeological findings, the earliest Neolithic cultures in East Asia occurred around 10,000 years ago, and one of them is located at the upper and middle Yellow River region of China, the so-called the Yang-Shao culture. The significance of this culture is highlighted by the advent of millet agriculture, and it is believed to be the mother culture of Sino-Tibetans (Fig. 3). The other two contemporary cultures are Long-Shan in the northeast and Hemudu in the southeast (Fig. 3). During the Yang-Shao epoch, the so-called Di-Qiang people, or Proto-Tibeto-Burman, were the dominant population living in the upper Yellow River basin, and formed a culture called "Ma-Jia-Yao" (Wang 1994). This culture was found to bear a strong similarity to the typical Yang-Shao culture, which is considered one of the three major cultures that led to the birth of Chinese civilization (Wang 1994).

Our conclusions, based on the Y chromosome haplotype distribution in the 31 Sino-Tibetan populations, are consistent with the historical records and literature of the origin of Sino-Tibetan people. It is shown that almost all the modern Sino-Tibetan populations share a common genetic signature, the high frequencies of M122C alleles defining haplotypes H6, H7 and H8, which are also present in the other five language families in East Asia at moderate to lower frequencies.

Fig. 3 The geographical distribution of China's three earliest cultures about 5,000 years ago. The early Yao-Shao started 8,500 years ago, while the Hemudu and Long-Shan started about 7,000 and 4,800 years ago, respectively (Wang 1994; Cavalli-Sforza and Piazza 1994)



Previous genetic studies using classical and microsatellite markers put Tibetans into the northern Mongoloid group, the group uniting all the populations living in the north to the Yangtze river, including northern Chinese, Japanese, Korean, Mongolian and Siberian (Zhao et al. 1986; Weng et al. 1989; Du and Xiao 1997; Chu et al. 1998). However, the mysterious occurrence of YAP+ in Tibetan populations always questions the East Asian origin of Tibetan populations because YAP+ is basically absent in almost all the other East Asian populations except in Japanese (Hammer et al. 1997). We recently reported our study on two Tibetan populations, the Zang (Tibetan-Lhasa) and Khamba (Tibetan-Yunnan). Our results showed that two predominant haplotype groups exist in the gene pool of Tibetan populations, one from East Asia (the M122C haplotypes) and the other from central Asia or southwest Siberia (the YAP+) (Qian et al. 2000). However, except for the moderate appearance in Yi and Bai populations, we did not observe YAP+ in Baric, Burmese-Lolo or Karen populations. In fact, both Yi and Bai peoples are neighbors of the Tibetans living in the Yunnan province. The occurrence of YAP+ in the Yi and Bai populations might result from recent gene flow from Tibetan populations as documented in the historical literature (Wang 1994), since only H3 (M15+, the derived haplotype under YAP+) is observed. Therefore, the absence of YAP+ in Baric, Burmese-Lolo and Karen populations indicates that the Y chromosome contribution to Tibetans from central Asia is a relatively recent event as compared to the early habitation of populations with dominant East Asian Y chromosome haplotypes. If that is the case, then Baric people were probably the first residents of the Himalayan region while Tibetans arrived later. This is again consistent with the historical literature (Cavalli-Sforza and Piazza 1994).

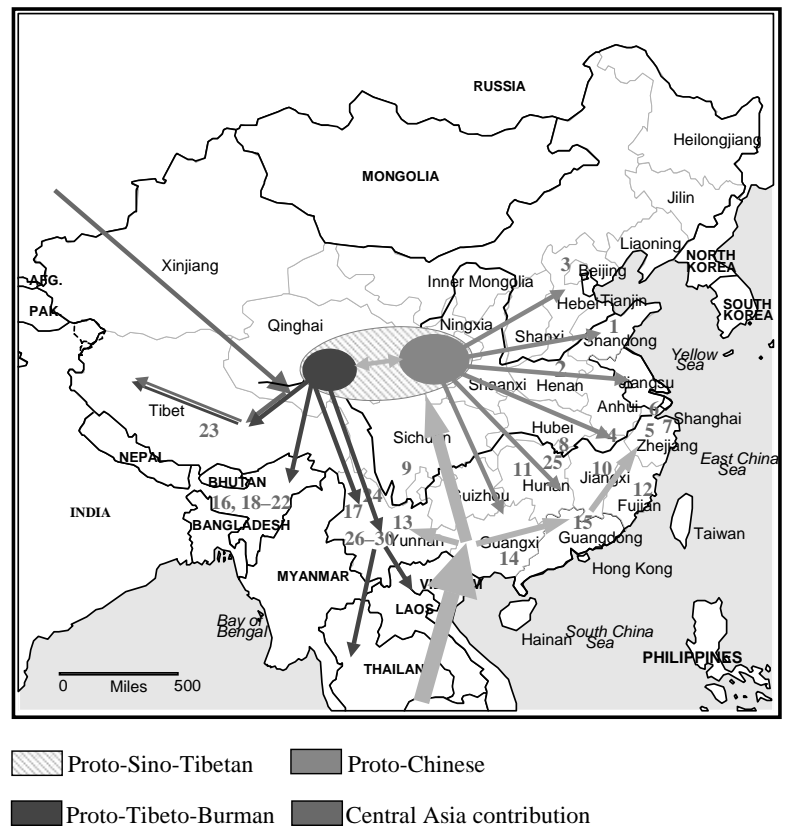
In order to estimate the age of the ancestral Sino-Tibetan population and, especially, the migration event to the Himalayas, we analyzed three Y chromosome microsatellites (DYS389, 390, 391) in those individuals with haplotype H8 (M134G) (138 individuals in total; Table 2). To minimize the effect of population substructure in the age estimation, we divided the 138 H8 individuals into two groups, the Chinese group (83 individuals) and the Tibeto-Burman group (55 individuals), and estimated them separately. By assuming an effective population size of 500–1000, a mutation rate of 0.18% (Heyer et al. 1997) and a generation time of 20 years, the age of H8 was calculated to be 20,000–40,000 years for the Chinese H8, and 5,200–5,900 years for the Tibeto-Burman H8. These estimations are based on the allele size variances of DYS390 (1.125 for Chinese and 0.401 for Tibeto-Burman), the largest among the estimations of the three microsatellites (Su et al. 1999). Therefore, we argue that these two estimations indicate two migration events leading to the peopling of the upper Yellow River basin by the ancestors of the Proto-Sino-Tibetan populations, and the peopling of the Himalayas by the Proto-Tibeto-Burman people. In fact, these two age estimations are quite consistent with the archaeological and lexical evidence (Wang 1998; Wu and Poirier 1995; Etlar 1996).

The close relationship between Sino-Tibetan populations and Altaic populations is intriguing. This is consistent with the earlier observations where northern East Asian populations include both Sino-Tibetan populations of northern origin and Altaic populations (Zhao et al. 1986; Weng et al. 1989; Du and Xiao 1997; Chu et al. 1998). The similarity between Sino-Tibetan and Altaic populations can be either interpreted as substantial gene flow between the populations (Chu et al. 1998) or a common ancestry (Du and Xiao 1997). The sporadic presence of M122C and M119C, which are of southern origin in the Altaic populations (Lell et al. 1997; Wells, personal communication), signifies the gene flow. Starostin (Wang 1995, 1996) proposed a northern origin of the Sino-Tibetans, based on linguistic evidence. However, this may not contradict our observation since a population of southern origin may speak a language of northern origin. The Sino-Tibetan language occurred tens of thousand years after the initial Yellow River settlement of populations originated from southern East Asia.

It would be interesting to compare the results of studies using the Y chromosome with previous studies on autosomal and mitochondrial DNA markers. Using 30 autosomal microsatellite markers, Chu et al. (1998) analyzed 28 East Asian populations. Among them, 11 populations were Sino-Tibetan speaking with 4 Han Chinese and 7 Tibeto-Burman populations. The phylogenetic analysis indicated a close genetic relationship among East Asian populations. The northern Sino-Tibetans cluster with northern East Asians, including Japanese, Korean and Altaic populations, and the southern Sino-Tibetans cluster with southern East Asians, including Daic, Austro-Asiatic, Hmong-Mien and Austronesian speaking populations, reflecting gene flow among geographically related East Asians. However, due to the high mutation rate of microsatellite markers and the unavailability of haplotype data in this study, it is difficult to infer the historical migrations of Sino-Tibetan populations. The mitochondrial DNA studies on Sino-Tibetan populations are sporadic without systematic comparisons so far. The appearance of Asian-specific haplogroups (e.g., A, B, C and D) in both Tibetans and Han Chinese also reflects the genetic affinity of the Chinese and Tibeto-Burman speaking populations (Kolman et al. 1996).

In summary, our genetic evidence on Y chromosome haplotype distributions outline an interesting picture about the origin and dispersion of Sino-Tibetan populations (Fig. 4). The ancestors of the Sino-Tibetan population were originally from southern East Asia (Su et al. 1999). Around 20,000–40,000 years ago, a population with dominant M122C Y chromosomes finally reached the upper and middle Yellow River basin. However, the exact geographic location of the first settlement of Sino-Tibetans is unknown. About 10,000 years ago, the Neolithic culture began to proliferate in this area with the advent of millet agriculture. Population growth triggered the need for exploring new habitats. The splitting of the two language subfamilies happened about 5000–6,000 years ago. A subgroup of the Proto-Sino-Tibetan, now called the Proto-Ti-

Fig. 4 The map of putative migration routes of the Sino-Tibetan populations. The numbers indicate the geographic locations of the Sino-Tibetan populations corresponding to the population numbers in Table 1



Proto-Tibeto-Burman people left their Yellow River homeland, marching westwards and then southward, leading to the inhabitation of the Himalayas. The ancient migrations of the Proto-Tibeto-Burman people probably followed the well-known route, called the Zang (Tibet)-Mien corridor (Wang 1994), the most frequently used route to enter the Himalayas from the east. This route starts in the upper Yellow River region, moves westward to Qinghai province and then southward to the Himalayas. The Baric branch moved its way southward, crossing the Himalayan mountains, and reached the southern Himalayan area, resulting in the peopling of Bhutan, Nepal, northeastern India and northern Yunnan. Subsequently, after having a substantial admixture with a population carrying YAP+, possibly from central Asia/southwest Siberia, the Bodic branch entered the Himalayas and eventually expanded all across Tibet. The Burmese-Lolo and Karen branches went southward into northwestern Yunnan and finally reached Vietnam, Laos and Thailand in the middle of the first millennium. In the past 5000 years, the Chinese subfamily expanded mainly to the east and south, and eventually inhabited all of China.

Acknowledgements We thank Dr. William S.-Y. Wang and James Zhu for inspiring comments on the early draft of the manuscript. B.S., R.C., and L.J. were supported by NIH grants. J.X. and D.L. were supported by a grant from National Natural Science Foundation of China.

References

- Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, et al (1998) Characterization of ancestral and derived Y-chromosomal haplotypes of New World native populations. *Am J Hum Genet* 63:1862–1871
- Cavalli-Sforza LL, Piazza MP (1994) The history and geography of human genes. Princeton University Press, Princeton, NJ
- Chu JY, Huang W, Kuang SQ, Wang JM, Xu JJ, Chu ZT, Yang ZQ, et al (1998) Genetic relationship of populations in China. *Proc Natl Acad Sci USA* 95:11763–11768
- Du RF, Xiao CJ (1997) Genetic distance between Chinese populations calculated on gene frequencies of 38 loci. *Science in China (Series C)* 40:613–621
- Etler DA (1996) The fossil evidence for human evolution in Asia. *Annu Rev Anthropol* 25:275–301
- Grimes BF (ed) (1999) The ethnologue: languages of the world. Summer Institute of Linguistics, <http://www.sil.org/ethnologue/>
- Hammer MF, Spurdle AB, Karafet T, Bonner MR, Wood ET, Novelletto A, Malaspina P, et al (1997) The geographic distribution of human Y chromosome variation. *Genetics* 145:787–805
- Heyer E, Puymirat J, Dieltjes P, Bakker E, Knijff P (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum Mol Genet* 6:799–803
- Jobling MA, Tyler-Smith C (1995) Fathers and sons: the Y chromosome and human evolution. *Trends Genet* 11:449–455
- Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, et al (1999) Ancestral Asian sources(s) of new world Y-chromosome founder haplotypes. *Am J Hum Genet* 64: 817–831

- Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, et al (1997) Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Legal Med* 110:125–133
- Kolman CJ, Sambuughin N, Bermingham E (1996) Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* 142:1321–1334
- Lell JT, Brown MD, Schurr TG, Sukernik RI, Starikovskaya YB, Torroni A, Moore LG, et al (1997) Y chromosome polymorphisms in native American and Siberian populations: identification of native American Y chromosome haplotypes. *Hum Genet* 100:536–543
- Martisoff JA (1991) Sino-Tibetan linguistics: present state and future prospects. *Annu Rev Anthropol* 20:469–504
- Qian YP, Qian BZ, Su B, Yu JK, Ke YH, Chu ZT, Shi L, et al (2000) Multiple origins of Tibetan Y chromosomes. *Hum Genet* 106:453–454
- Su B, Xiao JH, Underhill P, Deka R, Zhang WL, Akey J, Huang W, et al (1999) Y Chromosome evidence for a northward migration of modern humans in East Asia during the last ice age. *Am J Hum Genet* 65:1718–1724
- Su B, Jin L, Underhill P, Martinson J, Saha N, McGarvey ST, Shriver MD, et al (2000) Polynesian origins: insights from the Y chromosome. *Proc Natl Acad Sci USA* 97:8225–8228
- Underhill PA, Jin L, Zemans R, Oefner PA, Cavalli-Sforza LL (1996) A pre-Columbian human Y chromosome-specific transition and its implications for human evolution. *Proc Natl Acad Sci USA* 93:196–200
- Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, et al (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 7:996–1005
- Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, et al (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* 26:358–361
- Vollrath D, Foote S, Hilton A, Brown LG, Beer-Romero P, Bogan JS, Page DC (1992) The human Y chromosome: a 43-interval map based on naturally occurring deletions. *Science* 258:52–59
- Wang WSY (1995) The ancestry of Chinese: retrospect and prospect. *J Chinese Linguistics*, monograph no. 8
- Wang WSY (1996) Genes, dates, and the writing system. *Intl Rev Chinese Linguistics* 1:45–46
- Wang WSY (1998) In the Bronze Age and Early Iron Age peoples of Eastern Central Asia. University of Pennsylvania Museum Publications, pp 508–534
- Wang ZH (1994) History of nationalities in China. China Social Science Press
- Weng Z, Yuan Y, Du R (1989) Analysis of the genetic structure of human populations in China. *Acta Anthropol Sin* 8:261–268 (in Chinese)
- Wu XZ, Poirier FE (1995) Human evolution in China. Oxford University Press, Oxford
- Zhao T, Zhang G, Zhu Y, Zheng S, Liu D, Chen Q, Zhang X (1986) The distribution of immunoglobulin Gm allotypes in forty Chinese populations. *Acta Anthropol Sin* 6:1–8 (in Chinese)