Am. J. Hum. Genet. 74:856–865, 2004

Analyses of Genetic Structure of Tibeto-Burman Populations Reveals Sex-Biased Admixture in Southern Tibeto-Burmans

Bo Wen,[1] Xuanhua Xie,[1] Song Gao,[1] Hui Li,[1] Hong Shi,[2,3] Xiufeng Song,[1] Tingzhi Qian,[1] Chunjie Xiao,[2] Jianzhong Jin,[1] Bing Su,[3,4] Daru Lu,[1] Ranajit Chakraborty,[4] and Li Jin[1,4]

[1]State Key Laboratory of Genetic Engineering and Center for Anthropological Studies, School of Life Sciences, Fudan University, Shanghai; [2]Department of Biology and Human Genetics Center, Yunnan University, and [3]Key Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming; and [4]Center for Genome Information, Department of Environmental Health, University of Cincinnati, Cincinnati

An unequal contribution of male and female lineages from parental populations to admixed ones is not uncommon in the American continents, as a consequence of directional gene flow from European men into African and Hispanic Americans in the past several centuries. However, little is known about sex-biased admixture in East Asia, where substantial migrations are recorded. Tibeto-Burman (TB) populations were historically derived from ancient tribes of northwestern China and subsequently moved to the south, where they admixed with the southern natives during the past 2,600 years. They are currently extensively distributed in China and Southeast Asia. In this study, we analyze the variations of 965 Y chromosomes and 754 mtDNAs in >20 TB populations from China. By examining the haplotype group distributions of Y-chromosome and mtDNA markers and their principal components, we show that the genetic structure of the extant southern Tibeto-Burman (STB) populations were primarily formed by two parental groups: northern immigrants and native southerners. Furthermore, the admixture has a bias between male and female lineages, with a stronger influence of northern immigrants on the male lineages (~62%) and with the southern natives contributing more extensively to the female lineages (~56%) in the extant STBs. This is the first genetic evidence revealing sex-biased admixture in STB populations, which has genetic, historical, and anthropological implications.

Introduction

In the process of human evolution, genetically differentiated and geographically separated populations occasionally come in contact to form admixed populations. Understanding such admixture events is of great genetic, anthropological, and historical interest, since it reveals the genetic structure of such populations and might also be useful for disease mapping via admixture linkage disequilibrium (Chakraborty 1986; Chakraborty and Weiss 1988; Pritchard and Przeworski 2001; Smith et al. 2001; Ardlie et al. 2002). Genetic approaches have been proven prudent in revealing population admixture, as exemplified by the studies in Gypsies (Gresham et al. 2001), Turks (Di Benedetto et al. 2001), Icelanders (Helgason et al. 2001), Native Americans (Merriwether et al. 1997; Carvajal-Carmona et al. 2000; Mesa et al. 2000), and African Americans (Parra et al. 1998, 2001; Bortolini et al. 1999; Sans et al. 2002). More interesting are the observations of unequal contributions from male and female lineages of the parental populations to the admixed ones, such as Native American and African American populations, which is known as "sex-biased gene flow" or "directional mating" (Merriwether et al. 1997; Chakraborty 1998). Haplotype frequency distributions based on paternally transmitted Y chromosomes and maternally transmitted mtDNAs have been instrumental in detecting such sex-biased admixture processes (Merriwether et al. 1997; Chakraborty 1998; Parra et al. 1998, 2001; Bortolini et al. 1999; Mesa et al. 2000; Sans et al. 2002).

Tibeto-Burman (TB) is one of the two subfamilies of the Sino-Tibetan language family. There are 351 living languages in this subfamily, whose speakers are primarily distributed in nine countries in East, South, and Southeast Asia: China, Nepal, Bhutan, northeastern India, Pakistan, Myanmar, Bangladesh, Thailand, Vietnam, and Laos (Ethnologue Web site). Currently, in China, TB-speaking populations mainly reside in Qinghai in northern China, as well as in Tibet, Sichuan, Yunnan, and Hunan in southwestern China. According to historical records, the TB populations were derived from the ancient Di-Qiang tribes, which originally lived in north-

Received November 21, 2003; accepted for publication February 11, 2004; electronically published March 24, 2004.

Address for correspondence and reprints: Dr. Bo Wen, School of Life Sciences, Fudan University, Shanghai 200433, China, e-mail: lijin@fudan.edu.cn; or Dr. Li Jin, Center for Genome Information, University of Cincinnati, 3223 Eden Avenue, Cincinnati, OH 45267, e-mail: li.jin@uc.edu

0002-9297/2004/7405-0008$15.00

western China. In the Spring and Autumn Period (~2,600 years before present [BP]), the TB populations embarked on a large-scale southward migration, along the Tibetan-Burman Corridor, into an area probably densely populated by the Austro-Asiatic and possibly the Diac and Hmong-Mien populations, three groups that were native in south (Wang 1994). This is consistent with the genetic evidence, based on Y-chromosome markers, that almost all TB populations share a high frequency of M122-C and an extremely high frequency of M134-deletion, which was derived from M122-C (Su et al. 2000*b*). What remains to be further revealed is the genetic consequence of the admixture of the populations that were at least partially differentiated in terms of their allele frequency distributions.

In this study, we analyzed 10 Y-chromosome SNP markers in 965 individuals from 23 TB populations, as well as sequence variations in the mtDNA HVS-1 region and a set of diagnostic variants in the coding region of mtDNA in 754 individuals from 21 TB populations. We show that unequal contributions of male and female lineages from the parental populations—that is, the northern immigrants and southern native groups—played a significant role in shaping the gene pool of the extant southern TB populations.

Material and Methods

Samples

Blood samples of 622 unrelated anonymous individuals from 15 TB populations were collected in the Yunnan, Qinghai, and Hunan provinces of China. Genomic DNA was extracted by the phenol-chloroform method. The additional data were obtained from published reports on the Y chromosome (Su et al. 1999, 2000*b;* Qian et al. 2000; Karafet et al. 2001) and on mtDNA variation (Qian et al. 2001; Yao et al. 2002*a*, 2002*b*). These led to the final sample sizes for analysis, expanding them to 965 individuals (23 TB populations) for the Y chromosome and 754 individuals (21 TB populations) for mtDNA. These samples encompass most of the TB populations in China. Throughout this article, we refer the TB populations (not including Tibetans) in Yunnan, Sichuan, and Hunan as "southern Tibeto-Burmans" (STBs).

In addition, we used Y-chromosome data from 4 Austro-Asiatic–, 30 Daic-, and 23 Hmong-Mien–speaking populations and from 17 northern Han populations (Su et al. 1999, 2000*a,* 2000*b;* Karafet et al. 2001; L.J., unpublished data), as well as mtDNA data from 4 Austro-Asiatic–, 12 Daic-, 12 Hmong-Mien–speaking populations and from 10 northern Han populations (Kolman et al. 1996; Qian et al. 2001; Yao et al. 2000, 2002*a,* 2002*b,* 2002*c;* L.J., unpublished data). The detailed information on the populations studied, including their linguistic affinities, geographic distribution, and data sources, are given in table 1 and figure 1.

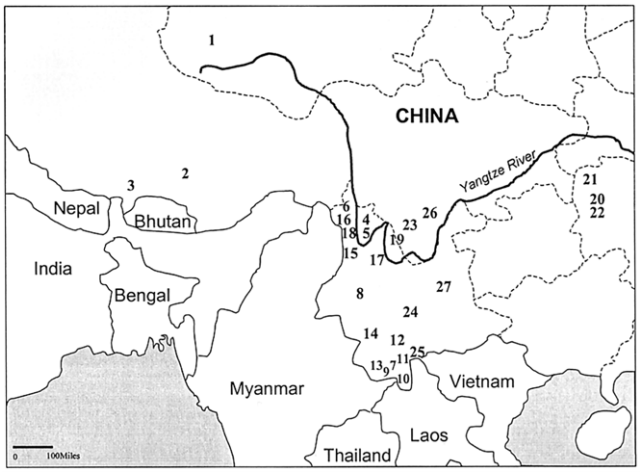

Figure 1 Geographic locations of TB populations sampled. Population numbers are defined in table 1.

Y-Chromosome Markers

Ten biallelic Y-chromosome markers—YAP, M15, M130, M89, M9, M122, M134, M119, M95, and M45—were typed by PCR-RFLP methods (Su et al. 2000*b*). These markers are highly informative in East Asians (Jin and Su 2000) and define 10 haplogroups, following the Y Chromosome Consortium (2002) nomenclature.

mtDNA Markers

The HVS-1 region of mtDNA was amplified by primers L15974 and H16488 (Yao et al. 2002*a*). Purified PCR products were sequenced using the BigDye terminator cycle sequencing kit and an ABI 3100 genetic analyzer (Applied Biosystems). Primers were designed for amplifying multiple fragments that contain haplogroup diagnostic polymorphisms in the coding regions. PCR products were then digested by restriction enzymes: 10397 *Alu*I, 5176 *Alu*I, 4831 *Hba*I, 13259 *Hin*cII, 663 *Hae*III, 12406 *Hpa*I, and 9820 *Hin*fI (Kivisild et al. 2002; Yao et al. 2002*a*). Primer sequences and PCR-RFLP conditions are available from L.J. on request. Both the HVS-1 motif and the coding region variations were used to infer haplogroups according to the classification of Kivisild et al. (2002). The HVS-1 sequences of 496 individuals from 14 TB populations have been submitted to GenBank (accession numbers AY397784–AY398279).

Data Analysis

Principal-component analysis (PCA) was conducted using mtDNA and Y-chromosome haplogroup frequen-

Table 1

Information for populations examined

Population[a]	Language Branch[b]	Location	Y Chromosome		mtDNA	
			n	Reference	*n*	Reference
Tibetan:						
1. Tibetan1	Himalayish	Qinghai	91	Present study	56	Present study
2. Tibetan2	Himalayish	Tibet	75	Hammer et al. 2001		
3. Tibetan3	Himalayish	Tibet	46	Qian et al. 2000		
4. Tibetan4	Himalayish	Diqing, Yunnan	27	Qian et al. 2000	24	Qian et al. 2001
5. Tibetan5	Himalayish	Zhongdian, Yunnan	49	Present study	35	Present study
6. Tibetan6	Himalayish	Deqing, Yunnan			40	Yao et al. 2002*a*
STB:						
7. Aini	Lolo-Burmese	Xishuanbanna, Yunnan	52	Present study	50	Present study
8. Bai1	Bai	Dali, Yunnan	63	Su et al. 2000*a;* present study	68	Yao et al. 2002*b;* present study
9. Bai2	Bai	Xishuanbanna, Yunnan	20	Present study	19	Present study
10. Hani	Lolo-Burmese	Xishuanbanna, Yunnan	34	Present study	33	Present study
11. Jino	Lolo-Burmese	Xishuanbanna, Yunnan	36	Su et al. 2000*b;* present study	18	Present study
12. Lahu1	Lolo-Burmese	Simao, Yunnan	13	Su et al. 2000*b*	32	Qian et al. 2001
13. Lahu2	Lolo-Burmese	Xishuanbanna, Yunnan	15	Present study	15	Present study
14. Lahu3	Lolo-Burmese	Lancang, Yunnan			35	Qian et al. 2001
15. Lisu1	Lolo-Burmese	Fugong, Yunnan	49	Present study		
16. Lisu2	Lolo-Burmese	Gongshan, Yunnan			37	Yao et al. 2002*b*
17. Naxi	Lolo-Burmese	Lijiang, Yunnan	40	Present study	45	Present study
18. Nu	Lolo-Burmese	Gongshan,Yunnan	28	Present study	30	Present study
19. Pumi	Tangut-Qiang	Ninglang, Yunnan	47	Present study	36	Present study
20. Tujia1	Tujia	Western Hunan	68	Present study	64	Present study
21. Tujia2	Tujia	Yongshun, Hunan	38	Present study	30	Present study
22. Tujia3	Tujia	Jishou, Hunan	49	Hammer et al. 2001		
23. Yi1	Lolo-Burmese	Liangshan, Sichuan	14	Su et al. 1999		
24. Yi2	Lolo-Burmese	Shuangbai, Yunnan	50	Present study	40	Present study
25. Yi3	Lolo-Burmese	Xishuanbanna, Yunnan	18	Present study	16	Present study
26. Yi4	Lolo-Burmese	Butuo, Sichuan	43	Hammer et al. 2001		
27. Yi5	Lolo-Burmese	Luxi, Yunnan			31	Present study

[a] Numbering of populations is that used in figure 1 and table 3.
[b] According to the Ethnologue Web site.

cies and SPSS10.0 software (SPSS). Results of PCA are presented by the plots of the first two PCs, which together account for 65.8% of the Y-chromosome and 51.0% of the mtDNA variation in these populations. The genetic structure of populations was investigated by the analysis-of-molecular-variance approach (AMOVA [Excoffier et al. 1992]), using Arlequin software (Schneider et al. 2000; Arlequin's Home on the Web). Detailed grouping designs are listed in table 4.

Admix 2.0 (Dupanloup and Bertorelle 2001; Admix Web site) and LEADMIX (Wang 2003) software was used to estimate the level of admixture of southern and northern groups in the STB populations, using the methods of Bertorelle and Excoffier (1998) (hereafter referred to as "BE") and Roberts and Hiorns (1965) (hereafter referred to as "RH"), respectively. These two estimators are less biased for single-locus data, as indicated by the simulation results of Wang (2003). The selection of parental populations is critical for appropriate estimation of admixture proportions (Chakraborty 1986, Sans et al. 1997), and we paid special attention to make them as unbiased as possible by using large data sets across East Asia and by taking historical records into account. In this analysis, the average haplogroup frequencies (for Y-chromosome or mtDNA markers, respectively) of Tibetan (arithmetic mean of Tibetan populations in Tibet, Qinghai, and northwestern Yunnan) and northern Han populations were taken for the northern parental population (northern East Asians [NEAs]), and the average haplogroup frequencies of Austro-Asiatic, Daic, and Hmong-Mien were taken for the southern parental population (southern East Asians [SEAs]).

Results

Distribution of Y-Chromosome and mtDNA Haplogroups

The Y-chromosome haplogroup frequency distributions of the TB populations are presented in table 2. Almost all TB populations, except for the Naxi and Pumi, showed high frequencies of O3* and O3e haplogroups, both carrying the M122-C mutation (haplogroup frequencies ranging from 20% to 86% with an average of 39.5% in TBs and 40.7% in STBs), as previously observed by Su et al. (2000*b*). The YAP+ hap-

Table 2

Y-SNP Haplogroup Frequency in TB Populations

		No. of Occurrences of Mutation/Haplogroup									
Population	Size	M130 C	YAP D*	M15 D1	M89 F*	M9 K*	M122 O3*	M134 O3e	M119 O1	M95 O2a	M45 P*
Tibetan:											
Tibetan1	92	13	19	2	13	20	5	13	1		6
Tibetan2	75	2	25	12	2	4	1	24			5
Tibetan3	46	4	11	8	2		2	16		1	2
Tibetan4	27		12		1	4	2	8			
Tibetan5	49	1	14	4	1	5	5	17		2	
STB:											
Bai1	61	5	1	3		11	10	21	3	7	
Lisu1	49				1	11	2	30		4	1
Naxi	40	1	15			3		1		20	
Nu	28		1			1	4	20		2	
Pumi	47	3	33	1		3	1	3	2		1
Yi1	14			2		6	3	1		2	
Yi2	50	4	1	5	19	8	5	5	1	2	
Yi4	43	1		7	2	15	2	12		4	
Aini	52	6			1	18	14	7	2	4	
Bai2	20	4				6	5	3	2	0	
Hani	34	4				12	11	5	1	1	
Jino	36	5			2	13	7	7		2	
Lahu1	13	2			4	2	2	2		1	
Lahu2	15				1	3	5	1	3	2	
Yi3	18	2			1	6	5	3		1	
Tujia1	68	10			2	7	20	18	5	6	
Tujia2	38	2		1		9	15	4	6		1
Tujia3	49	12		1		4	15	11	4		2
Total	964	81	132	46	52	171	141	232	30	61	18

logroups D* and D1, which are highly frequent in the Tibetan and Japanese populations, are prevalent in some populations in northwestern Yunnan (38% in the Naxi and 70.2% in the Pumi versus 15.7% in TBs and 9.8% in STBs). In contrast, haplogroups O1-M119 and O2a-M95, which are prevalent in SEAs but very rare in the north (Su et al. 1999; Karafet et al. 2001), were found in most of the TB populations, but at low frequencies (average 10.5% in TBs and 12.9% in STBs), except in the Naxi, which has high frequency of O2a-M95. However, haplogroups F*-M89 and P*-M45, which are predominant in Central Asia and northeastern Asia, were found in only some of the TB populations, at low frequencies (6.7% in TBs and 6.1% in STBs), except for F*-M89 in Yi2 (38%) and Lahu1 (31%). In general, the distribution of Y-chromosome haplogroups in the TB populations is more similar to that in the NEAs than it is to their southern neighbors such as the Austro-Asiatic, Daic, and Hmong-Mien.

In contrast, the distribution of mtDNA haplogroups in the TB populations is more complex. Almost all (sub)haplogroups found in East Asia are present in the TB populations (table 3). A (11%), B (12%), D (18%), F (20%), and M* (18%) are the predominant haplogroups in the TB populations, accounting for 77.6% and 77.7% of the total mtDNA lineages in TBs and STBs, respectively. In particular, haplogroup D is highly frequent in most TB populations, whereas the others show differential distributions across geographic regions. Frequencies of A and M* are higher in Qinghai (A, 21%; M*, 36%) and northwestern Yunnan (A, 14%; M*, 22%) than in southern Yunnan (A, 5%; M*, 12%) and Hunan (A, 9%; M*, 4%). Haplogroup A is one of the more prevalent mtDNA haplogroups in NEAs and Siberians (Kivisild et al. 2002), whereas M* might be a mixture of several M lineages whose specific subhaplogrouping is yet to be ascertained. On the basis of the north-to-south decrease in its frequencies, most of the TB M* mtDNAs might belong to a single unknown lineage. Haplogroup B and F are prevalent in Southeast Asia (Kivisild et al. 2002; Yao et al. 2002*a*). In the TB populations, these two lineages (B and F) are more frequent in southern Yunnan (B, 18%; F, 30%) and Hunan (B, 17%; F, 21%) than in Hingham (B, 2%; F, 6%) and northwest Yunnan (B, 10%; F, 15%).

Population Clustering as Revealed by PCA

The result of PCA of Y-chromosome haplogroup frequencies is presented in figure 2. NEAs (northern Han and Tibetans) and SEAs (Austro-Asiatic, Daic, and Hmong-Mien populations) show significantly different

Table 3

mtDNA Haplogroup Distribution in TB Populations Examined

	No. in Population[a]																				
Haplogroup	1	4	5	6	7	8	9	10	11	12	13	14	16	17	18	19	20	21	24	25	27
A	12	6	4	2	1	4	1	4	1				11	4	9	5	9		4	3	3
B*					2																1
B4*			2	1	1		2	1		3	1	1	1	3	3	1	2	2	5	1	2
B4a	1				4		1	1	1	5	1			5			5		1	1	
B4b1						1		1			1								1		
B5*													1							1	
B5a			4		1	2	1	2	1					3			2	2		1	
B5b								1									3				
C	3		2		1	2		1			1		5	4		8	8	1	1		2
D*	4		3	20	7	9	3	6		2	4	5	3	2	6	6	9	3	4	3	8
D5				2	3	4		2	1	1							1	2	1		2
D5a					2			1									4		1		
F*					1	1	1					9					2	6	2		1
F1a	1	1			6	6	2	6	2	10	3	18	1	8			2	4	2	1	
F1b		1				2	2	1	1		1		2	2		1	2	1	2	2	3
F1c	2					4					1						1	2			
F2a			1			4				1			4	1	5	2				1	
G	2													3	1	2	1	1	1		1
G2	1					5											1	1	1		
G2a	2				2																
G3	1		2													1	1				
M*	20	6	15	12	7	15	3	4	5	6		1	7	7	1	7	3	1	8	1	3
M7*						1												1			
M7b*					1	2					1						1		2		
M7b1					4	2	1		3					1			1	1	3	1	2
M8a					1						1								1		
M9									1												
M9a	5	4	1	3		3	1										1	1			1
N*	1				1	2											3				2
N9a	1								2							1	2	2			
R10					1	2		1		4											
R9a							1					1		1			2				
Y								1													
Z		6	1		1								2	1	5	1					

[a] Population numbers are defined in table 1.

distributions of PC2. The PC2 values are -0.41 ± 0.05 (range -0.95 to 0.31) for the northerners and 0.39 ± 0.02 (range 0.22 to 0.63) for the southerners. The STB populations were positioned between them, with the PC2 values being 0.23 ± 0.03 (range -0.04 to 0.43) after removing one outlier (Naxi, PC2 = -0.71), supporting the historical accounts that these populations are admixed.

However, PCA results of mtDNA haplogroups showed a different picture (fig. 3). SEAs and NEAs are still separated by PC2, but the distribution of the STB populations is more scattered into both the southern and northern groups. This observation with mtDNA is also consistent with the admixed nature of these populations, with contributions from both NEAs and SEAs.

Evaluation of Hierarchical Structure, using AMOVA

AMOVA was employed to evaluate genetic differentiation between STBs and other East Asian populations and to examine genetic structure within the STB group (table 4). For the Y chromosome, the F_{ct} (between-group divergence) between STBs and SEAs is significant ($P = .002$), whereas the divergence between STBs and northern populations (NEAs) is not ($P = .221$). The F_{ct} between the STBs and SEAs is >10 times as high as that between the STBs and NEAs (0.029 versus 0.002), indicating that the STBs bear a stronger similarity in their parental lineages to the NEAs than to the SEAs. However, the F_{ct} values of STB/north and STB/south are both significant and almost equal those for mtDNA (0.005 versus 0.007; P values .003 and <.001, respectively), which reflects a similar level of divergence between the STB and the two East Asian groups. In the TB populations, between-population divergence (F_{st}) for the Y chromosome is about twofold higher than that of mtDNA (0.115 vs. 0.057), suggesting a more extensive between-population differentiation in female lineages than in their male counterparts in the patrilocal popu-

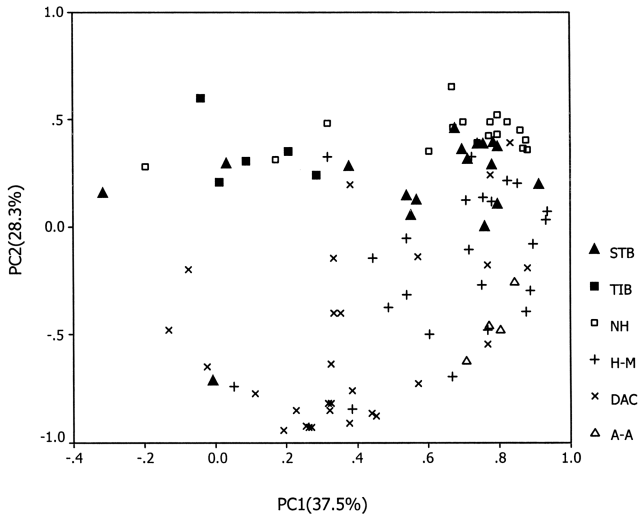

Figure 2 PC plots of Y-chromosome SNP haplogroup frequencies. Abbreviations for populations are as follows: STB = southern Tibeto-Burman; TIB = Tibetan; NH = northern Han; H-M = Hmong-Mien; DAC = Daic; A-A = Austro-Asiatic.

lations. This is consistent with previous observations in global populations (Seielstad et al. 1998), although the extent of contrast is less drastic in the TBs.

Estimation of Admixture in STBs

Table 5 presents the proportion of the contribution of SEAs in the different STB populations, based on both mtDNA and Y-chromosome markers, by using two different approaches. On the basis of either the Y chromosome or mtDNA, the contributions of the SEA gene pool (*M*) in the different STB populations vary. The estimates of *M* when the two different methods are used are fairly consistent (Pearson's correlation 0.931 for Y chromosomes [$P < .01$] and 0.928 for mtDNA [$P < .01$]). In table 5, *M* is listed as 0 if it is negative and as 1 if it is >100%.

In the STB populations—except for the Nu and Naxi, in which both Y-chromosome and mtDNA data are available—the contribution of SEAs is higher in the female lineages than in the male lineages (table 5). In other words, in the extant STB populations, the southern natives (i.e., the SEAs) made more of a contribution to the female lineages ($M_{BE} = 55.5\%$; $M_{RH} = 60.1\%$) than to the male lineages ($M_{BE} = 38.1\%$; $M_{RH} = 39.6\%$). Alternatively, the northern immigrants made more of a contribution to the male lineages (~61.9%) than to the female lineages (~44.5%). However, it is important to point out that the bias of male and female contributions is likely underestimated and will be discussed in the next section. The skewed distributions of Y-chromosome haplotypes in the Nu and Naxi (i.e., the extremely high frequency of O3e-M134 in the Nu and of O2a-M95 in the Naxi), probably due to strong bottleneck events, may explain the exceptional trend of their admixture estimates.

Conclusions and Discussion

By examining the haplotype group distributions of Y-chromosome and mtDNA markers and their principle components, we have shown that the genetic structure of the extant STB populations was primarily shaped by two parental groups: northern immigrants and native southerners (tables 2 and 3). Significant divergence has developed between the STBs and their parental populations for mtDNA and between the STBs and the southern natives for the Y chromosome, whereas the difference between the STBs and the northern immigrants remains insignificant (table 4). Furthermore, we showed that the admixture has a bias between male and female lineages, since, in the extant STBs, there is a stronger influence of northern immigrants on the male lineages, and the southern natives contribute more extensively to the female lineages (table 5).

Unequal admixture proportions of males and females have been previously observed in Native Americans and African Americans (Parra et al. 1998, 2001; Bortolini et al. 1999; Mesa et al. 2000; Sans et al. 2002). This bias was explained as being a result of "directional mating" in the last several centuries (Merriwether et al. 1997; Chakraborty 1998), which means that the influx of genes from ancestral to hybrid populations is directional. For example, admixture in Native American populations of South America mainly involved European men and native woman, owing to the asymmetric

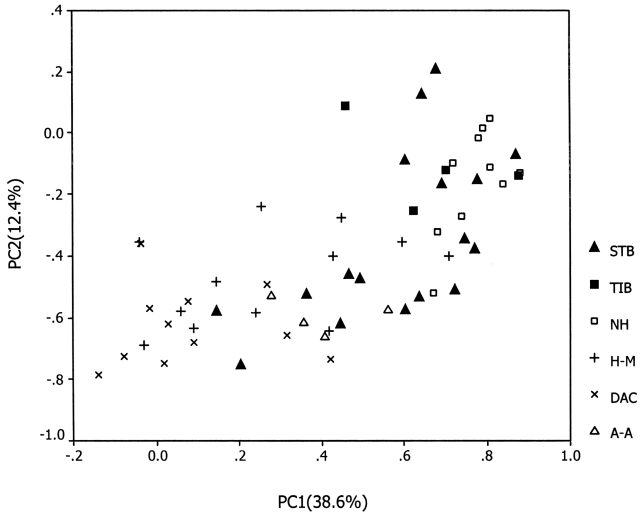

Figure 3 PC plots of mtDNA haplogroup frequencies. The population codes are defined in figure 2.

Table 4

AMOVA Results

Grouping	No. of Groups	No. of Populations	F_{ct} (P)[a]	F_{cs}(P)[a]	F_{st} (P)[a]
mtDNA:					
STB/NEA	2	31	.005 (.003)	.035 (.000)	.040 (.000)
STB/SEA	2	45	.007 (.000)	.035 (.000)	.042 (.000)
STB/NEA/SEA	3	59	.010 (.000)	.033 (.000)	.043 (.000)
TB as one group	1	21	…	…	.057 (.000)
Y chromosome:					
STB/NEA	2	40	.002 (.221)[b]	.084 (.000)	.086 (.000)
STB/SEA	2	77	.029 (.002)	.136 (.000)	.161 (.000)
STB/NEA/SEA	3	98	.035 (.000)	.110 (.000)	.141 (.000)
TB as one group	1	23	…	…	.115 (.000)

[a] The P values of F statistics were obtained by 3,000 permutations, under a hypothesis of no population structure.

[b] STB/NEA is the only comparison with an F value that is not significant; all others are statistically significant.

sex ratio of European colonizers and other political reasons (Carvajal-Carmona et al. 2000; Sans 2000).

In the STB populations, the sex-biased admixture pattern is different from that of America, not only in the extent of bias but also, probably, in the mechanism. In the admixed populations derived from American natives and European colonists, the contribution of native mtDNAs is almost 10 times higher than that of native Y chromosomes (Santos et al. 1999; Carvajal-Carmona et al. 2000), whereas the ratio of southern native mtDNA and Y-chromosome contribution in the STB populations is only ~1.5, which is much less drastic than that found in the present-day Native Americans. According to historical records, the southward migration of the ancestral TB populations began at ~2,600 years BP, because of the expansion of the Qing Kingdom (Wang 1994; Cang 1997). However, little evidence has indicated an asymmetry of the male/female ratio in the TB immigrants in historical literature. The less drastic bias between male and female lineages in the TBs may suggest that the southward migrations of the TBs likely occurred with the involvement of both sexes rather than as conquests involving expedition forces primarily consisting of male soldiers, as occurred in the American continents. Unfortunately, little description of the migration can be found in the historical record, largely because of the Sino-centric nature of the available historical literature.

The study of admixture patterns may shed light on our understanding of the historic and anthropological aspects of migrations of TB populations. For example, the variation of the admixture in different geographic regions may be associated with the variation of the time of the arrival of the immigrants and the ethnic constitution of the parental populations in the regions when admixture started to occur. According to historical records, the STB populations in southern Yunnan and Hunan (Aini, Hani, Jino, Lahu, and Tujia) were more ancient, whereas the immigration of the TBs to northwestern Yunnan (Tibetan, Lisu, Nu, and Pumi) were more recent, except for Bai, Naxi, and Yi, which were dominating native populations in Yunnan (Cang 1997). Accordingly, the contribution of the southern natives in the TBs in northwestern Yunnan are relatively smaller, except for Naxi and Bai, than they are in southern Yunnan and Hunan (table 5), areas dominated by the Daic, Hmong-Mien, and Austro-Asiatic populations. In fact, Tibetans are the most recent immigrants to this region (Cang 1997), and no contribution from southern natives was observed in them in this study (data not shown). Therefore, our genetic observation is consistent with historical records of the TB populations.

The observations made in this study may also shed light on the genetic structure of the individual populations studied. For instance, the levels of admixture are very different in the Yi populations in different geographic areas, indicating high heterogeneity of this ethnic group. This is consistent with the historical and linguistic observations that Yi encompasses a number of distinctive branches. However, for the Lahu and Tujia, the levels of admixture are similar across different regional populations, suggesting a genetic homogeneity of these ethnic groups. The Naxi have the highest contribution of southern natives for mtDNA in northwestern Yunnan and for the Y chromosome in all the TBs studied. This reflects a high level of genetic input from southern populations in its history, and this has not been well recognized by historians and anthropologists.

The estimations of the levels of admixture for the Y chromosome and mtDNA may not be reliable, because of the large variances of the estimations based on a single-locus system. However, three lines of evidence indicate that our conclusions are valid. First, two different statistics were used to estimate the contribution

Table 5

Admixture of SEA in STB Populations

	$M_{BE} \pm SE$[a]		M_{RH}[a]	
Population	mtDNA	Y Chromosome	mtDNA	Y Chromosome
Aini	.755 ± .234	.511 ± .100	.885	.521
Bai1	.878 ± .219	.542 ± .134	.849	.429
Bai2	.461 ± .373	.377 ± .096	.700	.369
Hani	.530 ± .271	.399 ± .079	.725	.424
Jino	.403 ± .408	.327 ± .103	.795	.363
Lahu1	1	.123 ± .163	1	.078
Lahu2	1	.677 ± .186	1	.831
Lahu3	1	ND	1	ND
Lisu1	ND	.200 ± .140	ND	.169
Lisu2	0	ND	0	ND
Naxi	.713 ± .252	.899 ± .281	.729	1
Nu	0	.096 ± .173	0	.159
Pumi	0	0	0	0
Tujia1	.688 ± .183	.544 ± .099	.520	.498
Tujia2	1	.530 ± .081	.914	.449
Tujia3	ND	.556 ± .089	ND	.360
Yi1	ND	.501 ± .199	ND	.681
Yi2	.327 ± .262	.133 ± .110	.382	.121
Yi3	.385 ± .395	.321 ± .099	.491	.419
Yi4	ND	.129 ± .114	ND	.258
Yi5	.295 ± .275	ND	.218	ND
Average	.555	.381	.601	.396

[a] Southern admixture proportion (M) is estimated by method of Bertorelle and Excoffier (1998) (M_{BE}) and Roberts and Hiorns (1965) (M_{RH}). SEs for the BE method were obtained by 1,000 bootstraps. ND = data not available.

of the SEA, and they are consistent with each other (with a correlation coefficient of 0.93 for both the Y chromosome and mtDNA). Second, the relative magnitudes of the estimations of M (1 for mtDNA and 1 for the Y chromosome) are consistent for both methods in all 14 populations in which data on both the Y chromosome and mtDNA are available. Third, the probability of observing higher contributions of female lineages than contributions of male lineages in 11 of 14 populations is 0.01, if we assume a null binomial distribution with equal male and female contribution. Furthermore, in the AMOVA analyses, the lack of significant differentiation between the STBs and their northern progenitors at the Y chromosome ($F_{ct} = 0.002$; $P = .221$) is also consistent with this conclusion.

It should also be noted that the estimated frequencies of haplotype groups in the parental populations might not be accurate. In estimating the level of admixture of the STB populations, the frequencies of haplotype groups of the parental populations, for both the Y chromosome and mtDNA, were estimated by taking the arithmetic mean of the frequencies of haplotype groups from individual populations. We did not try to compute a weighted average by taking the population size into account, which is important but practically impossible to achieve, because of the drastic fluctuation in size during the history of these populations. However, this simple treatment should not affect the difference of the contributions of male and female lineages we observed, since any scheme of weighting would have the same impact on both male and female lineages. The extant SEA populations (one of the parental populations) may have also been affected, more so in the male lineages than in the females, by the northern immigrants (data not shown). This would lead to an even more drastic bias between males and females in the STBs. Therefore, the sex-bias admixture in the STBs may, in fact, be more pronounced than what is shown in this study (table 5).

In conclusion, we have provided evidence for a general south/north admixture, and, more interestingly, asymmetric contributions of male and female parental lineages in the extant STB populations. The STB populations have preserved more male lineages of the northern immigrants and thus are genetically closer to the NEAs than they are to the SEAs. In contrast, more southern mtDNA lineages are in the gene pool of the extant STB populations. Although the admixture with a strong sex bias found in America is rare in other global human populations, lesser bias could be frequent, as exemplified in the STBs.

Acknowledgments

We thank all of the donors for making this work possible. This work is partially supported by Natural Science Foun-

dation of China grant 39993420. L.J. is also supported by National Science Foundation grant BCS-0213857, and L.J. and R.C. are supported by National Institutes of Health grant GM 41399.

Electronic-Database Information

Accession numbers and URLs for data presented herein are as follows:

Admix 2.0, http://www.unife.it/genetica/Isabelle/admix2_0.html
Arlequin's Home on the Web, http://lgb.unige.ch/arlequin/
Ethnologue, http://www.ethnologue.com/ (for Ethnologue languages database, 14th edition)
GenBank, http://www.ncbi.nlm.nih.gov/Genbank/ (for mtDNA HVS-1 sequences: accession numbers AY397784–AY398279)

Reference

Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 3:299–309

Bertorelle G, Excoffier L (1998) Inferring admixture proportions from molecular data. Mol Biol Evol 15:1298–1311

Bortolini MC, Da Silva WA Junior W, De Guerra DC, Remonatto G, Mirandola R, Hutz MH, Weimer TA, Silva MC, Zago MA, Salzano FM (1999) African-derived South American populations: a history of symmetrical and asymmetrical matings according to sex revealed by bi- and uni-parental genetic markers. Am J Human Biol 11:551–563

Cang M (1997) Study on the migration culture of the ethnic groups in Yunan. Yunnan Nationality Press, Kunming (in Chinese)

Carvajal-Carmona LG, Soto ID, Pineda N, Ortiz-Barrientos D, Duque C, Ospina-Duque J, McCarthy M, Montoya P, Alvarez VM, Bedoya G, Ruiz-Linares A (2000) Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. Am J Hum Genet 67:1287–1295

Chakraborty R (1986) Gene admixture in human populations: models and predictions. Yearb Phys Anthropol 29:1–43

——— (1998) Molecular evidence of directional mating for gene migration in human populations. Abstract from the XVIIIth International Congress of Genetics, Beijing, August 10–15, 1998, p 102

Chakraborty R, Weiss KM (1988) Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc Natl Acad Sci USA 85:9119–9123

Di Benedetto G, Erguven A, Stenico M, Castri L, Bertorelle G, Togan I, Barbujani G (2001) DNA diversity and population admixture in Anatolia. Am J Phys Anthropol 115: 144–156

Dupanloup I, Bertorelle G (2001) Inferring admixture proportions from molecular data: extension to any number of parental populations. Mol Biol Evol 18:672–675

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

Gresham D, Morar B, Underhill PA, Passarino G, Lin AA, Wise C, Angelicheva D, Calafell F, Oefner PJ, Shen P, Tournev I, de Pablo R, Kucinskas V, Perez-Lezaun A, Marushiakova E, Popov V, Kalaydjieva L (2001) Origins and divergence of the Roma (gypsies). Am J Hum Genet 69:1314–1331

Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura SL (2001) Hierarchical patterns of global human Y-chromosome diversity. Mol Biol Evol 18:1189–1203

Helgason A, Hickey E, Goodacre S, Bosnes V, Stefansson K, Ward R, Sykes B (2001) mtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. Am J Hum Genet 68:723–737

Jin L, Su B (2000) Natives or immigrants: modern human origin in East Asia. Nat Rev Genet 1:126–133

Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL, Hammer MF (2001) Paternal population history of east Asia: sources, patterns, and microevolutionary process. Am J Hum Genet 69:615–628

Kivisild T, Tolk H-V, Parik J, Wang Y, Papiha SS, Bandelt H-J, Villems R (2002) The emerging limbs and twigs of the East Asian mtDNA tree. Mol Biol Evol 19:1737–1751

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

Merriwether DA, Huston S, Iyengar S, Hamman R, Norris JM, Shetterly SM, Kamboh MI, Ferrell RE (1997) Mitochondrial versus nuclear admixture estimates demonstrate a past history of directional mating. Am J Phys Anthropol 102:153–159

Mesa NR, Mondragon MC, Soto ID, Parra MV, Duque C, Ortiz-Barrientos D, Garcia LF, Velez ID, Bravo ML, Munera JG, Bedoya G, Bortolini MC, Ruiz-Linares A (2000) Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. Am J Hum Genet 67:1277–1286

Parra EJ, Kittles RA, Argyropoulos G, Pfaff CL, Hiester K, Bonilla C, Sylvester N, Parrish-Gause D, Garvey WT, Jin L, McKeigue PM, Kamboh MI, Ferrell RE, Pollitzer WS, Shriver MD (2001) Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. Am J Phys Anthropol 114:18–29

Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD (1998) Estimating African American admixture proportions by use of population-specific alleles. Am J Hum Genet 63: 1839–1851

Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. Am J Hum Genet 69:1–14

Qian Y, Qian B, Su B, Chu Z, Ke Y, Yu J, Shi L, Chu J, Lu D, Jin L (2000) Multiple origins of Tibetan Y chromosomes. Hum Genet 106:453–454

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

Roberts DF, Hiorns RW (1965) Methods of analysis of the genetic composition of a hybrid population. Hum Biol 37: 38–43

Sans M (2000) Admixture studies in Latin America: from the 20th to the 21st century. Hum Biol 72:155–177

Sans M, Salzano FM, Chakraborty R (1997) Historical genetics in Uruguay: estimates of biological origins and their problems. Hum Biol 69:161–170

Sans M, Weimer TA, Franco MH, Salzano FM, Bentancor N, Alvarez I, Bianchi NO, Chakraborty R (2002) Unequal contributions of male and female gene pools from parental populations in the African descendants of the city of Melo, Uruguay. Am J Phys Anthropol 118:33–44

Santos FR, Pandya A, Tyler-Smith C, Pena SD, Schanfield M, Leonard WR, Osipova L, Crawford MH, Mitchell RJ (1999) The central Siberian origin of native American Y chromosomes. Am J Hum Genet 64:619–628

Schneider S, Kueffer J-M, Roessli D, Excoffier L (2000) Arlequin version 2.000: a software for population genetic analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva

Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. Nat Genet 20:278–280

Smith MW, Lautenberger JA, Shin HD, Chretien JP, Shrestha S, Gilbert DA, O'Brien SJ (2001) Markers for mapping by admixture linkage disequilibrium in African American and Hispanic populations. Am J Hum Genet 69:1080–1094

Su B, Jin L, Underhill P, Martinson J, Saha N, McGarvey ST, Shriver MD, Chu J, Oefner P, Chakraborty R, Deka R (2000*a*) Polynesian origins: new insights from the Y-chromosome. Proc Natl Acad Sci USA 97:8225–8228

Su B, Xiao C, Deka R, Seielstad MT, Kangwanpong D, Xiao J, Lu D, Underhill P, Cavalli-Sforza LL, Chakraborty R, Jin L (2000*b*) Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. Hum Genet 107:582–590

Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza L, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z, Jin L (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

Wang J (2003) Maximum-likelihood estimation of admixture proportions from genetic data. Genetics 164:747–765

Wang ZH (1994) History of nationalities in China. China Social Science Press, Beijing (in Chinese)

Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res 12:339–348

Yao Y-G, Kong Q-P, Bandelt H-J, Kivisild T, Zhang Y-P (2002*a*) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet 70:635–651

Yao Y-G, Lu X-M, Luo H-R, Li W-H, Zhang Y-P (2000) Gene admixture in the silk road of China: evidence from mtDNA and melanocortin 1 receptor polymorphism. Genes Genet Syst 75:173–178

Yao Y-G, Nie L, Harpending H, Fu YX, Yuan ZG, Zhang Y-P (2002*b*) Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. Am J Phys Anthropol 118:63–76

Yao Y-G, Zhang Y-P (2002*c*) Phylogeographic analysis of mtDNA variation in four ethnic populations from Yunnan province: new data and a reappraisal. J Hum Genet 47:311–318