AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 134:481–488 (2007)

Mitochondrial DNA Diversity and Population Differentiation in Southern East Asia

Hui Li,[1,2] Xiaoyun Cai,[1] Elizabeth R. Winograd-Cort,[2] Bo Wen,[1] Xu Cheng,[1] Zhendong Qin,[1] Wenhong Liu,[1] Yangfan Liu,[1] Shangling Pan,[3] Ji Qian,[1] Chia-Chen Tan,[1] and Li Jin[1,4]*

[1]*MOE Key Laboratory of Contemporary Anthropology and Center for Evolutionary Biology, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai 200433, China*
[2]*Department of Genetics, School of Medicine, Yale University, New Haven, CT 06520*
[3]*Department of Pathophysiology, Guangxi Medical University, Nanning 530021, China*
[4]*Center for Genome Information, Department of Environmental Health, University of Cincinnati, Cincinnati, OH 45267*

KEY WORDS mtDNA; Daic; Austro–Asiatic; ethnic differentiation; East Asia

ABSTRACT Mitochondrial DNA (mtDNA) polymorphism has been studied systematically in the Han, Tibeto–Buman, and Hmong–Mien ethnic families of southern East Asia. Only two families in this region, Daic and Austro–Asiatic, were still uninvestigated. Daic is a major ethnic family in South China and Southeast Asia and has a long history. To study mtDNA polymorphism within this family, all the Daic populations of China and some of Vietnam (774 individuals from 30 populations) were typed by HVS-1 region sequencing and by PCR-RFLP assays. The observed high Southern type frequencies (B, F, M7, R) confirmed Daic as a typical Southern group. mtDNAs of other populations (126 individuals from 14 populations) from Austro–Asiatic ethnic families neighboring the Daic were also typed. Networks of mtDNA haplogroups in South China were traced from these new data and those from the literature. Ethnic families share many haplogroups, indicating their common origin. However, the two largest families in South China, Daic, and Hmong-Mien, polarized into several ethnic family specific haplogroups. Haplogroup ages were estimated in the networks of high-frequency haplogroups (B, F, M7, R), and they were found to originate about 50,000 years ago. In contrast, ethnic family specific haplogroups all originated around 20,000 years ago. We therefore conclude that modern humans have lived in South China for a long time, inside-ethnogenesis was a rather late event, and frequent inmixing was taking place throughout. MtDNA data of Daic, Austro-Asiatic and other populations in South China has therefore proven pivotal for studying the human history of East Asia. Am J Phys Anthropol 134:481–488, 2007.

Modern humans have lived in South China for at least 30,000 years, since the age of the *Liujiang* Man, discovered in Guangxi, China, in 1958 (Wu, 1959). Genetic studies show that populations in North China derived from populations in South China (Su et al., 1999). A large number of anthropological studies suggested that South China is the motherland of many ethnic groups throughout East Asia and the Pacific. There have been migrations from South China to Southeast and Northeast Asia and other neighboring areas during prehistory and history. The most famous among these migrations was the migration of Austronesian populations, which are believed to have left the coast of South China about 6,000 years before the present (kybp) and spread throughout the Pacific (Zhang, 1987; Diamond, 1988; Bellwood et al., 1995). The genetic structure of populations in this area is, therefore, pivotal in forming of ethnic patterns in east Eurasia.

There are many ethnic families besides Han Chinese in South China, such as the Daic, Hmong–Mien (HM), Tibeto–Burman (TB), Austro–Asiatic (AA), Taiwan aborigines (TA) and the Malayo–Polynesians (MP). Among these populations, the Daic (also called *Baiyue* in Chinese) is the largest and most widespread. The total population of the nine official Daic nationalities in China is 25.8 million (2000 census). However, the population of the Daic descendants assimilated by Han Chinese is most probably even larger than this number (Song, 1991). Moreover, there are about 80 million (Grimes, 2002) Daic in Thailand, Laos, Vietnam, Myanmar, and India who migrated from South China. So in size, the Daic population is second only to the Han, and is far larger than any other ethnic family in South China. Until the arrival of the Han (Wen et al., 2004a), the Daic were the aboriginal ethnic group of China's southeast coastal zone, which extends from Shanghai to Hanoi

This article contains supplementary material available via the Internet at http://www.interscience.wiley.com/jpages/0002-9483/suppmat.

Grant sponsors: National Natural Science Foundation of China (NSFC), the Science and Technology Commission of Shanghai Municipality (STCSM), NSF.

*Correspondence to: Li Jin, School of Life Sciences, Fudan University, 220 Handan Road, 200433, Shanghai, China.
E-mail: lijin@fudan.edu.cn

Received 3 March 2007; accepted 21 June 2007

DOI 10.1002/ajpa.20690
Published online 31 July 2007 in Wiley InterScience (www.interscience.wiley.com).

Fig. 1. Ethnic divisions in South China before the Han's southward expansion. Dashed lines are the ethnic divisions. Dotted lines are the province borders. Dashed dotted lines are the country borders. Solid lines are rivers and sea lines. Important cordilleras between ethnic areas are marked in the map.

(*Kuaiji* to *Jiaozhi*, Song, 1991), although most of the Daic populations reside to the west of Hong Kong now. Thus, the genetic structure of the Daic is one of the most important parts of the study of South China's genetic structure.

Mitochondrial DNA is a powerful tool in population history studies, and it claims a vital role in global human population research (Shriver and Kittles, 2004). mtDNA in South Chinese populations is also being explored. Population data from certain areas have been reported in succession, especially data from the minorities of the Yunnan province in Southwest China (Yao and Zhang, 2002). Ethnic families are also being systematically investigated, and the first family investigation to be completed was that of TB. mtDNA polymorphisms of most of the TB populations in China were found to have a northern origin, while also having had absorbed a large proportion of southern lineages (Wen et al., 2004b). Han populations were also studied systematically, and a mixed structure, similar to that of TB, was found (Wen et al., 2004a). Similar studies were performed in the HM family, but these data show less mixing of northern lineages with the southern lineages (Wen et al., 2005). Data of TA have also been reported (Tajima et al., 2003).

Overall, mtDNA research in South China is relatively complete, except for that of the major group, the Daic, and for the AA group in outlying regions. As is widely known, the Han and TB groups moved from Northern China (Su et al., 2000), and while the Han settled throughout South China, the TB settled mainly in remote areas of Southwestern China. The TA and MP groups, natives of South China, have smaller populations and now reside in the Southeast islands far from the mainland. Therefore, the ethnic groups of most relevance to our study in South China are the HM and Daic. Knowing their genetic structure would contribute to a full understanding of the original genetic structure of Southern China. Although the AA group is far to the Southwest, it is still of significant value as a reference population considering it is so ancient a group (the ancient distribution of groups mentioned above refers to Fig. 1). As a result, we will analyze the matrilineal genetic structure of South China by exploring the mitochondrial DNA polymorphisms of all Daic populations and many AAs, eventually combining our data with that from other groups.

Through the analysis of mtDNA, we expect to discover which mtDNA polymorphisms represent the matrilineage of South China, when they were formed, and how and when they branched out geographically. In this way, we hope to ascertain the timetable of modern human settlement and fragmentation in South China, and thus to provide a sturdy foundation for the original genetic structure of the whole of East Asia.

MATERIALS AND METHODS

Blood samples of 774 individuals from 30 Daic populations across South China and Vietnam were collected with FTA cards (Whatman), covering almost all of the Daic populations in China. There are also 126 individuals from 13 AAs and one Sino–Tibetan population, bringing the total sample size to 900. All of the donors are unrelated and gave their consent to the study. The data from the populations studied is shown in Table 1.

An mtDNA fragment containing the HVS-1 region was amplified by primers L15974 and H16488 (Yao et al., 2002a), and the purified PCR product was sequenced using the BigDye terminator cycle sequencing kit and an ABI 3100 genetic analyzer (Applied Biosystem). Primers were also designed for amplifying multiple fragments containing 12 RFLP polymorphisms in the coding regions, and most of the PCR products were digested by restriction enzymes: 9bp deletion, 10397 AluI, 10394 DdeI, 663 HaeIII, 5176 AluI, 4831 HhaI, 12406 HpaI, and 9824 HinfI (Kivisild et al., 2002; Yao et al., 2002a). The additional variations, 3010, 4715, 5417, and 10310, were genotyped by directional sequencing or PCR-RFLP assay by engineering restriction sites in the primers (primer information and genotyping protocol for all the coding region variants are shown in Supplementary Table 1).

The 464bp long HVS-1 sequences (16024-16488) were edited and aligned against the revised CRS (Andrews et al., 1999) using DNASTAR software (DNASTAR). All of the 900 HVS-1 sequences from 53 populations have been submitted to Genbank (Accession number: EF654716-EF655616).

Haplogroup affiliation of each mtDNA sequence was inferred by combined use of the HVS-1 motif and diagnostic variants in the coding regions following Kivisild et al. (2002) and Kong et al. (2003a). Median joining networks (Bandelt et al., 1999) were constructed by NETWORK software (www.fluxus-engineering.com) to investigate detailed lineage relationship within each haplogroup. Coalescence time and its standard error of the haplogroups were calculated by the methods developed by Forster et al. (1996) and Saillard et al. (2000), respectively. The reference data of other groups in East and Southeast Asia used in networks were obtained from the literature (Horai et al., 1996; Wen et al., 2003, 2004a,b, 2005; Qian et al., 2001; Yao and Zhang, 2002; Kivisild et al., 2002; Yao et al., 2002a,b, 2003, 2004; Tajima et al., 2003; Kong et al., 2003b), including that for the Daic, AAs, HMs, TAs, Sino–Tibetans, and some Altaic populations.

RESULTS

Distribution and specificity of haplogroups

Almost all of the samples were genotyped and their haplogroups were confirmed (Supplementary Table 2) (Table 2). The most common Daic haplogroups are B4a,

TABLE 1. General information about the populations studied

Code	Ethnic name	ISO639-3	Family	Sub-family	Branch	Population	Size	Country	Province	County	Township	Village	Hamlet	Long.(E)	lat.(N)
AC	Ai–Cham	AIH	Daic	Kam–Tai	Kam–Sui	2,300	6	China	Guizhou	Libo	Boyao	Xinqiao		107.7	25.4
BG	Bugan	BBH	Austro–Asiatic	Mon–Khmer	Unclassified	3,000	32	China	Yunnan	Xichou	Jijie		Manlong	104.8	23.7
BN	Bana	BDQ	Austro–Asiatic	Mon–Khmer	Eastern	137,000	3	Vietnam	Kontum					108.3	12.7
BU	Buyang	BYU	Daic	Kadai	Yang–Biao	3,000	31	China	Yunnan	Guangnan	Dihe	Punong	Yanglian	104.4	24.2
CL	Caolan	MLC	Daic	Kam–Tai	Be–Tai	114,000	30	China	Guangxi	Fangcheng	Nadong	Banmeng		108	21.8
CT	Chut	SCB	Austro–Asiatic	Mon–Khmer	Viet–Muong	1,500	1	Vietnam	Quangbinh					105.8	16.9
CU	Cun	CUQ	Daic	Kadai	Yang–Biao	70,000	30	China	Hainan	Dongfang				108.8	19.1
CX	Zhuang–N	CCX	Daic	Kam–Tai	Be–Tai	10,000,000	25	China	Guangxi	Tianlin				106	24.2
CY	Zhuang-S	CCY	Daic	Kam–Tai	Be–Tai	4,000,000	12	China	Guangxi	Chongzuo/Shangsi				107	22.5
DA	Dai-lu	KHB	Daic	Kam–Tai	Be–Tai	770,000	56	China	Yunnan	Jinghong				100.7	21.6
DE	Die	JEH	Austro–Asiatic	Mon–Khmer	Eastern	10,000	2	Vietnam	Kontum					108.4	12.7
DG	Danga	YUE*	Daic	Unclassified		1,000,000	40	China	Hainan	Lingshui	Xincun	Xincungang		110.1	18.5
DN	Dong	DOC	Daic	Kam–Tai	Kam–Sui	907,560	10	China	Hubei	Enshi	Bajiao	Huangnitang		119.4	30.1
GA	Qau	GIO*	Daic	Kadai	Ge–Chi	3,000	12	China	Guizhou	Bijie	Puyi	Huoma		105.6	27.7
GL	Blue–Gelao	GIO	Daic	Kadai	Ge–Chi	1,700	30	China	Guangxi	Longlin	Dee			105.4	24.6
HL	Hlai–Qi	LIC	Daic	Kadai	Hlai	747,000	34	China	Hainan	Tongza	Nansheng	Yananxia		109.5	18.8
HR	Hre	HRE	Austro–Asiatic	Mon–Khmer	Eastern	94,000	1	Vietnam	Quangngai					108.7	13.8
HS	Sui	SWI	Daic	Kam–Tai	Kam–Sui	345,993	30	China	Guangxi	Rongshui	Yongle	Beigao	Mengcun	109	25
JM	Jiamao	JIO	Daic	Kadai	Hlai	52,300	27	China	Hainan	Baoting	Jiamao	Jiada		109.7	18.6
KH	Halang	HAL	Austro–Asiatic	Mon–Khmer	Eastern	10,000	1	Vietnam	Kontum					108.2	12.7
KN	Kinh	VIE	Austro–Asiatic	Mon–Khmer	Viet–Muong	65,051,000	43	Vietnam	Hue					108.2	15.7
KT	Katu	KTV	Austro–Asiatic	Mon–Khmer	Eastern	37,300	2	Vietnam	Quangnam					107.4	15.1
LC	Lachi	LBT	Daic	Kadai	Ge–Chi	9,016	30	China	Yunnan	Maguan	Jiahanjing	Niulongshan	Oldville	104.3	23
LG	Lingao	ONB	Daic	Kam–Tai	Be–Tai	520,000	31	China	Hainan	Lingao				109.6	19.9
LL	Lolo	YIG	Sino–Tibetan	Tibeto–Burman	Lolo–Burmese	800,000	4	China	Guizhou	Dafang				105.5	27.2
LQ	Pubiao	LAQ	Daic	Kadai	Yang–Biao	307	25	China	Yunnan	Malipo	Tiechang	Dongdu	Pufeng	104.9	23.2
MG	Mnong	CMO	Austro–Asiatic	Mon–Khmer	Eastern	50,000	3	Vietnam	Daklak					107.9	12
MK	Mak	MKG	Daic	Kam–Tai	Kam–Sui	10,000	33	China	Guizhou	Libo	Jialiang	Fangcun		107.6	25.6
ML	Mulam	MLM	Daic	Kam–Tai	Kam–Sui	159,328	39	China	Guangxi	Luocheng				108.9	24.8
MN	Maonan	MMD	Daic	Kam–Tai	Kam–Sui	37,000	32	China	Guangxi	Huanjiang	Xianan			108	24.95
MO	Mollao	GIO*	Daic	Kadai	Ge–Chi	30,000	29	China	Guizhou	Majiang	Xiasi			107.7	26.5
MQ	DornQdayc	WUU*	Daic	Unclassified		500,000	17	China	Shanghai	Minhang	Maqiao			121.4	31
MT	Man–Thanh	TMM	Daic	Kam–Tai	Be–Tai	Small	2	Vietnam	Hatinh					106.3	16
PK	Pacoh	PAC	Austro–Asiatic	Mon–Khmer	Eastern	15,000	3	Vietnam	Quangtri					108.5	14
PO	Pou	BYK	Daic	Kam–Tai	Kam–Sui	20,000	34	China	Guangdong	Huaiji	Shidong			112.1	23.6
PY	Palyu	PLY	Austro–Asiatic	Mon–Khmer	Palyu	10,000	30	China	Guangxi	Longlin	Changfa	Xinhe	Mouzitun	105.4	24.6
RG	Red-Gelao	GIR	Daic	Kadai	Ge–Chi	1,500	31	China	Guizhou	Dafang	Pudi	Hongfeng		105.7	27.3
SD	Sedang	SED	Austro–Asiatic	Mon–Khmer	Eastern	40,000	1	Vietnam	Kontum					107.9	13.8
TI	Trieng	STG	Austro–Asiatic	Mon–Khmer	Eastern	27,000	2	Vietnam	Kontum					107.7	14.1
TN	Then	TCT	Daic	Kam–Tai	Kam–Sui	20,000	30	China	Guizhou	Pingtang	Liudong			107.3	25.7
TY	Tay	TYZ	Daic	Kam–Tai	Be–Tai	1,190,000	4	Vietnam	Gialai					106	22
WG	White–Gelao	GIW	Daic	Kadai	Ge–Chi	1,200	14	China	Yunnan	Malipo	Tiechang	Dongdu	Chongchong	104.9	23.1
WS	E	EEE	Daic	Kam–Tai	Be–Tai	30,000	33	China	Guangxi	Rongshui	Yongle	Xiaqin	Baima	109.1	25
YR	Yerong	YRN	Daic	Kadai	Bu–Rong	400	15	China	Guangxi	Napo	Ronghe	Renhe	Rongtun	106	23.4

ISO639-3 is the international standard devised to enable the uniform identification of all known languages in a wide range of applications, particularly including information systems. For our sample populations these linguistic tags can be used to search Ethnologue (http://www.ethnologue.com) for information on the populations. Those codes with stars stand for the languages used by several populations beside our sample populations. The original languages of those sample populations have been replaced by the coded languages.

TABLE 2. Frequencies of the mtDNA haplogroups of the populations studied (%)

CODE	A	B*	B4*	B4a	B4b1	B5*	B5a	B5b	C	D*	D5	F*	F1a	F1b	F1c	F2a	F3	F*16218	G*	G2a	M*	M7*	M7b*	M7b1	M7b2	M7c1	M8*	M8a	M9a	N*	N9a	R*	R9b	R9c	Z	Un
AC				33.3																			16.7	33.3									16.7			
BG			3.1						3.1	18.8		3.1	25.0	12.5			12.5		12.5		3.1		6.3													
BN						33.3							33.3																			33.3				
BU		2.9		8.8	5.9		2.9		8.8				20.6	2.9			8.8		5.9	5.9			14.7	2.9						5.9			2.9			
CL				10.0	3.3	3.3	6.7			6.7		3.3	6.7			3.3			3.3		10.0	6.7		10.0		3.3				13.3	3.3		3.3	3.3		
CT										1 in.																									
CU		3.3	6.7	13.3	3.3				10.0	6.7			3.3					10.0	3.3		16.7	3.3	3.3	6.7			3.3						6.7			
CX		4.0		4.0	4.0	4.0			4.0	12.0			8.0								8.0		4.0	24.0		4.0				4.0	4.0		8.0	4.0		
CY			8.3	16.7			25.0			8.3			8.3	8.3									8.3								16.7					
DA	5.4	1.8		7.1	3.6		5.4		8.9	7.1	3.6	1.8	8.9		1.8		1.8				12.5	3.6	7.1	3.6							1.8	1.8	5.4	3.6	3.6	
DE										50.0			50.0																							
DG	5.0	2.5	10.0	7.5			17.5		7.5			2.5	2.5				2.5		5.0		7.5			10.0		2.5				10.0	2.5		2.5		2.5	
DN	30.0			10.0			20.0								10.0				20.0		10.0														
GA				8.3			8.3			25.0			16.7										33.3									8.3				
GL	3.3	6.7		3.3		3.3			13.3	16.7			10.0			3.3	3.3				13.3		3.3	16.7						3.3					
HL		5.9		8.8	8.8		8.8			2.9		2.9						5.9	8.8		23.5			5.9		5.9		2.9		2.9	2.9		2.9			
HR													1 in.																							
HS				3.3	6.7		20.0				3.3	3.3	20.0				13.3				6.7		6.7	6.7				6.7					3.3			
JM			7.7	11.5				3.8				15.4				11.5					3.8		3.8	7.7				7.7		3.8			19.2	3.8		
KH																					1 in.														
KN		4.2	6.3	4.2		2.1	4.2		4.2	6.3		10.4	18.8						6.3		4.2		8.3	8.3				2.1	2.1			2.1	6.3			
KT			2 in.																																	
LC			8.8	11.8	5.9		5.9		8.8				2.9			2.9			2.9		11.8		20.6	5.9								2.9	5.9			2.9
LG	6.7	3.3		3.3	3.3	3.3			6.7	3.3		10.0	6.7	3.3		3.3	3.3	3.3	6.7		6.7		6.7	3.3				6.7			3.3			3.3		3.3
LL	25.0	25.0	25.0																25.0																	
LQ			4.0	4.0			4.0			8.0				8.0			8.0		8.0		16.0		8.0	20.0									4.0		8.0	
MG							66.7						33.3																							
MK			3.0	6.1	3.0		9.1			3.0	3.0		9.1			3.0	9.1		3.0		3.0	6.1	15.2	15.2	3.0								3.0			3.0
ML	5.3			7.9			10.5			7.9		2.6	7.9	5.3	2.6	2.6	7.9				10.5	2.6	7.9	2.6	2.6	2.6	2.6				2.6				2.6	2.6
MN				15.6	15.6		3.1		3.1	3.1			9.4								3.1	3.1	6.3	15.6			6.3			3.1	12.5					
MO	3.4		3.4	6.9			10.3		3.4	13.8			3.4				13.8				10.3		13.8							10.3		6.9				
MQ	5.9	11.8	17.6	5.9	5.9		5.9						17.6		5.9						11.8					5.9										5.9
MT																					50.0												50.0			
PK				33.3									33.3	33.3																						
PO	2.9			5.9			5.9			5.9	2.9	2.9	23.5				20.6				8.8		8.8	5.9				2.9			2.9					
PY	3.3								23.3	13.3			6.7										13.3	20.0						13.3	3.3		3.3			
RG			2.8							2.8					8.3		2.8		44.4				11.1					2.8				8.3	13.9			2.8
SD													1 in.																							
TI						50.0							50.0																							
TN	6.7		6.7	20.0		3.3	3.3		6.7	3.3	3.3		3.3								10.0		13.3	16.7				3.3								
TY													50.0																				25.0	25.0			
WG				12.5		6.3			25.0	6.3					6.3						18.8		6.3											6.3			12.5
WS			3.1	15.6	3.1		9.4			9.4			12.5	9.4			6.3						3.1	3.1						3.1	9.4		9.4		3.1	
YR		6.7	6.7	13.3			6.7			13.3			6.7		6.7						20.0		6.7								6.7			6.7		

F*16218: the subgroup under haplogroup F with 16218 derived allele. Un: haplogroup undetermined. 1 in.: single individual for the sample population.

F1a, M7b1, B5a, M7b*, M*, R9a, and R9b, in order of frequency, and the total percentage of these common haplogroups is 48.8%. In the remote AA populations, the most common haplogroups in order of frequency are F1a, M*, D*, F1b, N*, C, M7b*, M7b1, and F1a1a. This list is noticeably but unsurprisingly different from that of the Daic's. The percentage of the first three haplogroups for the AA populations alone totals 50.8%. The Daic's haplogroup list is similar, however, to that of the HM's (B5a, B4a, M*, M7b*, C, B4b1, M7b1, F1a, B4*, and R9b, totaling 50.6%), and is based on these two Southern populations, the Daic and the HM (Wen et al., 2005), we can conclude that B, M7, F, and R are the most characteristically Southern haplogroups. Among Daic populations, the frequencies of these four Southern specific haplogroups total 66.4%, which is higher than the totals in either the AA population (48.9%) or the HM population (58.9%). Moreover, the frequency of these haplogroups is decreased in more northern populations, such as in the Han (40.8%), the more northern Tibeto–Burman (37.5%), and the northernmost Altaic (16.3%). These four haplogroups are, therefore, essential to the study of matrilineage in South China.

Network analysis of related groups

To create networks in the Daic, AA, MP, HM, Sino–Tibetan, and Altaic populations, we used HVS-1 motifs (see Supplementary Table 2) from the haplogroups with high frequencies in South China (Fig. 2) as well as previously published haplogroup and HVS-1 data (Horai et al., 1996; Wen et al., 2003, 2004a,b, 2005; Qian et al., 2001; Yao and Zhang, 2002; Kivisild et al., 2002; Yao et al., 2002a,b, 2003, 2004; Tajima et al., 2003; Kong et al., 2003b). We marked each ethnic family in the network by color, but we separated the coastal Han, those from Shanghai to Guangxi, from the remaining population, taking into consideration that there may be a large number of Daic descendents in the coastal Southeastern Han populations, especially among the Southern specific haplogroup ones.

The networks contain large nodes with primary, secondary, and sometimes further branching representing shared haplogroups among ethnic families. The networks clearly show greater differences in haplogroups between Northern and Southern populations than between ethnicities. For example, some branches contain a variety of unrelated ethnic families who are geographically close. However, some haplogroups, such as M7b and B4-16140 (subhaplogroup under B4 with 16140 derived allele), remain approximately specific to one ethnic group, in this case the Daic. Han samples from the Southeast coastal zone are the closest to the Daic samples phylogenetically, sharing the most haplotypes. Although linkages, which is to say, haplogroup similarities, between the Daic and the HM are also common, there are still fewer linkages than between the Southeastern Han and Daic. The linkages between the Daic and the Han exist mostly in smaller branches, however; those between the Daic and the HM are more primary. Differentiation among the subfamilies of Daic is difficult. The Tai subfamily is especially varied, indicating a history of strong population expansion.

Time estimate

We conducted time estimates of the approximate age of several relatively high-frequency branches in specific haplogroups of the South China populations. The results have a standard deviation of about 1/3. R is a primary haplogroup that branches off about 70,000 years ago, while branches for B, F, and M7 are each around 50,000 years old. Most other ancillary branches are 40,000 years old. There is not much contradiction for the corresponding relationship between the phylogenesis of haplogroup and branch age. For the age of each branch and the phylogenic tree refer to Figure 3.

In the root nodes at the very base of the network, the HM and the Daic are not differentiable. However, in the first ancillary branches where differentiations begin to occur, the groupings are extremely HM or Daic specific, with no more than one outlier per branch. All of the haplogroups with ethnic distinctions differentiated within the last 30,000 years, and most differentiated within the last 20,000. The largest ethnic specific haplogroups became distinguishable no more than 10,000 ago, which is also about when the most terminal ends of the networks differentiated. The population fragmentation in the terminal ends is obvious. According to the time table in Figure 3, the mtDNA lineage history generally falls into three phases, the first around 55–25 kybp, the second from 25–9 kybp, and the third from 9 kybp present.

DISCUSSION

The Daic contain most Southern specific mtDNA haplotypes

mtDNA haplogroup distribution in South and North China is obviously unequal (Yao et al., 2002a; Kong et al., 2003a; Wen et al., 2004a,b, 2005), and the haplogroups B, M7, F, R, which are common in South China, are clearly of Southern origin. These Southern haplogroups exist in high frequencies in the Daic, HM, and AA populations, indicating that these three ethnic families are native to South China. The Daic are acknowledged to be the descendants of the *Baiyue*, a famous, ancient ethnic family, which according to Chinese historical records lived in the coastal zone that now exists between Shanghai and Hanoi 2,000 years ago. This family, in turn, descended from the most powerful ethnic family in South China 8–2,000 years ago (Song, 1991). Their ancestors had comparatively advanced cultures (Song, 1991) (Hemudu Culture, Liangzhu Culture, etc) in these areas during prehistory, and they may have lived in South China for at least 30,000 years. When the Han began to expand southward in 2 kybp (Wen et al., 2004a), a large number of the *Baiyue* were assimilated by the Han. Others migrated westwards to become today's Daic populations. Thus, we can find a high proportion of mtDNA haplogroups similar to the Daic in southeastern Han populations.

The Daic population has been shown to have a prominent place in population studies of South China, and ancient DNA from archaeological sites in the southeastern coastal zone can be used to better trace the footprints of the Daic's ancestors. These studies could reveal much of the population history in the entire Southern region.

Among the Daic, the HM, and the AA groups, the Southern haplogroup frequency is highest in the Daic and lowest in the AAs, although the AAs are nevertheless regarded as a native Southwestern population. The overall frequencies of B, M7, F, R are high in the Southeast but become lower in the populations to the northwest. We therefore conclude that the Daic, the HM, and

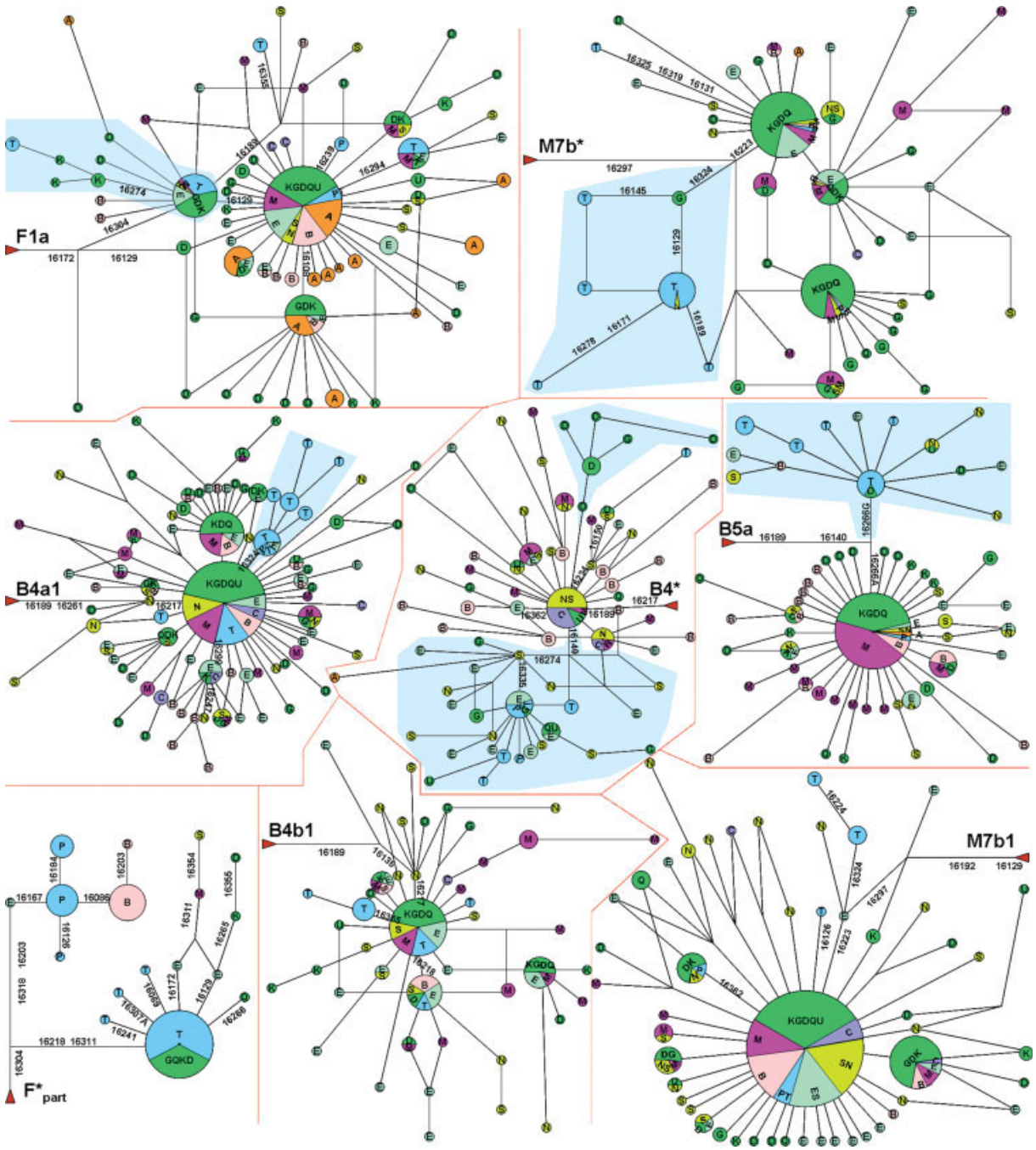

Fig. 2. Networks of the mtDNA haplogroups that are in high-frequency in South China derived from East Asian population data. Triangles represent the networks' roots. Dark green nodes represent Daic individuals (G: Gelao; Q: Hlai; K: Kam-Sui; D: Tai; U: Unclassified); orange: Austro–Asiatic (A); amaranthine: Hmong–Mien (M); blue: Taiwanese (T) and Malayo–Polynesian (P); light green: southeast coastal Han (E); lime green: Han–Mandarin (S: South; N: North); incarnadine: Tibeto–Burman (B); light indigotic: Altaic (C). The blue background marks the ethnic specific sub-haplogroups. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

the AA haplogroups are Southeast specific rather than South China specific.

One explanation why these Southern specific haplogroups seem to gather in the Southeast might be that the AA population in Southwest China was affected by TB returning from the North, thus reducing much of the Southern specificity. However, the Han and the TB populations share a geographic origin and were indistinguishable before their southward migration, so any affect the TB population may have had on the AA should be relatively equal to the effect the Northern Han had on the Daic. Noticeable differences between the AA and the Daic should therefore not occur, but they do. Another explanation may be that because modern humans entered South China by more than one route and because the few original populations in South China were very small,

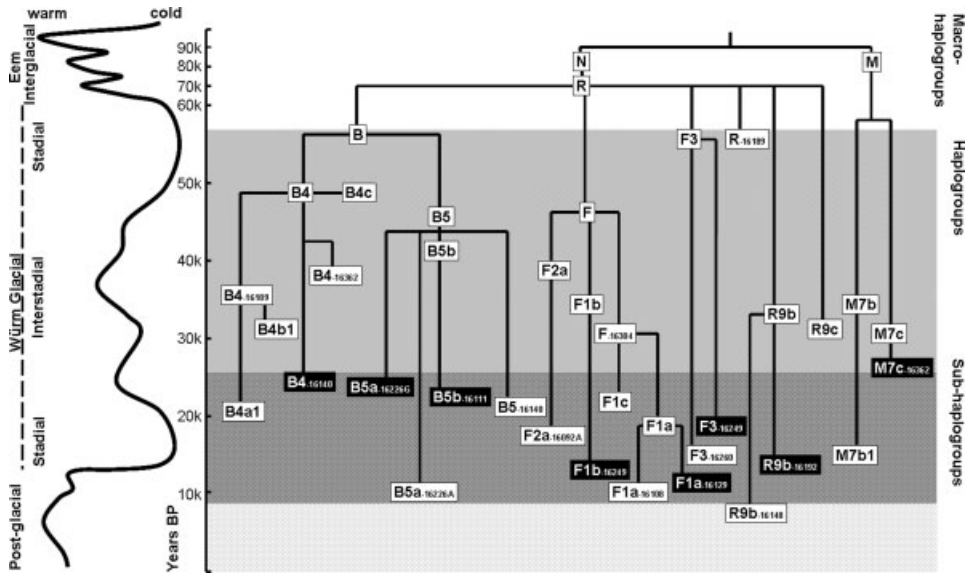

Fig. 3. The ages of the mtDNA lineages in South China and the climatic variation during this period. Sub-haplogroups written in black boxes are ethnic specific. Three different backgrounds indicate three periods of ethnicity development; the darkest background represents the ethnic differentiation period. The temperature change is show on the left. All of the ethnic specific sub-haplogroups appeared in the last stadial of the Würm Glacial.

independent genetic drift would have taken place in the populations of East and West transmigrants. Southern specific haplogroups then randomly became plentiful in the East while there were relatively few in the West. It is possible that we, therefore, find many fewer southern haplogroups in the Northern populations because they arose from the Western populations, and the difference in frequencies of Southern specificity is traceable to this day.

History of modern East Asian's entry into South China

Because the four main haplogroups (B, M7, F, R) are all around 50,000 years old, the ancestor groups of modern East Asians likely differentiated at that time when the original mutations were formed. However, these ancestors still had probably not settled in East Asia at that point. The genetic data must be compared to other anthropological evidence to give an estimate for the time of the modern East Asians' entry into South China. According to palaeoanthropological evidence, this must have taken place more than 30 kybp. As a result, the migration should have taken place in the period between 50 and 30 kybp, which is consistent with the research on the East Asian Y chromosome (Jin and Su, 2000). After the first settlement of modern East Asians in South China, there was a long pause before the diversification of populations occurred.

Population fragmentation

Basing on our research, the fragmentation of Southern Chinese populations began around 20 kybp, when ethnically specific haplogroups were formed. Low population density, which allowed all of the original migrants to remain in the southernmost part of China, may be one reason that the population did not differentiate earlier or upon their arrival. Population growth around 20 kybp may then have caused the subsequent differentiation. Inmixing between groups, however, was still quite common. Therefore mtDNA haplogroups, even some of those branching later than 20, 000 years ago, were still found in a wide range of ethnic groups. In fact, the populations may well have diverged geographically before 20 kybp, but the high frequency of inmixing may have homogenized the population's haplotypes. However, after about 20 kybp, the genetic exchange was almost cut off, either by natural or social factors or a mixture of the two. No specific social cause is known, but a natural disaster in the form of the peak of a glacial age may well have been the cause of the sudden end of genetic inmixing (Fig. 3, Shi et al., 1989). The freezing weather very likely not only reduced communication between populations but also may have altered their conditions of survival. The original modern East Asian was probably forced to move to different areas segregated by the cordilleras (Fig. 1), and this segregation was apt to spur the gradual shaping of the diversity between populations afterwards. The differentiation was culturally magnified over 10,000 years concomitant with ever less genetic communication, and the presently observed diversity of populations, represented by the Daic, the HM, and the AA of South China, was finally formed. Two most important archaeological cultures, Hemudu and Daxi, formed in the region of the Daic and the HM, respectively, around 8 kybp. Based on mtDNA data, population mixing has never fully stopped, so the differences that exist among the AA, HM, and Daic are still increasing, albeit very slowly. Unsurprisingly, then, the groups closest geographically are also comparatively close in mtDNA haplotype data. This interchange not only takes place among Southern groups, but also occurs between Southern and Northern groups and is likely more frequent between the Northern groups moving southwards and the Southern aboriginal groups already there.

Much information can be obtained about the original settling of South China and the subsequent differentiations of its populations using mtDNA polymorphisms. However, mtDNA only shows the matrilineal genetic history of South China. The genetic information to be gleaned from South Chinese Y-chromosomes might reflect a very different genetic structure, and is, therefore, a vital complement to mitochondria research.

ACKNOWLEDGMENTS

We thank all of the donors for making this work possible. The Ethnic Affairs Committee of Guangxi Zhuang

Autonomous Region, Institute of Ethnology in Guizhou, Viêtnam Huê Medical College, Xishuangbanna Prefecture Committee of C.C.Youth League, Wenshan Prefecture Committee of C.C.Youth League, Research Society of Hainan Ancient Migrants, and Mr Shi Shi of Chongqing Teacher's University offered help in sample collection. Dr. Angelika Hofmann of Yale University revised the paper and gave important suggestions on the scientific writing.

LITERATURE CITED

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.

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48.

Bellwood PA, James JF, Barrell T. 1995. The Austronesians: historical and comparative perspectives. Canberra. Department of Anthropology, Research School of Pacific and Asian Studies, Australian National University.

Diamond JM. 1988. Express train to Polynesia. Nature 336: 307–308.

Forster P, Harding R, Torroni A, Bandelt H-J. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59:935–945.

Grimes BF. 2002. Ethnologue: languages of the world, 14th ed. Dallas: International Academic Bookstore.

Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara S, Park KS, Omoto K, Pan IH. 1996. mtDNA polymorphism in East Asian populations, with special reference to the peopling of Japan. Am J Hum Genet 59:579–590.

Jin L, Su B. 2000. Natives or immigrants: origin and migrations of modern humans in East Asia. Nat Rev Genet 1:126–133.

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.

Kong QP, Yao YG, Sun C, Bandelt HJ, Zhu CL, Zhang YP. 2003a. Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. Am J Hum Genet 73:671–676.

Kong QP, Yao YG, Liu M, Shen SP, Chen C, Zhu CL, Palanichamy MG, Zhang YP. 2003b. Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. Hum Genet 113:391–405.

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.

Saillard J, Forster P, Lynnerup N, Bandelt H-J, Nørby S. 2000. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am J Hum Genet 67:718–726.

Shi YF, Cui ZJ, Li JJ. 1989. Quaternary glacier in eastern China and the climate fluctuation. Beijing: Science Press.

Shriver MD, Kittles RA. 2004. Genetic ancestry and the search for personalized genetic histories. Nat Rev Genet 5:611–618.

Song SH. 1991. Bai-Yue. Changchun: Jilin Education Press.

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 LL, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z, Jin L. 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, Xiao C, Deka R, Seielstad MT, Kangwanpong D, Xiao J, Lu D, Underhill P, Cavalli-Sforza LL, Chakraborty R, Jin L 2000. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. Hum Genet 107:582–590.

Tajima A, Sun CS, Pan IH, Ishida T, Saitou N, Horai S. 2003. Mitochondrial DNA polymorphisms in nine aboriginal groups of Taiwan: implications for the population history of aboriginal Taiwanese. Hum Genet 113:24–33.

Wen B, Shi H, Ren L, Xi H, Li K, Zhang W, Su B, Si S, Jin L, Xiao C. 2003. The origin of Mosuo people as revealed by mtDNA and Y chromosome variation. Sci China C Life Sci 33:375–384.

Wen B, Li H, Lu DR, Song XF, Zhang F, He YG, Li F, Gao Y, Mao XY, Zhang L, Qian J, Tan JZ, Jin JZ, Huang W, Deka R, Su B, Chakraborty R, Jin L. 2004a. Genetic evidence supports demic diffusion of Han culture. Nature 431:302–305.

Wen B, Xie XH, Gao S, Li H, Shi H, Song XF, Qian TZ, Xiao CJ, Jin JZ, Su B, Lu DR, Chakraborty R, Jin L. 2004b. Analyses of genetic structure of Tibeto–Burman populations reveals sex-biased admixture in southern Tibeto–Burmans. Am J Hum Genet 74:856–865.

Wen B, Li H, Gao S, Mao XY, Gao Y, Li F, Zhang F, He YG, Dong YL, Zhang YJ, Huang WJ, Jin JZ, Xiao CJ, Lu DR, Deka R, Chakraborty R, Su B, Jin L. 2005. Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. Mol Biol Evol 22:1–10.

Wu RC. 1959. Human fossils found in Liujiang Guangxi. Acta Anthropolagica Sinica 1:97–104.

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.

Yao YG, Kong QP, Bandelt H-J, Kivisild T, Zhuang YP. 2002a. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet 70:635–651.

Yao YG, Nie L, Harpending H, Fu YX, Yuan ZG, Zhang YP. 2002b. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. Am J Phys Anthropol 118:63–76.

Yao YG, Kong QP, Man XY, Bandelt HJ, Zhang YP. 2003. Reconstructing the evolutionary history of China: a caveat about inferences drawn from ancient DNA. Mol Biol Evol 20:214–219.

Yao YG, Kong QP, Wang CY, Zhu CL, Zhang YP. 2004. Different matrilineal contributions to genetic structure of ethnic groups in the silk road region in china. Mol Biol Evol 21:2265–2280.

Zhang GZ. 1987. Archaeology on the coast of Southeast China and origin of Austronesian. Southern ethnic archaeology 1. Chengdu: Sichuan University Press.