

## Research Article

## Genetic evidence for the multiple origins of Pinghua Chinese

<sup>1</sup>Yan LU<sup>†</sup> <sup>2</sup>Shang-Ling PAN<sup>†</sup> <sup>2</sup>Shu-Ming QIN <sup>1</sup>Zheng-Dong QIN  
<sup>1</sup>Chuan-Chao WANG <sup>2</sup>Rui-Jing GAN <sup>1</sup>Hui LI\* the Genographic Consortium<sup>‡</sup>

<sup>1</sup>(State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China)

<sup>2</sup>(Department of Pathophysiology and Institute of Cardiovascular Disease, Guangxi Medical University, Nanning 530021, China)

**Abstract** Linguistics and genetics always reach similar results in phylogenetic studies of human populations. A previous study found that populations speaking Han Chinese dialects have closer genetic relationships to each other than to neighboring ethnic groups. However, the Pinghua Chinese population from Guangxi is an exception. We have reported that northern Pinghua people are genetically related to populations speaking Daic languages. In this study, we further studied the southern Pinghua population. The Y chromosome and mitochondrial DNA haplogroup components and network analysis indicated that northern and southern Pinghua populations were genetically different. Therefore, we concluded that the Pinghua speakers may have various origins, even though Pinghua dialects are similar. Pinghua dialects might have originated when the Daic or Hmongic speakers from different regions learnt to speak the same Chinese dialect hundreds of years ago. Speakers of one language do not always have just one origin.

**Key words** mitochondrial DNA, multiple origins, Pinghua Chinese, Y chromosome.

Spoken by one-fifth of the world's populations, Chinese is one of the most ancient written languages in the world and has diverged into different dialects in the past few thousand years. Traditional Chinese dialect classification comprises seven groups, in which Pinghua is not included. With further linguistic study during the past 20 years, Chinese is now classified into 10 major dialects, Mandarin, Jin, Cantonese (Yue), Wu, Hui, Xiang, Hakka, Gan, Min, and Pinghua (Cao, 2008). Most Pinghua speakers are Han Chinese; however, small sections of Guangxi indigenous ethnic groups such as Yao in Longsheng and Fuchuan counties are also taking Pinghua as their native language now (Liang & Zhang, 1999). There are various Chinese dialects in Guangxi, including Cantonese, Mandarin, Hakka, and Pinghua. Pinghua was the main language for communication among Han Chinese and southern ethnic groups in Guangxi before the arrival of the other Chinese dialects (Zhang, 1988). The history of Pinghua Chinese might be older than those of Hakka, Cantonese, or Min Chinese. Now, the population speaking Pinghua is around 3–4 million, mainly

distributed in Guangxi, some also scattered in Hunan, Yunnan, and Guangdong.

As one of the four dialects (Pinghua, Hakka, Mandarin, and Cantonese) in Guangxi, Pinghua is considered to be distinct from Hakka and Mandarin; however, the relation between Pinghua and Cantonese has always been a controversial subject in linguistics (Zhang & Liang, 1996). Some linguists insisted that Pinghua is different from Cantonese, therefore the two dialects are not closely related (Wei, 1996; Zhang, 1997; Liang & Zhang, 1999; Li, 2007). The Pinghua dialect is classified into the northern and southern subgroups, geographically divided by the city of Liuzhou (Yang et al., 1985). Some other linguists proposed to divide northern Pinghua and southern Pinghua according to the Cantonese influence on them (Wu, 2001; Xie, 2001; Qin, 2002). However, the characters of all the Pinghua dialects are apparent (Wang, 2001; Zhan et al., 2003).

Our previous study on the Y chromosome and mitochondrial DNA diversity of northern Pinghua populations (Gan et al., 2008) indicated that Pinghua people were much closer to indigenous ethnic groups than to other Han Chinese. Pinghua populations are not descendants of Han migrants but of ethnic groups that have accepted the Han culture and language. The Pinghua dialect might have developed from an early Chinese dialect learnt by the Guangxi ethnic groups, while also under the influence of their mother language.

Received: 12 October 2012 Accepted: 23 December 2012

<sup>†</sup> These authors contributed equally to this work.

<sup>‡</sup> Members of the Genographic Consortium are listed in the Appendix.

\* Author for correspondence. E-mail: lihui.fudan@gmail.com. Tel.: 86-21-55664574. Fax: 86-21-55664885.

In this paper, we studied the southern Pinghua populations to verify the conclusion of the indigenous origin of Pinghua people and to investigate the genetic relationship between the two Pinghua subgroups.

## 1 Material and methods

### 1.1 Population samples

We collected blood samples of 344 healthy and unrelated individuals (181 males and 163 females) from southern Pinghua populations in Nanning, Pingguo, and Tiandong, Guangxi Zhuang Autonomous Region, China, with approval from the Ethics Committee of the Fudan School of Life Sciences (Fudan University, Shanghai, China). All individuals signed informed consent. Genomic DNA was extracted through the classic phenol–chloroform method.

### 1.2 Y chromosome markers

Twenty-one Y chromosome biallelic loci (single nucleotide polymorphism, SNP) were typed in the 181 male samples. Genotyping of DE\*-YAP was carried out by agarose gel electrophoresis directly after polymerase chain reaction (PCR). C-M130, F\*-M89, K\*-M9, O1a\*-M119, O1a1a-M101, O1a2-M110, O2\*-P31, O2a1\*-M95, O2a1a-M88, O3\*-M122, O3a2b-M7, O3a2b1-M188, O3a1b-M164, and O3a2a-M159 were typed by TaqMan (Applied Biosystems, Foster City, CA, USA). O\*-M175, O3a-M324, O3a1a-M121, O3a2c1-M134, O3a2c1a-M117, D1-M15, and 16 Y chromosome short-tandem repeat polymorphisms (including DYS456, DYS389I, DYS389II, DYS390, DYS458, DYS19, DYS385, DYS393, DYS391, DYS439, DYS635, DYS392, YGATAH4, DYS437, DYS438, and DYS448) were typed using fluorescence-labeled primers for PCR amplification (Gan et al., 2008). The denatured products were electrophoresed on an acrylamide gel using an ABI 3100 genetic analyzer (Applied Biosystems). The Y chromosome haplotypes were determined according to the classification of Y-DNA Haplogroup Tree (Karafet et al., 2008; Yan et al., 2011).

### 1.3 Mitochondrial DNA markers

The hypervariable segment I (HVS-I) of the control region was amplified by primers L15974 and H16488 (Yao et al., 2002). Purified PCR products were sequenced using the BigDye terminator cycle sequencing kit and an ABI 3130XL genetic analyzer (Applied Biosystems). A SNaPshot assay was used for typing SNPs in the coding regions to confirm haplogroup identity. This assay was designed as a multiplex panel

including 21 coding region SNPs and a one-length variation marker (Qin et al., 2010). Both the HVS-I motif and the coding region variations were used to infer haplogroups according to the classification of Kivisild et al. (2002). The HVS-I sequences of 344 individuals have been submitted to GenBank (JX826640–JX826981).

### 1.4 Statistical analyses

Neighbor-joining unrooted trees based on  $F_{ST}$  statistics were carried out using MEGA 4.0 (Tamura et al., 2007) to visualize relationships among Pinghua populations and other southern populations.  $F_{ST}$  statistics were calculated from haplotype frequencies using ARLEQUIN 3.11 (Excoffier et al., 2005). Networks of the Y chromosome short tandem repeat (Y-STR) haplotypes and mitochondrial DNA (mtDNA) HVS-I motifs were constructed by the median-joining method (Bandelt et al., 1999) using NETWORK 4.1 ([www.fluxus-engineering.com](http://www.fluxus-engineering.com)).

Reference population data on the Y chromosome (Li et al., 2003; Wen et al., 2004a, 2004b; Li, 2005; Lin et al., 2007) and mtDNA (Tsai et al., 2001; Wen et al., 2004a, 2004b, 2005a; Li, 2005; Li et al., 2007) were retrieved from published reports. Data of northern Pinghua populations in our previous study (Gan et al., 2008) were also included in comparative analyses.

## 2 Results

### 2.1 Y chromosome haplogroup frequencies

Thirteen Y chromosome haplogroups were determined among the southern Pinghua samples (Table 1). The southern Pinghua showed similarities to the northern Pinghua in haplogroup patterns. The major haplogroups of northern Pinghua populations, O2a1\*-M95, O3a2c1a-M117, and O\*-M175, are also prevalent in southern Pinghua. O3a2c1a-M117 is frequent in Han Chinese (Su et al., 1999; Wen et al., 2004a, Shi et al., 2005). O2a1\*-M95 is frequent in Daic populations (Li, 2005) and Hmong-Mien populations (Feng, 2007); the highest frequency of O2a1\*-M95 in Pinghua populations suggest that the Pinghua population is closer to southern ethnic groups than to the other Han Chinese populations (Gan et al., 2008). However, there are also differences between southern and northern Pinghua populations. Haplogroups O1a\*-M119, O1a1a-M101, and O1a2-M110 appear in southern Pinghua populations, but these haplogroups are absent in northern Pinghua populations. O1a-M119 is the major haplogroup among the Daic and western

**Table 1** Y chromosome haplogroup frequencies of southern Pinghua populations

No.	Population	Size	Haplogroups (%)												
			D1	O*	O1a*	O1a1a	O1a2	O2*	O2a1*	O2a1a	O3*	O3a2b1	O3a2b	O3a2c1a	O3a2a
			M15	M175	M119	M101	M110	P31	M95	M88	M122xM324	M188	M7	M117	M159
1	Nanning	38		2.63		7.89		5.26	36.84	13.16	18.42	2.63		10.53	2.63
2	Pingguo	123		8.94	4.88	2.44	21.9	8.13	19.5	3.25	20.3	3.25	1.63	4.07	1.63
3	Tiandong	20	5	15	5	5		10	30	10				20	

Blank cells indicate zero values.

Austronesian populations (Li et al., 2008), indicating a Daic background of the southern Pinghua populations.

**2.2 Mitochondrial DNA haplogroup frequencies**

Thirty-seven mtDNA haplogroups were found in the 344 samples. M7b1, F1a, M\*, R10, M7b\*, and B4a are frequent in southern Pinghua populations. Haplogroups B4a, M\*, F1a, and M7b1 present at high frequencies in both northern and southern Pinghua populations (Table 2). This indicates that the two Pinghua subgroups are similar to each other in terms of maternal lineage.

According to the reference population data, the Daic populations are characterized by B4a, B5a, F1a, M7b1, M7b\*, M\*, R9a, and R9b (Li et al., 2007), the Hmong-Mien populations by B4a, B5a, M\*, M7b\*, C, B4b1, M7b1, F1a, B4\*, and R9b (Wen et al., 2005), and the Han Chinese by A, C, D, G, M8a, Y, and Z (Wen et al., 2004a). Both of the Pinghua subgroups are closer to southern indigenous populations than to other Han Chinese. However, genetic differences also exist between northern and southern Pinghua populations. Haplogroup R10, which is rare in northern Pinghua populations, appears at a moderate frequency in southern Pinghua; haplogroups B5a and N\* present

at high frequencies in northern Pinghua but are rare in southern Pinghua.

**2.3 Comparison of Pinghua populations and other Han Chinese**

Judging from the diversity patterns of Y chromosome and mtDNA haplogroups among Han Chinese populations (Fig. 1), we found genetic coherence was maintained in all nine Han Chinese branches except the Pinghua populations. The Y chromosome haplogroup pattern of Pinghua populations is quite different from that of other Han Chinese. Haplogroups O3-M122 and O3a2c1a-M117 are dominant in Han Chinese but less frequent in Pinghua populations. O2-P31 is rare in Han Chinese but prevalent in Pinghua populations. Similar to the northern Pinghua population, the southern Pinghua are not of Han Chinese origin either. The patterns of Y chromosome and mtDNA haplogroups are also different between the two Pinghua subgroups. The frequencies of Y haplogroups O3a2c1a-M117 and O2a1\*-M95 are much lower in southern Pinghua than northern Pinghua. The biggest distinction is that the Y haplogroup O1a-M119 exists in southern Pinghua in a moderate frequency, but is absent in all northern Pinghua populations. Mitochondrial DNA also leads to

**Table 2** Mitochondrial DNA haplogroup frequencies of southern Pinghua populations

No.	Size	Haplogroups (%)												
		A	B*	B4*	B4a	B4b1	B5*	B5a	B5b	C	D*	D4*	D4a	D4b
1	68			1.5	4.4	2.9				1.5			4.41	
2	226	3.1	0.44		4.9	1.3	0.89	0.89	0.9	2.2	2.7	1.3	4.87	0.9
3	50	2.04			6.1	2		2.04		2		4.1		

No.	Size	Haplogroups (%)												
		D5	F*	F1a	F1c	F3	G2a	M*	M7*	M7a	M7b*	M7b1	M7c	M8*
1	68		8.82	13	1.5	1.5		10.3	3		5.9	13		
2	226	3.1	2.61	11		3.1	1.77	6.2	5.3	0.4	4.4	13	0.89	1.8
3	50	2.04	2.04	16		10		12.3			4.1	14		

No.	Size	Haplogroups (%)												
		M8a	M9a	M10	M12	N*	N9a	R*	R9*	R9a	R9c	R10	Y	Z
1	68	1.47	1.47			8.8	5.88	2.94		1.5	1.5	2.9		
2	226		0.44	0.4	3.1	0.4	2.66		0.4	1.8	2.7	6.2	0.44	1.8
3	50	2.04				4.1	2.04			2		8.2	2.04	

Blank cells indicate zero values.

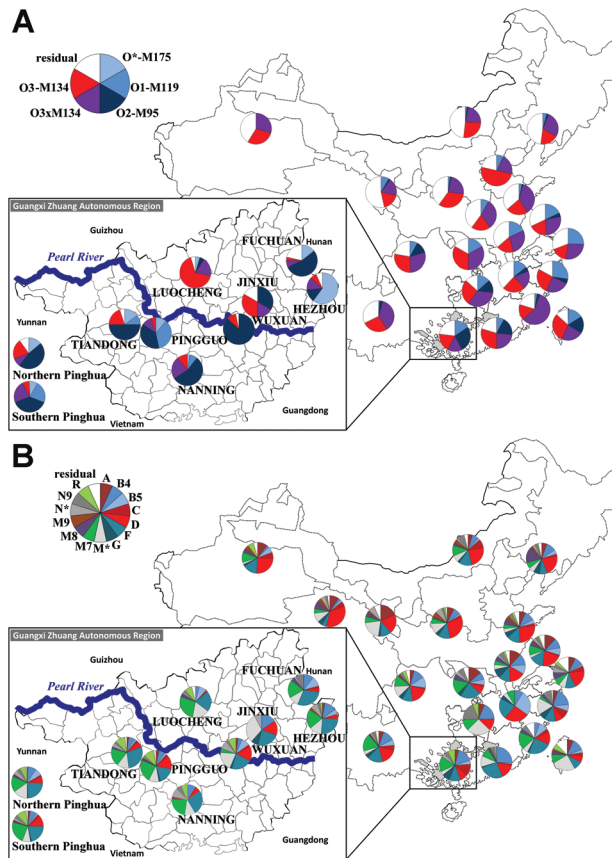


Fig. 1. Y chromosome (A) and mitochondrial DNA (B) dominant haplogroups of Pinghua populations and other Han Chinese populations. Maps of southwest China were modified from Google Maps.

the same conclusion. Haplogroups A, C, and D are less frequent in Pinghua populations than other Han Chinese. The mtDNA lineage pool of Pinghua populations is much closer to the southern ethnic groups than to other Han Chinese.

#### 2.4 Cluster analysis between Pinghua and neighboring populations

To examine the genetic relationship between Pinghua populations and other southern ethnic groups, data of Daic and Hmong-Mien populations living around Pinghua populations in Guangxi and Guangdong were used to construct a phylogenetic tree (Fig. 2). In the tree based on Y chromosome haplogroup frequencies, all the three southern Pinghua populations clustered tightly, indicating the highly coherent paternal genetic structure of southern Pinghua populations. The northern Pinghua populations do not cluster with southern Pinghua populations. This confirmed that the paternal lineages of northern and southern Pinghua populations are different. In the mitochondrial DNA

tree, three southern Pinghua populations and three of the five northern Pinghua populations grouped together. Therefore, Pinghua people have more similarity in maternal lineages than in paternal lineages. However, incoherence of the genetic structure of Pinghua subgroups is apparent.

#### 2.5 Network analyses of major haplogroups

Difference in SNP haplogroup frequency is not sufficient to support the different origins of the Pinghua subgroups. Therefore, we further investigated the Y-STR and mtDNA HVS-I diversities among each of the dominant haplogroups between the two subgroups by analyzing the networks (Fig. 3). In Y-STR networks (Fig. 3: A), the haplotypes of the southern Pinghua samples (Table S1) were compared with those of northern Pinghua samples and Han, Daic, and Hmong-Mien (Li, 2005; Chen et al., 2006; Feng, 2007; Li et al., 2007; Gan et al., 2008). In the networks of O1a2-M110 and O2a1a-M88, the southern Pinghua samples share many haplotypes with the Hmong-Mien samples. In the network of O2a1\*-M95, southern Pinghua and northern Pinghua samples seldom share haplotypes; however, both of them share many more haplotypes with the southern ethnic groups than with other Han Chinese. The southern and northern Pinghua samples are in different subclades, suggesting different origins for the southern and northern Pinghua populations. In the network of O3\*-M122, most Daic samples are clustered in the left section, whereas Hmong-Mien samples are in the right section, indicating that the network is ethnically informative. Pinghua haplotypes appear in both sections. There is no haplotype shared or connected between southern and northern Pinghua populations in O3\*-M122. In the network of O3a2c1a-M117, Pinghua samples are related more closely to the Hmong-Mien and Daic samples, and genetic differences still exist in northern and southern Pinghua populations.

The mtDNA networks (Fig. 3: B) were drawn using HVS-I motif and restriction fragment length polymorphism information (Table S2). The reference data for the mtDNA HVS-I motifs are from published reports (Wen et al., 2005; Li et al., 2007; Hill et al., 2007; Gan et al., 2008). In the network of B4a, most Daic and Hmong-Mien samples are in the center whereas Pinghua and other Han Chinese samples are in the margin, indicating that most of the Pinghua people with the B4a haplogroup were derived from southern ethnic groups. Southern and northern Pinghua populations do not share haplotypes with each other. Most northern Pinghua samples share haplotypes with or connect to Daic or Han samples, whereas southern Pinghua samples only share haplotypes with Daic

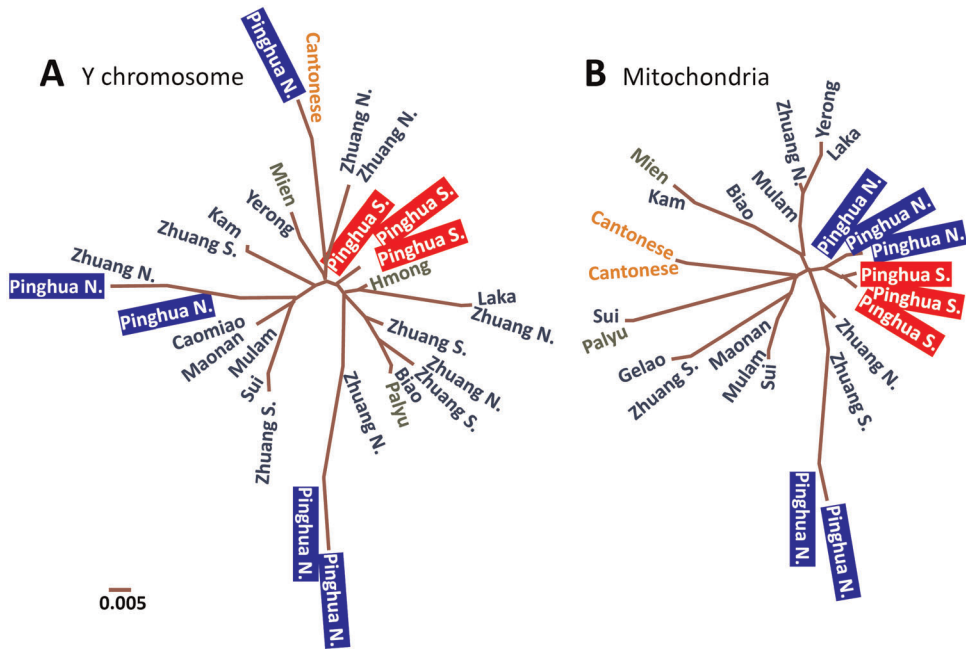


Fig. 2. Neighbor-joining trees of the Guangxi and Guangdong populations based on Y chromosome (A) and mitochondrial DNA (B) data.

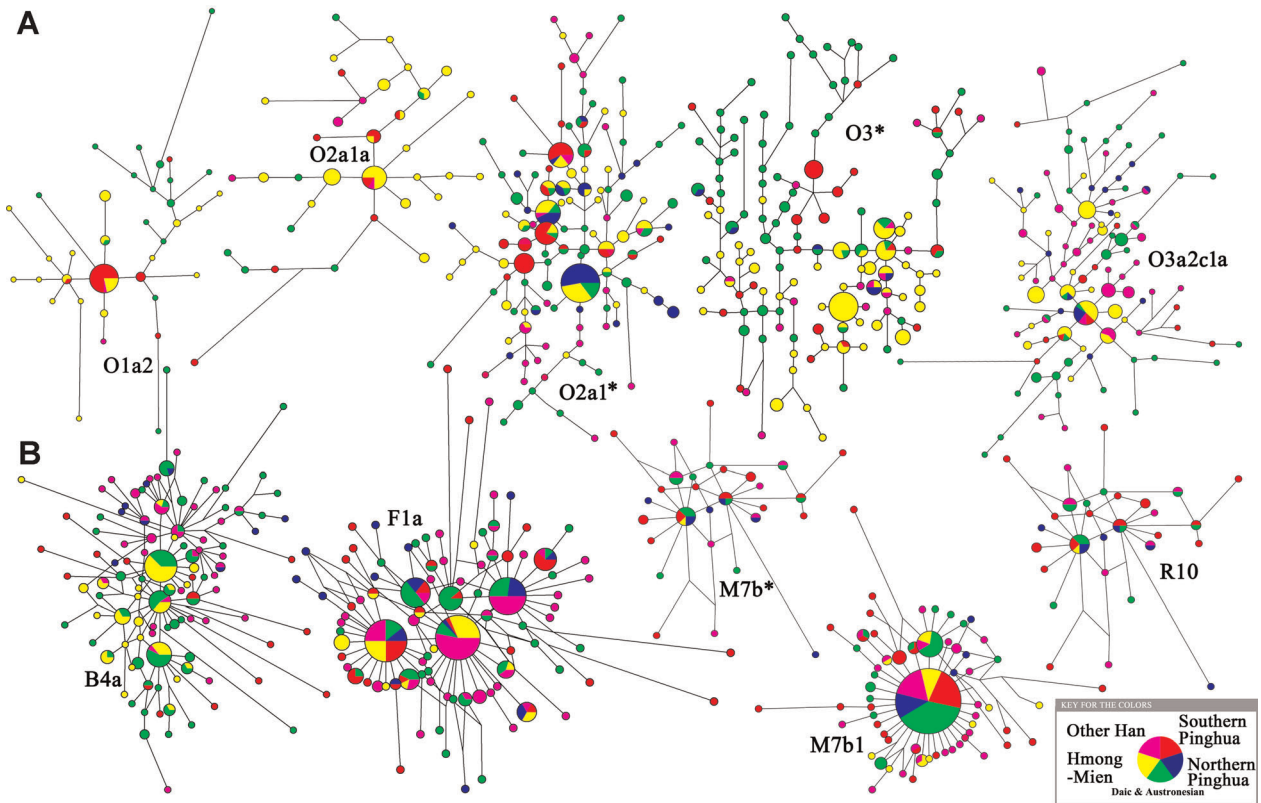


Fig. 3. Networks/shortest trees of Y chromosome (A) and mitochondrial DNA (B) haplogroups in southwest China.

samples. Interactions among the linguistic families can frequently be observed in the networks of F1a and M7b1, with almost all of the linguistic families sharing the central haplotypes, and fewer sharing the marginal haplotypes. Overall, the maternal lineages of both northern and southern Pinghua populations are closer to those of the southern ethnic groups and genetic distinctions are also observed in maternal lineages between the two Pinghua subgroups.

### 3 Discussion

In this study, both southern and northern Pinghua populations showed patterns of Y chromosome and mtDNA diversities different from other Han Chinese. Han Chinese were derived from the ancient Sino-Tibetan population who lived in the upriver basin of the Yellow River approximately 5000–6000 years ago. Subsequently, some Han Chinese populations migrated to the south in several waves because of warfare and famine. Although some of the southern indigenous populations were assimilated into the Han Chinese, the coherent genetic structure of Han Chinese can still be observed (Wen et al., 2004a). However, there were also southward Han migrations with small populations. When a small group of Han migrants arrived in the region mainly occupied by indigenous populations, such as Guangxi, the cultural assimilation of indigenous people by Han Chinese would take place. This phenomenon reflects a prominent concept in linguistics, which is language shift — members of a region in which more than one language is spoken abandon their original vernacular language in favor of another (Tsitsipis, 2004). In this case, the language shift of Pinghua people was probably “elite dominated” (Renfrew, 1991), whereby the language of a small invading group is adopted by the larger resident population, either because it is imposed by force or because it is considered socially desirable to speak the language of the invaders. As the migrant groups were very small and/or did not mix extensively with the resident groups, in any event, the migrant groups had a negligible genetic impact on the resident groups (Nasidze et al., 2003). One to two thousand years ago, the Han culture was more advanced; the southern ethnic groups accepted the Han culture, learnt Han Chinese for communication, and subsequently became part of the Han population. Neither northern nor southern Pinghua Han are of Han origin, but originated in southern indigenous populations living in Guangxi before the expansion of Han populations. The indigenous origin of Pinghua populations is also supported by

linguistic evidence. Earlier linguists noted the consistency of northern Pinghua and southern Pinghua dialects and indicated that the two Pinghua dialects are in the same linguistic unit (Li, 1989; Liang, 1997; Liang & Zhang, 1999). Although Pinghua populations with different genetic structures and mother languages learnt Han Chinese, the differences between what they finally mastered are not significant. A new dialect of Han Chinese with pronounced Guangxi indigenous accent finally occurred, which was Pinghua.

Despite the internal consistency within the Pinghua Chinese dialect, genetic non-conformity of northern and southern Pinghua populations was also observed, which might suggest that the origins of northern and southern Pinghua populations were not a single but multiple sources. This result was consistent with some historical documents. Since the Qin and Han Dynasties, there have been continuous waves of migrations from North China to Guangxi. A portion of migrations were caused by military deployment. Since the First Emperor of the Qin Dynasty (221–206 B.C.) dredged the Lingqu Canal to ensure the transportation for his army into northern Guangxi, the northern Pinghua populations had gradually formed from the Qin Dynasty to the Sui Dynasty. During the Song Dynasty, an army led by General Di Qing arrived in southern Guangxi for a counterinsurgency, leaving the soldiers in Guangxi. The Song Dynasty was an important period in the formation of southern Pinghua populations. Another source of migrants were civilians who intended to flee the war. Different waves of migrants were from different regions and would arrive in different destinations, and they were assimilated by different southern indigenous populations, such as Zhuang and Kam people of the Daic group, and Mien and Hmong people of the Hmongic group. As a result, the genetic structures of northern and southern Pinghua populations were inherited from various southern indigenous populations. That is to say, northern and southern Pinghua populations experienced independent processes of language shift toward the same destination.

**Acknowledgements** This work was partly supported by the National Natural Science Foundation of China (Grant Nos. 31071098, 91131002), the National Excellent Youth Science Foundation of China (Grant No. 31222030), Shanghai Rising-Star Program (Grant No. 12QA1400300), Shanghai Commission of Education Research Innovation Key Project (Grant No. 11zz04), and Shanghai Professional Development Funding (Grant No. 2010001).

## References

- Bandelt HJ, Forster P, Rohlf A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16: 37–48.
- Cao ZY. 2008. Linguistic atlas of Chinese dialects. Beijing: Commercial Press.
- Chen J, Li H, Qin ZD, Liu WH, Lin WX, Yin RX, Jin L, Pan SL. 2006. Y-chromosome genotyping and genetic structure of Zhuang populations. *Acta Genetica Sinica* 33: 1060–1072.
- Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1: 47–50.
- Feng H. 2007. General history of Yao nationality. Beijing: The Ethnic Publishing House.
- Gan RJ, Pan SL, Mustavich LF, Qin ZD, Cai XY, Qian J, Liu CW, Peng JH, Li SL, Xu JS, Jin L, Li H. 2008. Pinghua population as an exception of Han Chinese's coherent genetic structure. *Journal of Human Genetics* 53: 303–313.
- Hill C, Soares P, Mormina M, Macaulay V, Clarke D, Blumbach PB, Vizuete-Forster M, Forster P, Bulbeck D, Oppenheimer S, Richards M. 2007. A mitochondrial stratigraphy for island southeast Asia. *American Journal of Human Genetics* 80: 29–43.
- Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Research* 18: 830–838.
- Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt HJ, Villems R. 2002. The emerging limbs and twigs of the East Asian mtDNA tree. *Molecular Biology and Evolution* 19: 1737–1751.
- Li H. 2005. Genetic structure of Austro-Tai populations. Shanghai: Fudan University Press.
- Li H, Cai X, Winograd-Cort ER, Wen B, Cheng X, Qin Z, Liu W, Liu Y, Pan S, Qian J, Tan CC, Jin L. 2007. Mitochondrial DNA diversity and population differentiation in southern East Asia. *American Journal of Physical Anthropology* 134: 481–488.
- Li H, Pan WY, Wen B, Yang NN, Jin JZ, Jin L, Lu DR. 2003. Origin of Hakka and Hakkane: A genetics analysis. *Acta Genetica Sinica* 30: 873–880.
- Li H, Wen B, Chen SJ, Su B, Pramojnago P, Liu Y, Pan S, Qin Z, Liu W, Cheng X, Yang N, Li X, Tran D, Lu D, Hsu MT, Dekar, Marzuki S, Tan CC, Jin L. 2008. Paternal genetic affinity between western Austronesians and Daic populations. *BMC Evolutionary Biology* 8: 146.
- Li LJ. 2007. The distribution, subgrouping and category of Pinghua language. *Fangyan (Dialect)* 1: 71–78. (in Chinese)
- Li R. 1989. Subgrouping of Chinese dialects. *Fangyan (Dialect)* 4: 241–259. (in Chinese)
- Liang JR. 1997. Studies in phonetics of Pinghua in North Guangxi. Guangzhou: Jinan University Press. (in Chinese)
- Liang M, Zhang JR. 1999. Introduction of Guangxi Pinghua. *Fangyan (Dialect)* 1: 24–32. (in Chinese)
- Lin YW, Hsu LC, Kuo PL, Huang WJ, Chiang HS, Yeh SD, Hsu TY, Yu YH, Hsiao KN, Cantor RM, Yen PH. 2007. Partial duplication at AZFc on the Y chromosome is a risk factor for impaired spermatogenesis in Han Chinese in Taiwan. *Human Mutation* 28: 486–494.
- Nasidze I, Sarkisian T, Kerimov A, Stoneking M. 2003. Testing hypotheses of language replacement in the Caucasus: Evidence from the Y-chromosome. *Human Genetics* 112: 255–261.
- Qin YX. 2002. Study of Pinghua in North Guangxi. Guangzhou: Jinan University. (in Chinese)
- Qin Z, Yang Y, Kang L, Yan S, Cho K, Cai X, Lu Y, Zheng H, Zhu D, Fei D, Li S, Jin L, Li H. 2010. A mitochondrial revelation of early human migrations to the Tibetan Plateau before and after the last glacial maximum. *American Journal of Physical Anthropology* 143: 555–569.
- Renfrew C. 1991. Before Babel: Speculations on the origins of linguistic diversity. *Cambridge Archaeological Journal* 1: 13–23.
- Shi H, Dong YL, Wen B, Xiao CJ, Underhill PA, Shen PD, Chakraborty R, Jin L, Su B. 2005. Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. *American Journal of Human Genetics* 77: 408–419.
- Su B, Xiao J, Underhill P, Dekar, 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. *American Journal of Human Genetics* 65: 1718–1724.
- Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24: 1596–1599.
- Tsai LC, Lin CY, Lee JC, Chang JG, Linacre A, Goodwin W. 2001. Sequence polymorphism of mitochondrial D-loop DNA in the Taiwanese Han population. *Forensic Science International* 119: 239–247.
- Tsitsipis LD. 2004. A phenomenological view of language shift. *Collegium Antropologicum* 28 (Suppl. 1): 55–62.
- Wang FT. 2001. The attribution of Pinghua, South Hunan and North Guangdong native dialects. *Fangyan (Dialect)* 2: 107–118. (in Chinese)
- Wei SG. 1996. Preliminary exploration of the position of Pinghua in Chinese dialects. *Language Research* 2: 95–101. (in Chinese)
- Wen B, Li H, Gao S, Mao X, Gao Y, Li F, Zhang F, He Y, Dong Y, Zhang Y, Huang W, Jin J, Xiao C, Lu D, Chakraborty R, Su B, Dekar, Jin L. 2005. Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. *Molecular Biology and Evolution* 22: 725–734.
- Wen B, Li H, Lu D, Song X, Zhang F, He Y, Li F, Gao Y, Mao X, Zhang L, Qian J, Tan J, Jin J, Huang W, Dekar, Su B, Chakraborty R, Jin L. 2004a. Genetic evidence supports demic diffusion of Han culture. *Nature* 431: 302–305.
- Wen B, Xie X, Gao S, Li H, Shi H, Song X, Qian T, Xiao C, Jin J, Su B, Lu D, Chakraborty R, Jin L. 2004b. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *American Journal of Human Genetics* 74: 856–865.
- Wu W. 2001. Discussion on the Cantonese lineage of Pinghua in South Guangxi. *Fangyan (Dialect)* 2: 133–141. (in Chinese)

- Xie JY. 2001. The study of Pinghua in Guangxi. Beijing: Chinese Academy of Social Sciences. (in Chinese)
- Yan S, Wang C-C, Li H, Li S-L, Jin L. Genographic Consortium. 2011. An updated tree of Y-chromosome haplogroup O and revised phylogenetic positions of mutations P164 and PK4. *European Journal of Human Genetics* 19: 1013–1015.
- Yang HD, Liang ZS, Li PY, Liu CH. 1985. Chinese dialects in Guangxi. *Fangyan (Dialect)* 3: 181–190. (in Chinese)
- Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP. 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *American Journal of Human Genetics* 70: 635–651.
- Zhan BH, Cui SH, Liu XZ, Yang W. 2003. The ascription of Pinghua in Guangxi. *Linguistic Researches* 3: 47–52.
- Zhang JR. 1988. The influence of Pinghua Chinese on local Daic populations' languages. *Minority Languages of China* 3: 51–56. (in Chinese)
- Zhang JR, Liang M. 1996. Guangxi Pinghua. *Guangxi Ethnic Studies* 2: 96–101. (in Chinese)
- Zhang ZX. 1997. Rereading the atlas of Chinese language. *Fangyan (Dialect)* 4: 241–248. (in Chinese)

## Supplementary Material

The following supplementary material is available for this article at <http://onlinelibrary.wiley.com/doi/10.1111/jse.12003/supinfo>:

**Table S1.** Y chromosome haplotype information for southern Pinghua populations.

**Table S2.** Mitochondrial DNA haplotype information for southern Pinghua populations.

## Appendix I Members of the Genographic Consortium

- Janet S. ZIEGLE (Applied Biosystems, Foster City, CA, USA)
- Li JIN (Fudan University, Shanghai, China)
- Hui LI (Fudan University, Shanghai, China)
- Shilin LI (Fudan University, Shanghai, China)
- Pandikumar SWAMIKRISHNAN (IBM, Somers, NY, USA)
- Laxmi PARIDA (IBM, Yorktown Heights, NY, USA)
- Daniel E. PLATT (IBM, Yorktown Heights, NY, USA)
- Ajay K. ROYYURU (IBM, Yorktown Heights, NY, USA)
- Lluís QUINTANA-MURCI (Institut Pasteur, Paris, France)

- R. John MITCHELL (La Trobe University, Melbourne, Australia)
- Marc HABER (Lebanese American University, Beirut, Lebanon)
- Pierre A. ZALLOUA (Lebanese American University, Beirut, Lebanon)
- Syama ADHIKARLA (Madurai Kamaraj University, Madurai, India)
- ArunKumar GANESHPRASAD (Madurai Kamaraj University, Madurai, India)
- Ramasamy PITCHAPPAN (Madurai Kamaraj University, Madurai, India)
- Varatharajan Santhakumari ARUN (Madurai Kamaraj University, Madurai, India)
- R. Spencer WELLS (National Geographic Society, Washington DC, USA)
- Himla SOODYALL (National Health Laboratory Service, Johannesburg, South Africa)
- Elena BALANOVSKA (Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russia)
- Oleg BALANOVSKY (Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russia)
- Chris TYLER-SMITH (The Wellcome Trust Sanger Institute, Hinxton, UK)
- Daniela R. LACERDA (Universidade Federal de Minas Gerais, Belo Horizonte, Brazil)
- Fabício R. SANTOS (Universidade Federal de Minas Gerais, Belo Horizonte, Brazil)
- Pedro Paulo VIEIRA (Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil)
- Jaume BERTRANPETIT (Universitat Pompeu Fabra, Barcelona, Spain)
- David COMAS (Universitat Pompeu Fabra, Barcelona, Spain)
- Begoña MARTÍNEZ-CRUZ (Universitat Pompeu Fabra, Barcelona, Spain)
- David F. SORIA-HERNANZ (Universitat Pompeu Fabra, Barcelona, Spain)
- Christina J. ADLER (University of Adelaide, Adelaide, Australia)
- Alan COOPER (University of Adelaide, Adelaide, Australia)
- Clio S. I. Der SARKISSIAN (University of Adelaide, Adelaide, Australia)
- Wolfgang HAAK (University of Adelaide, Adelaide, Australia)
- Matthew E. KAPLAN (University of Arizona, Tucson, AZ, USA)
- Nirav C. MERCHANT (University of Arizona, Tucson, AZ, USA)



Colin RENFREW (University of Cambridge, Cambridge, UK)

Andrew C. CLARKE (University of Otago, Dunedin, New Zealand)

Elizabeth A. MATISOO-SMITH (University of Otago, Dunedin, New Zealand)

Matthew C. DULIK (University of Pennsylvania, Philadelphia, PA, USA)

Jill B. GAIESKI (University of Pennsylvania, Philadelphia, PA, USA)

Amanda C. OWINGS (University of Pennsylvania, Philadelphia, PA, USA)

Theodore G. SCHURR (University of Pennsylvania, Philadelphia, PA, USA)

Miguel G. VILAR (University of Pennsylvania, Philadelphia, PA, USA)