

## SHORT COMMUNICATION

# Gene flow between Zhuang and Han populations in the China–Vietnam borderland

Qing Zhao<sup>1,2,4</sup>, Shangling Pan<sup>1,4</sup>, Zhendong Qin<sup>2</sup>, Xiaoyun Cai<sup>2</sup>, Yan Lu<sup>2</sup>, Sara E Farina<sup>2</sup>, Chengwu Liu<sup>1</sup>, Junhua Peng<sup>1</sup>, Jieshun Xu<sup>3</sup>, Ruixing Yin<sup>1</sup>, Shilin Li<sup>2</sup>, Li Jin<sup>2</sup> and Hui Li<sup>2</sup>

In epidemiological studies, it is easier to glean information from isolated populations with simple genetic backgrounds than from populations with relatively complex genetic backgrounds. Regarded as an isolated population, the Minz (Hei-Yi Zhuang) in Napo County (in the China–Vietnam borderland) has been the subject of numerous recent studies. To assess the genetic isolation of this population, we studied Y chromosome and mitochondrial DNA from Minz and Han in Napo. Both populations exhibited southern Chinese genetic characteristics in both paternal and maternal lineages. Using NETWORK analyses, we found pronounced gene flow between these two populations only in maternal lineages. We also observed gene flow from other populations into the Minz gene pool. We thus concluded that the Minz is not a genetically isolated population and it is increasingly difficult to find isolated populations in mainland East Asia.

*Journal of Human Genetics* (2010) 55, 774–776; doi:10.1038/jhg.2010.107; published online 2 September 2010

**Keywords:** isolated population; Minz Zhuang; mitochondrial DNA; Southeast Asians; Y chromosome

Gene flow among populations in mainland China is very common; thus, it is difficult to find a population segregated from gene flow (that is, a population that maintains its simple genetic background). In molecular epidemiological studies, genetic samples from populations with simpler genetic backgrounds are more useful than those of populations with complex genetic diversity. Epidemiologists therefore constantly seek out isolated populations. About 10 years ago, many researchers began to take note of a special population, the Hei-Yi Zhuang (Black Clothes Zhuang).<sup>1</sup> This population resides in the mountain area of the China–Vietnam borderland, and it has maintained most of the Zhuang ethnic group's cultural traditions though other Zhuang populations have not. This group also claims to be strictly endogamous, which could result in genetic isolation. Therefore, many epidemiological studies have been conducted in this group.<sup>2</sup> However, the isolation of this population has never been genetically proven.

The Hei-Yi Zhuang, alternatively called 'Minz',<sup>3</sup> resides in Napo County, located in southwest Guangxi Zhuang Autonomous Region. The neighboring Pinghua Han Chinese, Mien, Kimmun, Hmong, Lolo, and Yerong<sup>3</sup> can also be found in different areas of Napo County. To detect possible gene flow between the Minz Zhuang and Han (Gaoshan Pinghua) in Napo (Supplementary Figure 1), we studied Y chromosome and mitochondrial DNA (mtDNA) polymorphisms from these two populations. Samples from 130 Minz (63 males and

67 females) and 205 Han (82 males and 123 females) were used in this study. Our study was approved by the Fudan School of Life Sciences Ethics Committee. All subjects signed informed consents. We employed the same methods used in our previous Pinghua population study.<sup>4</sup> Thirty single-nucleotide polymorphisms and 14 short tandem repeats (STRs) on the Y chromosome were typed (Supplementary Table 1). The hypervariable segments of mtDNA were sequenced, and 21 coding region single-nucleotide polymorphisms were typed using a SNaPshot assay (Applied Biosystems, Foster City, CA, USA; Supplementary Table 2). All sequences have been submitted to the GenBank (accession numbers: GU108620–947). The haplogroups were determined according to the most updated nomenclature for Y chromosome<sup>5</sup> and mtDNA.<sup>6</sup>

The Y chromosomes from the Minz samples were clustered into 12 haplogroups (Table 1). Four of these haplogroups, O1a\*, O2a\*, C3\* and D1, have high frequencies in the Minz samples. O1a\* and O2a\* are dominant in the Daic ethnic group, which includes the Zhuang,<sup>7</sup> suggesting that the Minz are a typical Daic population. C3\* is distributed widely across East Asia with a low frequency in most populations except for North Asian Altaic populations.<sup>8</sup> D1 is common in Tibet and neighboring areas, but is very rare in Southeast Asia.<sup>9</sup> The moderate frequencies of these two haplogroups among the Minz may result from the genetic drift of certain ancestral contributors to the Minz. The Han sample from Napo only formed eight haplogroups (Table 1), and all of these haplogroups are shared with

<sup>1</sup>Department of Pathophysiology and Institute of Cardiovascular Disease, Guangxi Medical University, Nanning, China; <sup>2</sup>MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China and <sup>3</sup>School of Ethnology and Sociology, Guangxi University for Nationalities, Nanning, China

<sup>4</sup>These authors contributed equally to this work.

Correspondence: Professor H Li, MOE Key Laboratory of Contemporary Anthropology, Fudan School of Life Sciences, 220 Handan Road, 308 Yifu Keji Building, Shanghai 200433, China.

E-mail: LiHui.Fudan@gmail.com

Received 13 April 2010; revised 29 June 2010; accepted 4 August 2010; published online 2 September 2010

**Table 1** Y-chromosome and mitochondrial DNA (mtDNA) haplogroup frequencies among Minz Zhuang and Napo Han populations

|              |             | Y-chromosome haplogroups (%) |       |        |       |       |       |        |       |      |         |        |        |      |      |
|--------------|-------------|------------------------------|-------|--------|-------|-------|-------|--------|-------|------|---------|--------|--------|------|------|
| Population   | Sample size | C3*                          | D1    | N1*    | O*    | O1a*  | O2*   | O2a*   | O2a1* | O3*  | O3a3b1* | O3a3c* | O3a3c1 |      |      |
|              |             | M130                         | M15   | LLY22g | M175  | M119  | P31   | M95    | M88   | M122 | M188    | M134   | M117   |      |      |
| Zhuang, Minz | 63          | 12.70                        | 11.11 | 6.35   | 4.76  | 19.05 | 1.59  | 19.05  | 3.17  | 6.35 | 4.76    | 3.17   | 7.94   |      |      |
| Han, Napo    | 82          |                              |       |        | 9.76  | 9.76  |       | 53.66  | 1.22  | 4.88 | 2.44    | 1.22   | 17.07  |      |      |
|              |             | mtDNA haplogroups (%)        |       |        |       |       |       |        |       |      |         |        |        |      |      |
| Population   | Sample size | A                            | B*    | B4*    | B4a   | B4b1  | B4c1* | B4c1b3 | B5*   | B5a  | B5b     | C      | D*     | D4   | D5   |
|              |             | Zhuang, Minz                 | 130   | 6.15   |       |       | 6.92  |        | 3.08  | 0.77 |         | 3.85   | 0.77   | 2.31 | 2.31 |
| Han, Napo    | 205         |                              | 0.49  | 1.46   | 4.88  | 1.95  |       |        | 0.49  | 4.88 |         | 1.95   | 1.46   | 0.98 | 4.88 |
|              |             | mtDNA haplogroups (%)        |       |        |       |       |       |        |       |      |         |        |        |      |      |
| Population   | Sample size | F*                           | F1*   | F1a*   | F1a1* | F1a1a | F1a3  | F1a4   | F3*   | F3a  | G*      | G2     | M*     | M12  | M7*  |
|              |             | Zhuang, Minz                 | 130   | 3.85   |       | 3.08  | 4.62  | 3.08   | 0.77  |      |         |        | 2.31   | 6.15 | 7.69 |
| Han, Napo    | 205         | 1.46                         | 4.88  | 5.85   | 0.98  | 4.88  |       | 0.49   | 0.98  | 0.98 |         | 0.98   | 8.29   | 0.98 | 7.80 |
|              |             | mtDNA haplogroups (%)        |       |        |       |       |       |        |       |      |         |        |        |      |      |
| Population   | Sample size | M7a1                         | M7b*  | M7b3   | M7c1a | M8*   | M9a   | N9*    | N9a   | R*   | R9*     | R9b    | R9c    | Y    |      |
|              |             | Zhuang, Minz                 | 130   | 0.77   | 6.92  |       |       | 0.77   |       | 3.85 |         | 0.77   | 2.31   | 8.46 | 0.77 |
| Han, Napo    | 205         |                              | 17.07 | 0.49   | 0.98  |       | 2.93  | 0.49   | 2.44  | 0.49 | 4.39    | 7.80   | 0.98   |      |      |

the Minz. The haplogroup specific to Southeast Asia, O2a\*, reaches its highest frequency among the Napo Han. O3a3c1 is the second highest occurrence in this population. A previous study proved that the Han Chinese originated in North China, and the Y chromosomes of Han contain mostly haplogroup O3.<sup>10</sup> The Y-chromosome diversity of Napo Han indicated that this population has common Han Chinese characteristics but was also strongly influenced by indigenous populations. The genetic affinity between Napo Han and the indigenous Daic populations were further proven by principal component analysis using the SPSS 15.0 program. The principal component plot (Supplementary Figure 2a) distinguished a Han cluster (upper left quadrant) and a Daic cluster (lower right quadrant). The Minz and Napo Han are both included in the Daic cluster, showing that the major paternal proportions of these two population samples are indigenous groups from the south.

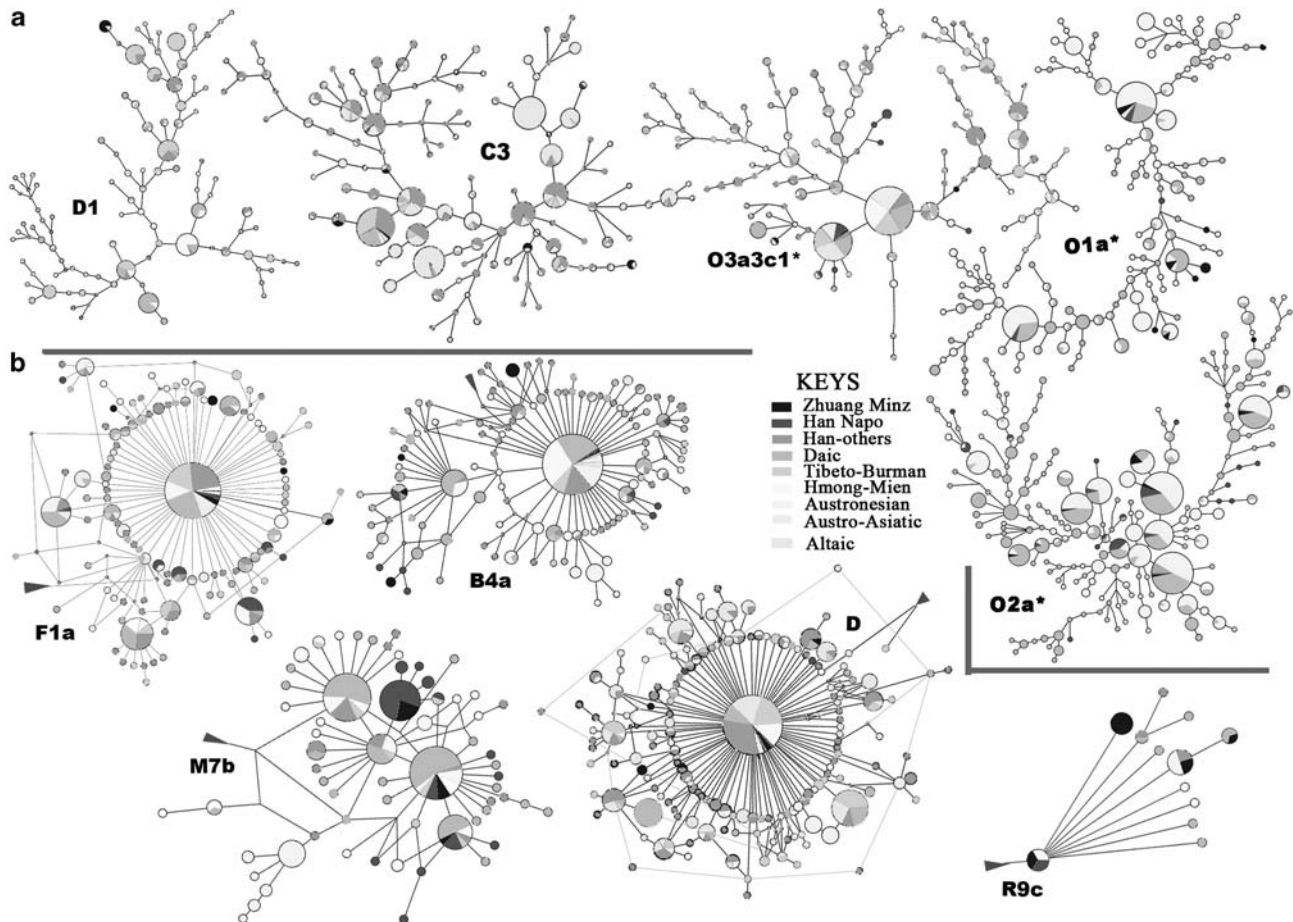
On the maternal side, mtDNA haplogroups for the Minz are mainly D4, R9c, M\*, M7b\* and B4a, whereas those for the Napo Han are mainly M7b\*, M\*, M7\* and R9b. These haplogroups are predominantly derived from southern China<sup>11</sup> and are not Han dominant (for example, A, C, D, G, M8a, Y and Z<sup>10</sup>). The principal component plot (Supplementary Figure 2b) also illustrates that the Minz and Napo Han are in the southern China cluster rather than the Han Chinese cluster. Notably, the Minz and Napo Han are very close to each other in both plots, indicating possible gene flow between the two groups.

To further investigate gene flow, we applied network analyses (NETWORK 4.516) to both Y-chromosome STR haplotypes and mtDNA hypervariable segment motifs within the major haplogroups of the two populations (Figure 1). The Y haplogroup D1 in the Minz population only formed one STR haplotype. This haplotype is shared with the Hmong from Laos and is linked upstream to Sino-Tibetan populations, indicating that D1 lineages in the Minz might have undergone genetic drift through a circuitous route out of Sino-Tibetan

populations. Y haplogroup C3\* in the Minz formed five STR haplotypes suggesting an absence of genetic drift similar to that of D1. Other Y and mtDNA haplogroups show higher diversity in Napo populations. Interestingly, a Napo-specific clade was found in the mtDNA haplogroup M7b. The hypervariable segment motif for this clade is 16129-16192-16223-16297. This clade includes five haplotypes and forms a sun-shaped structure. All five haplotypes can be found among Napo Han, but only the central haplotype can be found among the Minz. This suggests gene flow from the Napo Han to the Minz. We counted the unique shared and connected haplotypes among populations in these networks to assess gene flow (Table 2). The result showed that no paternal gene flow can be detected between the Minz and Napo Han, but maternal gene flow is pronounced. Gene flow from the other Daic or Han populations to the Minz and Napo Han was also observed.

Genetic diversities of both populations were calculated. The Y-STR diversity of the Minz is lower than that of the Napo Han, but mtDNA diversities are almost the same (Supplementary Table 3), indicating that the gene flow in maternal lineages is stronger than that in paternal lineages.

Thus, we show that gene flow between the Han and Zhuang populations in Napo is pronounced, especially in maternal lineages. Cultural difference is not a barrier to gene flow. The Minz Zhuang is not a genetically isolated population. Their lineages may be from Daic populations, as well as from Sino-Tibetan and Hmong-Mien populations. Although the borderland between China and Vietnam is moderately segregated geographically, populations arrived in this area from various regions. Therefore, the Minz might not be a truly isolated population for epidemiological studies. It might be better to examine autosomal DNA polymorphisms to further assess the population admixture. However, genetic studies on the Minz are still very relevant, as the Minz can be viewed as a half-way point population in the migration of the Thai-Lao population from China.



**Figure 1** Networks of the major haplogroups in Napo Han and Zhuang. (a) Y chromosome, (b) mitochondrial DNA.

**Table 2** Unique shared and upstream connected haplotypes counted from the networks

|            | Mitochondrial DNA |           | Y    |           |
|------------|-------------------|-----------|------|-----------|
|            | Minz              | Han, Napo | Minz | Han, Napo |
| Minz       | /                 | 17        | /    | 0         |
| Han, Napo  | 7 <sup>a</sup>    | /         | 0    | /         |
| Other Daic | 3                 | 2         | 2    | 7         |
| Other Han  | 1                 | 2         | 0    | 5         |

<sup>a</sup>Individual numbers of the Minz Zhuang in the haplotypes shared only with or connected only upstream to the Napo Han. Other data were counted in the similar way.

## ACKNOWLEDGEMENTS

We especially thank all of those who donated samples. Our research was supported by the Natural Science Foundation of Shanghai (10ZR1402200) to HL; the National Natural Science Foundation of China (39993420) and Guangxi Natural Science Foundation (GSN0728160, GSY0542044) to SLP; the National Outstanding Youth Science Foundation of China (30625016), National Science Foundation of China (30890034), 863 Program (2007AA02Z312) and Shanghai Leading Academic Discipline Project (B111) to LJ; and National Science Foundation of China (30360038) to RY.

- He, M. T., Li, Y. T. & Li, Q. W. *Field Study of Heiyi Zhuang* (Guangxi Ethnic Press, Nanning, 1999).
- Yin, R. X., Li, H., Wu, J. Z., Lin, W. X., Yang, D. Z., Pan, S. L. *et al.* Effects of alcohol consumption and other lifestyle behaviors on blood pressure for the middle-aged and elderly in the Guangxi Hei Yi Zhuang and Han populations. *Alcohol* **41**, 541–550 (2007).
- Gordon Jr. R. G. *Ethnologue: Languages of the World*, 15th edn (SIL International, Dallas, Texas, 2005).
- Gan, R. J., Pan, S. L., Mustavich, L. F., Qin, Z. D., Cai, X. Y., Qian, J. *et al.* Pinghua population as an exception of Han Chinese's coherent genetic structure. *J. Hum. Genet.* **53**, 303–313 (2008).
- Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L. & Hammer, M. F. New binary polymorphisms reshape and increase resolution of the human Y chromosomal Haplogroup tree. *Genome Res.* **18**, 830–838 (2008).
- Kivisild, T., Tolk, H. V., Parik, J., Wang, Y., Papiha, S. S., Bandelt, H. J. *et al.* The emerging limbs and twigs of the East Asian mtDNA tree. *Mol. Biol. Evol.* **19**, 1737–1751 (2002).
- Li, H., Wen, B., Chen, S. J., Su, B., Pramoongjago, P., Liu, Y. *et al.* Paternal relationships between Austronesian and Daic populations. *BMC Evol. Biol.* **8**, 146 (2008).
- Su, B., Xiao, C., Deka, R., Seielstad, M. T., Kangwanpong, D., Xiao, J. *et al.* Y-chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum. Genet.* **107**, 582–590 (2000).
- Shi, H., Zhong, H., Peng, Y., Dong, Y. L., Qi, X. B., Zhang, F. *et al.* Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol.* **6**, 45 (2008).
- Wen, B., Li, H., Lu, D., Song, X., Zhang, F., He, Y. *et al.* Genetic evidence supports demic diffusion of Han culture. *Nature* **431**, 302–305 (2004).
- Li, H., Cai, X. Y., Winograd-Cort, E. R., Wen, B., Cheng, X., Qin, Z. D. *et al.* Mitochondrial DNA diversity and population differentiation in southern East Asia. *Am. J. Phys. Anthropol.* **134**, 481–488 (2007).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)