

ORIGINAL ARTICLE

Genetic origin of Kadai-speaking Gelong people on Hainan island viewed from Y chromosomes

Dongna Li¹, Yuantian Sun^{1,2}, Yan Lu², Laura F Mustavich³, Caiying Ou¹, Zhenjian Zhou¹, Shilin Li², Li Jin² and Hui Li^{2,3}

The government of China defined 56 official ethnicities for the ethnic groups in China for political purposes; however, there are many more than 56 ethnic groups. Therefore, similar groups must be pooled for registry, and the so-called ethnicity identification is an important official mission in China. Here, we showed how genetics can help in the ethnicity identification for the Gelong people on Hainan Island. The Gelong speak a Kadai language whose other speakers (officially of the Gelao ethnicity) are all far in the southwest of China. Being registered as a Han ethnicity, the Gelong lost all the benefits assigned to the minorities. Y-chromosome typing was performed in a sample of 78 individuals. Twenty single nucleotide polymorphisms (SNPs) and seven short tandem repeats (STRs) were typed and eight haplogroups were detected, among which haplogroup O1a* was the most dominant. Compared with the Y haplogroups of the populations in south China, the Gelong were found to be closest to the Gelao and the Hlai using principal components (PCs) analysis, dendrogram clustering and STR networks. The genetic similarity between the Gelong and the Hlai may have resulted from the gene flow during thousands of years of neighboring history on Hainan Island, whereas the similarity between the Gelong and the Gelao may have resulted from their common ancestry because there is less possibility of gene flow over such a far distance. As both linguistic and genetic evidence support the similarity between the Gelong and the Gelao, we suggest that the Gelong register as Gelao for their official ethnicity. However, this identification is invalid until it is accepted by the Gelong people themselves and the Hainan government.

Journal of Human Genetics (2010) 55, 462–468; doi:10.1038/jhg.2010.50; published online 20 May 2010

Keywords: China minorities; ethnicity identification; Tai-Kadai; Y chromosome

INTRODUCTION

Officially, there are 56 ethnicities in mainland China; however, this number is much smaller than the ethnic diversity in China. It is very difficult to define an ethnic population. Distinctive language may be one of the important criteria. At least 300 different languages can be found within mainland China.¹ Many populations with distinctive languages are not in the list of ‘Official Nationalities’ of China. (Here, the word ‘nationality’ does not mean the status of belonging to a particular nation, but is misused in China for a certain ethnicity recognized by the government. The idea of using ‘nationality’ was learned from the Socialist Nation theory of Stalin, and is still insisted upon by the Chinese government. Here, ‘Official Ethnicity’ may be a better word.) The Gelong people who speak the Cun language on the west coast (lower reaches of the Changhua River; Figure 1) of Hainan Island are one of the best examples.^{1–3} The Gelong are currently registered as Han Chinese; however their language, Cun, belongs to the Kadai branch of the Daic (also called Tai-Kadai) linguistic family, which is unrelated to the Chinese language. It will be very important for this population of ~80 000 people to determine their ‘Official Ethnicity’ in the registry, because minorities are given special treat-

ment in China. No additional minority groups are allowed to be added to the list of ‘Official Ethnicities.’ Therefore, the Gelong must choose to be one of the 55 official minorities.

As their language belongs to the Kadai linguistic group, Gelong people have two choices of official ethnicities: Gelao or Hlai. Hlai is the major ethnicity on Hainan Island,⁴ and their languages were formerly classified within Kadai, but were recently distinguished as an independent branch from Kadai.⁵ Although the Cun and Hlai languages are very different, they share some characters. Therefore, in the 1950s, some linguists suggested Gelong should be registered as Hlai. However, the distinct difference of culture and language between Hlai and Gelong made this suggestion unacceptable to the Gelong people. The speakers of the other Kadai languages, with the exception of Hlai, were mostly registered as the Gelao official ethnicity, except for some mistaken registrations; for example, the Buyang were pooled into Zhuang (a large Kam-Tai population); Yerong into Yao (a Hmong-Mien population); Lachi and Qabiao into Yi (a Tibeto-Burman population). Taxonomically, it is reasonable to register all the Kadai (non-Hlai) populations, including the Gelong, as Gelao. The problem for this claim is that all the other Kadai populations

¹Department of Biology, Hainan Medical College, Haikou, Hainan, China; ²State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China and ³Department of Genetics, School of Medicine, Yale University, New Haven, CT, USA
Correspondence: Professor H Li, MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, 220 Handan Road, Shanghai 200433, China.
E-mail: LiHui.Fudan@gmail.com

Received 5 February 2010; revised 14 April 2010; accepted 18 April 2010; published online 20 May 2010

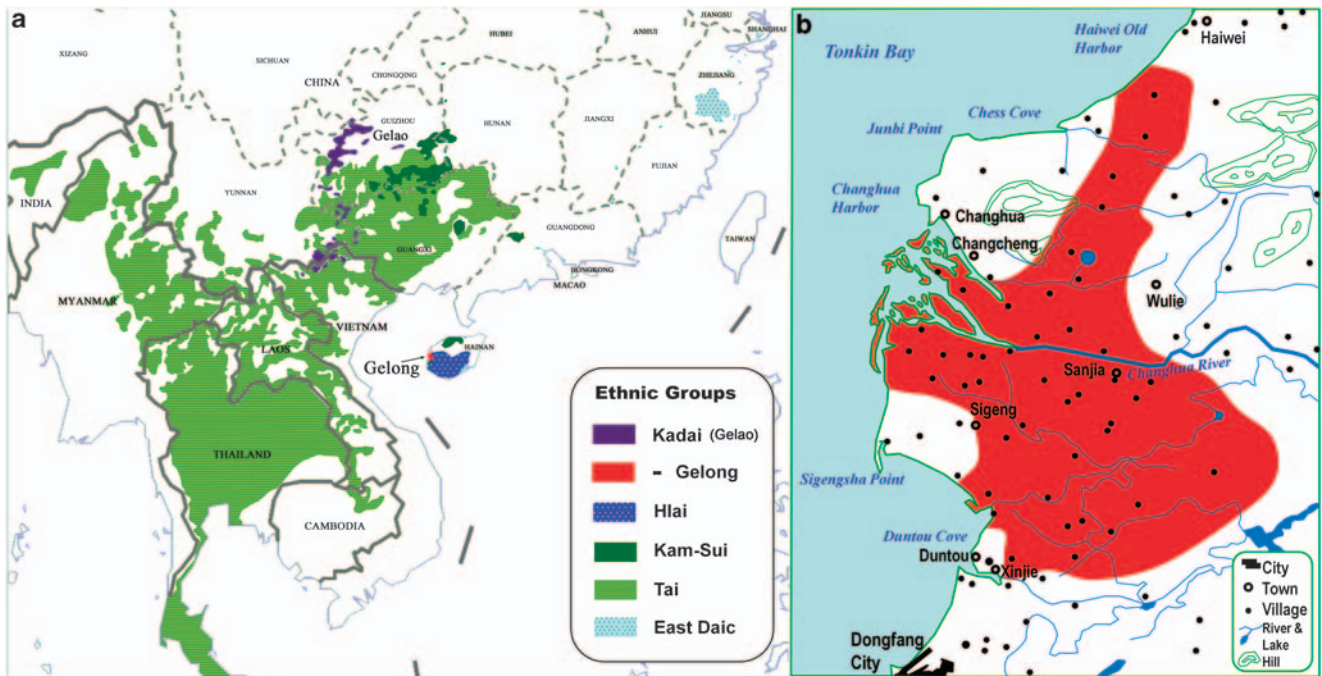


Figure 1 Geographic distribution of the subgroups of the Daic ethnic family (a) and the traditional distribution of the Gelong villages in the west of Hainan Island (b).

under the name of Gelao are in southwest China (Guizhou province, borderland of Vietnam, Yunnan and Guangxi) far away from Hainan Island; Figure 1a). Most of the Gelong people have never even heard of Gelao. Therefore, the Gelong do not identify with them.

Interestingly, the Gelong people use Han Chinese words to refer to male relatives and Cun words to refer to females. Evidence from some pedigree records further led some researchers to argue that the Gelong were mainly formed by Han Chinese male ancestry and indigenous female ancestry.^{2,3} This opinion led the Gelong people to be registered as Han Chinese. However, there has never been solid evidence to support the Han Chinese origin of Gelong paternal ancestry. Some Gelong people have also appealed to reconsider their official ethnicity registration. Genetic evidence may help to resolve this problem, and cluster the Gelong to Han, Hlai, Gelao, or some others.

The origin of the population includes the emergence of the culture, language and so on, and the population itself. Therefore, genetic evidence will be the most important for ethnic recognition. There are three types of genetic materials: maternal mitochondrial DNA, paternal Y chromosome and bi-lineage autosomal and X chromosome DNA. The non-recombining portion of the Y chromosome (NRY) is strictly inherited paternally, and is the best material to trace the paternal lineage of the population. The small effective population size, low mutation rate, sufficient single nucleotide polymorphism (SNP) and short tandem repeat (STR) markers, make NRY a strong tool for ethnic recognition.⁶⁻⁸ In this paper, we typed the relevant Y-chromosome markers of a Gelong population sample and analyzed the paternal origin of the Gelong people.

MATERIALS AND METHODS

Population sample

A demographic census was done among the traditional Gelong villages in Dongfang and Changjiang Counties in 2007 (Figure 1b). Blood samples of 78 male Gelong people were collected from Dongfang County, Hainan Province,

China. None of the subjects are related, or have ancestry of non-Gelong within the three generations. All the subjects signed the informed consents.

DNA preparation and Y-chromosome typing

The protocol of DNA preparation and genotyping of this study are the same as our earlier studies.⁴ We followed the nomenclature of the Y-Chromosome Consortium (<http://ycc.biosci.arizona.edu>).^{9,10} Fourteen SNPs in the Y-chromosome non-recombining portion were typed in the collected samples by polymerase chain reaction-RFLP (M130, M89, M9, M45, M119, M110, M101, P31, M95, M88, M122, M164, M159 and M7). Four SNPs (M48, M8, M217 and M356) were typed by Taqman (Applied Biosystems, Foster City, CA, USA). Seven SNPs (YAP, M15, M175, M111, M134, M117 and M121) and seven STR polymorphisms (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392 and DYS393) were typed by using fluorescently labeled primers for polymerase chain reaction amplification. Denatured products were separated by acrylamide gel electrophoresis through the use of an ABI 3100 genetic analyzer (Applied Biosystems) to distinguish the alleles. More than five repeat individual samples were typed for each marker to be quality controls. Inconsistency has never been seen among the repeat samples.

Data analyses

The pattern of Y-chromosome haplogroups of Gelong people was compared with those of the other population samples.^{4,11-15} The referential Gelao sample¹⁵ was collected from Longlin County in northwest Guangxi Province, which is believed to be the location of origin of the Gelao's expansion.¹⁶ The referential Hlai samples were collected from different branches of Hlai in Hainan Island.⁴ SPSS 13.0 was used to do the correlation analysis, principal component (PC) analysis, and dendrogram clustering (between-group clustering). NETWORK 4.510¹⁷ was used to draw the shortest tree of the O1a* STR haplotypes.

RESULTS

Demographic census of Gelong people

We investigated the population size in the traditional area of residence of the Gelong people in Dongfang and Changjiang Counties in the lower reaches of the Changhua River (Table 1). The Gelong people

Table 1 Population size of the Gelong villages in 2007

| City/county township | Village (number) | Family number | Individual number | |
|---------------------------------|------------------|----------------------|-------------------|------|
| <i>Dongfang city and county</i> | | | | |
| Xinjie | Baopo | 224 | 1264 | |
| | Tianzhuang | 181 | 795 | |
| | Yixing | 139 | 558 | |
| | Wentong | 184 | 1082 | |
| | Nadeng | 415 | 2409 | |
| | Laoguan (2) | 286 | 1195 | |
| | Pingling | 196 | 938 | |
| | Changyi | 11 | 60 | |
| | Yuzhang | 291 | 1600 | |
| | Sigeng | Tudi (2) | 207 | 1242 |
| | | Changshan | 448 | 2557 |
| | | Lainan | 234 | 1015 |
| | | Chikan | 299 | 1705 |
| | | (greater and lesser) | | |
| | | Daxin | 362 | 2181 |
| | | Rixin | 248 | 1357 |
| | | Juduo | 234 | 1248 |
| Yingxian | | 450 | 2260 | |
| Shacun | | 90 | 449 | |
| Sanjia | Danchang | 248 | 1343 | |
| | Danchangyuan (2) | 247 | 1035 | |
| | Yaoshang | 215 | 931 | |
| | Juhou | 365 | 1669 | |
| | Suanmei | 1002 | 4542 | |
| | Xiaosuanmei | 164 | 735 | |
| | Shuiliudong | 273 | 1500 | |
| | Daijiu | 212 | 892 | |
| | Hongcao | 604 | 2707 | |
| | Sanjia | 1140 | 5038 | |
| | Wanglao | 260 | 1172 | |
| | Yuxiong | 1492 | 6331 | |
| | Shangrong | 98 | 542 | |
| Xiarong | 75 | 508 | | |
| <i>Changjiang county</i> | | | | |
| Changhua | Jiuxian | 564 | 3104 | |
| | Naicun | 576 | 3166 | |
| | Dafeng | 654 | 3594 | |
| | Langbing | 269 | 1481 | |
| | Guangtian | 246 | 1355 | |
| | Huangjiang | 157 | 863 | |
| | Xianfeng | 74 | 409 | |
| | Haiwei | Jindong | 114 | 624 |
| Baisha | | 153 | 843 | |
| Total | | | 68299 | |

Non-Gelong families and individuals were excluded. Families living in the city were not counted.

mainly reside in this area with very few Han Chinese (Minnamese) and Hlai immigrants. The Gelong area is surrounded by Minnamese in the south and north and by Hlai in the east. Marriages between different ethnic people were found occasionally.

There are also some Gelong people living out of the traditional area, mainly in the downtowns and suburbs of Dongfang City and Changjiang County. Therefore, the total population of Gelong is approximately > 80000. There are 19 official ethnicities with populations smaller than the Gelong, according to the census of 2000 in China. The Gelao official ethnicity has a population of 579 357 (2000 census) in China, much more than the Gelong.

Table 2 Y-chromosome SNP haplogroup frequencies of the Gelong population

| Haplogroup | C3 | O1a* | O2* | O2a | O3* | O3a1 | O3a3c* | O3a3c1 |
|----------------|------|------|------|-----|------|------|--------|--------|
| Diagnostic SNP | M217 | M119 | P31 | M95 | M122 | M121 | M134 | M117 |
| Frequency (%) | 5.1 | 57.7 | 17.9 | 3.8 | 9.1 | 1.3 | 3.8 | 1.3 |

Abbreviation: SNP, single nucleotide polymorphism.

Y-chromosome SNP haplogroups of Gelong compared with other populations

According to the nomenclature of YCC,¹⁰ eight SNP haplogroups were determined from the 78 Gelong individual samples, among which O1a*, O2* and O3* are most dominant in decreasing order (Table 2). Over half of the samples were determined to be O1a*, which is specific to the Daic and Austronesian populations,^{15,18} indicating that the Gelong has a clear Daic genetic background.

Some SNPs that we typed were seldom typed for the Chinese population samples reported in the literature. To make the data comparable, some haplogroups of our sample were pooled; for example, O2*-P31 into K-M9, O3a3c1-M117 into O3a3c-M134. In the phylogenetic tree of Y-chromosome haplogroups, O2 is a branch of O, and O is a branch of K. If the markers P31 and M175, defining haplogroups O2 and O, are not typed for the actual O2* individuals, only the derived allele of M9 can be seen, so the individual samples will be counted in haplogroup K. O3a3c1 is pooled into O3a3c for the same reason. The frequencies of O3a3c1 and O3a3c* are very low in the Gelong, Gelao and Hlai samples; therefore, this pooling may not lead to serious bias in the subsequent analyses. We recently typed P31 in K-M9 individuals of our Han Chinese and Gelao populations, but did not find any O2* individuals in Han populations. Similar to the Gelong, the Gelao have haplogroup O2*, but do not have haplogroups K* or O*. The comparable haplogroup frequencies of Gelong and other reference population samples^{4,11–15} are in Table 3. Clearly, the haplogroup pattern of the Gelong is most similar to that of the Gelao, and also fairly similar to the Taiwan aborigines.

Correlation analysis of the SNP haplogroup patterns

To estimate the similarity of the haplogroup patterns of the population samples, correlation analysis was performed (Table 4). The patterns of the Gelong and the Gelao are significantly correlated ($r=0.962$, $P<0.001$), indicating that these two populations have almost the same Y-chromosome SNP patterns. The Taiwan aborigines are also significantly correlated to the Gelong. No significant correlations can be found between the Gelong and the other populations we studied, including the Hlai, the neighboring population of the Gelong. The Han Chinese from Fujian and Mien from Guangxi are the most different populations from the Gelong.

PC analysis of SNP haplogroup patterns

PC analysis can show the relationships among the populations based on the frequency data. As recombination never occurs in the region of the Y chromosome that we studied, the Y chromosome can be regarded as one single locus. PC analysis cannot deal with the individual data of a single locus. Therefore, we applied the analysis to the haplogroup frequency data of the population samples. Plots of the populations by the largest four components were illustrated in Figure 2. The first two components explain 61.9% of the data variance, and the first four components explain 86.6%. Every component is associated with different ethnic groups; for example, the first

Table 3 Comparison of Gelong to other populations using Y-SNP haplogroup frequencies

| Population | Sample size | Haplogroup frequency (%) | | | | | | | | | | | |
|------------|-------------|--------------------------|------|------|------|-----|------|------|------|------|------|-------|-------|
| | | C | D | F | K | P* | O1a* | O1a2 | O2a* | O2a1 | O3* | O3a3b | O3a3c |
| | | M130 | YAP | M89 | M9 | M45 | M119 | M110 | M95 | M88 | M122 | M7 | M134 |
| Gelong | 79 | 5.1 | | | 17.9 | | 57.7 | | 3.8 | | 10.4 | | 5.1 |
| Gelao | 30 | | | | 16.7 | | 60.0 | | 16.7 | | 3.3 | | 3.3 |
| Hlai | 11 | | | | | | 27.3 | | 54.6 | 9.1 | | | 9.1 |
| Atayal | 24 | | | | | | 54.2 | 8.3 | | | 29.2 | 4.2 | 4.2 |
| Paiwan | 11 | | | | | | 54.6 | 27.3 | | | | | 18.2 |
| Bouyei | 49 | 7.0 | | 4.4 | 17.7 | | 4.4 | | 46.7 | 11.1 | 4.4 | 2.2 | 2.2 |
| Kam | 10 | 20.0 | | | | | 20.0 | 10.0 | 20.0 | | 10.0 | | 20.0 |
| Sui | 92 | | | | 5.4 | | 31.5 | | 58.7 | | | | 4.4 |
| Hmong | 49 | 8.2 | 2.1 | | 28.6 | | 12.3 | | 16.3 | 2.1 | 24.5 | 4.1 | 2.1 |
| Bunu | 10 | 20.0 | 30.0 | | | | | | 40.0 | | 10.0 | | |
| Zhuang | 28 | 3.6 | 3.6 | 7.1 | 3.6 | | 17.9 | | 25.0 | 10.7 | 3.6 | | 25.0 |
| Han-JX | 21 | 4.8 | 4.8 | 9.5 | 19.1 | | 14.3 | | 4.8 | | 19.1 | | 23.8 |
| Han-FJ | 13 | 7.7 | | | 7.7 | | | | | 7.7 | 38.5 | | 38.5 |
| Han-HaN | 15 | 6.7 | | | 6.7 | | 6.7 | | 13.3 | | 33.3 | | 33.3 |
| Han-HeN | 28 | 7.1 | | 3.6 | 25.0 | 7.1 | 10.7 | | | | 32.1 | | 14.3 |
| Thai | 20 | | | 5.0 | 5.0 | 5.0 | 5.0 | 5.0 | 45.0 | 20.0 | 5.0 | 5.0 | |
| Cambodian | 26 | 3.8 | 3.8 | 11.5 | 11.5 | 7.7 | 3.8 | 3.8 | 23.1 | 11.5 | 3.8 | | 15.4 |
| Malaysian | 13 | | | 7.7 | 7.7 | | 7.7 | 23.1 | 7.7 | | 30.8 | | 15.4 |

Abbreviations: FJ, Fujian province; HaN, Hainan province; HeN, Henan province; JX, Jiangxi province; SNP, single nucleotide polymorphism.

Table 4 Correlation analysis of Y-haplogroup frequencies between the Gelong and other populations

| Pairs | R | P-value |
|---------------|--------|---------|
| Gelong-Gelao | 0.962 | <0.001 |
| Gelong-Hlai | 0.329 | 0.296 |
| Gelong-Zhuang | 0.321 | 0.309 |
| Gelong-Bouyei | 0.003 | 0.993 |
| Gelong-Kam | 0.414 | 0.180 |
| Gelong-Sui | 0.401 | 0.196 |
| Gelong-Atayal | 0.866 | <0.001 |
| Gelong-Paiwan | 0.776 | 0.003 |
| Gelong-Han-JX | 0.443 | 0.149 |
| Gelong-Han-FJ | -0.023 | 0.944 |
| Gelong-Hmong | 0.417 | 0.177 |
| Gelong-Mien | -0.187 | 0.561 |

Abbreviations: FJ, Fujian province; JX, Jiangxi province.

component with the Daic, the second with the Taiwan aborigines, the third with the Han Chinese and the fourth with the Hmong-Mien. In both plots, the Gelong is very close to the Gelao, especially in the second plot of the third and fourth components, where the two populations are almost overlapped. However, in the plot of PC1-PC2, Gelong is also close to the Taiwan aborigines (Paiwan and Atayal) and Jiangxi Han. Therefore, it is better to draw conclusions from the results of multiple analyses.

Dendrogram of the populations

Dendrogram clustering can display the relationships among the populations in a different way from PC analysis by giving an overall structure. We applied the dendrogram clustering analysis to the same data set as in the PC analysis (Figure 3). Four clusters were created. The Kam-Tai branch of Daic, including the Sui, Hlai, Bouyei and Thai

fell into one cluster, with Mien close to them. The other two Kam-Tai populations, Zhuang and Kam, were clustered with the Cambodians. The Han Chinese and Hmong were clustered with Malays. The last cluster contained the Gelong, Gelao and two Taiwan aboriginal populations. The Gelong and the Gelao are the closest to each other in the dendrogram. This result agrees with the results of the correlation analysis and PC analysis.

STR diversity and network

The seven STR polymorphisms were typed and the haplotype information was integrated with the SNP haplogroups (Table 5). Different SNP haplogroups never share the same STR haplotype, showing that these seven STR polymorphisms provided enough information for the haplotype diversification. The STR haplotypes were not equally distributed in the population sample. Two of the haplotypes under O1a* and O2* made up half of the samples. These two SNP-STR haplotypes might be the original haplotypes of the Gelong people, and therefore require more detailed analyses. We recently typed O2*-P31 in our population samples and found derived alleles in Gelao, Hlai and other Daic populations, providing comparable data for further analyses.

We constructed two shortest trees using the NETWORK program (Figure 4). These trees showed the most probable relationships among the STR haplotypes within haplogroups O2* and O1a*. In the O1a* tree, most of the Hlai haplotypes are in a clade that is distinct from the mainland haplotypes. Interestingly, the largest node (Hap02) of the Gelong is just between the Hlai clade and the mainland clade, connecting the Hlai and the Kadai in the tree. In the O2* tree, Kadai and Hlai nodes are scattered in different clades. The largest node (Hap19) of the Gelong is between a Kadai node and a Sui (in Kam-Tai subfamily) node. In both trees, the Gelong nodes are all close to the Kadai nodes. We subsequently counted the shared haplotypes and the neighboring haplotypes between the Gelong and other Daic ethnic branches (Table 6). Similar haplotypes between the Gelong and other

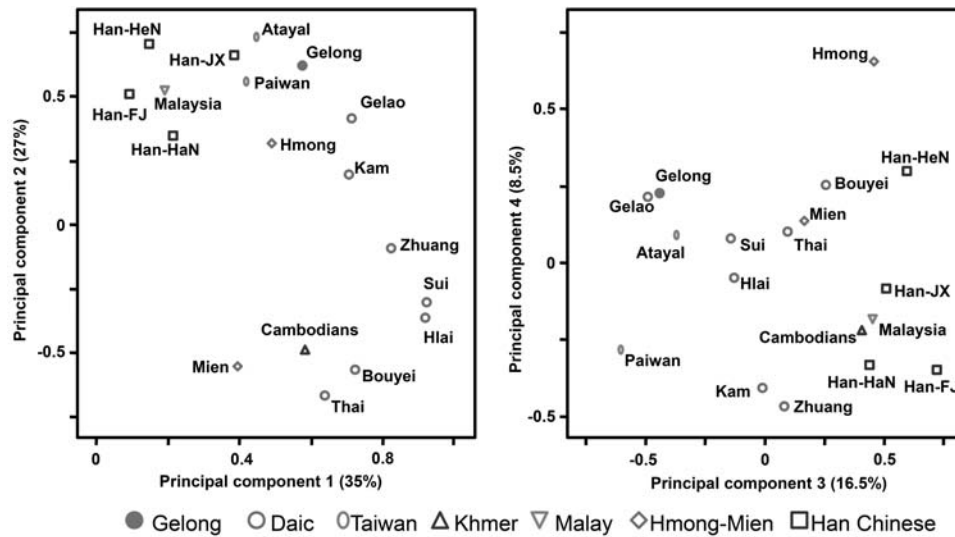


Figure 2 Principal component plots of the populations concerned.

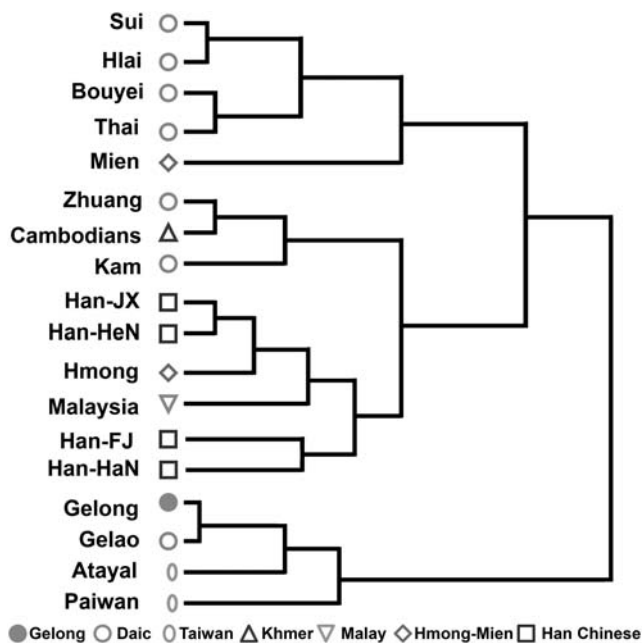


Figure 3 Dendrogram of the populations based on the Y-SNP haplogroup frequencies.

Kadai populations are possessed by 48 Gelong individuals, whereas fewer Gelong individuals are similar to the Hlai and other branches. However, the similarity between Gelong and Hlai is also pronounced, which is reasonable because they have been neighbors for thousands of years, and gene flows between them must have taken place. Overall, the Gelong are much more related to the Kadai than to the other Daic branches, which cannot be explained by recent gene flow, as they are geographically very far from each other. Therefore, the Gelong and the other Kadai people may have shared common ancestry before they departed to different regions.

DISCUSSION

Recent gene flow or common ancestral population

The Y-chromosome structure of the Gelong sample is similar to both the Hlai and the Kadai (Gelao). Although the Gelong people are now registered under the Han ethnicity temporarily, their genetic structure is not similar to other Han Chinese samples. The dominant haplogroups of Han Chinese, O3a3c^{*}-M134 and O3a3c1-M117^{12,19,20} only make up a proportion of 5% in the Gelong. However, this proportion probably resulted from the recent gene flow from the neighboring Han migrants, similar to the O3a3c proportion in Hlai⁴ and many other minorities in China.²¹ Although the Gelong people have a nomenclature for paternal lineages in Chinese, they do not seem to have a pronounced Han Chinese paternal origin.

Genetic similarity between the two populations may result from various population history events; for example, recent gene flow between neighboring populations, two populations originating from a common ancestral population, or even random genetic drift. Because of the complexity of the marker system of the Y chromosome, it is highly unlikely that random genetic drift can lead to population similarity. This leaves recent gene flow or a common ancestral population as the only possible explanations for the similarity between Gelong and relevant populations. Recent gene flow can change the population affiliation of a population only if it changed most of the genetic proportion of its members. On the other hand, if two populations originated from one common ancestral population, their genetic structures will be similar, and they are most reasonable to be classified into one group.

According to our results, the origin of the Gelong must be either Hlai or Gelao, or both. Some Chinese linguists have suggested that they should be a branch of the Hlai, as they share 42.1% of the vocabulary with the Hlai.^{2,3} Some ethnologists also suggested that the Gelong people register as Hlai in the 1950s; however, the Gelong people themselves could not accept the identity of Hlai, a neighbor they knew was very different from them. In this paper, we do see genetic similarity between the Gelong and the Hlai. Owing to thousands of years of neighboring history, gene flow between the two populations must have left strong effects. Despite this, the Gelong

Table 5 STR haplotypes of the Gelong

| No. | Size | DYS19 | DYS389-1 | DYS389-2 | DYS390 | DYS391 | DYS392 | DYS393 | Diagnostic SNP | Haplogroup |
|-------|------|-------|----------|----------|--------|--------|--------|--------|----------------|------------|
| Hap01 | 1 | 15 | 12 | 28 | 26 | 10 | 14 | 12 | M117 | O3a3c1 |
| Hap02 | 28 | 15 | 13 | 28 | 23 | 10 | 14 | 13 | M119 | O1a* |
| Hap03 | 4 | 15 | 13 | 29 | 23 | 10 | 14 | 13 | M119 | O1a* |
| Hap04 | 1 | 15 | 14 | 29 | 23 | 10 | 14 | 13 | M119 | O1a* |
| Hap05 | 2 | 15 | 14 | 30 | 23 | 10 | 14 | 13 | M119 | O1a* |
| Hap06 | 2 | 16 | 12 | 27 | 23 | 10 | 12 | 13 | M134 | O3a3c* |
| Hap07 | 1 | 16 | 12 | 28 | 22 | 9 | 14 | 13 | M119 | O1a* |
| Hap08 | 4 | 16 | 12 | 28 | 23 | 10 | 14 | 13 | M119 | O1a* |
| Hap09 | 2 | 16 | 12 | 28 | 25 | 11 | 15 | 13 | M119 | O1a* |
| Hap10 | 1 | 16 | 12 | 29 | 23 | 10 | 12 | 12 | M134 | O3a3c* |
| Hap11 | 3 | 16 | 12 | 29 | 23 | 11 | 13 | 13 | M119 | O1a* |
| Hap12 | 1 | 16 | 13 | 28 | 25 | 10 | 13 | 12 | M122 | O3* |
| Hap13 | 1 | 16 | 13 | 29 | 23 | 10 | 14 | 13 | M119 | O1a* |
| Hap14 | 1 | 16 | 13 | 29 | 26 | 11 | 13 | 14 | M95 | O2a* |
| Hap15 | 1 | 16 | 13 | 30 | 25 | 10 | 11 | 13 | M217 | C3 |
| Hap16 | 1 | 16 | 13 | 30 | 24 | 11 | 11 | 14 | M217 | C3 |
| Hap17 | 1 | 16 | 13 | 31 | 23 | 11 | 13 | 12 | M121 | O3a1 |
| Hap18 | 1 | 16 | 14 | 30 | 23 | 11 | 11 | 14 | M217 | C3 |
| Hap19 | 11 | 16 | 14 | 30 | 24 | 10 | 13 | 14 | P31 | O2* |
| Hap20 | 2 | 16 | 14 | 30 | 24 | 11 | 13 | 14 | P31 | O2* |
| Hap21 | 1 | 17 | 13 | 29 | 25 | 10 | 13 | 14 | M95 | O2a* |
| Hap22 | 4 | 17 | 12 | 27 | 25 | 10 | 13 | 12 | M122 | O3* |
| Hap23 | 2 | 17 | 12 | 28 | 25 | 11 | 13 | 12 | M122 | O3* |
| Hap24 | 1 | 17 | 13 | 29 | 23 | 10 | 11 | 15 | M217 | C3 |
| Hap25 | 1 | 17 | 13 | 30 | 24 | 11 | 13 | 14 | M95 | O2a* |

Abbreviations: SNP, single nucleotide polymorphism; STR, short tandem repeat.

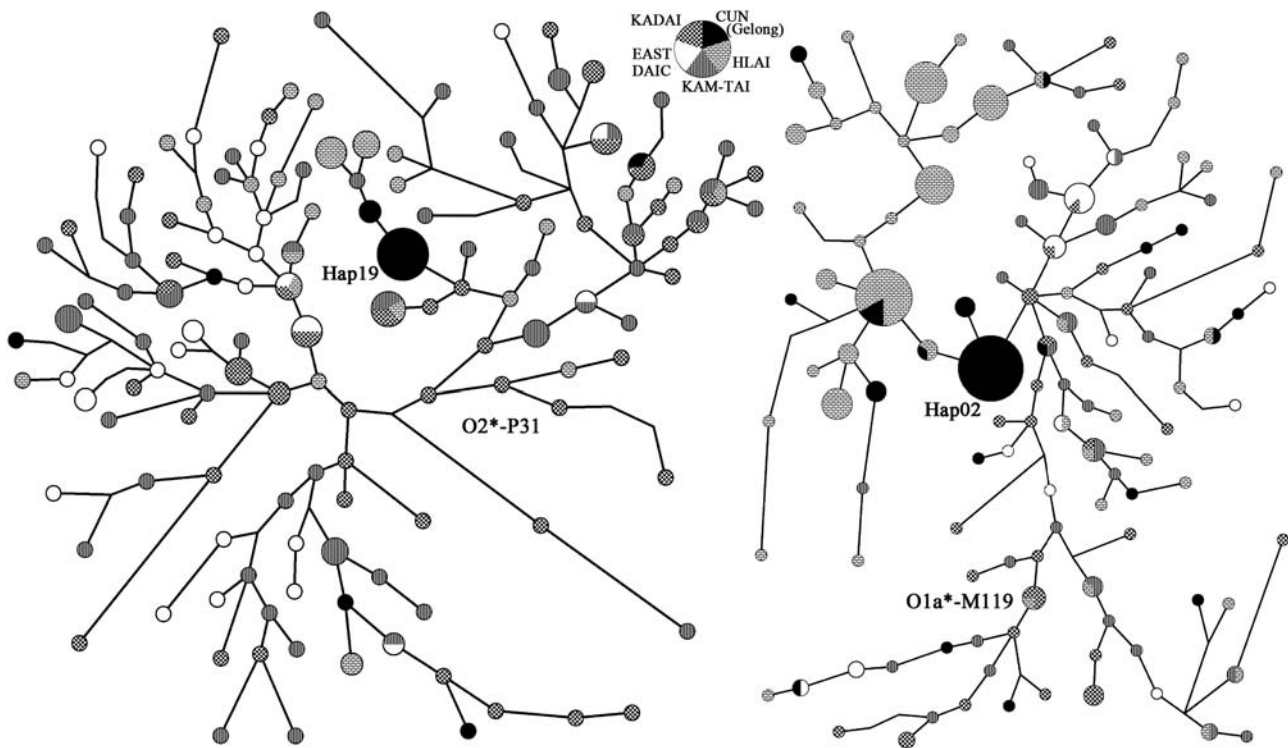


Figure 4 Shortest tree of the STR haplotypes. The length of the lines between nodes is proportional to the mutation steps.

is still much closer to the Gelao than to the Hlai. Therefore, although some Gelong individuals have Hlai ancestry, the genetic similarity cannot be evidence for a Hlai origin of the Gelong population.

On the contrary, the genetic similarity between the Gelong and the Gelao could not have resulted from recent gene flow because of geographic segregation. The Gelao live in an area at least 600 km away

Table 6 Observed STR haplotype similarities between the Gelong and other Daic groups in the networks

| | Shared haplotype | | Neighboring haplotype | |
|-----------|------------------|-----|-----------------------|-----|
| | O1a* | O2* | O1a* | O2* |
| Kadai | 1 | 1 | 31 | 15 |
| Kam-Sui | 1 | 0 | 2 | 1 |
| Tai | 0 | 0 | 0 | 0 |
| East Daic | 1 | 0 | 1 | 2 |
| Hlai | 7 | 0 | 35 | 0 |

Abbreviation: STR, short tandem repeat.

from the Gelong, across the sea, mountains and the regions of the Kam-Tai people, which are not that similar to the Gelong. There has never been any evidence of recent contact between the Gelong and the Gelao. Therefore, the genetic similarity indicates that the Gelong and the Gelao most probably have a common ancestral population. There is also linguistic evidence for the similarity between the Gelao and the Gelong, classifying their languages into the same group of Kadai.^{1,16} Thus, we showed that the Gelong should be a distant relative of the Gelao, and that these two populations may have a common ancestral population.

Culture or genetics

The debate whether culture or genetic background is more important in recognizing a so-called 'ethnicity' has lasted for decades in China. However, for the case of the Gelong, both culture (language¹ and burial custom²²) and genetics support their relationship to the Gelao. They also have similar names, which makes it easier for the Gelong people to accept the title of the Gelao official ethnicity.

On the other hand, there is less cultural similarity between the Gelong and the Hlai, although they both live on the Hainan Island. Archaeological evidence revealed that the Gelong people arrived in their present home about 6000 years ago and left the *Dongfang Xinjie* Shell-heap archaeological site, which is the earliest Neolithic site on Hainan Island.²³ This culture is quite different from the archaeological culture found in the region of the Hlai people. The genetic estimate showed that the Hlai people have been isolated on Hainan Island for about 10 000 to 20 000 years.⁴ Therefore, the Gelong people must have arrived in Hainan much later with a new culture. The Gelong and Gelao people have departed and segregated from each other for at least 6000 years. Thus, cultural and genetic differences should be pronounced. However, we can still see clear resemblance between these two populations. This may be explained by the relative isolation of these two groups. The Gelao people live in the mountains, whereas the Gelong live in a corner of the island. Both of them are less influenced by the exotic populations than the Kam-Tai people are.

Ethnicity identification and benefits

Ethnicity identification in China is quite complicated; it is not only ethnological, but also psychological, and especially political. Ethnologically, there is no doubt for the Gelong people to be identified as the Gelao ethnicity, but the government is more concerned with their self-identification; their acceptance will be more important. The political affairs about the ethnicity identification are also related to the benefits that the government can offer. If the Gelong people can successfully register as the Gelao ethnicity, should they set up an autonomous county in their traditional area according to certain laws in China?

Can they share the seats of the Gelao in the National People's Congress? There is still a long way to go to resolve the ethnicity identification of the Gelong. Regardless, being correctly labeled as a minority rather than Han Chinese will help in keeping their cultures in the current political environment. We hope our study will be a good example of how genetics can help the populations in China in their identification and that the Gelong people will be able to identify themselves.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (No. 30460056), and the Ministry of Education of PRC Science and Technology Research Key Project Fund (No. 204160). The Governments, the Bureaus of Ethnic and Religious Affairs, and the Bureaus of Statistics of both Dongfang City and Changjiang County assisted our field work. We especially thank the many individuals who volunteered to provide samples for this study.

- Gordon, R. G. Jr. *Ethnologue: Languages of the World* 15 edn (SIL International, Dallas, Texas, 2005).
- Fu, C. Z. *Hainan Cun Language* (South China University of Technology Press, Guangzhou, Guangdong, 1996).
- Ouyang, J. Y. *Studies of Cun Language* (Shanghai Far East Press, Shanghai, 1998).
- Li, D., Li, H., Ou, C., Lu, Y., Sun, Y., Yang, B. *et al*. Paternal genetic structure of Hainan aborigines isolated at the entrance to East Asia. *PLoS ONE* **3**, e2168 (2008).
- Burusphat, S. A comparison of general classifiers in Tai-Kadai languages. *Mon-Khmer Studies: J. Southeast Asian Lang. Cultures* **37**, 129–153 (2007).
- Jobling, M. A. & Tyler-Smith, C. Father and sons: the Y chromosome and human evolution. *Trends Genet.* **11**, 449–456 (1995).
- Jobling, M. A. & Tyler-Smith, C. New uses for new haplotypes: the human Y chromosome, disease and selection. *Trends Genet.* **16**, 356–362 (2000).
- Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H. *et al*. Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **26**, 358–361 (2000).
- Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* **12**, 339–348 (2002).
- 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).
- Su, B., Xiao, J., Underhill, P., Deka, R., Zhang, W., Akey, J. *et al*. 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 (1999).
- Ke, Y., Su, B., Xiao, J., Chen, H., Huang, W., Chen, Z. *et al*. Y-chromosome haplotype distribution in Han Chinese populations and modern human origin in East Asians. *Sci. China C Life. Sci.* **44**, 225–232 (2001).
- Li, Y. N., Zuo, L., Wen, B., Ke, Y. H., Huang, W. & Jin, L. Origins and Migrations of Bouyei people in China—insights from Y Chromosome and Mitochondrion. *Acta Genetica Sinica* **29**, 196–200 (2002).
- He, Y., Wen, B., Shan, K. R., Huang, W., Xie, Y., Wu, C. X. *et al*. The distribution of Y-chromosome haplotypes of Shui Ethnic in Sandu, Guizhou. *Hereditas (Beijing)* **25**, 249–252 (2003).
- Li, H., Wen, B., Chen, S. J., Su, B., Pramoongjago, P., Liu, Y. *et al*. Paternal genetic affinity between Western Austronesians and Daic populations. *BMC Evol. Biol.* **8**, 146 (2008).
- Li, J. F. & Zhou, G. Y. *Studies on Outlier Kam-Tai*. (Central University for Nationalities Press, Beijing, 1999).
- Bandelt, H. J., Forster, P. & Rohl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).
- Li, H., Huang, Y., Mustavich, L. F., Zhang, F., Tan, J. Z., Wang, L. E. *et al*. Y chromosomes of prehistoric people along the Yangtze River. *Hum. Genet.* **122**, 383–388 (2007).
- 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).
- 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).
- Shi, H., Dong, Y. L., Wen, B., Xiao, C. J., Underhill, P. A., Shen, P. D. *et al*. Y chromosome evidence of southern origin of the East Asian specific haplogroup O3-M122. *Am. J. Hum. Genet.* **77**, 408–419 (2005).
- Li, H. & Li, D. Analyzing Courtyard tombs found among the Gelong in Hainan. *Commun. Contemp. Anthropol.* **4**, 1–4 (2010).
- Hao, S. D. & Wang, D. X. Retrospection and prospection of Archaeology in Hainan. *Archaeology (China)* **4**, 291–299 (2003).