



ORIGINAL ARTICLE

Forensic characteristics and phylogenetic analyses of one branch of Tai-Kadai language-speaking Hainan Hlai (Ha Hlai) via 23 autosomal STRs included in the HuaxiaTM Platinum System

Wenhui Li^{1,2} | Xianwen Wang³ | Xiehong Wang³ | Fenfen Wang⁴ | Zhengming Du⁴ | Fangshu Fu⁵ | Wenlong Wu⁴ | Shuya Wang⁶ | Ziqing Mu⁷ | Chunwei Chen⁸ | Xiaomin Hu⁹ | Jiuyang Ding¹⁰ | Yunle Meng¹¹ | Pingming Qiu^{11,12}  | Haoliang Fan^{1,2,11,12} 

¹School of Basic Medicine and Life Science, Hainan Medical University, Haikou, China

²Forensic Science Center of Hainan Medical University, Hainan Medical University, Haikou, China

³Criminal Technical Detachment, Haikou City Public Security Bureau, Haikou, China

⁴First Clinical Medical College, Hainan Medical University, Haikou, China

⁵School of Biomedical Information and Engineering, Hainan Medical University, Haikou, China

⁶School of Public Health, Hainan Medical University, Haikou, China

⁷School of Management, Hainan Medical University, Haikou, China

⁸Public Security and Judicial Appraisal Center of Sanya City, Sanya, China

⁹Hainan Zhujian Center for Molecular Cytogenetic Clinical Testing, Haikou, China

¹⁰School of Forensic Medicine, Guizhou Medical University, Guiyang, China

¹¹School of Forensic Medicine, Southern Medical University, Guangzhou, China

¹²Multi-Omics Innovative Research Center of Forensic Identification, Department of Forensic Genetics, School of Forensic Medicine, Southern Medical University, Guangzhou, China

Correspondence

Haoliang Fan and Pingming Qiu, School of Forensic Medicine, Southern

Abstract

Background: Hainan Island, located in the South China Sea and separated from the Leizhou Peninsula by Qiongzhou Strait, is the second largest island after Taiwan in China. With the expansion of Han Chinese and the gradual formation of “South Hlai and North Han”, nowadays, Hainan Hlai is the second largest population after Han Chinese in Hainan Island. Ha Hlai, distributed in southwest and southern Hainan Island, is the dominant branch of Hlai and speaks Ha localism.

Methods: We utilized the HuaxiaTM Platinum PCR Amplification System (including 23 autosomal STRs and 2 sex-linked markers) to obtain the first STR profiling batch of 657 Ha Hlai individuals (497 males and 160 females). In order to explore the genetic relationships between the studied Ha Hlai and other reference populations with different language families, population genetic analyses, including PCA, MDS, STRUCTURE, and phylogenetic analysis, were conducted based upon the raw data and allelic frequencies of the polymorphic autosomal STR markers.

Results: In total, 271 distinct alleles were observed at the 23 STR loci. The number of diverse alleles ranged from 7 at TPOX locus to 23 at FGA locus, and the allelic frequencies varied from 0.0008 to 0.5533. In addition, the CPE and CPD were $1-7.39 \times 10^{-10}$ and $1-3.13 \times 10^{-28}$, respectively. The phylogenetic analyses indicated that Ha Hlai is a Tai-Kadai language-speaking and relatively isolated population which has a close genetic and geographical relationship with Hainan Hlai, and M95 is the dominant haplogroup in Ha Hlai (56.18%).

Conclusion: The 23 autosomal STR genetic markers were highly polymorphic as well as potentially useful for forensic applications in Hainan Ha Hlai population. The phylogenetic analyses demonstrated that small geographic scale gene flows could not be ignored and the shaping of the unique gene pool for each population was the combination effects of geographic, language, and cultural isolations.

Wenhui Li and Xianwen Wang contributed equally to this paper.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2020 The Authors. *Molecular Genetics & Genomic Medicine* published by Wiley Periodicals LLC.

Medical University, Guangzhou 510515, Guangdong, China.
E-mail: fanhaoliang198931@163.com (H. F.) and qiupmfy@126.com (P. Q.).

Funding information

National Undergraduate Innovation and Entrepreneurship Training Program, Grant/Award Number: 201911810008, 201911810023 and X201911810028; National Natural Science Foundation of China, Grant/Award Number: 81971786; Program of Hainan Association for Science and Technology Plans to Youth R&D Innovation, Grant/Award Number: QCXM201705

KEYWORDS

autosomal STR, forensic characteristics, Hainan Ha Hlai, phylogenetic analyses, Tai-Kadai language

1 | INTRODUCTION

Hainan province, which means “shore of pearls”, is the smallest province of China in terms of its land areas (totally 35.4 thousand square kilometers) which is coextensive with Hainan Island and a handful of nearby offshore islands (Xisha, Zhongsha, and Nansha archipelagoes), yet the largest province in its maritime territories (plus 2 million square kilometers), which located in the South China Sea and separated from the Leizhou Peninsula of southern Guangdong province to the north by the shallow and narrow Qiongzhou Strait, while the west coast of Hainan Island is about 320 kilometers east of northern Vietnam, across the Beibu Gulf (Liang, 2013; Zhao, Wang, & Yuan, 2007). In Chinese history, the mass landings of Han Chinese and Lingao people dated from BC 214 in the Qin Dynasty and about BC 500 during the Spring and Autumn period (Yan, 2013; Zhang, 1974). While, with the establishment of Zhuya and Dan'er prefectures on Hainan Island, Hainan Island was officially and formally incorporated into the Chinese empire in Han Dynasty (BC 110) (Weimin Zhou, 2018; Zhong, 2010). Due to constant rebellions staged by the indigenous people, the island remained firmly in the hands of the indigenous Li ancestors and the effective government was not reintroduced until Tang Dynasty (618–907 CE), even though the island remained nominally under Chinese sovereignty during that period (Du & Yip, 1993; Weimin Zhou, 2018; Zhong, 2010). Until 12–17th centuries, the massive immigrants of Han Chinese from Fujian and Guangdong provinces began settling in the northern uplands and plains, displacing and pushing the indigenous Li farther into the central and southern highlands of Hainan Island, which gradually formed the present situations of “South Hlai and North Han” (Weimin Zhou, 2018; Xing, 2004; Yan, 2013; Zhong, 2010).

Li ethnic minority, the aborigines of Hainan Island, had five dominant subgroups, Ha, Gei, Zwn, Moifau, and Jiamao,

which had distinct cultures and languages that cannot be understood by each other although they all belong to the most primordial branches of the Tai-Kadai (also called Daic) linguistic phylum (Burusphat, 2007; Norquest, 2015; Thurgood, 1994; Wang XP & Xing, 2004; Weera, 2017). Nowadays, the Li people (Hlai) are the second largest population after Han Chinese in Hainan Island with a population of 1,514,780 according to the Population of Ethnicity by Region (2018) of Hainan Statistical Year Book 2019. Old Hlai did not have their own writing system and the Li language was inherited from mouth to mouth, until 1957, the writing system of Li minority that was made up of 33 initial consonants, 99 simple or compound vowels, 6 tones, and 5 tone symbols, which was not marked tone symbols for the first tone (Shu Tone with the tone value 53) and had three overlapping letter symbol tones, designed and created by linguists, ethnologists, and anthropologists at the Scientific Conference on the Language and Writing System of Hainan Li Minority conducted at Tongshi city (now Wuzhishan city in the center of Hainan Island) which reached a consensus that the writing system was based on the local localism of the Ha branch of Hlai (Ha Hlai) and the pronunciations of the language used by the Ha Hlai who lived in Baoding village, Baoyou Town, Ledong Li Autonomous County as the fundamental dialect and received pronunciation of Hainan Li, respectively (Wang XP & Xing, 2004).

Ha Hlai, also called Xiao Li or Xiao Hlai, is the most populous and most extensive branch of Hainan Li and mainly distributed in the southwest and southern Hainan Island including the most regions of Ledong Li Autonomous County, Changjiang Li Autonomous County, Baisha Li Autonomous County, and Lingshui Li Autonomous County, and the partial areas of Dongfang and Sanya with about seven hundred thousand populations totally (Du & Yip, 1993; Li, Li, et al., 2008; Sun, Yang, Ou, Chen, et al., 2007; Sun, Yang, Ou, Zhou, et al., 2007; Xing, 2004; Zhong, 2010). Ha Hlai speaks Ha localism

(Xiao localism) which belongs to Tai-Kadai language, a crucial branch of the Sino-Tibetan language family. Hlai, as the aborigines who lived at the south entrances (Hainan Island) to East Asia which might be the border between China and Indo-China Peninsular countries, had been separated since about 20 thousand years ago, after two dominant haplogroups, O1 and O2, entered East Asia (31–36 thousand years ago). However, the origin of Hlai in Hainan Island is still unsolved, one hypothesis was the origin from Luoyue tribes in southeast China (now in Guangxi Zhuang Autonomous Region of mainland China) because gene flow from mainland China played a nonnegligible role in shaping the Hlai maternal pool (Song et al., 2019), and another was for the immigrations from the indigenous peoples of Southeast Asia (Bao Maohong & Wenzhe, 2014; Jin & Su, 2000; Mou, 2007; Osborne, 2012; Yingming, 2010). Even though the genome-wide results of ancient DNA from eighteen Southeast Asian individuals spanning from the Neolithic period through the Iron Age (4100–1700 years ago) supported archaic hominids in Southeast Asia were likely to be descended from southern Chinese immigrants, however, the genetic evidence were insufficient. Our previous studies reported the allele frequencies of 19 autosomal short tandem repeats (STRs) and 27 Y-STRs of the whole Hainan Hlai, mainly exploring the forensic parameters of these STRs and Y-STRs to potential forensic applications (Fan, Wang, Chen, Zhang, et al., 2018; Fan, Wang, Ren, et al., 2019). The tentative explorations and general impressions for Hainan Li based on forensic genetic markers were conducted, whereas, the structures of the segmented populations are still ambiguous due to the limitations to the sample sizes for the branches of Hainan Li. Hence, as the important branch of the Hlai who inhabited at the most significant “transfer” location (the south entrances to East Asia), we collected the first batch of autosomal STR data from the branch, namely Ha Hlai, and conducted a series of population genetic analyses with other populations with diverse language families to depict the STR landscape of Ha Hlai and clear up the population genetic structures.

2 | MATERIALS AND METHODS

2.1 | Ethnic standards and sample preparation

This study and the procedures were approved by the Institutional Review Boards of Hainan Medical University and the Medical Ethics Committee of Hainan Medical University (No. HYLL-2020-012). The humane and ethical research principles recommended by Hainan Medical University were followed in this study.

Blood samples were collected using the FTA cards with the written informed consent from 657 unrelated healthy Ha Hlai individuals (497 males and 160 females) from the districts of Ha localism. All self-declared Ha Hlai indigenous subjects were needed to be no intermarriage with people of other populations and other subgroups of Hlai, used the Ha localism as their mother tongue and resided in the regions of Ha dialect at least three generations.

2.2 | DNA extractions and quantification

Genomic DNA was extracted using the QIAamp DNA Blood Mini Kit (Qiagen, Germany) according to the manufacturer's protocol. The quantity of the DNA template was determined using Qubit 3.0 Fluorometer (Invitrogen™, Life Technologies™, USA) according to the manufacturer's instructions. Based on quantitative results, samples were normalized to 1.0 ng/μl and stored at –20°C until amplification.

2.3 | Multiplex amplification and STR genotyping

The Huaxia™ Platinum PCR Amplification System (Applied Biosystems®, Life Technologies™, USA) which consisted of 23 autosomal STRs, two sex-linked inherited Y-InDel (rs2032678), and Amelogenin for STR genotyping and co-amplified in one multiplex PCR on the Veriti® 96 Well Thermal Cycler System (Applied Biosystems®, Life Technologies™, USA) in accordance with the manufacturer's instructions.

The amplified products were separated by capillary electrophoresis on an Applied Biosystems 3500 Genetic Analyzer (Applied Biosystems®, Life Technologies™, USA). Allele designation was conducted using GeneMapper® ID-X software (Applied Biosystems®, Life Technologies™, USA) by comparing the fragment size with the corresponding allelic ladder provided by the Huaxia™ Platinum PCR Amplification System.

2.4 | Quality control

We strictly followed the recommendations of the Chinese National Standards and Scientific Working Group on DNA Analysis Methods (SWGDM) (“Scientific Working Group on DNA Analysis (SWGDM),” 2010) and employed the typical control DNA of 9947A as positive, and the ddH₂O as negative in each batch of PCR amplification and electrophoresis. In addition, the laboratory has been accredited in accordance with ISO/IEC 17025:2005 and the China National Accreditation Service for Conformity Assessment (CNAS) (Registration No. CNAS L10088).

2.5 | Statistical and phylogenetic analysis

Allelic frequencies and forensic parameters of 23 autosomal STR loci were calculated by SAS[®] 9.4 software and the online tool of forensic statistics analysis toolbox (FORSTAT) (Ristow & D'Amato, 2017). The forensic genetic parameters contained genotype count (N), heterozygotes (H_e), random match probability (PM), polymorphism information content (PIC), power of discrimination (PD), power of exclusion (PE), genetic diversity (GD), and typical paternity index (TPI). Subsequently, the probability values of Hardy–Weinberg equilibrium (HWE) were assessed using the Arlequin v3.5 software (Excoffier & Lischer, 2010).

The principal component analysis (PCA) and the Multi-Dimensional Scaling (MDS) were performed with SPSS 22.0 (Hansen, 2000) based on the allelic frequencies of autosomal STR loci, respectively, which were used to explore the extent of correlated genetic relationships. To further investigate the genetic structure of populations, a Bayesian model-based clustering approach was adopted with STRUCTURE 2.3.4 (Pritchard, Stephens, & Donnelly, 2000). Population structures were inferred by setting the value of the clusters (K) from 2 to 8 with 10 runs and 100,000 iterations. STRUCTURE HARVESTER (Earl & vonHoldt, 2012) and DISTRUCT (Rosenberg, 2003) were performed to identify the optimum K value and show the results of population stratification, respectively. The pairwise fixation index F (F_{st}) (Nei, 1977) and corresponding P values between the studied population and the other 72 reference populations were estimated using the Arlequin v3.5 software (Excoffier & Lischer, 2010). Additionally, phylogenetic relationships among different populations were depicted in the Molecular Evolutionary Genetics Analysis 7.0 (MEGA 7.0) software (Kumar, Stecher, & Tamura, 2016) using the neighbor-joining (N-J) phylogenetic tree (Saitou & Nei, 1987) based upon the F_{st} genetic distance matrix, further, the phylogenetic relationships were visualized by the Interactive Tree of Life v4 (iTOL) (Letunic & Bork, 2019). Except the result of STRUCTURE, other graph results in this study were visualized using the R programming language.

3 | RESULTS AND DISCUSSION

In this study, the STR profiling of 657 Ha Hlai individuals (497 males and 160 females) was genotyping utilizing the Huaxia[™] Platinum PCR Amplification System, all samples were obtained the integrated genotypes with no triallelic or null alleles. Furthermore, we evaluated the forensic characteristics and efficiencies of genetic markers included in the Huaxia[™] Platinum PCR Amplification System for forensic applications in Ha Hlai population, to make genetic differences between Ha Hlai and other populations with diverse

language-speaking populations, the bioinformatics analyses were conducted to clarify the genetic structures from different scales.

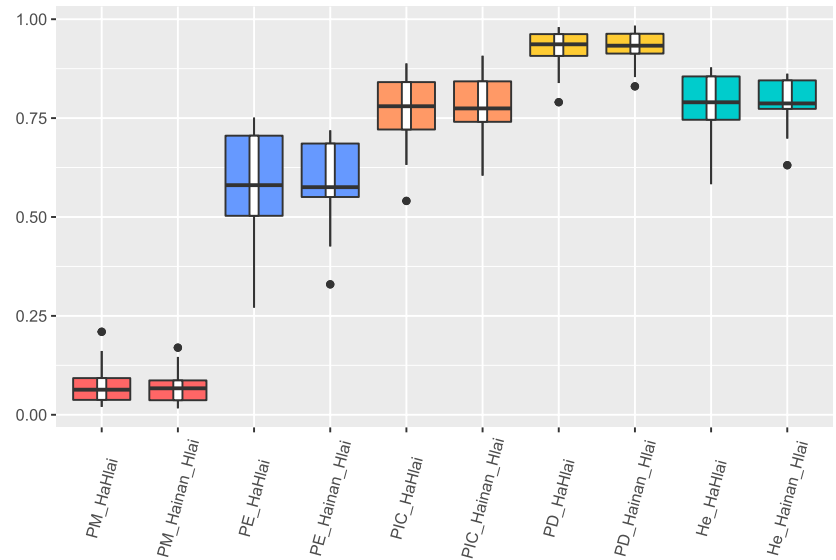
3.1 | Forensic characteristics of the 23 autosomal STRs

The first STR profiling batch of the Ha Hlai was reported in our present study, a total of 657 Ha Hlai (497 men and 160 women) were successfully genotyped for 23 autosomal STRs and two sex-linked inherited markers (rs2032678 and Amelogenin). As shown in Tables S1 and S2, totally, 271 distinct alleles were observed at the 23 STR loci. The number of diverse alleles ranged from 7 at TPOX locus to 23 at FGA locus, and the allelic frequencies varied from 0.0008 to 0.5533. The forensic related parameters of 23 autosomal STRs for Ha Hlai are presented in Table S2, the H_e , PM, PIC, PD, PE, GD, and TPI, spanned from 0.1218 (Penta E) to 0.4170 (TPOX), 0.0200 (Penta E) to 0.2098 (TPOX), 0.5406 (TPOX) to 0.8882 (Penta E), 0.7902 (TPOX) to 0.9800 (Penta E), 0.2709 (TPOX) to 0.7512 (Penta E), 0.6005 (TPOX) to 0.8970 (Penta E), and 1.1989 (TPOX) to 4.1063 (Penta E), respectively. In addition, the CPE and CPD were $1-7.39 \times 10^{-10}$ and $1-3.13 \times 10^{-28}$, respectively.

Furthermore, to evaluate the performances of forensic-associated autosomal STRs between Ha Hlai and Hainan Hlai, we made comparisons for both which are presented in Table S3. Except for D1S1656, D2S441, D10S1248, and D22S1045, the remaining 19 autosomal STRs, which were also genotyped in our previous study on Hainan Hlai ($n = 653$) (Fan, Wang, Ren, et al., 2019), had 227 diverse alleles for Ha Hlai, while for the same loci, Hainan Hlai had 270 different alleles. At D18S51 and FGA loci, Ha Hlai had 17 and 23 alleles, more than 16 and 21 in Hainan Hlai, respectively, while the numbers of other loci in Ha Hlai were less than or equal to Hainan Hlai, especially, at CSF1PO, the allele diversity had onefold in Hainan Hlai (18 vs. 9 in Ha Hlai), Hainan Hlai had obvious polymorphisms at D2S1338 and Penta E loci (21 and 32) when compared with the same loci in Ha Hlai (13 and 22, respectively). The comparisons of forensic statistical parameters between Ha Hlai and Hainan Hlai are visualized in Figure 1. Even though for the numbers of most loci, Hainan Hlai had higher diversity, there was no apparent difference for forensic-associated parameters and overall system effectiveness compared with Ha Hlai.

As a result, the analyses of all these allelic frequencies and forensic statistical parameters demonstrated that the 23 autosomal STR genetic markers included in Huaxia[™] Platinum PCR Amplification System (Applied Biosystems[®], Life Technologies[™], USA) were highly polymorphic as well as potentially useful for forensic applications in Hainan Ha Hlai population.

FIGURE 1 Forensic statistical parameters and performances of 19 autosomal STRs between Ha Hlai and Hainan Hlai



3.2 | PCA

Ha Hlai, a branch of Hainan Hlai, speaks Ha Hlai localism which is a branch of Tai-Kadai language generally belongs to the Sino-Tibetan language family. To illustrate the genetic differences of the Sino-Tibetan language-speaking populations, especially for the Ha Hlai, we established a database which included 49 populations with Sino-Tibetan language (30 Chinese/Sinitic, 8 Tai-Kadai, 6 Tibeto-Burman, and 5 Hmong-Mien language-speaking population, totally 60,347 individuals) to perform PCA based on allelic frequencies of the same 19 autosomal STRs (Bai et al., 2017; Chan, Chiu, Tsui, Wong, & Fung, 2005; Chen Chunbao, Tian Xin, & Hanhua, 2017; Chen et al., 2017; Chen, Zhu, Shiming, You, & Ma, 2010; Deng et al., 2007; Fan, Wang, Chen, et al., 2019; Fan, Wang, Ren, et al., 2019; Ferdous et al., 2010; Guo, 2017; Guo, Li, Wei, Ye, & Chen, 2017; He et al., 2017; Hongdan et al., 2017; Huihui Lian, Lin, & Li, 2015; Jingzhou Wang, Zhai, Wang, Zhang, &

Wang, 2015; Kraaijenbrink, van Driem, Tshering of Gaselô, & de Knijff, 2007; Le Wang, Zhang, Bai, & Ye, 2013; Li, Zheng, & Jun, 2017; Li et al., 2015, 2018; Liao et al., 2019; Lili Zhang et al., 2017; Liu, Liu, & Wang, 2006a, 2006b; Liu, Chen, Huang, et al., 2017; Liu, Chen, Mei, et al., 2017; Meng Pan et al., 2012; Qiuling Liu, Huiling, & Chen, 2003; Shen, Kang, Dong, Guo, & Wang, 2015; Sun et al., 2017; Sun, Zhang, Wu, Shen, & Wu, 2015; Wenming Han, 2016; Xiao, Zhang, Wei, Pan, & Huang, 2016; Xie et al., 2014; Xiuzi & Changchun, 2017; Xu, Feng, & Yao, 2017; Xu, Xu, Wang, & Yao, 2017; Yang et al., 2017, 2018; Yao et al., 2016; Zhang, 2015; Zhang, Zhao, Guo, Liu, & Wang, 2015; Zhang et al., 2013, 2018; Zhang, Du, et al., 2017; Zhang, Hu, Du, Nie, et al., 2017; Zhang, Hu, Du, Zheng, et al., 2017; Zou et al., 2017). PCA, an effective dimensionality reduction method, could visualize the essential patterns of genetic relationships, and recognize and depict the dominant patterns within a multivariate database, it could indicate that the populations with closer geographical distances have more

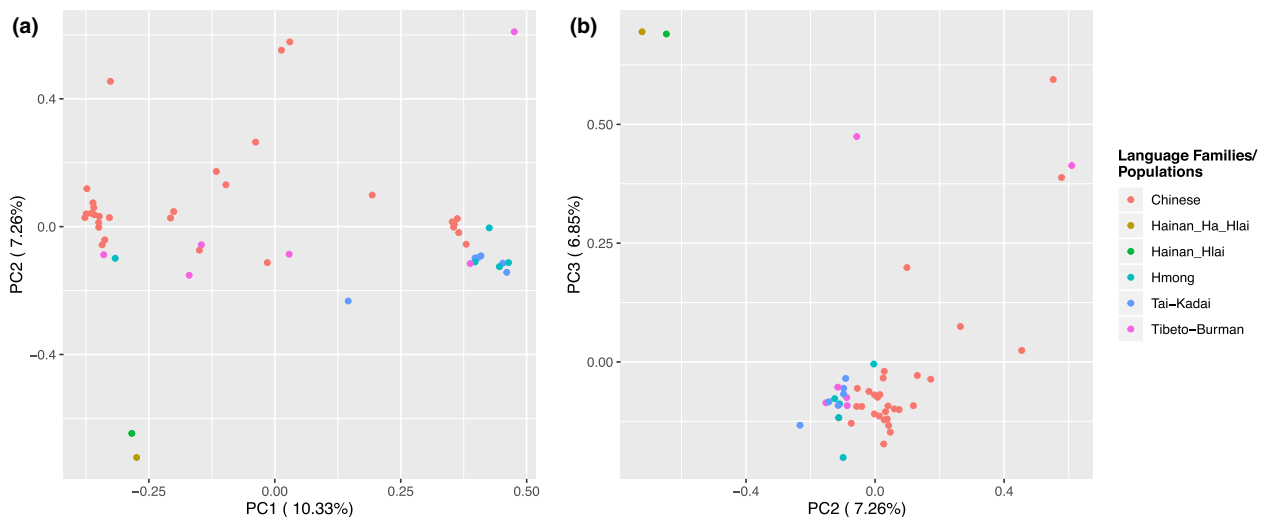


FIGURE 2 PCA based on allelic frequencies of 49 Sino-Tibetan language-speaking populations. (a. PC1 and PC2; b. PC2 and PC3)

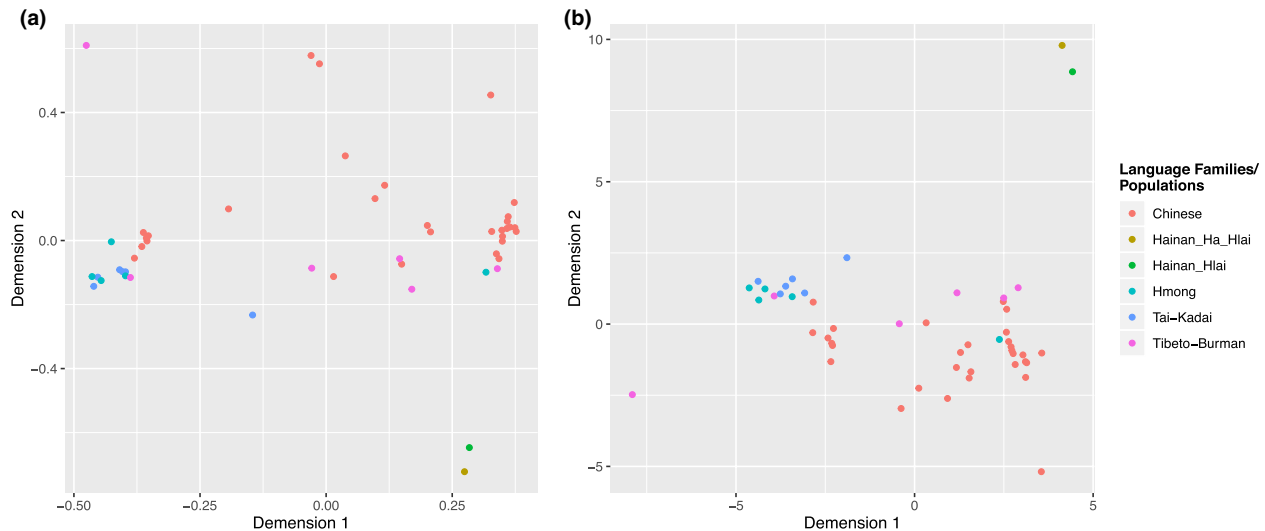


FIGURE 3 MDS based on allelic frequencies of 49 Sino-Tibetan language-speaking populations. (a. MDS based on Euclidean distance; b. MDS based on Manhattan distance)

intimate relationships in forensic genetics. As shown in Figure 2, the proportions of first, second, and third components (PC1, PC2, and PC3) were 10.33%, 7.26%, and 6.85% of the total variance observed within these Sino-Tibetan language-speaking populations, in total, the cumulative variance contribution of PC1, PC2, and PC3 was 24.44%. Intuitively speaking, Ha Hlai clustered with Hainan Hlai located at the bottom left corner in Figure 2a, while Ha Hlai and Hainan Hlai got together in the top left corner in Figure 2b. From the PCA plots, the Tai-Kadai and Hmong-Mien (Hmong) language-speaking populations clustered together and had distances with other populations, and the Tibeto-Burman language-speaking populations relatively clustered with Chinese/Sinitic language-speaking populations; however, the distributions of the Chinese/Sinitic language-speaking populations were relatively dispersive and scattered. Even though, Hlai language is a branch of Tai-Kadai language from the perspective of glossologies and linguistics (Ostapirat, 2005; Thurgood, 1994; Wang XP & Xing, 2004), the cluster of Ha Hlai and Hainan Hlai did not get together with Tai-Kadai-Hmong cluster, which indicated that Hainan Hlai is an isolated population because of the geographic and culture isolations.

3.3 | MDS

While in order to make further confirmation about the relationships among the Sino-Tibetan language-speaking populations, MDS, which could accelerate the speed of algorithm execution and reduce the complexity of complicated and multidimensional data simultaneously, were performed to depict the forensic genetic landscape of the Sino-Tibetan language-speaking populations on the basis of Euclidean and Manhattan distances, respectively. No matter the Euclidean MDS (Figure 3a) or the Manhattan MDS (Figure 3b),

the genetic relationships between Ha Hlai and other Sino-Tibetan populations were in accordance with the results of PCA, Ha Hlai was clustered with Hainan Hlai and gathered in the corners of MDS diagrams which separated with other populations relatively.

As a whole, the Sino-Tibetan language-speaking populations had relatively close distances, the Chinese/Sinitic and Tibeto-Burman language-speaking populations gathered together, while the Tai-Kadai and the Hmong language-speaking populations clustered, respectively. In China, the Tai-Kadai and Hmong language-speaking populations lived in southern China with the Chinese/Sinitic and Tibeto-Burman language-speaking populations at least for three thousand years, the genetic makeup of natives in southern China which included the populations who spoke Tai-Kadai, Hmong, and Austroasiatic had influenced by three massive southward waves of Han Chinese (Western Jin Dynasty, AD 265-316; Tang Dynasty, AD 618-907; Southern Song Dynasty, AD1127-1279), the expansions of Han Chinese contributed to the gene flows with southern natives (Fei, 1999; Ge, Wu, & Chao, 1997; Wen et al., 2004). Therefore, the Tai-Kadai-Hmong populations clustered together and had relatively close genetic distances with Han Chinese (Chinese/Sinitic language-speaking populations). However, as the geographical isolations by the Qiongzhou Strait, limited gene flows with other Sino-Tibetan populations depicted the characteristic genetic landscape of Hainan Hlai, especially for Ha Hlai with relatively far genetic and geographic distances.

3.4 | STRUCTURE

STRUCTURE analysis is commonly recognized to be capable of inferring population structure and assigning individuals to populations using multi-locus genotypic data (Pritchard

et al., 2000). STRUCTURE clustering analysis was performed to reflect the memberships of biogeographic ancestry components for Ha Hlai and other Sino-Tibetan (Tai-Kadai, Tibeto-Burman, and Chinese), Indo-European, Altaic, and Semito-Hamitic language-speaking populations (Fan, Wang, Chen, et al., 2019; Fan, Wang, Ren, et al., 2019; He, Wang, Liu, Hou, & Wang, 2018; Martinez-Cortes et al., 2019; Munoz et al., 2012; Tillmar, Backstrom, & Montelius, 2009; Wang et al., 2016, 2018) with the number of hypothetic populations (K) defined at 2-8 in the present study. As shown in Figure 4, population names and corresponding language families were marked on the bottom and the top of the graph and the width of each bar represented the sample size for each population. With the STRUCTURE Harvester, we observed a platform of the estimated posterior probability at $K = 4$. Thus, $K = 4$ was the most appropriate value for STRUCTURE analysis based on the guidelines in the STRUCTURE manual.

When K at 4, nine Sino-Tibetan language-speaking populations, three Indo-European populations (two Mexico populations and one Argentina population), two Altaic language-speaking Uyghur (Xinjiang and Xinjiang Kumul), and Semito-Hamitic language-speaking Somalia could be

distinguished. However, the Chinese Indo-European language-speaking Hui and Altaic language-speaking Uyghurs had the same components compared with other Sino-Tibetan language-speaking populations. In the inner of Sino-Tibetan language-speaking populations, Tai-Kadai, Tibeto-Burman, and Chinese could distinguish from each other on the basis of different components of red and green. In addition, Ha Hlai had the same compositions with Hainan Li and Lingao which all belong to the Tai-Kadai language family (YI, 2015; Min, 1997; Wang XP & Xing, 2004; Weera, 1998). In conclusion, the relatively small geographic scale gene flows, especially for the multi-ethnic mixed areas, could not be ignored, and the shaping of the unique gene pool for each population the co-effects of geographic, language and cultural isolations, which seemed that as if the effect of geographic isolation was a little stronger than that of language and cultural isolations.

3.5 | F_{st} and phylogenetic analysis

In order to estimate the genetic affiliations, we constructed a database including 73 populations (71,388 individuals in

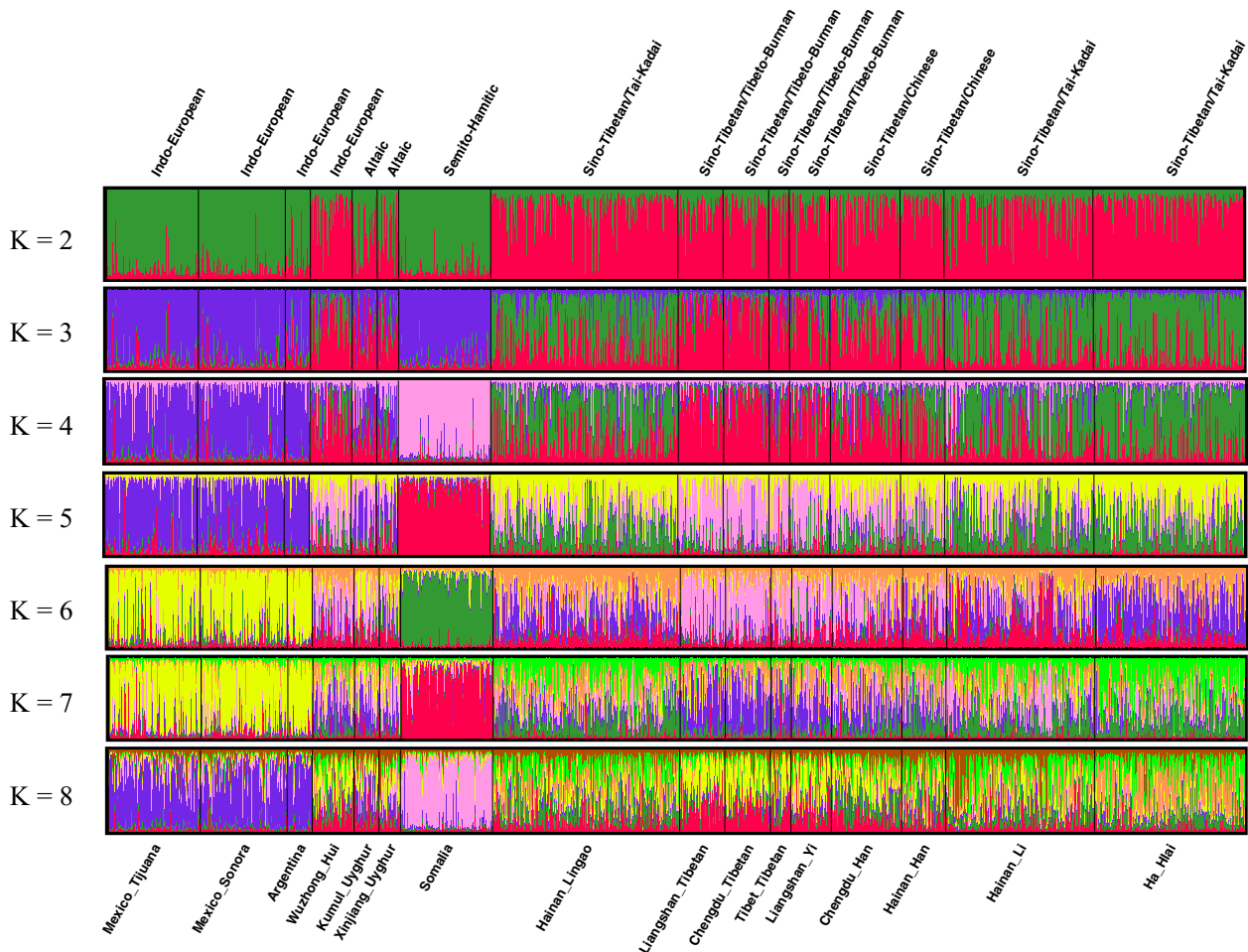


FIGURE 4 STRUCTURE Clustering analysis conducted at population level based upon genotyping data of the 19 autosomal STRs for 16 populations with four different language families

Tree scale: 0.001

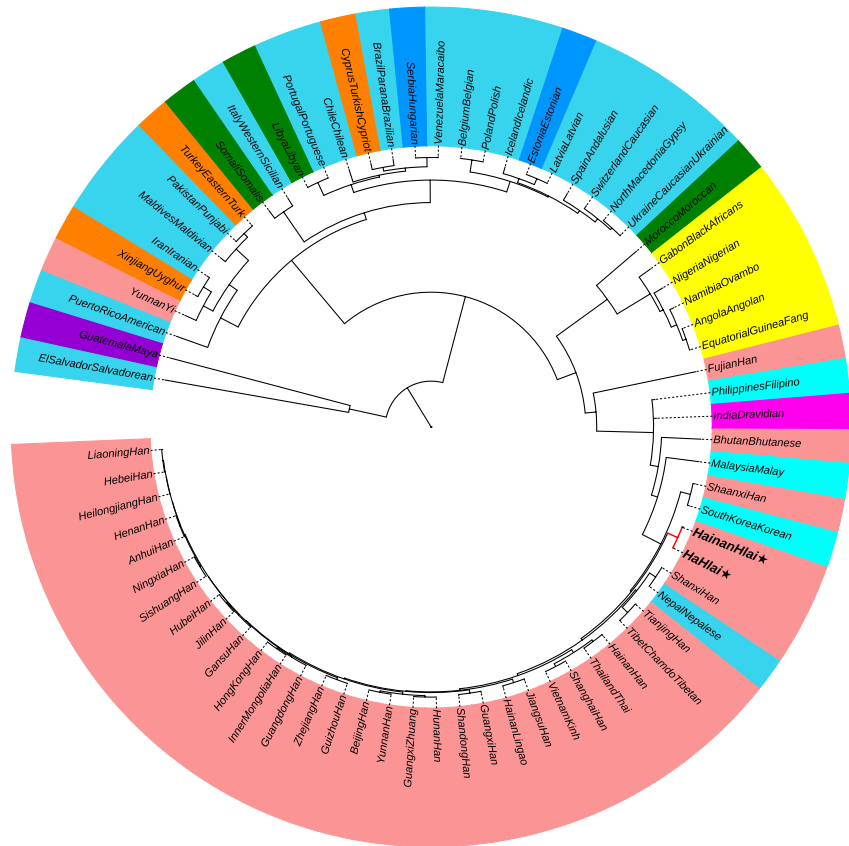


FIGURE 5 A phylogenetic N-J tree constructed by F_{st} values of Ha Hlai and other 72 populations with nine diverse language families

total) (Ali et al., 2012; Alves et al., 2014; Andreassen, Pereira, Dupuy, & Mevaag, 2010; Bernal et al., 2006; Bindu, Trivedi, & Kashyap, 2007; Calzada et al., 2005; Cardoso, Sevillano, Illescas, & de Pancorbo, 2016; Coudray, Calderon, et al., 2007; Coudray, Guitard, el-Chennawi, Larrouy, & Dugoujon, 2007; Coudray, Guitard, Keyser-Tracqui, et al., 2007; Decorte et al., 2004; Gehrig et al., 2014; Gurkan, Demirdov, Yamaci, & Sevay, 2015; Havas, Jeran, Efremovska, Dordevic, & Rudan, 2007; Hedjazi, Nikbakht, Hosseini, Hoseinzadeh, & Hosseini, 2013; Jemeljanova et al., 2015; Jin et al., 2017; Jingzhou Wang et al., 2015; Khodjet-el-Khil et al., 2012; Kraaijenbrink, van Driem, Opgenort, Tuladhar, & de Knijff, 2007; Kraaijenbrink, van Driem, et al., 2007; Le Wang et al., 2013; Li et al., 2015, 2017, 2018, 2020; Liao et al., 2019; Lili Zhang et al., 2017; Liu, Liu, & Wang, 2006a, 2006b; Liu, Chen, Huang, et al., 2017; Liu, Chen, Mei, et al., 2017; Lopes et al., 2009; Maruyama, Minaguchi, Takezaki, & Nambiar, 2008; Melo et al., 2010; Meng Pan et al., 2012; Munoz, Pinto de Erazo, Baeza, Arroyo-Pardo, & Lopez-Parra, 2015; Muro et al., 2008; Okolie et al., 2018; Petric, Draskovic, Zgonjanin-Bosic, Budakov, & Veselinovic, 2012; Qiuling Liu et al., 2003; Rodriguez, Salvador, Calacal, Laude, & De Ungria, 2015; Sadam et al., 2015; Serga et al., 2017; Shan et al., 2016; Shen et al., 2015; Shotivaranon, Chirachariyavej, Leetrakool, & Rerkamnuaychoke, 2009; Soltyszewski et al., 2014; Steinlechner, Schmidt, Kraft, Utermann, & Parson,

2002; Sun et al., 2015, 2017; Tillmar et al., 2009; Tokdemir, Tuncez, & Vicdanli, 2016; Toscanini et al., 2015; Tran et al., 2019; Ventura Spagnolo et al., 2017; Wenming Han, 2016; Xiao et al., 2016; Xie et al., 2014; Xiuzi & Changchun, 2017; Xu, Feng, et al., 2017; Xu, Xu, et al., 2017; Yang et al., 2017, 2018; Yao et al., 2016; Yoo et al., 2011; Zhang, 2015; Zhang et al., 2013, 2018; Zhang, Du, et al., 2017; Zhang, Hu, Du, Zheng, et al., 2017; Zhang, Liao, et al., 2015; Zhu et al., 2007; Zou et al., 2017; Zuniga et al., 2006) with nine language families consisting of Sino-Tibetan (38), Indo-European (19), Niger-Congo (5), Semito-Hamitic (3), Altaic (3), Uralic (2), Austronesian (3), Dravidian (1), and Native American (1). The detailed pairwise F_{st} and corresponding P values between Ha Hlai and the other 72 populations are presented in Table S4. As shown in Figure 5, two main branches could be clearly identified in the phylogenetic tree. The upper branch was composed of three clusters: most Indo-European language-speaking populations, except for Nepalese, clustered in the upper clade with three Altaic, two Uralic, two Semito-Hamitic and one Native American language-speaking populations, while the most Sino-Tibetan (except for Yunnan Yi) and Niger-Congo language-speaking populations got together in the bottom clade clustering with three Austronesian and one Dravidian populations, and the Semito-Hamitic language-speaking Moroccan clustered with the Niger-Congo cluster. In addition, in the Sino-Tibetan cluster, Ha Hlai

clustered with Hainan Hlai ($F_{st} = 0.0011$) and got together other Sino-Tibetan language-speaking populations, which indicated that Ha Hlai is a Sino-Tibetan population and has a relatively close relationship with Hainan Hlai. The clustering results of the dendrogram focusing on the Hainan Hlai and Ha Hlai populations were roughly congruent with the former analyses such as the results of PCA and MDS analyses.

3.6 | Predictions of Y haplogroups

As we all know, no effective algorithm has a strong and accurate ability to distinguish the same or very similar haplotypes and assign them into diverse haplogroups and the convergence of Y chromosome STR haplotypes among different haplogroups has compromised the accuracy of haplogroup prediction. Hence, we applied our in-house database which included 37,754 pieces of Y SNP/STR data and 109,142

Y-STR totally mainly from East and Southeast Asia, to make more precise predictions (Wang et al., 2015).

In the present study, the 27 Y-STR profiles applied to make the Y-haplogroup estimation were collected from 185 Ha Hlai male individuals which came from our previous study that included 422 Hainan Hlai males (185/422) in total (Fan, Wang, Chen, Zhang, et al., 2018). As a result, 178 out of the 185 genotyped Y-STRs (96.22%) were observed the Y-haplogroup and four haplogroups observed in our Ha Hlai samples, which belonged to major clades O1 and O2. The predominant haplogroups were O1b1a1a-M95 (56.18%), O1a-M119 (29.21%), O1b1a1a1a1-P203 (10.67%), and O2-M122 (3.93%), which were determined according to ISOGG 2019. Our results were in accordance with previous studies (Li, Li, et al., 2008; Sun, Yang, Ou, Chen, et al., 2007), O1b1a1a-M95 was the predominant haplogroup in the population of Ha Hlai, which were in accordance with previous studies that the M95 proportions ranged from 46% to 91.94% in different branches of Hainan Hlai.

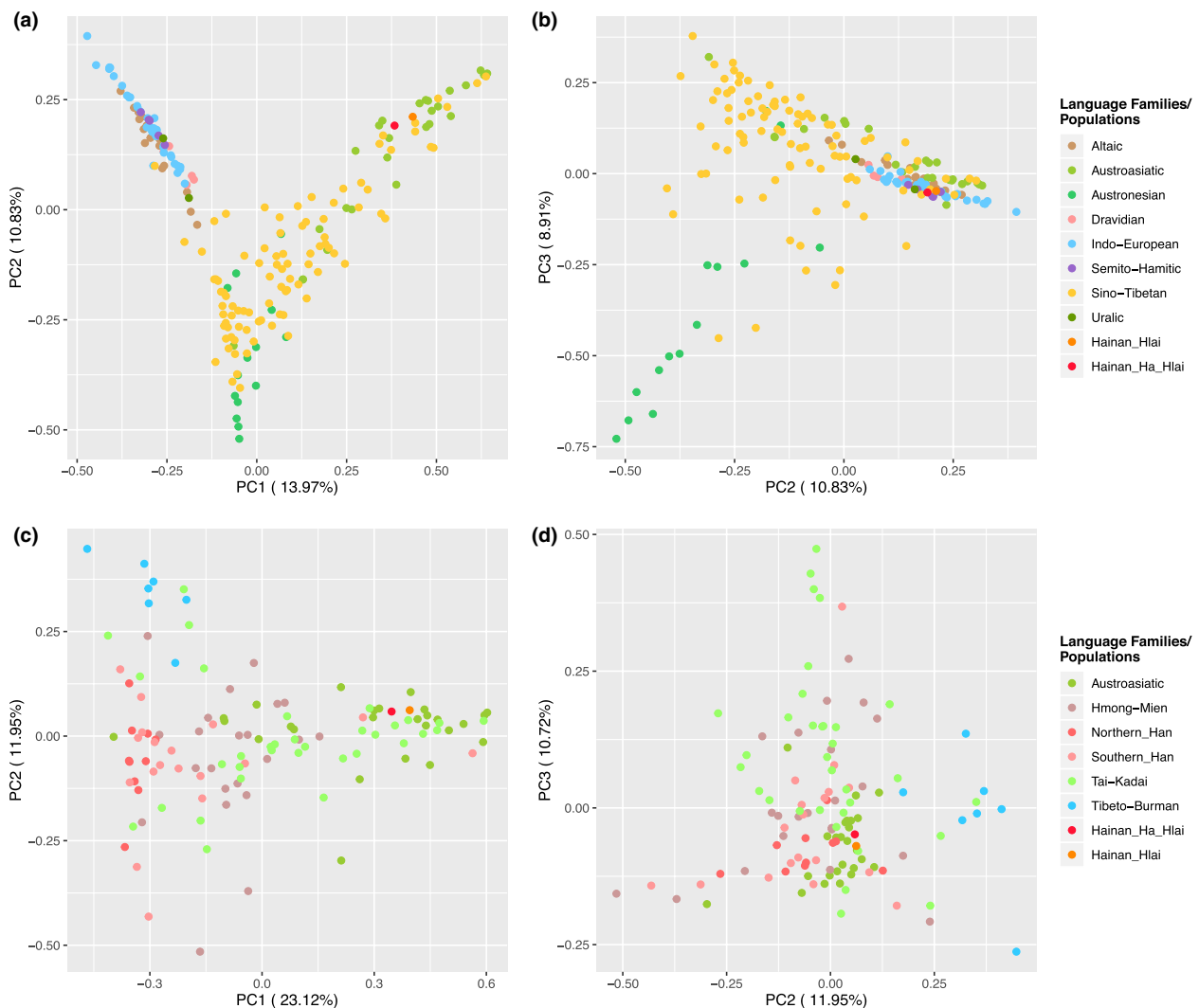


FIGURE 6 Principal component analyses based on the frequencies of Y haplogroups in dataset 1 (193 populations) and dataset 2 (118 populations). (Dataset 1: a. PC1 and PC2, b. PC2 and PC3; Dataset 2: c. PC1 and PC2, d. PC 2 and PC3)

To discern the detailed relationship between the Ha Hlai and other language-speaking populations with different language families, we performed a PCA which included 193 populations (Cadenas, Zhivotovsky, Cavalli-Sforza, Underhill, & Herrera, 2008; Cai et al., 2011; Chennakrishnaiah et al., 2013; Deng et al., 2013; Fan et al., 2020; Fan, Wang, Chen, Long, et al., 2018; Fan, Zhang, et al., 2018; Gan et al., 2008; Gayden et al., 2007; Khurana et al., 2014; Kwon, Lee, Lee, Yang, & Shin, 2015; Li, Wen, et al., 2008; Regueiro, Cadenas, Gayden, Underhill, & Herrera, 2006; Rowold et al., 2016, 2019; Wells et al., 2001; Wen et al., 2004) from all over the world with eight language families, 11,197 individuals in total (dataset 1). The PC1, PC2, and PC3 of the total variances accounted for 13.97%, 10.83%, and 8.91%, respectively. From the dendrogram of Figure 6a,b, the populations with different language families were separated relatively, and Hainan Hlai and Ha Hlai clustered together and had close relationships with Sino-Tibetan and Austroasiatic cluster. In addition, we collected all Sino-Tibetan and Austroasiatic populations (dataset 2) to make further genetic differences, as shown in Figure 6c,d, Ha Hlai clustered with Hainan Hlai and located in Tai-Kadai-Austroasiatic cluster. The M95 haplogroup was shown to be prevalent in Austroasiatic language-speaking populations in Southeast Asia (74-87%) (Arunkumar et al., 2015; Majumder, 2010; Tattersall, 2009; Zhang, Liao, et al., 2015). All Austroasiatic populations in our datasets belonged to the Mon-Khmer group which were spread over Southeast Asia (Vietnam and Cambodia), Yunnan, and Guangxi provinces (Sciences, 2012; Yingming, 2010). However, Hlai localism is a sort of Tai-Kadai language from the perspective of linguistics, and the Ha Hlai inhabited in Hainan Island has close distances with our in-house Mon-Khmer populations from the geographical scale. Therefore, Ha Hlai which has a close genetic and geographical relationship with Hainan Hlai is a Tai-Kadai language-speaking population, and the M95 which is ubiquitous in Mon-Khmer populations is the dominant haplogroup in Ha Hlai (56.18%).

4 | CONCLUSIONS

In the present study, the first batch of 23 autosomal STR profiles of Ha Hlai, one important branch of Hainan Li (Hlai), were obtained and reported by genotyping 657 Ha Hlai individuals (497 males and 160 females) utilizing the Huaxia™ Platinum PCR Amplification System. In total, two hundred and seventy-one distinct alleles were observed at the 23 STR loci with the allelic frequencies from 0.0008 to 0.5533. For the numbers of most loci, Hainan Hlai had higher diversity, however, there was no apparent difference for forensic-associated parameters and overall system effectiveness compared with Ha Hlai. The CPE and CPD were $1-7.39 \times 10^{-10}$ and $1-3.13 \times 10^{-28}$, respectively. The analyses and comparisons of

all these allelic frequencies and forensic statistical parameters demonstrated that the 23 autosomal STR genetic markers included in Huaxia™ Platinum PCR Amplification System were highly polymorphic as well as potentially useful for forensic applications in Hainan Ha Hlai population. In addition, we applied the polymorphic genetic markers to explore the genetic relationships between the studied Ha Hlai and other reference populations with different language families. The phylogenetic analyses indicated that Ha Hlai is a Tai-Kadai language-speaking and relatively isolated population which has a relatively close relationship with Hainan Hlai, and the haplogroup M95 is the dominant haplogroup in Ha Hlai (56.18%), which demonstrated that small geographic scale gene flows, especially for the multi-ethnic mixed areas, could not be ignored and the shaping of the unique gene pool for each population the co-effects of geographic, language and cultural isolations.

ACKNOWLEDGMENTS

We would like to thank all the volunteers who contributed samples for this study. This study was supported by the Program of Hainan Association for Science and Technology Plans to Youth R&D Innovation (QCXM201705), and the National Undergraduate Innovation and Entrepreneurship Training Program (No. 201911810008, No. 201911810023 and No. X201911810028), and the National Natural Science Foundation of China (NSFC, No. 81971786).

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

AUTHOR CONTRIBUTIONS

Conceptualization, Haoliang Fan and Pingming Qiu; Resources, Wenhui Li, Xianwen Wang, and Xiehong Wang; Software, Haoliang Fan; Investigation, Wenhui Li, Xianwen Wang, Xiehong Wang, Fenfen Wang, Zhengming Du, Shuya Wang, and Ziqing Mu; Validation, Chunwei Chen and Xiaomin Hu; Formal Analysis, Haoliang Fan, Fangshu Fu, Wenlong Wu, Jiuyang Ding, and Yunle Meng; Data Curation, Wenhui Li and Xianwen Wang; Writing – Original Draft Preparation, Haoliang Fan; Writing – Review & Editing, Haoliang Fan; Visualization, Haoliang Fan; Supervision, Pingming Qiu; Project Administration, Wenhui Li and Xianwen Wang; Funding Acquisition, Haoliang Fan, Fenfen Wang, Zhengming Du, Shuya Wang, and Pingming Qiu. All authors reviewed the manuscript.

ORCID

Pingming Qiu  <https://orcid.org/0000-0002-5579-1124>
Haoliang Fan  <https://orcid.org/0000-0002-3214-0177>

REFERENCES

- Ali, I., Rasheed, A., Saleem, M. M., Saeed, A., Rasheed, H., Nasih, A., ... Sameer, H. (2012). Allele frequencies of 15 STR loci using AmpF/STR Identifier kit in the Maldivian population.

- Forensic Science International: Genetics*, 6(5), e136. <https://doi.org/10.1016/j.fsigen.2012.01.007>
- Alves, H. B., Leite, F. P., Sotomaior, V. S., Rueda, F. F., Silva, R., & Moura-Neto, R. S. (2014). STR data for 15 autosomal STR markers from Parana (Southern Brazil). *International Journal of Legal Medicine*, 128(2), 269–270. <https://doi.org/10.1007/s00414-013-0859-y>
- Andreassen, R., Pereira, L., Dupuy, B. M., & Mevaag, B. (2010). Icelandic population data for the STR loci in the AMPFISTR SGM Plus system and the PowerPlex Y-system. *Forensic Science International: Genetics*, 4(4), e101. <https://doi.org/10.1016/j.fsigen.2009.08.011>
- Arunkumar, G. P., Wei, L.-H., Kavitha, V. J., Syama, A., Arun, V. S., Sathua, S., ... The Genographic Consortium. (2015). A late Neolithic expansion of Y chromosomal haplogroup O2a1-M95 from east to west. *Journal of Systematics and Evolution*, 53(6), 546–560.
- Bai, R., Zhao, L., Sun, J., Tian, Q., Wang, J., Zhang, J., ... Ma, S. (2017). Genetic variation of 18 STR loci in the Changsha Han population from Hunan Province, South Central China. *Forensic Science International: Genetics*, 31, e43–e45. <https://doi.org/10.1016/j.fsigen.2017.07.008>
- Bao Maohong, L. Y., & Wenze, B. (2014). *A collection of studies on the history and culture of southeast Asia*. Fujian, China: Xiamen University Press.
- Bernal, L. P., Borjas, L., Zabala, W., Portillo, M. G., Fernández, E., Delgado, W., ... García, O. (2006). Genetic variation of 15 STR autosomal loci in the Maracaibo population from Venezuela. *Forensic Science International*, 161(1), 60–63. <https://doi.org/10.1016/j.forsciint.2005.07.011>
- Bindu, G. H., Trivedi, R., & Kashyap, V. K. (2007). Allele frequency distribution based on 17 STR markers in three major Dravidian linguistic populations of Andhra Pradesh, India. *Forensic Science International*, 170(1), 76–85. <https://doi.org/10.1016/j.forsciint.2006.04.012>
- Burusphat, S. (2007). A comparison of general classifiers in Tai-Kadai languages. *Mon-Khmer Studies. A Journal of Southeast Asian Languages and Cultures*, 37, 129–153.
- Cadenas, A. M., Zhivotovsky, L. A., Cavalli-Sforza, L. L., Underhill, P. A., & Herrera, R. J. (2008). Y-chromosome diversity characterizes the Gulf of Oman. *European Journal of Human Genetics*, 16(3), 374–386. <https://doi.org/10.1038/sj.ejhg.5201934>
- Cai, X., Qin, Z., Wen, B. O., Xu, S., Wang, Y. I., Lu, Y., ... Li, H. (2011). Human migration through bottlenecks from Southeast Asia into East Asia during Last Glacial Maximum revealed by Y chromosomes. *PLoS One*, 6(8), e24282. <https://doi.org/10.1371/journal.pone.0024282>
- Calzada, P., Suarez, I., Garcia, S., Barrot, C., Sanchez, C., Ortega, M., ... Gene, M. (2005). The Fang population of Equatorial Guinea characterised by 15 STR-PCR polymorphisms. *International Journal of Legal Medicine*, 119(2), 107–110. <https://doi.org/10.1007/s00414-004-0505-9>
- Cardoso, S., Sevillano, R., Illescas, M. J., & de Pancorbo, M. M. (2016). Analysis of 16 autosomal STRs and 17 Y-STRs in an indigenous Maya population from Guatemala. *International Journal of Legal Medicine*, 130(2), 365–366. <https://doi.org/10.1007/s00414-015-1195-1>
- Chan, K. M., Chiu, C. T., Tsui, P., Wong, D. M., & Fung, W. K. (2005). Population data for the Identifier 15 STR loci in Hong Kong Chinese. *Forensic Science International*, 152(2–3), 307–309. <https://doi.org/10.1016/j.forsciint.2004.10.018>
- Chen Chunbao, S. Z., Tian Xin, W. U., & Hanhua, W. J. (2017). Genetic polymorphisms of 19 STR loci in Han population from Hainan and its application in paternity testing. *Chinese Journal of Forensic Medicine*, 32(3), 276–280. <https://doi.org/10.13618/j.issn.1001-5728.2017.03.012>
- Chen, P., Wang, B., He, G., Song, F., Luo, H., Zhang, L., ... Han, Y. (2017). Population genetic analyses of 20 autosomal STR loci in Chinese Han population from Tianjin in Northern China. *Forensic Science International: Genetics*, 27, 184–185. <https://doi.org/10.1016/j.fsigen.2016.11.010>
- Chen, Y. (2015). Obstruents in Proto-Ong-Be: A reconstruction. *Paper presented at the 25th Annual Meeting of the Southeast Asian Linguistics Society (SEALS)*, Chiang Mai, Thailand, May 27–29, 2015 at Payap University.
- Chen, Y., Zhu, Y., Shiming, L. U., You, J., & Ma, Y. (2010). Establishment and application of 18-STR database in Han population from Zhejiang Province. *Journal of Sun Yat-sen University(Medical Sciences)*, 31(1), 122–128. [https://doi.org/10.13471/j.cnki.j.sun.yat-sen.univ\(med.sci\).2010.0011](https://doi.org/10.13471/j.cnki.j.sun.yat-sen.univ(med.sci).2010.0011)
- Chennakrishnaiah, S., Perez, D., Gayden, T., Rivera, L., Regueiro, M., & Herrera, R. J. (2013). Indigenous and foreign Y-chromosomes characterize the Lingayat and Vokkaliga populations of Southwest India. *Gene*, 526(2), 96–106. <https://doi.org/10.1016/j.gene.2013.04.074>
- Coudray, C., Calderon, R., Guitard, E., Ambrosio, B., Gonzalez-Martin, A., & Dugoujon, J. M. (2007). Allele frequencies of 15 tetrameric short tandem repeats (STRs) in Andalusians from Huelva (Spain). *Forensic Science International*, 168(2–3), e21–e24. <https://doi.org/10.1016/j.forsciint.2007.01.016>
- Coudray, C., Guitard, E., El-Chennawi, F., Larrouy, G., & Dugoujon, J.-M. (2007). Allele frequencies of 15 short tandem repeats (STRs) in three Egyptian populations of different ethnic groups. *Forensic Science International*, 169(2–3), 260–265. <https://doi.org/10.1016/j.forsciint.2006.03.017>
- Coudray, C., Guitard, E., Keyser-Tracqui, C., Melhaoui, M., Cherkaoui, M., Larrouy, G., & Dugoujon, J. M. (2007). Population genetic data of 15 tetrameric short tandem repeats (STRs) in Berbers from Morocco. *Forensic Science International*, 167(1), 81–86. <https://doi.org/10.1016/j.forsciint.2006.01.003>
- Decorte, R., Engelen, M., Larno, L., Nelissen, K., Gilissen, A., & Cassiman, J. J. (2004). Belgian population data for 15 STR loci (AmpFISTR SGM Plus and AmpFISTR profiler PCR amplification kit). *Forensic Science International*, 139(2–3), 211–213. <https://doi.org/10.1016/j.forsciint.2003.11.007>
- Deng, Q.-Y., Wang, C.-C., Wang, X.-Q., Wang, L.-X., Wang, Z.-Y., Wu, W.-J., ... the Genographic Consortium. (2013). Genetic affinity between the Kam-Sui speaking Chadong and Mulam people. *Journal of Systematics and Evolution*, 51(3), 263–270. <https://doi.org/10.1111/jse.12009>
- Deng, Q., Xu, L., Gong, J., Zhou, L., Li, S., Deng, X., ... Xie, X. (2007). Genetic relationships among four minorities in Guangxi revealed by analysis of 15 STRs. *Journal of Genetics and Genomics*, 34(12), 1072–1079. [https://doi.org/10.1016/S1673-8527\(07\)60122-2](https://doi.org/10.1016/S1673-8527(07)60122-2)
- Du, R., & Yip, V. F. (1993). *Ethnic groups in China*. Beijing, China: Science Press.
- Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and

- implementing the Evanno method. *Conservation Genetics Resources*, 4, 359–361.
- Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
- Fan, H., Du, Z., Wang, F., Wang, X., Wen, S. Q., Wang, L., ... Qiu, P. (2020). The forensic landscape and the population genetic analyses of Hainan Li based on massively parallel sequencing DNA profiling. *bioRxiv*. <https://doi.org/10.1101/2020.03.27.011064>
- Fan, H., Wang, X., Chen, H., Li, W., Wang, W., & Deng, J. (2019). The Ong Be language-speaking population in Hainan Island: genetic diversity, phylogenetic characteristics and reflections on ethnicity. *Molecular Biology Reports*, 46(4), 4095–4103. <https://doi.org/10.1007/s11033-019-04859-8>
- Fan, H., Wang, X., Chen, H., Long, R., Liang, A., Li, W., ... Deng, J. (2018). The evaluation of forensic characteristics and the phylogenetic analysis of the Ong Be language-speaking population based on Y-STR. *Forensic Science International: Genetics*, 37, e6–e11. <https://doi.org/10.1016/j.fsigen.2018.09.008>
- Fan, H., Wang, X., Chen, H., Zhang, X., Huang, P., Long, R., ... Deng, J. (2018). Population analysis of 27 Y-chromosomal STRs in the Li ethnic minority from Hainan province, southernmost China. *Forensic Science International: Genetics*, 34, e20–e22. <https://doi.org/10.1016/j.fsigen.2018.01.007>
- Fan, H., Wang, X., Ren, Z., He, G., Long, R., Liang, A., ... Deng, J. (2019). Population data of 19 autosomal STR loci in the Li population from Hainan Province in southernmost China. *International Journal of Legal Medicine*, 133(2), 429–431. <https://doi.org/10.1007/s00414-018-1828-2>
- Fan, H., Zhang, X., Wang, X., Ren, Z., Li, W., Long, R., ... Deng, J. (2018). Genetic analysis of 27 Y-STR loci in Han population from Hainan province, southernmost China. *Forensic Science International: Genetics*, 33, e9–e10. <https://doi.org/10.1016/j.fsigen.2017.12.009>
- Fei, X. T. (1999). *The pattern of diversity in unity of the Chinese Nation*. Beijing, China: Central University for Nationalities Press.
- Ferdous, A., Ali, M. E., Alam, S., Hasan, M., Hossain, T., & Akhteruzzaman, S. (2010). Allele frequencies of 10 autosomal STR loci from Chakma and Tripura Tribal populations in Bangladesh. *Molecular Biology International*, 2010, 740152. <https://doi.org/10.4061/2010/740152>
- Gan, R.-J., Pan, S.-L., Mustavich, L. F., Qin, Z.-D., Cai, X.-Y., Qian, J. I., ... Li, H. (2008). Pinghua population as an exception of Han Chinese's coherent genetic structure. *Journal of Human Genetics*, 53(4), 303–313. <https://doi.org/10.1007/s10038-008-0250-x>
- Gayden, T., Cadenas, A. M., Regueiro, M., Singh, N. B., Zhivotovskiy, L. A., Underhill, P. A., ... Herrera, R. J. (2007). The Himalayas as a directional barrier to gene flow. *American Journal of Human Genetics*, 80(5), 884–894. <https://doi.org/10.1086/516757>
- Ge, J. X., Wu, S. D., & Chao, S. J. (1997). *Zhongguo yimin shi (The Migration History of China)*. Fuzhou, China: Fujian People's Publishing House.
- Gehrig, C., Balitzki, B., Kratzer, A., Cossu, C., Malik, N., & Castella, V. (2014). Allelic proportions of 16 STR loci-including the new European Standard Set (ESS) loci—In a Swiss population sample. *International Journal of Legal Medicine*, 128(3), 461–465. <https://doi.org/10.1007/s00414-013-0949-x>
- Guo, F. (2017). Genetic variation of 17 autosomal STR loci in the Dong ethnic minority from Guangxi Zhuang Autonomous Region, South China. *International Journal of Legal Medicine*, 131(6), 1537–1538. <https://doi.org/10.1007/s00414-017-1576-8>
- Guo, F., Li, J., Wei, T., Ye, Q., & Chen, Z. (2017). Genetic variation of 17 autosomal STR loci in the Zhuang ethnic minority from Guangxi Zhuang Autonomous Region in the south of China. *Forensic Science International: Genetics*, 28, e51–e52. <https://doi.org/10.1016/j.fsigen.2017.03.015>
- Gurkan, C., Demirdov, D. K., Yamaci, R. F., & Sevay, H. (2015). Population genetic data for 15 autosomal STR markers in Turkish Cypriots from Cyprus. *Forensic Science International: Genetics*, 14, e1–e3. <https://doi.org/10.1016/j.fsigen.2014.09.004>
- Hansen, J. (2000). Using SPSS for windows and Macintosh: Analyzing and understanding data. *American Statistician*, 59, 113.
- Havas, D., Jeran, N., Efremskova, L., Dordevic, D., & Rudan, P. (2007). Population genetics of 15 AmpliSTR Identifier loci in Macedonians and Macedonian Romani (Gypsy). *Forensic Science International*, 173(2–3), 220–224. <https://doi.org/10.1016/j.forsciint.2006.10.027>
- He, G., Li, Y. E., Wang, Z., Liang, W., Luo, H., Liao, M., ... Wu, J. (2017). Genetic diversity of 21 autosomal STR loci in the Han population from Sichuan province, Southwest China. *Forensic Science International: Genetics*, 31, e33–e35. <https://doi.org/10.1016/j.fsigen.2017.07.006>
- He, G., Wang, M., Liu, J., Hou, Y., & Wang, Z. (2018). Forensic features and phylogenetic analyses of Sichuan Han population via 23 autosomal STR loci included in the Huaxia Platinum System. *International Journal of Legal Medicine*, 132(4), 1079–1082. <https://doi.org/10.1007/s00414-017-1679-2>
- Hedjazi, A., Nikbakht, A., Hosseini, M., Hoseinzadeh, A., & Hosseini, S. M. (2013). Allele frequencies for 15 autosomal STR loci in Fars province population, southwest of Iran. *Legal Medicine (Tokyo)*, 15(4), 226–228. <https://doi.org/10.1016/j.legalmed.2013.01.005>
- Hongdan, W., Bing, K., Ning, S. U., Miao, H. E., Bo, Z., Yuxin, G., ... Zhaoshu, Z. (2017). Evaluation of the genetic parameters and mutation analysis of 22 STR loci in the central Chinese Han population. *International Journal of Legal Medicine*, 131(1), 103–105. <https://doi.org/10.1007/s00414-016-1389-1>
- Huihui Lian, W. G., Lin, F., & Li, B. (2015). Genetic Polymorphisms of 21 Autosomal STR Loci of Fujian Han Population. *Journal of Forensic Medicine*, 31(3), 211–214. <https://doi.org/10.3969/j.issn.1004-5619.2015.03.010>
- Jemeljanova, V., Gobrusjonoka, O., Bergere, O., Latisheva, K., Podovsovnik Axelsson, E., & Zupanic Pajnic, I. (2015). Population data for 15 autosomal STR loci from Latvia. *International Journal of Legal Medicine*, 129(4), 739–740. <https://doi.org/10.1007/s00414-015-1205-3>
- Jin, L., & Su, B. (2000). Natives or immigrants: Modern human origin in east Asia. *Nature Reviews Genetics*, 1(2), 126–133. <https://doi.org/10.1038/35038565>
- Jin, X., Wei, Y., Chen, J., Kong, T., Mu, Y., Guo, Y., ... Zhu, B. (2017). Phylogenetic analysis and forensic genetic characterization of Chinese Uyghur group via autosomal multi STR markers. *Oncotarget*, 8(43), 73837–73845. <https://doi.org/10.18632/oncotarget.17992>
- Jingzhou Wang, Q. Z., Zhai, Z., Wang, H., Zhang, J., & Wang, Y. (2015). Genetic polymorphisms of 19 STRs in Inner Mongolia Han population. *Chinese Journal of Forensic Medicine*, 30(1), 82–84. <https://doi.org/10.13618/j.issn.1001-5728.2015.01.026>
- Khodjet-el-Khil, H., Fadhlaoui-Zid, K., Gusmao, L., Alves, C., Benammar-Elgaaied, A., & Amorim, A. (2012). Allele frequencies for 15 autosomal STR markers in the Libyan population. *Annals*

- of *Human Biology*, 39(1), 80–83. <https://doi.org/10.3109/03014460.2011.630678>
- Khurana, P., Aggarwal, A., Mitra, S., Italia, Y. M., Saraswathy, K. N., Chandrasekar, A., & Kshatriya, G. K. (2014). Y chromosome haplogroup distribution in Indo-European speaking tribes of Gujarat, western India. *PLoS One*, 9(3), e90414. <https://doi.org/10.1371/journal.pone.0090414>
- Kraaijenbrink, T., van Driem, G. L., Opgenort, J. R., Tuladhar, N. M., & de Knijff, P. (2007). Allele frequency distribution for 21 autosomal STR loci in Nepal. *Forensic Science International*, 168(2–3), 227–231. <https://doi.org/10.1016/j.forsciint.2006.02.014>
- Kraaijenbrink, T., van Driem, G. L., Tshering of Gaseló, K., & de Knijff, P. (2007). Allele frequency distribution for 21 autosomal STR loci in Bhutan. *Forensic Science International*, 170(1), 68–72. <https://doi.org/10.1016/j.forsciint.2006.04.006>
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33(7), 1870–1874. <https://doi.org/10.1093/molbev/msw054>
- Kwon, S. Y., Lee, H. Y., Lee, E. Y., Yang, W. I., & Shin, K. J. (2015). Confirmation of Y haplogroup tree topologies with newly suggested Y-SNPs for the C2, O2b and O3a subhaplogroups. *Forensic Science International: Genetics*, 19, 42–46. <https://doi.org/10.1016/j.fsigen.2015.06.003>
- Le Wang, X. Z., Zhang, J., Bai, X., & Ye, J. (2013). Genetic polymorphisms of 18 STRs for Han population in Xinjiang Bazhou District. *Chinese Journal of Blood Transfusion*, 26(10), 1015–1017. <https://doi.org/10.13303/j.cjbt.issn.1004-549x.2013.10.022>
- Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, 47(W1), W256–W259. <https://doi.org/10.1093/nar/gkz239>
- Li, D., Li, H., Ou, C., Lu, Y., Sun, Y., Yang, B. O., ... Jin, L. I. (2008). Paternal genetic structure of Hainan aborigines isolated at the entrance to East Asia. *PLoS One*, 3(5), e2168. <https://doi.org/10.1371/journal.pone.0002168>
- Li, H., Wen, B. O., Chen, S.-J., Su, B., Pramoongjago, P., Liu, Y., ... Jin, L. I. (2008). Paternal genetic affinity between Western Austronesians and Daic populations. *BMC Evolutionary Biology*, 8, 146. <https://doi.org/10.1186/1471-2148-8-146>
- Li, L., Zhang, Y., Wang, W., Zhao, K., Li, G., Huang, Y., & Xu, B. (2020). Allele frequencies of 20 autosomal STR loci for 207 unrelated individuals of the Blang people in China. *International Journal of Legal Medicine*, 134(3), 987–988. <https://doi.org/10.1007/s00414-019-02192-0>
- Li, X. N., Zheng, J. L., & Jun, Y. (2017). Genetic polymorphisms of 15 autosomal STR loci in 3962 individuals from the Han population of Jiangxi, Southeast China. *Forensic Science International: Genetics*, 31, e57–e58. <https://doi.org/10.1016/j.fsigen.2017.08.010>
- Li, Y. I., Hong, Y., Li, X., Yang, J., Li, L., Huang, Y., ... Xu, B. (2015). Allele frequency of 19 autosomal STR loci in the Bai population from the southwestern region of mainland China. *Electrophoresis*, 36(19), 2498–2503. <https://doi.org/10.1002/elps.201500129>
- Li, Z., Zhang, J., Zhang, H., Lin, Z., & Ye, J. (2018). Genetic polymorphisms in 18 autosomal STR loci in the Tibetan population living in Tibet Chamdo, Southwest China. *International Journal of Legal Medicine*, 132(3), 733–734. <https://doi.org/10.1007/s00414-017-1740-1>
- Liang, G. (2013). Eight major evidence of Hainan Island's spinning drift apart from China's Beibu Gulf (in Chinese). *Acta Geologica Sinica*, 7(supplementary issue), 73–76.
- Liao, Y., Chen, L., Huang, R., Wu, W., Liu, D., & Sun, H. (2019). Genomic portrait of Guangdong Liannan Yao population based on 15 autosomal STRs and 19 Y-STRs. *Scientific Reports*, 9(1), 2141. <https://doi.org/10.1038/s41598-018-36262-x>
- Lili Zhang, S. Z., Yin, G., Zhai, Y., Xia, M., Zhai, Q., Ronghua, D. U., & Zhao, Q. (2017). Genetic polymorphisms of 21 autosomal STR Loci of Anhui Han population. *Chinese Journal of Forensic Medicine*, 32(3), 314–315. <https://doi.org/10.13618/j.issn.1001-5728.2017.03.023>
- Liu, C., Liu, C., & Wang, H. (2006a). Genetic diversity at 15 STR loci in two tribal populations in Southern China. *Forensic Science International*, 162(1–3), 28–32. <https://doi.org/10.1016/j.forsciint.2006.06.023>
- Liu, C., Liu, C., & Wang, H. (2006b). STR data for the 15 loci from three minority populations in Guangxi municipality in South China. *Forensic Science International*, 162(1–3), 49–52. <https://doi.org/10.1016/j.forsciint.2006.06.015>
- Liu, Q. L., Chen, Y. F., Huang, X. L., Liu, K. Y., Zhao, H., & Lu, D. J. (2017). Population data and mutation rates of 19 STR loci in seven provinces from China based on Goldeneye DNA ID System 20A. *International Journal of Legal Medicine*, 131(3), 653–656. <https://doi.org/10.1007/s00414-016-1441-1>
- Liu, Y.-S., Chen, J.-G., Mei, T., Guo, Y.-X., Meng, H.-T., Li, J.-F., ... Zhang, L.-P. (2017). Genetic variation and forensic characteristic analysis of 25 STRs of a novel fluorescence co-amplification system in Chinese Southern Shaanxi Han population. *Oncotarget*, 8(33), 55443–55452. <https://doi.org/10.18632/oncotarget.19317>
- Lopes, V., Serra, A., Gamero, J., Sampaio, L., Balsa, F., Oliveira, C., ... Carvalho, M. (2009). Allelic frequency distribution of 17 STRs from Identifiler and PowerPlex-16 in Central Portugal area and the Azores archipelago. *Forensic Science International: Genetics*, 4(1), e1–e7. <https://doi.org/10.1016/j.fsigen.2008.12.001>
- Majumder, P. P. (2010). The human genetic history of South Asia. *Current Biology*, 20(4), R184–R187. <https://doi.org/10.1016/j.cub.2009.11.053>
- Martínez-Cortés, G., Zuñiga-Chiquette, F., Celorio-Sánchez, A. S., Ruiz García, E., Antelo-Figueroa, A. B., Dalpozzo-Valenzuela, V., ... Rangel-Villalobos, H. (2019). Population data for 21 autosomal STR loci (GlobalFiler kit) in two Mexican-Mestizo population from the northwest, Mexico. *International Journal of Legal Medicine*, 133(3), 781–783. <https://doi.org/10.1007/s00414-018-1950-1>
- Maruyama, S., Minaguchi, K., Takezaki, N., & Nambiar, P. (2008). Population data on 15 STR loci using AmpF/STR Identifiler kit in a Malay population living in and around Kuala Lumpur, Malaysia. *Legal Medicine (Tokyo)*, 10(3), 160–162. <https://doi.org/10.1016/j.legalmed.2007.11.002>
- Melo, M. M., Carvalho, M., Lopes, V., Anjos, M. J., Serra, A., Vieira, D. N., ... Corte-Real, F. (2010). Genetic study of 15 STRs loci of Identifiler system in Angola population. *Forensic Science International: Genetics*, 4(5), e153–e157. <https://doi.org/10.1016/j.fsigen.2010.03.010>
- Meng Pan, Z. C., Ding, X., Xiaobin, J. U., Chen, Q. I., Zhou, R., & Zhou, H. (2012). A genetic polymorphism analysis on fifteen short tandem repeat loci among Han population in Jiangsu district. *Jiangsu Medical Journal*, 38(5), 561–564. <https://doi.org/10.19460/j.cnki.0253-3685.2012.05.024>
- Min, L. (1997). *A study of Lingao*. Shanghai, China: Shanghai Far Eastern Publishing House.
- Mou, L. (2007). The formation and distribution of ethnic minorities in Southeast Asia (in Chinese). *Southeast*, 2, 47–59.

- Munoz, A., Albeza, M. V., Acreche, N., Castro, J. A., Ramon, M. M., & Picornell, A. (2012). Allele frequencies of 15 STRs in the Calchaqui Valleys population (North-Western Argentina). *Forensic Science International: Genetics*, 6(1), e58–e60. <https://doi.org/10.1016/j.fsigen.2011.05.002>
- Munoz, P., Pinto de Erazo, E. L., Baeza, C., Arroyo-Pardo, E., & Lopez-Parra, A. M. (2015). Genetic polymorphism of 15 STR loci in El Salvador. *International Journal of Legal Medicine*, 129(5), 991–993. <https://doi.org/10.1007/s00414-015-1148-8>
- Muro, T., Fujihara, J., Imamura, S., Nakamura, H., Yasuda, T., & Takeshita, H. (2008). Allele frequencies for 15 STR loci in Ovambo population using AmpFISTR Identifiler Kit. *Legal Medicine (Tokyo)*, 10(3), 157–159. <https://doi.org/10.1016/j.legalmed.2007.10.003>
- Nei, M. (1977). F-statistics and analysis of gene diversity in subdivided populations. *Annals of Human Genetics*, 41(2), 225–233. <https://doi.org/10.1111/j.1469-1809.1977.tb01918.x>
- Norquest, P. (2015). *A phonological reconstruction of Proto-Hlai*. Leiden, the Netherlands: Brill.
- Okolie, V. O., Cisana, S., Schanfield, M. S., Adekoya, K. O., Oyedeji, O. A., Podini, D., & Akanmu, A. S. (2018). Population data of 21 autosomal STR loci in the Hausa, Igbo and Yoruba people of Nigeria. *International Journal of Legal Medicine*, 132(3), 735–737. <https://doi.org/10.1007/s00414-017-1722-3>
- Osborne, M. (2012). *Southeast Asia: An introductory history*. Beijing, China: The Commercial Press.
- Ostapirat, W. (2005). Kra-Dai and Austronesian: Notes on phonological correspondences and vocabulary distribution. In L. Sagart, R. Blench, & A. Sanchez-Mazas (Eds.), *The peopling of East Asia: Putting together archaeology, linguistics and genetics*. London, UK: RoutledgeCurzon.
- Petric, G., Draskovic, D., Zgonjanin-Bosic, D., Budakov, B., & Veselinovic, I. (2012). Genetic variation at 15 autosomal STR loci in the Hungarian population of Vojvodina Province, Republic of Serbia. *Forensic Science International: Genetics*, 6(6), e163–e165. <https://doi.org/10.1016/j.fsigen.2012.05.003>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959.
- Qiuling Liu, D. L., Huiling, L. U., & Chen, L. (2003). Genetic polymorphisms of 15-STR loci Taiwan Han population. *Journal of Law and Medicine*, 10(4), 245–247.
- Regueiro, M., Cadenas, A. M., Gayden, T., Underhill, P. A., & Herrera, R. J. (2006). Iran: Tricontinental nexus for Y-chromosome driven migration. *Human Heredity*, 61(3), 132–143. <https://doi.org/10.1159/000093774>
- Ristow, P. G., & D'Amato, M. E. (2017). Forensic statistics analysis toolbox (FORSTAT): A streamlined workflow for forensic statistics. *Forensic Science International: Genetics Supplement Series*, 6, e52–e54. <https://doi.org/10.1016/j.fsigs.2017.09.006>
- Rodriguez, J. J., Salvador, J. M., Calacal, G. C., Laude, R. P., & De Ungria, M. C. (2015). Allele frequencies of 23 autosomal short tandem repeat loci in the Philippine population. *Legal Medicine (Tokyo)*, 17(4), 295–297. <https://doi.org/10.1016/j.legalmed.2015.02.005>
- Rosenberg, N. A. (2003). Distruct: A program for the graphical display of population structure. *Molecular Ecology Notes*, 4, 137–138.
- Rowold, D. J., Gayden, T., Luis, J. R., Alfonso-Sanchez, M. A., Garcia-Bertrand, R., & Herrera, R. J. (2019). Investigating the genetic diversity and affinities of historical populations of Tibet. *Gene*, 682, 81–91. <https://doi.org/10.1016/j.gene.2018.09.043>
- Rowold, D. J., Perez Benedico, D., Garcia-Bertrand, R., Chennakrishnaiah, S., Alfonso-Sanchez, M. A., Gayden, T., & Herrera, R. J. (2016). Ladakh, India: The land of high passes and genetic heterogeneity reveals a confluence of migrations. *European Journal of Human Genetics*, 24(3), 442–449. <https://doi.org/10.1038/ejhg.2015.80>
- Sadam, M., Tasa, G., Tiidla, A., Lang, A., Axelsson, E. P., & Pajnic, I. Z. (2015). Population data for 22 autosomal STR loci from Estonia. *International Journal of Legal Medicine*, 129(6), 1219–1220. <https://doi.org/10.1007/s00414-014-1089-7>
- Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4), 406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>
- Sciences, C. A. O. S. (2012). *Language Atlas of China* (2nd ed.). Beijing, China: The Commercial Press (Chinese dialect volume).
- Scientific Working Group on DNA Analysis (SWGDM). (2010). *SWGDM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories*.
- Serga, S. V., Dombrovskiy, I. V., Maistrenko, O. M., Ostapchenko, L. I., Demydov, S. V., Krivda, R. G., & Kozeretska, I. A. (2017). Allele frequencies for 15 STR loci in the Ukrainian population. *Forensic Science International: Genetics*, 29, e40–e41. <https://doi.org/10.1016/j.fsigen.2017.05.004>
- Shan, M. A., Hussain, M., Shafique, M., Shahzad, M., Perveen, R., & Idrees, M. (2016). Genetic distribution of 15 autosomal STR markers in the Punjabi population of Pakistan. *International Journal of Legal Medicine*, 130(6), 1487–1488. <https://doi.org/10.1007/s00414-016-1351-2>
- Shen, J., Kang, H., Dong, F., Guo, J., & Wang, R. (2015). Research of Genetic Polymorphism on 19 STR Loci in Han Nationality in Shanxi Province and Forensic Medicine Application. *Journal of Shanxi Police Academy*, 23(2), 63–66.
- Shotvaranon, J., Chirachariyavej, T., Leetrakool, N., & Rerkamnuaychoke, B. (2009). DNA database of populations from different parts in the Kingdom of Thailand. *Forensic Science International: Genetics*, 4(1), e37–e38. <https://doi.org/10.1016/j.fsigen.2009.02.009>
- Soltyszewski, I., Pepinski, W., Wolanska-Nowak, P., Maciejewska, A., Paszkowska, R., Abreu-Glowacka, M., ... Ploski, R. (2014). Polish population data on 15 autosomal STRs of AmpFISTR NGM PCR kit. *Forensic Science International: Genetics*, 9, 142–149. <https://doi.org/10.1016/j.fsigen.2013.12.002>
- Song, M., Wang, Z., Zhang, Y., Zhao, C., Lang, M., Xie, M., ... Hou, Y. (2019). Forensic characteristics and phylogenetic analysis of both Y-STR and Y-SNP in the Li and Han ethnic groups from Hainan Island of China. *Forensic Science International: Genetics*, 39, e14–e20. <https://doi.org/10.1016/j.fsigen.2018.11.016>
- Steinlechner, M., Schmidt, K., Kraft, H. G., Utermann, G., & Parson, W. (2002). Gabon black population data on the ten short tandem repeat loci D3S1358, VWA, D16S539, D2S1338, D8S1179, D21S11, D18S51, D19S433, TH01 and FGA. *International Journal of Legal Medicine*, 116(3), 176–178. <https://doi.org/10.1007/s004140100234>
- Sun, L., Shi, K. E., Tan, L. U., Zhang, Q., Fu, L., Zhang, X., ... Cong, B. (2017). Analysis of genetic polymorphisms and mutations at 19 STR loci in Hebei Han population. *Forensic Science International: Genetics*, 31, e50–e51. <https://doi.org/10.1016/j.fsigen.2017.07.016>
- Sun, M., Zhang, X., Wu, D., Shen, Q., & Wu, Y. (2015). Population genetic data of 15 STR loci in Gansu Han population from China.

- International Journal of Legal Medicine*, 129(4), 735–736. <https://doi.org/10.1007/s00414-014-1132-8>
- Sun, Y., Yang, B., Ou, C., Chen, L., Su, Z., & Li, D. (2007). Investigation into the origin of Li ethnic group in China by genetic analysis of Y chromosome single nucleotide polymorphism. *China Tropical Medicine*, 7, 1527–1529.
- Sun, Y., Yang, B., Ou, C., Zhou, Z., Su, Z., & Li, D. (2007). Origins of the three minority populations in Hainan Island as seen from Y-SNP. *Science and Technology Review*, 25, 44–47.
- Tattersall, I. (2009). Out of Africa: Modern human origins special feature: Human origins: Out of Africa. *Proceedings of the National Academy of Sciences of the United States of America*, 106(38), 16018–16021. <https://doi.org/10.1073/pnas.0903207106>
- Thurgood, G. (1994). Tai-Kadai and Austronesian: The nature of the historical relationship. *Oceanic Linguistics*, 33, 345–368.
- Tillmar, A. O., Backstrom, G., & Montelius, K. (2009). Genetic variation of 15 autosomal STR loci in a Somali population. *Forensic Science International: Genetics*, 4(1), e19–e20. <https://doi.org/10.1016/j.fsigen.2009.01.004>
- Tokdemir, M., Tuncez, F. T., & Vicdanli, N. H. (2016). Population Genetic data for 15 Autosomal STR markers in Eastern Turkey. *Gene*, 586(1), 36–40. <https://doi.org/10.1016/j.gene.2016.03.014>
- Toscanini, U., Moreno, F., Pantoja-Astudillo, J. A., Morales, E. A., Bustos, P., & Salas, A. (2015). A reference frequency database of 15 autosomal STRs in Chile. *Forensic Science International: Genetics*, 19, 35–36. <https://doi.org/10.1016/j.fsigen.2015.05.016>
- Tran, H. L., Nguyen, H. T., Pham, T. T., Nguyen, M. H., Hoang, H., & Chu, H. H. (2019). Allele frequencies for 22 autosomal STRs in the Kinh population in Vietnam. *International Journal of Legal Medicine*, 133(6), 1761–1762. <https://doi.org/10.1007/s00414-018-01996-w>
- Ventura Spagnolo, E., Mondello, C., Procaccianti, S., Marretta, S., Zerbo, S., Asmundo, A., & Argo, A. (2017). Allele frequencies distribution of 16 forensic STR loci in a Western Sicilian population sample. *Egyptian Journal of Forensic Sciences*, 13(7), 1–5. <https://doi.org/10.1186/s41935-017-0013-y>
- Wang, C.-C., Wang, L.-X., Shrestha, R., Wen, S., Zhang, M., Tong, X., ... Li, H. (2015). Convergence of Y Chromosome STR haplotypes from different SNP haplogroups compromises accuracy of haplogroup prediction. *Journal of Genetics and Genomics*, 42(7), 403–407. <https://doi.org/10.1016/j.jgg.2015.03.008>
- Wang, M., Wang, Z., He, G., Jia, Z., Liu, J., & Hou, Y. (2018). Genetic characteristics and phylogenetic analysis of three Chinese ethnic groups using the Huaxia Platinum System. *Scientific Reports*, 8(1), 2429. <https://doi.org/10.1038/s41598-018-20871-7>
- Wang XP, W. J., & Xing, G. Y. (2004). *China Hlai*. Beijing, China: The Ethnic Publishing House.
- Wang, Z., Zhou, D., Jia, Z., Li, L., Wu, W., Li, C., & Hou, Y. (2016). Developmental validation of the Huaxia Platinum System and application in 3 main ethnic groups of China. *Scientific Reports*, 6, 31075. <https://doi.org/10.1038/srep31075>
- Weera, O. (1998). A mainland Bê language. *Journal of Chinese Linguistics*, 26, 338–344.
- Weera, O. (2017). *Kra-Dai in Southern China*. Tianjin, China: Invited talk in the Nankai University.
- Weimin Zhou, L. T. (2018). *Hainan general history (in Chinese)*. Beijing: People's Publishing House.
- Wells, R. S., Yuldasheva, N., Ruzibakiev, R., Underhill, P. A., Evseeva, I., Blue-Smith, J., ... Bodmer, W. F. (2001). The Eurasian heartland: A continental perspective on Y-chromosome diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 98(18), 10244–10249. <https://doi.org/10.1073/pnas.171305098>
- Wen, B. O., Li, H., Lu, D., Song, X., Zhang, F., He, Y., ... Jin, L. I. (2004). Genetic evidence supports demic diffusion of Han culture. *Nature*, 431(7006), 302–305. <https://doi.org/10.1038/nature02878>
- Wenming Han, Q. Z. (2016). Genetic polymorphisms of 24 STRs for Beijing and Ningbo Han populations. *Chinese Journal of Forensic Medicine*, 31(6), 621–625.
- Xiao, C., Zhang, W., Wei, T., Pan, C., & Huang, D. (2016). Population data of 21 autosomal STR loci in Chinese Han population from Hubei province in Central China. *Forensic Science International: Genetics*, 20, e13–e14. <https://doi.org/10.1016/j.fsigen.2015.11.002>
- Xie, J., Shao, C., Zhou, Y., Zhu, W., Xu, H., Liu, Z., & Zhao, Z. (2014). Genetic distribution on 20 STR loci from the Han population in Shanghai, China. *Forensic Science International: Genetics*, 9, e30–e31. <https://doi.org/10.1016/j.fsigen.2013.08.007>
- Xing, G. Y. (2004). Discursion from the earliest governmentalization of Hainan Island. *Hainan Archives*, 3, 43–47.
- Xiuzi, M. A., & Changchun, H. A. O. (2017). Polymorphism of 15 short tandem repeat loci in Hui and Han nationality of Ningxia district. *Journal of Shanxi Normal University (Natural Science Edition)*, 45(5), 66–71. <https://doi.org/10.15983/cnki.snu.2017.05.354>
- Xu, X. M., Feng, Z., & Yao, J. (2017). Population data of 15 autosomal STR loci in Chinese Han population from Jilin Province, Northeast China. *Forensic Science International: Genetics*, 27, 187–188. <https://doi.org/10.1016/j.fsigen.2016.12.012>
- Xu, X. M., Xu, Z. H., Wang, D. W., & Yao, J. (2017). Population data of 18 autosomal STR loci in the Chinese Han population from Heilongjiang Province, Northeast China. *Forensic Science International: Genetics*, 29, e33–e34. <https://doi.org/10.1016/j.fsigen.2017.04.002>
- Yan, G. (2013). *History and culture of Hainan (in Chinese)*. Beijing, China: Social Sciences Academic Press.
- Yang, L., Zhang, X., Zhao, L., Sun, Y., Li, J., Huang, R., ... Nie, S. (2018). Population data of 23 autosomal STR loci in the Chinese Han population from Guangdong Province in southern China. *International Journal of Legal Medicine*, 132(1), 133–135. <https://doi.org/10.1007/s00414-017-1588-4>
- Yang, M., Ren, Z., Ji, J., Zhou, H., Zhang, H., Dai, J., ... Huang, J. (2017). Population genetic data and mutations of 22 autosomal STR loci in Guizhou Han population. *Forensic Science International: Genetics*, 29, e29–e30. <https://doi.org/10.1016/j.fsigen.2017.03.008>
- Yao, J., Wang, L. M., Gui, J., Xing, J. X., Xuan, J. F., & Wang, B. J. (2016). Population data of 15 autosomal STR loci in Chinese Han population from Liaoning Province, Northeast China. *Forensic Science International: Genetics*, 23, e20–e21. <https://doi.org/10.1016/j.fsigen.2016.04.012>
- Yingming, L. (2010). *The history of Southeast Asia (in Chinese)*. Beijing, China: People's Publishing House.
- Yoo, S. Y., Cho, N. S., Park, M. J., Seong, K. M., Hwang, J. H., Song, S. B., ... Chung, K. W. (2011). A large population genetic study of 15 autosomal short tandem repeat loci for establishment of Korean DNA profile database. *Molecules and Cells*, 32(1), 15–19. <https://doi.org/10.1007/s10059-011-2288-4>
- Zhang, L. (2015). Population data for 15 autosomal STR loci in the Bouyei ethnic minority from Guizhou Province, Southwest China. *Forensic Science International: Genetics*, 17, 108–109. <https://doi.org/10.1016/j.fsigen.2015.04.006>
- Zhang, L., Zhao, Y., Guo, F., Liu, Y., & Wang, B. (2015). Population data for 15 autosomal STR loci in the Miao ethnic minority from

- Guizhou Province, Southwest China. *Forensic Science International: Genetics*, 16, e3–e4. <https://doi.org/10.1016/j.fsigen.2014.11.002>
- Zhang, M. X., Han, S. Y., Gao, H. M., Sun, S. H., Xiao, D. J., Liu, Y., & Wang, Y. S. (2013). Genetic polymorphisms of 19 STR loci in Shandong Han population. *Journal of Forensic Medicine*, 29(6), 440–446. <https://doi.org/10.3969/j.issn.1004-5619.2013.06.010>
- Zhang, T. Y. (1974). *Ming Shi (The history of Ming dynasty, volume three hundred and four)*. Beijing, China: Zhonghua Book Company.
- Zhang, X., Du, L., Huang, Z., Gu, T., Hu, L., & Nie, S. (2017). Genetic variation of 20 autosomal STR loci in three ethnic groups (Zhuang, Dai and Hani) in the Yunnan province of southwestern China. *Forensic Science International: Genetics*, 31, e41–e42. <https://doi.org/10.1016/j.fsigen.2017.06.005>
- Zhang, X., Hu, L., Du, L., Nie, A., Rao, M., Pang, J. B., ... Nie, S. (2017). Genetic analysis of 20 autosomal STR loci in the Miao ethnic group from Yunnan Province, Southwest China. *Forensic Science International: Genetics*, 28, e28–e29. <https://doi.org/10.1016/j.fsigen.2017.02.003>
- Zhang, X., Hu, L., Du, L., Zheng, H., Nie, A., Rao, M., ... Nie, S. (2017). Population data for 20 autosomal STR loci in the Yi ethnic minority from Yunnan Province, Southwest China. *Forensic Science International: Genetics*, 28, e43–e44. <https://doi.org/10.1016/j.fsigen.2017.02.017>
- Zhang, X., Liao, S., Qi, X., Liu, J., Kampuansai, J., Zhang, H., ... Su, B. (2015). Y-chromosome diversity suggests southern origin and Paleolithic backwave migration of Austro-Asiatic speakers from eastern Asia to the Indian subcontinent. *Scientific Reports*, 5, 15486. <https://doi.org/10.1038/srep15486>
- Zhang, X., Liu, L., Xie, R., Wang, G., Shi, Y., Gu, T., ... Nie, S. (2018). Population data and mutation rates of 20 autosomal STR loci in a Chinese Han population from Yunnan Province, Southwest China. *International Journal of Legal Medicine*, 132(4), 1083–1085. <https://doi.org/10.1007/s00414-017-1675-6>
- Zhao, H., Wang, L., & Yuan, J. (2007). Origin and time of Qiongzhou strait. *Marine Geology & Quaternary Geology*, 27, 33–40.
- Zhong, Y. (2010). *Historical Hainan (in Chinese)*. Hainan, China: Hainan Publishing House.
- Zhu, J., Shen, C., Ma, Y., He, Y., Zhao, J., Li, X., & Liu, Y. (2007). Genetic polymorphisms of 15 STR in Chinese Salar ethnic minority group. *Forensic Science International*, 173(2–3), 210–213. <https://doi.org/10.1016/j.forsciint.2007.01.013>
- Zou, X., Li, Y. G., Li, P., Nie, Q. Y., Wang, T., Hu, Y., ... Tang, R. K. (2017). Genetic polymorphisms for 19 autosomal STR loci of Chongqing Han ethnicity and phylogenetic structure exploration among 28 Chinese populations. *International Journal of Legal Medicine*, 131(6), 1539–1542. <https://doi.org/10.1007/s00414-017-1577-7>
- Zúñiga, J., Ilzarbe, M., Acunha-Alonzo, V., Rosetti, F., Herbert, Z., Romero, V., ... Yunis, E. J. (2006). Allele frequencies for 15 autosomal STR loci and admixture estimates in Puerto Rican Americans. *Forensic Science International*, 164(2–3), 266–270. <https://doi.org/10.1016/j.forsciint.2005.11.022>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Li W, Wang X, Wang X, et al. Forensic characteristics and phylogenetic analyses of one branch of Tai-Kadai language-speaking Hainan Hlai (Ha Hlai) via 23 autosomal STRs included in the Huaxia™ Platinum System. *Mol Genet Genomic Med*. 2020;8:e1462. <https://doi.org/10.1002/mgg3.1462>