Legal Medicine 53 (2021) 101954

Contents lists available at ScienceDirect

Legal Medicine

journal homepage: www.elsevier.com/locate/legalmed

Announcement of Population Data

Forensic characteristics and phylogenetic analysis of the Chinese Han population from Chongqing Municipality, Southwest China

Chi-Zao Wang [a,b,c,d,1], Shuang Li [e,1], Xu Hao [e], Chuan-Chao Wang [f], Ru-Feng Bai [g,*], Hui Li [d,*], Shu-Hua Ma [a,b,c,*]

[a] Department of Radiology, the First Affiliated Hospital of Shantou University Medical College, Shantou 515041, China
[b] Shantou University Medical College, Shantou, Guangdong 515041, China
[c] Guangdong Key Laboratory of Medical Molecular Imaging, Shantou 515041, China
[d] MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200438, China
[e] Shapingba Branch of Chongqing Public Security Bureau, Chongqing 400030, China
[f] Department of Anthropology and Ethnology, Institute of Anthropology, National Institute for Data Science in Health and Medicine, and School of Life Sciences, Xiamen University, Xiamen 361005, China
[g] Criminal Justice College of China University of Political Science and Law, Beijing 100088, China

ARTICLE INFO

Keywords:
Chongqing Han
Forensic characteristics
Phylogenetic analysis

ABSTRACT

Chongqing Han is an important southern Han group, but investigations on its paternal genetic structure are still limited. Here, we analyzed the forensic and phylogenetic characteristics of the Chongqing Han population based on 27 Y-STR and predicted Y-SNP markers. Based on AMOVA, haplogroup distribution and network analysis, we explored the genetic relationship between Chongqing Han, other Chinese groups and some southern indigenous groups (speaking Kra-Dai, Austronesian, etc).

1. Introduction

Chongqing is the largest city in southwestern China. It was once considered a barbaric land in historical Chinese literature, with many ancient groups, such as Pu and Ba, residing in this area. With the strong diffusion of the Han population and the Han culture in the past two thousand years, Han Chinese has become the main population in Chongqing. According to the latest census (2010), the Han population of Chongqing is 26,909,061, accounting for 93.28% of the total population. Two widely used markers on the male-specific region of the Y chromosome, STRs and SNPs (single-nucleotide polymorphisms), play an important role in human evolution, population history, genealogy, forensics and male medical genetics [1]. Although some forensic genetic analyses have been conducted on Chongqing Han people for familial searching [2], regional lineage accumulations and spatial comparisons among groups were still neglected. In order to investigate the phylogenetic history and forensic characteristics of the Han nationality in Chongqing, a total of 559 unrelated male individuals were recruited and genotyped by the Yfiler® Plus Kit (Thermo Fisher Scientific, Waltham, MA, USA).

2. Materials and methods

All samples were collected after receiving informed consent, and individuals were considered autochthonous if their ancestors had lived in Chongqing municipality for at least three generations. The Ethical Committee of Fudan University, Shanghai, People's Republic of China approved the study (No. 14012). Experimental methods were shown in Table S1. Data were submitted to the YHRD (Ychromosomal haplotype reference database, https://yhrd.org [3]) under accession number YA004232 for the Chongqing Han population. Based on the Y chromosome database in our laboratory [4], we predicted the haplogroup information of the Chongqing Han group. In addition, we corrected the predicting results with reference to the Whit Atheys' haplogroup predictor (http://www.hprg.com/hapest5/). Haplotype and Haplogroup frequencies were estimated by direct gene-counting. Haplotype diversities were calculated according to Nei's formula [5]. The discrimination capacity (GD) and match probability (MP) were also calculated as important forensic parameters [6]. Pairwise genetic distances of Rst and corresponding P-values between different populations were evaluated by analysis of molecular variance (AMOVA) and visualized using

* Corresponding authors.
E-mail addresses: bairufeng@cupl.edu.cn (R.-F. Bai), LHCA@fudan.edu.cn (H. Li), shuhua6699@163.com (S.-H. Ma).
[1] Chi-Zao Wang and Shuang Li contributed equally to this work.

https://doi.org/10.1016/j.legalmed.2021.101954
Received 19 August 2020; Received in revised form 20 March 2021; Accepted 4 August 2021
Available online 8 August 2021
1344-6223/

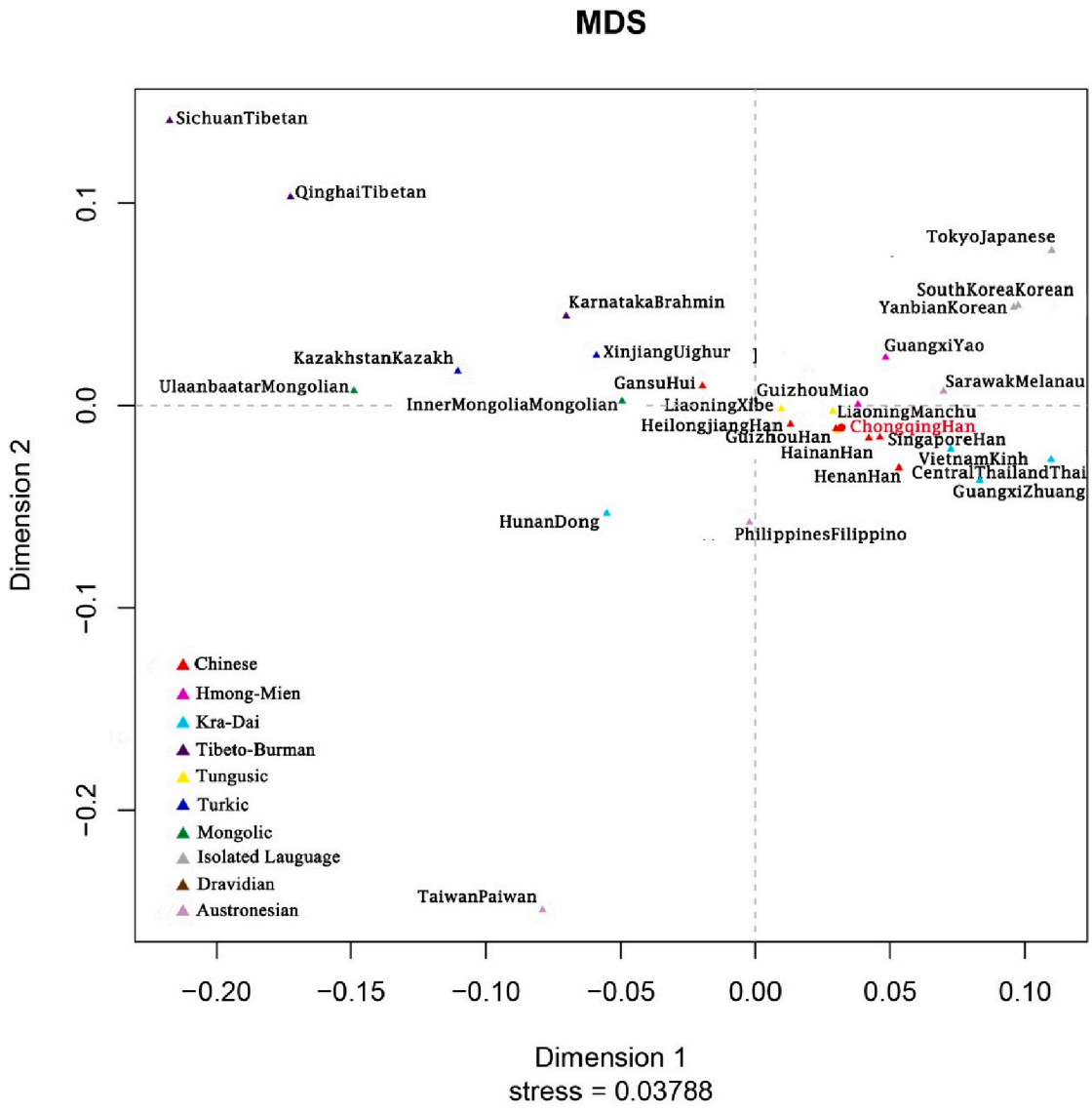

Fig. 1. Multidimensional scaling (MDS) analysis for the Chongqing Han and 27 other reference populations based on Rst values. The red dot represents the studied population. All populations are marked with different colours according to the language classification, as shown in the lower left corner of the Figure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

multidimensional scaling (MDS) plots using statistical online tool available on the YHRD website. The *P*-values were then calculated at a significance level of 0.05 using 10,000 permutations. A Bonferroni correction was also applied to adjust for potential type I errors [7]. With reference to Y-DNA haplogroup Tree 2019 (https://isogg.org/tree/), we constructed a simplified phylogenetic tree that showed the distribution of Chongqing Han samples. NETWORK 5.0.1.1 was imported to present median-joining trees of the Chongqing Han ethnic group [8].

3. Results and discussion

The allele frequencies and the GD values for 23 single-copy STR loci and two multi-copy STR were summarized in Table S2. 320 alleles were detected across the 27 Y-STR loci tested, with frequencies between 0.0018 and 0.7352. The two multi-copy Y-STRs, DYS385 a/b and DYF387S1 a/b, were the most diverse markers with gene diversity (GD) values of 0.9644 and 0.9556, respectively. Among 23 single-copy loci, DYS449 was the most informative locus with the GD of 0.8831. In addition, we observed double peaks occurred in the loci of DYS389I (13,14), DYS389 II (28,29), DYS439 (12,13) and DYS635 (21,23) allele ranged in a male sample. Double peaks were confirmed by using a domestic commercial Microreader™ 29Y Direct ID System (Microread Genetics, Suzhou, Jiangsu, China) (Fig. S1). The same 27 Y-STR haplotype was also found in his son. These duplicated alleles were verified by Sanger sequencing (Fig. S2). The most likely explanation for this phenomenon was the presence of segment duplication involving at least these four loci located in the same Y-chromosome region, the 780 kb AZFa segment. Bosch and Jobling described the AZFa duplications as the reciprocal product of recombination between HERVs flanking the region [9,10]. Therefore, the presence of double alleles does not always indicate the presence of multiple male profiles. It is important to know the incidence of duplication mutations involving Y-STRs in forensic investigations.

A total of 559 distinct haplotypes of the 27 Y-STRs were observed in 559 Chongqing Han individuals and which gave HD value of 1.0000 (Table S3). The result indicates that the Yfiler® Plus kit offers excellent discrimination capabilities and can be useful in forensic investigations and paternal lineage identification in this Han population. To extensively illustrate the genetic relationship of these markers, we compared the 17 Yfiler haplotype data of the Chongqing Han with those of 27

neighboring populations according to the language classification (published and referenced in the YHRD) (n = 16,550) (Table S4 and Fig. S3a). Table S5 showed that there were no significant differences ($P > 0.0018$ after Bonferroni's correction) between the Chongqing Han and Han Chinese from Guizhou (Rst = 0.0003, p = 0.2317), Hainan (Rst = 0.0049, p = 0.0024) and Heilongjiang (Rst = 0.0090, p = 0.0371), indicating little genetic difference. Although significant, low Rst values were obtained between the Chongqing Han and the other three Chinese-speaking populations (Henan-Han, Singapore-Han and Gansu-Hui) (Rst: 0.0103 ~ 0.0379) and eight other groups in East Asia (Manchu, Xibe, Melanau, Thai, Filipino, Yao, Zhuang and Miao) (Rst: 0.0079 ~ 0.0462). The low Rst values between the Chongqing Han and the other southern groups represented by Sarawak-Melanau might suggest the genetic connections between the populations in southern China and Southeast Asia due to the Austronesian expansion [11]. In comparisons between the Chongqing Han and the remaining populations, highly significant distances were observed (Rst: 0.0648 ~ 0.3038, $P < 0.0018$). The MDS plot (Fig. 1) structured from Rst distance matrix showed that Heilongjiang-Han, Guizhou-Han, Hainan-Han, Henan-Han, Singapore-Han and Chongqing Han were gathered in the lower right quadrant, confirming their genetic affinity. In addition, some groups from southern China and Southeast Asia, including Guizhou-Miao (Hmong-Mien), Guangxi-Yao (Hmong-Mien), Guangxi-Zhuang (Kra-Dai), Vietnam-Kinh (Kra-Dai), Central Thailand-Thai (Kra-Dai) and Sarawak-Melanau (Austronesian), were also close to the Chongqing Han, which showed the genetic affinities among populations could be well reflected by geographic locations.

After analysing the phylogenetic relationship, 40 different terminal haplogroups were predicted in the dataset (Fig. S4). The majority (74.24%) of Chongqing Han individuals was divided into haplogroup O-M175, the branch where most Han individuals lie [12]. Haplogroups C (8.05%), N (6.98%), D (4.29%) and Q (4.11%) were the remaining majority haplogroups. Three Han dominant lineages encompass 32.56% of the Chongqing Han Chinese in total (estimated 12.88% for O2a2b1a1-M117, 10.02% for O2a2b1a2a1-F444, and 9.66% for O2a1b-002611). This distribution ratio is close to that of the total Han nationality, indicating that the Chongqing Han nationality has a high consistency with other Han nationalities. Meanwhile, the haplogroup C2b-F1067 and Q1a1a1-M120 are also important paternal lineages of the Han nationality [13,14], accounting for 2.68% and 1.43% of the Chongqing Han nationality. Haplogroup O1a-M119, including its most important downstream branch, O1a1a-P203, is the important paternal types shared by Austronesian and the Kra-Dai group [15], a distinct southern indigenous type. It is also relatively high frequency in the Han nationality in Chongqing, accounting for 14.49% of the total population. Haplogroup O1b1a1-M95, containing its downstream branch O1b1a1a1a1a1-M88, is a distinct southern indigenous lineage, too. It is not only the main patrilineal lineage of the Austroasiatic group, but also represents a high proportion in the Hmong-Mien, Kra-Dai and Austronesian population [15,16]. 9.66% of the Chongqing Han carry the M95 genetic marker. In addition to the haplogroup O1a-M119 and O1b1a1-M95, D1a1-M15 (4.29%) and O2a2a1a2-M7(3.58%) also represents a large number of the genetic contribution from southern indigenous groups to Chongqing Han nationality, including groups speaking Hmong-Mien, Tibeto-Burman and so on. In a word, the haplogroup distribution of the Chongqing Han nationality confirms that it is an admixture of Han immigrants and southern natives.

Network analyses were performed based on Y-STR haplotypes of the five most dominant haplogroups (Fig. S5). Among them, O2a2b1a1-M117 (12.88%), O2a2b1a2a1-F444 (10.02%) and O2a1b-002611 (9.66%) were typical high frequency types of Han nationality while O1a-M119 (14.49%) and O1b1a1-M95 (9.66%) had distinctive southern indigenous characteristics [17,18]. Overall, Samples from the Chongqing Han spread over the entire network, rather than being concentrated with certain secondary branches, indicating that their sources are extremely diverse. More details were shown in Fig. S5. Geographically, Chongqing is located in the plain of the upper reaches of the Yangtze River where well-developed rice farming can support a large population [19] (Fig. S3b). Besides, in the development process of the Han population, great importance was attached to cultural identity and a large number of indigenous populations were integrated. This may be important reasons for the formation of the Chongqing Han population structure today.

In conclusion, these data in Chongqing Han population could be potentially useful for the regional specific and prerequisite reference to the forensic, genealogical, and evolutionary purposes. To further reveal the genetic information of the Han populations, more studies with larger sample sizes and marker sets are necessary.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank all sample donors for their contributions to this work and all those who helped with sample collection. This study was supported by the National Natural Science Foundation of China (NSFC No. 81671874, 31801040).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.legalmed.2021.101954.

References

[1] M.A. Jobling, C. Tyler-Smith, Human Y-chromosome variation in the genome-sequencing era, Nat. Rev. Genet. 18 (2017) 485–497, https://doi.org/10.1038/nrg.2017.36.
[2] X.-T. Mo, J. Zhang, W.-H. Ma, X. Bai, W.-S. Li, X.-C. Zhao, J. Ye, Developmental validation of the DNATyper™ Y26 PCR amplification kit: An enhanced Y-STR multiplex for familial searching, Foren. Sci. Int. Genetics 38(2019) 113-120, https://doi.org/10.1016/j.fsigen.2018.10.008.
[3] S. Willuweit, L. Roewer, The new Y Chromosome Haplotype Reference Database, Foren. Sci. Int. Genetics 15 (2015) 43–48, https://doi.org/10.1016/j.fsigen.2014.11.024.
[4] C.-C. Wang, L.-X. Wang, R. Shrestha, S. Wen, M. Zhang, X. Tong, L. Jin, H. Li, Convergence of Y chromosome STR haplotypes from different SNP haplogroups compromises accuracy of haplogroup prediction, J. Genet. Genomics 42 (2015) 403–407, https://doi.org/10.1016/j.jgg.2015.03.008.
[5] M. Nei, Molecular evolutionary genetics, Science 235 (1987) 599, https://doi.org/10.1016/j.jgg.2015.03.008.
[6] S. Ip, S.-W. Lin, T.-T. Lam, Haplotype data of 27 Y-STR loci in Hong Kong Chinese, Foren. Sci. Int. Genetics 38 (2019) e14–e15, https://doi.org/10.1016/j.fsigen.2018.11.001.
[7] Y. Hochberg, A sharper Bonferroni procedure for multiple tests of significance, Biometrika 75 (1988) 800–802, https://doi.org/10.1093/biomet/75.4.800.
[8] H.J. Bandelt, P. Forster, A. Rohl, Median-joining networks for inferring intraspecific phylogenies, Mol. Biol. Evol. 16 (1999) 37–48, https://doi.org/10.1093/oxfordjournals.molbev.a026036.
[9] E. Bosch, M.A. Jobling, Duplications of the AZFa region of the human Y chromosome are mediated by homologous recombination between HERVs and are compatible with male fertility, Hum. Mol. Genet. 12 (2003) 341–347, https://doi.org/10.1093/hmg/ddg031.
[10] J.M. Butler, A.E. Decker, M.C. Kline, P.M. Vallone, Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation, J. Foren. Sci. 50 (2005) 1–7, https://doi.org/10.1520/JFS2004481.
[11] M. Lipson, P.R. Loh, N. Patterson, P. Moorjani, Y.C. Ko, M. Stoneking, B. Berger, D. Reich, Reconstructing Austronesian population history in Island Southeast Asia, Nat. Commun. 5 (2014) 4689, https://doi.org/10.1038/ncomms5689.
[12] Y. Xue, T. Zerjal, W. Bao, S. Zhu, Q. Shu, J. Xu, Du R, S. Fu, P. Li, M.E. Hurles, H. Yang , C. Tyler-Smith, Male demography in East Asia: a north-south contrast in human population expansion times, Genetics 172 (2006) 2431–2439, https://doi.org/10.1534/genetics.105.054270.
[13] H. Zhong, H. Shi, X.-B. Qi, C.-J. Xiao, L. Jin, R.Z. Ma, B. Su, Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia, J. Hum Genet. 55 (2010) 428–435, https://doi.org/10.1038/jhg.2010.40.

[14] Y.-Z. Huang, H. Pamjav, P. Flegontov, V. Stenzl, S.-Q. Wen, X.-Z. Tong, C.-C. Wang, L.-X. Wang, L.-H. Wei, J.-Y. Gao, L. Jin, H. Li, Dispersals of the Siberian Y-chromosome haplogroup Q in Eurasia, Mol. Genet. Genomics 293 (2018) 107–117, https://doi.org/10.1007/s00438-017-1363-8.

[15] H. Li, B. Wen, S.-J. Chen, B. Su, P. Pramoonjago, Y. Liu, S. Pan, Z. Qin, W. Liu, X. Cheng, N. Yang, X. Li, D. Tran, D. Lu, M.-T. Hsu, R. Deka, S. Marzuki, C.-C. Tan, L. Jin, Paternal genetic affinity between Western Austronesians and Daic populations, Bmc. Evol. Biol. 8 (2008) 146, https://doi.org/10.1186/1471-2148-8-146.

[16] G.P. Arunkumar, L.-H. Wei, V.J. Kavitha, A. Syama, V.S. Arun, S. Sathua, R. Sahoo, R. Balakrishnan, T. Riba, J. Chakravarthy, B. Chaudhury, P. Panda, P. K. Das, P. K. Nayak, H. Li, R. Pitchappan, A late Neolithic expansion of Y chromosomal haplogroup O2a1-M95 from east to west, J. Syst. Evol. 53 (2015) 546–560, https://doi.org/10.1111/jse.12147.

[17] S. Yan, C.-C. Wang, H.-X. Zheng, W. Wang, Z.-D. Qin, L.-H. Wei, Y.i. Wang, X.-D. Pan, W.-Q. Fu, Y.-G. He, L.-J. Xiong, W.-F. Jin, S.-L. Li, Y.u. An, H. Li, L. Jin, B. Su, Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers, PLoS ONE 9 (8) (2014) e105691, https://doi.org/10.1371/journal.pone.0105691.

[18] J. A. Trejaut, E. S. Poloni, J.-C. Yen, Y.- H. Lai, J.-H. Loo, C.-L. Lee, C.-L. He, M. Lin, Taiwan Y-chromosomal DNA variation and its relationship with Island Southeast Asia, Bmc. Genet. 15(2014) 77, https://doi.org/10.1186/1471-2156-15-77.

[19] P. Bellwood, The checkered prehistory of rice movement southwards as a domesticated cereal—from the Yangzi to the Equator, Rice 4 (2011) 93–103, https://doi.org/10.1007/s12284-011-9068-9.