

SCIENTIFIC REPORTS



OPEN

Genomic Portrait of Guangdong Liannan Yao Population Based on 15 Autosomal STRs and 19 Y-STRs

Yaoqi Liao¹, Ling Chen², Runze Huang¹, Weibin Wu², Dayu Liu² & Huilin Sun¹

Here we studied the genetic polymorphism and evolutionary differentiation of the Guangdong Liannan Yao population based on 15 autosomal STR loci and 19Y chromosomal STR loci. The blood card DNA of 302 unrelated individuals from the Yao Autonomous County of Liannan was directly amplified using an Expressmarker 16 + 19Y kit and genotyped using a 3500XL Genetic Analyzer. For the autosomal STR loci, the CPD value was over 0.999 999 999 999, while the CPE value was over 0.9999. The population comparison revealed a closer relationship between the Liannan Yao population and the She ethnic population than other reported Chinese populations. For the Y-STRs, a total of 102 unique haplotypes were obtained, 87 of which were observed only once. Both R_{ST} pairwise analysis and a multidimensional scaling plot showed that the Liannan Yao population is closely related to the Fujian She ethnic population and is significantly different from other Chinese ethnic populations. The results show that the 15 autosomal STR and 19Y-STR loci are valuable for forensic applications and human genetic studies in the Liannan Yao population.

The Yao ethnic group in China is primarily distributed across the Guizhou, Hunan, Yunnan, Guangdong, and Guangxi provinces. In 2010, the population of this group consisted of 2, 796, 003 individuals (2010 census, www.stats.gov.cn). According to differences in the language, dress, customs, and living environment of the Yao population, the Yao group can be divided into Pan Yao, Guo-Shan Yao, Gaoshan Yao, Pingdi Yao, Landian Yao, Eight-Row Yao, Chashan Yao, and Baikou Yao¹. The Yao Autonomous County of Liannan, located in Guangdong province, is home to the world's only Eight-Row Yao population². Liannan Yao has a small population of about 80,000. The Yao Autonomous County of Liannan is within northwest Guangdong province. According to folklore and historical records^{3,4}, its inhabitants mainly came from the area of the Xiangjiang River, the middle and lower reaches of the Yuanjiang River, and Dongting Lake in Hunan. During the Sui-Tang Dynasties, their ancestors migrated into the Liannan Mountain district via Chenzhou and Daozhou and settled there. Genetic polymorphisms of Guangdong Liannan Yao have not yet been reported. Guangdong Liannan Yao is a special ethnic population in Guangdong in which the incidence of some diseases such as diabetes differs from those in the Guangdong Han group⁵. Through the present study, we aimed to gather autosomal STR and Y-STR genetic data for the Guangdong Liannan Yao population. Moreover, we hoped to reveal the relationship between the Guangdong Liannan Yao population and other populations.

Autosomal STR and Y-STR have been widely used in forensic evidence examinations, historical investigations, and genealogical research worldwide⁶. Autosomal STR loci have also been used to uncover the population's genetic backdrop and structure⁷. Here we used an Expressmarker 16 + 19Y kit to detect 302 samples collected from unrelated Liannan Yao individuals (122 males and 180 females) and analyze the genetic polymorphism of 15 autosomal STR loci and 19 Y chromosome STR loci as well as assess the efficiency of this kit in the individual identification and paternity testing for the studied population. The Expressmarker 16 + 19Y kit comprises all 13 core Combined DNA Index (CODIS) STR loci (CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11), Amelogenin, two non-CODIS STR loci (D19S433, D2S1338) and 19 widely used Y-STR loci (DYS635, DYS456, DYS385a/b, DYS437, DYS458, DYS389I, DYS392, DYS439, DYS390, DYS393, DYS391, DYS438, DYS448, Y-GATA-H4, DYS19, DYS389II, DYS527a/b), among which, DYS385a/b and DYS527a/b are the double copy loci.

¹Department of Endocrinology, The First Affiliated Hospital of Guangdong Pharmaceutical University, Guangzhou, 510515, China. ²School of Forensic Medicine, Southern Medical University, Guangzhou, 510515, China. Correspondence and requests for materials should be addressed to H.S. (email: sun-hui-lin@126.com)

Materials and Methods

Sample collection. Bloodstains of 302 unrelated Liannan Yao individuals (122 males, 180 females) in Liannan Yao Autonomous County were collected. All participants were interviewed to confirm their ethnic origins and sign the Informed consents. The study was carried out in accordance with the relevant guidelines and regulations and was approved by the Ethics Committee of the First Affiliated Hospital of Guangdong Pharmaceutical University, Medical ethics review [2016] (No. 76).

Polymerase chain reaction amplification and STR typing. Samples from the filter paper were punched using a 1.2-mm BSD600-DUET stiletto instrument (BSD, Australia) and placed into 96-well plates for direct polymerase chain reaction (PCR) amplification using an Expressmarker 16 + 19Y kit (AGCU ScienTech Incorporation, China). The kit's details are shown in Table S1.

PCR amplification was performed using an Applied Biosystems Gene Amp PCR System 9700 thermal cycler. The reaction was performed in a 25 μ L volume system containing 1X Reaction Mix (including 2.5 mM magnesium ion, 0.25 mM dNTPs, 10 mM Tris-HCl, 50 mM KCl, etc.), 1X EX16 + 19Y primer set, 2U hotstart C-Taq DNA polymerase, 1.2 mm filter paper of genomic DNA, and sdH₂O supplying the rest of the volume. The PCR amplification cycling parameters of Expressmarker 16 + 19Y were as follows: 95 °C for 2 min; 15 cycles of 94 °C for 10 s, 60 °C for 40 s, and 72 °C for 1 min and 15 cycles of 90 °C for 10 s, 59 °C for 1 min, and 65 °C for 80 s; final extension at 60 °C for 20 min; and a final hold at 4 °C.

The PCR products were separated on a 3500XL Genetic Analyzer (Life Technologies, USA). Samples were prepared by adding 1 μ L of PCR product (amplicons or allele ladder) to 12 μ L of deionized formamide and 0.25 μ L of marker (SIZ-500 internal lane standard). The mixture was denatured by heating 95 °C for 3 min followed by quick chilling on ice for 3 min. Standard run parameters included sample injection for 28 s at 1.2 kV. The sample DNA was genotyped using GeneMapper-ID-X software version 1.3 (Life Technologies) with the peak amplitude threshold set at 150 relative fluorescent units.

Quality control. The study was conducted following the recommendations of DNA Commission of the International Society for Forensic Genetics (ISFG) as described by Carracedo *et al.*⁸ on DNA polymorphism analysis.

Data analysis. For autosomal STR loci, we used the PowerStatsV12.xls software⁹ (<https://www.promega.com.cn/products/genetic-identity/>) to calculate the allelic frequencies, discrimination power (DP), power of exclusion, polymorphism information content (PIC), and typical paternity index. The p-values of the exact test of Hardy-Weinberg's equilibrium, expected heterozygosity, and observed heterozygosity were calculated using Genepop software¹⁰ (<http://genepop.curtin.edu.au/>). Pairwise genetic distance (F_{ST}) and p-values were calculated for each locus between populations using Arlequin V3.5 software¹¹. Further, a phylogenetic tree and a principal component analysis (PCA) plot showing the interpopulation relationship were constructed using Poptree 2¹² and Past 3.11¹³ software, respectively, according to the allelic frequency data of Liannan Yao population and 8 other reported populations (Guangxi Yao¹⁴, Jiangxi Han¹⁵, Sichuan Han¹⁶, Fujian Han¹⁷, Hunan Han¹⁸, Fujian She¹⁹, Guangdong Han²⁰, and Hebei Han²¹ populations).

For the Y-STR loci, the allelic frequencies of each 19 Y-STR loci were calculated by the direct counting method. Haplotype frequencies were calculated using Arlequin SoftwareV3.5¹¹. The genetic diversity or haplotype diversity was calculated as $GD/HD = n(1 - \sum P_i^2)/(n - 1)$ (P_i indicates the frequency of the i th allele or haplotype; n indicates the number of samples). The discrimination capacity (DC) of the haplotypes was calculated as $DC = m/n$ (m indicates the number of different haplotypes; n indicates the total number of samples). Analysis of molecular variance (AMOVA) and multidimensional scaling (MDS) tests were conducted with the online tools in YHRD (<http://www.yhrd.org>), and our population data were compared to those of other Chinese reported populations including Chinese Han populations (Guangdong Han²², Shandong Han²³, Hunan Han²⁴, Guangxi Han [YA004218], and Henan Han²⁵⁻²⁷), the She ethnic population from Fujian (YA004031), and the Yao ethnic population from Guangxi (YA004221). A neighbor-joining phylogenetic tree based on R_{ST} values was built using MEGA 6.0 software²⁸.

Results and Discussion

Allelic frequencies and forensic parameters of 15 autosomal STR in the Liannan Yao population. The allelic frequencies and forensic parameters of the Liannan Yao population on the basis of the 15 autosomal STR loci are shown in Table S2 and Figure S1. An STR locus can be considered highly polymorphic when its DP value is >0.80 or its power of exclusion value is >0.50^{29,30}. The results of this study showed that most of the examined loci were highly polymorphic. The examined autosomal STRs had a high DP that ranged from 0.7785 (TPOX) to 0.9646 (FGA). The combined discrimination power (CPD) for the 15 loci was 0.999 999 999 999. The power of exclusion ranged from 0.2559 (TPOX) to 0.7293 (D8S1179), while the combined probability of excluding paternity (CPE) for the 15 loci was 0.999 995. A multiple STR amplification system can be considered as reliable for individual identification if its CPD value is over 0.999 999 999 999, and for paternity testing if its CPE value is over 0.9999. On this basis, the Expressmarker 16 + 19Y qualifies for the application of forensic individual testing and paternity testing among the Liannan Yao population.

In addition, all the STR loci were considered highly informative because they had high PIC values ($PIC > 0.5$ is considered highly polymorphic³⁰) that ranged from 0.5300 (TPOX) to 0.8484 (FGA). The most polymorphic and discriminatory locus in the studied population was FGA, with values of 0.8484 (PIC) and 0.9646 (DP). The highest allelic frequency (0.5629) was observed at allele 8 of the TPOX locus. The most common allele and least common allele for each locus are listed in Table S3. No significant deviations from Hardy-Weinberg expectations were observed after Bonferroni correction for any studied locus ($P > 0.05$).

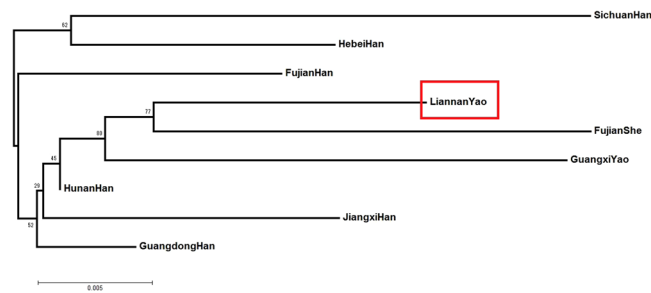


Figure 1. The phylogenetic tree based on allele frequencies of 15 autosomal STR loci for Liannan Yao population and 8 other reported populations (Guangxi Yao, Jiangxi Han, Sichuan Han, Fujian Han, Hunan Han, Fujian She, Guangdong Han, and Hebei Han populations).

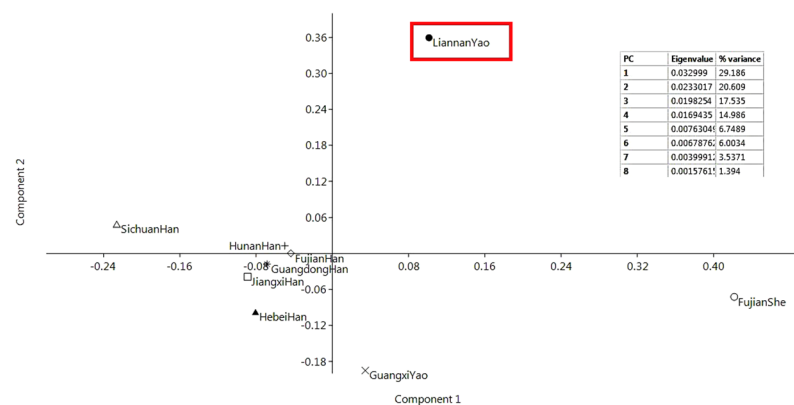


Figure 2. Principal component analysis plot based on allele frequencies of 15 autosomal STR loci for the Liannan Yao population and 8 other reported populations (Guangxi Yao, Jiangxi Han, Sichuan Han, Fujian Han, Hunan Han, Fujian She, Guangdong Han, and Hebei Han populations).

Interpopulation comparisons based on the 15 autosomal STRs. For autosomal STR loci, the F_{ST} distance was calculated based on the allelic frequencies and was used to compare the studied population and 8 other reported populations (Table S4). The geographical locations of the reference populations were shown in Figure S3.

Statistically significant differences ($p < 0.05/15$) were found between the Liannan Yao and Guangxi Yao populations at two STR loci, the Fujian She population at five loci, the Han population from Sichuan and Hunan at six STR loci, the Fujian Han and Jiangxi Han populations at seven STR loci, and the Guangdong Han and Hebei Han populations at eight STR loci. The genetic portrait of the Liannan Yao population is closer to that of the Guangxi Yao and Fujian She populations.

Phylogenetic analyses based on the 15 autosomal STRs. The neighbor-joining phylogenetic tree (Fig. 1) mirrors the historical and geographical backgrounds of the studied population. The phylogenetic tree showed that the Liannan Yao and Fujian She populations were clustered in one branch, while the other Chinese populations were clustered in another branch. The Liannan Yao population in this study was closest to the Fujian She ethnic group, followed by the Guangxi Yao and Hunan Han populations.

Historically, the Yao and She ethnic groups originated from the same ancient ethnic group. These ancient peoples migrated from Xiangnan district, currently known as Hunan province. Some of them moved eastward to the junction of Jiangxi, Guangdong, and Fujian provinces and evolved into the She ethnic groups, whereas the others moved into northern Guangxi and Guangdong and evolved into the Yao ethnic groups. Later, the She ethnic groups extended to eastern Fujian and southern Zhejiang, while the Yao people extended to the west and south of Guangxi and Guangdong. Eventually, the overall distribution pattern of the present Yao and She ethnic groups was formed. Our study results according to the phylogenetic analyses based on the 15 autosomal STRs support the migration history of the two populations. Moreover, these results supported that the Fujian She population and Guangdong Liannan Yao population have a remote geographic distance but a limited genetic distance.

PCA based on the 15 autosomal STRs. A PCA plot was drawn based on the allelic frequencies of 15 STRs (Fig. 2). The Liannan Yao, Guangxi Yao, and Fujian She populations were clustered in the right side; the Sichuan Han population was clustered in the left upper quadrant; and the Chinese Han populations from Hebei, Hunan, Guangdong, Jiangxi and Fujian province were clustered lower near to the center.

As described above, the Yao and She ethnic groups share the same origin. The genetic distance between the Liannan Yao and Fujian She populations is the closest, followed by the Guangxi Yao and Hunan Han populations.

Groups	Liannan Yao	Guangdong Han	Guangxi Han	Henan Han	Hunan Han	Shandong Han	Fujian She	Guangxi Yao
Liannan Yao	—	0.0000	0.0000	0.0000	0.0000	0.0000	0.0020	0.0000
Guangdong Han	0.0882	—	0.1480	0.0000	0.0542	0.0000	0.0004	0.0000
Guangxi Han	0.1077	0.0005	—	0.0000	0.0606	0.0000	0.0000	0.0000
Henan Han	0.1078	0.0117	0.0108	—	0.0000	0.0003	0.0001	0.0000
Hunan Han	0.0997	0.0018	0.0015	0.0078	—	0.0001	0.0009	0.0000
Shandong Han	0.0867	0.0124	0.0144	0.0034	0.0086	—	0.0008	0.0000
Fujian She	0.0363	0.0307	0.0406	0.0326	0.0292	0.0238	—	0.0022
Guangxi Yao	0.0812	0.0429	0.0502	0.0426	0.0418	0.0363	0.0352	—

Table 1. Population pairwise genetic distances (R_{ST}) and associated probability values (p-values) based on 19 Y-STRs between the Liannan Yao population and 7 other populations (Guangdong Han, Shandong Han, Hunan Han, Guangxi Han, Henan Han, Fujian She, and Guangxi Yao). p-values are shown above the diagonal line and R_{ST} values are shown below it.

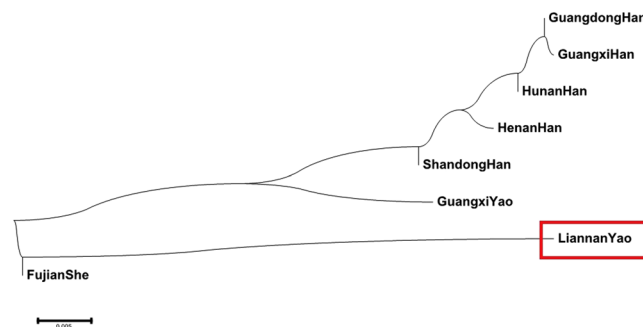


Figure 3. A Neighbor-Joining tree based on 19 Y-STRs showing the relationship between Liannan Yao group and other 7 populations (Guangdong Han, Shandong Han, Hunan Han, Guangxi Han, Henan Han, Fujian She, and Guangxi Yao populations).

Haplotypic structure of 19 Y-STR loci in the Liannan Yao population. The data were submitted to the Y-STR Haplotype Reference Database (YHRD) for checking as Liannan China [Yao], and can be retrieved using the accession number YA004311 (<http://www.yhrd.org>). The 122 unrelated individuals were genotyped with 19 Y-STR loci, and allelic frequencies of each loci are presented in Table S5. 30 haplotypes were found for DYS385a/b, while 19 were observed for DYS527a/b. A total of 118 alleles were observed, and the corresponding allelic frequencies ranged from 0.0041 (DYS458) to 0.8770 (DYS391). The most common allele and least common allele for each locus are listed in the Table S3, as autosomal STRs before. Similarly, a total of 102 Y-STR haplotypes (Table S6) were observed, of which 87 were unique and 15 were observed for more than two individuals. The genetic diversity (Figure S2) values of the 19 loci ranged from 0.2211 (DYS391) to 0.8862 (DYS385a/b). The haplotype diversity and DC was 0.9891 and 0.8197, respectively.

Inter-population comparisons based on the 15 shared Y-STRs. For the Y-STR loci, the studied data were compared to those of other Chinese reported populations. The geographical locations of the reference populations were shown in Figure S3. The R_{ST} values and corresponding p-values were computed using AMOVA (Table 1). The results indicate that after applying Bonferroni's correction, R_{ST} values did not differ significantly among the Liannan Yao and Fujian She population ($R_{ST} = 0.0363$; $p > 0.0018$, 28 pairs), but they differed significantly for following populations: Guangdong Han, Guangxi Han, Henan Han, Hunan Han, Shandong Han, and Guangxi Yao ($p < 0.0018$).

Phylogenetic analyses based on the 15 shared Y-STRs. In the neighbor-joining tree, the Liannan Yao population first clustered with the Fujian She ethnic population, second with the Guangxi Yao ethnic population, and third with the Guangdong Han and Hunan Han populations (Fig. 3). As shown in the MDS plot, there were clear differences between the Liannan Yao and 7 other populations (Fig. 4). Moreover, compared with other Chinese Han populations, the Liannan Yao population was closer to the Fujian She population, followed by the Guangxi Yao population. This shows Liannan Yao population has a close genetic distance with the She and other Yao ethnic populations, which support the historian's perspective that the Liannan Yao and Fujian She ethnic populations share the same origin.

There was one potential limitation in the present study that the included groups for population comparison between autosomal STR and Y-STR are different, due to the limited available relevant data. However, these analyses all indicate that Liannan Yao population has a close relationship with Fujian She and Guangdong Yao population, also support the migration history of the two populations that Yao and She have the same origin. Moreover,

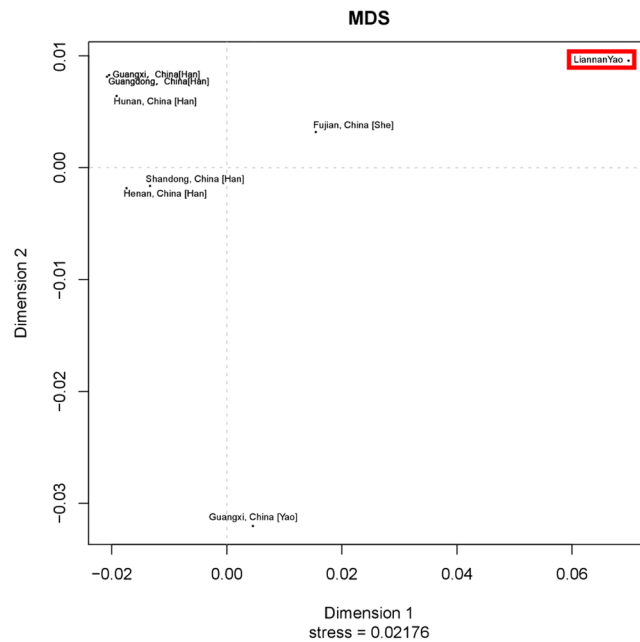


Figure 4. A MDS plot based on 19 Y-STRs between Liannan Yao population and other 7 populations (Guangdong Han, Shandong Han, Hunan Han, Guangxi Han, Henan Han, Fujian She, and Guangxi Yao).

there are clear differences between the Liannan Yao and other Chinese Han population but Guangdong Han population. This result might indicate that the Liannan Yao integrated gradually with natives, such as Guangdong Han population, following its geographical migration, which is also corresponded with the historical records^{3,4}.

Conclusion

These results demonstrate that the Liannan Yao population is an independent exclusive ethnicity and has some unique genetic characteristics. 15 autosomal STR and 19 Y-STR loci for Liannan Yao population are informative, and Expressmarker 16 + 19Y kit is suitable for personal identification and paternity testing within this population. These data could be helpful for inferring the genetic genealogy evolution and ancient human migration patterns of the Liannan Yao population.

References

- Xu, Z. P. & Pai, Y. Row, Eight Rows and Yao Discrimination Title. *Study of Ethnicity in Guangxi*, <https://doi.org/10.3969/j.issn.1004-454X.2011.01.024> (2011).
- Zheng, S. A Study of Yao's Architecture in Liannan: from a Humanity Perspective. *Journal of Huizhou University* **38**, 90–93, <https://doi.org/10.16778/j.cnki.1671-5934.2018.03.017> (2018).
- Chen, Hui. & Ma, J. The history and development of Liannan Yao embroidery. *Journal Of Guangdong Polytechnic Normal University* **32**, 23–26, <https://doi.org/10.3969/j.issn.1672-402X.2011.02.006> (2011).
- Xie, J. The history of Liannan Pai Yao population. *Journal of guangdong institute for nationalities*, 27–37, <https://doi.org/10.13408/j.cnki.gjxb.1990.03.005> (1990).
- Liu, Z. Z. *et al.* Relationship between Serum AHSG Level and Pregnancy Outcome in Liannan Yao Autonomous County Pregnant Women. *Journal of Sun Yat-sen University(Medical Sciences)* (2013).
- Larmuseau, M. H. *et al.* High Y-chromosomal diversity and low relatedness between paternal lineages on a communal scale in the Western European Low Countries during the surname establishment. *Heredity* **115**, 3–12, <https://doi.org/10.1038/hdy.2015.5> (2015).
- Black, M. L., Wise, C. A., Wang, W. & Bittles, A. H. Combining genetics and population history in the study of ethnic diversity in the People's Republic of China. *Human biology* **78**, 277–293, <https://doi.org/10.1353/hub.2006.0041> (2006).
- Carracedo, A. *et al.* Update of the guidelines for the publication of genetic population data. *Forensic science international. Genetics* **10**, A1–2, <https://doi.org/10.1016/j.fsigen.2014.01.004> (2014).
- Tereba, A. *Tools for Analysis of Population Statistics*. vol. 2 (1999).
- Rousset, F. Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular ecology resources* **8**, 103–106, <https://doi.org/10.1111/j.1471-8286.2007.01931.x> (2008).
- Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources* **10**, 564–567, <https://doi.org/10.1111/j.1755-0998.2010.02847.x> (2010).
- Takezaki, N., Nei, M. & Tamura, K. POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. *Mol Biol Evol* **27**, 747–752, <https://doi.org/10.1093/molbev/msp312> (2010).
- Hammer, O., Harper, D. & Ryan, P. *PAST: Paleontological Statistics Software Package for Education and Data Analysis*. vol. 4 (2001).
- Xu, Y. *et al.* Genetic Polymorphism of 18 STR Loci in Guangxi Yao and Miao Population. *Forensic Science and Technology* **41**, 74–76, <https://doi.org/10.16467/j.1008-3650.2016.01.016> (2016).
- Huang, Y., Xie, B., Chen, H. & Tang, J. Genetic polymorphism of 19 STR loci in Han population in Ganzhou, Jiangxi. *Chinese Journal of Forensic Medicine* **30**, 293–294, <https://doi.org/10.13618/j.issn.1001-5728.2015.03.019> (2015).
- He, G. *et al.* Genetic diversity of 21 autosomal STR loci in the Han population from Sichuan province, Southwest China. *Forensic science international. Genetics*. <https://doi.org/10.1016/j.fsigen.2017.07.006> (2017).

17. Liu, J., Lin, T. & Liu, J. Genetic polymorphism of 15 STR loci in Han Chinese population in Fujian. *Chinese Journal of Forensic Medicine* **29**, 479–480, <https://doi.org/10.13618/j.issn.1001-5728.2014.05.024> (2014).
18. Chang, Y. F. *et al.* Genetic polymorphism of 17 STR loci in Chinese population from Hunan province in Central South China. *Forensic science international. Genetics* **6**, e151–153, <https://doi.org/10.1016/j.fsigen.2012.02.011> (2012).
19. Yuan, L. *et al.* Population genetics analysis of 38 STR loci in the She population from Fujian Province of China. *Legal medicine* **16**, 314–318, <https://doi.org/10.1016/j.legalmed.2014.05.008> (2014).
20. Yang, L. *et al.* Population data of 23 autosomal STR loci in the Chinese Han population from Guangdong Province in southern China. *International journal of legal medicine* **132**, 133–135, <https://doi.org/10.1007/s00414-017-1588-4> (2017).
21. Zhang, Y., Li, A., Zhao, L., Liu, J. & Song, J. Genetic polymorphism of 19 STR loci in Hebei Han population. *Forensic Science And Technology*, 48–50, <https://doi.org/10.3969/j.issn.1008-3650.2011.02.023> (2011).
22. Wang, Y. *et al.* Genetic polymorphisms and mutation rates of 27 Y-chromosomal STRs in a Han population from Guangdong Province, Southern China. *Forensic science international. Genetics* **21**, 5–9, <https://doi.org/10.1016/j.fsigen.2015.09.013> (2016).
23. Xu, J. *et al.* Genetic analysis of 17 Y-STR loci in Han population from Shandong Province in East China. *Forensic science international. Genetics* **22**, e15–e17, <https://doi.org/10.1016/j.fsigen.2016.01.016> (2016).
24. Jiang, W. *et al.* Population genetics of 26 Y-STR loci for the Han ethnic in Hunan province, China. *International journal of legal medicine* **131**, 115–117, <https://doi.org/10.1007/s00414-016-1411-7> (2017).
25. Wang, L. *et al.* Genetic population data of Yfiler Plus kit from 1434 unrelated Hans in Henan Province (Central China). *Forensic science international. Genetics* **22**, e25–e27, <https://doi.org/10.1016/j.fsigen.2016.02.009> (2016).
26. Shi, M. *et al.* Analysis of 24 Y chromosomal STR haplotypes in a Chinese Han population sample from Henan Province, Central China. *Forensic science international. Genetics* **17**, 83–86, <https://doi.org/10.1016/j.fsigen.2015.04.001> (2015).
27. Bai, R. *et al.* Analysis of 27 Y-chromosomal STR haplotypes in a Han population of Henan province, Central China. *International journal of legal medicine* **130**, 1191–1194, <https://doi.org/10.1007/s00414-016-1326-3> (2016).
28. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30**, 2725–2729, <https://doi.org/10.1093/molbev/mst197> (2013).
29. Shriver, M. D. *et al.* A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Mol Biol Evol* **12**, 914–920 (1995).
30. KuangSq, Z. Y. & ChenZ. A powerful tool for genome scanning and gene mapping of inherited disease. *Chinese Journal of Medical Genetics*, 41–45 (1997).

Acknowledgements

The authors are very grateful to all sample donors for their contributions to this work and all those who helped with sample collection. This work was supported by a grant from Public Welfare Research and Capacity Building Project of Guangdong Province (2014A020212461), Natural Science Foundation of Guangdong Province (Grant no. 2014A030310025 and Grant no. 2017ZC0203) and Natural Science Foundation of China (Grant no. 81501627).

Author Contributions

Ling Chen wrote the main manuscript text and conducted the experiment, Yaoqi Liao, Runze Huang, Weibin Wu, Dayu Liu and Huilin Sun did the data processing and the manuscript modification. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-36262-x>.

Competing Interests: No conflict of interest exists in the submission of this manuscript, and manuscript is approved by all authors for publication. I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part. All the authors listed have approved the manuscript that is enclosed.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019