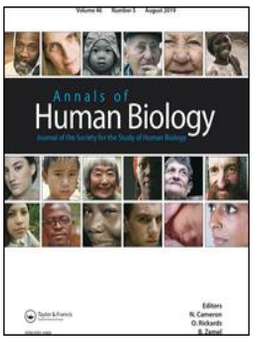

Annals of Human Biology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/iahb20

Forensic features and genetic structure of 20 autosomal STR loci in the Han population of Ningde City, Southeastern China

Hongbo Wang , Cairui Xin & Hongyi Cao

To cite this article: Hongbo Wang , Cairui Xin & Hongyi Cao (2021): Forensic features and genetic structure of 20 autosomal STR loci in the Han population of Ningde City, Southeastern China, Annals of Human Biology, DOI: 10.1080/03014460.2020.1854344

To link to this article: https://doi.org/10.1080/03014460.2020.1854344

ANNALS OF HUMAN BIOLOGY
https://doi.org/10.1080/03014460.2020.1854344

RESEARCH PAPER

Forensic features and genetic structure of 20 autosomal STR loci in the Han population of Ningde City, Southeastern China

Hongbo Wang[a], Cairui Xin[b] and Hongyi Cao[c]

[a]Department of Anatomy, Shenyang Medical College, Shenyang, China; [b]Tibetan Police Academy, Lhasa, China; [c]Department of Pathology, College of Basic Medical Science, China Medical University, Shenyang, China

ABSTRACT

Background: Short tandem repeat (STR) loci are widely used in forensic medicine and population genetics.

Aim: To profile 20 autosomal STR loci using the SureID® 21 G Human STR Identification Kit.

Subjects and methods: Samples were obtained from 1412 unrelated Chinese Han individuals from Ningde City, Southeastern China and 20 autosomal STR loci were profiled using the SureID® 21 G Human STR Identification Kit.

Results: A total of 261 alleles were observed among 1412 unrelated individuals and the corresponding allelic frequencies ranged from 0.5464 to 0.0004. The combined power of discrimination and exclusion of the 20 autosomal STR loci were 0.99999999999999999922 and 0.999999340285752, respectively. There was no significant deviation from Hardy–Weinberg equilibrium (HWE) and minimal departure from linkage equilibrium (LE) for two pairwise combinations of loci after sequential Bonferroni correction. In the population comparison, phylogenetic analysis was performed between the Han population and other relevant populations on the basis of the shared autosomal STR genotyping. Moreover, the neighbor-joining tree and principal component analysis were analysed based on the Nei's standard genetic distance.

Conclusion: The population comparison revealed that the structure of the Ningde Han population was similar to the structure of southern Han populations in China and was significantly different to the other Chinese ethnic groups, such as Kyrgyz, Uzbek, Kazakh, Uyghur, Manchu from Xinjiang and Mongols.

ARTICLE HISTORY
Received 5 June 2020
Revised 20 October 2020
Accepted 28 October 2020

KEYWORDS
STR genotyping; genetic polymorphism; forensic medicine; SureID® 21G

Introduction

Short tandem repeats (STRs) are widely used in forensic medicine and population genetics as genetic markers. They can provide more information when compared with single nucleotide polymorphisms (SNPs) (Zhu et al. 2015; Yao and Wang 2016). Due to their high diversity, they can play an important role in ascertaining population structure and tracking "human population biodiversity" (Chen et al. 2017; Guo 2017; He et al. 2018). Moreover, STR loci can be used to investigate the genetic background of populations (Li et al. 2006; Eaaswarkhanth et al. 2009; EL Ossmani et al. 2010; Lowery et al. 2011; Zhang et al. 2015). In forensic medicine, STRs show powerful discrimination from individual to individual so their application includes personal identification and paternity testing cases (Adnan et al. 2016; 2018; Zhan et al. 2018).

Ningde City, part of Fujian Province, is located in Southeast China. The land area is 13,400 square kilometres and the directly adjacent sea area is 44,600 km^2. It is bordered by East China Sea in the east, Nanping in the west, Zhejiang Province in the north, and Fuzhou City in the south. To date, the genetic polymorphisms of the autosomal STR loci have been reported in the Han populations of Putian City, Zhangzhou City, and Xiamen City, which are close to Ningde City (Lu et al. 2017; Wu et al. 2017; Li et al. 2019; Wu et al. 2020). In the current study, in order to understand the genetics and structural background of the Ningde Han population, we compared our population with other reference populations from across the globe. Population comparisons including Nei's genetic distance, neighbor-joining tree, and principal component analysis (PCA) were carried out between the Ningde Han population and different ethnic groups to better understand the genetic background and structure of the Ningde Han population.

Subjects and methods

Study population

Ningde had a population of over 2.90 million in 2017 (www.stats.gov.cn). Han is the most dominant ethnic group of Ningde city followed by other minority groups such as She, Hui, and Zhuang.

CONTACT Hongbo Wang 18640249580@163.com Department of Pathology, College of Basic Medical Science, China Medical University, No.146, Huanghe North Street, Shenyang, Liaoning 110034, China.

Supplemental data for this article is available online at https://doi.org/10.1080/03014460.2020.1854344.

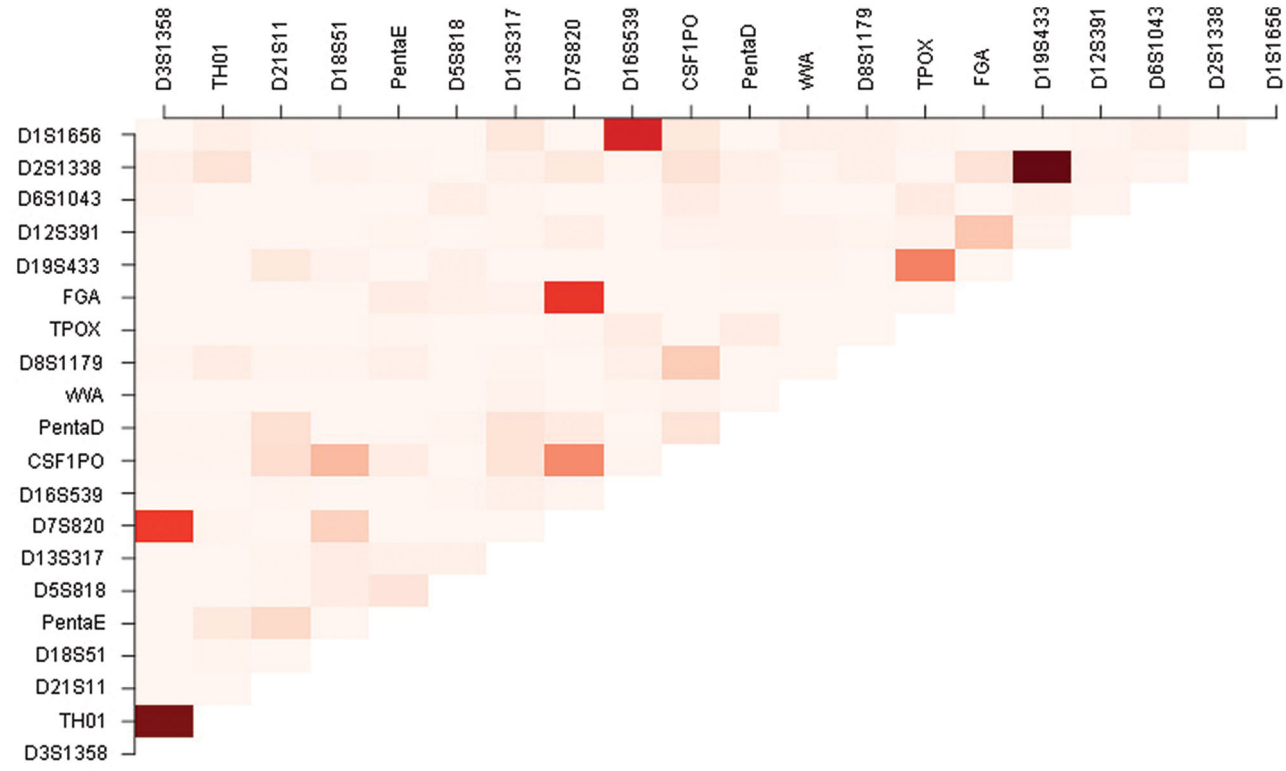

Figure 1. Pairwise LD *p* values matrix for 19 autosomal STR loci for Ningde Han population.

Saliva swabs were collected from 1412 unrelated healthy individuals living in Ningde City, Fujian Province, Southeastern China (Figure 1). All participants gave their informed consent either orally and with thumb prints (in case they could not write) or in writing after the study aims and procedures were carefully explained to them. The study was approved by the ethical review board of Shenyang Medical College, Shenyang, Liaoning Province, People's Republic of China ([2016]063) and in accordance with the standards of the Declaration of Helsinki.

DNA extraction, PCR amplification, and genotyping

Genomic DNA was extracted from saliva swabs by using the Chelex-100 method (Walsh et al. 1991). We used the SureID® 21 G Human STR Identification Kit to explore the genetic characteristics of 1412 Han individuals from Ningde City, Southeastern China (Lu et al. 2017). A total of 20 autosomal STR loci (D3S1358, TH01, D21S11, D18S51, Penta E, D5S818, D13S317, D7S820, D16S539, CSF1PO, Penta D, vWA, D8S1179, TPOX, FGA, D19S433, D12S391, D6S1043, D2S1338, and D1S1656) were amplified simultaneously by using the SureID® 21 G Human STR Identification Kit (HEALTH Gene Technologies, Ningbo, China) on a ProFlexTM PCR system (Thermo Fisher Scientific, USA) following the manufacturer's protocol. Internal controls (H_2O as a negative and 9947 A DNA as a positive control) were genotyped along with each batch of samples to ensure that the results were reproducible and accurate. Detection and separation of amplified PCR products were carried out using Applied Biosystems 3500 Genetic Analyser (Thermo Fisher Scientific, USA) with POP-4 polymer (Life Technologies, USA). Allele identification was conducted using the SureID® 21 G Human STR Identification Kit (HEALTH Gene Technologies, Ningbo, China) panels, bin sets, stutter files and a 175 relative fluorescence units (RFU) threshold, unless otherwise stated, and the results were compared with the allele ladder provided by the corresponding kit via Applied Biosystems GeneMapper ID-X version 1.2 software.

Data analysis

The allele frequencies of samples, exact tests of Hardy-Weinberg equilibrium (HWE), and pair linkage disequilibrium tests were calculated with PowerMarker V3.25 (Liu and Muse 2005). The values for matching probability (MP), power of discrimination (PD), polymorphism information content (PIC), power of exclusion (PE), typical paternity index (TPI), gene diversity (GD), and heterozygosity (He) were calculated using the PowerStats software (Ver 1.2, Promega, Madison, WI, USA) (Tereba 1999). Nei's standard genetic distance between populations was computed using Phylip 3.69 package (Felsenstein 2009).

Principal component analysis (PCA) was conducted based on allele frequency correlation among 52 populations using the MVSP v3.22 software. Additionally, multidimensional scaling plots (MDS) were carried out on the basis of the Nei or Fst genetic distance matrix using the SPSS software (IBM SPSS, version 19.0, Chicago). The phylogenetic tree was constructed using Mega X software and ancestry component

dissection was explored using the STRUCTURE v.2.3.4 software. The model-based analysis employed the length of burn-in period of 100,000 and Markov Chain Monte Carlo (MCMC) step of 100,000 under the 'independent allele frequencies' and 'LOCPRIOR' models with k values ranging from 2 to 10.

Results and discussions

Allele frequencies and forensic parameters of the 20 autosomal STR loci

The allele frequencies for the 20 studied STR loci are summarised in Supplementary Table S1. A total of 261 alleles were observed with the allele frequencies ranging from 0.5464 (TPOX) to 0.0004 (D3S1358, D21S11, D18S51, Penta E, D7S820, D16S539, vWA, TPOX, FGA, D19S433, D12S391, D6S1043, and D1S1656). Forensic efficiency and statistical parameters across the 20 STR loci are shown in Table 1. PIC represents the possibility that a certain allelic mark of an offspring comes from the same allelic mark of its father/mother. A high degree of genetic variation was observed by all STR loci in the Ningde Han population. The values of MP, PD and PE ranged from 0.2145 (TPOX) to 0.0133 (Penta E), 0.9867 (Penta E) to 0.7855 (TPOX), and 0.8102 (Penta E) to 0.2984 (TPOX), respectively. GD can represent the degree of the allelic variation and He can measure the degree of genetic variation within a population (Kumar et al. 2020). The values of GD and He spanned from 0.9153 (Penta E) to 0.6053 (TPOX) and from 0.9072 (Penta E) to 0.6062 (TPOX), respectively. PIC can be used to estimate the polymorphism of the gene as an indicator of the degree of DNA variation (Seyoum et al. 2018). Except for D3S1358 (0.6638), TH01 (0.6191), and TPOX (0.5450), all STR loci were highly polymorphic (PIC $>$ 0.7), with the Penta E locus having the highest degree of polymorphism (0.9092). The combined powers of PD and PE were 0.99999999999999999922 and 0.999999340285752, respectively.

Hardy-Weinberg equilibrium (HWE)

Initially eighteen loci were in Hardy-Weinberg Equilibrium (HWE) and D16S539 and D3S1358 were out of HWE ($p < 0.05$). However, when a sequential Bonferroni's correction was applied, all 20 loci were in Hardly-Weinberg equilibrium (Supplementary Table S2).

Linkage equilibrium (LE)

Linkage disequilibrium (LD) indicates the association between qualitative random variables corresponding to alleles at different STRs. Measuring the levels of linkage disequilibrium is important for gene mapping and it helps in the understanding of genome structure. Exact tests for linkage equilibrium (LE) between 190 pairs showed that the values of only 21 pairs were below 0.05 and thus displaying LD (Figure 1). When sequential Bonferroni's correction was applied, two pairs were still displaying LD (TH01/D3S1358 and D2S1338/D19S433) (Supplementary Table S3). It may be due to the population substructure or natural selection that the paired loci are in different chromosomes (Yuan et al. 2014). Application of the "product rule" for calculation of random match probabilities across multiple loci is fully justified in the Ningde Han population. To check the existence of population hierarchy in Ningde Han ethnic group and 11 other Asian populations (Kumul Yughurs, Xianjiang Yughurs, Liangshan Tibetan, Chengdu Tibetan, Tibetan, Liangshan Yi, Chengdu Han, Wuzhong Hui, Hainan Han, Artux Uyghur and Pakistani Hazara) with raw genotypic data of the STR loci, we checked for genetic homozygosity or heterozygosity via principal component analysis (PCA). A total of 1.94% genetic variations can be extracted by the first three PCs, which indicates that there is minor genetic homology between Ningde Han and 11 other populations from China (Figure 2). The above mentioned PCA results were later confirmed with pairwise *Fst* genetic distances (Supplementary Table S4) and phylogenetic relationship reconstruction (Figure 3). The

Table 1. Forensic efficiency and statistical parameters on 20 autosomal STR loci in the Ningde Han population ($n = 1412$).

Locus	GD	PIC	PM	PD	Hobs	PE	TPI
CSF1PO	0.7429	0.7010	0.1107	0.8893	0.7465	0.5037	1.9721
D12S391	0.8496	0.8314	0.0404	0.9596	0.8470	0.6890	3.2685
D13S317	0.7992	0.7695	0.0708	0.9292	0.7967	0.5930	2.4599
D16S539	0.7842	0.7500	0.0801	0.9199	0.7826	0.5671	2.2997
D18S51	0.8632	0.8484	0.0332	0.9668	0.8385	0.6724	3.0965
D19S433	0.8071	0.7829	0.0607	0.9393	0.8123	0.6221	2.6642
D1S1656	0.8262	0.8066	0.0493	0.9507	0.8237	0.6437	2.8353
D21S11	0.8215	0.7997	0.0527	0.9473	0.8081	0.6141	2.6052
D2S1338	0.8641	0.8490	0.0340	0.9660	0.8711	0.7369	3.8791
D3S1358	0.7158	0.6638	0.1378	0.8622	0.7309	0.4776	1.8579
D5S818	0.7744	0.7408	0.0853	0.9147	0.7875	0.5761	2.3533
D6S1043	0.8754	0.8620	0.0284	0.9716	0.8661	0.7269	3.7354
D7S820	0.7566	0.7199	0.0936	0.9064	0.7500	0.5098	2.0000
D8S1179	0.8490	0.8300	0.0420	0.9580	0.8513	0.6974	3.3619
FGA	0.8580	0.8423	0.0372	0.9628	0.8676	0.7298	3.7754
PentaD	0.8013	0.7777	0.0638	0.9362	0.8102	0.6181	2.6343
PentaE	0.9156	0.9092	0.0133	0.9867	0.9072	0.8102	5.3893
TH01	0.6638	0.6191	0.1605	0.8395	0.6707	0.3844	1.5183
TPOX	0.6055	0.5450	0.2145	0.7855	0.6062	0.2984	1.2698
vWA	0.7987	0.7680	0.0707	0.9293	0.7882	0.5774	2.3612

GD: gene diversity; PIC: polymorphism information content; Hobs: observed heterozygosity; PD: power of discrimination; PM: matching probability; PE: power of exclusion; TPI: typical paternity index.

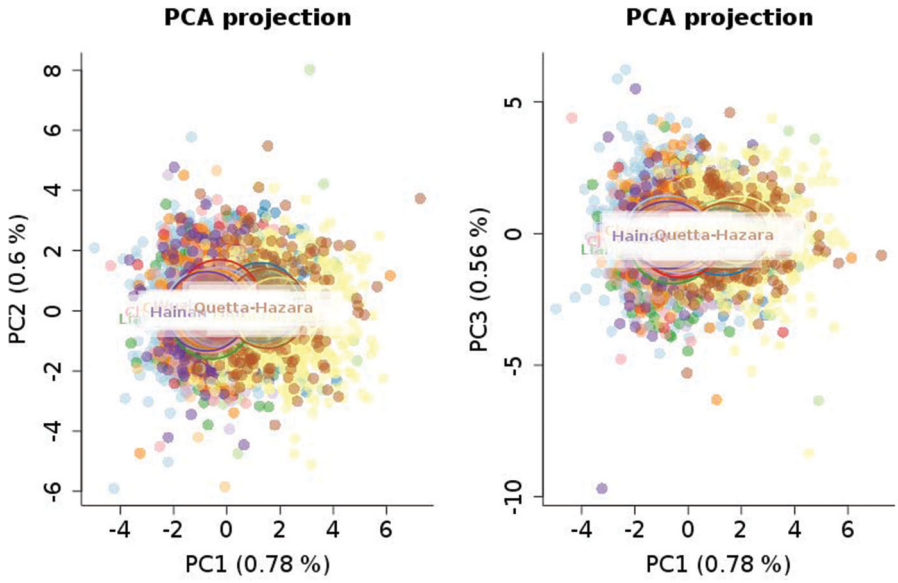

Figure 2. Genetic homology between Ningde Han and 11 other population from China revealed by principle component analysis.

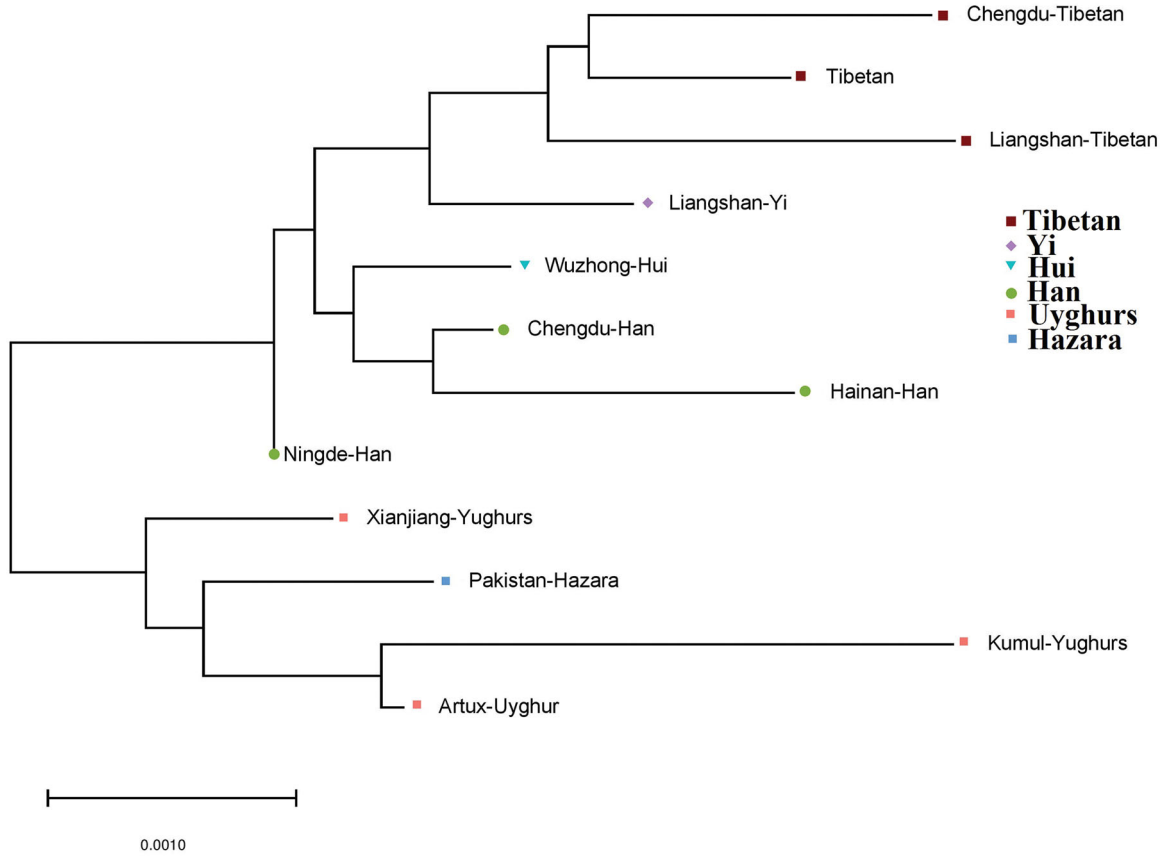

Figure 3. Genetic homology between Ningde Han and 11 other populations revealed by NJ phylogenetic tree based on Fst values.

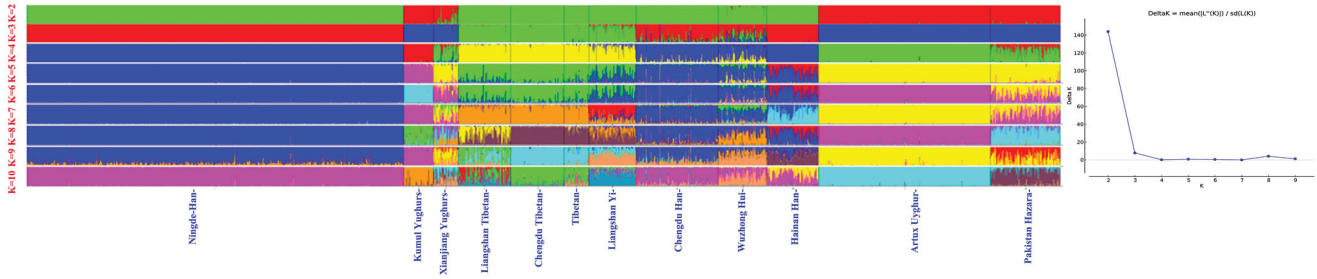

Figure 4. Estimated population genetic structure of 12 different populations at K = 2 to K = 10.

results showed that the Ningde Han population has unique genetic characteristics that are different to the other compared groups.

Cluster analysis with structure

The genetic landscapes of the Ningde Han were further dissected by employing the model-based clustering algorithm in STRUCTURE software in the context of the genetic variations from Kumul Yughurs, Xianjiang Yughurs, Liangshan Tibetan, Chengdu Tibetan, Tibetan, Liangshan Yi, Chengdu Han, Wuzhong Hui, Hainan Han, Artux Uyghur and Pakistani Hazara. As shown in Figure 4, we identified the best optimal predefined populations in five ($K = 5$). Ningde Han shared most of the genetic components with Chengdu Han and Hainan Han (blue component) and shared few genetic alleles or genetic drift with Wuzhong Hui and Liangshan Yi. In total, four genetic clusters were observed: Han, Tibetans, Kumul Yughurs, while Artux Uyghurs, Xianjiang Uyghurs and Hazara's shared the same genetic cluster. Our STRUCTURE results demonstrate that the Ningde Han are genetically closer with Han populations than other East Asian groups.

Clearly, our results show that the genetic structure of the Ningde Han population is similar with that of the Xiamen Han population. Furthermore, this phenomenon is also reflected by the phylogenetic trees. Additionally, the structure of the Ningde Han is also similar to that of the Southern Han population in China, which is different from that of the Han in central and northern regions and the other ethnic populations. Therefore, the Ningde Han population has its own unique Southern Han genetic characteristics.

Conclusions

In the current study, we genotyped 1412 Ningde Han individuals for 20 autosomal STR loci. The genetic variation in the Ningde Han population and its comparison to other relevant groups were analysed using different statistical tests. Overall, our study demonstrated that the SureID® 21 G Human STR Identification Kit could be used in the Ningde Han population for forensic analysis. The acquired genetic data of Ningde Han population can provide useful information for differentiation studies. In addition, the population comparison analysis showed that Ningde Han population has a closer genetic relationship with Han populations from other regions in China.

Acknowledgements

We thank all volunteers who participated in this project.

Ethical approval

All participants gave their informed consent in writing after the study aims and procedures were carefully explained to them in their own language. The study was approved by the ethical review board of the Shenyang Medical College, Shenyang Liaoning Province, People's Republic of China, and was in accordance with the principles of the Declaration of Helsinki.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Author contributions

Hongyi Cao and Hongbo Wang developed the idea. Hongbo Wang analysed the results and wrote the manuscript. Hongyi Cao and Cairui Xin conducted the experiment. All authors reviewed the manuscript.

ORCID

Hongbo Wang http://orcid.org/0000-0002-5030-5529

References

Adnan A, Rakha A, Noor A, VAN Oven M, Ralf A, Kayser M. 2018. Population data of 17 Y-STRs (Yfiler) from Punjabis and Kashmiris of Pakistan. Int J Legal Med. 132(1):137–138.

Adnan A, Ralf A, Rakha A, Kousouri N, Kayser M. 2016. Improving empirical evidence on differentiating closely related men with RM Y-STRs: a comprehensive pedigree study from Pakistan. Forensic Sci Int Genet. 25:45–51.

Chen L, Duan L, Yuan L, Shen Z, He W, Zhai D, Huang Y, Xu B. 2017. Genetic polymorphisms of 19 autosomal STR loci in the China Burmese immigrants. Forensic Sci Int Genet. 31:e46–e47.

Eaaswarkhanth M, Dubey B, Meganathan PR, Ravesh Z, Khan FA, Singh L, Thangaraj K, Haque I. 2009. Diverse genetic origin of Indian Muslims: evidence from autosomal STR loci. J Hum Genet. 54(6): 340–348.

EL Ossmani H, Bouchrif B, Aboukhalid R, Bouabdillah M, Gazzaz B, Zaoui D, Chafik A, Talbi J. 2010. Assessment of phylogenetic structure of Berber-speaking population of Azrou using 15 STRs of Identifiler kit. Leg Med (Tokyo). 12(1):52–56.

Felsenstein J. 2009. PHYLIP (Phylogeny Inference Package) Version 3.69. Seattle (WA): Department of Genome Sciences, University of Washington.

Guo F. 2017. Genetic polymorphism of 17 autosomal STR loci in the Lahu ethnic minority from Yunnan Province, Southwest China. Forensic Sci Int Genet. 31:e52–e53.

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. Int J Legal Med. 132(4): 1079–1082.

Kumar D, Chhokar V, Sheoran S, Singh R, Sharma P, Jaiswal S, Iquebal MA, et al. 2020. Characterization of genetic diversity and population structure in wheat using array based SNP markers. Mol Biol Rep. 47(1):293–306.

Li XN, Al-Sabri A, Ameen F, Al-Dayan N, Adnan A, Yao J. 2019. Genetic polymorphisms of 20 autosomal STR loci in the Han population of Putian City, Southeastern China. Ann Hum Biol. 46(6):509–513.

Liu K, Muse SV. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. 21(9):2128–2129.

Li SL, Yamamoto T, Yoshimoto T, Uchihi R, Mizutani M, Kurimoto Y, Tokunaga K, et al. 2006. Phylogenetic relationship of the populations within and around Japan using 105 short tandem repeat polymorphic loci. Hum Genet. 118(6):695–707.

Lowery RK, Herrera KJ, Barrett DA, Rodriguez R, Hadden LR, Harutyunyan A, Margaryan A, et al. 2011. Regionalized autosomal STR profiles among Armenian groups suggest disparate genetic influences. Am J Phys Anthropol. 146(2):171–178.

Lu Y, Song P, Huang JC, Wu X. 2017. Genetic polymorphisms of 20 autosomal STR loci in 5141 individuals from the Han population of Xiamen, Southeast China. Forensic Sci Int Genet. 29:e31–e32.

Seyoum M, Du XM, He SP, Jia YH, Pan Z, Sun JL. 2018. Analysis of genetic diversity and population structure in upland cotton (*Gossypium hirsutum* L.) germplasm using simple sequence repeats. J Genet. 97(2): 513–522.

Tereba A. 1999. Tools for analysis of population statistics. Profiles in DNA. 3:14–16.

Walsh PS, Metzger DA, Higuchi R. 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques. 10:506–513.

Wu L, Pei B, Ran P, Song X. 2017. Population genetic analysis of Xiamen Han population on 21 short tandem repeat loci. Leg Med (Tokyo). 26: 41–44.

Wu X, Zheng JL, Lou Y, Wei XH, Wang BJ, Yao J. 2020. Genetic polymorphisms of 20 autosomal STR loci in the Han population of Zhangzhou City, Southeastern China. Leg Med (Tokyo). 46:101726.

Yao J, Wang BJ. 2016. Genetic variation of 25 Y-chromosomal and 15 autosomal STR Loci in the Han Chinese Population of Liaoning Province, Northeast China. PLoS One. 11(8):e0160415.

Yuan JY, Wang XY, Shen CM, Liu WJ, Yan JW, Wang HD, Pu HW, et al. 2014. Genetic profile characterization and population study of 21 autosomal STR in Chinese Kazak ethnic minority group. Electrophoresis. 35(4):503–510.

Zhan X, Adnan A, Zhou Y, Khan A, Kasim K, Mcnevin D. 2018. Forensic characterization of 15 autosomal STRs in four populations from Xinjiang, China, and genetic relationships with neighboring populations. Sci Rep. 8(1):4673.

Zhang YD, Tang XL, Meng HT, Wang HD, Jin R, Yang CH, Yan JW, et al. 2015. Genetic variability and phylogenetic analysis of Han population from Guanzhong region of China based on 21 non-CODIS STR loci. Sci Rep. 5:8872.

Zhu BF, Zhang YD, Shen CM, Du WA, Liu WJ, Meng HT, Wang HD, et al. 2015. Developmental validation of the AGCU $21+1$ STR kit: a novel multiplex assay for forensic application. Electrophoresis. 36(2):271–276.