ORIGINAL ARTICLE

Genetic variations and forensic characteristics of Han Chinese population residing in the Pearl River Delta revealed by 23 autosomal STRs

Guanglin He[1] · Zheng Wang[1] · Mengge Wang[1] · Xing Zou[2] · Jing Liu[1] · Shouyu Wang[1] · Yiping Hou[1]

Abstract

Pearl River Delta, known as the Chinese "South Gate", locates adjacent to Hong Kong and Macao, and across the sea from the Southeast Asian region. Herein, 4325 unrelated Han Chinese individuals residing in the Pearl River Delta region were recruited and genotyped with 23 autosomal STRs (the expanded CODIS core loci plus D6S1043, Penta D and Penta E). No evidence of deviation from Hardy–Weinberg equilibrium after Bonferroni correction was observed. The combined match probability and combined power of exclusion were 1.7829×10^{-28} and 0.9999999996, respectively. Next, population comparisons among ethnically, linguistically, and geographically diverse populations (11 Chinese populations based on genotypes of same 23 autosomal STRs, 46 Chinese nationwide populations based on allele frequency distributions of 19 autosomal STRs and 51 worldwide populations on the basis of 20 autosomal STRs) were performed via Structure, MDS, PCA and neighbor-joining tree. Genetic heterogeneities among Chinese nationalities along ethno-linguistic boundaries (Turkic-speaking, Tibeto-Burman-speaking, and Chinese-speaking populations) and geographic divisions (North-Han and South-Han) have been illustrated by first two autosomal microsatellite datasets. The results from worldwide population genetic relationship exploration based on 20 autosomal STRs demonstrated that genetic affinity has existed within ethnical and geographical close populations.

Keywords Forensic genetics · Genetic polymorphisms · Han Chinese · STR · Population comparison

Guanglin He and Zheng Wang have contributed equally to this work and should be considered as co-first authors.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11033-018-4264-y) contains supplementary material, which is available to authorized users.

✉ Yiping Hou
forensic@scu.edu.cn; profhou@yahoo.com

Guanglin He
2017324010012@stu.scu.edu.cn

Zheng Wang
wangzhengtim@scu.edu.cn

Mengge Wang
2016224010014@stu.scu.edu.cn

Xing Zou
2015110076@stu.cqmu.edu.cn

Jing Liu
2015224010021@stu.scu.edu.cn

Shouyu Wang
2017324010021@stu.scu.edu.cn

1 Institute of Forensic Medicine, West China School of Basic Medical Sciences and Forensic Medicine, Sichuan University, Chengdu 610041, Sichuan, China

2 Department of Forensic Medicine, College of Basic Medicine, Chongqing Medical University, Chongqing 400016, China

Introduction

Short tandem repeats (STRs), known as microsatellites normally consisting of 2–6 base pairs in length, are tandemly repeated sequences of DNA which widely exist in the human genome, numbering over 1.6 million [1, 2]. STRs conveying high information contents have played an important role in population genetics, anthropology, genetic genealogy, and forensics. The Combined DNA Index System (CODIS) core

loci working group advocates the wiedly used the CODIS system (13 loci) and expanded CODIS (20 loci) to establish the forensic search database to assist in individual identification and facilitate greater power of discrimination. PowerPlex Fusion Systems [3] (Promega, WI, USA) and GlobalFiler PCR Amplification Kit [4] (Thermo Fisher Scientific, MA, USA) have been released to achieve data sharing locally and internationally. Recently, to get the harmonization of Chinese National Database and expanded CODIS system in Chinese forensic biology laboratories, a 25-locus, six-dye system has been developed and validated [5–7].

Pearl River Delta (PRD), which is geographically situated at the crossroads of Chinese inland and the special administrative regions of Hong Kong and Macao, has been an indispensable piece to understand Chinese population genetic diversity and population structure. Today, the Pearl River Delta has a population of approximately 57.2 million and is divided into nine administrative districts: Guangzhou, Shenzhen, Foshan, Dongguan, Huizhou, Zhongshan, Zhuhai, Jiangmen and Zhaoqing. STRs for Identity ENFSI Reference Database (STRidER) [8] is committed to improving autosomal STR data quality. Currently, most of autosomal STR data included in the STRidER is collected from European populations, however, no data is available for China which has a population over 1.3 billion consisting of the largest world Han Chinese group (91.51%) and 55 officially recognized minority ethnicities (8.49%). Hitherto, the genetic variants and forensic characteristics, as well as genetic relationships of Pearl River Delta Han and other nationwide or worldwide populations are largely unknown.

Here, in continuation to our previous studies [5–7], we investigated the genetic variations and forensic features of 23 STRs in Pearl River Delta Han nationality and analyzed population structures of Chinese populations across 6 provinces using a new autosomal STR dataset with a total of 7097 samples, of which 4325 are newly reported here, from 12 populations. Additionally, we briefly discussed the hierarchy of autosomal microsatellites in the broader populations in East Asia and the whole world within and between ethnic/continental boundaries via principal component analysis, multidimensional scaling analysis, and phylogenetic relationship reconstruction on the basis of two datasets (46 national populations based the 19 autosomal STRs and 51 worldwide populations based on 20 autosomal STRs).

Methods and materials

Sample collection

4325 Blood stains (667 females and 3658 males) from Pearl River Delta in the south China were obtained with informed written contents. This work was reviewed and approved by the Medical Ethics Committee of the Sichuan University (K2015008) and it followed by the corresponding humane and ethical research principles. All included participants were the offspring of aboriginal members over three generations.

DNA extractions and quantification

The PureLink Genomic DNA Mini Kit (Thermo Fisher Scientific) was utilized to extract and isolate the human genomic DNA on the basis of the manufacturer's instructions. The Quantifiler Human DNA Quantification kit (Thermo Fisher Scientific) was used to assess the nucleotide quantity and quantification on a 7500 Real-time PCR System (Thermo Fisher Scientific) according to the manufacturer's protocol. The DNA was diluted to 2.0 ng/μL and stored at −20 °C until subsequently used.

DNA amplification and genotyping

25 genetic markers, including D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, Penta E, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, D6S1043, D10S1248, D1S1656, D12S391, D2S1338, Penta D, as well as two loci of Y-InDel (rs2032678) and Amelogenin for sex typing, were co-amplified using the Huaxia Platinum System (Thermo Fisher Scientific) on a ProFlex PCR System (Thermo Fisher Scientific) according to the manufacturer's recommendations. Fragments were separated and detected using the Applied Biosystems 3500 Genetic Analyzer and sized with GeneScan 600 LIZ Size Standard v2.0 (Thermo Fisher Scientific) on the basis of the user's manual instructions. Allele calling was carried out using GeneMapper ID-X v.1.4 software.

Statistical analysis

Allelic frequencies of 23 autosomal STRs and forensic statistical parameters were calculated using the online tool of STR Analysis for Forensics (STRAF) [9] and Forensic statistics analysis toolbox (FORSTAT) [10], and further consistently evaluated using the STRidER [8], including matching probability (MP), power of discrimination (PD), expected heterozygosity (He), observed heterozygosity (Ho), power of exclusion (PE), typical paternity index (TPI), polymorphism information content (PIC), combined match probability (CMP), combined power of discrimination (CPD), and combined power of exclusion (CPE). Arlequin v3.5 was used to evaluate p value of exact tests of Hardy–Weinberg equilibrium (HWE) and Linkage Disequilibrium (LD). Comprehensive population comparisons on the basis of three datasets (11 Chinese population based on the genotype data of 23 autosomal STRs, 46 Chinese nationwide populations

based on the allele frequency distributions of 19 autosomal STRs and 51 worldwide populations according to genetic variations of 20 autosomal STRs) were performed. The pairwise Nei's genetic distances and Fst values were estimated using the STRAF [9] and Phylip 3.9 [11]. Principal component analysis (PCA) based on allele frequencies and genotypes was respectively conducted using the MVSP [12] and STRAF [9] to assess the population genetic structures and relationships with the reference populations. Multidimensional scaling plots (MDS) based on the genetic distance matrixes were carried out using the IBM SPSS Statistics 23 software [13]. Neighbor-joining (N-J) trees were constructed using the Mega 7.0 [14]. Structure [15] with length of Burn-in of 100,000 steps and number of MCMC replications of 100,000 was used to explore 11 Chinese population structures. 10 runs, 'Admixture' and 'correlated allele frequencies' models were employed. Structure Harvester [16] was used to choose the most likely value of K. CLUMPP v.1.1.222 [17] and online tool in the PGG Population database [18] were used to visualize the results of Structure.

Quality control

This work was strictly followed the Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories published by Scientific Working Group on DNA Analysis Method (SWGDAM) [19] and was performed in the Institute of Forensic Medicine, West China School of Basic Medical Sciences and Forensic Medicine, Sichuan University, which is certificated by China National Accreditation Service for Conformity Assessment Laboratory Accreditation (CNAS) and ISO 17025. The positive control (007) and negative control (ddH2O and kit blank) were used in each batch of genotyping. The data of 4325 unrelated Han Chinese in present study have been submitted to the STRidER [8] and received the accession number of STR000136.

Results and discussion

Forensic efficacy

4325 Pearl River Delta Han individuals were successfully genotyped in the present study. No deviations from HWE are observed among 23 autosomal STRs after Bonferroni correction of the multiplex test (Table S1). The LD tests show p value below 0.0002 (a threshold after Bonferroni correction) at one pairwise locus, D5S818 and D8S1179 (Table S2), which may be caused by the results of natural selection, mutation or recombination, founder effects, genetic admixture and so on. A total of 354 alleles are observed with the allele frequencies ranging from 0.0001 to 0.5450 (Table S3). Table S1 presents the forensic statistical

parameters. PIC in PRD Han Chinese population varies from 0.5398 to 0.9016 and all 23 studied STRs are observed to be fairly informative (PIC > 0.5). The locus of Penta E (0.9016) is found to be the most informative, followed by D6S1043 (0.8638) and FGA (0.8567). He and Ho span from 0.6022 to 0.9083, and from 0.5910 to 0.9003, respectively. MP and PD vary from 0.0150 to 0.2198, and 0.7802 to 0.9849, respectively. PE ranges from 0.2802 to 0.7961, and TPI spans from 1.2224 to 5.0174. The combined probability of match is 1.7829×10^{-28}, and the combined probability of exclusion is 0.9999999996. Our results combined with our previous findings [5–7] illustrated that STR loci included in the Huaxia Platinum System are high diversity and can be used as a powerful tool in forensic individual identification and parentage testing in the Chinese populations along different ethnical, geographical or administrative divisions.

Population comparisons

The population genetic characteristics of the Huaxia Platinum System

To investigate the genetic landscapes of Chinese populations, we performed a series of analyses based on the genotype data of the Huaxia Platinum System among our investigated 10 populations [5–7] and one previously published population [20]. PCA based on the genetic variations of PC1 (0.58%), PC2 (0.48%) and PC3 (0.47%) reveals the South-Han Chinese populations from both Pearl River Delta and Hainan constitute the same genetic cluster with Sichuan Han (Fig. 1a and Fig. S1), implying Han Chinese along geographical and administrative divisions shares the common ancestry. Besides, the clusters of Chinese minorities (Hui, Yi, Uyghur and Tibetan) tend to partially overlap with Han Chinese populations and partially extend towards a separated group, suggesting the possible genetic admixture and high level of gene flow between and within Han Chinese populations and minority ethnicities. The Fst pairwise genetic distances among 11 populations are presented in Table S4. Pearl River Delta Han has a close genetic relationship with Hainan Han (0.00023) and a relative genetically distant relationship with Urumqi Uyghur (0.00283). In our MDS analysis, four Tibeto-Burman populations are located in the right upper part and three Uyghur populations are located in the left part (Fig. 1b and Fig. S2). Pearl River Delta Han and Urumqi Uyghur are clustered together. Phylogenetic relationship is constructed on the basis of Fst genetic matrix and N-J algorithm (Fig. 1c). The phylogeny reveals a clear genetic differentiation between Turkic-speaking populations and Sino-Tibetan-speaking populations. Three Uyghur populations are grouped together and clearly separated with others. Additionally, the genetic distinctions between the language subfamilies (Tibeto-Burman-speaking
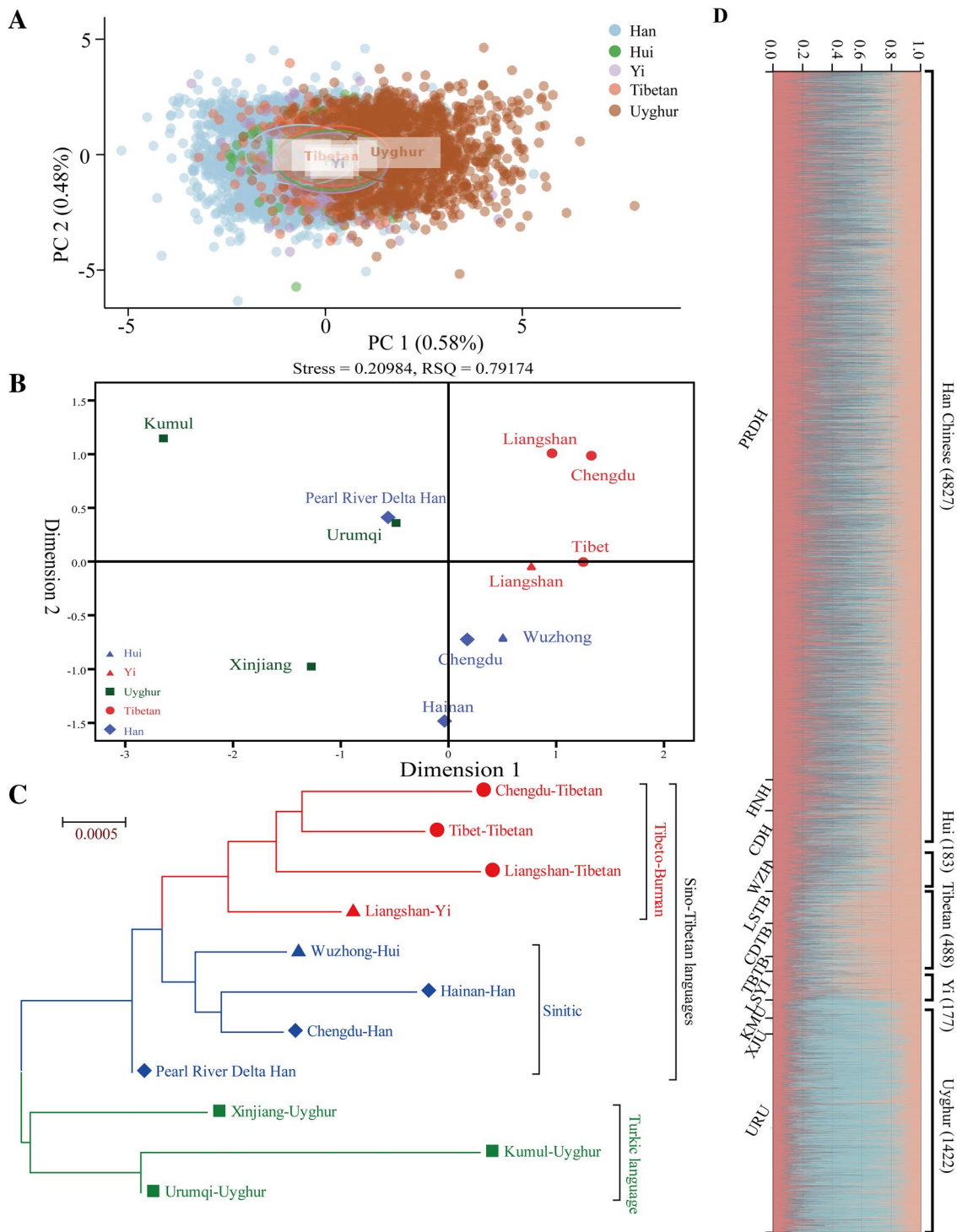
Fig. 1 Genetic relationships and population genetic admixture structure among 11 Chinese populations based on the genetic variation of 23 autosomal STRs included in the Huaxia Platinum six-dye amplification system. **a** Principle component analysis (PCA) based on PC1 and PC2; **b** multidimensional scaling plots (MDS) constructed by Fst genetic distance matrix; **c** phylogenetic tree constructed by using the neighbor-joining (N-J) algorithm; **d** genetic components in 11 population at K = 4

populations and Chinese/Sinitic speaking populations) are also observed. The model-based ancestry-estimation method Structure was used to analyze population structure of 11

groups consisting of 7097 individuals. The optimal separation of four ancestral components (K = 4) is identified (Fig. S3). We defined the highest proportion of ancestral

components in one ethnicity as ethnic-dominant ancestry component (EDAC). Four EDACs can be observed: Han-Chinese-dominant ancestry component (HCDAC), Uyghur-dominant ancestry component (UDAC), Tibetan-Burman-dominant ancestry component (TBDAC) and Hui-dominant ancestry component (HDAC). We observe that three Han populations (HCDAC, 20.51%), four Tibeto-Burman populations (TBDAC, 31.43%) and three Uyghur groups (UDAC, 28.57%) have a dominantly higher corresponding ethnic-dominant ancestry component, whereas Hui exhibits four ancestral ingredients evenly (Fig. 1d).

The current analysis suggests that genetic affinity and distinction are accompanied with the linguistic boundaries and ethnic categories. As our findings in the PCA based on the genotype data of 23 autosomal STRs shows a large overlap of populations, which is obviously not visible when these populations are assigned and then represented as one data point in a PCA, MDS or phylogenetic tree. These results should be carefully used in the study design of case-control study to avoid population stratification and the interpretation of the process of human origin, migration and genetic admixture. The population genetic structures need to be confirmed or carefully dissected using the high-density genome wide SNP chip or the whole genome sequencing, such as the population history reconstruction in Tibetans and Sherpas [21].

Genetic differentiation along Chinese administrative and ethnic divisions

To gain an overview of the broad genetic similarities and differences of more Chinese populations, we projected to combine the allele frequency distributions of Pearl River Delta Han with previously investigated 45 populations from 11 different ethnicities using PCA, MDS, and phylogenetic relationship reconstruction. The Nei's pairwise genetic distances between the studied PRD Han and 45 reference populations are listed in Table S5. PRD Han has the close genetic relationships with adjacent populations of Guangxi Han (0.0030) and Western Fujian Han (0.0031) and has the genetically distant relationships with Turkic-speaking populations (Xinjiang Uyghur1: 0.0354; Xinjiang Uyghur2: 0.0488; Xinjiang Uyghur3: 0.0568; and Xinjiang Kazakh: 0.0563), followed by Tibeto-Burman populations (Liangshan Tibetan: 0.0339; and Tibet Tibetan: 0.0349). We performed a principal component analysis in which the first three PCs are found to account for 28.519%, 16.530%, and 6.308%, respectively, of the total variations. As shown in Fig. 2a, b, PC1 can differentiate four Turkic-speaking populations from others and PC2 displays a clear separation between the South-Han and North-Han. In our MDS analysis, we see that 31 Han Chinese populations from different administrative regions fall into two distinct clusters, respectively associated

with North-Han and South-Han defined by Yangzi River boundary (Fig. 2c), which is in line with their positions on the two-dimension PCA established by the combination of PC1 and PC2, as well as PC2 and PC3. Other Chinese minorities are scattered around the profile in MDS and PCA. Similar compositions of genetic affinity clusters are identified in the N-J-based phylogenetic relationship reconstruction. Figure 2d presents four clusters: Turkic-speaking populations cluster, Chinese minorities cluster, North-Han Chinese cluster and South-Han Chinese cluster. Four Yunnan minorities in the Southwest China are grouped with the South-Han Chinese cluster and located at the farthest end of the evolutionary tree. Based on these results, our studied PRD Han Chinese population fall within the South-Han Chinese cluster and first group with Guangxi Han and Yunnan Yi, and then group together with Hainan Han and Yunnan Hani cluster. Collectively, this population genetic comparison showed that the autosomal STR gene pool in China is comparatively homogeneous. But at the same time, it revealed that the small genetic differentiations between the subgroups (Turkic-speaking and Sino-Tibetan-speaking populations; southern and northern populations as well as Han Chinese populations and Chinese minorities) exist although large-scale population migration occurred in the East Asian in the prehistoric and historic period. Our findings are characterized by a significant association between genetic and ethnical origin, language boundary and geographic distance.

Population structures among worldwide populations

To the best of our knowledge, it is rare to find examples of studies that have used autosomal STRs to provide insights into global population relationships [22, 23]. We subsequently assessed the genetic relationships between the PRD Han and a panel of 50 worldwide reference populations (22 East Asians, 3 South Asians, 4 Oceanians, 4 Africans, 6 Europeans and 11 Americans) based on the genetic polymorphisms of 20 overlapped autosomal STRs among Huaxia Platinum System [6], GlobalFiler Express PCR amplification kit [4] and PowerPlex Fusion System [3] using PCA, MDS and N-J algorithm based clustering method. Table S6 presents the Nei's standard genetic distances among 51 worldwide populations. There are largest genetic distances between PRD Han and South African residing in AmaXhosa (0.2015), followed by African living in AmaZulu (0.1834), as well as the smallest genetic distance with Guangdong Han (0.0032).

Principal component analysis based on the first three components (extracting 57.772% genetic variations) reveals that our studied PRD Han population is more closely related to Han Chinese populations (Central Chinese Han, Guangdong Han and Guizhou Han) than any other reference populations, followed by Chinese minority populations and Japanese and
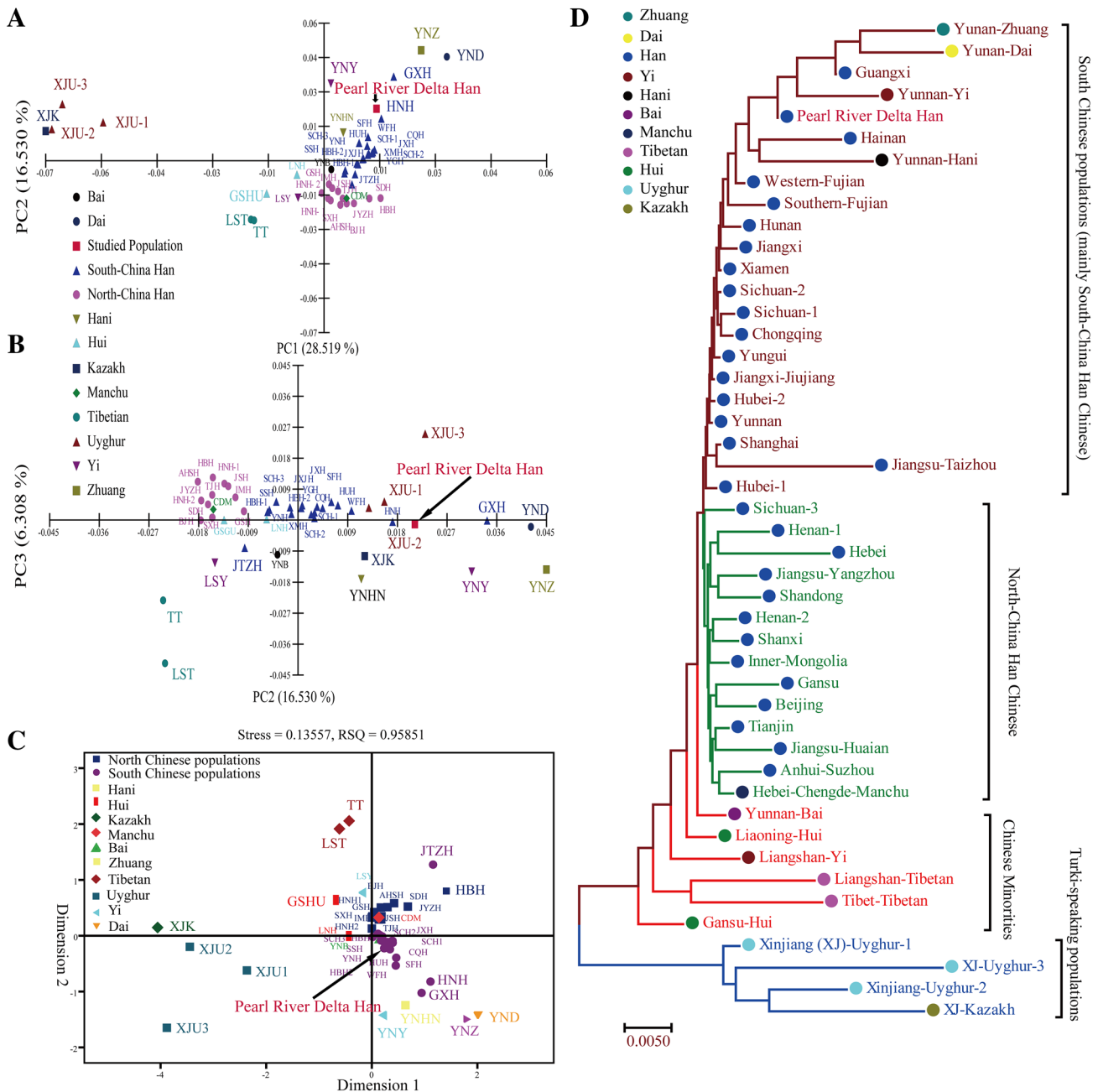
Fig. 2 PCA, N-J tree and MDS analyses reflecting genetic relationships among 46 Chinese nationwide populations. **a** Principal components analysis of 46 populations based on the genetic variation of PC1 and PC2; **b** two-dimension PCA on the basis of genetic variation of PC2 and PC3; **c** multidimensional scaling plots was constructed according to the Nei's genetic distance matrix; **d** phylogeny was estimated with Nei's genetic distance using neighbor-joining (N-J) method. Group name abbreviations are in accordance with Table S5

Korean populations (Fig. 3a). PC1 (31.477%) from a global scale shows a genetic differentiation between East Asians (with the exception of 5 Chinese Turkic-speaking populations) and others. Americans can be partially separated on the basis of PC2 (16.113%), but completely distinguished via PC3 (10.182%). Two-dimensional plots analysis shows the consistent population distribution patterns with the observed structure in the PCA. East Asians combined with three South

Asians are separated with other four continental populations and assigned at the upper position of the MDS. Most of the American and European groups show a strong genetic affinity within the groups, but there is a partially confused aggregation of some of them. Africans and Oceanians are scattered due to the relatively low number of populations. To clearly explore the phylogeny relationship among 51 populations, we reconstructed the N-J tree based on the pairwise
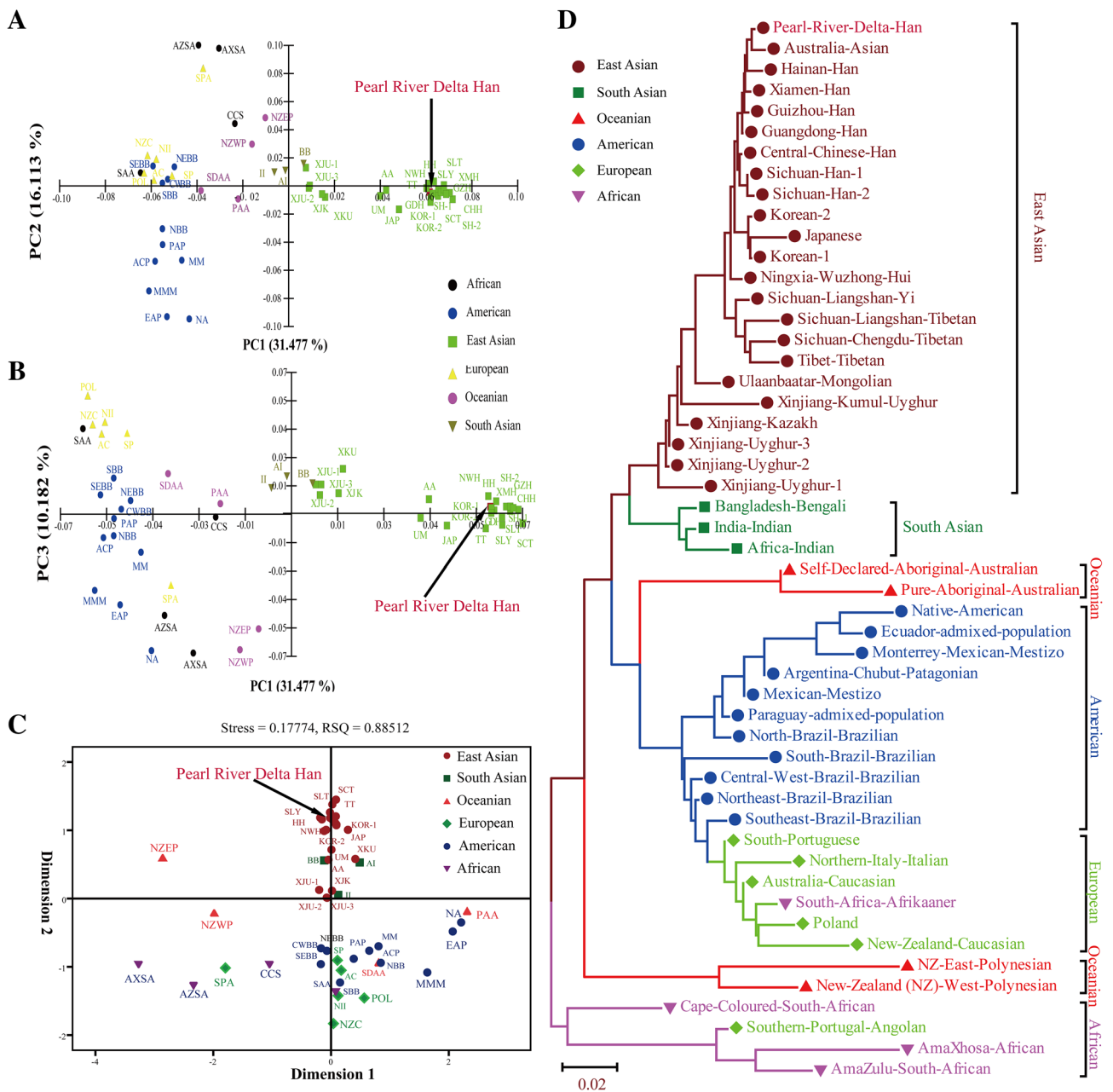
Fig. 3 Genetic relationship of 51 worldwide populations. **a** Principal component analysis of global populations on the basis of genetic variation of PC1 and PA2; **b** PCA constructed on the combination of PC1 and PC3; **c** two-dimensional scaling plots constructed using SPSS software; **d** N-J tree indicate the worldwide population phylogenetic relationship. Group name abbreviations are in accordance with Table S6

Nei's genetic distances (Fig. 3d). Notably, seven clusters are observed, which is in close agreement with our previous findings based on genetic variants of 165 ancestry informative single nucleotide polymorphisms [24, 25]. Clusters in the phylogenetic tree from bottom to top accordingly are Africans, New Zealand Oceanians, Europeans, Americans, Aboriginal Australians, South Asians and East Asians. African Afrikaans is clustered with European groups and

Southern Portugal Angolan is clustered with African populations, which is consistent with the ethnic origins (Afrikaans people are of European origin, and Angolans are of African origin). These findings imply that some geographically close inter-continental populations have a close genetic affinity. Our investigated PRD Han herein forms a clade with Han Chinese sub-cluster, which is first grouped with Australia Asian and then clustered with Hainan Han and Xiamen

Han. Our analysis provides genetic evidence that modern human populations can be assigned into six main clusters, which are consistent with the ethnic origin and geographical categories.

Conclusion

This study submitted the first batch of genetic population data from Pearl River Delta Han to the STRidER and presented the comprehensive population comparisons to dissect Chinese population structures and genetic admixtures as well as worldwide population relationships based on microsatellites. Our analyses combined with our previous investigated results revealed that 23 autosomal STRs included in Huaxia Platinum System are informative and polymorphic in Han groups and Chinese minority populations and can be used as a useful tool in the forensic individual identification and parentage testing after the establishment of a region-specific database. Genetic structure investigation among 7097 Chinese individuals from 11 populations based on Huaxia Platinum panel revealed that genetic differences are observed among different language families or subfamilies. Additionally, population comparisons among 46 Chinese population revealed a significant genetic distinction between the North-Han Chinese and South-Han Chinese, as well as genetic differentiation is mainly existed between Uyghurs, Tibetans and Han. Finally, worldwide genetic structures among 51 populations based on 20 autosomal STRs identified six main genetic affinity clusters in accordance with the continent or geographic boundaries. Collectively, Han Chinese population residing in the Pearl River Delta has the genetically close relationships with South-China Hans (most predominantly Guangdong Han and Hainan Han).

Compliance with ethical standards

Conflict of interest The authors claim that they have no competing interests.

References

1. Willems T, Gymrek M, Highnam G, Mittelman D, Erlich Y, 1000 Genomes Project (2014) The landscape of human STR variation. Genome Res 24(11):1894–1904. https://doi.org/10.1101/gr.177774.114

2. Gymrek M, Willems T, Reich DE, Erlich Y (2017) Interpreting short tandem repeat variations in humans using mutational constraint. Nat Genet. https://doi.org/10.1101/092734

3. Oostdik K, Lenz K, Nye J, Schelling K, Yet D, Bruski S, Strong J, Buchanan C, Sutton J, Linner J, Frazier N, Young H, Matthies L, Sage A, Hahn J, Wells R, Williams N, Price M, Koehler J, Staples M, Swango KL, Hill C, Oyerly K, Duke W, Katzilierakis L, Ensenberger MG, Bourdeau JM, Sprecher CJ, Krenke B, Storts DR (2014) Developmental validation of the PowerPlex((R)) Fusion System for analysis of casework and reference samples: A 24-locus multiplex for new database standards. Forensic Sci Int Genet 12:69–76. https://doi.org/10.1016/j.fsigen.2014.04.013

4. Hennessy LK, Mehendale N, Chear K, Jovanovich S, Williams S, Park C, Gangano S (2014) Developmental validation of the GlobalFiler((R)) express kit, a 24-marker STR assay, on the RapidHIT((R)) System. Forensic Sci Int Genet 13:247–258. https://doi.org/10.1016/j.fsigen.2014.08.011

5. He G, Wang M, Liu J, Hou Y, Wang Z (2017) Forensic features and phylogenetic analyses of Sichuan Han population via 23 autosomal STR loci included in the Huaxia Platinum System. Int J Legal Med. https://doi.org/10.1007/s00414-017-1679-2

6. 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. Sci Rep 6:31075. https://doi.org/10.1038/srep31075

7. 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. Sci Rep 8(1):2429. https://doi.org/10.1038/s41598-018-20871-7

8. Bodner M, Bastisch I, Butler JM, Fimmers R, Gill P, Gusmao L, Morling N, Phillips C, Prinz M, Schneider PM, Parson W (2016) Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER). Forensic Sci Int Genet 24:97–102. https://doi.org/10.1016/j.fsigen.2016.06.008

9. Gouy A, Zieger M (2017) STRAF-A convenient online tool for STR data evaluation in forensic genetics. Forensic Sci Int Genet 30:148–151. https://doi.org/10.1016/j.fsigen.2017.07.007

10. Ristow PG, D'Amato ME (2017) Forensic statistics analysis toolbox (FORSTAT): a streamlined workflow for forensic statistics. Forensic Sci Int Genet Supple Ser 6:e52–e54. https://doi.org/10.1016/j.fsigss.2017.09.006

11. Cummings MP (2004) PHYLIP (phylogeny inference package). Wiley, Hoboken

12. Kovach WL (2007) MVSP-A multivariate statistical package for Windows, ver. 3.1. Kovach Computing Services, Pentraeth

13. Hansen J (2005) Using SPSS for Windows and Macintosh: analyzing and understanding data. Am Stat 59(1):113–113

14. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33(7):1870–1874. https://doi.org/10.1093/molbev/msw054

15. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14(8):2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

16. Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4(2):359–361. https://doi.org/10.1007/s12686-011-9548-7

17. Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23(14):1801–1806. https://doi.org/10.1093/bioinformatics/btm233

18. Zhang C, Gao Y, Liu J, Xue Z, Lu Y, Deng L, Tian L, Feng Q, Xu S (2017) PGG. Population: a database for understanding the genomic diversity and genetic ancestry of human populations. Nucleic Acids Res. https://doi.org/10.1093/nar/gkx1032

19. Scientific Working Group on DNA Analysis (SWGDAM) (2017) Interpretation guidelines for autosomal STR typing by forensic DNA testing laboratories. https://docs.wixstatic.com/ugd/4344b4340_4350e2749756a2242528e2746285a2749755bb2749478f2749754c.pdf

20. Jin X, Wei Y, Chen J, Kong T, Mu Y, Guo Y, Dong Q, Xie T, Meng H, Zhang M, Li J, Li X, Zhu B (2017) Phylogenic 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

21. Zhang C, Lu Y, Feng Q, Wang X, Lou H, Liu J, Ning Z, Yuan K, Wang Y, Zhou Y (2017) Differentiated demographic histories and local adaptations between Sherpas and Tibetans. Genome Biol 18(1):115

22. Chu JY, Huang W, Kuang SQ, Wang JM, Xu JJ, Chu ZT, Yang ZQ, Lin KQ, Li P, Wu M, Geng ZC, Tan CC, Du RF, Jin L (1998) Genetic relationship of populations in China. Proc Natl Acad Sci USA 95(20):11763–11768

23. Buckleton J, Curran J, Goudet J, Taylor D, Thiery A, Weir BS (2016) Population-specific FST values for forensic STR markers: a worldwide survey. Forensic Sci Int Genet 23:91–100. https://doi.org/10.1016/j.fsigen.2016.03.004

24. Wang Z, He G, Luo T, Zhao X, Liu J, Wang M, Zhou D, Chen X, Li C, Hou Y (2018) Massively parallel sequencing of 165 ancestry informative SNPs in two Chinese Tibetan-Burmese minority ethnicities. Forensic Sci Int Genet 34:141–147. https://doi.org/10.1016/j.fsigen.2018.02.009

25. He G, Wang Z, Wang M, Luo T, Liu J, Zhou Y, Gao B, Hou Y (2018) Forensic ancestry analysis in two Chinese minority populations using massively parallel sequencing of 165 ancestry-informative SNPs. Electrophoresis. https://doi.org/10.1002/elps.201800019