

Genetic and structural characterization of 20 autosomal short tandem repeats in the Chinese Qinghai Han population and its genetic relationships and interpopulation differentiations with other reference populations

Zhanhai Wang^a, Bin Lu^a, Xiaoye Jin^{b,c,d}, Jiangwei Yan^e, Haotian Meng^{b,c,d*} and Bofeng Zhu^{b,c,d*}

^aThe Public Security Bureau of Qinghai Province, Xining, China; ^bKey Laboratory of Shaanxi Province for Craniofacial Precision Medicine Research, College of Stomatology, Xi'an Jiaotong University, Xi'an, China; ^cClinical Research Center of Shaanxi Province for Dental and Maxillofacial Diseases, College of Stomatology, Xi'an Jiaotong University, Xi'an, China; ^dResearch Center of Stomatology, Stomatological Hospital, Xi'an Jiaotong University, Xi'an, China; ^eBeijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

ABSTRACT

China is a multinational country composed of 56 ethnic groups of which the Han Chinese accounts for 91.60%. Qinghai Province is located in the northeastern part of the Qinghai–Tibet Plateau, has an area of 72.12 km², and is the fourth largest province in China. In the present study, we investigated the genetic polymorphisms of 20 short tandem repeat (STR) loci in a Qinghai Han population, as well as its genetic relationships with other populations. A total of 273 alleles were identified in 2 000 individuals at 20 loci, and the allelic frequency ranged from 0.0002 to 0.5327. The 20 STR loci showed a relatively high polymorphic rate in the studied group. Observed and expected heterozygosities ranged 0.613 0–0.907 5 and 0.614 8–0.920 0, respectively. The combined power of discrimination, and the probability of exclusion in duo and trio cases were 0.999 999 999 999 999 999 999 999 34, 0.999 996 0 and 0.999 999 996 5, respectively. Analyses of interpopulation differentiation revealed that the most significant differences were found between the Qinghai Han and Malaysian, while no significant differences were found between the Qinghai Han and Han people from Shaanxi and Jiangsu. The results of principal component analysis, multidimensional scaling analysis and phylogenetic reconstructions also suggested the close relationships between the Qinghai Han and other two Han populations. The present results, therefore, indicated that these 20 STR loci could be used for paternity testing and individual identification in forensic applications, and may also provide information for the studies of genetic relationships between Qinghai Han and other groups.

ARTICLE HISTORY

Received 1 April 2018
Accepted 3 June 2018

KEYWORDS

Genetic polymorphisms;
forensic genetics;
phylogenetic reconstruction;
short tandem repeat

Instruction

The Han population is the largest of the 56 officially recognized ethnic groups in China. Findings from the 6th National Population Census of 2010 suggest that they make up 91.60% of the overall Chinese population with a population of 1 220 844 520, and they are also distributed worldwide. The Chinese language, used as the spoken and written language of Han people, belongs to the Sino–Tibetan language family. The appellation ‘Han’ can be traced back to the Han dynasty of the second and third centuries and represents the majority of the Chinese population to date.

Qinghai Province has the fourth largest land area in China. It is located in the northeastern part of

the Qinghai–Tibet Plateau, which has an altitude over 3 000 m above sea level. The history of Qinghai Province began during the Han dynasty when General Huo Qubing built the military fortress known as Xipingting. This was the former site of Xining, which appeared during the Ming dynasty between 1368 and 1644. In 2010, Qinghai had a population of 5 626 723, of which 39% was taken up by minorities including Tibetans, Mongolians, Kazaks and the Hui, Tu and Salar.

In the present study, the genetic distributions of 20 short tandem repeat (STR) loci and the sex-determining locus amelogenin were studied in the Qinghai Han population. Additionally, the phylogenetic relationships between the Qinghai Han and

CONTACT Haotian Meng menghaotian0803@126.com

*These authors contributed equally to this work.

Supplemental data for this article can be accessed on the [publisher's website](#).

© 2018 The Author(s). Published by Taylor & Francis Group on behalf of the Academy of Forensic Science. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

other reference populations [1–13] were studied using 13 overlapping STR loci including *D8S1179*, *D21S11*, *D18S51*, *vWA*, *D3S1358*, *FGA*, *TH01*, *D5S818*, *D13S317*, *D7S820*, *CSF1PO*, *D16S539* and *TPOX* loci.

Materials and methods

Sample collection and DNA extraction

Blood samples were collected from 2 000 unrelated healthy Han individuals living in Qinghai Province, China, whose ancestors over the past three generations were Han individuals who had not migrated or interbred with other ethnic groups. All of the participants signed an informed consent form and completed a questionnaire providing information about their direct blood relatives over three generations. The experimental procedures conformed to the human and ethical research principles of Xi'an Jiaotong University Health Science Center, China. Genomic DNA extraction was performed using the Chelex-100 procedure as described by Walsh et al. [14].

Polymerase chain reaction (PCR) amplification and STR typing

PCR amplification was performed using the PowerPlex[®] 21 System (Promega, Madison, WI, USA). The total volume of PCR reactions was 25 μ L, containing 5 μ L PowerPlex[®] 21 5 \times Master Mix, 5 μ L PowerPlex[®] 21 5 \times Primer Pair Mix, 1 ng template DNA and amplification grade water. Amplification was carried out using a GeneAmp PCR System 9700 Thermal Cycler (Applied Biosystems, Foster City, CA, USA) under the manufacturer's specifications. The AB PRISM 3130 Genetic Analyzer (Applied Biosystems) was used to obtain sample genotypes. Raw data were analysed using GeneMapper ID 3.2 software (Applied Biosystems). 9947A DNA was used as a positive control.

Statistical analysis

Allelic frequencies of 20 STRs, their forensic relevant parameters, and *P*-values for exact tests of Hardy–Weinberg equilibrium were calculated using the modified Powerstat (version 1.2) spreadsheet. Linkage disequilibrium (LD) analysis of pairwise STR loci was calculated using Genepop version 4.0.10 [15]. Based on genetic data of the 13 overlapping STRs, genetic differentiation comparisons (*P*-values) between the Qinghai Han population and other referenced populations were conducted using Arlequin software version 3.1 with the method of analysis of molecular variance (AMOVA) [16].

Population genetic structure analysis among the Qinghai Han population and other populations was performed using Structure software version 2.2 [17]. The pairwise genetic distance (D_A) and fixation index (F_{st}) of the studied Han group and other populations were calculated using the DISPAN program [18] and Arlequin software version 3.1, respectively. Heatmaps of D_A and F_{st} between these populations were plotted by *R* software (<https://www.r-project.org/>). Principal components analysis (PCA) of these populations was drawn using MVSP software version 3.1 [19] based on the allelic frequencies of the 13 overlapping STRs. Multidimensional scaling (MDS) analysis of the Qinghai Han population and other compared populations was plotted using IBM SPSS version 18.0 (IBM Co., Armonk, NY, USA). Two different phylogenetic trees were constructed by MEGA software version 5.0 [20] and PHYLIP software version 3.6 to determine the phylogenetic relationships between Qinghai Han and other populations.

Results and discussion

Allelic distributions and forensic parameter analysis of 20 STR loci

The allelic frequencies of 20 autosomal STR loci and their corresponding forensic relevant parameters are shown in Table 1. A total of 273 alleles were found in the studied Han population within the 20 loci (Table 1). The minimum allelic frequency was 0.000 2 and the maximum was 0.532 7. The lowest values of the power of discrimination (*DP*), and the probability of exclusion (*PE*) in duo and trio cases were 0.792 6, 0.203 5 and 0.306 8, respectively, at the *TPOX* locus, while the highest values were 0.987 5, 0.720 5 and 0.810 8, respectively, at the *Penta E* locus. With the exception of locus *TPOX*, the polymorphism information content (*PIC*) of all remaining loci reached above 0.6. Highest observed heterozygosity (*Ho*) and expected heterozygosity (*He*) values were observed at the *Penta E* locus, while the lowest *Ho* and *He* values were at the *TPOX* locus. The combined *DP*, and *PE* in duo and trio cases were 0.999 999 999 999 999 999 999 34, 0.999 996 0 and 0.999 999 996 5, respectively. These results revealed that these 20 STR loci are highly polymorphic, and have the potential to be used in both forensic human identification and paternity testing in the Qinghai Han population.

LD analysis

The results of LD tests are shown in Table 2. After Bonferroni correction, the exact *P*-values of two

Table 1. The allelic frequencies and forensic statistical parameters of the 20 STR loci in Chinese Qinghai Han population ($n = 2,000$).

Allele	D8S1179	D21S11	D18S51	vWA	D3S1358	FGA	TH01	D5S818	D13S317	D7S820	CSFIPO	D16S539	TPOX	D2S1338	D19S433	Penta D	Penta E	D6S1043	D1580	D12S391
5							0.1035	0.0002								0.0050	0.0532			
6							0.2552	0.0140	0.0012	0.0012	0.0022		0.0002			0.0072	0.0002			
7							0.0598	0.0035	0.2662	0.1455	0.0028	0.0090	0.0002			0.0072	0.0022			
8	0.0008						0.5192	0.0680	0.1383	0.0630	0.0532	0.2665	0.1222			0.0425	0.0060	0.0008		
9	0.0002	0.0002								0.0028					0.0002	0.3132	0.0082	0.0020		
9.1							0.0390													
9.3							0.0228	0.1930	0.1475	0.1635	0.2305	0.1190	0.0215			0.1140	0.0422	0.0278		
10	0.1032	0.0018	0.0018				0.0005	0.3378	0.2415	0.3390	0.2422	0.2588	0.2918		0.0030	0.1600	0.1348	0.0985	0.0620	
10.1										0.0008										
11	0.0725	0.0018	0.0018				0.0005	0.3378	0.2415	0.3390	0.2422	0.2588	0.2918		0.0030	0.1600	0.1348	0.0985	0.0620	
11.1										0.0002										
11.2										0.0002										
11.3										0.0002										
12	0.1258	0.0292	0.0292		0.0012		0.2408	0.1595	0.1595	0.2380	0.3830	0.2148	0.0288		0.0400	0.1810	0.1068	0.1300	0.0398	
12.2															0.0048					
12.3																				
13	0.2355	0.2005	0.2005	0.0028	0.0010	0.0002	0.1315	0.0357	0.0415	0.0740	0.0740	0.1158	0.0018		0.2872	0.1242	0.0503	0.1362	0.1008	
13.2															0.0492					
14	0.2020	0.2192	0.2192	0.2525	0.0442		0.0098	0.0095	0.0042	0.0107	0.0107	0.0145	0.0008		0.2442	0.0420	0.0830	0.1380	0.0792	
14.2															0.1122					
15	0.1725	0.1785	0.1785	0.0330	0.3738		0.0015	0.0005	0.0002	0.0012	0.0012	0.0015	0.0002		0.0670	0.0090	0.0955	0.0162	0.3025	0.0160
15.2															0.1448					
15.3																				
16	0.0730	0.1235	0.1235	0.1855	0.3315									0.0078	0.0122	0.0018	0.0775	0.0025	0.0008	0.0085
16.2															0.0290				0.2272	0.0085
16.3																			0.0090	0.0002
16.4																			0.0855	0.0938
17	0.0115	0.0703	0.0703	0.2432	0.1840	0.0005								0.0648	0.0018		0.0002			
17.2															0.0028					
17.3																				
18	0.0030	0.0430	0.0430	0.1842	0.0562	0.0185								0.1078	0.0005		0.0790	0.1778	0.0552	0.2298
18.2																			0.0115	0.0002
18.3																			0.0002	0.0002
18.4																			0.0230	0.0010
19		0.0402	0.0402	0.0808	0.0072	0.0472								0.1590			0.0008	0.1582	0.0010	0.2245
19.2																	0.0595		0.0025	0.0002
19.3																			0.0025	0.0002
19.4																			0.0025	0.0002
20		0.0365	0.0365	0.0175	0.0008	0.0545								0.1270			0.0008	0.0528		0.1645
20.2																	0.0445			
20.3																				
21		0.0280	0.0280	0.0005		0.1038								0.0272			0.0305	0.0107		0.1128
21.2																				
21.3																				
22		0.0138	0.0138			0.1668								0.0440			0.0215	0.0005		0.0780
22.2						0.0062														
22.3						0.0002														
23		0.0040	0.0040			0.2175								0.2195			0.0088	0.0002		0.0440

(continued)

Table 1. Continued.

Allele	D8S1179	D21S11	D18S51	vWA	D3S1358	FGA	TH01	D5S818	D13S317	D7S820	CSFIPO	D16S539	TPOX	D2S1338	D19S433	Penta D	Penta E	D6S1043	D1580	D12S391	
23.2						0.009 5															
24			0.005 5			0.188 0								0.159 5			0.004 8			0.016 5	
24.2						0.005 2															
25			0.003 0			0.110 8								0.063 0			0.001 0			0.007 5	
25.2						0.004 8															
26			0.000 2			0.046 8								0.016 5						0.002 2	
26.2						0.001 2															
27			0.002 8			0.010 2								0.003 5						0.000 2	
27.2						0.002 8								0.000 5							
28			0.043 0			0.000 2															
28.2			0.009 5																		
29			0.249 8																		
29.2			0.003 2																		
29.3			0.000 2																		
30			0.296 0																		
30.2			0.015 0			0.000 2															
30.3			0.005 2																		
31			0.100 8																		
31.2			0.071 8																		
32			0.029 8																		
32.2			0.126 5																		
33			0.004 5																		
33.2			0.037 2																		
34.2			0.003 5																		
35			0.000 2																		
35.2			0.000 5																		
36			0.000 2																		
Ho	0.848 0	0.817 0	0.859 0	0.810 5	0.712 0	0.860 0	0.649 5	0.772 0	0.815 0	0.771 0	0.754 5	0.797 0	0.613 0	0.872 0	0.806 5	0.801 0	0.907 5	0.864 5	0.821 5	0.830 0	
He	0.836 8	0.814 2	0.853 0	0.800 8	0.711 4	0.858 5	0.648 9	0.768 5	0.803 1	0.774 8	0.733 1	0.788 1	0.614 8	0.862 2	0.814 7	0.811 3	0.920 0	0.873 3	0.823 9	0.839 6	
MP	0.049 2	0.058 3	0.038 8	0.071 0	0.135 4	0.035 2	0.169 4	0.089 9	0.069 1	0.084 0	0.119 5	0.079 8	0.207 4	0.034 6	0.058 0	0.058 6	0.012 5	0.029 7	0.051 0	0.044 9	
DP	0.950 8	0.941 7	0.961 2	0.929 0	0.864 6	0.964 8	0.830 6	0.910 1	0.930 9	0.916 0	0.880 5	0.920 2	0.792 6	0.965 4	0.942 0	0.941 4	0.987 5	0.970 3	0.949 0	0.955 1	
PIC	0.816 2	0.791 6	0.836 6	0.771 1	0.660 8	0.843 0	0.602 7	0.733 0	0.774 3	0.741 7	0.689 6	0.755 2	0.554 5	0.847 2	0.791 4	0.787 5	0.914 4	0.860 0	0.804 1	0.820 2	
PE(D)	0.503 4	0.470 3	0.546 6	0.427 5	0.294 4	0.558 0	0.243 0	0.376 0	0.431 9	0.388 7	0.324 1	0.403 5	0.203 5	0.564 9	0.467 5	0.458 5	0.720 5	0.589 7	0.492 0	0.513 6	
PE(T)	0.690 9	0.631 0	0.712 7	0.618 7	0.447 1	0.714 7	0.354 5	0.548 1	0.627 2	0.546 4	0.517 5	0.593 5	0.306 8	0.738 7	0.611 2	0.600 9	0.810 8	0.723 6	0.639 5	0.655 9	
P	0.000 5	0.879 0	0.528 0	0.616 5	0.925 7	0.756 0	0.598 2	0.400 8	0.622 0	0.701 1	0.040 9	0.615 2	0.913 3	0.383 4	0.816 4	0.645 4	0.007 9	0.127 0	0.064 2	0.479 7	

Ho: observed heterozygosity; He: expected heterozygosity; MP: matching probability; DP: power of discrimination; PIC: polymorphism information content; PE(D): probability of exclusion in duo cases; PE(T): probability of exclusion in trio cases; P: probability values of exact tests for Hardy-Weinberg equilibrium.

Table 2. The P -values of linkage disequilibrium of all pairwise STR loci in Chinese Qinghai Han population.

Loci	D12S391	D1580	D6S1043	Penta E	Penta D	D19S433	D2S1338	TPOX	D16S539	CSFIPO	D7S820	D13S317	D5S818	TH01	FGA	D3S1358	vWA	D18S51	D21S11
D1580	0.1551																		
D6S1043	0.2771	0.7982																	
Penta E	0.6945	0.8152	0.2802																
Penta D	0.0254	0.7055	0.0066	0.4194															
D19S433	0.0802	0.2771	0.0615	0.2612	0.3048														
D2S1338	0.4423	0.2852	0.6501	0.2553	0.3115	0.0797													
TPOX	0.3112	0.7976	0.9223	0.7566	0.0263	0.7542	0.8294												
D16S539	0.4738	0.4484	0.1465	0.0187	0.8278	0.9281	0.9635	0.1410											
CSFIPO	0.0005	0.4037	0.0170	0.5313	0.4707	0.3215	0.9315	0.7630	0.1539										
D7S820	0.8008	0.0564	0.5788	0.0007	0.9507	0.2528	0.2625	0.8995	0.1266	0.3115									
D13S317	0.1829	0.2837	0.9988	0.0258	0.1934	0.3382	0.1933	0.2390	0.4045	0.2021	0.4754								
D5S818	0.3253	0.3675	0.1428	0.8053	0.2074	0.3828	0.7188	0.9048	0.7329	0.4737	0.0830	0.6974							
TH01	0.0394	0.0142	0.1700	0.1010	0.8470	0.0107	0.8234	0.2202	0.4836	0.6192	0.3897	0.7108	0.4159						
FGA	0.0836	0.0244	0.2564	0.4916	0.6420	0.0000	0.8458	0.6329	0.3200	0.4592	0.0139	0.4040	0.0829	0.0000					
D3S1358	0.4837	0.6710	0.3565	0.2641	0.4658	0.0315	0.4711	0.3951	0.9057	0.3557	0.7964	0.3134	0.4852	0.2107	0.2971				
vWA	0.1037	0.0136	0.0214	0.0929	0.2552	0.5436	0.2276	0.5666	0.8062	0.0760	0.1615	0.0746	0.4124	0.7572	0.3077	0.7083			
D18S51	0.9656	0.2961	0.1556	0.7498	0.0785	0.6570	0.7123	0.2370	0.6535	0.4275	0.3850	0.2380	0.3816	0.1496	0.3399	0.4786	0.0124		
D21S11	0.8931	0.4014	0.0457	0.0506	0.0781	0.5753	0.9311	0.7897	0.4295	0.1856	0.2982	0.2974	0.1831	0.0151	0.8255	0.6209	0.7366	0.0037	
D8S1179	0.8549	0.1744	0.3066	0.1323	0.2231	0.2582	0.0007	0.3514	0.2052	0.3670	0.0161	0.7393	0.6912	0.7875	0.8631	0.0437	0.6134	0.2325	0.7850

(*D19S433* and *FGA*; *TH01* and *FGA*) out of the 190 pairwise comparisons were below the significant level (0.000263). LD can be influenced by many factors, such as selection, the rate of recombination, the mutation rate, genetic drift, the system of mating, population structure and genetic linkage. As the loci are located on different chromosomes, genetic linkage cannot explain the observed LD. However, additional studies are required to determine the role of other factors in LD.

Interpopulation differentiations between the Qinghai Han population and other compared populations

The P -values of genetic differentiation comparisons are shown in Table 3. Significant differences ($P < 0.05$) were observed between the Qinghai Han and the following groups: the Malaysian at 13 loci, the Tibetan and She at 11 loci, the Uyghur and Shui at 10 loci, the Zhuang at five loci, the Dong and Yi at three loci and the Hui, Guangdong Han and Russian at two loci. No significant differences were observed among the Qinghai Han, Jiangsu Han, and Shaanxi Han groups. The highest ethnic diversity was observed at the *D18S51* locus, where significant differences were found in eight out of 13 compared groups. The lowest ethnic diversity was observed at *CSFIPO* and *TPOX* loci, where significant differences were found in only three out of 13 compared groups.

Population structure clustering analysis of the 14 populations

We explored the population structures of the Qinghai Han and other published populations by the Structure program. Different K settings ($K = 2-7$) for the 13 overlapping loci dataset are shown in Supplementary Figure S1. Population components of the 14 populations at $K = 3$ are shown in Figure 1. Similar population component distributions were observed for all groups analysed, indicating a lack of population structure among these populations. Hence, the Qinghai Han group and other referenced populations showed no clear population structure in the present analysis.

Genetic distance (D_A and F_{st}) analysis among the 14 populations

Pairwise genetic distances of the Qinghai Han population and other reference populations are shown in Figure 2 and Supplementary Tables S1 and S2. As shown in Figure 2(A), close relative

Table 3. The *P*-values of pairwise comparisons between Chinese Qinghai Han population and other groups at 13 overlapping STR loci based on the method of AMOVA.

Populations	CSF1PO	D13S317	D16S539	D18S51	D21S11	D3S1358	D5S818	D7S820	D8S1179	FGA	TH01	TPOX	vWA
Hui	0.786 9	0.891 5	0.995 1	0.008 8	0.663 7	0.988 3	0.164 2	1.000 0	0.508 3	1.000 0	0.291 3	0.081 1	0.028 4
She	0.000 0	0.000 0	0.001 0	0.012 7	0.000 0	0.481 9	0.003 9	0.000 0	0.000 0	0.001 0	0.002 0	0.307 9	0.000 0
Shui	0.843 6	0.001 0	0.002 0	0.000 0	0.000 0	0.049 9	0.052 8	0.045 9	0.000 0	0.000 0	0.006 8	0.001 0	0.075 3
Dong	0.275 7	0.021 5	0.060 6	0.255 1	1.000 0	0.000 0	0.754 6	0.702 8	0.000 0	0.142 7	0.334 3	0.785 9	0.560 1
Uygur	0.000 0	0.000 0	0.004 9	0.000 0	0.000 0	0.002 0	0.000 0	0.002 9	0.404 7	0.223 9	0.000 0	0.421 3	0.000 0
Yi	0.656 9	0.881 7	0.081 1	0.149 6	0.000 0	0.000 0	0.050 8	0.380 3	0.112 4	0.017 6	0.106 6	1.000 0	1.000 0
Zhuang	0.819 2	0.763 4	0.581 6	0.000 0	1.000 0	0.021 5	0.937 4	0.027 4	0.000 0	0.199 4	0.011 7	1.000 0	0.098 7
Guangdong Han	0.960 9	1.000 0	1.000 0	0.134 9	1.000 0	0.313 8	0.870 0	0.339 2	0.004 9	0.084 1	0.027 4	0.098 7	0.373 4
Jiangsu Han	0.927 7	0.793 7	0.128 1	0.181 8	0.698 9	1.000 0	1.000 0	1.000 0	0.764 4	0.910 1	0.993 2	1.000 0	1.000 0
Malaysian	0.034 2	0.000 0	0.000 0	0.000 0	0.001 0	0.010 8	0.000 0	0.000 0	0.000 0	0.000 0	0.000 0	0.000 0	0.000 0
Russian	0.185 7	0.003 9	0.253 2	0.013 7	0.320 6	0.277 6	1.000 0	0.928 6	0.271 8	0.339 2	1.000 0	0.472 1	0.296 2
Shaanxi Han	0.779 1	0.303 0	0.833 8	0.696 0	0.999 0	0.154 5	0.772 2	0.962 9	1.000 0	1.000 0	0.893 5	0.336 3	0.371 5
Tibetan	1.000 0	0.000 0	0.000 0	0.000 0	0.004 9	0.029 3	0.004 9	0.000 0	0.006 8	0.056 7	0.002 0	0.001 0	0.000 0

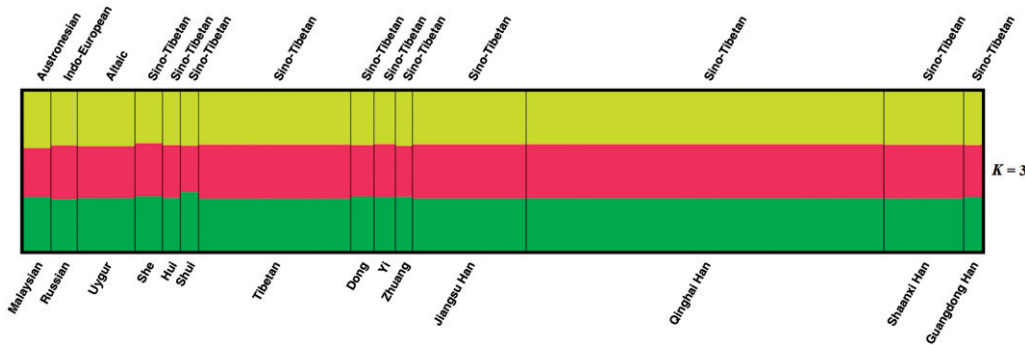


Figure 1. Cluster analysis by structure assuming $K=3$. Population names are labeled beneath the plot and population languages above the plot.

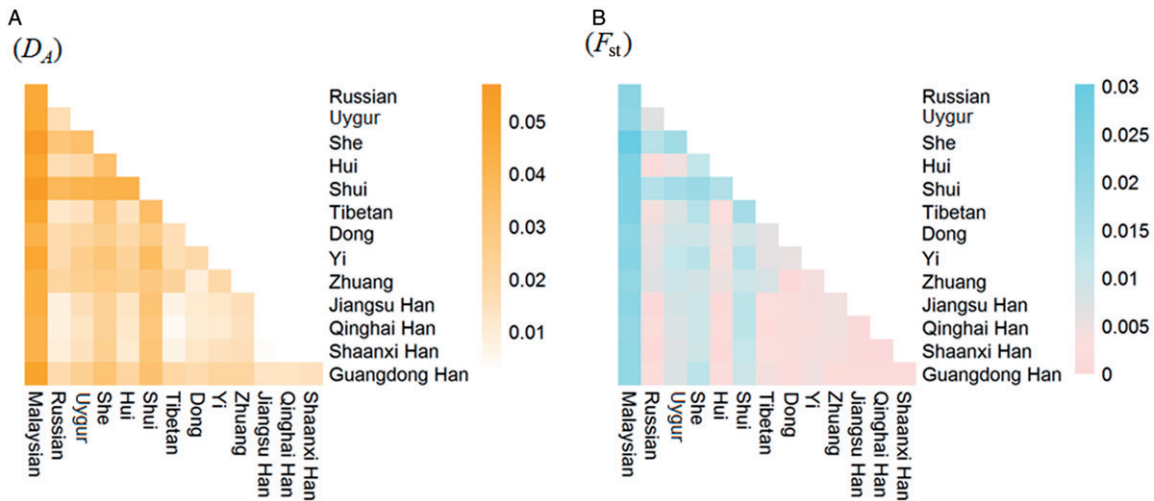


Figure 2. The genetic distance analysis among Qinghai Han and other reference populations. (A) Heatmaps of genetic distance (D_A) and (B) fixation index (F_{st}) of the Qinghai Han population and other reference populations.

genetic distances were observed between the Qinghai Han and Shaanxi Han and Jiangsu Han populations, while the largest distance was seen with the Malaysian (0.0422). Similar results can be discerned from Figure 2(B), which indicates that the Qinghai Han have close relationships with Shaanxi Han and Jiangsu Han populations.

PCA and MDS analysis of the Qinghai Han population and other reference populations

PCA of the Qinghai Han population and 13 other reference groups was performed according to the normalized allelic frequencies of the shared STR loci. As shown in Figure 3(A), the first two principal components contributed to 47.06% of the total

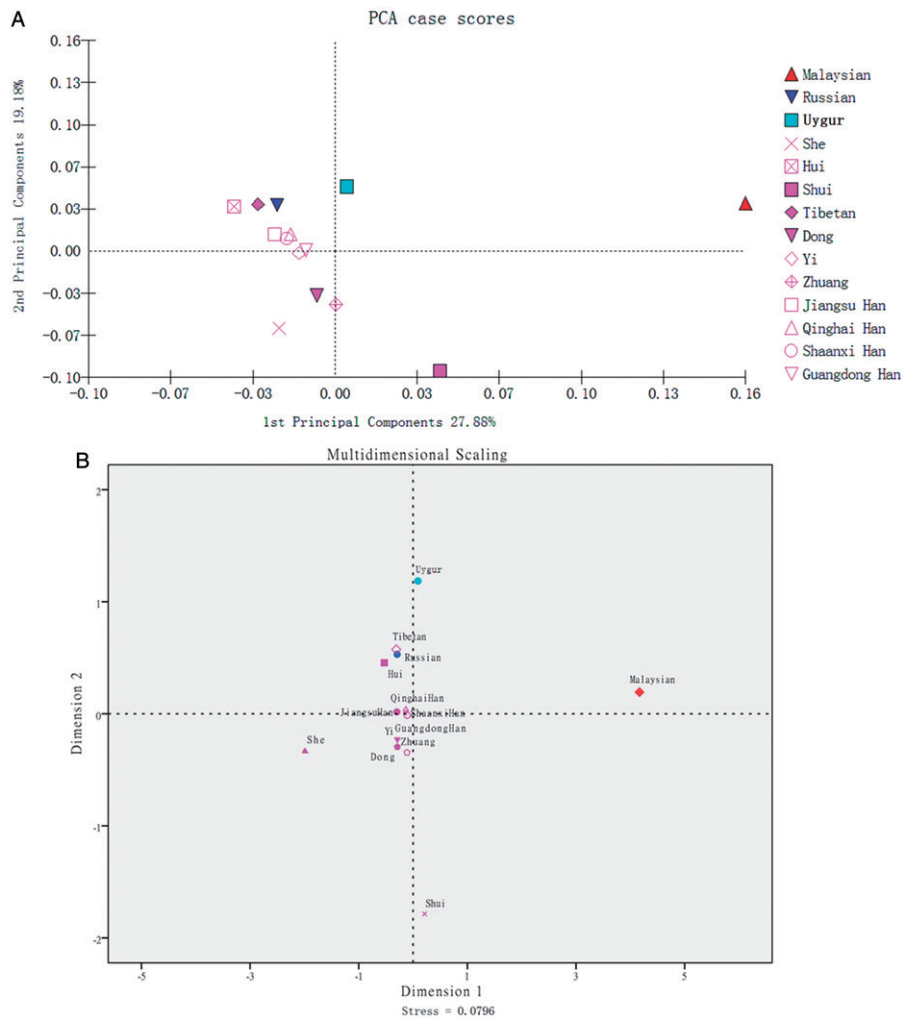


Figure 3. Principal components analysis (A) and multidimensional scaling analysis (B) of the Qinghai Han population and 13 other groups.

variance, with the first and second components accounting for 27.88% and 19.18%, respectively. In Figure 3(A), the Qinghai Han group is clustered in the upper left quadrant, near the centre, close to the Shaanxi Han and Jiangsu Han groups. Similarly, most Han populations, including the Qinghai Han population, formed one cluster, which is located in the centre of the MDS plot. The proximity of these Han populations indicates their similar genetic components.

Phylogenetic reconstructions of the Qinghai Han population and other Chinese populations

Phylogenetic analysis was used to explore genetic relationships between the Qinghai Han group and other populations, as shown in Figure 4. The Neighbor-Joining tree method (Figure 4(A)) demonstrated that the Qinghai Han population formed a sub-branch of the tree with Shaanxi Han and Jiangsu Han populations. Another phylogenetic tree (Figure 4(B)) also showed that the Qinghai

Han population was close to Shaanxi Han and Jiangsu Han populations. These phylogenetic results are consistent with the findings of the inter-population differentiation study, genetic distance analysis, PCA and MDS described above, which likely reflects similar genetic distributions of Qinghai Han, Shaanxi Han and Jiangsu Han populations.

Conclusion

The present study indicated that the loci examined were highly polymorphic in the studied Qinghai Han population. We also found that the Qinghai Han population has the close genetic relationships with Shaanxi Han and Jiangsu Han populations. These results suggest that the loci can be used for paternity testing and forensic human identification, and could also provide information about the genetic relationships between Qinghai Han and other groups.

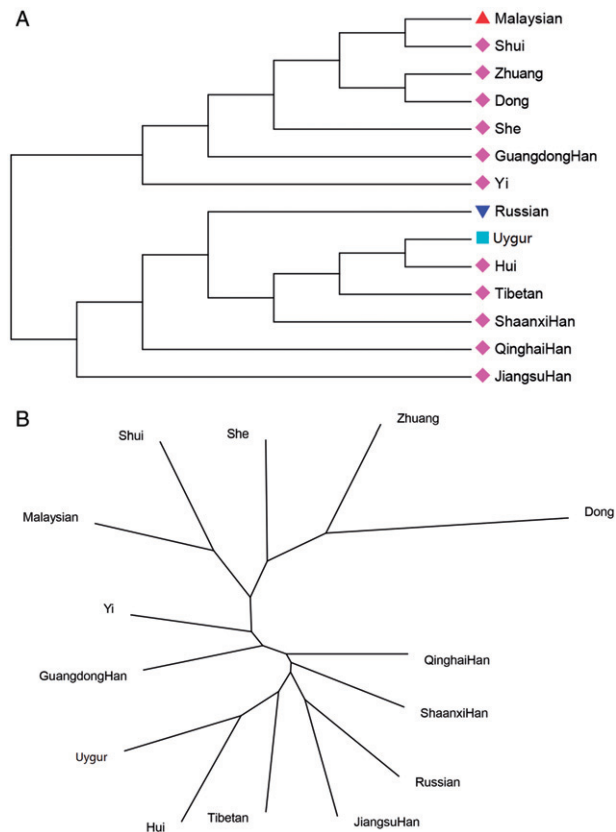


Figure 4. Phylogenetic reconstructions of Qinghai Han population and other 13 populations based on allelic frequencies of the same STR loci. The Neighbor-Joining tree of these 14 populations was constructed using MEGA software (A); another phylogenetic tree was plotted using PHYLIP software (B).

Compliance with ethical standards

The experimental protocol in the present research involving human blood samples conformed to the human and ethical research principles of Xi'an Jiaotong University Health Science Center.

Disclosure statement

No conflict of interest was declared by the authors.

Funding

This project was supported by the Fundamental Research Funds for the Central Universities, China [grant number xj2018165] and the Key Project for Science Research and Development of Shaanxi Province (2018SF-119).

References

- [1] Seah LH, Jeevan NH, Othman MI, et al. STR data for the AmpFI STR identifier loci in three ethnic groups (Malay, Chinese, Indian) of the Malaysian population. *Forensic Sci Int.* 2003;138:134. doi: [10.1016/j.forsciint.2003.09.005](https://doi.org/10.1016/j.forsciint.2003.09.005).
- [2] Yan JW, Shen CM, Li YZ, et al. Genetic analysis of 15 STR loci on Chinese Tibetan in Qinghai Province. *Forensic Sci Int.* 2007;169:E3–E6. doi: [10.1016/j.forsciint.2007.02.008](https://doi.org/10.1016/j.forsciint.2007.02.008).

- [3] Yuan L, Ou Y, Liao QX, et al. Population genetics analysis of 38 STR loci in the She population from Fujian Province of China. *Legal Med.* 2014;16:314–318. doi: [10.1016/j.legalmed.2014.05.008](https://doi.org/10.1016/j.legalmed.2014.05.008).
- [4] Zhu BF, Wang ZY, Wu QJ, et al. Genetic analysis of 15 STR loci of Chinese Uigur ethnic population. *J Forensic Sci.* 2005;50:1235–1236.
- [5] Zhu BF, Shen CM, Wu QJ, et al. Population data of 15 STR loci of Chinese Yi ethnic minority group. *Legal Med (Tokyo).* 2008;10:220–224. doi: [10.1016/j.legalmed.2007.12.004](https://doi.org/10.1016/j.legalmed.2007.12.004).
- [6] Deng W, Yunshao HE, Gao J, et al. Polymorphisms of fifteen short tandem repeat loci in Ningxia Hui ethnic population. *Shanghai J Med Lab Sci.* 2005;20:562–564.
- [7] Li Y, Wang SB, Liu C, et al. [Genetic studies of 13 loci in Guangdong Han population by two multiplex PCR systems and its forensic application]. *Fa Yi Xue Za Zhi.* 2001;17:82–85.
- [8] Zhu BF, Yang G, Shen CM, et al. Population genetic analysis of 15 autosomal STR loci in the Russian population of northeastern Inner-Mongolia, China. *Mol Biol Rep.* 2010;37:3889–3895. doi: [10.1007/s11033-010-0045-y](https://doi.org/10.1007/s11033-010-0045-y).
- [9] Pan M, Chen Z, Ding X. A genetic polymorphism analysis on fifteen short tandem repeat loci among Han population in Jiangsu district. *Jiangsu Med J.* 2012;38:561–564.
- [10] Wu YM, Zhang XN, Zhou Y, et al. Genetic polymorphisms of 15 STR loci in Chinese Han population living in Xi'an city of Shaanxi Province. *Forensic Sci Int Genet.* 2008;2:e15–e18. doi: [10.1016/j.fsigen.2007.11.003](https://doi.org/10.1016/j.fsigen.2007.11.003).
- [11] Wang P, Zhu SJ, Liu C, et al. [Genetic analysis of 15 STR loci of 3 Chinese minority groups from Guangxi Province]. *Zhong Guo Fa Yi Xue Za Zhi.* 2003;18:299–300. Chinese.
- [12] Zhai D, Chen LF, Wu DL, et al. [Polymorphisms of 15 STR loci in Yunnan Zhuang ethnic population]. *Zhong Guo Fa Yi Xue Za Zhi.* 2015;30:509–510. Chinese.
- [13] Wang W, Sun QF, Zhang T, et al. [Genetic polymorphisms of 18 STR loci in Dai population of Yunnan and Dong population of Guangxi]. *Shan Xi Yi Ke Da Xue Xue Bao.* 2015;46:991–995. Chinese.
- [14] Walsh PS, Metzger DA, Higuchi R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques.* 1991;10:506–513.
- [15] Rousset F, Ccedil, OIS. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour.* 2008;8:103–106.
- [16] Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 2010;10:564–567. doi: [10.1111/j.1755-0998.2010.02847.x](https://doi.org/10.1111/j.1755-0998.2010.02847.x).
- [17] Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155:945–959.
- [18] Ota T. Program DISPAN: Genetic distance and phylogenetic analysis. 1993.
- [19] Kovach WL. MVSP: A Multivariate Statistical Package for Windows, version 3.1. Pentraeth, Wales, UK: Kovach Computing Services; 2007.
- [20] Tamura K, Peterson D, Peterson N, et al. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol.* 2011;28:2731–2739. doi: [10.1093/molbev/msr121](https://doi.org/10.1093/molbev/msr121).