



## Letter to the Editor

**Allele frequencies of 15 STRs in five ethnic groups (Han, Gelao, Jing, Shui and Zhuang) in South China**

Dear Editor,

Chinese consists of 56 ethnic groups, including a largest ethnic group of Han, and other 55 minor ethnic groups. Among these groups, there are significant differences in language, culture, custom and history. Some studies have been carried out on the genetic relationships among the different ethnic groups [1–3].

Ethnic minorities Gelao, Jing, Shui and Zhuang are the major ethnic groups in South China. Gelao ethnic group, an aboriginal population residing in Southwest China, has a population of 579 thousand. The Jing is the smallest Chinese ethnic minority with about 21 thousand population, mostly the people of Jing inhabits in the Guangxi Province. Shui ethnic group has a population of 400 thousand living in the Sandu Shui Autonomous County of Guizhou Province. The populations of Gelao, Jing and Shui are living in relatively isolated environments, and seldom contact with other nationalities. Zhuang ethnic group, the second largest official ethnic group in China, currently has a population of nearly 17 million mainly living in Guangxi, Guangdong and Yunnan Provinces of South China. This paper reports the genetic relationships among the five ethnic groups based on the data of STRs.

The blood samples of 2022 unrelated individuals from five ethnic groups were collected. They include Han ( $N = 1000$ ), Gelao ( $N = 316$ ), Jing ( $N = 312$ ), Shui ( $N = 204$ ) and Zhuang ( $N = 190$ ) from the provinces of Guangxi and Guangdong of South China. The locations of the two provinces are shown in the map of China (Fig. 1).

Genomic DNA was extracted from each of the blood samples by Chelex-100 method (Bio-rad Company) as described by Walsh et al. [4]. DNA samples were amplified using AmpFISTR<sup>®</sup>.

Identifiler<sup>™</sup> commercial kits (Applied Biosystems) [5]. PCR amplifications were carried out according to the manufacturer's instructions. Amplified products were carried out on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems) according to the manufacturer's recommendations.

Fragment size determination was carried out using Genescan v3.7 software. The observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ) and  $p$ -values of exact test for Hardy–Weinberg equilibrium ( $p$ ) were calculated using Genepop v3.4 software [6]. Polymorphism information content (PIC), power of discrimination (PD) and power of exclusion (PE) were calculated using Powerstats v1.2 software (Promega Corp.) ([http://www.promega.com/genetici-dtools/pow\\_erstats/](http://www.promega.com/genetici-dtools/pow_erstats/)). Two methods which include analysis of molecular variance (AMOVA) and population pairwise genetic distances ( $F_{st}$ ) were performed with Arlequin v3.1 software [7]. Neighbor-joining tree was built from Nei genetic distance matrix using the R software [8]. A new principal component analysis method (multiple co-inertia analysis) was performed among the ethnic groups using the R software adegent package [9].

The allele frequencies and six statistic parameters for the 15 STR loci of the five ethnic groups in South China were shown in Table 1. The Nei genetic distances between the populations of the five ethnic groups were shown in Table 2. Two other ethnic groups from references (Tujia ethnic group [10] and Monggol ethnic group [11]) had been added in the neighbor-joining phylogenetic tree (Fig. 2) for comparisons. The principal components and synthetic variables were shown in Table 3, and the principal component analysis diagram was shown (Fig. 3).

All the STR loci analyzed reach the Hardy–Weinberg equilibrium ( $p > 0.05$ ) in this study except for the D13S317 ( $p = 0.0112$ ) locus in Gelao ethnic group. For the five ethnic groups, the combined power of discrimination and the combined power of exclusion for the 15 STR loci were higher than 0.99999 (Table 2A). Based on heterozygosity ( $H_e$ ) and polymorphic informative content (PIC), FGA may be the most informative locus.

Nei genetic distances were calculated based on the gene frequencies for 15 STRs to find out the genetic relationships among the five ethnic groups. The result (Table 2) showed an obvious separation between (Shui), (Jing) groups and (Zhuang, Han, Gelao) groups. The Nei's genetic distances between populations of Han, Jing and Shui were more than 0.03 and Nei's genetic distances between populations of Han, Zhuang and Gelao were less than 0.02. Originally, it was found that Zhuang was closer to Han than other groups (Fig. 2A). After adding two other ethnic groups (Monggol and Tujia), it was found that Zhuang and Tujia were closer to Han than any other groups (Fig. 2B). The principal



Fig. 1. Geographic distribution of the focused populations in South China. Guangdong (shadow on the right side): Han; Guangxi (shadow on the left side): Gelao, Jing, Shui and Zhuang.

**Table 1**  
Allele frequencies and six parameters of 15 STR.

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
In Han group (N=1000)															
6						0.0967						0.002			
7			0.0025	0.0045		0.2771	0.0015					0.0015		0.0335	
8	0.0015		0.1433	0.0025		0.0696	0.3000	0.0030				0.5420		0.0035	
9	0.0005		0.0686	0.0356		0.4740	0.1455	0.2560				0.1115		0.0696	
9.3						0.0321									
10	0.1365		0.1723	0.2415	0.0005	0.0501	0.1405	0.1250				0.0270		0.2147	
11	0.1030		0.3347	0.2325		0.0005	0.2365	0.2635		0.0030		0.2915	0.0050	0.3073	
12	0.1235		0.2265	0.3898	0.0005		0.1370	0.2345		0.0510		0.0230	0.0438	0.2392	
12.2										0.0050					
13	0.1940		0.0496	0.0807	0.0015		0.0310	0.1055		0.2780	0.0020	0.0015	0.1675	0.1221	
13.2			0.0025							0.0400					
14	0.1850			0.0105	0.0420		0.0080	0.0125		0.2525	0.2750		0.2123	0.0100	
14.2										0.0910					
15	0.1660			0.0025	0.3435					0.0700	0.0285		0.1816		
15.2										0.1525					
16	0.0790				0.3165				0.0150	0.0145	0.1780		0.1358		0.0010
16.2										0.0370					
17	0.0110				0.2365				0.0686	0.0010	0.2140		0.0916		0.0020
17.2										0.0045					
18					0.0540				0.1016		0.1965		0.0518		0.0280
19					0.0050				0.1937		0.0925		0.0397		0.0596
20									0.1201		0.0135		0.0221		0.0455
21									0.0335				0.0166		0.1271
21.2															0.0025
22									0.0526				0.0196		0.1662
22.2															0.0050
23									0.1847				0.0076		0.2242
23.2															0.0085
24									0.1627				0.0050		0.1582
24.2															0.0095
25									0.0591						0.0911
25.2															0.0035
26									0.0085						0.0420
26.2															0.0030
27		0.0030													0.0170
28		0.0460													0.0060
28.2		0.0085													
29		0.2590													
29.2		0.0020													
30		0.2685													
30.2		0.0100													
31		0.0980													
31.2		0.0670													
32		0.0365													
32.2		0.1460													
33		0.0045													
33.2		0.0435													
34.2		0.0075													
Ho	0.8520	0.8010	0.7830	0.7210	0.7050	0.7050	0.7110	0.8140	0.8460	0.8280	0.7940	0.6120	0.8590	0.7660	0.8640
He	0.8520	0.8190	0.7710	0.7360	0.7260	0.6720	0.7920	0.7850	0.8650	0.8150	0.8000	0.6060	0.8620	0.7820	0.8620
P	0.6109	0.1939	0.3412	0.0752	0.7560	0.4275	0.1143	0.0608	0.2070	0.4617	0.0849	0.2936	0.0969	0.7525	0.4892
PD	0.9580	0.9470	0.9170	0.8820	0.8750	0.8510	0.9240	0.9160	0.9660	0.9410	0.9290	0.7800	0.9630	0.9200	0.9670
PE	0.6990	0.6010	0.5680	0.4610	0.4360	0.4360	0.6200	0.6250	0.6940	0.6520	0.5880	0.3060	0.7130	0.5380	0.7230
PIC	0.8300	0.8000	0.7500	0.6800	0.6700	0.6400	0.7600	0.7500	0.8500	0.7900	0.7700	0.5500	0.8400	0.7500	0.8500
In Gelao group (N=316)															
6						0.0696									
7				0.0411		0.2753	0.0032							0.0190	
8			0.1930			0.0570	0.3671					0.5411			
9			0.0380	0.0063		0.5475	0.1139	0.1741				0.0823		0.0728	
9.2															
9.3						0.0317									
10	0.2089		0.1804	0.2500		0.0190	0.1582	0.1108				0.0032		0.2025	
11	0.1171		0.4177	0.2405			0.1867	0.2658				0.3513		0.3006	
12	0.0728		0.1519	0.3861			0.1298	0.2089				0.0222	0.0224	0.2722	
12.2	0.0000									0.0190					
13	0.1646		0.0190	0.0665			0.0411	0.1994				0.2563		0.1795	0.1298
13.2	0.0000									0.0506					
14	0.1741			0.0095	0.0253			0.0253		0.2437	0.2690		0.2212	0.0032	
14.2	0.0000									0.1203					
15	0.1772				0.3576			0.0158		0.1013	0.0348		0.1667		
15.2	0.0000									0.1487					
16	0.0665				0.2943				0.0190	0.0032	0.2184		0.1699		0.0032
16.2	0.0000									0.0127					
17	0.0095				0.2184				0.0475		0.1456		0.0545		



Table 1 (Continued)

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
33.2		0.0385													
34.2		0.0096													
Ho	0.8910	0.8010	0.7820	0.7120	0.7440	0.7240	0.7500	0.7950	0.8330	0.8080	0.7180	0.5710	0.8330	0.7560	0.8780
He	0.8560	0.8410	0.7790	0.7220	0.6980	0.6950	0.7760	0.7780	0.8590	0.8620	0.7930	0.6100	0.8410	0.7860	0.8640
P	0.7815	0.5218	0.0937	0.6550	0.7413	0.2137	0.0732	0.6920	0.0767	0.1116	0.0854	0.3728	0.0646	0.3007	0.8142
PD	0.9560	0.9520	0.8600	0.8780	0.8320	0.8440	0.9140	0.9100	0.9580	0.9610	0.9250	0.7890	0.9450	0.9200	0.9600
PE	0.7770	0.6010	0.3500	0.4620	0.4990	0.4670	0.5100	0.5900	0.6620	0.6130	0.4570	0.2570	0.6620	0.5210	0.7510
PIC	0.8400	0.8200	0.6400	0.6800	0.6400	0.6400	0.7400	0.7400	0.8400	0.8400	0.7600	0.5500	0.8200	0.7500	0.8500
In Shui group (N=204)															
6						0.1618									
7				0.0147		0.2941								0.0196	
8			0.1814			0.0294	0.3922				0.3971				
9			0.0686	0.0049		0.4216	0.1078	0.3677			0.1177			0.0098	
9.3						0.0441									
10	0.0931		0.1177	0.2304		0.0490	0.1716	0.0343				0.0098		0.2108	
11	0.1078		0.4118	0.2500			0.2451	0.2647				0.3578		0.2892	
12	0.2108		0.1814	0.3922			0.0490	0.2451		0.0294		0.1177	0.0343	0.3137	
13	0.3186		0.0392	0.0931			0.0294	0.0833		0.3775			0.0588	0.1569	
13.2										0.0245					
14	0.0686			0.0147	0.0441		0.0049	0.0049		0.1275	0.2500		0.2451		
14.2					0.0000					0.0539					
15	0.1422				0.2990					0.0588	0.0539		0.1961		
15.2					0.0000					0.2745					
16	0.0294				0.3333				0.0049		0.1078		0.1961		
17	0.0294				0.2549				0.0784	0.0539	0.2108		0.1177		
18					0.0490				0.0539		0.1912		0.0490		0.0735
19					0.0196				0.1912		0.1029		0.0196		0.0343
20									0.0931		0.0686		0.0637		0.0294
21									0.0392		0.0147				0.1716
22									0.0539						0.1912
22.2															0.0098
23									0.1667				0.0147		0.1128
23.2															0.0049
24									0.2451						0.0588
24.2															0.0147
25									0.0245						0.1422
26									0.0490						0.1275
27													0.0049		0.0294
28		0.0245													
29		0.2206													
30		0.1520													
30.2		0.0098													
31		0.2010													
31.2		0.0980													
32		0.0098													
32.2		0.1569													
33.2		0.1029													
34.2		0.0245													
Ho	0.7750	0.8240	0.7450	0.7550	0.7550	0.6180	0.7200	0.6760	0.8730	0.7060	0.8330	0.6960	0.8140	0.7250	0.8630
He	0.8110	0.8460	0.7480	0.7250	0.7330	0.7080	0.7470	0.7300	0.8550	0.7590	0.8310	0.6900	0.8420	0.7520	0.8770
P	0.2561	0.6793	0.5311	0.2777	0.0529	0.1672	0.1770	0.8437	0.3376	0.1512	0.7002	0.5639	0.4132	0.4539	0.2649
PD	0.9330	0.9510	0.8940	0.8550	0.8650	0.8660	0.8950	0.8850	0.9510	0.9040	0.9260	0.8190	0.9500	0.8880	0.9630
PE	0.5530	0.6430	0.5010	0.6250	0.5180	0.3130	0.4640	0.3930	0.6620	0.4370	0.8200	0.2340	0.6250	0.4690	0.7200
PIC	0.7800	0.8200	0.7100	0.6700	0.6800	0.6600	0.7100	0.6800	0.8300	0.7200	0.8000	0.6300	0.8200	0.7000	0.8600
In Zhuang group (N=190)															
6						0.0684									
7			0.0053	0.0158		0.2790	0.0105					0.0053		0.0263	
8			0.1158			0.0632	0.2790					0.5632			
9			0.0842	0.0211		0.4316	0.1684	0.2211				0.1053		0.0474	
9.2			0.0053												
9.3						0.0526									
10	0.1737		0.1895	0.2316		0.1053	0.1579	0.1053				0.0316		0.1632	
11	0.1737		0.4105	0.2421			0.2368	0.2790				0.2842	0.0053	0.3790	
12	0.1263		0.1526	0.4158			0.1053	0.2632			0.0368	0.0105	0.0790	0.2421	
12.2											0.0105				
13	0.1263		0.0316	0.0474			0.0421	0.1263			0.2895		0.0790	0.1316	
13.2											0.0211				
14	0.1632		0.0053	0.0263	0.0160			0.0053			0.2105	0.3105	0.2158	0.0105	
14.2											0.1053				
15	0.1684				0.2819						0.0632	0.0316	0.2632		
15.2											0.1526				
16	0.0421				0.3830				0.0053		0.0211	0.1316	0.1316		
16.2											0.0790				
17	0.0158				0.2447				0.0737		0.1947		0.0895		
17.2											0.0105				

**Table 1** (Continued)

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
18	0.0105				0.0638				0.0895		0.2000		0.0421		0.0211
19					0.0106				0.2421		0.1158		0.0368		0.0579
20									0.0790		0.0105		0.0316		0.0421
21									0.0316		0.0053		0.0105		0.1263
22									0.0579				0.0053		0.1947
22.2															0.0158
23									0.1842				0.0053		0.1368
24									0.1632				0.0053		0.2000
24.2															0.0158
25									0.0579						0.1053
25.2															0.0105
26									0.0158						0.0368
26.2															0.0053
27															0.0211
28		0.0474													0.0105
29		0.2684													
30		0.2737													
30.2		0.0263													
31		0.0947													
31.2		0.0632													
32		0.0526													
32.2		0.1421													
33		0.0053													
33.2		0.0211													
34.2		0.0053													
Ho	0.8420	0.7680	0.7890	0.7680	0.6490	0.7890	0.7680	0.7680	0.8630	0.8000	0.7890	0.5930	0.8530	0.6950	0.8000
He	0.8550	0.8180	0.7550	0.7150	0.7130	0.7170	0.8040	0.7810	0.8580	0.8290	0.7980	0.5790	0.8460	0.7550	0.8730
P	0.6624	0.1669	0.1726	0.4515	0.8482	0.1484	0.2273	0.9210	0.3958	0.8457	0.7345	0.4706	0.3581	0.5844	0.1787
PD	0.9530	0.9330	0.8940	0.8590	0.8680	0.8470	0.9260	0.9120	0.9560	0.9480	0.9240	0.7730	0.9490	0.9000	0.9620
PE	0.6790	0.5420	0.5800	0.5420	0.3540	0.5800	0.5420	0.5420	0.7210	0.5990	0.5800	0.2660	0.7000	0.4200	0.5990
PIC	0.8300	0.7900	0.7200	0.6600	0.6600	0.6700	0.7700	0.7400	0.8400	0.8000	0.7600	0.5300	0.8200	0.7100	0.8500

Ho: observed heterozygosity; He: expected heterozygosity; PIC: polymorphism information content; PD: power of discrimination; PE: power of exclusion; P: the p-values of exact test of Hardy–Weinberg equilibrium.

component analysis showed that the 15 STR loci of five components contribute to the variation. The first two components account for 82.39% of the total sample variation (Table 3A). As shown in Fig. 2A, Zhuang was found to be closer to Han population than other three populations (Gelao, Jing, Shui). The first principal axis of the analysis differentiates the three groups (Han, Zhuang, Gelao) and the other two groups (Jing, Shui) from the five STRs (D16S539, D21S11, D19S433, TH01, and D8S1179), while the second principal component differentiates the Han, Zhuang two groups and Gelao group from the four STRs (FGA, D13S317, TPOX and D18S51). This result also supported the conclusion above.

**Table 2A**

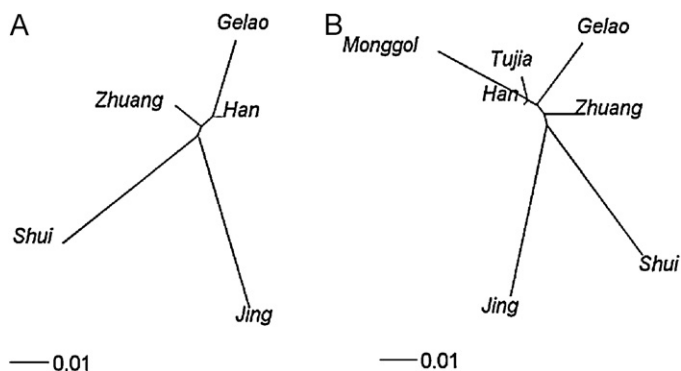
The information of the five ethnic groups in South China.

Ethnic groups	Han	Gelao	Jing	Shui	Zhuang
Region	Guangdong	Guangxi	Guangxi	Guangxi	Guangxi
The number of population	1000	316	312	204	190
Combined power of exclusion	0.999998	0.999998	0.999998	0.999995	0.999996
Combined matching probability	7.93E–18	5.83E–17	3.94E–17	1.44E–16	4.74E–17

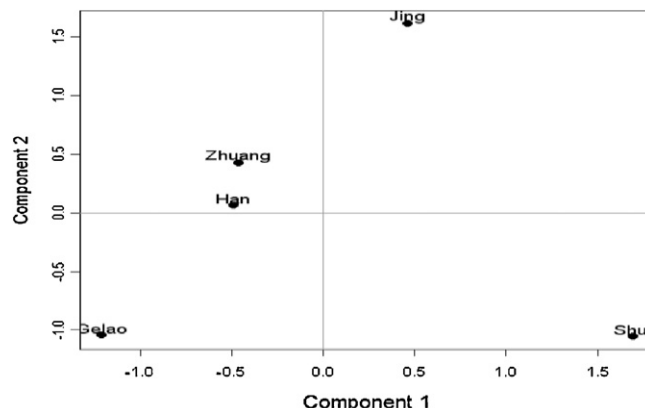
**Table 2B**

Pairwise Nei genetic distance coefficients between the five ethnic groups.

	Han	Gelao	Zhuang	Jing	Shui
Han	0.0000				
Gelao	0.0186	0.0000			
Zhuang	0.0102	0.0201	0.0000		
Jing	0.0393	0.0497	0.0414	0.0000	
Shui	0.0301	0.0376	0.0339	0.0645	0.0000



**Fig. 2.** Neighbor-joining tree based on Nei genetic distance calculated between five populations (A) and seven populations (B).



**Fig. 3.** Principal component analysis based on the allele frequencies for 15 STRs.

**Table 3A**

The five principal components for the 15 STR loci.

	PC1	PC2	PC3	PC4	PC5
D8S1179	0.4970	0.1880	0.0238	0.0480	0.0001
D21S11	0.5964	0.2953	0.0576	0.0117	0.0001
D7S820	0.1734	0.2129	0.2315	0.0410	0.0001
CSF1PO	0.1597	0.2163	0.1078	0.0832	0.0005
D3S1358	0.2171	0.0661	0.0793	0.0256	0.0001
TH01	0.5493	0.1104	0.1144	0.0126	0.0006
D13S317	0.3541	0.4196	0.1040	0.0132	0.0000
D16S539	0.7362	0.1003	0.0813	0.0307	0.0001
D2S1338	0.4390	0.1988	0.2208	0.0251	0.0000
D19S433	0.5711	0.2273	0.0847	0.0191	0.0000
vWA	0.3816	0.2917	0.2070	0.0202	0.0011
TPOX	0.4211	0.4133	0.0152	0.0002	0.0004
D18S51	0.2962	0.3290	0.0811	0.0665	0.0000
D5S818	0.2933	0.2268	0.1471	0.0156	0.0002
FGA	0.1704	0.4763	0.0380	0.0392	0.0087

PC: principal component.

**Table 3B**

The synthetic variables for the five ethnic groups.

	SynVar1	SynVar2	SynVar3	SynVar4	SynVar5
Han	-0.4882	0.0615	0.2541	-1.9218	0.0024
Gelao	-1.2109	-1.0447	-1.0020	0.6621	0.2931
Zhuang	-0.4596	0.4263	1.6863	0.8737	-2.1221
Jing	0.4654	1.6141	-1.0382	0.3165	0.5181
Shui	1.6933	-1.0573	0.0998	0.0695	-0.3776

This paper follows the ISFG recommendations [12] and the guidelines for publication of population data requested by the journal [13].

### Acknowledgment

This work was supported by National Natural Science Foundation of China (No. 91130009).

### References

- [1] W. Wu, L. Pan, H. Hao, X. Zheng, J. Lin, D. Lu, Population genetics of 17 Y-STR loci in a large Chinese Han population from Zhejiang Province, Eastern China, *Forensic Sci. Int. Genet.* 5 (1) (2011) 11–13.
- [2] C. Li, L. Li, Z. Zhao, Y. Lin, T. Que, Y. Liu, J. Xue, Genetic polymorphism of 17 STR loci for forensic use in Chinese population from Shanghai in East China, *Forensic Sci. Int. Genet.* 3 (4) (2009) 117–118.
- [3] Y. Zhu, S. Lu, Z. Xie, Y. Chen, J. You, Genetic analysis of 15 STR loci in the population of Zhejiang Province (Southeast China), *Forensic Sci. Int. Genet.* 3 (4) (2009) 139–140.
- [4] P.S. Walsh, D.A. Metzger, R. Higuchi, Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material, *Biotechniques* 10 (4) (1991) 506–513.
- [5] S. Mitra, On Nei and Roychoudhury's sampling variances of heterozygosity and genetic distance, *Genetics* 80 (1) (1975) 223–226.
- [6] M. Raymond, F. Rousset, GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism, *J. Hered.* 86 (3) (1995) 248–249.
- [7] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0): an integrated software package for population genetics data analysis, *Evol. Bioinform. Online* 1 (2005) 47–50.
- [8] M. Nei, A.K. Roychoudhury, Sampling variances of heterozygosity and genetic distance, *Genetics* 76 (2) (1974) 379–390.
- [9] T. Jombart, Adegenet: a R package for the multivariate analysis of genetic markers, *Bioinformatics* 24 (11) (2008) 1403–1405.
- [10] S.X. Deng, Genetic polymorphism of 15 STR loci in Chongqing Tujia population, *J. Chongqing Med. Univ.* 34 (2009) 1499–1500.
- [11] Z.H. Wang, B. Lu, Genetic polymorphism of 15 STR loci in Qinghai Monggol population, *Chin. J. Forensic Med.* 21 (2006) 18–19.
- [12] B. Olaisen, W. Bär, B. Brinkmann, B. Budowle, A. Carracedo, P. Gill, P. Lincoln, W.R. Mayr, S. Rand, DNA recommendations 1997 of the International Society for Forensic Genetics, *Vox Sang.* 74 (1) (1998) 61–63.

- [13] A. Carracedo, J.M. Butler, L. Gusmao, W. Parson, L. Roewer, P.M. Schneider, Publication of population data for forensic purposes, *Forensic Sci. Int. Genet.* 4 (2010) 145–147.

Linlin Yang<sup>a,b</sup><sup>a</sup>Department of Pediatrics, Guangdong Women and Children's Hospital, Guangzhou 510010, China<sup>b</sup>Life Sciences School, Sun Yat-sen University, Guangzhou 510275, China

Yanchao Zhao

Changhui Liu

Life Sciences School, Sun Yat-sen University, Guangzhou 510275, China

Daniel Wai Tin Chan

Department of Building Services Engineering, The Hong Kong Polytechnic University, Kowloon 999077, Hong Kong

Manhim Chan

Department of Architecture, Cardiff University, Cardiff, Wales, United Kingdom

Miao He\*

Life Sciences School, Sun Yat-sen University, Guangzhou 510275, China

\*Corresponding author. Tel.: +86 20 84110036  
E-mail address: lsshem@mail.sysu.edu.cn (M. He)

21 February 2012