Genetic Analysis of 15 STR Loci in Chinese Han Population from West China

Ya-Jun Deng[1,2#*], Jiang-Wei Yan[1#], Xiao-Guang Yu[2], Yuan-Zhe Li[1], Hao-Fang Mu[2], Yan-Qing Huang[2], Xiao-Tie Shi[2], and Wei-Min Sun[2]

[1] *Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 101300, China;* [2] *Center of Forensic Sciences, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 101300, China.*

Allele frequencies for 15 short tandem repeat (STR) loci (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818, and FGA) were obtained from 7,636 unrelated individuals of Chinese Han population living in Qinghai and Chongqing, China. Totally 206 alleles were observed, with the corresponding allele frequencies ranging from 0.0001–0.4982. Chi-square test showed that all of the STR loci agreed with the Hardy-Weinberg equilibrium. We also compared our data with previously published population data of other ethnics or areas. The results are valuable for human identification and paternity testing in Chinese Han population.

Key words: forensic science, DNA typing, population genetic analysis, short tandem repeats, Chinese Han population

Introduction

Short tandem repeat (STR) loci have been widely used as human identification markers in forensic sciences. In this study we obtained and analyzed allele frequencies for 15 STR loci in Chinese Han population from West China, with the aim of setting up a primary forensic DNA database and acquiring more reliable statistic data for human identification and paternity testing.

Results and Discussion

Allele frequencies and statistical parameters regarding the 15 STR loci in Chinese Han population were shown in Table 1. A total of 206 alleles and 930 genotypes of the 15 STR loci were detected, with the corresponding allele frequencies ranging from 0.0001–0.4982. The combined probability of exclusion (P_e), power of discrimination (PD), and probability of match (P_m) for all the 15 STR loci were 0.999999, 0.9999999999, and 5.85×10^{-11}, respectively. The observed and expected genotype frequencies were evaluated by χ^2 test and the heredity of all STR loci followed the Hardy-Weinberg equilibrium ($P>0.05$).

We also compared these data with previous population data observed at the same 15 STR loci by using the method of χ^2 test after Bonferroni's correction of SPSS 11.5 (www.spss.com). As a result, no significant differences were observed between our data and the Chinese Han population living in Guangdong Province (*1*), Hui (*2*) and Uigur (*3*) ethnic groups of China, Korean population (*4*), and Kosovo population (*5*). However, we found that there were significant differences between our data and the Chinese Miao population (*6*) at locus vWA (P=0.027), Maghreb population (*7*) at locus D13S317 (P<0.001), Polish population (*8*) at loci D8S1179 (P=0.043) and D7S820 (P=0.001), and Otomi population (*9*) at loci vWA, D7S820, CSF1PO, FGA, and D8S1179 (P<0.001), respectively. The results of the present study are valuable for human identification and paternity testing in Chinese Han population.

Materials and Methods

Blood-stained samples were obtained with informed consent from 7,636 unrelated individuals of Chinese Han population living in Qinghai Province and Chongqing Municipality in West China. Genomic DNA was extracted using the Chelex-100 method (*10*). Fifteen STR loci (D8S1179, D21S11, D7S820,

[#]Contributed equally to this work.
***Corresponding author.**
E-mail: dengyj@genomics.org.cn

Table 1 Allele frequencies and statistical parameters for the 15 STR loci in Chinese Han population

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	DS19S433	vWA	TPOX	D18S51	D5S818	FGA
5	–	–	–	–	–	0.0001	–	–	–	–	–	0.0001	–	–	–
5.3	–	–	–	–	–	0.0001	–	–	–	–	–	–	–	–	–
6	–	–	–	–	–	0.0940	–	0.0001	–	–	–	0.0002	–	0.0001	–
6.3	–	–	–	–	–	0.0001	–	–	–	–	–	–	–	–	–
7	–	–	0.0016	0.0037	–	0.2786	0.0013	0.0001	–	–	–	0.0007	–	0.0203	–
8	0.0012	–	0.1602	0.0020	–	0.0563	0.2692	0.0095	–	–	–	0.5270	0.0001	0.0024	–
8.3	–	–	–	–	–	0.0002	–	–	–	–	–	–	–	–	–
9	0.0004	–	0.0678	0.0522	–	0.4982	0.1360	0.2765	–	0.0002	–	0.1271	0.0003	0.0726	–
9.2	–	–	–	–	–	–	–	–	–	0.0001	–	–	–	–	–
9.3	–	–	–	–	–	0.0435	–	–	–	–	–	–	–	–	–
10	0.1127	–	0.1644	0.2368	–	0.0286	0.1465	0.1239	–	0.0001	0.0001	0.0221	0.0010	0.1958	–
10.2	–	–	–	–	–	–	–	–	–	0.0001	–	–	–	–	–
11	0.0768	–	0.3379	0.2420	0.0001	0.0005	0.2319	0.2544	–	0.0033	–	0.2950	0.0039	0.3296	–
11.2	–	–	–	–	–	–	–	–	–	0.0006	–	–	–	–	–
12	0.1261	–	0.2268	0.3760	0.0007	–	0.1659	0.2205	–	0.0393	0.0003	0.0256	0.0364	0.2316	–
12.2	–	–	–	–	–	–	–	–	–	0.0058	–	–	–	–	–
13	0.2290	–	0.0364	0.0760	0.0014	–	0.0395	0.1001	–	0.2735	0.0020	0.0021	0.1971	0.1363	–
13.2	–	–	–	–	–	–	–	–	–	0.0484	–	–	–	–	–
14	0.1946	–	0.0047	0.0101	0.0430	–	0.0097	0.0142	0.0001	0.2441	0.2367	0.0003	0.2143	0.0098	–
14.2	–	–	–	–	–	–	–	–	–	0.1120	–	–	–	–	–
15	0.1703	–	0.0002	0.0013	0.3566	–	0.0001	0.0008	0.0001	0.0740	0.0296	–	0.1708	0.0013	–
15.2	–	–	–	–	–	–	–	–	–	0.1434	–	–	–	–	–
16	0.0735	–	0.0001	–	0.3223	–	–	0.0001	0.0084	0.0167	0.1870	–	0.1270	0.0001	0.0006
16.2	–	–	–	–	–	–	–	–	–	0.0322	–	–	–	–	–
17	0.0136	–	–	–	0.2010	–	–	–	0.0640	0.0017	0.2492	–	0.0771	0.0001	0.0011
17.2	–	–	–	–	–	–	–	–	–	0.0039	–	–	–	–	0.0001
18	0.0018	–	–	–	0.0692	–	–	–	0.1113	0.0001	0.1887	–	0.0475	–	0.0251
18.2	–	–	–	–	–	–	–	–	–	0.0005	–	–	–	–	–
19	–	–	–	–	0.0054	–	–	–	0.1837	–	0.0884	–	0.0461	–	0.0479
20	–	–	–	–	0.0003	–	–	–	0.1192	–	0.0167	–	0.0336	–	0.0490
21	–	–	–	–	–	–	–	–	0.0304	–	0.0011	–	0.0207	–	0.1047
21.2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.0031
22	–	–	–	–	–	–	–	–	0.0491	–	0.0001	–	0.0136	–	0.1738
22.2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.0067
23	–	–	–	–	–	–	–	–	0.2003	–	–	–	0.0063	–	0.2191
23.2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.0093
24	–	–	–	–	–	–	–	–	0.1560	–	–	–	0.0027	–	0.1835
24.2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.0095
25	–	–	–	–	–	–	–	–	0.0620	–	–	–	0.0011	–	0.1053
25.2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.0039

Table 1 *Continued*

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	DS19S433	vWA	TPOX	D18S51	D5S818	FGA
26	–	0.0003	–	–	–	–	–	–	0.0123	–	–	–	0.0004	–	0.0426
26.2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.0014
27	–	0.0027	–	–	–	–	–	–	0.0025	–	–	–	0.0001	–	0.0097
27.2	–	0.0001	–	–	–	–	–	–	0.0006	–	–	–	–	–	0.0002
28	–	0.0450	–	–	–	–	–	–	0.0001	–	–	–	–	–	0.0026
28.2	–	0.0093	–	–	–	–	–	–	–	–	–	–	–	–	–
29	–	0.2636	–	–	–	–	–	–	–	–	–	–	–	–	0.0007
29.2	–	0.0016	–	–	–	–	–	–	–	–	–	–	–	–	–
30	–	0.2805	–	–	–	–	–	–	–	–	–	–	–	–	0.0001
30.2	–	0.0132	–	–	–	–	–	–	–	–	–	–	–	–	–
31	–	0.1019	–	–	–	–	–	–	–	–	–	–	–	–	–
31.2	–	0.0764	–	–	–	–	–	–	–	–	–	–	–	–	–
32	–	0.0261	–	–	–	–	–	–	–	–	–	–	–	–	–
32.2	–	0.1266	–	–	–	–	–	–	–	–	–	–	–	–	–
33	–	0.0030	–	–	–	–	–	–	–	–	–	–	–	–	–
33.2	–	0.0436	–	–	–	–	–	–	–	–	–	–	–	–	–
34	–	0.0005	–	–	–	–	–	–	–	–	–	–	–	–	–
34.2	–	0.0052	–	–	–	–	–	–	–	–	–	–	–	–	–
35	–	0.0001	–	–	–	–	–	–	–	–	–	–	–	–	–
35.2	–	0.0002	–	–	–	–	–	–	–	–	–	–	–	–	–
36	–	0.0001	–	–	–	–	–	–	–	–	–	–	–	–	–
H_o	0.1653	0.1868	0.2453	0.2612	0.2859	0.3556	0.2077	0.2274	0.1385	0.1852	0.2027	0.3833	0.1460	0.2219	0.1469
H_e	0.8347	0.8132	0.7547	0.7388	0.7141	0.6444	0.7923	0.7726	0.8615	0.8148	0.7973	0.6167	0.8540	0.7781	0.8531
PIC	0.8207	0.7917	0.7426	0.6920	0.6732	0.5571	0.7762	0.7509	0.8489	0.7797	0.7728	0.5571	0.8407	0.7410	0.8432
PD	0.8406	0.8147	0.7758	0.7354	0.7141	0.6180	0.7923	0.7846	0.8639	0.8218	0.8023	0.6180	0.8565	0.7751	0.8586
P_e	0.6951	0.6582	0.5896	0.5271	0.5045	0.3845	0.6320	0.5984	0.7378	0.6684	0.6272	0.3845	0.7265	0.5871	0.7302
P	0.1668	0.2037	0.5809	0.3765	0.3050	0.2019	0.6789	0.6275	0.1471	0.1994	0.2026	0.8353	0.1686	0.4018	0.1772

H_o, observed heterozygosity; H_e, expected heterozygosity; PIC, polymorphism information content; PD, power of discrimination; P_e, probability of exclusion; P, probability values of the exact tests for Hardy-Weinberg equilibrium.

CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818, and FGA) were amplified in five fluorescence-based multiplex polymerase chain reactions (PCRs) using the AmpFLSTR^{TM} Identifiler kit (Applied Biosystems, Foster City, USA). The amplification reactions in a total volume of 10 μL contained 2.0 μL primer set, 0.2 μL AmpliTaq Gold DNA Polymerase (5 U/μL), 2.9 μL deionized water, 4.0 μL dNTP, and 0.9 μL genomic DNA. Thermal cycling was conducted with the following conditions: 95°C for 11 min; 28 cycles of 94°C for 1 min, 59°C for 1 min, 72°C for 1 min; and a final extension of 60°C for 45 min using the GeneAmp PCR system 9700 (Applied Biosystems). Detection and genotyping of all the PCR products were accomplished using the ABI 3130xl DNA Genetic Analyzer (Applied Biosystems). Allele designation was performed using GeneMapper ID Software v3.2 (Applied Biosystems). Quality control was adhered to laboratory internal control standards and kit controls. Allele and genotype frequencies were estimated by direct counting.

The complete dataset is available from the corresponding author via e-mail requests.

Acknowledgements

We thank Mr. Chuan-Hai Wang for his hard work on collecting blood samples.

Authors' contributions

XGY, HFM, YQH, XTS, and WMS carried out the DNA typing. YZL performed the statistical analysis. YJD and JWY conceived of the study, participated in its design and coordination, and drafted the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors have declared that no competing interests exist.

References

1. Zhu, J., *et al.* 2005. Population data of 15 STR in Chinese Han population from north of Guangdong. *J. Forensic Sci.* 50: 1510-1511.
2. Wang, Z., *et al.* 2005. Genetic polymorphisms of 15 STR loci in Chinese Hui population. *J. Forensic Sci.* 50: 1508-1509.
3. Zhu, B., *et al.* 2005. Genetic analysis of 15 STR loci of Chinese Uigur ethnic population. *J. Forensic Sci.* 50: 1235-1236.
4. Kim, Y.L., *et al.* 2003. Allele frequencies of 15 STR loci using AmpFLSTR Identifiler kit in a Korean population. *Forensic Sci. Int.* 136: 92-95.
5. Kubat, M., *et al.* 2004. Population genetics of the 15 AmpFlSTR identifiler loci in Kosovo Albanians. *Int. J. Legal Med.* 118: 115-118.
6. Liu, C., *et al.* 2003. Genetic studies of 15 STR loci in Guangxi Miao population. *J. Forensic Med.* 19: 204-206.
7. Farfan, M.J., *et al.* 2001. STR data for the AmpFlSTR Profiler Plus and COfiler loci from the Maghreb (North Africa). *Forensic Sci Int.* 121: 199-200.
8. Szczerkowska, Z., *et al.* 2004. Northern Polish population data and forensic usefulness of 15 autosomal STR loci. *Forensic Sci. Int.* 144: 69-71.
9. Barrot, C., *et al.* 2005. Characterisation of three Amerindian populations from Hidalgo State (Mexico) by 15 STR-PCR polymorphisms. *Int. J. Legal Med.* 119: 111-115.
10. Walsh, P.S., *et al.* 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 10: 506-513.