

Genetic data provided by 21 autosomal STR loci from Chinese Tujia ethnic group

Guo-lian Yuan · Chun-mei Shen · Hong-dan Wang · Wen-juan Liu ·
Guang Yang · Jiang-wei Yan · Hai-Xia Qin · Tong Xie · Hangyan Ran ·
Jie Yuan · Zuochun Liu · Bofeng Zhu

Received: 9 February 2012 / Accepted: 30 September 2012 / Published online: 13 October 2012
Springer Science+Business Media Dordrecht 2012

Abstract The aim of this study was to investigate allelic frequency distribution and forensic genetic parameters of autosomal short tandem repeats (STR) loci of the population samples from 107 Tujia individuals from Chinese Hubei Province. Twenty-one autosomal STR genetic markers (D9S1122, D6S474, D6S1017, D5S2500, D4S2408, D3S4529, D2S441, D2S1776, D22S1045, D20S482, D1S1677, D1S1627, D1GATA113, D19S433, D18S853, D17S1301, D11S4463, D12ATA63, D10S1248, D10S1435 and D14S1434) were simultaneously amplified in a new multiplex polymerase chain reaction system. 155 alleles for all the STR

loci from the Tujia population were observed and the corresponding allelic frequencies ranged from 0.005 to 0.589. Expected heterozygosity, polymorphic information content, power of discrimination and power of exclusion of the 21 STR loci in the Tujia population were from 0.579 to 0.824, from 0.525 to 0.802, from 0.773 to 0.945 and from 0.257 to 0.641, respectively. Our results indicate that the autosomal STRs multiplex system provides highly informative STR data and could be useful in forensic individual identification and parentage testing in this region.

Keywords Short tandem repeat · Chinese Tujia ethnic group · Allelic diversity

Scientific Research and Experiment Center, The Second Affiliated Hospital, School of Medicine, Xi'an Jiaotong University, Xi'an 710004, People's Republic of China
e-mail: yuanguolian2013@yeah.net

G. Yuan · B. Zhu
The Key Laboratory of Environment and Gene Related to Diseases (Xi'an Jiaotong University), Ministry of Education, Xi'an 710061, People's Republic of China
e-mail: bfzhu2008@yahoo.com.cn

C. Shen
Blood Center of the Shaanxi Province,
Shaanxi 710061, People's Republic of China

H. Wang
Medical Genetic Institute of Henan Province,
People's Hospital of Henan Province, Zhengzhou 450003,
People's Republic of China

H. Wang · W. Liu
School of Medicine, Xi'an Jiaotong University,
Xi'an 710061, People's Republic of China

Introduction

Short tandem repeat (STR) loci, are well widespread throughout the human genome and are the rich source of highly polymorphic genetic markers [1, 2]. STRs show sufficient variability among individuals and have become important analytical tools in several fields of study (e.g., genetic mapping, linkage analysis, human identification, and paternity testing) [3, 4]. With their short PCR fragment, efficient simple typing, and capacity for multiple amplification, STR loci are considered to be superior to other genetic markers. STR typing was also suitable for analyzing degraded and minute amounts of human DNA. Thus, STR technology has been widely used in forensic casework. Moreover, STR loci are employed in population genetics, as population data of STR loci are very useful in predicting genetic relationships among populations and in investigating genetic diversity of different populations [5, 6].

In recent years, more and more new STR loci have been discovered, and these loci have been increasingly used in research and other applications. The aim of this study was to create a set of new STR loci data representing Chinese Tujia ethnic group. We studied the allelic diversity and forensic statistical parameters of 21 new autosomal STR loci (D6S474, D12ATA63, D22S1045, D10S1248, D1S1677, D11S4463, D1S1627, D3S4529, D2S441, D6S1017, D4S2408, D17S1301, D1GATA113, D18S853, D20S482, D14S1434, D9S1122, D2S1776, D10S1435, D19S433 and D5S2500), and provided novel 21 STR data for 107 volunteers of the Tujia ethnic group from Chinese Hubei Province. The present study can provide basic and valuable data of 21 new STR loci for population genetics, human identification and paternity testing in forensic sciences.

Materials and methods

Samples collections and DNA extraction

Bloodstained samples were obtained from 107 unrelated healthy individuals of Tujia group in Enshi Tujia and Miao Autonomous Prefecture in south Hubei Province. The human genomic DNA was extracted using the Chelex-100 protocol as described by Walsh et al. [7]. All the individuals provided their written informed consent for the collection of the samples and subsequent analysis, and the investigation was conducted in accordance with human and ethical research principles of Xi'an Jiaotong University, China. This study was approved by the Ethics Committee of Xi'an Jiaotong University, China.

PCR amplification and DNA typing

The 21 autosomal STR loci were amplified by using the AGCU 21+1 STR Fluorescence Assay Kit (AGCU ScienTech

Incorporation, Wuxi, Jiangsu, China) following manufacturer's instructions in 25 μ l reactions containing 0.5–2 ng genomic DNA, 10 μ l reaction mix, 5 μ l 21+1 primers, 0.5 μ l HS-Taq DNA polymerase and ddH₂O. Thermal cycling was performed using the GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA, USA).

PCR products were separated by capillary electrophoresis on an ABI 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) using 1 μ l PCR product or 21+1 Allelic Ladder was mixed with 12 μ l Hi-Di formamide and 0.5 μ l AGCU Marker SIZ-500 internal lane standard (AGCU ScienTech Incorporation, Wuxi, Jiangsu, China). The mixture was denatured at 95 °C for 3 min; followed by immediate chilling on ice for 3 min. Genotyping was performed by comparison with allelic ladders using GeneMapper ID 3.2 software. According to the number of repeat units present as recommended by the DNA Commission of the Society for Forensic Genetics [8], the alleles of all 21 STR loci were named. In addition, control DNA (9947A) (Promega, Madison, WI, USA) was included in the kit for quality control. All experimental steps were carried out according to the laboratory internal control standards and kit controls.

Statistical and phylogenetic analysis

Hardy–Weinberg equilibrium expectations at all STR loci were evaluated using the Exact test by the modified PowerStat version 1.2 spreadsheet (Promega, Madison, WI, USA) [9]. The polymorphic information content (PIC), power of discrimination (PD), power of exclusion (PE), observed heterozygosity (HO) and expected heterozygosity (HE) were calculated using Excel 2003 and the PowerStat version 1.2 spreadsheet (Promega, Madison, WI, USA) [10]. Locus-by-locus allelic frequencies were compared between the Tujia group and other previously published populations using Analysis of molecular variance method (AMOVA, based on *F*-statistics), which was performed with ARLEQUIN version 3.1 software [11]. The *p*-values of linkage disequilibrium (LD) between all the loci in our study were estimated by Genepop version 4.0.10 (<http://genepop.curtin.edu.au/>).

Results and discussions

The Tujia, with a total population of over 8 million, is the 6th largest ethnic minority in People's Republic of China. They live in Wuling Mountains, straddling the common borders of Hunan, Hubei and Guizhou Provinces, and Chongqing Municipality. Tujia is a Tibeto-Burman language and is usually considered an isolate within this group. Although there are different accounts of their origins, the Tujia may trace their history back over twelve centuries, and possibly beyond, to the

Table 1 Allelic frequencies and forensic statistical parameters of 21 STR loci from Tujia ethnic group in Hubei Province, China (*n* = 107)

Allele	D9S1122	D6S474	D6S1017	D5S2500	D4S2408	D3S4529	D2S441	D2S1776	D22S1045	D20S482	D1S1677
7											
8			0.192		0.234						
9			0.019		0.350			0.126			
9.1							0.023				
9.2											
10	0.051		0.397		0.271		0.280	0.065		0.028	
11	0.182	0.005	0.056		0.131		0.248	0.257	0.224	0.014	0.005
11.3							0.089				
12	0.290		0.229		0.009		0.168	0.416	0.009	0.033	0.019
12.2											
13	0.397		0.103	0.005	0.005	0.238	0.014	0.103	0.009	0.266	0.112
13.2											
14	0.070	0.299	0.005	0.425		0.229	0.168	0.028	0.023	0.477	0.467
14.2											
15	0.005	0.425				0.346	0.009	0.005	0.327	0.117	0.350
15.2											
16	0.005	0.117				0.173			0.243	0.061	0.042
16.2											
17		0.121		0.304		0.014			0.121	0.005	0.005
18		0.028		0.201					0.033		
19		0.005		0.005					0.009		
20				0.056							
23				0.005							
HO	0.701	0.673	0.682	0.617	0.720	0.692	0.785	0.738	0.822	0.729	0.645

Table 1 continued

Allele	D1S1627	D1GATA113	D19S433	D18S853	D17S1301	D11S4463	D12ATA63	D10S1248	D10S1435	D14S1434
17						0.028	0.336	0.014		
18							0.103			
19							0.019			
20										
23										
HO	0.570	0.589	0.785	0.692	0.729	0.757	0.738	0.785	0.804	0.654
HE	0.579	0.611	0.824	0.718	0.716	0.756	0.751	0.758	0.767	0.654
PD	0.773	0.792	0.945	0.866	0.873	0.893	0.894	0.895	0.906	0.826
PIC	0.525	0.552	0.802	0.675	0.675	0.717	0.711	0.720	0.733	0.604
PE	0.257	0.278	0.572	0.415	0.474	0.522	0.490	0.572	0.606	0.361
<i>p</i>	0.814	0.590	0.243	0.486	0.827	0.946	0.688	0.569	0.417	0.955

HO observed heterozygosity, HE expected heterozygosity, PD power of discrimination, PIC polymorphic information content, PE probability of exclusion, *p* probability values of exact tests for Hardy–Weinberg equilibrium

ancient Ba people who occupied the area around modern-day Chongqing some 2,500 years ago. Today, traditional Tujia customs can only be found in the most remote areas [12].

Allelic frequencies and forensic statistical parameters of the observed 21 STR loci are listed in Table 1. In the Tujia ethnic group from Chinese Hubei Province, 155 alleles were observed with corresponding allelic frequencies ranging from 0.005 to 0.589. No deviations from Hardy–Weinberg equilibrium were observed at 21 loci. The HO, HE, PIC, PE, PD ranged from 0.570 (D1S1627 locus) to 0.822 (D22S1045 locus), 0.579 (D9S1122 locus) to 0.824 (D19S433 locus), 0.525 (D1S1627 locus) to 0.802 (D19S433 locus), 0.257 (D1S1627 locus) to 0.641 (D22S1045 locus), 0.773 (D1S1627 locus) to 0.945 (D19S433 locus), respectively. The combined probability of exclusion, power of discrimination, probability of matching value for all 21 STR loci reached 0.9999977307, 0.9999999999999999889026 and 1.10974×10^{-19} , respectively.

Linkage disequilibrium (LD) test is a statistic that refers to a non-random association of alleles at different loci. If the loci on the same human chromosome are closely linked, the LD has to be evaluated for their later practical application. Some articles have reported that LD manifestation between markers distanced more than 5 cM (genetic distance) or 5 Mb (physical distance) is unlikely [13, 14]. The selected loci on the same chromosome in our study are at least 50 Mb apart from each other, which means the 21 STR loci are not linked with each other. However, LD can occur between unlinked loci because of population substructure, natural selection, mutation, random genetic drift, founder effect etc. Hence, we tested the *p*-values of linkage disequilibrium between all the loci in our study. In 210 pairwise comparisons, 20 pairs were found with *p* values below 0.05 (D10S1248/D10S1435, *p* = 0.0001; D10S1248/D11S4463, *p* = 0.0014; D10S1248/D2S441,

p = 0.0156; D10S1248/D9S1122, *p* = 0.0292; D12ATA63/D11S4463, *p* < 0.0001; D12ATA63/D1GATA113, *p* = 0.0044; D12ATA63/D4S2408, *p* = 0.0202; D12ATA63/D5S2500, *p* = 0.0050; D18S853/D19S433, *p* = 0.0001; D18S853/D2S441, *p* = 0.0295; D18S853/D4S2408, *p* = 0.0291; D18S853/D17S1301, *p* = 0.0102; D1GATA113/D1S1677, *p* < 0.0001; D2S441/D2S1776, *p* < 0.0001; D3S4529/D1GATA113, *p* = 0.0052; D4S2408/D3S4529, *p* = 0.003; D5S2500/D6S1017, *p* < 0.0001; D6S474/D9S1122, *p* = 0.0009; D6S474/D3S4529, *p* = 0.0151; D17S1301/D14S1434, *p* < 0.0001). If more than three pairs of LD are due to a single locus, this locus may be removed from the panel (e.g., if A and B, A and C, A and D are in LD, removing A can make the panel better). As is showed above, the three STR loci D10S1248, D12ATA63, and D18S853 result in four pairs of LD. In the population from the Tibetan ethnic minority group residing in Lhasa region, Tibet Autonomous Region of China, three pairs of LD are due to a single locus D18S853 [15]. In the individuals from the Salar ethnic group in Xunhua Salar Autonomous County of Qinghai province of China, the D10S1248, D12ATA63, D6S1017, and D14S1434 loci were in linkage disequilibrium with 4, 6, 4 and 8 STR loci, respectively [16]. The LD results indicate that 5 STR loci (i.e. D10S1248, D12ATA63, D6S1017, D14S1434 and D6S1017) have limited values for forensic applications in the Salar group. These results show that the loci with limited values in the above populations are not the same sites and such LDs may be mainly caused by population substructure.

The allelic frequency distributions in the Tujia group were compared with data which were available for the same set of 21 STR loci studied in the Tibetan, Salar and Northern Han groups [15–17], and the *F_{st}* and *p* values were listed in Table 2. The AMOVA results showed statistically significant differences at four STR loci between Tujia group and the Tibetan group, i.e. D12ATA63 (*F_{st}* = 0.0285; *p* = 0.001), D22S1045 (*F_{st}* = 0.0250; *p* < 0.001), D2S441 (*F_{st}* = 0.0104; *p* = 0.0264), and

Table 2 Comparison of F_{st} and p values between Chinese Tujia and other groups at the same set of 21 STR loci

Populations	Index	D9S1122	D6S474	D6S1017	D5S2500	D4S2408	D3S4529	D2S441	D2S1776	D22S1045	D20S482	D1S1677
Tujia vs. North Han	F_{st}	-0.0019	0.0060	-0.0009	-0.0008	0.0036	0.0040	0.0033	-0.0018	0.0382	0.0041	-0.0007
	p	1.0000	0.0909	0.9189	0.8446	0.2209	0.1769	0.1838	0.9990	0.0000	0.1486	0.8221
Tujia vs. Salar	F_{st}	0.0011	0.0181	0.0053	0.0049	0.0110	0.0014	0.0012	0.0055	0.0769	0.0013	0.0137
	p	0.5200	0.0078	0.1730	0.1896	0.0381	0.4946	0.5249	0.1056	0.0000	0.4663	0.0332
Tujia vs. Tibetan	F_{st}	-0.0027	0.0030	0.0256	-0.0009	0.0045	0.0055	0.0104	-0.0018	0.0250	-0.0022	-0.0025
	p	1.0000	0.3333	0.0020	0.8133	0.2424	0.1701	0.0264	0.9785	0.0000	1.0000	1.0000
Populations	Index	D1S1627	D1GATA113	D19S433	D18S853	D17S1301	D11S4463	D12ATA63	D10S1248	D10S1435	D14S1434	
Tujia vs. North Han	F_{st}	-0.0021	-0.0033	0.0005	-0.0010	-0.0011	-0.0003	0.0016	-0.0016	-0.0017	0.0082	
	p	1.0000	1.0000	0.6940	0.9071	0.9560	0.7879	0.4115	0.9980	1.0000	0.0518	
Tujia vs. Salar	F_{st}	0.0014	0.0074	-0.0020	-0.0015	0.0057	0.0053	-0.0010	0.0033	-0.0007	0.0086	
	p	0.4438	0.1114	1.0000	0.9550	0.1369	0.1584	0.8856	0.2659	0.8602	0.0723	
Tujia vs. Tibetan	F_{st}	0.0002	0.0046	0.0026	0.0063	-0.0012	-0.0029	0.0285	0.0041	0.0035	0.0044	
	p	0.5689	0.1994	0.3842	0.1359	0.9355	1.0000	0.0010	0.2297	0.2630	0.2131	

Table 3 Comparison of F_{st} and p values between Chinese Tujia group and other groups at the overlapping STR markers

Populations	Index	D3S4529	D6S474	D2S1776	D10S1435	D12ATA63	D6S1017	D1S1627	D5S2500	D10S1248	D2S441	D22S1045	D1S1677	D14S1434
Tujia vs. Malay	F_{st}	0.0141	0.0230	0.0052	0.0015	0.0303	0.0077	0.0422	0.0173					
	p	0.0586	0.0148	0.2838	0.9450	0.0026	0.2520	0.0048	0.0696					
Tujia vs. Indian	F_{st}	0.0192	0.0032	0.0030	0.0115	0.0220	0.1396	0.0002	0.1516					
	p	0.0230	0.4106	0.4336	0.5672	0.0074	0.0120	0.0480	0.0164					
Tujia vs. Afrikaner	F_{st}						0.1062	0.0014	0.0716					
	p						0.0040	0.0145	0.0165	0.0437	0.0419			
Tujia vs. Asian Indian	F_{st}						0.2936	0.0190	0.0180	0.0000	0.0006			
	p						0.0032	0.0421	0.0153					
Tujia vs. Mixed Ancestry	F_{st}						0.5038	0.0006	0.0872					
	p													
Tujia vs. Individuals from (RS), Southern Brazil	F_{st}													
	p													
Tujia vs. Individuals from Maghreb	F_{st}													
	p													

D6S1017 ($F_{st} = 0.0256$; $p = 0.0020$); Also at four STR loci between Tujia group and the Salar group, i.e. D1S1677 ($F_{st} = 0.0137$; $p = 0.0332$), D22S1045 ($F_{st} = 0.0769$; $p < 0.001$), D4S2408 ($F_{st} = 0.0110$; $p = 0.0381$), and D6S474 ($F_{st} = 0.0181$; $p = 0.0078$) locus; Only at one STR loci between Tujia group and Northern Han, i.e. D22S1045 ($F_{st} = 0.0382$; $p < 0.001$). From the results we can see there is only one STR marker showed the statistically significant difference

22. Zhu BF, Yang G, Shen CM, Qin HX, Fan SL, Deng LB, Chen F, Chen LP, Wang HD, Wang ZY, Lucas R, Lian ZM (2010) Population genetic analysis of 15 autosomal STR loci in the Russian population of northeastern Inner-Mongolia, China. *Mol Biol Rep* 37:3889–3895
23. Wu YM, Zhang XN, Zhou Y, Chen ZY, Wang XB (2008) Genetic polymorphisms of 15 STR loci in Chinese Han population living in Xi'an city of Shaanxi Province. *Forensic Sci Int Genet* 2:e15–e18
24. Zhu J, Li J, Guo Y, Liu K, Zhu B, Liu Y (2005) Population data of 15 STR in Chinese Han population from north of Guangdong. *J Forensic Sci* 50:1510–1511
25. Zhu B, Wang Z, Wu Q, Sedike ME, Zhu J, Huang P, Xu Y, Liu Y (2005) Genetic analysis of 15 STR loci of Chinese Uigur ethnic population. *J Forensic Sci* 50:1235–1236
26. Yong RY, Gan LS, Chua EM, Yap EP (2009) Polymorphism studies of six miniSTR loci for three ethnic populations in Singapore. *Leg Med (Tokyo)* 11(4):195–197
27. Abrahams Z, D'Amato ME, Davison S, Benjeddou M (2011) Allele frequencies of six non-CODIS miniSTR loci (D1S1627, D3S4529, D5S2500, D6S1017, D8S1115 and D9S2157) in three South African populations. *Forensic Sci Int Genet* 5(4):354–355
28. Raimann PE, de Oliveira AP, Rodenbusch R, Picanço JB, Albuquerque TC, Alho CS (2012) Genetic data for D1S1677, D2S441, D4S2364, D10S1248, D14S1434 and D22S1045 miniSTR loci from the state of Rio Grande do Sul, Southern Brazil. *Forensic Sci Int Genet* 6(1):42–43
29. Cortellini V, Cerri N, Verzeletti A (2011) Genetic variation at 5 new autosomal short tandem repeat markers (D10S1248, D22S1045, D2S441, D1S1656, D12S391) in a population-based sample from Maghreb region. *Croat Med J* 52(3):368–371