

Population data of 21 non-CODIS STR loci in Han population of northern China

Li Yuan · Jianye Ge · Di Lu · Xue Yang

Received: 27 October 2011 / Accepted: 27 December 2011 / Published online: 13 January 2012
© Springer-Verlag 2012

Abstract Allele frequencies and forensic statistics of 21 autosomal short tandem repeat loci (i.e., D6S474, D12ATA63, D22S1045, D10S1248, D1S1677, D11S4463, D1S1627, D3S4529, D2S441, D6S1017, D4S2408, D19S433, D17S1301, D1GATA113, D18S853, D20S482, D14S1434, D9S1122, D2S1776, D10S1435 and D5S2500) were estimated in Han population from northern China ($n=220$). Significant deviation from Hardy–Weinberg equilibrium was detected only for D22S1045. The observed heterozygosity, the expected heterozygosity, the discrimination power, the probability of paternity exclusion in trios, the probability of paternity exclusion in duos and the polymorphic information content ranged from 0.591 to 0.836, 0.594 to 0.830, 0.762 to 0.948, 0.341 to 0.659, 0.189 to 0.487 and 0.535 to 0.807, respectively. Triallelic patterns were observed at D19S433 and D10S1435. Mutations occurred at D22ATA63, D10S1248, D19S433 and D14S1434 loci with all single-step mutations. The expected mutation rates of these four loci are 0.0042 with 95% confidence interval [0.0001, 0.0232] in a total of 238 meioses. Our results show that these 21 non-CODIS STR loci are highly polymorphic and can be useful

for human identification and kinship analysis in Northern Han population in China.

Keywords Short tandem repeats · Non-CODIS · Genotyping · Population data · Han population · Linkage · Linkage Disequilibrium

Introduction

Short tandem repeats (STRs) are important genetic markers and widely used in forensic applications because of their highly polymorphic characteristics [1, 2]. The STRs in the Combined DNA Index System (CODIS) are commercially available in many multiplex amplification kits (e.g., AmpF/STR Identifier®). Testing with these loci usually can meet the requirements of human individual identification and standard triopaternity test. However, duo paternity testing [3, 4], complex kinship analysis cases (e.g., full-sib, grandparents–grandchildren, etc.) [5], and cases with mutations [6] may need more STR loci to obtain reliable identifications. Extra STR loci can provide additional genetic information and are a complementary tool to the conventional STR analysis [7–10]. There have been studies on seeking additional polymorphic STR loci independent from the current CODIS loci [11, 12]. This study is to validate a panel of 21 autosomal non-CODIS STR loci for Northern Han population in China.

Materials and methods

Population and DNA samples

Han population is a native ethnic group in China and, by most modern definitions, the largest single ethnic group in the world. In this study, 220 healthy unrelated Han volunteers in northern China were sampled. One hundred fifty-eight two-generation families including 80 father–child–mother trios

L. Yuan (✉)
Shanghai Key Laboratory of Forensic Medicine
(Institute of Forensic Science, Ministry of Justice),
Shanghai 200063, PR China
e-mail: yuanliwcy@yahoo.com.cn

L. Yuan
West China School of Preclinical and Forensic Medicine,
Sichuan University,
Chengdu, Sichuan 610041, PR China

L. Yuan · D. Lu · X. Yang
China University of Political Science and Law's Key
Laboratory of Evidence Science, Ministry of Education,
Beijing 100040, People's Republic of China

J. Ge
Institute of Applied Genetics, University of North Texas,
Health Science Center,
Fort Worth, TX, USA

and 78 mother–child or father–child duos were collected to estimate the mutation rates of the STR loci. All samples were genotyped with a AmpF Φ STR Identifiler ® multiplex STR kit (Applied Biosystems, Foster City, CA, USA). Paternity index of each family is at least 10,000 to confirm the relationships based on the STRs in Identifiler.

DNA extraction and quantification

Genomic DNA was extracted by using the Chelex-100 protocol as described by Walsh et al. [13]. The quantity of

recovered DNA was determined by Qubit ® Quantitation System (Invitrogen, CA, USA) according to the manufacturer's specifications.

DNA amplification

Amplification of STRs was carried out using a multiplex PCR system AGCU 21+1 fluorescence amplification reagents (AGCU ScienTech Incorporation, Wuxi, Jiangsu, China), which includes Amelogenin, D6S474, D12ATA63, D22S1045, D10S1248, D1S1677, D11S4463, D1S1627,

Table 1 The genomic mapping information of 21 non-CODIS loci with some loci in the Identifiler ® kit which locate on the same chromosome with the non-CODIS loci

Chromosome	Locus	UCSC STS id	Band	Physical (bp)	Genetic (cM)	Reference	9947A genotype
1	<i>D1S1677</i>	5370	1q23.3	163 559 700—163 560 041	175.62	–	13, 14
	<i>D1S1627</i>	5280	1p21.1	106 963 665—106 963 777	139.02	–	13, 14
	<i>D1GATA113</i> ^a	–	1p36.23 ^a	–	17.377 Mb ^a	–	11, 12
2	<i>D2S441</i>	5578	2p14	68 239 016—68 239 157	86.82	–	10, 14
	<i>D2S1776</i>	5575	2q24.3	169 645 172—169 645 775	173.00	–	10, 10
	TPOX	104909	2p25.3	1 493 368—1 493 481	–	[22]	
	D2S1338	5591	2q35	218 879 369—218 879 717	215.78	[22]	
3	<i>D3S4529</i>	5699	3p12.1	85 852 474—85 852 736	112.42	–	13, 13
	D3S1358	162309	3p21.31	45 582 205—45 582 335	–	–	
4	<i>D4S2408</i>	5889	4p15.1	31 304 231—31 304 513	45.97	–	9, 10
	FGA	212610	4q31.3	155 508 848—155 509 043	–	–	
5	<i>D5S2500</i>	6096	5q11.2	58 697 041—58 697 348	69.23	–	14, 23
	D5S818	22478	5q23.2	123 111 125—123 111 402	127.29 ^b	–	
	CSF1PO	168109	5q32	149 455 735—149 456 053	–	–	
6	<i>D6S474</i>	6246	6q21	112 878 950—112 879 283	118.64	–	14, 18
	<i>D6S1017</i>	6276	6p21.1	41 677 174—41 677 509	63.28	–	9, 10
9	<i>D9S1122</i>	6686	9q21.2	79 688 594—79 688 849	75.88	–	12, 13
10	<i>D10S1248</i>	6816	10q26.3	131 092 374—131 092 796	165.27	–	13, 15
	<i>D10S1435</i>	15291	10p15.3	2 243 220—2 243 552	–	–	10, 11
11	<i>D11S4463</i>	116998	11q25	130 872 239—130 872 721	–	–	12, 13
	TH01	167016	11p15.5	2 192 277—2 192 522	–	[22]	
12	<i>D12ATA63</i> ^a	–	12q23.3 ^a	55 349—55 437 ^a	106.825 Mb ^a	–	13, 13
	vWA	7162	12p13.31	6 093 104—6 093 253	14.23	[22]	
14	<i>D14S1434</i>	7278	14q32.13	95 308 123—95 308 685	113.17	–	11, 13
17	<i>D17S1301</i>	7525	17q25.1	72 680 786—72 681 109	100.02	–	12, 12
18	<i>D18S853</i>	24104	18p11.31	3 990 524—3 990 853	–	–	11, 14
	D18S51	7683	18q21.33	60 948 678—60 949 364	95.46	[22]	
19	<i>D19S433</i>	7730	19q12	30 416 990—30 417 261	51.88	–	14, 15
20	<i>D20S482</i>	20134	20p13	4 506 248—4 506 466	13.21 ^b	–	14, 15
22	<i>D22S1045</i>	7934	22q12.3	37 536 285—37 536 570	42.81	–	11, 14

21 non-CODIS loci were denoted in italics

Genetic (cM) Marshfield genetic map position (genetic distance from p-telomere)

^a The information of D1GATA113 (chromosomal position 17.377 Mb) and D12ATA63 (chromosomal position 106.825 Mb) in this study see http://www.cstl.nist.gov/biotech/strbase/pub_pres/Promega2006_Hill.pdf

^b deCODE genetic map position (genetic distance from p-telomere)

Table 2 Allele frequencies and relevant forensic statistics of the 21 non-CODIS STR loci in northern China Han population ($n=220$)

Allele	D6S474	D12ATA63	D22S1045	D10S1248	D1S1677	D11S4463	DIS1627	D3S4529	D2S441	D6S1017	D4S2408	D19S433	D17S1301	D1GATA113	D18S853	D20S482	D14S1434	D9S1122	D2S1776	D10S1435	D5S2500		
7						0.0023							0.0136	0.5205						0.0023		0.0046	
8						0.2341				0.2650			0.0136	0.0068								0.0482	
9						0.0250				0.2696			0.0364								0.0023	0.1644	0.0023
10	0.0068					0.0364				0.2696			0.0568								0.0023	0.0594	0.0482
11	0.0023	0.1182	0.0091			0.0093				0.1590			0.1841	0.1477	0.3785	0.0045	0.1399				0.1438	0.2580	0.1261
11.1										0.0023			0.0023										
11.2																							
11.3									0.0409														
12	0.0023	0.3545	0.1114	0.0568	0.0295	0.0651	0.1000	0.0023	0.1659	0.2409	0.0369	0.0411	0.4227	0.2750	0.0467	0.0523	0.0344	0.3242	0.4041	0.3601			
12.2												0.0068											
13	0.0023	0.0045	0.0045	0.3818	0.1455	0.2302	0.5614	0.1591	0.0432	0.0886		0.2877	0.2205	0.0455	0.2220	0.2818	0.2982	0.3676	0.0822	0.2546			
13.2												0.0548											
14	0.3386	0.0205	0.0136	0.2386	0.4750	0.2837	0.2818	0.2705	0.1159	0.0045		0.2329	0.0455	0.0023	0.2687	0.3977	0.3876	0.0799	0.0320	0.1252	0.3934		
14.2												0.0890											
15	0.3318	0.0045	0.1432	0.2250	0.2909	0.2651	0.0182	0.3795	0.0091			0.0845	0.0068		0.0678	0.1773	0.0528	0.0205	0.0046	0.0229			
15.2						0.0023						0.1370											
16	0.1727	0.1818	0.2432	0.0795	0.0545	0.1349	0.0023	0.1409				0.0251			0.0500			0.0023		0.0069	0.0024		
16.2												0.0274											
17	0.1136	0.3341	0.2114	0.0091	0.0023	0.0070		0.0386				0.0091			0.0045	0.0023					0.2891		
17.2																							
18	0.0341	0.0773	0.1386			0.0023									0.0023						0.1896		
19	0.0045	0.0136	0.0136																		0.0024		
20																					0.1043		
23																					0.0190		
<i>N</i>	4	5	6	5	4	5	3	4	4	4	4	6	5	3	4	5	5	5	5	5	4	4	
<i>H_o</i>	0.768	0.718	0.759	0.750	0.659	0.772	0.627	0.759	0.786	0.741	0.742	0.836	0.746	0.591	0.724	0.723	0.656	0.712	0.680	0.716	0.711		
<i>H_e</i>	0.731	0.723	0.830	0.737	0.665	0.774	0.594	0.736	0.774	0.713	0.758	0.823	0.732	0.630	0.728	0.725	0.730	0.729	0.733	0.771	0.715		
PIC	0.685	0.676	0.807	0.695	0.611	0.737	0.535	0.694	0.741	0.665	0.715	0.802	0.695	0.572	0.683	0.682	0.688	0.685	0.693	0.739	0.665		
DP	0.878	0.874	0.945	0.883	0.836	0.914	0.762	0.880	0.918	0.862	0.902	0.948	0.895	0.805	0.888	0.878	0.886	0.883	0.892	0.919	0.878		
PE(D)	0.319	0.311	0.487	0.330	0.248	0.379	0.189	0.328	0.390	0.295	0.348	0.487	0.336	0.214	0.316	0.319	0.327	0.323	0.331	0.390	0.296		
PE(T)	0.493	0.483	0.659	0.507	0.414	0.557	0.341	0.504	0.568	0.468	0.525	0.658	0.515	0.373	0.490	0.494	0.503	0.497	0.509	0.570	0.468		
<i>P</i>	0.7317	0.1635	0.0000	0.1747	0.6138	0.7636	0.2807	0.1665	0.9306	0.7063	0.5289	0.4815	0.6265	0.4217	0.6842	0.3017	0.0047	0.6296	0.3543	0.1427	0.9969		

N the number of high frequency allele (frequencies ≥ 0.05) in each locus, Table 1 Allele frequencies for 21 STR in Han populations of North China ($n=236$)

Allele frequencies for six STR with a Midi-6 multiplex system in four Japanese populations ($n=198$ in Akita, 200 in Nagoya, 175 in Oita, and 196 in Okinawa)
H_o observed heterozygosity, *H_e* expected heterozygosity, *PIC* polymorphism information content, *DP* discrimination power, *PE(D)* probability of paternity exclusion in duos, *PE(T)* probability of paternity exclusion in trios, *P* probability values of exact tests for Hardy–Weinberg disequilibrium

Table 3 The significant Linkage Disequilibrium values of 13 pair of loci after pairwise comparison

Locus	D2S441	D10S1248	D11S4463	D12ATA63	D14S1434	D17S1301	D20S482
D1S1627	0.0237						
D2S1776	0.0000			0.0353		0.0075	
D3S4529	0.0019						
D4S2408				0.0225			
D6S1017			0.0162				
D6S474		0.0294			0.0462		
D10S1435		0.0184					
D18S853							0.0033
D19S433				0.0000			0.0300

After Bonferroni correction, only two pairs (D2S441 and D2S1776, D12ATA63 and D19S433) were still significant

D3S4529, D2S441, D6S1017, D4S2408, D19S433, D17S1301, D1GATA113, D18S853, D20S482, D14S1434, D9S1122, D2S1776, D10S1435 and D5S2500. A multiplex PCR amplification was performed with a total volume of 10.0 μ l containing 0.2–1.0 ng genomic DNA, 4 μ l Reaction Mix, 2 μ l 21+1 Primers, 1 U HS-Taq DNA polymerase and ddH₂O. PCR was conducted with a GeneAmp PCR System 9700 Thermal Cycler (Applied Biosystems, Foster City, CA) following a protocol with an initial denaturation step at 95°C for 11 min, followed by 10 cycles at 94°C for 1 min, 62°C for 1 min, 72°C for 1 min, and 20 cycles at 90°C for 1 min, 60°C for 1 min, 72°C for 1 min, a terminal extension step at 60°C for 60 min. The experiments were conducted in accordance with quality control measures. Cell lines 9947A (Promega, Madison, WI, USA) were used as positive standard reference materials [14] and ddH₂O was used as negative control.

Genotyping

The PCR products were separated by Capillary Electrophoresis on ABI PRISM 3130 Genetic Analyzers (Applied Biosystems, Foster City, CA, USA). Raw data were analyzed using GenemapperID 3.2 software (Applied Biosystems, Foster City, CA, USA), and alleles were determined by comparing with the allele ladder (AGCU ScienTech Incorporation, Wuxi, Jiangsu, China). If off-ladder peaks, triallelic patterns [15, 16] or mutations between children and parents were encountered, the samples would be typed again to confirm the genotypes. Off-ladder alleles were determined by the method described by Gill et al. [17].

Statistical analysis

Hardy–Weinberg Equilibrium (HWE) and expected heterozygosity (H_e) of each locus as well as Linkage Disequilibrium (LD) between each pair of loci were tested with the Genepop Version 4.0.10 software package (<http://genepop.curtin.edu.au>). Polymorphism information content (PIC), observed heterozygosity (H_o), discrimination power (DP)

and probability of paternity exclusion (PE) of Han population were calculated by PowerStats program (<http://www.promega.com/geneticidtools/>). Further, the Han population data were compared with Tibet population [18] by the shuffling testing method described in Ref. [19].

Quality control

All laboratory procedures are accredited according to ISO17025. Furthermore, laboratory internal control standards were employed according to recommendation published by the Paternity Testing Commission of the International Society for Forensic Genetics [20].

Table 4 Shuffling test for Han and Tibet populations (10,000 shuffles). The method is described in Ref. [19]

Marker	P-value
D6S474	0.2713
D12ATA63	0.0000
D22S1045	0.0000
D10S1248	0.0844
D1S1677	0.2708
D11S4463	0.5578
D1S1627	0.2586
D3S4529	0.0000
D2S441	0.0360
D6S1017	0.0001
D4S2408	0.0040
D19S433	0.0782
D17S1301	0.0106
D1GATA113	0.0022
D18S853	0.0005
D20S482	0.3515
D14S1434	0.2492
D9S1122	0.7346
D2S1776	0.7275
D10S1435	0.0004
D5S2500	0.0017

References

1. Butler JM (2007) Short tandem repeat typing technologies used in human identity testing. *Biotechniques* 43:ii–v
2. Gill P, Werrett DJ, Budowle B, Guerrieri R (2004) An assessment of whether SNPs will replace STRs in national DNA databases—Joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDM). *Sci Justice* 44:51–53
3. Von Wurmb-Schwark N, Malyusz V, Simeoni E et al (2006) Possible pitfalls in motherless paternity analysis with related putative fathers. *Forensic Sci Int* 159:92–97
4. Poetsch M, Ludcke C, Repenning A et al (2006) The problem of single parent/child paternity analysis: practical results involving 336 children and 348 unrelated men. *Forensic Sci Int* 159:98–103
5. Egeland T, Sheehan N (2008) On identification problems requiring linked autosomal markers. *Forensic Sci Int Genet* 2:219–225
6. America Association of Blood Banks. Annual report summary for testing in 2002 [DB/OL]. (2003-11-01) <http://www.aabb.org/sa/facilities/Documents/ptanrpt02.pdf>
7. Uchihi R, Yamamoto T, Yoshimoto T et al (2007) Population data of six STR loci using hexaplex PCR amplification and typing system, “Midi-6” in four regional populations in Japan. *Forensic Sci Int* 169:255–259
8. Kuzniar P, Jastrzebska E, Ploski R (2006) Validation of nine non-CODIS STR loci for forensic use in a population from Central Poland. *Forensic Sci Int* 159:258–260
9. Becker D, Vogelsang D, Brabetz W (2007) Population data on the seven short tandem repeat loci D4S2366, D6S474, D14S608, D19S246, D20S480, D21S226 and D22S689 in a German population. *Int J Legal Med* 121:78–81
10. Hill CR, Butler JM, Vallone PM (2009) A 26plex autosomal STR assay to aid human identity testing. *J Forensic Sci* 4:1008–1015
11. Odriozola A, Aznar JM, Celorrio D et al (2011) Development and validation of I-DNA1: a 15-Loci multiplex system for identity testing. *Int J Legal Med* 125:685–694
12. Hwa HL, Chang YY, Lee JC et al (2011) Fourteen non-CODIS autosomal short tandem repeat loci multiplex data from Taiwanese. *Int J Legal Med* 125:219–226
13. Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 10:506–513
14. Szibor R, Edelmann J, Hering S et al (2003) Cell line DNA typing in forensic genetics—the necessity of reliable standards. *Forensic Sci Int* 138:37–43
15. Crouse CA, Rogers S, Amriott E et al (1999) Analysis and interpretation of short tandem repeat microvariants and three banded allele patterns using multiple allele detection system. *J Forensic Sci* 44:87–94
16. Lane AB (2008) The nature of tri-allelic TPOX genotypes in African populations. *Forensic Sci Int: Genetics* 2:134–137
17. Gill P, Urquhart A, Millican E et al (1996) A new method of STR interpretation using inferential logic-development of a criminal intelligence database. *Int J Legal Med* 109:14–22
18. Zhu BF, Shen CM, Wang HD et al (2011) Genetic diversities of 21 non-CODIS autosomal STRs of a Chinese Tibetan ethnic minority group in Lhasa. *Int J Legal Med* 125:581–585
19. Ge J, Budowle B, Planz JV et al (2010) US forensic Y-chromosome short tandem repeats database. *Leg Med* 12:289–295
20. Morling N, Allen RW, Carracedo A et al (2002) Paternity Testing Commission of the International Society of Forensic Genetics: recommendations on genetic investigations in paternity cases. *Forensic Sci Int* 129:148–157
21. Junge A, Brinkmann B, Fimmers R, Madea B (2006) Mutations or exclusion: an unusual case in paternity testing. *Int J Legal Med* 120:360–363
22. Lu DJ, Liu QL, Zhao H (2011) Genetic data of nine non-CODIS STRs in Chinese Han population from Guangdong Province, Southern China. *Int J Legal Med* 125:133–137