Int J Legal Med (2011) 125:133–137
DOI 10.1007/s00414-009-0388-x

SHORT COMMUNICATION

Genetic data of nine non-CODIS STRs in Chinese Han population from Guangdong Province, Southern China

De-Jian Lu · Qiu-Ling Liu · Hu Zhao

Received: 20 July 2009 / Accepted: 21 October 2009 / Published online: 20 January 2010

Abstract Nine non-combined DNA index system tetranucleotide short tandem repeat (STR) loci D2S1772, D6S1043, D7S3048, D8S1132, D11S2368, D12S391, D13S325, D18S1364, and GATA198B05 were amplified in a multiplex polymerase chain reaction system. The distribution of alleles of the nine STRs was reported from a Chinese Han population in Guangdong Province, Southern China. The combined power of exclusion in trios and duos for the nine loci was 0.999981 and 0.999025, respectively. Mutation rates range from 0 to 0.005618. Pairwise analysis of linkage disequilibrium, which included PowerPlex 16 System loci, did show statistically significant deviation from independence even though loci locate on the same chromosomes. The nine STRs are highly informative and suitable to extend the results obtained with other STRs commonly analyzed for difficult paternity and kinship analysis.

Keywords Short tandem repeat · Non-CODIS loci · Chinese Han · Population data · Forensic genetics

Introduction

Polymorphic short tandem repeats (STR) are powerful and widely used genetic markers for paternity testing. Some laboratories have only used STRs for this purpose. In forensic casework, ordinary parentage cases can be expected to be solved by a commercially available multiplexes of STRs genotyping kit, e.g., PowerPlex 16 System Kit (Promega, Madison, WI, USA) or AmpFℓSTR Identifiler polymerase chain reaction (PCR) amplification kit (Applied Biosystems, Foster City, CA, USA). However, for more difficult cases, such as deficient cases [1], related individuals are involved [2, 3], mutations are encountered [4], or complex kinship analysis [5, 6] and common STRs cannot provide enough discriminatory power [5, 7, 8]. Additional STRs are needed to complement conventional analysis for obtain more information. In this paper, we present the allele frequencies, forensic efficiency values of nine non-combined DNA index system (CODIS) autosomal STR loci which were analyzed in a Chinese population using a new multiplex PCR system.

Electronic supplementary material The online version of this article (doi:10.1007/s00414-009-0388-x) contains supplementary material, which is available to authorized users.

D.-J. Lu (✉) · Q.-L. Liu · H. Zhao
Faculty of Forensic Medicine, Zhongshan Medical School,
Sun Yat-sen University,
74# Zhongshan Road II,
Guangzhou 510089, People's Republic of China
e-mail: lvdejian@mail.sysu.edu.cn

Materials and methods

Five hundred six bloodstains samples deposited on qualitative filter papers were collected from unrelated Chinese Han individuals residing in Guangdong Province, Southern China. Genomic DNA was extracted from bloodstains using a Chelex-100 method. Of the bloodstain in 250 μl 5% Chelex-100 (*w*/*v*), 3×3-mm size cutting was incubated at 56°C for 2 h to overnight. Subsequently, the extract was heated at 98°C for 8 min, vortexed for 10 s, and centrifuged in a microcentrifuge.

Amplification of STRs was carried out using a multiplex PCR system STRtyper-10F/G kit (Condon, Zhuhai, China). This kit comprises nine STR loci and a sex determination gene: D2S1772, D6S1043, D7S3048, D8S1132, D11S2368, D12S391, D13S325, D18S1364, GATA198B05, and Amelogenin. The genomic mapping information and

genotypes of DNA from cell line 9947A (Promega, Madison, WI, USA) are presented in Table 1. These loci and their alleles have been confirmed by cloning sequencing. Multiplex PCR of the STRtyper-10F/G kit was performed using 2 µl of template DNA in 10 µl volume including 1 µl 10× Primer Set, 2 µl 5× PCR Reaction Mix, and 0.2 µl (1 U) HS-Taq DNA polymerase. PCR cycle condition was 95°C for 5 min, followed by 28 cycles at 94°C for 1 min, 60°C for 1 min, 72°C for 1 min, and final extension at 72°C for 20 min. All amplifications were done on GeneAmp PCR 9700 (Applied Biosystems, Foster City, CA, USA) or MJ PTC-100 (Bio-Rad, Hercules, CA, USA) thermal cyclers.

The PCR products were separated and detected by capillary electrophoresis using an ABI 3100 Genetic Analyzer (Applied Biosystems) according to the manufacturer's instructions. Briefly, 1 µl PCR product or Allelic Ladder was mixed with 10 µl Hi-Di formamide (with 5% internal lane standard), denatured at 95°C for 3 min, followed by chilling on ice for 3 min immediately. Then, the mixture was analyzed on the ABI 3100 Genetic Analyzer with Data Collection Software (version 2.0); POP™-4 (Applied Biosystems) was utilized for separations on a 36-cm array. Samples were injected electrokinetically for 10 s at 3 kV. Allele calls were made in GeneMapper ID v3.2 software by comparison with kit allelic ladders using the panel and bin files supplied by the kit manufacturer. If an off-ladder peak was encountered, the sample was typed repeatedly to confirm the typing. New allele was determined as described in [9] and added to the allelic bin set. Control DNA from cell line 9947A (Promega, Madison, WI, USA) was genotyped as standard reference in all experiments.

Allelic frequencies, expected heterozygosity, probability of exclusion in trios (PE_T), probability of exclusion in duos (PE_D), and probability of match (PM) were computed using GenAlEx (ver6.2) software package [10], Power of discrimination (PD) were calculated using formula PD=1−PM. The exact test (base on 5,000 shufflings) for Hardy–Weinberg equilibrium and linkage disequilibrium was carried out by using GDA ver. 1.1 (http://lewis.eeb.uconn.edu/lewishome/software.html) software. Exact test differentiation of allele frequency distribution between populations was performed by using Arlequin ver 3.11 software package [11].

Results and discussion

All samples showed completed genotyping profiles. The plot of Allelic Ladder and a sample profile was displayed in

Table 1 Physical and genetic mapping information of the studied nine short tandem repeat (STRs) and commonly used STRs on the same chromosome and genotyping of 9947 A DNA of nine STR markers in this study

Chro.	Marker	UCSC	Cytogenetic	Physical (bp)	Genetic (cM)	9947A
2	TPOX	104909	2p25.3	1493368-1493481	–	
	D2S1772[a]	5567	2p14	67051054-67051286	85.48	26,28
	D2S1338	5591	2q35	218879369-218879717	215.78	
6	D6S1043[a]	6245	6q15	92449867-92450188	100.91	12,18
7	D7S3048[a]	6360	7p15.3	21266641-21266915	34.15	24,24
	D7S820	6392	7q21.11	83789381-83789718	98.44	
8	D8S1132[a]	14549	8q23.1	107328821-107329072	119.22	19,21
	D8S1179	6559	8q24.13	125907064-125907405	135.08	
11	TH01	167016	11p15.5	2192277-2192522	–	
	D11S2368[a]	6957	11p15.1	19280960-19281402	22.56	19,21
12	vWA	7162	12p13.31	6093104-6093253	14.23	
	D12S391[a]	7052	12p13.2	12449874-12450226	26.23	18,20
13	D13S325[a]	7211	13q14.11	43173250-43173571	38.96	20,21
	D13S317	7213	13q31.1	82722033-82722314	63.90	
18	D18S51	7683	18q21.33	60948678-60949364	95.46	
	D18S1364[a]	24186	18q22.1	63400053-63400355	–	13,15
22	GATA198B05[a]	7939	22q11.1	17650650-17650795	1.79	18,19

Physical and genetic mapping data obtained from Human Genome Browser (http://www.genome.ucsc.edu). Genotype data of 9947A DNA supplied by the manufacturer

Chro number of chromosome, *UCSC* UCSC STS id, *cytogenetic*, cytogenetic localization, *physical*: physical localization (physical distance from p-telomere), *genetic* Marshfield genetic localization (genetic distance from p-telomere), *9947A* genotypes of 9947A cell line DNA, – no report

[a] Short tandem repeat markers in this study

Fig. S1. Microvariant such as N.1, N.2, and N.3 alleles were detected at loci D6S1043, D7S3048, D8S1132, D11S2368, D12S391, and D18S1364. The full typing data including PowerPlex 16 System is presented in Table S1. Allele frequencies and forensic parameters are presented in Table 2. Genetic polymorphism of D2S1772 [12], D6S1043 [12], D7S3048 [12], D8S1132 [7, 13, 14], D12S391 [8, 13–15], and D13S325 [16] have been published previously. Comparison of allele frequency distribution between the studied population and other

Table 2 Allele frequencies and forensic parameters for the nine non-combined DNA index system short tandem repeats used in this study

Allele	D2S1772	D6S1043	D7S3048	D8S1132	D11S2368	D12S391	D13S325	D18S1364	GATA198B05
9		0.0010							
10		0.0366							
11		0.1117							
12		0.1364						0.0524	
12.3		0.0010							
13		0.1245						0.1769	
14		0.1462						0.1621	0.0049
15		0.0178		0.0010	0.0049	0.0158	0.0010	0.1927	0.0267
16		0.0040		0.0128	0.0395	0.0069	0.0030	0.1976	0.1136
17	0.0079	0.0316	0.0119	0.0909	0.1532	0.0563	0.0099	0.0464	0.1709
17.2		0.0010							
17.3		0.0010							
18	0.0128	0.1640	0.1028	0.2095	0.1354	0.2233	0.0504	0.1008	0.0721
18.1				0.0010				0.0030	
18.2		0.0010							
18.3				0.0010					
19	0.0178	0.1393	0.0741	0.1917	0.1472	0.2055	0.2204	0.0553	0.0850
19.2					0.0010				
19.3				0.0010		0.0010			
20	0.1028	0.0682	0.2075	0.1571	0.1947	0.1759	0.2520	0.0128	0.0978
20.3		0.0020							
21	0.1275	0.0109	0.1344	0.1423	0.2105	0.1166	0.2510		0.2559
21.3		0.0020							
22	0.0879		0.0919	0.1018	0.0761	0.0998	0.1512		0.1462
22.1			0.0010						
23	0.0138		0.1304	0.0603	0.0217	0.0652	0.0375		0.0227
24	0.2866		0.1255	0.0198	0.0128	0.0277	0.0178		0.0040
25	0.0504		0.0879	0.0079	0.0020	0.0030	0.0030		
26	0.0366		0.0296	0.0020	0.0010	0.0030	0.0010		
27	0.1344		0.0030				0.0020		
28	0.0879								
29	0.0277								
30	0.0040								
31	0.0020								
p value	0.2052	0.4710	0.0506	0.5590	0.0310	0.4960	0.2512	0.2874	0.3744
Hom	0.1542	0.1304	0.1245	0.1304	0.1502	0.1344	0.1957	0.1621	0.1581
He	0.8530	0.8791	0.8737	0.8524	0.8472	0.8457	0.7985	0.8488	0.8486
PE_T	0.7139	0.7531	0.7434	0.7017	0.6910	0.6909	0.6017	0.6945	0.6981
PE_D	0.5523	0.6018	0.5889	0.5370	0.5241	0.5243	0.4246	0.5284	0.5327
PD	0.9639	0.9727	0.9707	0.9604	0.9576	0.9574	0.9290	0.9585	0.9596

p value: *p* value of exact test of Hardy–Weinberg equilibrium; *Hom*: observed homozygosity; *He*: unbiased expected heterozygosity; PE_T: probability of exclusion in trios (PE_T); PE_D: probability of exclusion in duos; *PD*: power of discrimination

published population was performed. The results showed significant differentiation between our data and other populations except Chinese Mongolian at D2S1772, German at D8S1132, and Southeast China Han at D12S391 (Table S2). D11S2368, D18S1364, and GATA198B05 have not been reported for forensic purpose. Data in Table 2 reveals the three loci have highly polymorphic information.

Base on the exact test, only the locus D11S2368 showed a departure from Hardy–Weinberg expectation (p=0.0310), but it was well below the significance threshold after employing a Bonferroni's correction [17].

Pairwise linkage disequilibrium tests were performed for all possible two-locus combinations, and one deviation was showed in 36 pairs (D2S1772/GATA198B05, p=0.0456), which is not to be significant after Bonferroni's correction. Considering the loci as independent, total PD for the nine loci is 0.999999999999753. Combined probability of exclusion in trios and combined probability of exclusion in duos is 0.999981 and 0.999025, respectively.

Besides the nine loci mentioned above, more loci may need in parentage analysis. So, samples were typed using PowerPlex 16 system simultaneously. Exact test indicated that only Penta D (p=0.0216) and Penta E (p=0.0417) did not meet Hardy–Weinberg expectations for the loci in PowerPlex 16 system. Pairwise linkage disequilibrium tests were also performed for all pairs of two loci, including the loci comprising in PowerPlex 16 System. Although some markers were located on the same chromosome (Table 1), linkage disequilibrium between these loci was not observed. Five deviations (including one described above) were detected in 276 pairwise comparisons (D2S1772/GATA198B05, p=0.0456; D3S1358/GATA198B05, p=0.0184; D11S2368/Penta D, p=0.0366; D13S325/D3S1358, p=0.0488; and D18S1364/Penta D, p=0.0004), with p values below the nominal 0.05 threshold. However, none of them remained significant after the Bonferroni's correction for multiple testing (resulting in a significance threshold p value of 0.0002). Furthermore, these markers of pairwise of deviations do not locate on the same chromosome. Overall, there is no linkage disequilibrium between the nine loci studied here and the Powerplex 16 system loci.

It is important to note that in cases where a mutation might have occurred, another mutation may arise when more STRs were used in parentage analysis. Therefore, mutation rate of the nine non-CODIS STRs were investigated. In total of 51 mother/father/child trios and 432 parent/child duos, single mismatch was founded in seven cases. The observed mutation rates ranged from 0 to 5.618×10^{-3} (Table 3) and are in the range reported by Brinkmann et al. [18], but the sample size is too small to draw a reliable mutation rate.

We used aliquots of the cell line 9947A DNA (200, 300, 400, 500, 750, 1,000, 1,500, 2,000 2,500, and 3,000 pg) to determine the minimum quantity of DNA required to achieve reliable results. With 500–1,500 pg of template DNA, allele typing at all nine STRs and Amel gene was very successful (peak highs at all loci exceeded 100 relative fluorescence units). Out of this range, unbalance or artifact peaks maybe appear.

In conclusion, our work has demonstrated the nine non-CODIS loci are highly polymorphic markers and highly discriminating tools. These markers are probably unlinked to common STRs included in commercially available kits (e.g., PowerPlex 16 system) and useful to obtain additional information in STR analysis for forensic casework, especially useful for complex kinship analysis, paternity testing in deficiency cases or within related individuals. In our practice, we have used the nine non-CODIS loci to solve the complicated cases, including all cases of parent/child duos and cases of suspected mutation.

Table 3 Observed mutations in a total of 51 mother/father/child trios and 432 parent/child duos for nine non-combined DNA index system short tandem repeats

Locus	Case number	μ($\times10^{-3}$)	95% CI ($\times10^{-3}$)	Origin	Type	Genotype		
						Father	Mother	Child
D11S2368	1	1.873	0.047–10.389	Paternal	+1	19,19		20,20
D13S325	1	1.873	0.047–10.389	Paternal	+1	18,20		21,21
D6S1043	3	5.618	1.160–16.330	Paternal	+1	14,19	11,11	11,15
				Paternal	+1	11,13		12,12
				Paternal	+1	12,19		20,20
D7S3048	2	3.745	0.45–13.46	Maternal	+2	18,20	19,24	18,21
				Paternal	−1	19,19		18,23

μ: mutation rate; *95% CI*: 95% confidence interval (CI) of mutation rate calculated by assuming a binomial distribution. (http://statpages.org/confint.html); +: repeat gain; −: repeat loss

References

1. Lee HS, Lee JW, Han GR, Hwang JJ (2000) Motherless case in paternity testing. Forensic Sci Int 114:57–65
2. Fung WK, Wong DM, Chung YK (2003) How well do serial tandem repeat loci perform in excluding paternity in relatives of the biologic father among immigration cases? Transfusion 43:982–983
3. von Wurmb-Schwark N, Malyusz V, Simeoni E, Lignitz E, Poetsch M (2006) Possible pitfalls in motherless paternity analysis with related putative fathers. Forensic Sci Int 159:92–97
4. 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
5. Wenk RE, Chiafari FA, Gorlin J, Polesky HF (2003) Better tools are needed for parentage and kinship studies. Transfusion 43:979–981
6. Gaytmenn R, Hildebrand DP, Sweet D, Pretty IA (2002) Determination of the sensitivity and specificity of sibship calculations using AmpF/STR Profiler Plus. Int J Legal Med 116:161–164
7. Betz T, Immel UD, Kleiber M, Klintschar M (2007) "Paterniplex", a highly discriminative decaplex STR multiplex tailored for investigating special problems in paternity testing. Electrophoresis 28:3868–3874
8. Grubwieser P, Zimmermann B, Niederstatter H, Pavlic M, Steinlechner M, Parson W (2007) Evaluation of an extended set of 15 candidate STR loci for paternity and kinship analysis in an Austrian population sample. Int J Legal Med 121:85–89
9. Phillips CP, Syndercombe Court D, Lareu MV, Hasskamp J, Carracedo A (1998) Band shift analysis of three base-pair repeat alleles in the short tandem repeat locus D12S391. Forensic Sci Int 93:79–88
10. Peakall R, Smouse P (2006) genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295
11. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50
12. Su XL, Li DX, Liu SP, Bi LF, Wang PY, Yun TYW, Yun S (2004) Population studies on two native Mongolian population groups in China using STR loci. Forensic Sci Int 141:197–199
13. Schmid D, Anslinger K, Rolf B (2005) Allele frequencies of the ACTBP2 (=SE33), D18S51, D8S1132, D12S391, D2S1360, D3S1744, D5S2500, D7S1517, D10S2325 and D21S2055 loci in a German population sample. Forensic Sci Int 151:303–305
14. 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
15. Shi M, Yu X, Baia R, Shu X, Zhu G, Lv J, Tu Y (2008) Genetic polymorphism of 14 non-CODIS STR loci for forensic use in southeast China population. Forensic Sci Int 174:76–79
16. Wang Z, Wang YY, Zhou B, Du B, Zhang BL, Zhang GQ, Jia Y, Zhu YH, Wei YS, Zhang L (2007) Allele frequency of D13S325 and D13S892 in two populations. J Forensic Sci 52:226–226
17. Weir BS (1996) Genetic data analysis II: methods for discrete population genetic data. Sinauer Associates, Sunderland
18. Brinkmann B, Klintschar M, Neuhuber F, Huhne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet 62:1408–1415