Journal of Forensic and Legal Medicine 44 (2016) 10–13

Contents lists available at ScienceDirect

Journal of Forensic and Legal Medicine

journal homepage: www.elsevier.com/locate/jflm

Research Paper

Population study and mutation analysis for 28 short tandem repeat loci in southwest Chinese Han population

Qin Su [a, 1], Bo Jin [a, b, 1], Haibo Luo [a], Yingbi Li [a], Jin Wu [a], Jing Yan [a], Yiping Hou [a], Weibo Liang [a, *], Lin Zhang [a, **]

[a] Department of Forensic Genetics, West China School of Basic Science and Forensic Medicine, Sichuan University (West China University of Medical Sciences), Chengdu 610041, PR China
[b] Department of Forensic Medicine, North Sichuan Medical College, Nanchong 637000, PR China

ARTICLE INFO

Article history:
Received 12 December 2015
Received in revised form
8 August 2016
Accepted 23 August 2016
Available online 24 August 2016

Keywords:
Short tandem repeats
Southwestern China
Population genetics
Mutations
Forensic practice

ABSTRACT

Short tandem repeat (STR) system is the most widely used genetic markers in modem forensic practice. Because of the relatively unstable molecular structure, STRs show a high mutation rate. In the current study, we report 169 mutation events of 13 CODIS and 15 non-CODIS STR loci that were found in 5569 cases of trios and duos paternity test. Our result indicated that locus-specific mutation rate varied among different populations, geometric means of the longest run of perfect repeats (LRPR) and heterozygosity. Along with previous published data, a forensic dataset for allele frequencies and locus-specific mutation rates of 13 CODIS and 15 non-CODIS STR loci from southwest Chinese Han population has been established. The mutation rate data have important implications in interpreting forensic individual identification and paternity testing.

1. Introduction

Short tandem repeat (STR) loci, also known as microsatellites, are abundant in human genome and widely used as first-line personal genetic markers in forensic practice. Based on the comparative theory, the identification of person via STR typing have been considered as a golden standard in forensic fields.[1,2] Despite of the great value of STR loci for forensic application, most STR loci showed a relatively higher mutation rates than coding genes[3] and lead to incorrect conclusion in forensic analysis. Thus, mutation events were suggested to be taken into the consideration during the calculation of cumulative paternity index (CPI), or additional STR loci should be analyzed for more comparable DNA intelligence to reach an undoubted identifications especially in complex and/or disputed cases.[4,5]

The Combined DNA Index System (CODIS), including 13 STRs, has been world widely used since it had been established by FBI in 1997.[6] STR loci including but not limited to CODIS have been used in China for over two decades. It is necessary to establish a data bank for mutation rates of CODIS and non-CODIS loci, which may have important implication in the interpretation of forensic cases. Mutation analyses of STRs of some Chinese populations have been made in recent years.[7] Many new STR loci with excellent distinguishing ability in Chinese population have been investigated and frequently used in our practice. However, their mutation rates have not been well studied. Zhu et al. have reported mutation rates and 95% CI of 28 STR loci, mutation steps and gender origin.[8] However, the factors that may contribute to the mutation have not be studied. In the current study, we analyzed 28 STR loci, including 13 CODIS STRs (D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D16S539, TH01, TPOX, CSF1PO and D7S820) and 15 additional loci (Penta E, D2S441, D2S1338, Penta D, D10S1248, D19S433, D6S1043, D12S391, D11S2368, D13S325, D18S1364, D2S1772, D7S3048, D8S1132 and D22-GATA198B05) in Han population living in Southwestern China and intend to analyze the factors that may impact the mutation rates. It was found that the locus-specific STR mutation rate is associated with different populations, geometric means of the longest run of perfect repeats (LRPR) and heterozygosity.

* Corresponding author.
** Corresponding author.
E-mail addresses: liangweibo@gmail.com (W. Liang), zhanglin@scu.edu.cn (L. Zhang).
[1] Qin Su and Bo Jin contributed equally to this work, and were considered as co-first author.

http://dx.doi.org/10.1016/j.jflm.2016.08.008
1752-928X/

2. Materials and methods

AS described in,[8] DNA samples from 565 parents/child-trio, 3558 father/child-duo and 1446 mother/child-duo paternity cases were collected and received STR genotyping. The Chelex-100® protocol was used to extract genomic DNA from peripheral blood or buccal cotton swab samples. Amplification of 28 STR loci (D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D16S539, TH01, TPOX, CSF1PO, D7S820, Penta E, D2S441, D2S1338, Penta D, D10S1248, D19S433, D6S1043, D12S391, D11S2368, D13S325, D18S1364, D2S1772, D7S3048, D8S1132 and D22-GATA198B05) were applied using multiplex PCR system Goldeneye 20A kit (Peoplespot Incorporation, Beijing, China), AGCU Expressmarker 22 kit (AGCU ScienTech Incorporation, Wuxi, Jiangsu, China) according to manufacturer's instructions. PCR products were separated on an ABI PRISM 310/3130 Genetic Analyzers (Applied Biosystems, Foster City, CA, USA). Data were analyzed with GeneMapper ID v3.2 software (Applied Biosystems).

A mutation was first identified for, then the parental origin and amount of steps were determined according to the definition introduced by Brinkmann et al.[3] and Weber.[9] Null/silent alleles have been excluded from this study.[10]

The Fisher's exact test of Hardy–Weinberg's Equilibrium (HWE) of each locus and observed heterozygosity (Ho), expected heterozygosity (He) were estimated with the Arlequin Ver 3.5.1.3 (http://cmpg.unibe.ch/software/arlequin3/). Polymorphism information content (PIC), power of discrimination (PD) and probability of paternity exclusion (PE) of loci were calculated using the PowerStats Ver 1.2 (http://www.promega.com/geneticidtools/).

The mutation data of this study were compared with other reports using SPSS 22.0 (SPSS Incorporation, Chicago, Illinois, USA). The relevancy between each locus mutation rate and the longest run of perfect repeats (LRPR), expected heterozygosity were investigated through SPSS 22.0 too. Other statistical tests were described in the article, when p value < 0.05 a significant relationship between two variables was confirmed.

3. Results and discussion

3.1. General data of population genetics study

Based on the genotype of 28 STR loci in 6134 samples of unrelated individuals, deviations of Hardy–Weinberg's Equilibrium (HWE) were estimated at D12S391 ($p = 0.0000$), D18S1364 ($p = 0.0386$), D18S51 (p = 0.0109), Penta D ($p = 0.0444$) and TPOX ($p = 0.0466$). After Bonferroni correction (i.e., $0.05/28 = 0.00178$), only D12S391 was significant. Seventeen of 28 loci had the observed heterozygosity (Ho) higher than 0.8. The value of Ho ranged from 0.615 (TPOX) to 0.921 (Penta E). Penta E presented the most informative locus with PD = 0.987, while TPOX (PD = 0.792) was the least informative one. The combined PD of all 28 loci was reached at approximately 1. The probability of paternity exclusion (PE) varied between 0.306 at TPOX and 0.833 at Penta E. The typical paternity index (TPI, the harmonic mean of the paternity index) varied from 1.291 at TPOX to 6.131 at Penta E. Combined PE and TPI was 0.999999999999 and 1.26×10^{12}, respectively. According to the information showed in Table S1, most of these loci indicated useful potential for forensic application.

3.2. The mutation rates in familial trios versus duos

According to Zhu's,[8] 169 mutation events were observed from 565 parents/child-trio and 5004 parent/child-duo cases. Among them, there were 25 mutation events in trios and 144 mutation events in duos, and there were 6 cases where two STR exclusions were found. The implication of a combination of duos and trios might result in an underestimation of mutation rate when the child was heterozygous at a particular locus and the available parent had

Table 1
Comparison of mutation rates and 95% confidence interval (CI) with other datasets.

Locus	This study		Qian et al.		Hohoff et al.		Lotte Henke et al.		Ana Carolina Mardini et al.	
	Mutation Rate $\times 10^{-3}$	95% CI $\times 10^{-3}$	Mutation Rate $\times 10^{-3}$	95% CI $\times 10^{-3}$	Mutation Rate $\times 10^{-3}$	95% CI $\times 10^{-3}$	Mutation Rate $\times 10^{-3}$	95% CI $\times 10^{-3}$	Mutation Rate $\times 10^{-3}$	95% CI $\times 10^{-3}$
CSF1PO	1.1	0.4–2.5	1.2	0.7–1.8	0.0	0.0–7.6	1.9	1.0–3.3	1.5	1.0–2.1
D10S1248	0.4	0.0–2.4								
D11S2368	3.1	0.8–8.0								
D12S391	3.9	2.3–6.1	2.1	1.5–2.8			4.5	0.1–25		
D13S317	1.5	0.6–3.0	0.5	0.2–0.9	0.0	0.0–7.6	0.9	0.3–1.9	1.0	0.6–1.5
D13S325	1.6	0.2–5.6								
D16S539	0.9	0.3–2.2	0.5	0.2–0.9	4.3	0.1–23.7	1.2	0.7–1.9	1.1	0.7–1.6
D18S1364	2.3	0.5–6.7								
D18S51	2.3	1.2–4.0	1.8	1.2–2.5	0.0	0.0–11.0	1.4	0.9–2.2	1.7	1.2–2.3
D19S433	0.9	0.2–2.5	0.7	0.3–1.4	0.0	0.0–15.7	0.8	0.4–1.4		
D21S11	1.7	0.8–3.2	0.8	0.5–1.4	5.9	0.7–21.1	1.5	0.9–2.3	1.6	1.1–2.2
D2S1338	2.8	1.4–5.2	1.4	0.8–2.3	4.3	0.1–23.7	0.7	0.3–1.3		
D2S1772	0.8	0.0–4.3								
D2S441	0.4	0.0–2.5								
D3S1358	1.3	0.5–2.7	1.1	0.6–1.6	2.1	0.1–11.5	1.2	0.7–1.8	0.6	0.3–1.0
D5S818	0.9	0.3–2.2	1.1	0.6–1.6	0.0	0.0–7.6	1.1	0.5–2.2	1.5	1.0–2.1
D6S1043	2.8	1.3–5.1	0.7	0.3–1.2						
D7S3048	0.8	0.0–4.3								
D7S820	0.7	0.1–2.1	1.1	0.6–1.6	2.1	0.1–11.6	1.3	0.6–2.5	1.0	0.6–1.5
D8S1132	2.3	0.5–6.8					4.5	0.1–24.8		
D8S1179	0.9	0.3–2.4	1.3	0.8–1.9	2.9	0.1–16.3	0.9	0.5–1.5	1.5	1.0–2.1
FGA	5.9	3.8–8.7	2.6	1.9–3.5	2.1	0.1–11.5	2.4	1.7–3.3	2.3	1.7–3.0
GATA198B05	2.3	0.5–6.8								
Penta D	0.5	0.1–1.8	0.4	0.2–0.8			0.9	0.4–1.6		
Penta E	1.7	0.7–3.6	4.0	2.9–5.3			0.9	0.4–1.6		
TH01	0.7	0.2–2.2	0.0	0.0–0.2	0.0	0.0–7.6	0.2	0.0–0.7	0.0	0.0–0.3
TPOX	0.5	0.1–1.7	0.1	0.0–0.4	0.0	0.0–7.6	0.1	0.0–0.4	0.1	0.0–0.4
vWA	2.6	1.3–4.6	1.7	1.2–2.4	4.1	0.5–14.9	1.6	1.0–2.4	2.2	1.6–2.9