SHORT COMMUNICATION

Mutation analysis of 24 short tandem repeats in Chinese Han population

Dejian Lu · Qiuling Liu · Weiwei Wu · Hu Zhao

Abstract Germline mutations of 24 short tandem repeat (STR) loci (TPOX, D3S1358, FGA, D5S818, CSF1PO, D7S820, D8S1179, TH01, vWA, D13S317, Penta E, D16S539, D18S51, Penta D, D21S11, D2S1772, D6S1043, D7S3048, D8S1132, D11S2368, D12S391, D13S325, D18S1364, and GATA198B05) were studied for 6,441 parent–child meioses taken from the paternity testing cases in Chinese Han population. In total, 195 mutations were identified at 22 of the 24 loci. Among them, 189 (96.92%) mutations were one step, five mutations (2.56%) were two step, and one mutation (0.51%) was three step. No mutation was found at the TH01 and TPOX loci. The overall mutation rate estimated was 0.0013 (95% CI 0.0011–0.0015), and the locus-specific mutation rate estimated ranged from 0 to 0.0034. There was a bias in the STR mutations that repeat gains were more common than losses (~1.7:1). Mutation events in the male germline were more frequent than in the female germline (~4.3:1). Furthermore, loci with a larger heterozygosity tended to have a higher mutation rate. Mutation in short alleles was biased towards expansion, whereas mutation in long alleles favored contraction. The long alleles have a higher allelic mutational probability than short alleles.

Keywords Microsatellites · Chinese Han · Paternity testing · Mutations · Forensic genetics

D. Lu (✉) · Q. 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

W. Wu
Department of Criminal Investigation,
Zhejiang Provincial Public Security Bureau,
66# Minsheng Road,
310009 Hangzhou, People's Republic of China

Introduction

Short tandem repeats (STRs) have been widely used in paternity testing. The most commonly used STRs are the loci included in the commercially available amplification kits, such as AmpFℓSTR® Identifiler (Applied Biosystems, Inc., Foster City, CA, USA) and PowerPlex® 16 System (Promega Corp., Madison, WI, USA). But, extended loci may be needed in parentage testing, especially in deficiency cases of disputed paternity [1] and complex kinship testing [2, 3]. Additional STRs have been developed for such purpose [4, 5]. STRs have mutation rates ranging from 0 to 7×10^{-3} [6] and are hence more susceptible to mutation if only one or two mismatches occurred between parent and offspring [7, 8]. Mutations need to be considered in calculating the paternity probability [9, 10]. A reliable knowledge on the mutation rates and characteristics of STRs is very important for interpretation of paternity or kinship. Here, we report the mutations of 24 STRs from the PowerPlex 16 kit and additional loci in the Chinese Han population based on data obtained from cases of paternity testing.

Materials and methods

Genomic DNA was extracted from bloodstain samples using a Chelex-100® method. These 24 STRs were applied in routine paternity testing. Amplification of the 15 STRs

(TPOX, D3S1358, FGA, D5S818, CSF1PO, D7S820, D8S1179, TH01, vWA, D13S317, Penta E, D16S539, D18S51, Penta D, and D21S11) including in the Power-Plex® 16 System kit was carried out according to the instructions provided by the manufacturer. The nine non-CODIS loci (D2S1772, D6S1043, D7S3048, D8S1132, D11S2368, D12S391, D13S325, D18S1364, and GATA198B05) were amplified as described previously [11]. PCR products were analyzed in a 3100 Genetic Analyzer, and genotypes were generated using GeneMapper ID v3.2 software (Applied Biosystems Inc.).

Paternity cases of trios and duos were randomly selected from the Chinese Han population. All parental pairs self-declared that they were unrelated. The parenthood was considered to be confirmed if paternity/maternity index exceeded 10,000.

A mutation was assumed when there was an isolated Mendelian inconsistency with a single mutational step [6] between parent(s)/child provided the paternity/maternity index is >10,000,000, not taking the inconsistency into consideration. In the case of two- or three-step mutation, further analysis was performed to confirm the paternity/maternity using additional 21 autosomal STRs [12]. If the final paternity or maternity index attained or exceeded 1×10^{13} (ignoring the discrepant loci) after inclusion of the additional STRs, new mutations were considered to have occurred. If two or more genetic incompatibilities were observed, the parenthood was excluded.

To reduce the impact of genotyping errors, mutational cases were confirmed by re-genotyping both parents and offspring. If homozygous parent(s) or homozygous child was observed at a locus, the homozygote was verified by extended testing: homozygotes at TPOX, D3S1358, FGA, D5S818, CSF1PO, D6S1043, D7S820, D8S1179, TH01, D12S391, vWA, D13S317, Penta E, D16S539, D18S51, Penta D, or D21S11 were analyzed using AmpFlSTR® Identifiler kit, AmpFlSTR® Minifiler kit (Applied Biosystems, Inc., Foster City, CA, USA), AmpFlSTR® Sinofiler kit (Applied Biosystems, Inc., Foster City, CA, USA), or AGCU 17+1 STR kit (AGCU ScienTech Inc. Wuxi, China). Homozygotes at D2S1772, D7S3048, D8S1132, D11S2368, D13S325, D18S1364, or GATA198B05 were reanalyzed by singleplex PCR using primer sets in the website (http://genome.ucsc.edu/cgi-bin/hgGateway).

Because Mendelian discrepancies may come from a null/silent allele[13], we assumed that null alleles were presented in cases in which only a single mismatch between a homozygous parent and homozygous child at a locus when the other STR loci were consistent with paternity and/or maternity. Null alleles were removed from mutation analysis. The parental origin of the mutated "new" allele and the number of mutational steps were defined as described by Brinkmann et al. [6].

The mutation rate at each locus was calculated as the number of mutations divided by the number of allelic transfers from parent to child. The 95% confidence intervals (CI) for mutation rates were derived based on the binomial distribution and obtained via the website http://statpages.org/confint.html. Unrelated parent pairs were used to estimate allelic frequency and Hardy–Weinberg expectation using a software GDA ver. 1.1 (http://lewis.eeb.uconn.edu/lewishome/software.html). Spearman's test was performed using SPSS 13.0.

To assess the relationship of allele sizes and mutation rate, a modified category was used as described by Ge et.al. [14]. Briefly, allele sizes collected from the unrelated individuals were arbitrarily cut into three equal groups for short, moderate, and long allele sizes, respectively.

Results and discussion

The parentage cases included in this study were from 2,506 father–mother–child triplets, 857 father/child duos, and 572 mother/child duos, of which parenthood had been proven. This provided a total of 154,584 parent/child allele transfers at 24 loci (involving 6,441 parent–child meioses) for study.

In total, we observed 195 mutations at 22 of the 24 loci (Table S1). No mutation was found at two loci: TH01 and TPOX loci. The average mutation rate across all loci was 0.0013 (95% CI 0.0011–0.0015) per locus per gamete per generation. But there obviously existed a significant variation in mutation rates among STRs. The observed locus-specific mutation rates ranged from 0 to 0.0034 and were in the ranges reported by Brinkmann et al. [6] and Becker et al. [15]. The mutation counts and rates at each locus are presented in Table 1.

Based on the heterozygosity estimated from 3,890 unrelated individuals, Spearman's test was used to test the correlation between the mutation rate and heterozygosity at the 24 loci. The result implied that STR with larger heterozygosity may have a higher mutation rate (Table S2, $P=0.020$). This was not consistent with Leopoldino and Pena's study [16], which did not detect any association between mutation rate and heterozygosity at nine loci.

In comparing locus-specific mutation rates with the data from other studies [17–20], we found that the mutation rate differences between datasets vary with the loci (Table S3). For the CODIS loci [18], FGA had the highest mutation rate, and TPOX and TH01 had the lowest mutation rates in each dataset. D18S51 had a moderate mutation rate in [17], [19], and this study, while the highest mutation rate was shown in AABB (Association of Blood Banks) data [18]. Furthermore, differences could be observed within the Chinese Han population [17] at the following loci: D3S1358, D5S818, D18S51, D21S11, and Penta E

Table 1 Mutation rate and 95% CI for the 24 autosomal STRs studied in Chinese Han population

Locus	Paternal (no. of transmissions=3,363)			Maternal (no. of transmissions=3,078)			Unassigned	Total (no. of transmissions=6,441)		
	No. of mutations	Mutation rate	95% CI	No. of mutations	Mutation rate	95% CI	No. of mutations	No. of mutations	Mutation rate	95% CI
CSF1PO	4	0.0012	0.0003–0.0030	0	0.0000	0.0000–0.0012	2	6	0.0009	0.0003–0.0020
D11S2368	12	0.0036	0.0018–0.0062	2	0.0006	0.0001–0.0023	2	16	0.0025	0.0014–0.0040
D12S391	10	0.0030	0.0014–0.0055	1	0.0003	0.0000–0.0018	0	11	0.0017	0.0009–0.0031
D13S317	1	0.0003	0.0000–0.0017	0	0.0000	0.0000–0.0012	0	1	0.0002	0.0000–0.0009
D13S325	11	0.0033	0.0016–0.0058	5	0.0016	0.0005–0.0038	6	22	0.0034	0.0021–0.0052
D16S539	1	0.0003	0.0000–0.0017	1	0.0003	0.0000–0.0018	0	2	0.0003	0.0000–0.0011
D18S1364	8	0.0024	0.0010–0.0047	2	0.0006	0.0001–0.0023	0	10	0.0016	0.0007–0.0029
D18S51	7	0.0021	0.0008–0.0043	1	0.0003	0.0000–0.0018	0	8	0.0012	0.0005–0.0024
D21S11	2	0.0006	0.0001–0.0021	5	0.0016	0.0005–0.0038	1	8	0.0012	0.0005–0.0024
D2S1772	6	0.0018	0.0007–0.0039	1	0.0003	0.0000–0.0018	1	8	0.0012	0.0005–0.0024
D3S1358	5	0.0015	0.0005–0.0035	0	0.0000	0.0000–0.0012	2	7	0.0011	0.0004–0.0022
D5S818	7	0.0021	0.0008–0.0043	0	0.0000	0.0000–0.0012	1	8	0.0012	0.0005–0.0024
D6S1043	2	0.0006	0.0001–0.0021	0	0.0000	0.0000–0.0012	0	2	0.0003	0.0000–0.0011
D7S3048	5	0.0015	0.0005–0.0035	3	0.0010	0.0002–0.0028	1	9	0.0014	0.0006–0.0027
D7S820	4	0.0012	0.0003–0.0030	1	0.0003	0.0000–0.0018	1	6	0.0009	0.0003–0.0020
D8S1132	9	0.0027	0.0012–0.0051	4	0.0013	0.0004–0.0033	1	14	0.0022	0.0012–0.0036
D8S1179	4	0.0012	0.0003–0.0030	0	0.0000	0.0000–0.0012	1	5	0.0008	0.0003–0.0018
FGA	12	0.0036	0.0018–0.0062	2	0.0006	0.0001–0.0023	0	14	0.0022	0.0012–0.0036
GATA198B05	5	0.0015	0.0005–0.0035	0	0.0000	0.0000–0.0012	2	7	0.0011	0.0004–0.0022
PENTA_D	2	0.0006	0.0001–0.0021	0	0.0000	0.0000–0.0012	1	3	0.0005	0.0001–0.0014
PENTA_E	10	0.0030	0.0014–0.0055	3	0.0010	0.0002–0.0028	1	14	0.0022	0.0012–0.0036
TH01	0	0.0000	0.0000–0.0011	0	0.0000	0.0000–0.0012	0	0	0.0000	0.0000–0.0006
TPOX	0	0.0000	0.0000–0.0011	0	0.0000	0.0000–0.0012	0	0	0.0000	0.0000–0.0006
vWA	11	0.0033	0.0016–0.0058	1	0.0003	0.0000–0.0018	2	14	0.0022	0.0012–0.0036
Total	138	0.0017	0.0014–0.0020	32	0.0004	0.0003–0.0006	25	195	0.0013	0.0011–0.0015

(Table S3). However, the 95% CIs for mutation rates from AABB data [18] were in the ranges of our data at most loci, including CSF1PO, D16S539, D18S51, D21S11, D3S1358, D5S818, D7S820, D8S1179, FGA, and Penta E. Spearman's tests showed significant correlations ($P<0.05$) between our data and those of Yan et.al. [17], AABB [18], and Henke et al. [19]. Population differentiation of mutation rates at some loci may have resulted from the limited mutation events in our samples.

All mutations observed were either repeat losses or repeat gains, including 102 repeat gains, 59 repeat losses, and 34 unassigned (Table S4). The ratio of repeat gains versus repeat losses was ~1.7:1. Although this ratio may be overestimated (or underestimated) because of the 34 indeterminate mutations, a similar ratio should remain in the unassigned events. The ratio suggested that there is a bias in STR mutations that gains are more common than losses. A similar tendency has been observed by Leopoldino and Pena [16], Xu et.al. [21], and Brinkmann et al. [6]. None of the mutations observed was an addition or deletion of an incomplete repeat.

To investigate the relationship of allele sizes and repeat gains/losses, mutation progenitor alleles (in which mutation direction was determined certainly) were classified into short, moderate, and long allele sizes. The data showed that mutations with repeat gains were more frequent for short alleles, and repeat losses were more common for long alleles (Table S5). The data herein supported that the mutation in short alleles was biased towards expansion, whereas mutation in longer alleles favored contraction. Similar trends have been reported at Y STRs [14, 22]. However, Xu et al. [21] found the rate of repeat contraction, but not repeat expansion, to increase with increasing allele size at human autosomal tetranucleotide repeat markers.

An exact test for Hardy–Weinberg equilibrium (HWE) performed on the samples of 1,950 unrelated parent pairs indicated that the genotype distributions did not deviate from HWE at all 24 loci ($P>0.05$) (Table S6). Allele frequencies were stationary. Ignoring the multiple-step mutations (for more than 95% of mutations were single step, described below), allele mutation probability was proportional to the allele frequency [23]. To assess the relationship between allele sizes and allele mutation probability, a ratio of mutation counts versus allele proportion was calculated for short, moderate, and long allele sizes, respectively (Table S5). The results showed that the ratio of long alleles is greater than short and moderate alleles provided that mutation was observed in the long allele category. Considering that the loci without mutation of long alleles may result from limited mutational events (D13S317, D16S539, D3S1358, and Penta D) or too small allele proportion in the population (D13S325), we conclude that the long alleles have a higher allelic mutational probability than short alleles.

The vast majority of mutations (189/195=96.92%) were one-step events over all loci, while five mutations were two step (5/195=2.56%), and one mutation was three step (0.51%). Single-step changes were strongly favored over multiple-step changes (Table S4).

Although the origin of 25 mutations remained unclear, 138 paternal and 32 maternal mutations could be determined under 80,712 paternal and 73,872 maternal allelic transfers, respectively (Table S4). The overall ratio of paternal versus maternal mutations was ~4.3:1. This ratio was lower than that reported by some investigators [6, 15, 16, 20, 24], whereas an approximate ratio was observed by Xu et.al. [21] and Leopoldino and Pena [16]. These differences may have resulted from the variational characteristics at each STR loci.

In summary, the average mutation rate across all 24 loci in this study is compatible with other data [6, 15, 16]. Differentiation of locus-specific mutation rates between Chinese Han and other populations varies with loci. A correlation between the mutation rate and heterozygosity of STR was observed. Allelic mutational probability of long alleles is higher than short alleles. Mutation events were more frequent in the male than in the female germline. The vast majority of mutations can be explained by losses or gains of a single repeat unit. In general, gains are more common than losses, and there is a significant excess of losses in long alleles and gains in short alleles. With comparisons between relatives in relationship testing, mutational events can play an important role [25]. Our data in this study will be helpful for parentage testing and kinship analysis, such as deficient cases and mass disaster victim identification.

References

1. Poetsch M, Ludcke C, Repenning A, Fischer L, Malyusz V, Simeoni E, Lignitz E, Oehmichen M, von Wurmb-Schwark N (2006) The problem of single parent/child paternity analysis–practical results involving 336 children and 348 unrelated men. Forensic Sci Int 159(2–3):98–103
2. Nothnagel M, Schmidtke J, Krawczak M (2010) Potentials and limits of pairwise kinship analysis using autosomal short tandem repeat loci. Int J Legal Med 124(3):205–215
3. Wenk RE, Chiafari FA, Gorlin J, Polesky HF (2003) Better tools are needed for parentage and kinship studies. Transfusion 43(7):979–981
4. Betz T, Immel U-D, Kleiber M, Klintschar M (2007) "Paterniplex", a highly discriminative decaplex STR multiplex tailored for investigating special problems in paternity testing. Electrophoresis 28(21):3868–3874
5. 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(2):85–89
6. 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(6):1408–1415

7. Narkuti V, Vellanki RN, Anubrolu N, Doddapaneni KK, Gandhi Kaza PC, Mangamoori LN (2008) Single and double incompatibility at vWA and D8S1179/D21S11 loci between mother and child: implications in kinship analysis. Clin Chim Acta 395(1–2):162–165

8. Nutini AL, Mariottini A, Giunti L, Torricelli F, Ricci U (2003) Double incompatibility at human alpha fibrinogen and penta E loci in paternity testing. Croat Med J 44(3):342–346

9. Drabek J (2009) Validation of software for calculating the likelihood ratio for parentage and kinship. Forensic Sci Int Genet 3(2):112–118

10. Dawid AP, Mortera J, Pascali VL (2001) Non-fatherhood or mutation? A probabilistic approach to parental exclusion in paternity testing. Forensic Sci Int 124(1):55–61

11. 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(1):133–137

12. Zhu BF, Shen CM, Wang HD, Yang G, Yan JW, Qin HX, Guo JX, Huang JF, Jing H, Liu XS (2011) Genetic diversities of 21 non-CODIS autosomal STRs of a Chinese Tibetan ethnic minority group in Lhasa. Int J Legal Med 125(4):581–585

13. Clayton TM, Hill SM, Denton LA, Watson SK, Urquhart AJ (2004) Primer binding site mutations affecting the typing of STR loci contained within the AMPFlSTR SGM Plus kit. Forensic Sci Int 139(2–3):255–259

14. Ge J, Budowle B, Aranda XG, Planz JV, Eisenberg AJ, Chakraborty R (2009) Mutation rates at Y chromosome short tandem repeats in Texas populations. Forensic Sci Int Genet 3(3):179–184

15. Becker D, Bender K, Edelmann J, Gotz F, Henke L, Hering S, Hohoff C, Hoppe K, Klintschar M, Muche M, Rolf B, Szibor R, Weirich V, Jung M, Brabetz W (2007) New alleles and mutational events at 14 STR loci from different German populations. Forensic Sci Int Genet 1(3–4):232–237

16. Leopoldino AM, Pena SDJ (2003) The mutational spectrum of human autosomal tetranucleotide microsatellites. Hum Mutat 21(1):71–79

17. Yan J, Liu Y, Tang H, Zhang Q, Huo Z, Hu S, Yu J (2006) Mutations at 17 STR loci in Chinese population. Forensic Sci Int 162(1–3):53–54

18. Butler JM (2006) Genetics and genomics of core short tandem repeat loci used in human identity testing. J Forensic Sci 51(2):253–265

19. Henke L, Henke J (2006) Supplemented data on mutation rates in 33 autosomal short tandem repeat polymorphisms. J Forensic Sci 51(2):446–447

20. Hohoff C, Schurenkamp M, Borchers T, Eppink M, Brinkmann B (2006) Meiosis study in a population sample from Afghanistan: allele frequencies and mutation rates of 16 STR loci. Int J Legal Med 120(5):300–302

21. Xu X, Peng M, Fang Z (2000) The direction of microsatellite mutations is dependent upon allele length. Nat Genet 24(4):396–399

22. Dupuy BM, Stenersen M, Egeland T, Olaisen B (2004) Y-chromosomal microsatellite mutation rates: differences in mutation rate between and within loci. Hum Mutat 23(2):117–124

23. Vicard P, Dawid AP (2004) A statistical treatment of biases affecting the estimation of mutation rates. Mutat Res 547(1–2):19–33

24. Hohoff C, Schurenkamp M, Brinkmann B (2009) Meiosis study in a population sample from Nigeria: allele frequencies and mutation rates of 16 STR loci. Int J Legal Med 123(3):259–261

25. Leclair B, Fregeau CJ, Bowen KL, Fourney RM (2004) Enhanced kinship analysis and STR-based DNA typing for human identification in mass fatality incidents: the Swissair flight 111 disaster. J Forensic Sci 49(5):939–953