



Mutation rates at 42 Y chromosomal short tandem repeats in Chinese Han population in Eastern China

Weiwei Wu¹ · Wenyan Ren¹ · Honglei Hao¹ · Hailun Nan² · Xin He² · Qiuling Liu² · Dejian Lu²

Received: 1 October 2017 / Accepted: 17 January 2018 / Published online: 31 January 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Mutation analysis of 42 Y chromosomal short tandem repeats (Y-STRs) loci was performed using a sample of 1160 father–son pairs from the Chinese Han population in Eastern China. The results showed that the average mutation rate across the 42 Y-STR loci was 0.0041 (95% CI 0.0036–0.0047) per locus per generation. The locus-specific mutation rates varied from 0.000 to 0.0190. No mutation was found at DYS388, DYS437, DYS448, DYS531, and GATA_H4. DYS627, DYS570, DYS576, and DYS449 could be classified as rapidly mutating Y-STRs, with mutation rates higher than 1.0×10^{-2} . DYS458, DYS630, and DYS518 were moderately mutating Y-STRs, with mutation rates ranging from 8×10^{-3} to 1×10^{-2} . Although the characteristics of the Y-STR mutations were consistent with those in previous studies, mutation rate differences between our data and previous published data were found at some rapidly mutating Y-STRs. The single-copy loci located on the short arm of the Y chromosome (Yp) showed relatively higher mutation rates more frequently than the multi-copy loci. These results will not only extend the data for Y-STR mutations but also be important for kinship analysis, paternal lineage identification, and family relationship reconstruction in forensic Y-STR analysis.

Keywords Y chromosomal short tandem repeat (Y-STR) · Mutation · Microsatellites · Rapidly mutating Y-STRs (RM Y-STRs) · Chinese Han population

Knowledge of the mutability of Y chromosome short tandem repeats (Y-STRs) is necessary for the correct interpretation of genetic profiles, kinship calculations, and genealogical relationship reconstructions. However, the current information on Y-STR mutability is still limited, since empirical data are only available for a small number of particular loci commonly used in the forensic field. Furthermore, the accuracy of the Y-STR mutation rate estimates needs further analysis with large family data. In this paper, we examined the mutations of 42 Y-

STRs, including 17 newly developed loci that have not been embodied in YHRD (<https://yhrd.org/>) and lack accurate mutation rate data in the Chinese Han population. We reveal the mutation rate differences between the present data and other estimates at the rapidly mutating STRs (RM-STRs) and describe the characteristics of Y-STR mutations.

Buccal swab samples of 1160 father–son pairs were taken from individuals in an Eastern Chinese Han population (Zhejiang Province). Informed consent was obtained from the volunteers or legally authorized representatives of all father–son pairs. All procedures performed in this study were approved by the Medical Ethics Committee of the Zhongshan Medical School of Sun Yat-sen University in accordance with the 2013 Helsinki Declaration.

Genomic DNA was extracted using the DNA IQ system (Promega Corporation, Madison, USA). DNA samples were examined with 3 Y-STR multiplex PCR reactions: Yfiler® Plus PCR Amplification kit, AGCU GFS Y STR kit, and AGCU Y24 STR kit. The biological relationship of all father–son pairs was confirmed by analysis of 38 autosomal STRs with paternity index values above 100,000. The STR kits and STR loci are listed in Table S1. All assays were

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00414-018-1784-x>) contains supplementary material, which is available to authorized users.

✉ Dejian Lu
lvdejian@mail.sysu.edu.cn

¹ Zhejiang Key Laboratory of Forensic Science and Technology, Institute of Forensic Science of Zhejiang Provincial Public Security Bureau, 66# Minsheng Road, Hangzhou 310009, People's Republic of China

² Faculty of Forensic Medicine, Zhongshan Medical School of Sun Yat-sen University, 74# Zhongshan Road II, Guangzhou 510089, People's Republic of China

performed according to the manufacturer's recommendations. Control DNA (such as 007, 008, and 9947A) was genotyped as a standard reference.

In mutation counting, if changes of motifs occurred in both DYS389I and DYS389II (e.g., 12 → 13 and 29 → 30), then only one mutation was counted for DYS389I because the DYS389I PCR product was actually a subset of the DYS389II amplified product. For the multi-copy markers DYS385a/b, DYF387S1a/b, DYS459a/b, and DYS527a/b, the number of allele transmissions was counted as twice the number of meioses. The mutation rates and their confidence intervals (CIs) were calculated using the exact binomial probability distribution (see <http://statpages.info/confint.html>).

The number of repeat gains vs. losses and the different mutation rates were compared using an exact test to assess whether the different values significantly deviate from the null hypothesis of a 1:1 ratio. A statistical method similar to that described by Ge et al. [1] was applied to assess the relationship between the number of repeats and mutation rates. Statistical analysis was implemented in the R package "stats". An algorithm similar to that described by Ge et al. [1] was used to investigate the independence of mutations across the 42 loci using the R package "XNomial" to compute the *p* value.

A total of 38 single-copy and 4 multi-copy Y-STR markers were investigated in this study. Although the mutation rates in the Chinese Han population have been previously estimated for 25 of these 42 markers (DYF387S1, DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS449, DYS456, DYS458, DYS460, YS481, DYS518, DYS533, DYS570, DYS576, DYS627, DDYS635, and GATA_H4) [2], only a small sample (100 father–son pairs) was analyzed for the other 17 loci (DYS388, DYS443, DYS444, DYS446, DYS447, DYS459a/b, DYS510, DYS520, DYS522, DYS527a/b, DYS531, DYS552, DYS557, DYS587, DYS622, DYS630, and GATA_A10) [3], indicating a lack of accurate mutation rates for these 17 loci. The accuracy of the Y-STR mutation rate estimation needs more large family data.

In our study, the number of allele transmissions (which varied from 1153 to 2343 at each locus since null alleles and multi-alleles were detected), number of mutations, mutation rates, and their 95% CIs were calculated for each locus and are presented in Table S2. The average mutation rate across all loci was 0.0041 (95% CI 0.0036–0.0047) per locus per generation. The observed locus-specific mutation rates for the 42 Y-STRs ranged from 0.000 to 0.0190. No mutation was found at DYS388, DYS437, DYS448, DYS531, and GATA_H4. Comparison of the average mutation rates calculated in the present study with those reported by other authors yielded results comparable to those based on the direct count in father–son pairs [2, 4–6] and derivation from deep root pedigrees [7].

Mutation rate differences at rapidly mutating STR (RM Y-STR) loci were observed between our data and those in other

previous reports. The DYS518 locus in our study showed a relatively low average mutation rate (8.6×10^{-3}) compared with that reported by Ballantyne et al. [6], Wang et al. [2], and Oh et al. [4], in which the average mutation rate of DYS518 ranged from 13.6×10^{-3} to 18.4×10^{-3} . We also noticed that the mutation rate at DYF387S1a/b (5.13×10^{-3} , 95% CI $2.65\text{--}8.94 \times 10^{-3}$) in our study was significantly lower than that estimated by Ballantyne et al. [6] at 15.9×10^{-3} (95% CI $10.8\text{--}22.4 \times 10^{-3}$). In contrast, the present mutation rate was similar to that reported in a Southern Chinese Han population (4.8×10^{-3} , 95% CI $2.3\text{--}8.3 \times 10^{-3}$) [2] and the 95% confidence interval of Central Chinese Han population (8.7×10^{-3} , 95% CI $4.0\text{--}16.5 \times 10^{-3}$) [8]. Therefore, different mutation rates at DYF387S1a/b may exist among different populations.

In contrast to the low mutation rates at DYS552 ($< 2.84 \times 10^{-3}$) and DYS630 ($< 4.86 \times 10^{-3}$) calculated in Ref [6, 9], relatively high mutation rates of 7.76×10^{-3} and 8.62×10^{-3} , respectively, were detected at these two loci in the present study, consistent with the rates reported in the Southern Chinese Han population [2]. Although the mutation rates of the two loci were lower than those of RM Y-STRs [10], these values are much higher than those of conventional forensic analyses using Y-STRs and can be identified as moderately mutating loci [11]. Again, different Y-STR mutation rates were indicated between populations.

All observed mutation events were single-repeat changes, with the exception of the 5 two repeat changes that occurred at DYS481, DYS533, DYS570, and DYS627 (Table S2 and Table S3). The vast majority of mutations were one-step events ($214/219 = 97.72\%$), fitting to the stepwise mutation model (SMM). None of the observed mutations were additions or deletions of an incomplete repeat. Namely, STR mutations arise by replication slippage.

The overall ratio of repeat gains versus losses was ~ 1.3298:1 (125:94) (Table S2). The level of statistical significance (*p* value = 0.1081) indicated that mutations at these Y-STR markers have approximately equal gains and losses of repeats, consistent with the results of previous studies [1, 2, 6, 11]. However, gain bias of repeats [6] or loss bias of repeats [12, 13] has been observed.

The distribution of the mutation counts by allele size is shown in Fig. S1. The STRs showed a length dependency of the mutation rate. A significant difference in the mutation rates was observed among the short, moderate, and long alleles (*p* = 0.0002). Mutations at longer alleles were more common than those at shorter alleles (Table S4) [6].

We found that the mutation rates may be related to the locus position on the Y chromosome. DYS627, DYS570, DYS449, and DYS576 had the mutation rates exceeding 0.0100, and these STRs were single-copy markers located on the Y chromosome short arm (Yp). Notably, the single-copy loci on Yp showed relatively higher mutation rates more frequently than multi-

copy loci, such as DYS387S1a/b, DYS385a/b, DYS459a/b, and DYS527a/b. This finding may be related to the structure complexity of repetitive and surrounding regions [6].

In addition, the distribution of the number of repeat gains versus losses in short, moderate, and long alleles (Fig. S2) showed that longer alleles tended to lose repeats, whereas shorter alleles significantly and more frequently gained repeats (expansion). Furthermore, a significant correlation ($R = 0.5094$, $p = 4 \times 10^{-9}$) was observed between the gene diversities and the locus-specific mutation rates. Our results are consistent with the findings of previous studies [1, 6, 13].

Most of the father–son pairs with mutations had a mutation at one of the 42 Y-STRs, but 13 father–son pairs had mutations at two loci, and 3 father–son pairs had mutations at three loci (Table S3). Comparison of the observed counts of father–son pairs showed no, one, two, and three or more mutations with the expected numbers of father–son pairs involving zero, one, and two, and three or more mutations under the hypothesis of independence of mutation across loci was presented in Table S5. No significant deviation was shown between the observed and expected numbers ($p = 0.1707$), indicating that the mutations across the 42 STRs randomly occurred.

In summary, the average mutation rate across the 42 Y-STR loci was 0.0041 (95% CI 0.0036–0.0047) with locus-specific mutation rates ranging from 0.000 to 0.0190. Mutation rate differences between our data and previously published data were observed at some loci. The single-copy loci located on Yp showed relatively higher and more frequent mutation rates than the multi-copy loci. However, the characteristics of Y-STR mutations are consistent with those reported in previous studies [1, 6].

Funding information This work was financially supported by grants from the Key Research and Development Program of Zhejiang Province (2017C03026).

Compliance with ethical standards

Buccal swab samples of 1160 father–son pairs were taken from individuals in an Eastern Chinese Han population (Zhejiang Province). Informed consent was obtained from the volunteers or legally authorized representatives of all father–son pairs. All procedures performed in this study were approved by the Medical Ethics Committee of the Zhongshan Medical School of Sun Yat-sen University in accordance with the 2013 Helsinki Declaration.

Conflict of interest The authors declare that there are no conflicts of interest.

References

- 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. <https://doi.org/10.1016/j.fsigen.2009.01.007>
- Wang Y, Zhang YJ, Zhang CC, Li R, Yang Y, Ou XL, Tong DY, Sun HY (2016) Genetic polymorphisms and mutation rates of 27 Y-chromosomal STRs in a Han population from Guangdong Province, Southern China. *Forensic Sci Int Genet* 21:5–9. <https://doi.org/10.1016/j.fsigen.2015.09.013>
- Lang M, Ye Y, Li J, Zhang Y, Yuan T, Hou Y (2017) Comprehensive mutation analysis of 53 Y-STR markers in father-son pairs. *Forensic Sci Int: Genet Supp Ser* 6:e57–e58. <https://doi.org/10.1016/j.fsigs.2017.09.004>
- Oh YN, Lee HY, Lee EY, Kim EH, Yang WI, Shin KJ (2015) Haplotype and mutation analysis for newly suggested Y-STRs in Korean father-son pairs. *Forensic Sci Int Genet* 15:64–68. <https://doi.org/10.1016/j.fsigen.2014.09.023>
- Turrina S, Caratti S, Ferriani M, De Leo D (2015) Haplotype data and mutation rates for the 23 Y-STR loci of PowerPlex(R) Y 23 System in a Northeast Italian population sample. *Int J Legal Med* 129(4):725–728. <https://doi.org/10.1007/s00414-014-1053-6>
- Ballantyne KN, Goedbloed M, Fang R, Schaap O, Lao O, Wollstein A, Choi Y, van Duijn K, Vermeulen M, Brauer S, Decorte R, Poetsch M, von Wurmb-Schwark N, de Knijff P, Labuda D, Vezina H, Knoblauch H, Lessig R, Roewer L, Ploski R, Dobosz T, Henke L, Henke J, Furtado MR, Kayser M (2010) Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *Am J Hum Genet* 87(3):341–353. <https://doi.org/10.1016/j.ajhg.2010.08.006>
- Boattini A, Sarno S, Bini C, Pesci V, Barbieri C, De Fanti S, Quagliariello A, Pagani L, Ayub Q, Ferri G, Pettener D, Luiselli D, Pelotti S (2016) Mutation rates and discriminating power for 13 rapidly-mutating Y-STRs between related and unrelated individuals. *PLoS One* 11(11):e0165678. <https://doi.org/10.1371/journal.pone.0165678>
- Zhang W, Xiao C, Yu J, Wei T, Liao F, Wei W, Huang D (2017) Multiplex assay development and mutation rate analysis for 13 RM Y-STRs in Chinese Han population. *Int J Legal Med* 131(2):345–350. <https://doi.org/10.1007/s00414-016-1489-y>
- Burgarella C, Navascues M (2011) Mutation rate estimates for 110 Y-chromosome STRs combining population and father-son pair data. *Eur J Hum Genet* 19(1):70–75. <https://doi.org/10.1038/ejhg.2010.154>
- Ballantyne KN, Keerl V, Wollstein A, Choi Y, Zuniga SB, Ralf A, Vermeulen M, de Knijff P, Kayser M (2012) A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. *Forensic Sci Int Genet* 6(2):208–218. <https://doi.org/10.1016/j.fsigen.2011.04.017>
- Rogalla U, Wozniak M, Swobodzinski J, Derenko M, Malyarchuk BA, Dambueva I, Kozinski M, Kubica J, Grzybowski T (2015) A novel multiplex assay amplifying 13 Y-STRs characterized by rapid and moderate mutation rate. *Forensic Sci Int Genet* 15:49–55. <https://doi.org/10.1016/j.fsigen.2014.11.004>
- Hohoff C, Dewa K, Sibbing U, Hoppe K, Forster P, Brinkmann B (2007) Y-chromosomal microsatellite mutation rates in a population sample from northwestern Germany. *Int J Legal Med* 121(5):359–363. <https://doi.org/10.1007/s00414-006-0123-9>
- Weng W, Liu H, Li S, Ge J, Wang H, Liu C (2013) Mutation rates at 16 Y-chromosome STRs in the South China Han population. *Int J Legal Med* 127(2):369–372. <https://doi.org/10.1007/s00414-012-0786-3>