Title: Analysis of genetic polymorphisms and mutations at 23 autosomal STR loci in Guangdong Han population

Authors: Ning Qu, Xiaochun Zhang, Hao Liang, Xueling Ou

PII: S1872-4973(18)30489-7
DOI: https://doi.org/10.1016/j.fsigen.2018.11.003
Reference: FSIGEN 1995

To appear in: *Forensic Science International: Genetics*

Received date: 3 September 2018

Please cite this article as: Qu N, Zhang X, Liang H, Ou X, Analysis of genetic polymorphisms and mutations at 23 autosomal STR loci in Guangdong Han population, *Forensic Science International: Genetics* (2018), https://doi.org/10.1016/j.fsigen.2018.11.003

Analysis of genetic polymorphisms and mutations at 23 autosomal STR loci in Guangdong Han population

Ning Qu[1,2], Xiaochun Zhang[1,2], Hao Liang[1,2], Xueling Ou[1,2]

[1.] *Faculty of Forensic Medicine, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, P. R. China*

[2]. *Guangdong Province Translational Forensic Medicine Engineering Technology Research Center, Sun Yatsen University, Guangzhou 510089, P.R. China.*

✉ Xueling Ou

Faculty of Forensic Medicine, Zhongshan School of Medicine, Sun Yat-sen University

Address: No.74 Zhongshan Road II, Guangzhou 510080, PR. China.

Email: ouxueling1997@163.com

Tel: +86 20 87330564; fax: +86 20 87334353.

Dear Editor,

Nowadays, short tandem repeat (STR) analysis has been regarded as the most important tool for paternity testing and personal identification in forensic casework. However, the high mutation rate of STR may lead to the incorrect interpretation in paternity testing. Moreover, elaborate forensic parameters and mutation characteristics of STR loci should be investigated to ensure the accuracy of parentage or kinship testing. In the present study, we investigated the forensic parameters and mutation rate of Guangdong Han population at 23 STR loci included in the Goldeneye™ DNA ID System 25A, which was newly developed and increasingly utilized in routine paternity testing and personal identification. Guangdong, as the most populous province of the People's Republic of China, has 102.3 million Han accounting for 98.02% of the province's total population at the 2010 census [1], providing a great

resource to study genetic polymorphisms and mutation of STR loci.

The study has been carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans. After acquiring informed consents, blood samples were collected from a total of 1,962 paternity cases from Guangdong Han population, including 980 trios and 982 duos that were tested in the Department of Forensic Science, Sun Yat-sen University during 2017. Genomic DNA was extracted from the blood samples using Chelex-100 methods. Twenty-three autosomal STR loci, D2S441, TPOX, D22S1045, D7S820, D1S1656, Penta E, D10S1248, D8S1179, D5S818, D19S433, D16S539, CSF1PO, Penta D, D3S1358, vWA, D2S1338, D18S51, D6S1043, D13S317, THO1, D12S391, D21S11, and FGA, were co-amplified with the Goldeneye™ DNA ID system 25A kit (Peoplespot, Beijing, China) on a GeneAmp PCR 9700 thermal cycler (Thermo Fisher Scientific, Waltham, MA, USA) [2]. The amplification products were separated on an ABI 3500 Genetic Analyzer (Thermo Fisher Scientific) and analyzed with the GeneMapper ® ID v 3.2 software (Thermo Fisher Scientific). All testing was conducted in accordance with the ISFG (International Society for Forensic Genetics) recommendations regarding STR polymorphisms and their nomenclature [3, 4]. The 9947A DNA was used as a positive control and ultrapure water as a negative control during the STR typing analysis.

Altogether, 2,942 unrelated individuals (1,647 males and 1,295 females) were selected from the paternity cases to study the genetic polymorphism of 23 autosomal STRs. The Modified-PowerStats version 1.2 spreadsheet (Promega, Madison, WI, USA) was used to evaluate allele frequencies and other forensic parameters, including the power of discrimination (PD), power of exclusion (PE), typical paternity index (TPI), and polymorphism information content (PIC), as well as the observed heterozygosity (Ho), expected heterozygosity (He), minimum allele frequency (MAF), probability values (*p*) of the Hardy-Weinberg equilibrium. The exact test for linkage disequilibrium (LD) were performed in ARELIQUIN v3.5.2 software [5]. The mutation was estimated based on the paternity cases. The number of mutations was counted directly. The data was analyzed using Excel and SPSS v20.0 software, and the 95% confidence intervals (CI) were calculated via an online confint.xla program (http://statpages. info/confint.html).

Allele frequencies and forensic parameters for the 23 STR loci of 2,942 unrelated Guangdong Han individuals are shown in Supplementary Table S1. A total of 339 alleles were observed with the range of

9 at VWA, D3S1358 and TPOX to 28 at Penta E. The allele frequency ranged from 0.0002 to 0.5469. Only two loci, D6S1043 (p = 0.0101) and FGA (p = 0.0093), showed significant deviations from Hardy-Weinberg expectations with the significance level of 0.05, however, no deviation from HWE and LD was detected after the Bonferroni adjustment (p = 0.05/23 = 0.0022 and p = 0.05/253 = 0.00020, respectively). Based on the analysis of forensic parameters, the locus Penta E was assumed to be the most discriminating with PD of 0.9853, PIC of 0.9030, PE of 0.8130 and TPI of 5.4684, and TPOX was the least discriminating with PD of 0.7753, PIC of 0.5377, PE of 0.2937 and TPI of 1.2573. The cumulative power of discrimination (CPD) and cumulative power of exclusion (CPE) of the 23 STR loci were equal to1-1.9125×10^{-28} and 0.9999999998.

For paternity cases analyzed, the cumulative paternity index (CPI) of each case was over 10000 to ascertain the parenthood, and 18 additional independent loci (D6S477, D18S535, D19S253, D15S659, D11S2368, D20S470, D22-GATA198B05, D7S3048, D8S1132, D4S2366, D21S1270, D13S325, D9S925, D9S925, D3S3045, D14S608, D17S1290, D5S2500) included in Microreader™23sp ID System (Suzhou Microread Genetics Co., Ltd, Suzhou, Jiangsu Province, China) [6], were used if the CPI of 10000 was not reached. In a total of 67,666 allelic transmissions of 2,942 meiotic transfers analyzed (Supplementary Table S2), 94 mutation events were observed at 19 of 23 STR loci in 92 cases, except 4 loci (D2S441, TPOX, D5S818, and TH01). The average mutation rate for the 23 loci was estimated to be 0.1389% (94/67,666) per meiosis (95%CI 0.00%-0.27%). The D12S391 locus exhibited the highest mutation rate of 0.3739% (95%CI 0.15%-0.59%), followed by the D8S1179, D3S1358, vWA and FGA, at which the mutation rates were all 0.2719% (95%CI 0.08%-0.46%). The mutation rates showed significant differences among the studied 23 STR loci when analyzed with Chi-square tests (p < 0.001). It was reported that the mutation rates were impacted by the number of repeats, the length of allele sequence and the nucleotide composition of repeat motif. In another Chinese Han population based on seven Chinese provinces (GuangDong, GuangXi, HuNan, HeNan, HuBei, JiangXi, and SiChuan), the D12S391, FGA, and vWA were also the loci with the highest mutation rate (0.2923%, 0.2663%, and 0.2403%, respectively); similarly, the TPOX and TH01 were the lowest [7]. However, the highest mutation rate was discovered for D18S51 marker (0.255%) in the Southern Chinese Han population [8] and D13S317 marker (0.63 %) in the mainland China [9].

The mutation was defined as any observed Mendelian inconsistency that cannot be attributable to a

null allele. The original and mutated allelic states are scored assuming the most probable mutation event (the minimum number of repeat motif differences). The origin was recorded as unknown when the use of this rule does not lead to an assignment of the event as a paternal or maternal meiotic transfer, as well as a gain or loss of a repeat unit could not be distinguished in some mutation events. Among the 94 mutation events, there were 83 paternal and 16 maternal mutations observed in the 1,647 paternal and 1,295 maternal meiosis, in which 5 mutations of unknown parental origin were included in both germlines (Supplementary Table S2 and S3). The ratio of paternal mutations versus maternal mutations was approximately 5.2:1. The difference of mutational origin was significant when analyzed with Chi-square tests ($p < 0.001$). That result was comparable with that reported by Sival et al.(5:1) [10] and D.J. Lu et al.(4.6:1)[7], but higher than that reported by Zhao Z et al. (1.43:1)[9]. It is generally agreed that male germ cells undergo 10 times more cell divisions than the female germ cells do and therefore have a much higher probability of mutation. Most mutational events were 1-step (93/94 = 98.94%) and one mutation at D10S1248 was 2-step. No > 2-step mutation was observed, which was in accordance with the stepwise mutation model [11]. In addition, the observed mutation events included 44 expansion mutations, 37 contraction mutations and 13 unassigned. The ratio of expansion mutations versus contraction mutations was 1.19:1, which was similar with previous reports [12,13]. It was also observed that the mutation rate increased with the paternal age at child birth (Supplementary Figure S1), which was consistent with other studies [8,14].

This study followed the guidelines for publication requested in the journal [15,16], and the dataset used in this study has successfully passed STRidER quality control (STRidER dataset reference STR000156) [17].

Acknowledgments

This work was supported by a grant from the National Natural Science Foundation of the People's Republic of China (Grants 81471825) and General Program of Guangzhou Science Technology and Innovation Commission (Grants 201804010247).

References

[1] Population Census Office under the State Council Department of Population, Employment Statistics National Bureau of Statistics, Tabulation of the 2010 Population Census of the People's Republic of

China Statistics Press, Beijing, 2012.

[2] Y.D. Sun, L.P. Cao, Forensic validation of the Goldeneye™ DNA ID 25A Kit. J. Forensic Med. 31 (2015) 284-286.

[3] P.J. Lincoln, DNA recommendations-further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. Forensic Sci. Int. 87 (1997) 181-184.

[4] W. Bar, B. Brinkmann, B. Budowle, A. Carracedo, P. Gill, P. Lincoln, W. Mayr, B. Olaisen, DNA recommendations. Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. International Society for Forensic Haemogenetics. Int. J. Legal. Med. 110 (1997):175-176.

[5] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0): an integrated software package for population genetics data analysis, Evol. Bioinform. Online 1 (2005) 47-50.

[6] Q.L. Liu, Z.X. Chen, C.G. Chen, D.J. Lu, Genetic polymorphism of 22 autosomal STR markers in a Han population of Southern China, Forensic Sci. Int. Genet. 24 (2016) e14-e16.

[7] Q.L. Liu, Y.F. Chen, X.L. Huang, K.Y. Liu, H. Zhao, D.J. Lu, Population data and mutation rates of 19 STR loci in seven provinces from China based on Goldeneye™ DNA ID System 20A, Int. J. Legal Med. 131 (2017) 653-656.

[8] H. Sun, S. Liu, Y. Zhang, M.R. Whittle, Comparison of southern Chinese Han and Brazilian Caucasian mutation rates at autosomal short tandem repeat loci used in human forensic genetics, Int. J. Legal Med. 128 (2014) 1-9.

[9] Z. Zhao, J. Zhang, H. Wang, Z.P. Liu, M. Liu, Y. Zhang, L. Sun, H. Zhang, Mutation rate estimation for 15 autosomal STR loci in a large population from Mainland China, Meta Gene 5 (2015) 150-156.

[10] A.C. Mardini, R. Rodenbusch, S. Schumacher, F.G. Chula, C.T. Michelon, A.Z. Gastaldo, L.P. MacieL, S.E. de Matos Almeida, C.M. da Silva, Mutation rate estimates for 13 STR loci in a large population from Rio Grande do Sul, Southern Brazil, Int. J. Legal Med. 127 (2013) 45-47.

[11] G. Levinson, G.A. Gutman, Slipped-strand mispairing a major mechanism for DNA sequence evolution, Mol. Biol. Evol. 4 (3) (1987) 203-221.

[12] B. Brinkmann, M. Klintschar, F. Neuhuber, J. Huhne, B. Rolf, Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat, Am. J. Hum. Genet. 62 (6) (1998) 1408-1415.

13] M. Sun, X. Zhang, D. Wu, Q. Shen, Y. Wu, S. Fu, Mutations of short tandem repeat loci in cases of

paternity testing in Chinese, Int. J. Legal Med. 130 (5) (2015) 1203-1204.

[14] Q.X. Qin, C.Y. Yin, Q. Ji, K. Li, H.T. Fan, Y.F. Yu, F.L. Bu, L.L. Hu, J.W. Wang, H.F. Mu, et al, Mutation rate analysis at 19 autosomal microsatellites, Electrophoresis 36 (2015) 1633-1639.

[15] A. Carracedo, J.M. Butler, L. Gusma˜o, A. Linacre, W. Parson, L. Roewer, P.M. Schneider, New guidelines for the publication of genetic population data, Forensic Sci. Int. Genet. 7 (2013) 217-220.

[16] A. Carracedo, J.M. Butler, L. Gusma˜o, A. Linacre, W. Parson, L. Roewer, P.M. Schneider, Update of the guidelines for the publication of genetic population data, Forensic Sci. Int. Genet. 10 (2014) A1-A2.

[17] M. Bodner, I. Bastisch, J.M. Butler, R. Fimmers, P. Gill , L. Gusmão, N. Morling, C. Phillips, M. Prinz, P.M. Schneider, W. Parson, Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER), Forensic Sci. Int. Genet. 24 (2016) 97-102.