

ORIGINAL ARTICLE

Variants in linkage status at D5S818 detected by multiple STR kits comparison and Sanger sequencing

Chengchen Shao² | Yining Yao² | Xinwei Pan³ | Mengde Wu³ | Beilei Zhang⁴ |
Hongmei Xu² | Jianhui Xie² | Kuan Sun^{1,2} 

¹Shanghai Key Laboratory of Forensic Medicine (Academy of Forensic Science), Shanghai, 200063, China

²Department of Forensic Medicine, School of Basic Medical Sciences, Fudan University, Shanghai, 200032, China

³Shanghai Hengping Judicial Expertise Center, Shanghai, 200070, China

⁴Fujian Zhengtai Judicial Expertise Center, Xiamen, Fujian 361000, China

Correspondence

Kuan Sun, Shanghai Key Laboratory of Forensic Medicine (Academy of Forensic Science), Shanghai, China.
Email: daisydaisysun@163.com

Funding information

National Natural Science Foundation of China, Grant/Award Number: 81901925

Abstract

Background: D5S818 discrepancies have been reported in forensic parental testing due to null alleles. However, more cases may be ignored since proportional null alleles were missed without detection of heredity discrepancy between parents and offspring.

Results: In this study, null allele 12 at D5S818 was detected by the PowerPlex[®] 21 System with a higher occurrence rate on the basis of review on 2824 samples from the 1282 routine cases in Chinese Han population. Sequencing results revealed novel variant of guanine (G) into adenine (A) in the 7th [AGAT] repeats in the core repeat region accompanied by rs1187948322 in the samples with null allele 12.

Conclusions: Forensic STR typing may benefit from this discovery: (1) primer design of CE profiling system could be improved for sensitive population and (2) polymorphic information could be enriched for the accuracy and precision of NGS genotyping system. Peak area of D5S818 was also analyzed through different commercial STR kits. It is suggested that more attention should be paid on observed homozygosity with reduced peak area, especially for the samples from Chinese Han population.

KEYWORDS

D5S818, linkage status, null alleles, STR typing, variant

1 | INTRODUCTION

Short tandem repeats (STRs) are widely used in forensic applications of individual identification and paternity testing for its high polymorphism (Butler, 2011). Discrepancies have been reported in concordance studies when the same sample was genotyped with different STR kits (W. Chen et al., 2014; Li et al., 2014; Mizuno et al., 2008; Ricci et al., 2007; Tsuji et al., 2010). Null alleles, which also called silent alleles, may lead to mismatch between parents and children since defective amplification

resulted from variations at primer binding sites. Point mutations or insertion or deletion (InDels) at the flanking regions of an STR locus may potentially affect annealing and/or elongation of primers in the PCR procedure, resulting in the dropout of one or both of the alleles. However, null alleles may be ignored in case the sister chromatid with the normal allele instead of that with the null allele was inherited to the offspring. Therefore, the probability of occurrence for null alleles are often underestimated. On the other hand, current studies are mainly aimed at solving the problem of Mendelian discrepancy across generations. Therefore, most

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Molecular Genetics & Genomic Medicine* published by Wiley Periodicals LLC

researches focused on consistency study among different detection systems and sequence validation in the flanking regions. However, sequence characteristics and the relationship between the flanking variants and the core repeat region of the STR locus needs further exploration. Linkage study is considered as an excellent tool for the validation of GWAS results because of the additional genetic information reflected by the linkage status (Wen et al., 2014). While in the forensic science, the linkage information of the adopted makers may provide potential clues for forensic investigation, such as the inference of biogeographic ancestry (Gattepaille & Jakobsson, 2012). Furthermore, specifically to the sequence polymorphism of forensic makers detected by NGS platform, the explicit linkage relationship for a target population could enrich the database for forensic practice.

Null alleles at D5S818 were more often observed, especially in Chinese Han population (Chen et al., 2014, 2015). To date, five sequence variations at primer binding region of D5S818 locus including rs576058164 (Jiang et al., 2011), rs182073376 (Alves et al., 2003; Delamoye et al., 2004), rs25768 (Edwards & Allen, 2004; Fujii et al., 2016), rs951218455 (Chen et al., 2014), and rs1187948322 (Chen et al., 2015; Yao et al., 2018) have been reported to cause discrepancies in paternity testing since null alleles. In this study, null allele 12 was detected in D5S818 with a higher occurrence rate from 1282 routine forensic cases by the PowerPlex[®] 21 System. A novel variant from Chinese Han population named as “D5S818[CE12]-Chr5-GRCh38 123775556–123775599 [ATCT]₁₂123775578-A; 123775689-T” was identified based on Sanger sequencing validation. Besides, Expressmarker[®] 22 PCR amplification kit, Identifiler[®] Plus PCR Amplification Kit. and PowerPlex[®] Fusion System were employed, respectively, for comparison study. Exploration on the sequence characteristics of D5S818 at the core repeat sequence and the flanking region was conducted to further enrich the polymorphic information of STRs for forensic genetics.

2 | MATERIALS AND METHODS

2.1 | DNA samples

A total of 1282 routine cases (including 26 individual identifications, 286 father–mother–child trios, 734 father–child duos, and 236 mother–child duos) from Chinese Han population were included in this study. Written informed consent from all individuals was obtained before sample collection. Ethics Approval was obtained from the ethics committee of School of Basic Medical Sciences, Fudan University. Genomic DNA was extracted from FTA card using a QIAamp[®] DNA Investigator Kit (Qiagen). DNA

quantification was performed using a Qubit 3 Fluorometer together with Qubit dsDNA HS Assay Kit (Thermo Fisher).

2.2 | STR genotyping

STR loci such as D3S1358, D1S1656, D6S1043, D13S317, Penta E, D16S539, D18S51, D2S1338, CSF1PO, Penta D, TH01, vWA, D21S11, D7S820, D5S818, TPOX, D8S1179, D12S391, D19S433 and FGA in the PowerPlex[®] 21 System (PP21, Promega), D3S1358, D13S317, D7S820, D16S539, Penta E, D2S441, TPOX, TH01, D2S1338, CSF1PO, Penta D, D10S1248, D19S433, vWA, D21S11, D18S51, D6S1043, D8S1179, D5S818, D12S391 and FGA in the Expressmarker[®] 22 PCR amplification kit (EX22, AGCU), D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX and D18S51 in the Identifiler[®] Plus PCR Amplification Kit (ID+, ThermoFisher Scientific), and D3S1358, D1S1656, D2S441, D10S1248, D13S317, Penta E, D16S539, D18S51, D2S1338, CSF1PO, Penta D, TH01, vWA, D21S11, D7S820, D5S818, TPOX, DTS391, D8S1179, D12S391, D19S433, FGA and D22S1045 in the PowerPlex[®] Fusion System (Fusion, Promega) were genotyped strictly as recommended by respective manufacturers in this study. The polymerase chain reaction (PCR) was performed by Mastercycler[®] nexus GSX1 (Eppendorf) according to the manufacturers' recommendations. PCR products were separated by capillary electrophoresis in an Applied Biosystems 3130xL Gene Analyzer (ThermoFisher Scientific). Allele designation was determined according to allelic ladders by using the GeneMapper[®] ID v3.2 or ID-X v1.4 (ThermoFisher Scientific).

2.3 | PCR amplification and DNA sequencing

PCR primers were designed on the basis of the GenBank D5S818 sequence (Accession No. AC008512.8): forward (FP): 5'-ACTTTGAGCTATTAGGCATGGGAGAG-3' (26mer); and reverse (RP): 5'-GCCTGTATAGTCATGTCCCTCTGTGTAG-3' (28mer). The thermal cycling parameters were enzyme activation at 95°C for 30s, followed by 30 cycles of denaturation at 95°C for 30 s, annealing at 60°C for 30 s and extension at 72°C for 1min, with a final extension at 72°C for 10 min. The PCR products were separated on nondenaturing polyacrylamide gel (T = 6%, C = 3.3%) and purified by the QIAquick1 Gel Extraction Kit (Qiagen). The separated allele fragments were cloned as per the standard procedure. Positive clones were selected and sequenced separately by sanger sequencing. Alignments of the sequences were performed with the reference sequence for verification.

2.4 | Data analysis

Peak areas were analyzed for the suspected samples with the null alleles at locus D5S818. Peak areas of all 601 samples detected by PP21 were exported as combined table by GeneMapper[®] software and the ratios of peak areas between D5S818 and D7S820 (the locus next to D5S818) were calculated as follows:

$$R_a = \frac{\text{Area1}_{D5S818}}{\text{Area1}_{D7S820} + \text{Area2}_{D7S820}}$$

2.5 | Quality control

The main experiments were conducted at the Forensic Genetics Laboratory of Fudan University, P.R. China, in accordance with quality control measures. All methods were carried out in accordance with the approved guidelines of Fudan University, P.R. China. The laboratory has been accredited by the China National Accreditation Service for Conformity Assessment (CNAS) which is also approved by the International Laboratory Accreditation Cooperation (ILAC).

3 | RESULTS

3.1 | Concordance study of D5S818 among various profiling systems

Genotyping profiles of 2824 samples from the 1282 routine cases with PP21 System were reviewed. Amplification kits EX22, ID+, and Fusion were also employed on 601 samples with observed homozygosity on D5S818 for comparison study and validation. When amplified and analyzed with PP21 only, seven samples from three cases show heredity discrepancy between parent and offspring with the

observed occurrence rate of 1.165%. However, discordant alleles were detected in 11 samples at D5S818 (occurrence rate 1.830%) with EX22. Four more samples were detected with null alleles at D5S818 except for the result from PP21. To confirm this results, ID+ and Fusion were also adopted for further comparison. Genotypes from ID+ were consistent with that from EX22 at locus D5S818, and the genotypes from Fusion supported the results from PP21. No discordant was detected in other overlapped STR loci. The genotyping results of D5S818 for the 11 suspected samples were listed in Table 1, and the electropherograms of one sample named C5C2 were displayed in Figure 1 as a representation. In the electropherograms analyzed with PP21 and Fusion, homozygous allele 11 was detected with deficient peak height at locus D5S818 when comparing with adjacent locus (as shown in Figure 1a,d). Profiling results of the same sample with EX22 and ID+ (as shown in Figure 1b,c) revealed that, genotype of D5S818 were heterozygous alleles 11 and 12 with balanced peak height and peak area. Besides, a tiny peak was observed in the position of allele 12 in the electropherogram from the PowerPlex[®] Fusion System. It cannot be recognized by GeneMapper since neither the peak height nor the peak area could meet the detection threshold.

3.2 | Validation through clone sequencing with plasmid vector

Sanger sequencing was adopted to validate the variant sequence at locus D5S818. Amplicon with target DNA fragment from PCR products was loaded into plasmid vector and sequenced with clone sequencing. As a result, the genotypes of D5S818 from the 11 suspected samples were all sequenced as heterozygosity with a shared allele 12, which was not detected by PP21, in accordance with the profiling result using EX22 and ID+. Furthermore, a variant in the core repeat region was detected in all of the

TABLE 1 Genotypes of D5S818 from 11 suspected samples using different kits

Sample	Gender	Category	PP21	EX22	Id+	Fusion
C1AF	M	Alleged Father	11,-	11,12	11,12	11,-
C1C	F	Child	10,-	10,12	10,12	10,-
C2AF	M	Alleged Father	10,-	10,12	10,12	10,-
C3AF	M	Alleged Father	9,-	9,12	9,12	9,-
C4AF	M	Alleged Father	11,-	11,12	11,12	11,-
C5AF	M	Alleged Father	9,-	9,12	9,12	9,-
C5C1	F	Child	11,-	11,12	11,12	11,-
C5C2	F	Child	11,-	11,12	11,12	11,-
C6C	F	Child	13,-	12,13	12,13	13,-
C7AM	F	Alleged Mother	10,-	10,12	10,12	10,-
C7C	F	Child	11,-	11,12	11,12	11,-

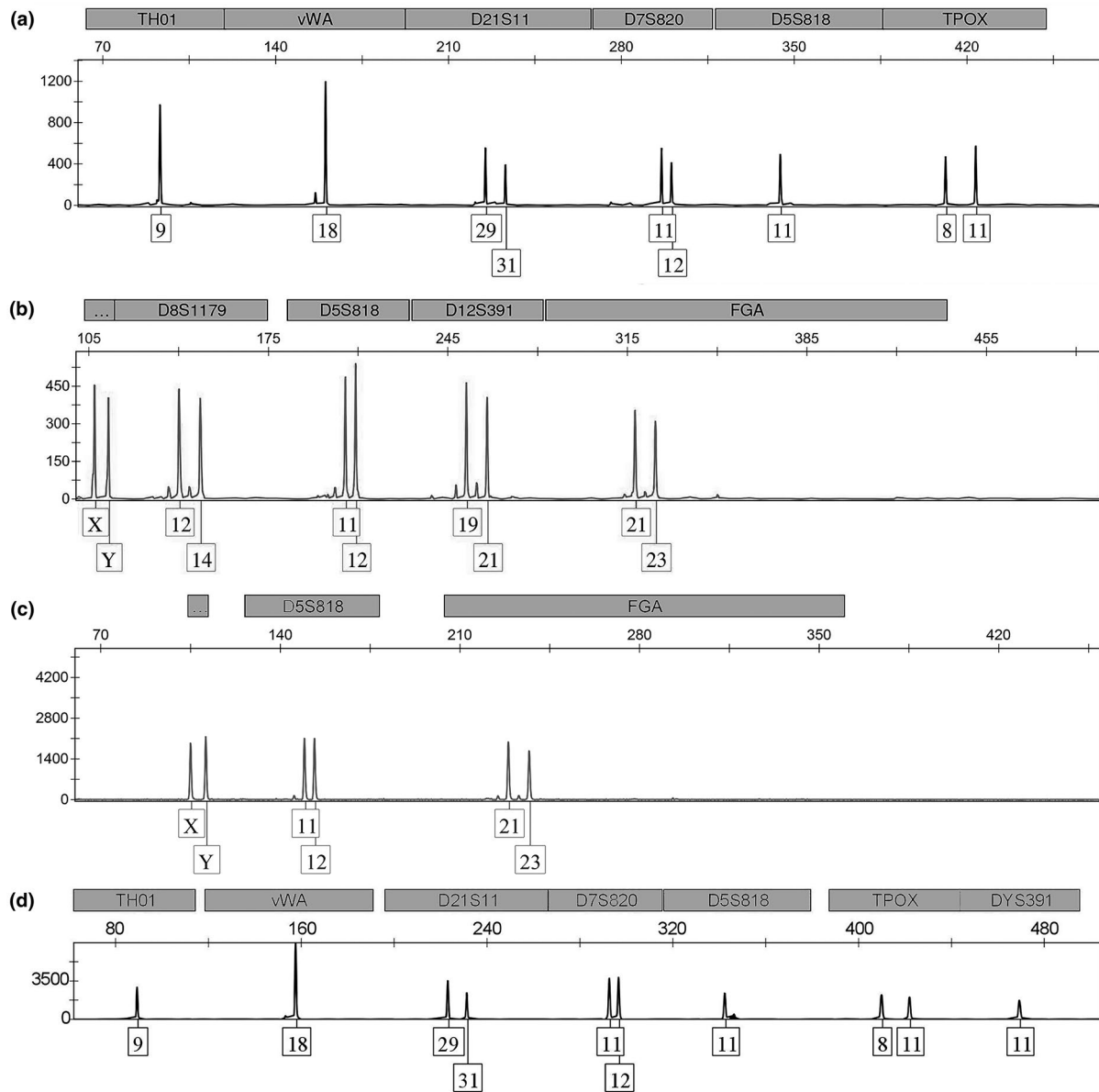


FIGURE 1 Electropherograms of D5S818 from sample C5C2 using four different multiplex kit. The fluorescence channels with D5S818 were shown for comparison. (a) Electropherograms of C5C2 by using PP21; (b) Electropherograms of C5C2 by using EX22; (c) Electropherograms of C5C2 by using ID+; (d) Electropherograms of C5C2 by using Fusion

clones with null allele 12. According to STRBase (Ruitberg et al., 2001), D5S818 is defined as a $[AGAT]_n$ simple repeats locus based on GenBank top strand. However, the core region of the null allele 12 were all sequenced as $[AGAT]_6AAAT[AGAT]_5$ rather than $[AGAT]_{12}$, as shown in Figure 2. Additionally, the null allele 12 from the 11 suspected samples present a C>A transversion at 90 bp upstream the 5' end of the repeat region when comparing with the sequence of common allele 12 (as shown in Figure 2). This point mutation has been termed as rs1187948322 in dbSNP. The two variants were also reported in the Genome Aggregation Database (gnomAD), but the linkage situation needs further verification (Lek et al., 2016). In our study, allele 12 with variant at rs1187948322 always show

$[AGAT]_6AAAT[AGAT]_5$ instead of $[AGAT]_{12}$ in the core repeat region of D5S818 locus. Latent linkage was supposed between these two variants, which may lead to erroneous decision when genotyping STR loci with CE.

3.3 | Peak area analysis in D5S818

Further investigation on the peak area was performed on D5S818. The ratios of peak areas between D5S818 and D7S820 (the locus next to D5S818) were calculated. And the relationship of the null alleles occurrence and the R_a values was analyzed, as shown by the violin plot in Figure 3. R_a of the common samples ranged from 0.799 to 1.556 while R_a

4 | DISCUSSION

In the concordance study of D5S818 among various profiling systems, four additional samples were detected with null alleles at locus D5S818 by re-genotyping, which may be easily ignored without extra genotyping since no discrepancy was detected between parent and offspring. The frequency of the null allele 12 at D5S818 among unrelated individuals was estimated as 0.1592% (4/2512) considering all reviewed samples in this study. Nevertheless, the frequency for the same null alleles in another Chinese Han population in Chen et al.'s study was about 0.0634% (3/4734) (Chen et al., 2015). Generally, null alleles are easily underestimated unless discrepant calls between alleged father/mother and child were observed. Another possible explanation may be the population difference of the polymorphism at the flanking region, which affects the primer binding in the process of DNA amplification for forensic detection. Null alleles may happen when the target DNA fragments are amplified inappropriately. For the same reason, the primer sequences in many common used amplification kits were mainly developed based on the sequence data from western populations. Mismatches occurred with a higher possibility in the primer binding process when the tested individual is from the Chinese Han population. Therefore, it is important to investigate and recognize the sequence characteristics in the flanking region of the STR loci for the target populations.

More importantly, despite the caution of repeated experiments with different profiling systems, the discovery of variants in linkage status on locus D5S818 may enrich the genetic data for interpretation of STR typing with NGS. By combing the sequence polymorphism with length polymorphism, STR analysis based on NGS show promising prospect in forensic genetics. The study on the polymorphic feature in the flanking region and its potential relationship with the core repeat sequence of the STR locus would benefit the forensic STR research and therefore, improve the resolution of forensic markers. Additional information such as biogeographic origin may also be reflected from the linkage status of variants if more reference populations were included. In 2016, International Society for Forensic Genetics (ISFG) recommended a new system of STR allele nomenclature based on sequence information (Parson et al., 2016). The allele described in this study could be named as "D5S818[CE12]-Chr5-GRCh38 123775556–123775599 [ATCT]₁₂123775574-T; 123775689-T" according to the recommendations.

Furthermore, the balance within a locus always got enough focus since discordant proportions between two alleles of heterozygotes may imply mixed samples. However, the balance between different loci were less concerned in consideration of specific features for different loci. Peak

area of D5S818 analyzed through 2824 samples (as listed in Figure 3) suggested that the balance between loci in a multiplex system should be paid more attention for the occurrence of null alleles.

5 | CONCLUSIONS

In this study, a novel variant was discovered in the core repeat region of D5S818 in Chinese Han population, and it is observed to be in linkage with a reported variation named rs1187948322. Polymorphic information of forensic STRs could be enriched for forensic application. Besides, additional attention was suggested on observed homozygosity with reduced peak area in D5S818, since null alleles occur more frequently in this situation than previously estimated.

ACKNOWLEDGMENT

The authors express their gratitude to Exome Consortium for supplying additional information from gnomAD. This work was supported by National Natural Science Foundation of China (No. 81901925).

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare that are relevant to the content of this article.

AUTHORS' CONTRIBUTIONS

C.S. and K.S. conceived the idea of the study; X.P., M.W. and B.Z. analyzed the data; Y.Y., J.X., and H.X. interpreted the results; C.S. and K.S. wrote the paper; all authors discussed the results and revised the manuscript.

COMPLIANCE WITH ETHICAL STANDARDS

Approval was obtained from the ethics committee of School of Basic Medical Sciences, Fudan University. The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

DATA AVAILABILITY STATEMENT

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

ORCID

Kuan Sun  <https://orcid.org/0000-0002-9087-438X>

REFERENCES

- Alves, C., Gusmao, L., Pereira, L., & Amorim, A. (2003). Multiplex STR genotyping: comparison study, population data and new sequence information. *Progress in Forensic Genetics*, 9(1239), 131–135. [https://doi.org/10.1016/S0531-5131\(02\)00623-4](https://doi.org/10.1016/S0531-5131(02)00623-4)

- Butler, J. M. (2011). *Advanced topics in forensic DNA typing: Methodology* (pp. 569–585). Elsevier Inc.
- Chen, L., Tai, Y., Qiu, P., Du, W., & Liu, C. (2015). A silent allele in the locus D5S818 contained within the PowerPlex(R)21 PCR Amplification Kit. *Leg Med (Tokyo)*, 17(6), 509–511. <https://doi.org/10.1016/j.legalmed.2015.10.012>
- Chen, W., Cheng, J., Ou, X., Chen, Y., Tong, D., & Sun, H. (2014). Identification of the sequence variations of 15 autosomal STR loci in a Chinese population. *Annals of Human Biology*, 41(6), 524–530. <https://doi.org/10.3109/03014460.2014.897754>
- Delamoye, M., Duverneuil, C., Riva, K., Leterreux, M., Taieb, S., & De Mazancourt, P. (2004). False homozygosities at various loci revealed by discrepancies between commercial kits: implications for genetic databases. *Forensic Science International*, 143(1), 47–52. <https://doi.org/10.1016/j.forsciint.2004.02.001>
- Edwards, M., & Allen, R. W. (2004). Characteristics of mutations at the D5S818 locus studied with a tightly linked marker. *Transfusion*, 44(1), 83–90. <https://doi.org/10.1111/j.0041-1132.2004.00621.x>
- Fujii, K., Watahiki, H., Mita, Y., Iwashima, Y., Kitayama, T., Nakahara, H., & Sekiguchi, K. (2016). Typing concordance between PowerPlex((R)) Fusion and GlobalFiler((R)) based on 1501 Japanese individuals and the causes of typing discrepancies. *Forensic Science International: Genetics*, 25, e12–e13. <https://doi.org/10.1016/j.fsigen.2016.07.023>
- Gattepaille, L. M., & Jakobsson, M. (2012). Combining markers into haplotypes can improve population structure inference. *Genetics*, 190(1), 159–174. <https://doi.org/10.1534/genetics.111.131136>
- Jiang, W., Kline, M., Hu, P., & Wang, Y. (2011). Identification of dual false indirect exclusions on the D5S818 and FGA loci. *Legal Medicine*, 13(1), 30–34. <https://doi.org/10.1016/j.legalmed.2010.08.006>
- Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., Tukiainen, T., Birnbaum, D. P., Kosmicki, J. A., Duncan, L. E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., ... MacArthur, D. G. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536(7616), 285–291. <https://doi.org/10.1038/nature19057>
- Li, F. R., Xuan, J. F., Xing, J. X., Ding, M., Wang, B. J., & Pang, H. (2014). Identification of new primer binding site mutations at TH01 and D13S317 loci and determination of their corresponding STR alleles by allele-specific PCR. *Forensic Science International-Genetics*, 8(1), 143–146. <https://doi.org/10.1016/j.fsigen.2013.08.013>
- Mizuno, N., Kitayama, T., Fujii, K., Nakahara, H., Yoshida, K., Sekiguchi, K., Yonezawa, N., Nakano, M., & Kasai, K. (2008). A D19S433 primer binding site mutation and the frequency in Japanese of the silent allele it causes. *Journal of Forensic Sciences*, 53(5), 1068–1073. <https://doi.org/10.1111/j.1556-4029.2008.00806.x>
- Parson, W., Ballard, D., Budowle, B., Butler, J. M., Gettings, K. B., Gill, P., Gusmão, L., Hares, D. R., Irwin, J. A., King, J. L., Knijff, P. D., Morling, N., Prinz, M., Schneider, P. M., Neste, C. V., Willuweit, S., & Phillips, C. (2016). Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. *Forensic Science International: Genetics*, 22, 54–63. <https://doi.org/10.1016/j.fsigen.2016.01.009>
- Ricci, U., Melean, G., Robino, C., & Genuardi, M. (2007). A single mutation in the FGA locus responsible for false homozygosities and discrepancies between commercial kits in an unusual paternity test case. *Journal of Forensic Sciences*, 52(2), 393–396. <https://doi.org/10.1111/j.1556-4029.2006.00357.x>
- Ruitberg, C. M., Reeder, D. J., & Butler, J. M. (2001). STRBase: A short tandem repeat DNA database for the human identity testing community. *Nucleic Acids Research*, 29(1), 320–322. <https://doi.org/10.1093/nar/29.1.320>
- Tsuji, A., Ishiko, A., Umehara, T., Usumoto, Y., Hikiji, W., Kudo, K., & Ikeda, N. (2010). A silent allele in the locus D19S433 contained within the AmpFISTR (R) Identifiler (TM) PCR Amplification Kit. *Legal Medicine*, 12(2), 94–96. <https://doi.org/10.1016/j.legalmed.2009.12.002>
- Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., Liu, J., Liu, H., Chen, W., Luo, J., & Yan, J. (2014). Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nature Communications*, 5, 3438. <https://doi.org/10.1038/ncomms4438>
- Yao, Y., Yang, Q., Shao, C., Liu, B., Zhou, Y., Xu, H., Zhou, Y., Tang, Q., & Xie, J. (2018). Null alleles and sequence variations at primer binding sites of STR loci within multiplex typing systems. *Legal Medicine*, 30, 10–13. <https://doi.org/10.1016/j.legalmed.2017.10.007>

How to cite this article: Shao, C., Yao, Y., Pan, X., Wu, M., Zhang, B., Xu, H., Xie, J., & Sun, K. (2021). Variants in linkage status at D5S818 detected by multiple STR kits comparison and Sanger sequencing. *Molecular Genetics & Genomic Medicine*, 9, e1765. <https://doi.org/10.1002/mgg3.1765>