Received: 30 August 2022 | Revised: 3 February 2023 | Accepted: 13 February 2023

DOI: 10.1002/elps.202200216

ELECTROPHORESIS

RESEARCH ARTICLE

# Analysis of five rare alleles at the STR loci D1S1656, D12S391, D13S317, Penta D, and D2S441

**Yequan Wang**[1] | **Ao Gao**[1] | **Zhenzhen Dong**[2] | **Dan Wang**[1]

[1]Institute of Forensic Medicine and Laboratory Medicine, Jining Medical University, Jining, Shandong, P. R. China

[2]Jining Public Security Bureau, Jining, Shandong, P. R. China

**Correspondence**
Dan Wang, Institute of Forensic Medicine and Laboratory Medicine, Jining Medical University, Jining 272067, Shandong, P. R. China.
Email: wangdanchn@163.com

**Color online**: See article online to view Figures 1–6 in color.

**Funding information**
Student's Platform for Innovation and Entrepreneurship Training Program, Grant/Award Numbers: S202010443001, cx2020001; Research Fund for Academician Lin He New Medicine, Grant/Award Number: JYHL2022MS23; Shanxi Provincial Key Laboratory of Forensic Science, China, Grant/Award Number: SFM2019002

**Abstract**

Short tandem repeat (STR) automatic typing technology is extensively used in forensic laboratories with commercial kits, in rare cases genotyping misinterpretations or mislabeling may occur due to unexpected rare alleles. This study refers to the investigation of several rare alleles observed from routine cases. Besides cross-kit verification with Goldeneye 25A (Beijing PeopleSpot Inc, China) and Huaxia platinum (Thermo Fisher Scientific, USA) kits, the next-generation sequencing technology by MiSeq FGx System (Illumina, USA) was applied to further validation. To solve the inconsistent outcomes reached by the above mentioned approaches at D2S441 locus, single gene amplification, gene cloning, and genetic sequencing was also performed. As a result, five rare alleles were detected. Two novel alleles of allele 3 at the D13S317 locus and allele 5 at the D2S441 locus were found; three previously reported alleles of allele 9 at D1S1656 locus, allele 19 at Penta D locus, and allele 28 at D12S391 locus in STRBase were initially supplemented with sequence information. We, therefore, propose that such uncommon observations with rare events should be carefully investigated and interpreted.

**KEYWORDS**
CE typing, cloning, next-generation sequencing, short tandem repeat, variant alleles

## 1 | INTRODUCTION

A short tandem repeat (STR) is a type of DNA genetic marker formed by repeats of 2–8 base pairs (bp) as the core sequence, which follow Mendel's law of inheritance [1]. Due to its high degree of genetic polymorphism, STR genotyping is widely used in forensic genetics for paternity testing and individual identification [2]. With the significantly increased number of tested populations and applied STR loci, the observation of rare alleles, including the pattern of off-ladder (OL) alleles is increasing accordingly. The occurrence of rare alleles often brings issues to the analysis of typing results. In this study, we collected cases from routine forensic investigations to describe several situations in which rare alleles occur. In addition to cross-kit verification by capillary electrophoresis (CE), next-generation sequencing (NGS), and molecular cloning techniques were applied to further validation in order to provide substantiation in the identification of these rare alleles.

**Abbreviations:** NGS, next-generation sequencing; OL, off-ladder alleles.

## 2 | MATERIALS AND METHODS

### 2.1 | Subjects

The five unrelated individuals involved in this study were collected from routine casework. The blood specimens

  

were collected on FTA cards (Whatman, GE Healthcare), left to dry naturally, and stored at room temperature. This study was carried out with approval from the Ethics Committee of Jining Medical University (jnmc-2019-zr-0027). Written informed consent was obtained from all participants.

## 2.2 | DNA extraction and sequencing analysis by CE typing

Genomic DNA was extracted from blood stains using a Chelex-100 protocol [3]. PCR (polymerase chain reaction) amplification was performed in a GeneAmp PCR System 9700 Thermal Cycler (Thermo Fisher Scientific, USA) with two different STR multiplex kits—Goldeneye 25A kit (Beijing PeopleSpot, China) and Huaxia platinum kit (Thermo Fisher Scientific), according to the manufacturers' instructions. Analyses of PCR products were performed on an ABI3500 Genetic Analyzer instrument (Thermo Fisher Scientific, USA). Data analysis was carried out using GeneMapper ID V3.2 software (Thermo Fisher Scientific). All steps were performed according to the laboratory internal control standards and kit controls.

## 2.3 | "Off-ladder" alleles calculations

For cases which OL appeared in the electropherograms using CE typing, and after determining which locus the OL belonged to, allele designation for OL was calculated by reference to the nearest allele size in the respective allelic ladders. The OL alleles were run under the same electrophoretic conditions as those of the variant allele. A decimal deviation was rounded to nearest whole number.

## 2.4 | Sequencing analysis by MiSeq FGx

Genomic DNA for NGS was extracted by QIAamp DNA Investigator Kit (Qiagen, USA) from blood stains according to the manufacturer's instructions. DNA concentrations were measured by Qubit 3.0 Fluorometer (Thermo Fisher Scientific, USA) and diluted to 0.2 ng/μL prior to PCR. The constructions of DNA libraries were performed using a ForenSeq DNA Signature Prep Kit (Verogen Inc., USA). The sequencing of DNA libraries was conducted on an integrated workflow of MiSeq FGx System (Verogen Inc.). Sequencing data were analyzed by ForenSeq Universal Analysis Software (Verogen Inc.).

## 2.5 | Cloning of PCR amplicon and Sanger sequencing for D2S441 locus

Nested-PCR [4] was further performed to amplify the D2S441 locus in case 5. The gene sequence of D2S441 was downloaded from NCBI (https://www.ncbi.nlm.nih.gov/), and primers covering this region were designed by Primer3Plus [5] and are as follows: outer primer forward: 5′-GCTTCCTGAACCCAGTCCTC-3′, outer primer reverse: 5′-CGTGCCATCATGTCCAGCTA-3′, inner primer forward: 5′-GTCCTCTTGGGGTTTGAGGG-3′, and inner primer reverse: 5′-CCCAGGCTGGTCTTGAACTT-3′. The first-round PCR was performed in a volume of 20.0 μL with 1.0 μL of DNA template (0.2 ng/μL), 1.0 μL of each outer primer (10 μM), 2.0 μL of PCR buffer (10×), 0.2 μL of DNA polymerase (5 units/μL), and 0.25 μL dNTP mix (20 mM each, pH 8.0) (Takara, Japan). The reaction conditions consisted of initial denaturation at 95°C for 10 min; 26 cycles of 95°C for 40 s, 54°C for 40 s, and 72°C for 40 s; and a final extension step for 10 min at 72°C. Subsequently, the second-round PCR was performed in 20.0 μL with 0.5 μL of DNA template derived from the product amplified by first-round PCR, 1.0 μL of each inner primer (10 μM), 2.0 μL of PCR buffer (10×), 0.2 μL of DNA polymerase (5 units/μL), and 0.25 μL dNTP mix (20 mM each, pH 8.0) (Takara). The reaction conditions consisted of initial denaturation at 95°C for 10 min; 32 cycles of 95°C for 35 s, 54°C for 35 s, and 72°C for 35 s; and a final extension step for 10 min at 72°C. The amplification was carried out in the GeneAmp PCR System 9700 Thermal Cycler (Thermo Fisher Scientific). The amplified products were then separated on a 1.2% agarose gel (Sigma-Aldrich, USA), and the two close bands were gel-cut together and purified using the TIANgel Midi Purification Kit (TIANGEN, China). PCR products were ligated to the pGEM-T Easy vector (Promega, USA) and cloned into *Escherichia coli* DH5α [5]. Positive clones were sequenced by Sanger sequencing (Sangon Biotech, China).

# 3 | RESULTS AND DISCUSSION

In this study, we aimed to report a variety of rare variant alleles that were detected during routine sample analysis.

For case 1, the D13S317 locus was observed as a tri-allele pattern of 8/11/16, followed by a homozygous allele 11 in D1S1656 by Goldeneye 25A typing. After typing with the Huaxia platinum kit, a heterozygote appeared at D13S317 locus with alleles 8/11 and D1S1656 locus with 9/11 (Figure 1). Combining the outcomes of the two genotyping systems, it can be inferred that the allele 16 assigned at D13S317 locus with Goldeneye 25A kit should be allele 9 at

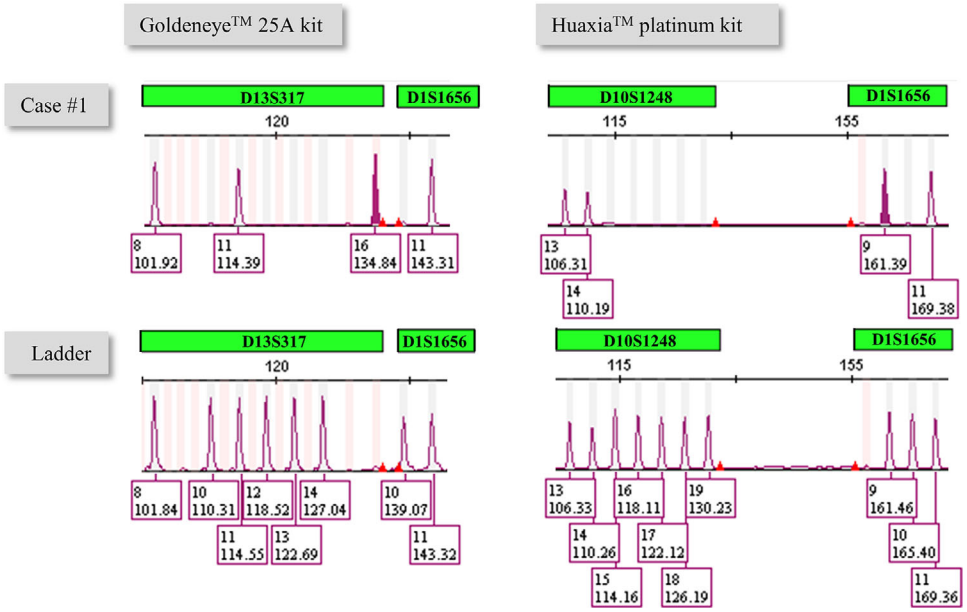


**FIGURE 1** Electropherograms for case 1 and its corresponding allelic ladder using two short tandem repeat (STR) kits. Variant allele with Goldeneye 25A kit appears as allele 16 within D13S317 locus, and verified as D1S1656 9 allele according to Huaxia platinum typing.

**TABLE 1** Sequencing results for variant alleles in our studied cases.

| Case ID | locus | allele | Repeat structure |
|---|---|---|---|
| 1 | D1S1656 | 9 | $[TAGA]_9$ |
| 2 | D12S391 | 28 | $[AGAT]_{19}[AGAC]_8[AGAT]_1$ |
| 3 | D13S317 | 3 | $[TATC]_3$ |
| 4 | Penta D | 19 | $[AAAGA]_{19}$ |
| 5 | D2S441 | 5 | $[TCTA]_5$ |

the D1S1656 locus. It was then corroborated by the MiSeq sequencing result with D1S1656 locus bearing an allele of nine [TAGA] complete repeat units as its core sequence (Table 1).

In case 2, an OL was observed in the gap between D12S391 and TPOX by Goldeneye 25A typing, with the former carrying an allele 20 and the latter 8/11. Huaxia platinum typing showed that an OL fell downstream of the D12S391 locus with allele 20 but upstream of the neighboring D2S1338 locus with alleles 23/24 (Figure 2). Accordingly, it is presumed that the OL should be a rare large size allele belonging to D12S391 locus. After calculation by comparing with the corresponding allelic ladder, the OL was 3.98 and 4.08 bp larger than allele 27 at D12S391 locus after drift correction in the two typing systems, respectively (Figure 2). As D12S391 is a complex STR with two variable sub-repeats of 4 bp, this OL was identified as 28. Further NGS analysis revealed it carrying complicated core sequence of $[AGAT]_{19}[AGAC]_8[AGAT]_1$ (Table 1).

Briefly, the typing pattern which rare allele invades into the adjacent locus or falls between two loci may cause misinterpretation or even lead to typing errors in forensic DNA investigations. The primary step was to ascertain the attribution of the rare allele before calculating its size. It can usually be solved by multi-kit review, and sometimes sequencing verification is required.

D13S317 is the locus with the smallest fragment in its fluorescence spectrum in the Goldeneye 25A kit. In case 3, a rare small peak appeared upstream of the bins of the D13S317 locus by Goldeneye 25A typing. By Huaxia platinum typing, an OL was found falling between the D5S818 and D13S317 loci, with the former being a 10/11 and the latter an 8 allele (Figure 3). Thus, it was speculated that this OL was a rare small allele at the D13S317 locus. After calculation, this OL was 15.98 bp smaller than allele 7 at D13S317 by Goldeneye 25A typing and 7.75 bp smaller than allele 5 by Huaxia platinum typing (Figure 3). It was accordingly assigned as allele 3 at the D13S317 locus. NGS analysis verified the core sequence of this allele consisting of three complete [TATC] repeated units (Table 1).

The Penta D locus appeared as the largest fragment in both Goldeneye 25A and Huaxia platinum kits in the framework of the standard allele ladder mixture of their corresponding fluorescence spectrum. In case 4, an extremely long OL downstream of the allele ladder of Penta D locus was observed with both kits. By comparing the largest allele of the ladder, it was found to be 15.23 bp larger than allele 16 and 9.67 bp larger than allele 17 in the two genotyping systems, respectively (Figure 4). In view of characteristics of 5 bp repeat in the core sequence of Penta D locus, the rare variant was calculated as 19 at

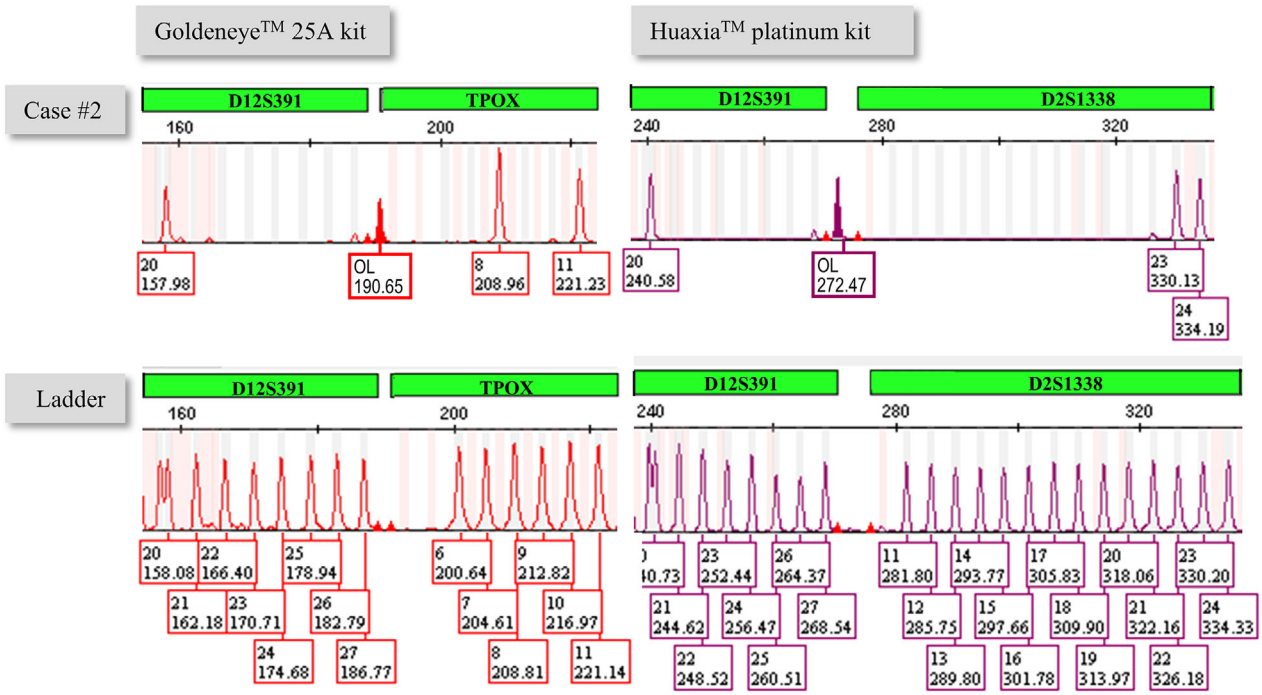


**FIGURE 2** Electropherograms for case 2 and its corresponding allelic ladder using two short tandem repeat (STR) kits. Variant allele with Goldeneye 25A kit appears as an off-ladder (OL) between D12S391 and TPOX loci, and an OL between D12S391 and D2S1338 loci with Huaxia platinum kit. It was calculated as allele 28 at D12S391 according to allelic ladder.

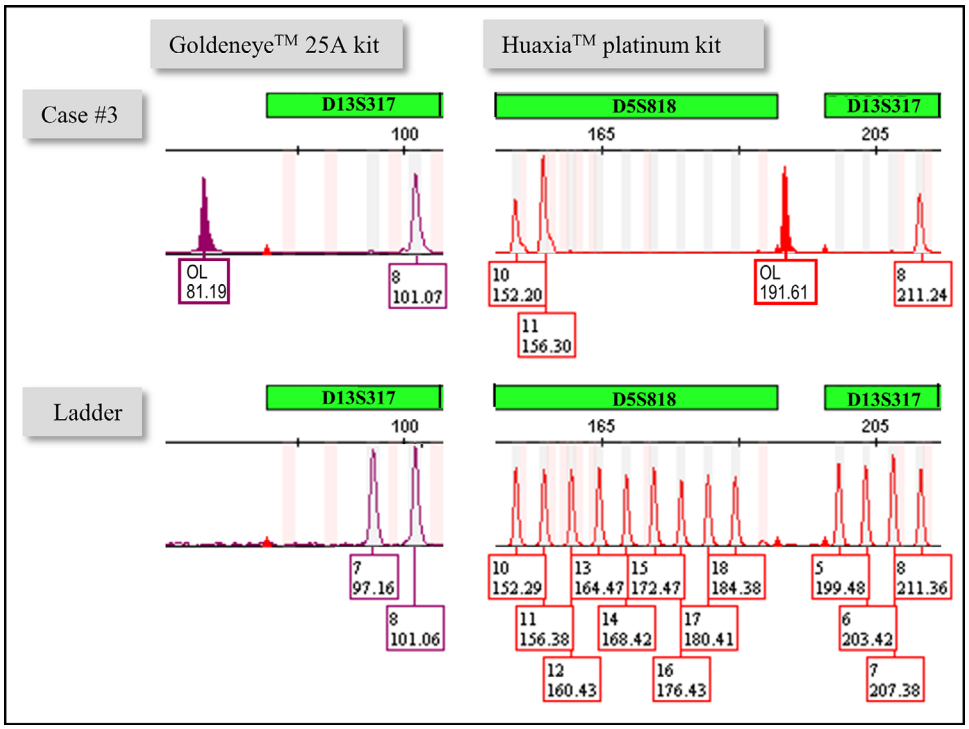


**FIGURE 3** Electropherograms for case 3 and its corresponding allelic ladder using two short tandem repeat (STR) kits. Variant allele appeared ahead of D13S317 locus with Goldeneye 25A kit and fell between D5S818 and D13S317 loci by Huaxia platinum typing. It was designated as allele 3 at D13S317 according to allelic ladder.

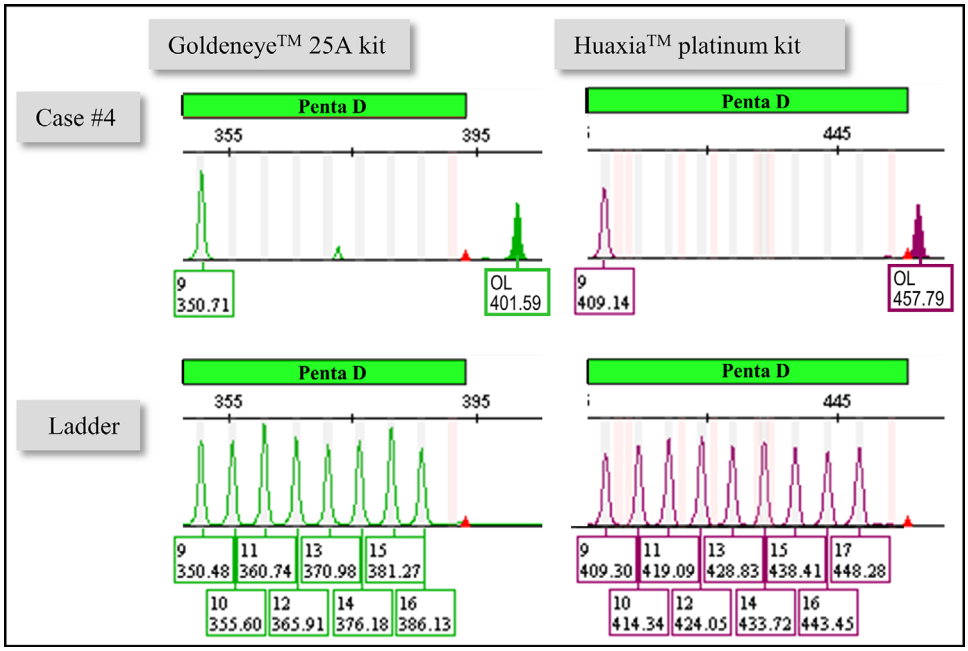


**FIGURE 4** Electropherograms for case 4 and its corresponding allelic ladder using two short tandem repeat (STR) kits. Variant allele appeared beyond Penta D locus with both Goldeneye 25A and Huaxia platinum kits. It was verified as Penta D 19 according to allelic ladder.

Penta D locus and was confirmed by NGS analysis with the repetitive motif of $[AAAGA]_{19}$ (Table 1).

For loci which were either downstream or upstream of the kit panel, it may be not difficult to identify extra OL alleles just as in the locus Penta D in case 4. However, when the allele was extremely beyond predefined analysis range, extrapolation of relative size outside of the size range of the allelic ladders can give erroneous results in CE. In this case, a further sequencing method should be considered to draw a conclusion.

For case 5, a three-band pattern of 11/12/OL appeared at CSF1PO locus with Goldeneye 25A typing, followed by a homozygote 11 in the neighboring locus D2S441. By Huaxia platinum typing, an upstream allele peak appeared before the size standard range which proved to belong to the locus D2S441 (Figure 5). It appeared that the rare allele was 13.54 bp smaller than allele 8.1 by Goldeneye 25A typing and 12.50 bp smaller than allele 8 by Huaxia platinum typing (Figure 5). The variant allele designation was thus calculated as 4.3. However, the subsequent NGS analysis showed a result of five complete [TCTA] repeated units in this allele, which had 1 bp discordance with the outcomes of 3500 typing. Considering that allelic size generated by CE typing also contains flanking sequence, we then assumed that the 1 bp loss might occur in its flanking regions other than the core sequence.

By designing primers and performing monogenic nested-PCR amplification, cloning, and Sanger sequencing, a fragment of 512 bp in D2S441 was obtained. It contained a flanking sequence of 268 and 224 bp in its 5′ and 3′ ends, respectively (Figure 6). A sequencing result was consistent with the NGS analysis that the core region which contained five complete [TCTA] repeated units, whereas no base deletion occurred in the entire region (Figure 7). Taken together, the correct designation of the rare allele should be a 5 allele at the D2S441 locus.

According to above results, the rare allele of 4.3 designated by CE typing in case 5 was verified to be an error. This result revealed the potential issues when applying a rounding-off method to calculate the precise size of rare variants. Simply rounding a number with decimals close to 0.5 up or down to the nearest whole number may lead to mistaken results. For such cases, further NGS or single gene amplification and sequencing would be necessary to be performed to obtain the correct allelic designation.

## 4 | CONCLUDING REMARKS

Taken together, several rare alleles were investigated in our study. To our knowledge, two novel alleles with their sequence data were described for the first time, including allele 3 for D13S317 locus and allele 5 for D2S441 locus. We also complemented the sequence information of allele 9 for D1S1656 locus, allele 19 for Penta D locus, and allele 28 for D12S391 locus which had been reported in STRBase [6] without sequence information. Even though the occurrence of infrequent rare alleles usually causes misjudgments and affects information exchange between

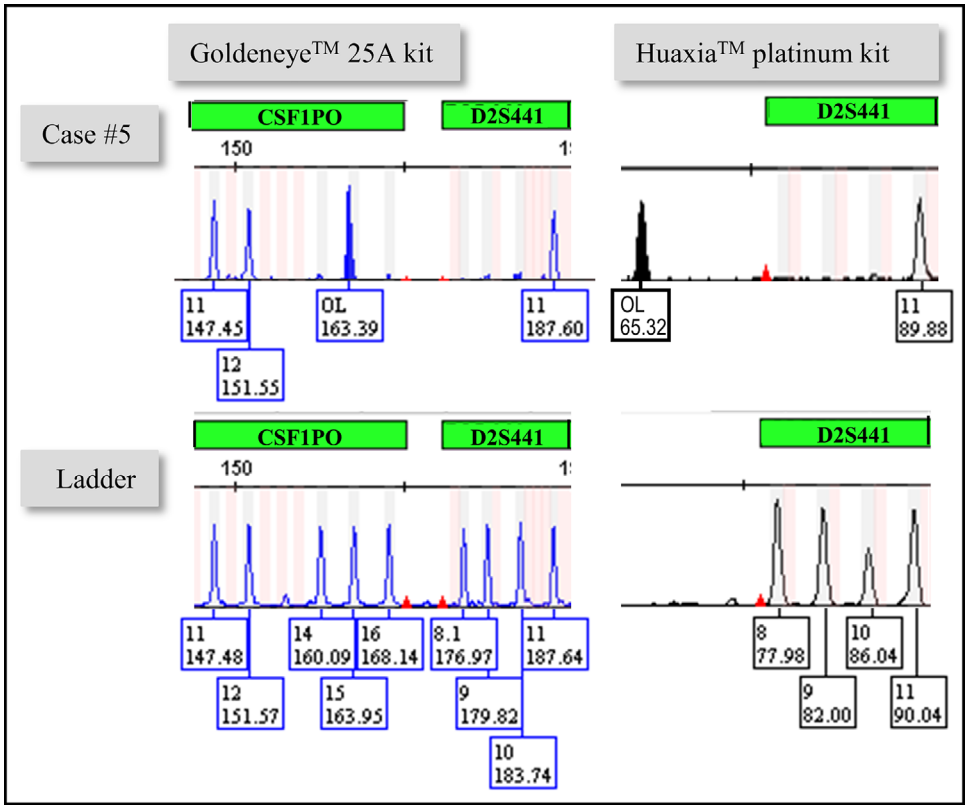


**FIGURE 5** Electropherograms for case 5 and its corresponding allelic ladder using two short tandem repeat (STR) kits. Variant allele appeared as an off-ladder (OL) at CSF1PO locus with Goldeneye 25A kit and fell upstream of the D2S441 locus by Huaxia platinum typing. It was calculated as allele 4.3 at D2S441 according to allelic ladder.

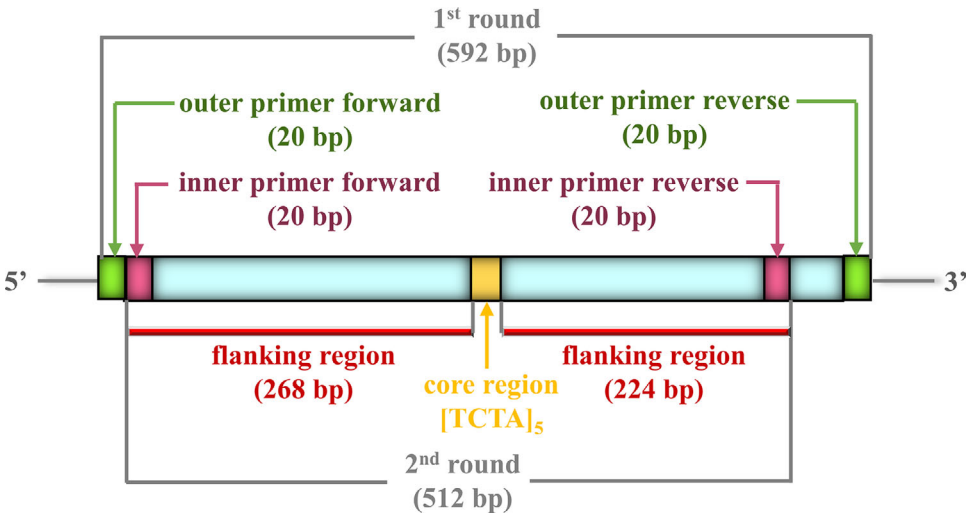


**FIGURE 6** Nested-PCR strategy for D2S441 in case 5. Two pairs of PCR primers were used. The first pair of primers (outer primer forward and out primer reverse; green bars) amplified as in a normal PCR with product of 592 bp in length. The second pair of primers (inner primer forward and inner primer reverse; pink bars) was the nested set internal to the first primer binding sites. The length of the final PCR product after the second round of PCR was 512 bp, including flanking sequence of 268 bp (pink plus blue bars on the left of yellow bar) and 224 bp (blue plus pink bars on the right of yellow bar) in its 5′ and 3′ ends, and a core region with five complete [TCTA] repeated units (yellow bar).

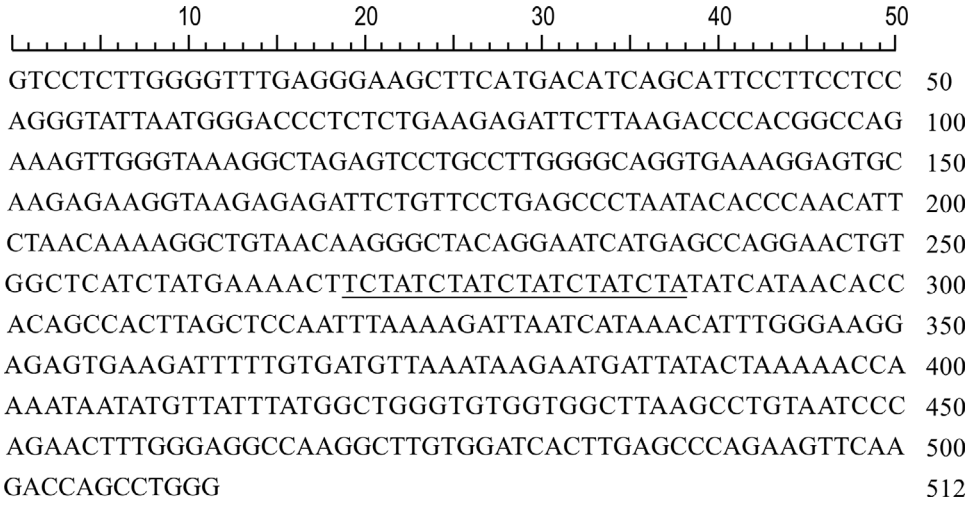


**FIGURE 7** Sanger sequencing result of the off-ladder (OL) allele at D2S441 locus in case 5 (the core sequence is underlined).

laboratories, it expanded the scope of STR loci in population and is of great benefit for individual identification and kinship testing. The rare variant events described above were technically caused by inadequate panel length designed for the corresponding locus of the kit. Therefore, kit manufacturers may take this into account to adjust the reading range for STR markers to ensure that newly discovered alleles could be detected. Moreover, we showed the importance of applying multiple techniques including different commercial kits and validation by CE typing, the NGS method, and possibly molecular cloning to confirm observations.

## AUTHOR CONTRIBUTIONS

Conceptualization and methodology: Yequan Wang. Software, investigation, writing—original draft: Ao Gao and Zhenzhen Dong. Supervision, validation, writing—review, and editing: Dan Wang.

### ACKNOWLEDGEMENTS

This work was supported by the Student's Platform for Innovation and Entrepreneurship Training Program (S202010443001, cx2020001), the Research Fund for Academician Lin He New Medicine (JYHL2022MS23), and an open project of Shanxi Provincial Key Laboratory of Forensic Science, China (SFM2019002). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### CONFLICT OF INTEREST STATEMENT

The authors have declared no conflict of interest.

### DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

### ORCID

*Dan Wang* https://orcid.org/0000-0002-6032-5638

### REFERENCES

1. Lygo JE, Johnson PE, Holdaway DJ, Woodroffe S, Whitaker JP, Clayton TM, et al. The validation of short tandem repeat (STR) loci for use in forensic casework. Int J Legal Med. 1994;107:77–89.
2. Butler JM. Short tandem repeat typing technologies used in human identify testing. Biotechniques. 2007;43:2–5.
3. Walsh PS, Metzger DA, Higuchi R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques. 1991;10:506–13.
4. Green MR, Sambrook J. Nested polymerase chain reaction (PCR). Cold Spring Harb Protoc. 2019;2:175–8.
5. Sambrook J, Russel DW. Molecular cloning: a laboratory manual. 3rd ed. New York: Cold Spring Harbor Laboratory Press; 2001.
6. Ruitberg CM, Reeder DJ, Butler JM. STRBase: a short tandem repeat DNA database for the human identity testing community. Nucleic Acids Res. 2001;29:320–2.

**How to cite this article:** Wang Y, Gao A, Dong Z, Wang D. Analysis of five rare alleles at the STR loci D1S1656, D12S391, D13S317, Penta D, and D2S441. Electrophoresis. 2023;1–7. https://doi.org/10.1002/elps.202200216