



Technical note: developmental validation of a novel 41-plex Y-STR system for the direct amplification of reference samples

Guangyao Fan¹ · Lipeng Pan² · Peizhi Tang³ · Yijun Zhou⁴ · Mengnan Liu⁴ · Xiangmin Luo³

Received: 12 March 2020 / Accepted: 28 May 2020 / Published online: 10 June 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

The SureID[®] PathFinder Plus is a new 6-dye, 41-plex Y-STR system that includes the 17 loci from the Yfiler[®] kit (DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, and Y-GATA-H4) plus 14 rapidly mutating Y-STR loci (DYS449, DYS481, DYS518, DYS527a/b, DYS533, DYS549, DYS570, DYS576, DYS627, DYF387S1a/b, and DYF404S1), and 10 low-medium mutation loci (DYS388, DYS444, DYS447, DYS460, DYS522, DYS557, DYS593, DYS596, DYS643, and DYS645). The inclusion of the 14 rapidly mutating Y-STR loci improves the discrimination of related individuals. Conversely, the 10 low-medium mutation loci are suitable not only for familial searching but also for providing a higher refinement in the construction of Y chromosome phylogenetic relationships among lineages. The 41-plex Y-STR system is designed for direct amplification of reference samples, such as blood samples on an FTA[®] Card, gauze, tissue, or cotton substrates as well as hair root or buccal samples on swabs. We performed developmental validation work including accuracy, stability, stutter precision, species specificity, sensitivity, PCR inhibitors, reproducibility, parallel testing of the system, and suitability for use on DNA mixtures. In addition, mutations of the loci were analyzed by 754 DNA-confirmed father–son pairs. The results demonstrate that this kit, developed in-house, is time-efficient, accurate, reliable, and highly informative for forensic database, familial searching, and distinguishing related males.

Keywords 41-plex Y-STR system · Y-STR · Developmental validation · Forensic genetics

Introduction

Y-chromosomal microsatellites (Y-STRs) have proven to be an excellent tool in sexual assault cases, as the male component of DNA mixtures can be selectively detected without interference of female victims' DNA [1]. The resulting Y-STR haplotypes also aid in the differentiation of paternal

lineages as well as in deficiency paternity cases. In situations of missing persons or paternity cases, a paternal male relative's profile can be used to support/exclude relatedness when a father's reference sample is not available [1]. However, due to lack of recombination on the Y chromosome, the discriminatory powers of a single multiplex Y-STR analysis system are insufficient, especially for the differentiation of male lineages in inbred populations [2, 3]. Thus, many studies have attempted to add candidate loci, with high gene diversity, to improve current Y-STR systems [1, 4–6]. Adding rapidly mutating (RM) Y-STR loci [4] will increase the potential to distinguish related males [7]. But it may also lead to unexpected exclusion when determining whether a suspect is contained in a paternal lineage [5, 8–10]. For familial searching, low-medium mutation Y-STR loci are a better choice and have already played a vital role in China [11].

In this study, a novel Y-STR system, SureID[®] PathFinder Plus, was developed by us. The SureID[®] PathFinder Plus Kit allows co-amplification of 41 loci in a 6-dye configuration. It includes all 17 loci from the Yfiler[®] kit [12] (DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392,

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00414-020-02326-9>) contains supplementary material, which is available to authorized users.

✉ Guangyao Fan
fanyoyo1983@163.com

¹ Forensic Center, College of Medicine, Shaoxing University, No. 508 Huancheng West Road, Shaoxing 312000, China

² Shaoxing Municipal Public Security Bureau, Shaoxing 312000, China

³ Ningbo Health Gene Technologies Co. Ltd., Ningbo 315040, China

⁴ College of Medicine, Shaoxing University, Shaoxing 312000, China

DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, and Y-GATA-H4) plus 24 Y-STR loci with reported mutation rates. Among the additional loci, DYS557 [13], DYS593 [14, 15], DYS596 [15, 16], and DYS645 [16] are highly polymorphic; DYS449, DYS518, DYS549, DYS570, DYS576, DYS627, DYS643, DYF387S1a/b, and DYF404S1 are rapidly mutating Y-STR loci [1, 17, 18]; and DYS388, DYS444, DYS447, DYS460, DYS481, DYS522, DYS527a/b, and DYS533 are low-medium mutation loci [11, 19]. Thus, this kit may become a novel application kit both for the discrimination of related males [8, 20] by including loci with high mutation rates and for familial searching due to the use of Y-STR loci with low-medium mutation rates. Moreover, the use of three Y-InDels—rs199815934, rs771783753, and rs759551978—that are least likely to mutate could also contribute to familial searching.

Through optimization of primer sequences and the master mix, the 41-plex Y-STR system can overcome challenges in direct amplification of buccal and blood reference samples on cards, gauzes, or swabs with a fast PCR cycle time of < 100 min. In addition, male DNA can be specifically amplified in the presence of large amounts of female DNA background. To verify the efficiency of the system, in this work, we performed a series of validation studies, including for sensitivity, species specificity, inhibition, precision, and DNA mixture assays as per the guidelines of the Scientific Working Group on DNA Analysis Methods (SWGAM) [21–23].

Materials and methods

DNA samples

Reference samples were collected from randomly selected subjects (834 males and 10 females) for the validation study. In addition, samples from 754 father–son pairs were obtained, and among them 734 pairs were composed of a father and a son, and 10 pairs contained a father and two sons. All donors were recruited from the Han ethnic group in Zhejiang province.

Concordance testing was performed with samples on a variety of substrates. Blood was spotted on FTA® Cards (GE Healthcare, Piscataway, NJ, USA), sterile gauzes, sterile cotton balls, and tissues. The samples on FTA® Cards were created following the manufacturer's instructions. The other samples included hair roots as well as buccal cells that were collected on sterile cotton-tipped swabs. Samples were stored for up to 24 h at room temperature until they were further processed (up to 2 weeks). For further direct amplification, 1.2-mm disks punched from FTA® Cards and 2-mm-long threads of gauze or hair roots were added directly into reactions. About 2 mm² of stains was clipped from the dried surfaces

of the swabs, blood-spotted cotton balls, or blood-spotted tissues for direct amplification. Because the entire swab is not consumed in the reaction, resampling was a feasible option if the profile was not acceptable.

The control DNAs 9948 and 9947A were purchased from Promega Corporation (Madison, WI, USA), while control DNA 007 was supplied with the Yfiler® Plus Kit. The 9948 DNA was used for the sensitivity and inhibition studies, and all three control DNAs were applied in mixture studies. DNA concentrations were validated by measuring the absorbance of samples at 260 nm by using a NanoDrop™ 2000 Spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE, USA). When necessary, DNA input was diluted with TE buffer (10 mM Tris-HCl pH 8.0 and 0.1 mM EDTA).

Species studies were conducted using blood samples from domestic animals (dog, chicken, horse, bovine, rabbit, duck, sheep, fish, and pig) and *Escherichia coli* amplified for 28 cycles.

All human samples were collected from healthy volunteer donors after informed consent was given according to the ethical guidelines of the Helsinki Declaration. This project was approved by the Ethics Committee of the Medical College, Shaoxing University, before the study.

PCR amplification and thermal cycling conditions

The SureID® PathFinder Plus Kit was optimized in a 10-μL total reaction volume consisting of 5.0-μL Master Mix and 2.5-μL Primer Mix. PCR samples were amplified in MicroAmp® Optical 96-well reaction plates (Thermo Fisher Scientific, Waltham, MA, USA) with a GeneAmp PCR system 9700 (Thermo Fisher Scientific) using the following conditions: enzyme activation at 95 °C for 5 min; 28 cycles of 94 °C for 10 s, 60 °C for 1 min, and 70 °C for 30 s; and 60 °C for 15 min, followed by a final hold at 4 °C. Cycle optimization was carried out using control DNA 9948. The optimal annealing temperature of the SureID® PathFinder Plus Kit was determined by conducting the annealing step at 58 °C, 59 °C, 60 °C, 61 °C, and 62 °C using control DNA 9948 or a bloodstain on an FTA® Card.

PCR primer set and PCR master mix components

The primers for each locus were designed with Primer Premier 5.0 (Premier Biosoft International, Palo Alto, CA, USA). The robustness of the primer set was determined by varying each primer pair in the multiplex in increments of 10% up to ± 20% compared with the optimized formulation and comparing performance based on intracolor balance, peak height, and alleles recovered. The five dyes used in the SureID® PathFinder Plus Kit to label amplified sample products were 6-FAM™ (blue), HEX™ (green), TAMRA™ (yellow), ROX™ (red), and

VIGTM (purple) (Fig. 1). The sixth dye, LIZTM (orange), was used to label the SIZE-580 internal size standard. The fragments of SIZE-580 were 75, 87, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 515, 530, 550, and 580 bp. The use of an additional dye channel (purple) enabled the addition of new loci while maintaining the marker range to less than 550 bp. All primers were synthesized by Ningbo Health Gene Technologies Co. Ltd., Ningbo, China.

The components of the SureID[®] PathFinder Plus Master Mix included DNA polymerase, buffer, salts, dNTPs, carrier protein, detergents, and 0.05% sodium azide. Three DNA samples on an FTA[®] Card were tested in triplicate for the standard PCR master mix formulation and six modified master mixes.

PCR product electrophoresis and data analysis

The Applied Biosystems 3500xL Genetic Analyzer (Thermo Fisher Scientific) was used as the default for most studies to separate and detect the PCR products unless mentioned otherwise. The dye set of HGT 6-Dye Matrix Standards (Ningbo Health Gene Technologies Co. Ltd.) was used for spectral calibration. Capillary electrophoresis (CE) was as follows: 1 μ L of amplified product or allelic ladder and 0.5- μ L SIZE-580 were added to 8.5 μ L of deionized Hi-DiTM Formamide (Thermo Fisher Scientific). Samples were denatured at 95 °C for 3 min then chilled on ice prior to electrophoresis. Sample injection was performed in a POP-4 polymer (Thermo Fisher Scientific) with the following run conditions: injection at 1.2 kV for 15 s and electrophoresis at 15 kV for 1500 s. GeneMapper[®] ID-X Software v1.2 (Thermo Fisher Scientific) was used to analyze the results with a 150 relative fluorescence unit (RFU) peak amplitude threshold for allele calls.

Sensitivity

To evaluate the sensitivity of the SureID[®] PathFinder Plus Kit, allele recovery performance was measured for control DNA inputs ranging from 0.125 to 2 ng. Samples were prepared by a serial dilution of male control DNA 9948 (2 ng/ μ L stock) in DNA Suspension Buffer. Total DNA inputs tested were 125, 250, 500, 1000, and 2000 pg. Non-template controls (NTC) were included.

Male specificity

All of our selected loci are located on the non-recombination region of the Y chromosome. Moreover, the Master Mix and PCR thermal cycling conditions of the SureID[®] PathFinder Plus Kit were optimized to achieve male specificity. Female control DNA 9947A and 10 female blood samples were utilized to validate the male specificity of the kit. In a reaction volume of 10 μ L for the mixture samples, a total of 1 ng of male DNA input in the presence of total 1 μ g female DNA input was maintained. The potential amplified artifacts from the females were analyzed.

DNA mixtures

Male–male mixtures of control DNAs 9948 and 007 in known ratios of 1:1, 1:2, 1:4, 1:8, and 1:15 were amplified, respectively. Each mixture was tested in triplicate to reduce the accidental error and ensure the accuracy of the results.

Species specificity

Genomic DNA from blood samples of nine male domestic animals were extracted and purified by organic extraction. *E. coli* DNA was isolated using the Ezup Column Bacteria

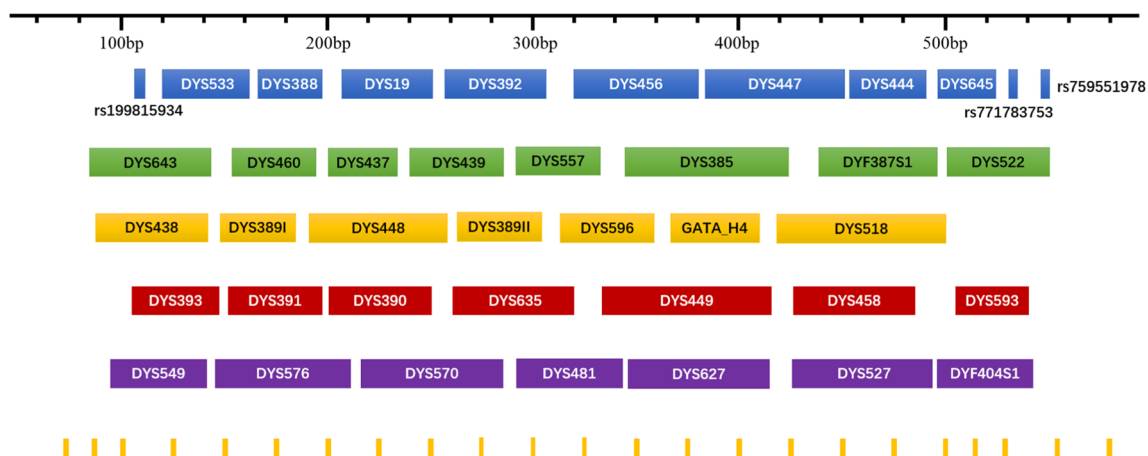


Fig. 1 Diagram illustrating PCR product size ranges of the 41 Y-STR loci in the SureID[®] PathFinder Plus PCR Amplification Kit. The five dye panels were 6-FAMTM (blue), HEXTM (green), TAMRATM (yellow), ROXTM (red), and VIGTM (purple). The sixth panel was for internal size

standard labels with an orange dye: SIZE-580 (75, 87, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 515, 530, 550, and 580 bp)

Genomic DNA Purification Kit (Sangon Biotech, Co., Ltd., Shanghai, China). As per the standard PCR protocol, 10 ng each from the 10 species was used to assess the species specificity of the kit. Non-template controls (NTC) were included. The corresponding products were analyzed in triplicate.

PCR inhibition models

Three common forensic inhibitors including humic acid (Sigma-Aldrich, Darmstadt, Germany), hematin (Sigma-Aldrich), and indigo (Sigma-Aldrich) were chosen to assess the performance of the SureID® PathFinder Plus Kit. Stock solutions of high concentration were prepared by dissolving each of the inhibitors in 0.1 N NaOH (hematin and indigo) or in DNA Suspension Buffer (humic acid). They were further diluted in water to achieve working stocks. One nanogram of control DNA 9948 along with each inhibitor in various concentrations was amplified with the SureID® PathFinder Plus Kit using the standard protocol.

Population and forensic statistical analysis

The genotypes of 834 unrelated male subjects from the Zhejiang Han population were determined using the SureID® PathFinder Plus Kit under standard conditions. The data have been submitted to Y chromosome STR haplotype reference database (<http://www.yhrd.org>) under accession number YA004678. Allele frequencies were determined using the direct counting method. The discrimination capacity (DC) was ascertained by dividing the number of different haplotypes by the total number of individuals in the population [19]. The match probability (MP) was calculated as the sum of the squared haplotype frequencies. Genetic diversities (GD) or haplotype diversities (HD) were calculated as HD or $GD = n(1 - \sum Pi^2)/(n - 1)$ according to Nei [24], where n stands for the total number of samples and Pi is the frequency of the i th haplotype.

In addition, the total number of meioses ($n = 754$) that were previously confirmed by autosomal STR was used to detect the mutation rate. These father–son pair samples, in the form of hair roots, buccal swabs, or blood samples cards, were obtained from the Zhejiang Han population. Mutation rates were calculated as the number of mutations observed divided by the total number of meioses. The 95% confidence intervals (CI) of the mutation rates were estimated from a frequentist approach assuming a binomial distribution using the website <http://statpages.org/confint.html>.

Performance testing

For performance testing of reference samples on a variety of substrates, blood spotted on FTA® Cards ($n = 53$), sterile gauze ($n = 76$), sterile cotton balls ($n = 3$), and tissues ($n =$

6) was amplified with the SureID® PathFinder Plus Kit as per the standard conditions. Buccal samples on swabs ($n = 5$) and hair roots ($n = 4$) were also amplified with the kit using standard conditions.

Sizing accuracy, precision, and stutter effects

Sizing accuracy was defined as the deviation in size of each sample allele from the corresponding allelic ladder allele. Bloodstain samples on 54 FTA® Cards and 76 gauze samples were used to measure sizing accuracy. Punches (1.2-mm disks) of blood on FTA® Cards or threads of blood on gauze (2 mm long) were amplified with the SureID® PathFinder Plus Kit as per the standard PCR conditions of 28 cycles on an Applied Biosystems 3500xL Genetic Analyzer. Allelic ladder sizing precision was measured by calculating the average base pair sizes and standard deviation for each allele from multiple injections of the SureID® PathFinder Plus allelic ladder on the same CE instrument.

Stutters of the SureID® PathFinder Plus Kit were measured by dividing the stutter peak height ($n \pm 1$ repeat units, or $n \pm 2$ nt) by that of the true allele peak height (n repeats) and are expressed as a percentage. The same batch of reference samples (54 blood samples on FTA® Cards and 76 blood samples collected with gauze) was run on an Applied Biosystems 3500xL Genetic Analyzer using a 150 RFU peak height minimum (minimum stutter peak height of 20 RFU) to calculate the stutter ratios. All saturating allele peaks were removed from the stutter calculation.

Statistical analysis

Intracolor balance was calculated as the lowest peak height in each dye channel divided by the highest peak height in the same dye channel and is reported as a percentage. For the four multi-copy loci (DYS385, DYF387S1a/b, DYS527, and DYF404S1), the peak heights were averaged prior to calculating the intracolor balance, and if the alleles were of the same length, the peak heights were halved prior to intracolor balance calculation. Statistical analyses were performed using Minitab® (Minitab Inc., State College, PA, USA) software.

Results and discussion

PCR reaction-cycle number

The SureID® PathFinder Plus Kit was tested over a range of amplifying cycle numbers. One nanogram of control DNA 9948 was amplified over 26, 27, 28, and 29 cycles in the 10- μ L reaction system. As a group, an increase in the cycle number generated a higher peak height as demonstrated in Supplementary Fig. S1. Full profiles were observed over 28

and 29 cycles. The results indicated that a PCR cycle number of 28 was optimal with respect to maximizing assay sensitivity while minimizing the possible occurrence of off-scale peaks or allelic drop-out events. A representative profile generated under the standard condition of 28 amplification cycles is shown in Supplementary Fig. S2 for control DNA 9948.

As mentioned in a previous study [1], many factors such as the collection method, sample type (blood vs. buccal), collection device, and even donors' white blood cell (WBC) counts or buccal-shedding levels can affect the amount of DNA available for direct amplification. Thus, for a better amplification effect, a laboratory may need to adjust the cycle number based on their samples and CE instruments. Representative profiles generated with in-house sample sets for six different substrates under optimal conditions of 28 amplification cycles are shown in Supplementary Fig. S3.

PCR reaction—denaturing, annealing, and final extension steps

Diverse key parameters in the thermal cycling conditions (denaturing temperature, annealing/extension temperature, and final extension time) can affect the robustness of a genotyping system. One nanogram of male control DNA 9948 and the FTA® Cards were used to search for the optimal thermal cycling conditions for the SureID® PathFinder Plus Kit. There was no significant alteration upon changing the denaturing temperature by ± 1.0 °C from the optimal 95 °C (data not shown). Temperatures of 58 °C, 59 °C, 60 °C (optimal condition), 61 °C, and 62 °C were tested for the annealing/extension step, and full profiles were obtained at each of the conditions. For male control DNA 9948, average peak heights at 58 °C, 59 °C, and 61 °C were relatively higher than those at 61 °C and 62 °C (Fig. S4). For 1 ng of 9948, it appeared to have a relatively intense distribution dispersal in intralocus regions and had better performance in intralocus balance at 60 °C, although elevations of the average peak heights over all dye channels were relatively low compared with the others. For blood samples on FTA® Cards, better intralocus and intralocus balances were observed at temperatures of 61 °C, and significant elevations of the average peak heights over all dye channels were observed at temperatures of 60 °C (Fig. S5). Thus, a denaturing temperature of 60 °C was considered to be the right balance between the two kinds of samples. Furthermore, a range of final extension times (5–20 min) was tested in order to ensure nucleotide addition on all amplified products. The results show that shortening the extension time (less than 8 min) will lead to incomplete adenosine addition, especially for loci DYS19, DYS390, and DYS448 (data not shown). The optimal final extension time was determined to be 15 min.

Sensitivity studies

The sensitivity of the SureID® PathFinder Plus Kit was tested with a serial dilution of DNA 9948 from 125 pg to 2 ng in triplicate. Full Y-STR profiles were obtained for DNA inputs greater than or equal to 250 pg (Fig. S6) with the analysis threshold set at 150 RFU peak height. When the 9948 DNA inputs decreased to 125 pg, allele drop-outs were observed. On this basis, we concluded that the SureID® PathFinder Plus Kit can provide reliable profiles at a threshold of 150 RFU with DNA inputs equal to or greater than 250 pg. The sensitivity under these conditions was determined to be 0.25 ng/10 μ L.

Male specificity

The SureID® PathFinder Plus Kit is designed to test only male samples. Using amplified pure female DNA as an input, even with 1 μ g of control DNA 9947A there were no meaningful profiles produced. However, biology samples from forensic casework often contain high levels of female DNA background such as in sexual assault mixtures. In such situations, female cross-reactive amplification products could be observed. To guarantee an accurate profile interpretation, the SureID® PathFinder Plus Kit was optimized to minimize the artifact peak height to be well below the 50 RFU peak amplitude threshold in the presence of female DNA inputs. One nanogram of control DNA 9948 was mixed with 1 μ g 9947A or 1 μ g DNA isolated from each of the 10 female blood samples for amplification. The results showed that none of the possible amplified artifacts from the above female samples would affect profile interpretation in the read region (82–540 bp). A representative profile generated from a male:female mixture in the presence of 1 ng 9948 and 1 μ g 9947A is shown in Supplementary Fig. S7.

DNA mixture studies

Male:male DNA mixtures are often encountered in forensic casework. In order to determine the performance of the SureID® PathFinder Plus Kit when dealing with such samples, the minor and major male contributors were distinguished. Results showed that as the mixture ratio decreased, the peak height of minor alleles decreased. Full alleles of the minor contributor at non-overlapping and non-stutter positions were obtained from a 1:1 to a 1:4 ratio (125-pg minor contributor DNA input). A representative profile generated at a ratio of 1:4 is shown in Supplementary Fig. S8. General losses of alleles were observed for ratios of 1:8 and 1:15, and the unique minor profile was called for an average of 80% and 50%, respectively. Thus, the SureID® PathFinder Plus Kit requires all alleles to be correctly detected for 1:4 male:male DNA mixtures. However, direct amplification of mixtures is not

Table 1 Haplotype diversity, match probability, discrimination capacity values as well as number of unique and shared haplotypes observed of the SureID[®] PathFinder Plus were calculated and compared with three other Y-STR Kits

| | PathFinder Plus | Yfiler Plus | Yfiler | PP23 |
|-------------------------|-----------------|-------------|---------|---------|
| Haplotype Diversity | 0.99996 | 0.99995 | 0.99980 | 0.99995 |
| Match Probability | 0.00124 | 0.00125 | 0.00139 | 0.00125 |
| Discrimination Capacity | 98.32% | 98.08% | 94.00% | 98.08% |
| 1 | 807 | 803 | 746 | 803 |
| 2 | 12 | 14 | 31 | 14 |
| 3 | 1 | 1 | 3 | 1 |
| 4 | 0 | 0 | 3 | 0 |
| 5 | 0 | 0 | 1 | 0 |

PathFinder Plus, SureID[®] PathFinder Plus Amplification Kit (41 Y-STR loci); *Yfiler*, AmpF/STR[®] Yfiler[®] PCR Amplification Kit (17 Y-STR loci); *Yfiler Plus*, Yfiler[®] Plus PCR Amplification Kit (27 Y-STR loci); *PP23*, PowerPlex[®] Y23 System (23 Y-STR loci)

recommended when the original DNA ratios of contributors in the starting material are unknown.

Species specificity

Non-human genomic DNA samples from common animal species (dog, chicken, horse, bovine, rabbit, duck, sheep, fish, pig, and *E. coli*) were amplified using the SureID[®] PathFinder Plus Kit. Results show that no cross-reactive peaks were observed within the genotyping range above the peak amplitude threshold of 150 RFU (Fig. S9). For fish, an “OL” peak (peak height 140 RFU) was observed for locus DYS635, which may disturb genotyping analysis. However, on this basis, the novel Y-STR system is unlikely to be affected by the presence of genetic material from these animal species.

Models of inhibition

The wide range of forensic sample types and variety of sampling conditions encountered make forensic analysis particularly difficult. For as long as forensic scientists

have used the PCR, PCR inhibitors that often lead to locus dropout have been an obstacle to success [25]. PCR inhibitors exert their effects mainly through direct interaction with DNA or interference with thermostable DNA polymerases [26]. Common sample types that are known to contain inhibitors include blood, tissues, fabrics, and soil. Other pivotal sources of inhibitors are the materials and chemistries that come into contact with samples during the upstream extraction.

The SureID[®] PathFinder Plus Kit is designed for the direct amplification of pristine reference samples such as blood and buccal samples. Therefore, DNA purification, which is the method used most often to remove inhibitors, is not used. However, actions have been taken to cope with samples from a variety of substrates and inhibitors such as increasing the amount of DNA polymerase in the reaction and using BSA that has been rigorously quality-tested. Mock inhibition samples with 1 ng of male control DNA 9948 containing varying concentrations of humic acid [27, 28], hematin [28], and indigo [29] were prepared to evaluate the inhibitor tolerance of the SureID[®]

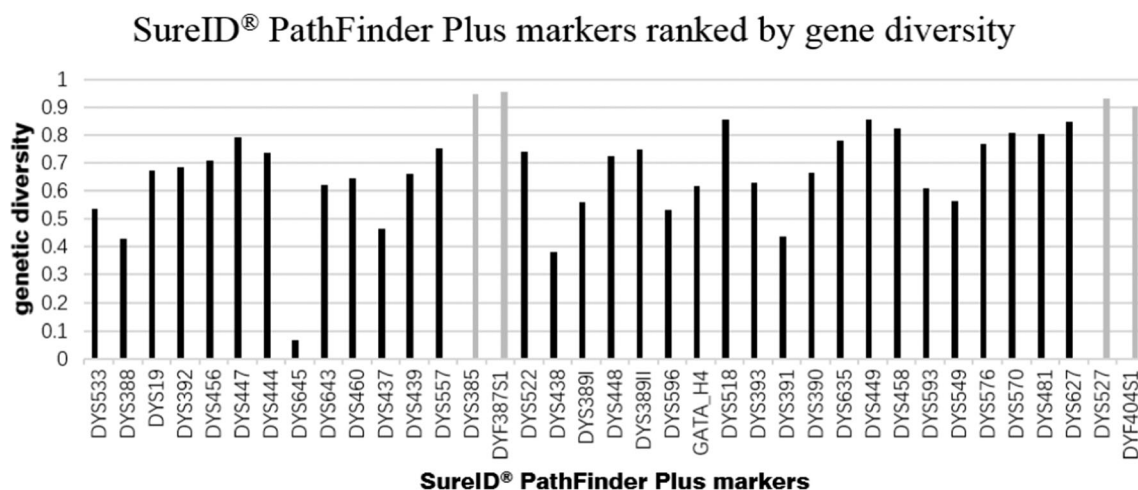


Fig. 2 Y-STR markers of the SureID[®] PathFinder Plus Kit ranked by genetic diversity. Gray bars indicate the multi-copy loci

Table 2 The mutation rates of the 41 Y-STR loci in the Han nationality of Zhejiang province

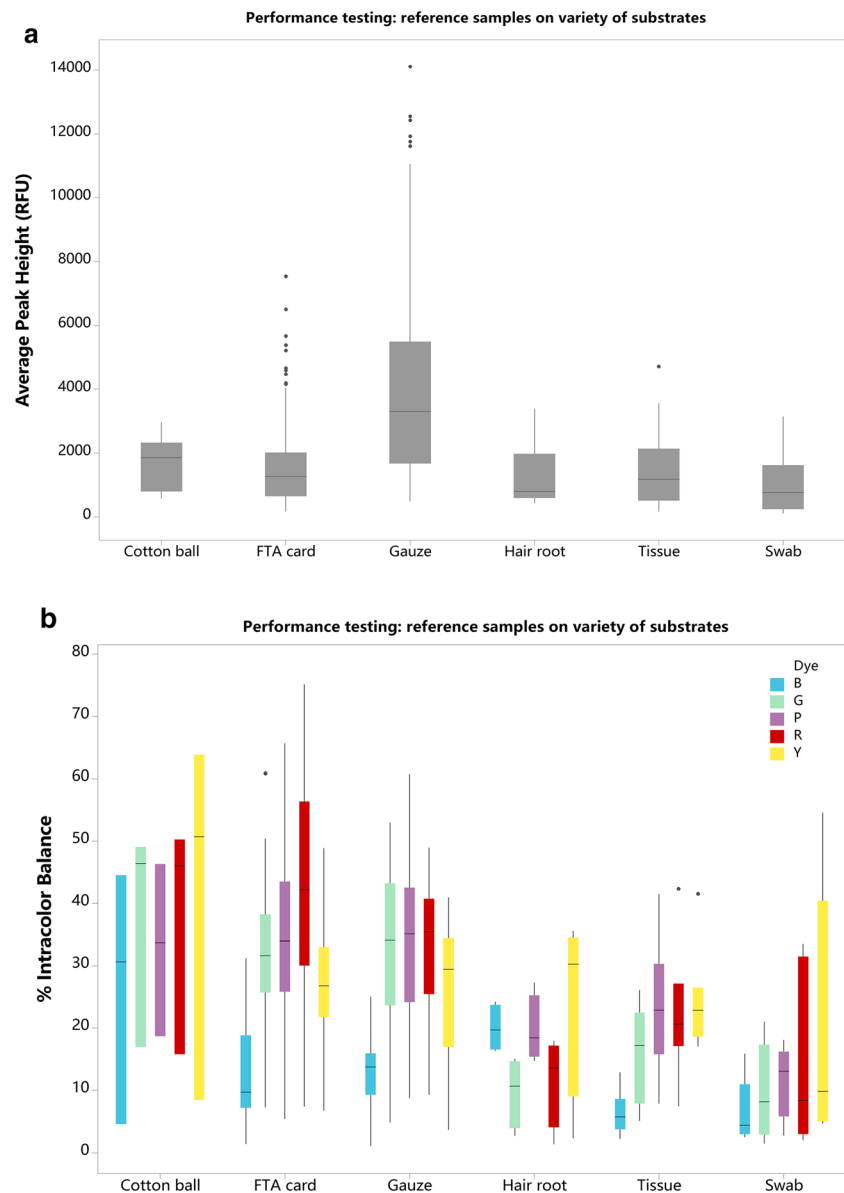
| Locus | Meiotic transfers | $n - 2$ | $n - 1$ | $n + 1$ | $n + 2$ | No. of mutations | Mutation rate ($\times 10^{-3}$) | CI ($\times 10^{-3}$) |
|-------------|-------------------|---------|---------|---------|---------|------------------|------------------------------------|-------------------------|
| DYS533 | 754 | 0 | 2 | 3 | 0 | 5 | 6.63 | 2.2–15.4 |
| DYS388 | 752 | 0 | 0 | 0 | 0 | 0 | 0 | 0–4.9 |
| DYS19 | 752 | 0 | 0 | 0 | 0 | 0 | 0 | 0–4.9 |
| DYS392 | 750 | 0 | 1 | 0 | 0 | 1 | 1.33 | 0–7.4 |
| DYS456 | 743 | 0 | 2 | 0 | 0 | 2 | 2.69 | 0.3–9.7 |
| DYS447 | 712 | 0 | 1 | 0 | 0 | 1 | 1.40 | 0–7.8 |
| DYS444 | 733 | 0 | 2 | 0 | 0 | 2 | 2.73 | 0.3–9.8 |
| DYS645 | 712 | 0 | 0 | 0 | 0 | 0 | 0 | 0–5.2 |
| DYS643 | 752 | 0 | 0 | 0 | 0 | 0 | 0 | 0–4.9 |
| DYS460 | 752 | 0 | 1 | 2 | 0 | 3 | 3.99 | 0.8–11.6 |
| DYS437 | 748 | 0 | 0 | 0 | 0 | 0 | 0 | 0–4.9 |
| DYS439 | 751 | 0 | 1 | 0 | 0 | 1 | 1.33 | 0–7.4 |
| DYS557 | 748 | 0 | 0 | 0 | 0 | 0 | 0 | 0–4.9 |
| DYS385a/b | 738 | 0 | 4 | 3 | 0 | 7 | 9.49 | 3.8–19.4 |
| DYF387S1a/b | 716 | 0 | 3 | 5 | 0 | 8 | 11.17 | 4.8–21.9 |
| DYS522 | 702 | 0 | 1 | 0 | 0 | 1 | 1.42 | 0–7.9 |
| DYS438 | 752 | 0 | 0 | 0 | 0 | 0 | 0 | 0–4.9 |
| DYS389I | 752 | 0 | 0 | 0 | 0 | 0 | 0 | 0–4.9 |
| DYS448 | 745 | 0 | 1 | 0 | 0 | 1 | 1.34 | 0–7.5 |
| DYS389II | 743 | 0 | 2 | 3 | 0 | 5 | 6.73 | 2.2–15.6 |
| DYS596 | 749 | 0 | 0 | 0 | 0 | 0 | 0 | 0–4.9 |
| GATA-H4 | 750 | 0 | 0 | 0 | 0 | 0 | 0 | 0–4.9 |
| DYS518 | 741 | 0 | 1 | 3 | 0 | 4 | 5.40 | 1.5–13.8 |
| DYS393 | 753 | 0 | 0 | 0 | 0 | 0 | 0 | 0–4.9 |
| DYS391 | 753 | 0 | 1 | 1 | 0 | 2 | 2.66 | 0.3–9.6 |
| DYS390 | 753 | 0 | 0 | 0 | 0 | 0 | 0 | 0–4.9 |
| DYS635 | 752 | 0 | 1 | 4 | 0 | 5 | 6.65 | 2.2–15.4 |
| DYS449 | 746 | 0 | 5 | 4 | 0 | 9 | 12.06 | 5.5–22.8 |
| DYS458 | 725 | 0 | 5 | 4 | 0 | 9 | 12.41 | 5.7–23.4 |
| DYS593 | 722 | 0 | 0 | 0 | 0 | 0 | 0 | 0–5.1 |
| DYS549 | 749 | 0 | 2 | 2 | 0 | 4 | 5.34 | 1.5–13.6 |
| DYS576 | 748 | 0 | 2 | 3 | 0 | 5 | 6.68 | 2.2–15.5 |
| DYS570 | 750 | 1 | 3 | 3 | 0 | 7 | 9.33 | 3.8–19.1 |
| DYS481 | 747 | 1 | 1 | 2 | 0 | 4 | 5.35 | 1.5–13.7 |
| DYS627 | 743 | 0 | 6 | 6 | 0 | 12 | 16.15 | 8.4–28.0 |
| DYS527a/b | 739 | 0 | 5 | 4 | 0 | 9 | 12.18 | 5.6–23.0 |
| DYF404S1 | 722 | 0 | 7 | 5 | 0 | 12 | 16.62 | 8.6–28.9 |

PathFinder Plus Kit. Full profiles were obtained with humic acid concentrations up to 200 ng/ μ L, hematin concentrations up to 40 μ M, and indigo concentrations up to 10 mM (Fig. S10). Allelic dropouts were observed in loci with long fragment size alleles when the concentrations exceeded these levels. Meanwhile, smaller loci in the kit were preferentially amplified. It should be noticed that inhibited samples may mistakenly be assumed to be degraded, as the same pattern is typical of highly degraded DNA templates.

Population studies and mutation rates

The population data were used to calculate GD, HD, DC, and MP values. The number of unique haplotypes observed was higher with the SureID[®] PathFinder Plus Kit compared with three former Y-STR kits (Table 1), and our novel kit showed improved haplotype diversity and discrimination capacity. Using the SureID[®] PathFinder Plus Amplification Kit, 807 (98.32%) unique Y-STR haplotypes (Table 1) were generated from the Zhejiang Han population. The increased number of

Fig. 3 Performance comparison of reference samples on a variety of substrates. Blood spotted on FTA® Cards ($n = 53$), sterile gauze ($n = 76$), sterile cotton balls ($n = 3$), and tissues ($n = 6$) was amplified with the SureID® PathFinder Plus Kit. Buccal samples on swabs ($n = 5$) and hair roots ($n = 4$) were also amplified with the kit using the same standard conditions. Electrophoresis was conducted on the same 3500xL Genetic Analyzer for all samples. Each box represents the 25th through the 75th percentile of the observed values, with the center line indicating the median. The range of the data was indicated by whiskers. The black dots represent outliers that are different from the mean by more than twice the pooled standard deviation



analyzed loci improved the DC and HD. Compared with Yfiler®, the SureID® PathFinder Plus Kit provided higher power forensic discrimination, which raised the DC from 94.00 to 98.32%. Results showed GD values above 0.4 for all markers of the SureID® PathFinder Plus Kit, except for DYS645 and DYS438 (Fig. 2). The multi-copy loci DYS385a/b, DYF387S1a/b, and DYS527a/b exceeded a GD value of 0.9.

A total of 754 father–son pairs were analyzed at 41 Y-STR loci as well as three Y-InDel markers. At all of these loci, 119 mutations were observed from 113 father–son pairs, while no mutations were found in rs199815934, DYS388, DYS19, DYS645, rs771783753, rs759551978, DYS643, DYS437, DYS557, DYS438, DYS389I, DYS596, Y-GATA-H4, DYS393, DYS390, or DYS593. Overall, there was no appreciable difference between the numbers of repeat gains (57)

and losses (62). The DYS458 locus had the highest mutation rate ($> 1 \times 10^{-2}$) among the single-allele loci, while DYF404S1 had the highest mutation rate among the multi-copy loci (Table 2). Additionally, among the 119 observed mutations, 117 were single-step mutations, and only two followed two-step mutations (father–son pairs no. 1 and no. 2), which are shown in Supplementary Fig. S11. Five of the father–son pairs had mutations at two loci, in which DYS456-DYS449, DYS533-DYS391, DYS389II-DYS570, DYS458-DYS527, and DYS527-DYF404S1 each appeared once. One of the mutations occurred during the meiotic transfer from the father to one of his two sons (father–son pair no. 3), which is also shown in Supplementary Fig. S11. In addition, a mutation was only observed at one locus in each of the meiotic transfers, except for a special father–son pair. In this case, for father–son pair no. 4 shown in Supplementary Fig. S11, both

highest minus ($n - 1$ repeat unit) stutter filter value. The differences in stutter occurrences between blood spotted on FTA® Cards and sterile gauze were discernible. Stutter effects of blood spotted on FTA® Cards were generally better than with gauze.

Conclusion

The SureID® PathFinder Plus Kit is compatible with the Yfiler® Kit, Yfiler® Plus Kit, and PowerPlex® Y23 System. Co-amplification of 41 loci and 3 Y-InDel markers enhances the multiplexing capability. Therefore, this novel kit has the power to distinguish related males and familial searching. In order to accelerate and simplify the process, optimization of primer sequences and master mix enables the SureID® PathFinder Plus Kit to deal with direct amplification challenges within 100 min for hair roots, blood, or buccal reference samples on different substrates. We performed a series of developmental validations of the SureID® PathFinder Plus Kit and found that it has stable allele sizing accuracy and precision, good intracolor balance, good repeatability, male specificity, and high sensitivity. Three external labs confirmed and reproduced the results of the SureID® PathFinder Plus Kit in beta-testing. All of the above results demonstrate the robustness and validity of the SureID® PathFinder Plus Kit for use with forensic reference samples. Quality Y-STR genotypes can be obtained with significantly less labor. This novel kit will be a dependable and convenient tool for distinguishing related males, for familial searching, and in DNA database construction.

Acknowledgments We thank all of the participants in this study. The authors would like to gratefully acknowledge the following laboratories for beta-testing of the SureID® PathFinder Plus Kit: Medical Research Center of Shaoxing University, Forensic Center of Shaoxing University, Human DNA Diagnostics Laboratory of Ningbo Health Gene Technologies Co. Ltd., and Forensic DNA laboratory of Shaoxing Municipal Public Security Bureau.

Compliance with ethical standards

Informed consent was obtained from all participating subjects, and this work was approved by the Ethics Committee of the Medical College, Shaoxing University.

Conflicts of interest The authors declare that they have no conflict of interest.

References

- Gopinath S, Zhong C, Nguyen V, Ge J, Lagace RE, Short ML, Mulero JJ (2016) Developmental validation of the Yfiler® Plus PCR Amplification Kit: an enhanced Y-STR multiplex for casework and database applications. *Forensic Sci Int Genet* 24:164–175
- Hedman M, Pimenoff V, Lukka M, Sistonen P, Sajantila A (2004) Analysis of 16 Y STR loci in the Finnish population reveals a local reduction in the diversity of male lineages. *Forensic Sci Int* 142:37–43
- Hedman M, Neuvonen AM, Sajantila A, Palo JU (2010) Dissecting the Finnish male uniformity: the value of additional Y-STR loci. *Forensic Sci Int Genet* 5:199–201
- Kayser M, Vermeulen M, Knoblauch H, Schuster H, Krawczak M, Roewer L (2007) Relating two deep-rooted pedigrees from Central Germany by high-resolution Y-STR haplotyping. *Forensic Sci Int Genet* 1:125–128
- Lin H, Ye Q, Tang P, Mo T, Yu X, Tang J (2020) Analyzing genetic polymorphism and mutation of 44 Y-STRs in a Chinese Han population of Southern China. *Legal Med* 42:101643
- Thompson JM, Ewing MM, Frank WE, Pogemiller JJ, Nolde CA, Koehler DJ, Shaffer AM, Rabbach DR, Fulmer PM, Sprecher CJ, Storts DR (2012) Developmental validation of the PowerPlex® Y23 System: a single multiplex Y-STR analysis system for casework and database samples. *Forensic Sci Int Genet* 7:240–250
- Westen AA, Kraaijenbrink T, Clarisse L, Grol LJW, Willemse P, Zuniga SB, Robles De Medina EA, Schouten R, van der Gaag KJ, Weiler NEC, Kal AJ, Kayser M, Sijen T, de Knijff P (2014) Analysis of 36 Y-STR marker units including a concordance study among 2085 Dutch males. *Forensic Sci Int Genet* 14:174–181
- 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* 6:208–218
- Butler JM, Schoske R (2005) U.S. population data for the multi-copy Y-STR locus DYS464. *J Forensic Sci* 50:975
- Robino C, Ralf A, Pasino S, De Marchi MR, Ballantyne KN, Barbaro A, Bini C, Carnevali E, Casarino L, Di Gaetano C, Fabbri M, Ferri G, Giardina E, Gonzalez A, Matullo G, Nutini AL, Onofri V, Piccinini A, Piglionica M, Ponzano E, Previderè C, Resta N, Scarnicci F, Seidita G, Sorçaburu-Cigliero S et al (2014) Development of an Italian RM Y-STR haplotype database: results of the 2013 GEFI collaborative exercise. *Forensic Sci Int Genet* 15:56–63
- Mo XT, Zhang J, Ma WH, Bai X, Li WS, Zhao XC, Ye J (2019) Developmental validation of the DNATyper™ Y26 PCR amplification kit: an enhanced Y-STR multiplex for familial searching. *Forensic Sci Int Genet* 38:113–120
- Mulero JJ, Chang CW, Calandro LM, Green RL, Li Y, Johnson CL, Hennessy LK (2006) Development and validation of the AmpF/STR® Yfiler™ PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system. *J Forensic Sci* 51:64–75
- Shi MS, Tang JP, Bai RF, Yu XJ, Lv JY, Hu B (2007) Haplotypes of 20 Y-chromosomal STRs in a population sample from southeast China (Chaoshan area). *Int J Legal Med* 121:455–462
- Dai HL, Wang XD, Li YB, Wu J, Zhang J, Zhang HJ, Dong JG, Hou YP (2004) Characterization and haplotype analysis of 10 novel Y-STR loci in Chinese Han population. *Forensic Sci Int* 145:47–55
- Hanson EK, Ballantyne J (2007) Population data for 48 'Non-Core' Y chromosome STR loci. *Legal Med* 9:221–231
- Jarve M, Zhivotovsky LA, Rootsi S, Help H, Rogaev EI, Khusnutdinova EK, Kivisild T, Sanchez JJ (2009) Decreased rate of evolution in Y chromosome STR loci of increased size of the repeat unit. *PLoS One* 4:e7276
- Du W, Feng P, Huang H, Wu W, Zhang L, Guo Y, Liu C, Liu H, Liu C, Chen L (2019) Technical note: developmental validation of a novel 6-dye typing system with 36 Y-STR loci. *Int J Legal Med* 133:1015–1027

18. Alghafri R, Goodwin W, Ralf A, Kayser M, Hadi S (2015) A novel multiplex assay for simultaneously analysing 13 rapidly mutating Y-STRs. *Forensic Sci Int Genet* 17:91–98
19. Baeta M, Nunez C, Villaescusa P, Ortueta U, Ibarbia N, Herrera RJ, Blazquez-Caeiro JL, Builes JJ, Jimenez-Moreno S, Martinez-Jarreta B, de Pancorbo MM (2018) Assessment of a subset of Slowly Mutating Y-STRs for forensic and evolutionary studies. *Forensic Sci Int Genet* 34:e7–e12
20. Ballantyne K, Ralf A, Aboukhalid R, Achakzai N, Anjos T, Ayub Q, Balažić J, Ballantyne J, Ballard D, Berger B, Bobillo C, Bouabdellah M, Burri H, Capal T, Caratti S, Cárdenas J, Cartault F, Carvalho E, Carvalho M, Cheng B, Coble M, Comas D, Corach D, D'Amato M, Davison S et al (2014) Toward male individualization with rapidly mutating Y-chromosomal short tandem repeats. *Hum Mutat* 35:1021–1032
21. SWGDAM (2004) Report on the current activities of the Scientific Working Group on DNA Analysis Methods Y-STR Subcommittee. Available at http://www2.fbi.gov/hq/lab/fsc/backissu/july2004/standards/2004_03_standards03.htm. Accessed 10.12.12
22. SWGDAM (2012) Validation Guidelines for DNA Analysis Methods. Available at http://swgdam.org/SWGDAM_Validation_Guidelines_APPROVED_Dec_2012.pdf. Accessed 04.05.15
23. SWGDAM (2014) Interpretation Guidelines for Y-Chromosome STR Typing. Available at http://swgdam.org/SWGDAM_YSTR_Guidelines_APPROVED_01092014_v_02112014_FINAL.pdf. Accessed 04.05.15
24. Nei M, Tajima F (1981) DNA polymorphism detectable by restriction endonucleases. *Genetics* 97:145–163
25. Hedman J, Nordgaard A, Rasmusson B, Ansell R, Radstrom P (2009) Improved forensic DNA analysis through the use of alternative DNA polymerases and statistical modeling of DNA profiles. *Biotechniques* 47:951–958
26. Opel KL, Chung D, McCord BR (2010) A study of PCR inhibition mechanisms using real time PCR. *J Forensic Sci* 55:25–33
27. Tsai YL, Olson BH (1992) Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. *Appl Environ Microbiol* 58:2292–2295
28. Akane A, Matsubara K, Nakamura H, Takahashi S, Kimura K (1994) Identification of the heme compound copurified with deoxyribonucleic acid (DNA) from bloodstains, a major inhibitor of polymerase chain reaction (PCR) amplification. *J Forensic Sci* 39:362
29. Shutler GG, Gagnon P, Verret G, Kalyn H, Korkosh S, Johnston E, Halverson J (1999) Removal of a PCR inhibitor and resolution of DNA STR types in mixed human-canine stains from a five year old case. *J Forensic Sci* 44:623–626
30. Schlotterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids Res* 20:211–215
31. Viguera E, Canceill D, Ehrlich SD (2001) In vitro replication slippage by DNA polymerases from thermophilic organisms. *J Mol Biol* 312:323–333
32. Walsh PS, Fildes NJ, Reynolds R (1996) Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Res* 24:2807–2812

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.