



OPEN Recombinase polymerase amplification of forensic short tandem repeat loci

Sonja Škevin¹, Liesl De Keyzer¹, Lynn De Waele¹, Olivier Tytgat¹, David Van Hoofstat², Dieter Deforce¹ & Filip Van Nieuwerburgh¹✉

Short tandem repeats (STRs) are highly polymorphic repetitive DNA sequences extensively used in forensic science for identification of individuals. STR genotyping is usually performed by capillary electrophoresis (CE) or next-generation sequencing (NGS) in centralized laboratories. However, there is an increasing need for a low cost, portable and rapid STR genotyping method. Multiple methods for miniaturization have been explored, all relying on polymerase chain reaction (PCR) for generating amplicons. PCR requires precise thermal cycling, which complicates the design of the STR genotyping microfluidic device. Recombinase Polymerase Amplification (RPA) is an isothermal DNA amplification method that operates between 37 °C and 42 °C and completes within 40 min. This, along with the robustness of reagents and reduced stutter rate compared to PCR makes RPA a suitable candidate for implementation in an STR genotyping microfluidics device as well as a part of the established STR genotyping work flows. In this proof-of-concept study, we evaluate RPA assay for amplification of forensically relevant STR loci. Thirteen core STR loci of the Combined DNA Index System (CODIS) were amplified using RPA in both singleplex and multiplex formats. The amplicons were then analyzed using three different methods: CE, Illumina and Oxford Nanopore Technologies (ONT). A subset of 5 loci was used for CE analysis. CE, Illumina and ONT sequencing of singleplex RPA each resulted in complete and correct STR profiles across all samples for the standard input of 1 ng. Sensitivity assessment demonstrated that complete and correct genotypes were achieved with DNA inputs of 62 pg and above for all but locus D8S1179. Attempts at multiplex RPA amplification resulted in incomplete or incorrect STR profiles. This outcome highlights a challenge in adapting RPA for simultaneous amplification of multiple STR loci, which is a standard requirement in forensic DNA profiling.

Keywords Recombinase polymerase amplification, Short tandem repeats, Microsatellites, Forensic genetics, Capillary electrophoresis, Massively parallel sequencing, Nanopore sequencing

Short tandem repeats (STRs), also known as microsatellites, consist of deoxyribonucleotide acid (DNA) sequences ranging from 1 to 6 base pairs, repeated multiple times in a head-to-tail manner. These sequences comprise approximately 3% of the human genome¹ and exhibit mutation rates 10,000 to 100,000 times higher than the average across the genome. While initially considered neutral genetic markers, STR expansions have since been linked to numerous neurological and developmental disorders. Beyond their pathological implications, STRs play a role in DNA replication and repair, chromatin organization, and gene expression regulation². Given their high polymorphism rate, STRs are extensively employed as genetic markers in humans, animals, and plants for various applications, including conservation, population genetics, kinship analysis, forensic DNA fingerprinting, and cancer diagnostics^{3,4}.

The gold standard for STR genotyping involves amplicon sizing using capillary electrophoresis (CE). CE analysis is, since recently, complemented by the introduction of massively parallel sequencing (MPS), which offers distinct benefits, being an increased throughput, isoallele typing and single nucleotide polymorphism (SNP) detection^{5,6}.

Even though both CE and MPS are established and reliable STR genotyping methods, there is an increasing demand for portable, rapid and affordable forensic genotyping⁷. Two main approaches to achieving speed and portability are currently pursued: a commercially available, portable CE device and a microfluidic-based STR genotyping approach⁸. Portable devices currently on the market are mostly CE-based, such as ANDE Rapid

¹Laboratory of Pharmaceutical Biotechnology, Ghent University, Ottergemsesteenweg 460, Gent 9000, Belgium. ²Eurofins Forensics Belgium NV, Lieven Bauwensstraat 6, Brugge 8200, Belgium. ✉email: Filip.VanNieuwerburgh@UGent.be

DNA (ANDE Corporations, Longmont, CO, USA) and RapidHIT™ ID System (Applied Biosystems, Thermo Fisher, Waltham, MA, USA). Although rapid and reliable, both devices come with an upfront cost of more than 200 000 USD. The price of the device, along with the high reagent cost, can be too high for many small, resource-scarce communities. This results in a significant backlog leading to delays in justice and closure for the victims⁹.

In addition to the established STR genotyping methods and aforementioned portable solutions, there is a growing body of research on STR genotyping using long read sequencing. Specifically, multiple recent studies have demonstrated the improvement of Oxford Nanopore Technologies (ONT) sequencing accuracy in STR genotyping^{10–12}. One of ONT's products is a cost-effective, portable, universal serial bus (USB) powered MinION device, which has been repeatedly tested for on-site applications^{13,14}. Therefore, MinION is another alternative for portable STR genotyping that does not come with a high upfront cost.

All of the described methods rely on PCR, which requires precise thermal cycling, which poses an additional obstacle for miniaturization. There are several isothermal amplification methods that are being researched with the purpose of circumventing the thermal cycling requirements of PCR, one of which is recombinase polymerase amplification (RPA). RPA is an isothermal DNA amplification method invented in 2006¹⁵ and it has primarily been researched in infectious disease diagnostics as a real-time pathogen detection method. RPA operates optimally in a temperature range between 37 and 42 °C, but can function at room temperature, albeit with a reduced efficacy¹⁶. The RPA reaction is completed in up to 40 min and has a sensitivity on-par to PCR^{17–19}. An additional benefit of the RPA assay is the robustness of the reagents, with no reduction in performance after storage at temperatures up to 45 °C, for multiple weeks^{20,21}. It has also been proven that RPA reduces stutter in cancer-related microsatellites compared to PCR³, a valuable characteristic in interpretation of low-input forensic samples and mixture samples.

The goal of this proof-of concept study is to explore the ability of RPA to amplify forensic STR loci with an envisioned application in either a portable microfluidic STR genotyping device, or as a part of an established forensic genotyping workflow. We used CE and MPS to explore the RPA amplicons of STR loci in order to address several key questions that arise when considering RPA for forensic STR amplification. These include the feasibility of working with fluorescently labeled markers, specificity, sensitivity, as well as speed and multiplexing capability. For this purpose, thirteen core STR loci of the Combined DNA Index System (CODIS) were amplified using RPA in singleplex and multiplexes. A singleplex, sixplex and thirteenplex set-up was analyzed using Illumina and Oxford Nanopore Technologies (ONT). A subset of 5 loci was used for singleplex CE analysis.

Materials and methods

Sample collection and DNA extraction

In total, nine samples were utilized in this study. Of those, four DNA samples were extracted from blood of healthy volunteers, three DNA samples were extracted from buccal swabs of healthy volunteers and two were commercially available forensic controls. Ethical approval was granted by the ethical review board of Ghent University Hospital (approval No. BC-05557). All volunteers provided a signed informed consent prior to participation and all experiments were performed in accordance with relevant guidelines and regulations. The commercial controls were 9947 A and 9948 forensic reference samples purchased from OriGene (Rockville, Maryland, USA). Eight samples were tested in the RPA followed by CE workflow. A subset of four DNA samples was used for the Illumina and Nanopore sequencing workflows; the two commercial controls and samples 3 and 4. Blood was collected via finger puncture using a 21G Minicollect® Lancelino safety lancet, with a penetration depth of 2.4 mm (Greiner Bio-One, Kremsmünster, Austria). The blood was collected into K3E K3EDTA Minicollect® collection tubes (Greiner Bio-One, Kremsmünster, Austria). DNA extraction from the collected blood samples was performed using the DNeasy® Blood and Tissue Kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. DNA was extracted from buccal swabs using chelex resin purchased from Bio-Rad (Bio-Rad Laboratories, CA, USA) by the protocol outlined by Sweet et al.²². Extracted DNA was quantified using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA).

Benchmarking using capillary electrophoresis

The benchmark genotyping of the DNA extracted from the donor samples was conducted using PCR followed by amplicon sizing by capillary electrophoresis (CE). For each sample, 1 ng of DNA was amplified using the AmpFISTR® Identifiler® Plus PCR Amplification Kit (Thermo Fisher Scientific, Waltham, MA, USA), according to the protocol provided by the manufacturer, opting for a 28 PCR cycles option. Resulting amplicons were analyzed using ABI3130xl Genetic Analyzer (Thermo Fisher Scientific, Waltham, MA, USA) equipped with 36 cm capillaries containing POP-4™ matrix (Applied Biosystems) according to the manufacturer's instructions. 1 µl of PCR sample amplicons was mixed with 12 µl of Hi-Di™ formamide (Applied Biosystems) and 0.2 µL GeneScan™ 600 LIZ internal lane standard (Applied Biosystems). A run module with a 4 s injection time at 1.2 kV injection voltage and a separation time of 1550 s at 13 kV was used. The resulting electropherograms were interpreted using GeneMapper ID-x 1.2 software (Thermo Fisher Scientific, Waltham, MA, USA). Relative fluorescent units (RFU) analytical threshold was set to 100. For reference, the genotypes and source material of all analyzed samples are available in Suppl. Tables S2 and S3.

RPA amplification, purification and pooling

Amplification of the samples was performed using the TwistAmp Liquid Basic Kit (TwistDx, Maidenhead, UK). The Reaction Buffer, 10x Basic E-mix, and 20x Core Reaction Mix were combined in the amounts required to achieve a 1X concentration in the final reaction mix. The concentration of dNTPs (Thermo Fisher Scientific, Waltham, MA, USA) in the reaction mix was 1.8 mM. The master mix was then added to the primer mixes, followed by the addition of 1 ng of DNA. The reaction was initiated by adding MgOAc at a reaction mix concentration of 14 mM. Five loci (Amelogenin, D18S51, D21S11, D8S1179, TH01) were selected for CE

analysis, based on their non-overlapping amplicon lengths and coverage of a wide size range, to demonstrate the feasibility of RPA amplification of STR loci using fluorescently labelled primers (Suppl. Table S4). Due to limited performance of the fiveplex with CE, a panel of different loci was selected to be amplified in multiplex and analyzed by sequencing. The non-overlapping lengths of Amelogenin, D13S317, D16S539, D7S820, FGA, vWA allowed for the preliminary optimization by agarose gel screening (Suppl. Figure S1). Along with the sixplex, an expanded panel of thirteen loci was used: Amelogenin, D13S317, D16S539, D18S51, D21S11, D3S1358, D5S818, D7S820, D8S1179, FGA, TH01, TPOX, vWA. These were amplified in singleplex and thirteenplex. Since lengths of the thirteen loci overlap, assessment of the multiplex reaction success was not possible on the agarose gel. Widely used PCR primer sequences acquired from STRbase²³ were adapted in length to match the recommended 30–35 nt for RPA (Suppl. Table S5). Sequences of the fluorescently labeled primers were the same for all loci except for D18S51 where two nucleotides were removed from the 5' end in order to prevent the quenching of the fluorophore by the guanine residue. Due to the sensitivity of the assay to oligonucleotide concentrations in the master mix, the primer quantities varied across the 13 loci, ranging from 62,5 nM to 250 nM for each forward and reverse primer. The final primer concentration used in the singleplex reaction was determined by testing multiple concentrations and examining the product on a precast 2% agarose gel (Thermo Fisher Scientific, Waltham, MA, USA) alongside the Invitrogen E-Gel™ 1 kb Plus DNA Ladder (Thermo Fisher Scientific, Waltham, MA, USA) size marker to aid the fragment sizing. The following concentrations were tested: 31,25 nM, 62,5 nM, 125 nM and 250 nM. For both positive and negative reactions, 3 µl of the amplification product was loaded on the gel. The concentration that gave the sharpest band and least visible smear on the agarose gel was selected for the experiment. The primer concentration ratios for the multiplex reactions were set based on their optimal concentration in the singleplex reaction and were adjusted not to exceed a maximal total oligo concentration of 2000 nM. Extensive optimization of multiplex primer concentrations was not conducted in this proof-of-concept study. Sequences and concentrations of individual primers can be found in Suppl. Table S4. Amplification was carried out for 40 min at 39 °C. Post-amplification, the samples were purified using Zymo Clean and Concentrator-5 kits (Zymo Research, Irvine, CA, USA) according to the manufacturer's instructions. Concentrations of the singleplex products were quantified using the PicoGreen assay (Thermo Fisher Scientific, Waltham, MA, USA), and subsequently pooled equimolarly. Since measuring of the DNA concentration of fluorescently labeled amplicons is not possible using Qubit or Picogreen techniques, pooling for CE analysis was performed by combining equal volumes of singleplex RPA product.

Sensitivity assessment and mixture samples

Sensitivity assessment of singleplex RPA assay was conducted on serial dilutions of sample 2. DNA input amounts were: 500 pg, 250 pg, 125 pg, 62 pg and 31 pg. Amplification, purification and detection by CE was performed in the same manner as for 1 ng input samples.

Illumina sequencing

The unlabeled RPA samples were sequenced on Illumina sequencing MiSeq platform, generally considered the golden standard for amplicon sequencing. The library preparation employed the PCR-free NEBNext® Ultra II for Illumina® kit (NEB, Ipswich, MA, USA), with input DNA ranging from 314 to 685 ng per sample. Following the ligation of TruSeq Unique Dual Index (UDI) adaptors, the libraries were purified using 0.9X AMPure XP beads (Beckman Coulter, High Wycombe, UK). Libraries were quantified using a Sequencing Library qPCR Quantification kit (Illumina, San Diego, CA, USA) and equimolarly pooled. Library quality was assessed using the Fragment Analyzer NGS Kit (Agilent Technologies, Santa Clara, CA, USA). The library pool was spiked with 25% PhiX. Paired-end 300 bp sequencing of the library pool was performed using a MiSeq Nano Flow Cell (Illumina, San Diego, CA, USA).

Nanopore sequencing

Library preparation was conducted using the SQK-NBD114.24 kit, version 14, according to the manufacturer's instructions (Oxford Nanopore Technology (ONT), Oxford, UK). For each library, 200 fmol of the purified RPA product was used, which corresponds to approximately 32 ng of DNA. Aside from reagents available in the ONT kit, additional reagents were used: NEB Next Ultra II End Repair/dA-Tailing Module was used for end preparation, the NEB Blunt/TA Ligase Master Mix for native barcode ligation, and the NEB Next Quick Ligation Module for adapter ligation (all from NEB, Ipswich, MA, USA). Quantification of the prepared libraries was performed using the Qubit dsDNA High Sensitivity Assay Kit on a Qubit fluorometer (Thermo Fisher, Waltham, MA, USA). A total of 12 distinct barcodes were used to create 12 separate libraries. For sequencing, 50 fmol of the pooled library was loaded onto a PromethION R10.4.1 flowcell, and the sequencing was conducted on a GridION device for a total duration of 40 h.

Data analysis

Nanopore sequencing reads were demultiplexed and filtered in real-time using Guppy basecaller (v7.2.13). Reads with a q score lower than 9 and shorter than 20 nt were filtered out. Reads passing the quality filter were later rebasecalled using the state-of-the-art Dorado basecaller (v0.5.1), with a high accuracy option. Paired-end Illumina reads were merged using the FLASH tool (v1.2.11)²⁴. Reads from both sequencing methods were processed using an in-house Python script pipeline as previously described¹². During the pipeline, amplicons are first extracted by matching to the RPA primer sequences using Fuzzy regex allowing two mismatches for nanopore data and zero mismatches for Illumina data. Extracted amplicons were then mapped using bwa mem (v0.7.17)²⁵ to the library containing all known alleles present in > 1% of the population, collected from STRbase 2.0²³. Nanopore sequencing is inherently more noisy than Illumina sequencing. In order to reduce the noise, custom filtering based on the alignment score (AS), previously described by our group¹² was applied to the

mapping data. Since the AS indicates the similarity between a read and its aligned reference, a higher AS typically indicates more accurate alignment to true alleles than to their + 1 and - 1 stutters. Conversely, reads affected by sequencing and basecalling errors tend to exhibit lower AS scores, potentially resulting in misalignment. The maximum achievable AS varies for each read and equals its span. The read span for each aligned read was determined based on the CIGAR string. All reads with an AS higher than 90% of the read span were retained for genotyping. No additional filtering was performed on Illumina data. Reads mapping to each allele were counted, and the allele with the highest count was called as present. If the allele with the second highest count was more than 50% of the highest count allele, it was also called as present¹². A different coverage threshold was set for each locus and it was calculated as 20 times the number of reference alleles for the given locus. If the coverage threshold was not met, genotyping was not performed.

Results

RPA assay results

The success of the RPA assay was first verified on a 2% agarose gel on sample 3 (Fig. 1). The singleplex RPA product displayed distinct bands at the anticipated amplicon lengths. Besides the amplicon bands, a band approximately 100 bp longer than the amplicon band is visible for several loci. Additionally, a smear is visible for most loci. A high level of primer artifacts is visible in the negative control lane, manifesting as a smear with equidistant bands. Thirteenplex RPA products exhibited a smeared profile with an unclear differentiation between bands (Suppl. Figure S1).

Capillary electrophoresis genotyping accuracy

All STR genotypes obtained by CE of the singleplex RPA product with the DNA input of 1 ng were concordant with the profiles obtained by the benchmarking method for all eight samples tested. Figure 2 presents an electropherogram of sample 4, while the electropherograms for the remaining samples amplified in singleplex can be found in the Suppl. Figure S4-S12. Electropherograms show distinct peaks corresponding only to the expected amplicon length, without discernible noise or artifact peaks. The presence and intensity of stutter products, a common artifact in PCR-based STR analysis, varied among the different loci amplified by RPA. Notably, no detectable stutter products were observed for the STR loci TH01 and D8S1179 across all samples analyzed. Small stutter peaks were visible in the electropherograms for loci D18S51 and D21S11 in six tested samples. These stutter peaks were below the level of quantification set in the CE analysis software, and therefore have no impact on genotyping accuracy.

Allelic balance was assessed by calculating the ratio between the peak heights of both alleles (RFU peak 1/ RFU peak 2) in the heterozygous samples across the loci. The average allelic balance was consistently between 0.72 and 0.82 in all samples and loci except for Amelogenin (Table 1).

Analysis of fiveplex RPA products by CE resulted in incomplete or incorrect STR profiles across all samples. The performance varied considerably among the different loci: D21S11 showed the highest success rate, with correct allele peaks observed in 6 out of 8 samples. Amelogenin, D8S1179 and D18S51 each displayed correct allele peaks in 5 out of 8 samples. Locus TH01 shows faint, but correct peaks in only one sample. The electropherograms of fiveplex RPA products exhibited variability in peak intensity both between samples and among loci within samples. Additionally, numerous off-ladder peaks were observed, possibly attributable to non-specific amplification and primer-dimer artifacts. Electropherograms of RPA fiveplex samples, are available in the Suppl. Figure S13-S21.

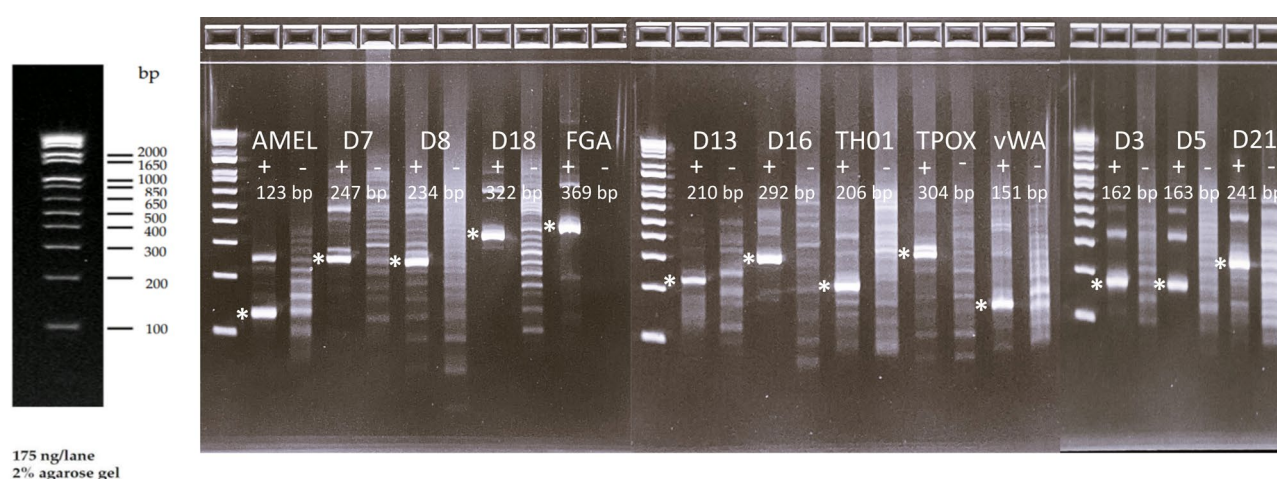


Fig. 1. Singleplex RPA product of sample 3 (“+” labeled lanes) on a 2% agarose gel. Lanes labeled with “-” contain negative controls. An Invitrogen E-Gel™ 1 kb Plus DNA Ladder (Thermo Fisher Scientific, Waltham, MA, USA) size marker is shown on the leftmost side of the figure. Expected amplicon sizes of the longer allele are indicated on the figure, underneath the shortened locus name. Anticipated locus-specific bands are denoted with an asterisk.

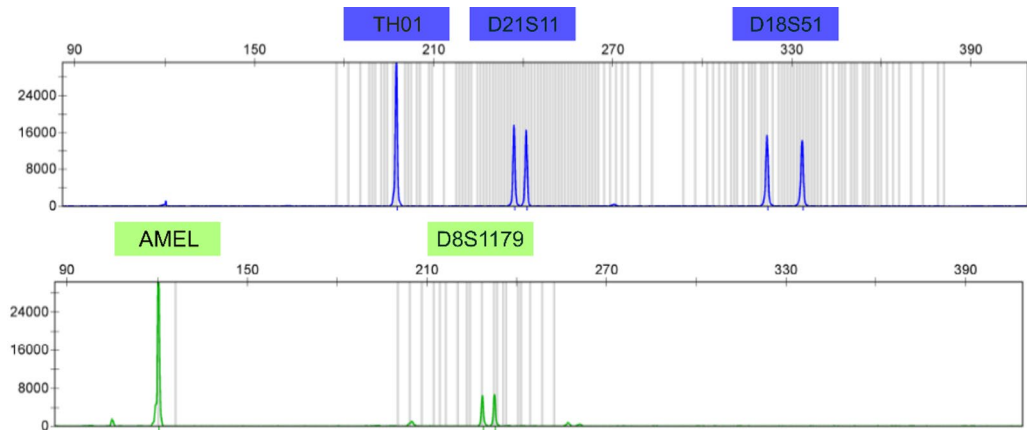


Fig. 2. Capillary electrophoresis electropherogram of the pooled singleplex RPA products of sample 4, generated using a DNA input of 1 ng. Small, off-ladder peaks are visible within the size ranges of loci Amelogenin, D8S1179 and D21S11.

Locus	Avg. AB	SD (AB)	Avg. PH/RFU	SD (PH)/RFU	Relative SD (PH)	n
Amel	0.45	0.016	18,272	9957	54.4%	2
D8S1179	0.72	0.26	7597	6601	86.9%	7
D18S51	0.76	0.22	10,227	5561	54.4%	8
D21S11	0.82	0.20	16,329	6042	37.0%	6
TH01	0.72	0.21	11,567	8198	70.9%	7

Table 1. Average allelic balance (Avg. AB) and average peak height (Avg. PH), with corresponding standard deviations, for heterozygous loci amplified in RPA singleplex reactions and analyzed by capillary electrophoresis. Allelic balance is expressed as the ratio of the lower to the higher peak height of the two alleles in heterozygous samples. The column ‘n’ indicates the number of heterozygous samples observed per locus.

Sensitivity assessment

Full STR profiles were consistently obtained with DNA inputs of 250 pg and 125 pg in both experimental replicates. One of the two 62 pg replicates showed allelic dropout at the D8S1179 locus, while the other replicate produced a complete profile. Input amounts of 500 pg and 31 pg resulted in D8S1179 peaks that were below the analytical threshold of the analysis software. Figure 3 shows the results of the sensitivity assessment.

Sequencing and genotyping results of RPA singleplex

On average, 164,193 (SD = 16080) reads were generated per singleplex sample during the Illumina sequencing run. Of these, 94.4% of generated pairs were successfully merged using FLASH tools. An average of 61.41% merged reads were assigned to a locus. Of the reads assigned to a locus, over 90% were uniquely mapped to an allele. The fraction of mapped reads mapping to a true allele was over 94% in all loci (Table 2).

Concordant STR profiles were obtained for all samples and loci in the singleplex setup (Fig. 4). The allelic balance (AB) in heterozygous samples was between 0.7 and 0.85 for all loci except FGA amplicons sequenced by ONT (Table 3).

On average, 8,705,546 (SD = 824405) passed-filter reads were generated per singleplex sample during the ONT sequencing run. Following rebasecalling, on average 63.2% of the passed-filter reads were assigned to a locus. More than 92% of the assigned reads were uniquely mapped to an allele. The fraction of reads mapping to a true allele ranged from 60.43% to 96.37% between loci. The percentages of reads mapping to a true allele are shown in Table 2. The STR profiles of samples amplified in singleplex were concordant with those obtained by benchmarking across all samples and loci. Similarly to the Illumina sequencing data, the allelic balance in heterozygous loci as measured with ONT sequencing, was between 0.7 and 0.85 for all loci, except for FGA (Table 3).

Multiplex RPA sequencing and genotyping

Illumina sequencing of the sixplex and thirteenplex samples resulted in on average 313,020 (SD = 25594) and 277,158 (SD = 20404) reads, respectively, of which over 94% were successfully merged by FLASH tool. On average, 53.66% and 27.40% of the merged reads were assigned to a locus in the sixplex and thirteenplex setups, respectively. Of the assigned reads, 93% and 89% were uniquely mapped to an allele in sixplex and thirteenplex, respectively. In the sixplex setup, only Amelogenin was incorrectly genotyped as X: X in sample 2, while the true genotype was X: Y. In the thirteenplex setup, only loci Amelogenin, D13S317 and D5S818 had sufficient coverage

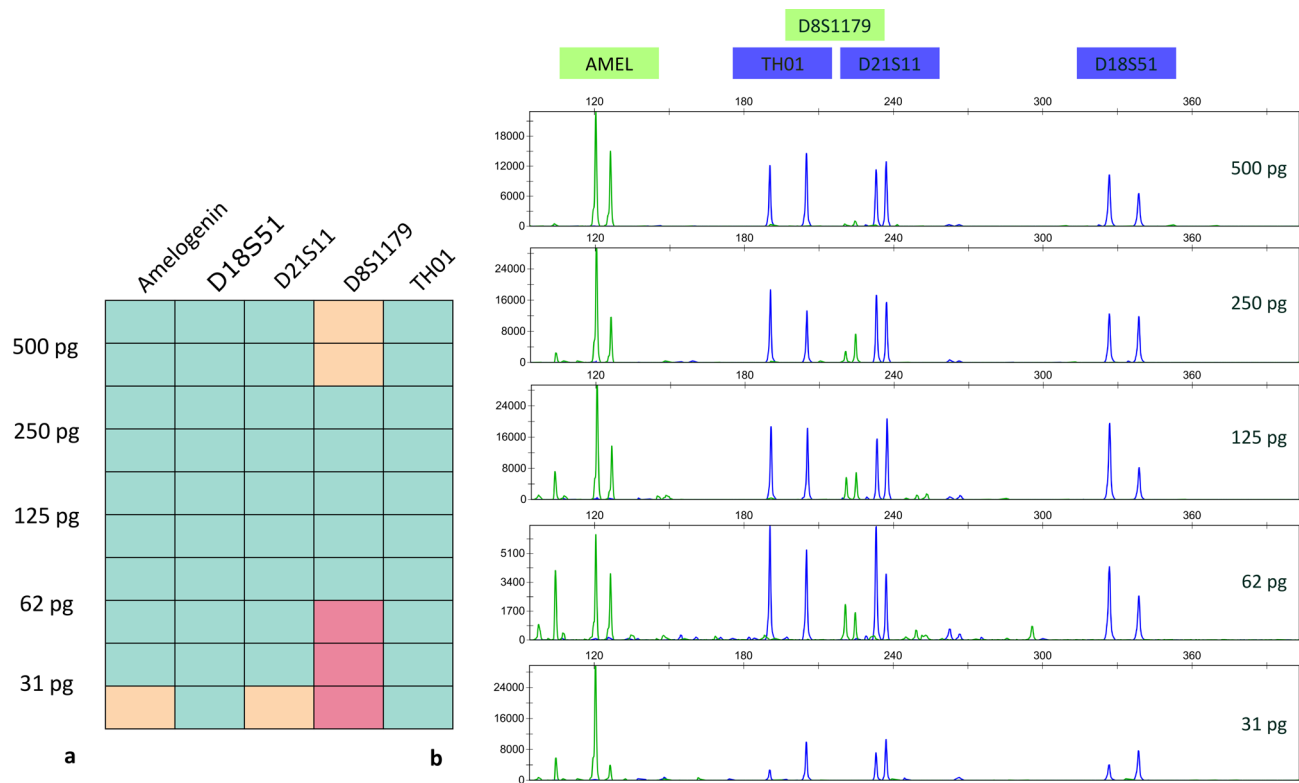


Fig. 3. Sensitivity assessment of RPA singleplex using a range of DNA input amounts (31–500 pg) of sample 2, performed in duplicate. **(a)** Genotyping concordance table: green fields indicate concordant genotypes for both alleles, orange fields indicate a single dropout, and red fields represent a double dropout. **(b)** Electropherograms of the first replicate of singleplex RPA products using different DNA input amounts.

	Illumina		ONT	
	Reads mapping to true allele (%)	Σ _{reads per locus}	Reads mapping to true allele (%)	Σ _{reads per locus}
AMEL	98.43	30,377	90.48	435,976
D13S317	97.02	28,523	90.77	1,065,401
D16S539	98.75	52,668	95.68	2,682,183
D18S51	94.42	39,748	76.50	3,731,515
D21S11	95.13	30,918	83.96	1,117,293
D3S1358	95.83	28,614	86.28	1,149,911
D5S818	95.78	23,193	76.16	950,446
D7S820	97.92	35,907	79.07	1,440,724
D8S1179	97.58	44,429	90.14	11,543
FGA	95.15	33,437	60.43	3,677,815
TH01	99.57	19,358	96.37	1,390,858
TPOX	99.43	8568	95.11	712,143
vWA	94.59	52,853	86.79	922,422

Table 2. Number of reads mapped per locus and the percentages of those reads that map to a true allele in singleplex RPA setup after illumina and ONT sequencing.

to establish a genotype in all samples, leading to correct genotypes for D13S317 and D5S818 and imbalanced results in the male samples for Amelogenin (Fig. 4).

To assess the amplification bias within the multiplex setups, the locus representation was calculated for each locus as a ratio of the number of reads assigned to the selected locus over the sum of all reads assigned to a locus. Graphic representation of the amplification bias for data generated via Illumina sequencing for all samples amplified in both sixplex and thirteenplex setups are shown in Fig. 5. The most abundant locus in both setups was Amelogenin, which is also the shortest locus. In the sixplex setup, the second most abundant locus was vWA, but its representation drops from 32.5% to 0.27% when transitioning from sixplex to thirteenplex setup.

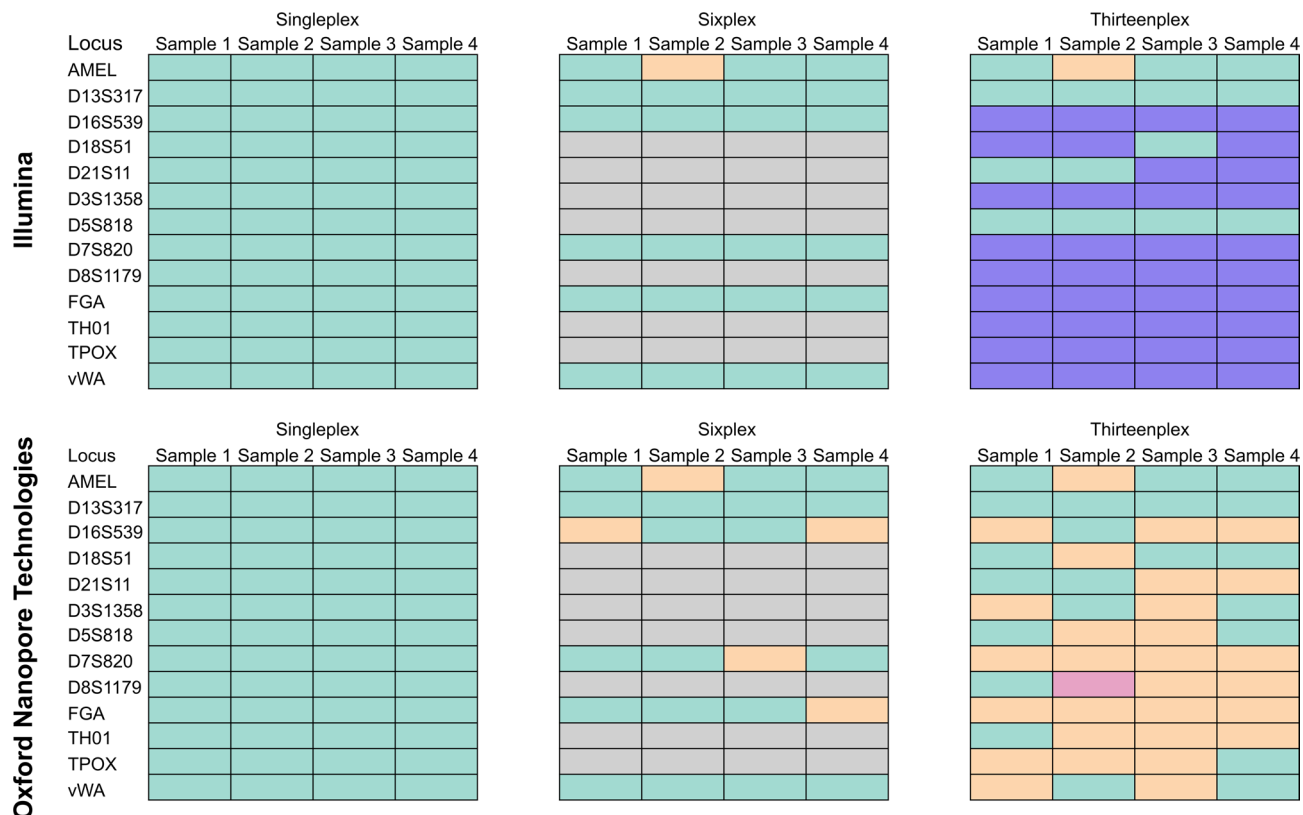


Fig. 4. Genotyping results of Illumina and ONT sequencing of singleplex, sixplex and thirteenplex setups. Green fields indicate concordant genotyping for both alleles, orange fields indicate one non-concordant allele and pink field indicates both alleles were non-concordant. Purple fields indicate the coverage is too low to establish a reliable genotype and grey indicates the loci that were not included in the RPA sixplex.

ONT sequencing metrics, as well as the locus amplification bias diagrams of sixplex and thirteenplex can be found in the Suppl. Table S1 and Suppl. Figure S2. Nanopore sequencing of the sixplex RPA product resulted in correct genotypes for loci D13S317 and vWA across all samples, while the loci Amelogenin, FGA and D7S820 were genotyped incorrectly in one of the samples and locus D16S539 was genotyped incorrectly in two samples. The thirteenplex RPA product exhibited higher error rates, with only locus D13S317 genotyped correctly across all samples (Fig. 4).

Discussion

STR genotyping of singleplex RPA amplicons

Distinct bands at anticipated lengths on an agarose gel of singleplex RPA product demonstrate that STR loci were successfully amplified using RPA. The presence of artifacts, including longer bands and smears on the agarose gels, could raise a concern regarding the specificity of the RPA product as non-specific products could complicate interpretation of results, especially in cases with limited or degraded DNA. Additional bands and smears on the agarose gels likely represent non-specific products that lack the fluorescently labeled primer and therefore do not appear in CE electropherograms. This interpretation is supported by the sequencing results, where 61% of the reads were assigned to a locus based on the presence of the primer sequences, and where a high proportion of these reads (90%) mapped to an allele. This implies that these artifacts neither carry complete primer sequences nor are they abundant enough to compromise the genotyping accuracy of singleplex RPA products at 1 ng input.

An increase in dropout frequency with decreasing DNA input is expected and has been consistently reported for amounts below 100 pg^{1,26,27}. Accordingly, we observed dropouts at Amelogenin and D21S11 loci with a 31 pg input, and at D8S1179 locus with both 62 pg and 31 pg inputs. Unexpectedly, dropouts were also observed at D8S1179 locus with the 500 pg input. At this stage, the underlying causes of dropouts at the 500 pg DNA input remain unclear. However, D8S1179 locus exhibited the lowest average peak height and the highest peak height variability (Table 1), suggesting a lack of robustness in the D8S1179 assay that may be affecting amplification efficiency. Therefore, changes in primer design should be implemented in order to increase the efficiency and robustness of the D8S1179 assay. However, only two replicates were included in this sensitivity study, and the dropouts may have occurred stochastically. More samples and replicates should be analyzed in order to have a robust assessment of the assay sensitivity. Extensive optimization for each locus specifically was not performed for this proof-of-concept study. Average AB across both sequencing methods and CE was more than 0.7 (SD

Locus	Illumina		ONT		n
	Avg. AB	SD	Avg. AB	SD	
AMEL	0.53	NA	0.68	NA	1
D13S317	0.75	NA	0.77	NA	1
D16S539	0.76	0.18	0.78	0.16	3
D18S51	0.74	0.09	0.76	0.12	4
D21S11	0.76	0.10	0.77	0.14	3
D3S1358	0.73	0.09	0.80	0.04	3
D5S818	0.74	0.11	0.70	0.06	3
D7S820	0.70	0.16	0.76	0.14	3
D8S1179	0.79	0.22	0.68	0.10	4
FGA	0.84	0.17	0.83	0.19	4
TH01	0.83	0.06	0.88	0.09	3
TPOX	0.72	0.23	0.70	0.21	3
vWA	0.83	0.08	0.91	0.003	2

Table 3. Average allelic balance (AB) of heterozygous singleplex samples. AB is represented as the number of reads mapping to the allele with a lower count, divided by the number of reads mapping to the higher count allele. The resulting ratio was averaged across the heterozygous loci of 4 samples amplified in singleplex. The number of heterozygous loci across the different samples is indicated in column n.

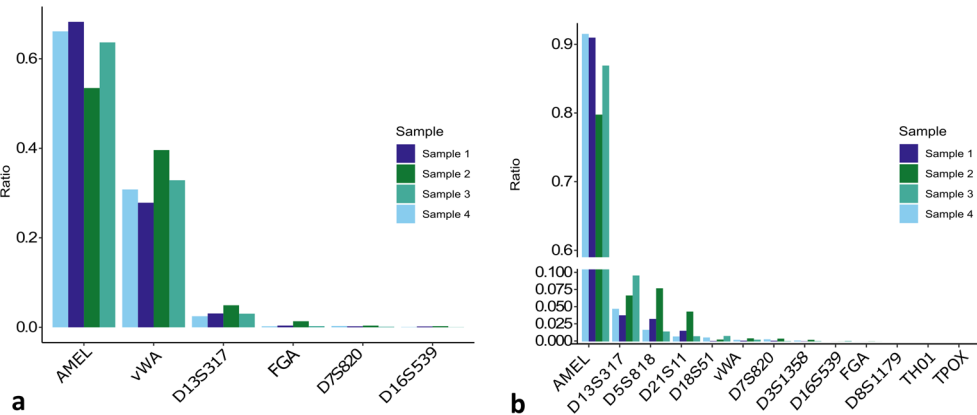


Fig. 5. Amplification bias in RPA multiplexes, as measured by Illumina sequencing. The Y-axis shows the ratio of reads assigned to a certain locus divided by the sum of all reads assigned to any locus in (a) Sixplex setup and (b) Thirteenplex setup. Notably, the analysis performed with ONT sequencing produced very similar results (Suppl. Figure S2).

= 0.2) in all loci amplified in singleplex except Amelogenin (0.45 – 0.78) (Tables 1 and 3). AB in widely used commercial kits, as evaluated in the 2016 study²⁸, is around 60%, which is comparable to the AB achieved using RPA singleplex. In profiles generated by Illumina sequencing, proportion of the reads mapping to the Y allele of Amelogenin locus was only 53% of the X allele read count. Similarly, CE profiles demonstrated a below average level of balance between X and Y peak height, with average Y allele peak height being 45% of X allele peak height. Further research is needed to establish the causes of the imbalance between the X and Y allele in Amelogenin locus when amplified using RPA. Nevertheless, in the electropherograms, the Y peak is still well above the analytical threshold and can therefore confidently be called as present.

Multiplexing of the RPA assay

In the sixplex setup, Illumina sequencing yields concordant genotypes for all loci except Amelogenin in sample 2, where incorrect genotyping stems from the allelic imbalance resulting in Y allele dropout. Due to high similarity between X and Y alleles, a small fraction of reads map to Y allele in female samples. The allelic balance between Y and X alleles in female samples were 0.1% and 0.003% for sixplex and thirteenplex, respectively, compared to 9.6% and 19.3% in the male sample. This suggests that a different allelic balance threshold should be considered for calling a male sample for Amelogenin in multiplex RPA setups.

Coverage for samples amplified in sixplex was above the set coverage threshold for all loci, although a considerable amount of amplification bias was observed. In the sixplex setup, the Amelogenin and vWA loci accounted for an average of 96.7% of reads, while the least represented locus, D16S539, was represented by only 0.1% of the reads, which still satisfied the read count threshold. In the thirteenplex setup, a shift occurred.

The fraction of vWA reads dropped from 32.5% to 0.2%, and the fraction of FGA reads dropped from 0.5%, which still provided sufficient coverage, to 0.009%. These changes in amplification levels are likely due to primer interactions introduced by the addition of primer sets to the singleplex and sixplex, as all loci amplified well in the singleplex and satisfactory in sixplex. Since different loci are used in the fiveplex and sixplex panels, the interactions between primers of different loci is likely the reason dropouts and incorrect peaks occur in the CE profiles of the fiveplex samples and not in the sequencing results of the sixplex panel. The absence of denaturation and annealing steps in RPA, combined with increased total oligonucleotide concentration in the thirteenplex, promotes the amplification of primer dimers which competes with the amplification of the target sequence²¹. These findings are consistent with the observed off-ladder peaks that were seen on the electropherograms of fiveplex samples (Suppl. Figure S13–S21).

In this proof-of-concept study, extensive optimization and validation of the RPA multiplexes was not conducted. However, adequate coverage and high genotyping concordance of the sixplex reaction prove the potential of RPA for multiplex amplification of forensic STR loci. At the moment, and to the best of our knowledge, sixplex described in our study is the multiplex with the highest number of loci amplified at once using RPA. Most published RPA multiplexes involve no more than four primer sets at once²⁹ and are primarily focused on pathogen detection in animals, humans, food and plants¹⁷. Therefore, literature reporting in-depth exploration of optimal primer composition for multiplexing and optimal RPA multiplex conditions is scarce. In a study by Tullayaprayouch et al.³⁰, during the primer design steps, the authors looked at the Gibbs free energy (ΔG) of the primer dimers and redesigned the primers which had the ΔG lower than -9 kcal/mol. This is potentially a good first step in the screening of the primers for multiplex. The optimizations of RPA multiplex should also entail adjustment of relative primer concentrations, as well as their length and sequence. Reaction temperature and duration, MgOAc and dNTP concentrations are variables that have a potential of improving the RPA multiplex performance, as well as the additives such as bovine serum albumin (BSA), gelatine and glycerol²⁰.

Even though our findings showcase promising early results for forensic RPA singleplex and sixplex, demonstrating RPA's full utility in forensics as a replacement or complement for PCR in forensic workflows would require multiple optimization and validation studies, which fall outside of the scope of this proof-of-concept research. We believe in the potential for integrating RPA into a microfluidic STR genotyping device, either in a singleplex or multiplex setup, but we are aware this remains a theoretical possibility. Developing a market-ready, lab-on-a-chip device for forensic RPA would require significant investment and effort, well beyond the scope of this proof-of-concept study. Nonetheless, we hope our manuscript inspires future initiatives to pursue this vision. Given the focus of our study, it would be inappropriate to compare our findings to state-of-the-art portable solutions such as ANDE or RapidHIT™ ID, which are fully integrated systems.

STR genotyping using nanopore sequencing

ONT sequencing of RPA singleplex product revealed that between 60.4% (for FGA) and 96.3% (for TH01) of mapped reads were mapped to a true allele. This is somewhat lower than for Illumina sequencing in which over 94% of mapped reads were mapped to correct alleles. This reflects the lower general sequencing accuracy of Nanopore sequencing data, which is especially pronounced in the repetitive and homopolymeric regions characterizing STR loci³¹. Suppl. Table S1 compares mapping percentages of ONT sequencing results generated in this study to mapping percentages of previously published PCR-based ONT sequencing of STR loci¹¹. The comparison shows that the proportion of total reads assigned to a locus and reads mapped to an allele is considerably higher in our study compared to the 2020 study¹¹. In the same study, the genotypes generated by Illumina sequencing were all in concordance with the ones generated by the benchmarking method, while some loci were consistently genotyped incorrectly by ONT sequencing. Consistent failing of certain STR loci in ONT sequencing was reported in two other studies^{10,12} in which nanopore sequencing was performed on samples prepared using the commercially available Verogen ForenSeq multiplex kit (Verogen, San Diego, USA). These results, along with earlier studies³² point to the lack of maturity of Nanopore sequencing technology for forensic STR genotyping at the time. Since correct genotypes in our study were generated from a pooled singleplex, in contrast to the multiplex assay in the 2020 Tytgat paper, parallels cannot be drawn between the two amplification methods due to added complexity of the multiplex in the PCR assay. However, concordant genotyping of RPA singleplex in this study points to the improvement of ONT sequencing in dealing with repetitive regions. Recent advancements in both sequencing technology and data analysis pipelines³³ are narrowing the performance gap between Illumina and ONT sequencing platforms for STR profiling. While Nanopore sequencing workflow is not considerably faster than Illumina sequencing or CE workflows, ONT's MinION device comes with a low upfront cost of around 2000 USD, considerably lower than Illumina or CE setups. Reliable forensic genotyping ONT sequencing workflow is a potential solution for low-resource communities that are in regular or occasional need for forensic STR genotyping. Beside the cost, MinION device is portable and USB powered. Recent pilot study³⁴ demonstrates fully of-grid forensic STR Nanopore sequencing in the field. If fully developed and validated, in-the-field Nanopore sequencing can become a valuable tool in the forensic genotyping toolbox, especially in specific situations such as disaster scenes in remote locations. Because of RPAs ability to reduce stutter³, its incorporation in ONT forensic sequencing workflow may lead to easier interpretation of the profiles.

Conclusion

This study evaluates RPA as an amplification method within the standard STR typing workflows: CE and MPS. Sequencing using both Illumina and ONT technologies, as well as CE of the singleplex RPA product resulted in complete and donor-concordant profiles across all samples and loci for the input amount of 1 ng. By conducting a sensitivity assessment we demonstrated the successful amplification of four out of five analyzed loci from as little as 62 pg of template DNA. Dropouts were observed in locus D8S1179 for several DNA inputs. Illumina

sequencing of the sixplex RPA product resulted in correct and complete STR profiles in all samples and loci except for one dropout in Amelogenin locus. These findings provide initial evidence for the feasibility of using RPA in multiplex STR amplification. However, the effort to simultaneously amplify all thirteen STR loci using RPA failed to produce complete and accurate profiles after DNA sequencing analysis. Similarly, CE analysis of a five-locus RPA multiplex failed to generate complete and accurate profiles. This outcome suggests that the current RPA chemistry may struggle with the increased complexity of a full forensic STR panel when amplified simultaneously. Even at a lower level of multiplexing, challenges remain in balancing amplification efficiency across multiple loci. Therefore, further optimization is required before RPA multiplex can be reliably used for generating multiplexed forensic STR profiles.

Data availability

Sequencing data related to this article is available at the NCBI Sequence Read Archive using the BioProject ID: PRJNA1152328. The data will be made available online on 1st September 2025. Prior to the release date, meta-data can be accessed via the following reviewer's link: <https://dataview.ncbi.nlm.nih.gov/object/PRJNA1152328?reviewer=mupkqtp3q8atcqn3dqlc88r1e>. For additional requests related to the sequencing data, please contact the corresponding author. For the in-house Python script used in the analysis of sequencing data, please contact the corresponding author.

Received: 3 July 2025; Accepted: 13 October 2025

Published online: 18 November 2025

References

- Butle, J. M. Advanced topics in forensic DNA typing: methodology. *Advanced topics in forensic DNA typing: methodology* (Elsevier Academic), <https://doi.org/10.1016/C2011-0-04189-3>. (2011).
- Dashnow, H. et al. STretch: detecting and discovering pathogenic short tandem repeat expansions. *Genome Biol.* **19**, 1–13 (2018).
- Daunay, A. et al. Low temperature isothermal amplification of microsatellites drastically reduces stutter artifact formation and improves microsatellite instability detection in cancer. *Nucleic Acids Res.* **47**, e141 (2019).
- Iyavoo, S. et al. Unveiling genetic diversity and forensic utility of SureID[®] human DNA identification kits: a comprehensive analysis of 44 autosomal STR loci in English and Irish populations. *Ann. Hum. Biol.* **50**, 399–406 (2023).
- Liu, Y. Y., Cheng, K., Just, R., Enke, S. & Bright, J. A. Sequencing-induced artefacts in NGS STR data. *Forensic Sci. Int. Genet.* **72**, 103086 (2024).
- Börsting, C. & Morling, N. Next generation sequencing and its applications in forensic genetics. *Forensic Sci. Int. Genet.* **18**, 78–89 (2015).
- van Asten, A. C. On the added value of forensic science and grand innovation challenges for the forensic community. *Sci. Justice.* **54**, 170–179 (2013).
- Bruijns, B., Van Asten, A., Tiggelaar, R. & Gardeniers, H. *Microfluidic Devices Forensic DNA Analysis: Rev.* <https://doi.org/10.3390/bios6030041>
- Advancing Justice Through DNA Technology. Using DNA To Solve Crimes. (2017). <https://www.justice.gov/archives/ag/advancing-justice-through-dna-technology-using-dna-solve-crimes>
- Ren, Z. L. et al. Forensic nanopore sequencing of STRs and SNPs using verogen's forenseq DNA signature Prep kit and minion. *Int. J. Legal Med.* **135**, 1685–1693 (2021).
- Tytgat, O. et al. Nanopore sequencing of a forensic STR multiplex reveals loci suitable for single-contributor STR profiling. *Genes (Basel)* **11**, 381 (2020).
- Tytgat, O., Škevin, S., Deforce, D. & Van Nieuwerburgh, F. Nanopore sequencing of a forensic combined STR and SNP multiplex. *Forensic Sci. Int. Genet.* **56**, 102621, (2022).
- Quick, J. et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).
- Tyler, A. D. et al. Field-based detection of bacteria using nanopore sequencing: method evaluation for biothreat detection in complex samples. *PLoS One* **18**, e0295028 (2023).
- Piepenburg, O., Williams, C. H. & Stemple, D. L. Armes, N. A. DNA detection using recombination proteins. *PLoS Biol.* **4**, 1115–1121 (2006).
- Zhang, T. T. et al. Rapid and simple detection of *Glaesserella parasuis* in synovial fluid by recombinase polymerase amplification and lateral flow strip. *BMC Vet. Res.* **15**, 1–8 (2019).
- Tan, M. et al. Recent advances in recombinase polymerase amplification: Principle, advantages, disadvantages and applications. *Front. Cell. Infect. Microbiol.* **12**, 1–13 (2022).
- Kim, J., Jung, J. S., Lee, S. & Yoo, J. Y. Applications of recombinase polymerase amplification in forensic science. *Biomed. Sci. Lett.* **31**, 99–114 (2025).
- Zheng, Y. et al. Development of specific and rapid detection of human DNA by recombinase polymerase amplification assay for forensic analysis. *Forensic Sci. Int. Genet.* **66**, 102903 (2023).
- Lillis, L. et al. Factors influencing recombinase polymerase amplification (RPA) assay outcomes at point of care. *Mol. Cell. Probes.* **30**, 74–78 (2016).
- Liljander, A. et al. Field-Applicable recombinase polymerase amplification assay for rapid detection of *Mycoplasma capricolum* subsp. *Capripneumoniae*. *J. Clin. Microbiol.* **53**, 2810 (2015).
- Sweet, D., Lorente, M., Valenzuela, A., Lorente, J. A. & Alvarez, J. C. Increasing DNA extraction yield from saliva stains with a modified chelex method. *Forensic Sci. Int.* **83**, 167–177 (1996).
- Butler, J. M. et al. NIST Short Tandem Repeat DNA Internet Database. *National Institute of Standards and Technology* (1997). <https://doi.org/10.18434/T44G6P>
- Magoč, T., Salzberg, S. L. & FLASH Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Gill, P., Whitaker, J., Flaxman, C., Brown, N. & Buckleton, J. An investigation of the rigor of interpretation rules for STRs derived from less than 100 Pg of DNA. *Forensic Sci. Int.* **112**, 17–40 (2000).
- Qu, W. et al. Internal validation of the GA118-24B genetic Analyzer, a stable capillary electrophoresis system for forensic DNA identification. *Int. J. Legal Med.* **138**, 361–373 (2024).
- Mattayat, D., Kitpipit, T., Phetpeng, S., Asawutmongkul, W. & Thanakiatkrai, P. Comparative performance of AmpFLSTR[®] Identifier[®] plus PCR amplification kit and QIAGEN[®] Investigator[®] idplex plus kit. *Sci. Justice.* **56**, 468–474 (2016).

29. Lin, Z. et al. Infectious disease diagnosis and pathogen identification platform based on multiplex recombinase polymerase Amplification-Assisted CRISPR-Cas12a system. *ACS Infect. Dis.* **9**, 2306–2315 (2023).
30. Tullayaprayouch, K. et al. Development of multiplex recombinase polymerase amplification for the rapid detection of five carbapenemase (blaKPC, blaNDM, blaOXA-48-like, blaIMP, and blaVIM) and 10 Mcr (mcr-1 to Mcr-10) genes in blood cultures. *Int J. Antimicrob. Agents* **66**, 107567 (2025).
31. Delahaye, C. & Nicolas, J. Sequencing DNA with nanopores: troubles and biases. *PLoS One* **16**, e0257521 (2021).
32. Asogawa, M. et al. Human short tandem repeat identification using a nanopore-based DNA sequencer: a pilot study. *J. Hum. Genet.* **65**, 21–24 (2020).
33. Hall, C. L. et al. Accurate profiling of forensic autosomal STRs using the Oxford nanopore technologies minion device. *Forensic Sci. Int. Genet.* **56**, 102629 (2022).
34. Luo, Y. et al. Pilot validation of on-field STR typing and human identity testing by minion nanopore sequencing. *Electrophoresis* **1–12** <https://doi.org/10.1002/elps.202300234> (2024).

Author contributions

S.Š. conceptualized the study and experimental setup, conducted part of the experiments and data analysis, and wrote and reviewed the manuscript. L.D.K., L.D.W., and O.T. performed parts of the experiments and contributed to the writing and editing of the manuscript. D.V.H. conducted the CE experiments and data analysis. D.D. and F.V.N. contributed to the conceptualization and supervision of the experiment and reviewed and edited the manuscript.

Funding

Research was funded by a PhD grant provided by Research Foundation Flanders [1SC6522N] awarded to Sonja Škevin.

Declarations

Competing interests

Sonja Škevin reports financial support in terms of PhD grant that was provided by Research Foundation Flanders [1SC6522N]. Other authors declare no conflict of interest.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-24283-2>.

Correspondence and requests for materials should be addressed to F.N.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025