Forensic Science International: Genetics 86 (2027) 103548

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

SiMSen-Seq STR sequencing across five laboratories enables lowered noise thresholds and improved allele calling

Arvid H. Gynnå [a], Maja Sidstedt [a], Kevin M. Kiesler [b], Adam Staadig [c,d], Åsa Lindgren [e], Firaol Tamiru Kebede [f], Joakim Håkansson [e,g], Carolyn R. Steffen [b], Tobias Österlund [f,h,i], Yalda Bogestål [d], Andreas Tillmar [c,d], Peter M. Vallone [b], Anders Ståhlberg [f,h,i,j], Johannes Hedman [a,k,*]

[a] National Forensic Centre, Swedish Police Authority, Linköping SE-581 94, Sweden
[b] National Institute of Standards and Technology, 100 Bureau Drive, M/S 8314, Gaithersburg, MD 20899, USA
[c] Department of Forensic Genetics and Forensic Toxicology, National Board of Forensic Medicine, Linköping SE-587 58, Sweden
[d] Department of Biomedical and Clinical Sciences, Faculty of Health Sciences, Linköping University, Linköping SE-581 83, Sweden
[e] RISE Research Institutes of Sweden, Methodology, Textile and Medical Device Technology, Brinellgatan 4, Borås 504 62, Sweden
[f] Department of Laboratory Medicine, Sahlgrenska Center for Cancer Research, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Medicinaregatan 1F, Gothenburg 41390, Sweden
[g] Department of Laboratory Medicine, Institute of Biomedicine, University of Gothenburg, Sahlgrenska Academy, Gothenburg, Sweden
[h] Wallenberg Centre for Molecular and Translational Medicine, University of Gothenburg, Medicinaregatan 1F, Gothenburg 41390, Sweden
[i] Department of Clinical Genetics and Genomics, Sahlgrenska University Hospital, Medicinaregatan 1D, Gothenburg, Region Västra Götaland 41390, Sweden
[j] Science for Life Laboratory, Institute of Biomedicine, University of Gothenburg, Medicinaregatan 1F, Gothenburg 41390, Sweden
[k] Division of Biotechnology and Applied Microbiology, Department of Process and Life Science Engineering, Lund University, Lund SE-221 00, Sweden

ARTICLE INFO

Keywords:
Forensic DNA analysis
Interlaboratory study
Massively parallel sequencing (MPS)
Short tandem repeats (STR)
Unique molecular identifiers (UMIs)

ABSTRACT

Unique molecular identifiers (UMI) can be used in forensic STR sequencing to reduce the level of analytical artifacts. Here, we perform an interlaboratory study across five independent sites where the previously developed UMI-based SiMSen-Seq STR assay is applied at each laboratory to both single-source and mixed samples. The assay showed consistent results between laboratories and more than 90% of the expected alleles were detected with 31 pg DNA of template. The assay tolerated ten times higher PCR inhibitor concentrations compared to an established commercial STR sequencing method. The combined results were used to determine stutter and noise thresholds, which were applied for allele calling, allowing an estimation of the sensitivity for minor contributor alleles in mixtures. We found that the SiMSen-Seq STR method is robust across laboratories and different types of PCR and sequencing equipment and that it allows for calling of alleles from contributors of smaller proportions compared to currently commercially available non-UMI STR sequencing methods.

1. Introduction

Short tandem repeat (STR) sequencing kits for human identification have been available for about a decade as an alternative to the widely used capillary electrophoresis (CE) methods [1,2]. Sequencing of STRs has theoretical advantages over CE, including a higher number of allele variants and the possibility to create larger multiplexes to analyze more markers simultaneously, both of which are expected to increase the discriminatory power [3]. The theoretical benefits have, however, been challenging to realize in practice due to PCR-induced noise and artifacts [4,5].

Since STR sequencing resolves the fragment sequence and not just the fragment length as CE does, more complex types of noise and artifacts will be detected [6,7]. Common artifacts in STR sequencing include stutter, i.e. the loss or gain of one or multiple STR repeats in PCR, and single base pair errors which may occur either during PCR or during sequencing. Substitution errors are not detected with CE, and while stutters exist, more stutter variants are discerned in sequencing, often complicating the interpretation of results [8]. For example, in the compound locus D12S391 there are two stretches of repeats (TAGA and

* Corresponding author at: National Forensic Centre, Swedish Police Authority, Linköping SE-581 94, Sweden.
E-mail address: johannes.hedman@tmb.lth.se (J. Hedman).

https://doi.org/10.1016/j.fsigen.2026.103548
Received 20 December 2025; Received in revised form 14 April 2026; Accepted 31 May 2026
Available online 1 June 2026

GAGA). If either of these stretches lose a repeat, a single minus one stutter will be detected in CE while different artifacts will be recorded in sequencing depending on where the repeat was lost. More confusingly, if one stretch loses a repeat while the other gains one on the same molecule, the expected length allele will be recorded in CE while a sequence with multiple errors can be detected by sequencing.

DNA samples from crime scenes often contain mixtures of biological material from multiple contributors. It is not unusual that the person of interest is a minor contributor, sometimes making up only a few percent of the total DNA, with the overwhelming DNA contribution coming from a victim or a person not connected to the crime. Reliable calling of the alleles of such minor contributors is a major challenge in forensic genetics [9]. Since allele calling thresholds are commonly set as percentages of the total number of reads at a marker, the level of noise and artifacts effectively determines the sensitivity for minor contributors. Several methods have been devised to decrease the level of errors. At the stage of PCR, much effort has been put into high fidelity DNA polymerases with low levels of substitution errors [10]. The rate of replication slippage – i.e. stuttering – can be lowered by choice of polymerase and reaction conditions, although the mechanisms remain unclear [11,12]. An engineered DNA polymerase with substantially reduced stuttering has recently been announced [13], but has, to our knowledge, not yet been used for sequencing. One commercially available STR sequencing kit employs reverse complementary PCR, where primers are synthesized during the amplification, to decrease the formation of unspecific PCR products [14]. In other current STR sequencing kits, standard library preparation approaches are used, with either two PCR steps [15,16] or a single amplification followed by adaptor ligation [17,18]. While these kits realize the nominal benefits of STR sequencing – larger multiplexes and higher numbers of possible alleles – studies investigating evidential values for minor contributors have often showed small improvements or sometimes even weaker conclusions compared to CE [4,5,19].

An alternative, or complement, to decreasing error rates biochemically is the use of bioinformatic methods. Unique molecular identifiers (UMIs), also known as barcodes, enable the removal of errors in bioinformatic post-processing [20]. UMIs are effective for reducing artifacts arising both in the amplification and sequencing, and also allow more accurate estimation of the proportions of different alleles in the DNA template [21,22]. In forensic genetics, UMIs have been shown to improve allele calling in a large-scale SNP panel and in microhaplotype analysis [23–25]. Forensic STR analysis can be especially challenging, due to the high levels of systematic artifacts. However, UMIs have been demonstrated to decrease noise and artifact levels also for STR amplicons [26]. Specifically, using the SiMSen-Seq method resulted in a tenfold decrease of the error rate compared to sequencing without UMIs [8].

In previous work, the SiMSen-Seq method for STR sequencing using UMIs was described and demonstrated to drastically decrease error rates and improve minor contributor detection [8]. However, one limitation was that interpretation thresholds for allele calling were not determined. The sensitivity of allele calling could thus not be compared to other STR typing methods. In this study, we continue to evaluate the SiMSen-Seq STR method by performing the analysis at five independent laboratories. By combining data from all laboratories, we propose noise and stutter thresholds and then compare sensitivity for minor contributors to a state-of-the art non-UMI STR sequencing method. Furthermore, the previous work is extended by validating the method across laboratories, with respect to sequencing quality parameters, concordance, error rates and sensitivity for minor contributors using several types of PCR and sequencing instruments. We also investigate PCR inhibitor tolerance and required read depth, which are two parameters of practical importance for casework.

2. Methods

2.1. Library preparation

Five laboratories, here referred to as L1 – L5, participated in the study. Each laboratory performed one round of library preparation and DNA sequencing according to the previously published 7-plex forensic SiMSen-Seq STR method [8]. The method is based on SiMSen-Seq, i.e. Simple, multiplexed, PCR-based barcoding of DNA for sensitive mutation detection using sequencing [27], where an initial marker-specific barcoding PCR is performed with primers including a UMI. The barcoding primers are used at low concentrations and have a stem-loop preventing binding to the UMI at temperatures below the stem-loop melting point, both measures intended to decrease the formation of unspecific PCR products. Four cycles of barcoding PCR are performed. Then, the UMI-labeled fragments are amplified with 30 cycles in a second PCR step with indexing primers, followed by bead purification, normalization, pooling and sequencing.

Two minor changes were applied compared to the published protocol [8]. As the D8S1179 locus was underrepresented among the consensus reads in the previous study, the concentration of primers for D8S1179 was increased from 0.08 to 0.10 μM (Supplementary table S 1). Also, the bead-based PCR cleanup was repeated twice with bead ratios 0.8 and 1X, respectively, to increase the proportion of reads that align to the target loci.

Differences in reagents and instrumentation between the laboratories are noted in Table 1. The barcoding primers were ordered as Ultramer DNA oligos (IDT Integrated DNA Technologies, Coralville, IA, USA) as three separate batches for laboratories L1, L2, L5 (batch A); L3 (B); and L4 (C), respectively. For the first batch (A), the primers for the different markers were mixed at two occasions, once for L1 and L5 (A1), and once for L2 (A2). Adapter primers were independently ordered by each laboratory as standard DNA oligos (IDT).

2.2. Study design

Each sequencing batch included 24 reactions that were identical across the laboratories, and 8 (Labs L1 – L4) or 24 (L5) reactions that were unique for each laboratory. Template DNA for the 24 identical reactions were prepared by the organizing lab, and shipped on ice to the participants where dilutions were performed. The identical samples were designed to evaluate concordance with other STR analysis methods, repeatability and detection of minor contributor alleles in mixtures. Additionally, individual laboratories studied limit of detection for single-source samples, PCR inhibitor tolerance and performance for DNA mixtures with 3–4 contributors.

Repeatability was evaluated through triplicate amplifications of 1 ng of standard DNA 2800 M (Promega Corp., Madison, WI, USA). These reactions were also used to study concordance together with NIST

Table 1
Reagents and instruments used at each laboratory.

Lab	Primer batch	Primer mix[a]	No of reactions	PCR instrument	Sequencing instrument	Reagent type[b]
L1	A	A1	32	BioRad T100	MiSeq	V3
L2	A	A2	32	Veriti	MiSeq FGx	FGx
L3	B	B1	32	ProFlex	MiSeq FGx	V3
L4	C	C1	32	Veriti/ ProFlex[c]	MiSeq FGx	V3
L5	A	A1	48	BioRad T100	NextSeq 2000	P1

[a] Preparation of primer stocks.
[b] Type of MiSeq or NextSeq sequencing reagents, according to Illumina nomenclature.
[c] Veriti was used for the barcoding PCR, ProFlex for the indexing PCR.

standard reference material 2391d components A-D, which were analyzed at 1 ng and 250 pg at each laboratory [28]. The sensitivity for minor contributors in DNA mixtures was evaluated using mixtures of 2800 M and standard DNA NA24385 (Verogen, San Diego, CA, USA) at proportions 4:1, 9:1, 19:1 and 49:1, which were sequenced with total DNA template amounts of 1 ng, 500 pg and 250 pg at each laboratory, and additionally at two laboratories the 4:1 mixture at 125 pg, 62 pg and 31 pg DNA in total. Furthermore, one laboratory analyzed four 1 ng three- and four contributor mixtures from the Forensic DNA Open Dataset in duplicates (mixtures B1_3P-A_1–1–98, A2_3P-B_1–49–50, A6_4P-A_1–1–1–97, C6_4P-A_5–5–5–85) [29].

The PCR inhibitor tolerance was evaluated by addition of humic acid (1, 2, 4, 8 or 16 ng/μL; Sigma-Aldrich, Munich, Germany) and hematin (50, 100, 200 or 400 μM; Apollo Scientific, Manchester, UK) to 8 μL sample containing 1 ng 2800 M template DNA and analyzing in duplicate reactions at one laboratory. The amplified libraries were analyzed by electrophoresis using a 5200 Fragment Analyzer System with the HS Small Fragment kit (Agilent Technologies, Santa Clara, CA, United States). Two concentrations of each inhibitor were selected for sequencing based on electrophoresis results.

2.3. DNA sequencing and data analysis

For sequencing, a MiSeq instrument (Illumina, San Diego, CA, USA) was used at all laboratories except L5, which used a NextSeq 2000 sequencer (Illumina). Sequencing results were transferred as FastQ files to the organizing lab. Deduplication and allele calling were performed with the UMIec Forensics pipeline v 0.2.1 (https://github.com/agynna/UMIec_forensics) and results were evaluated using custom Python scripts (available on request). The reads aligned to markers prior to consensus formation are here referred to as “STR reads”, while the reads representing each UMI family are called “consensus reads”. For generation of a consensus read, the most common sequence among the STR reads sharing a UMI (a *UMI family*) was selected. Furthermore, only UMI families with at least three STR reads and at least 50% purity (i.e. proportion of reads identical to the consensus read) were accepted for further analysis. Where indicated below, a machine learning model (ML) was used to determine which consensus sequences to accept, instead of the fixed criteria described above. As described previously, a neural network was then used to score UMI families with acceptance criteria tuned per marker to accept a similar rate of UMI families as the fixed numeric criteria [8].

Stutter ratio was defined as the number of reads for the stutter product divided by the number of reads for the main allele. Only stretches of five or more repeats were considered for stuttering. When multiple stutter products were observed from the same allele, as is seen with compound and complex markers, separate ratios were calculated for each product. If no consensus reads of any stutter product were observed from an allele, a stutter artifact with zero reads was inserted. Stutter ratios and noise levels in the consensus reads were calculated from 94 single source samples with 1000 or 250 pg DNA template, combined from this work and Sidstedt et al. [8].

In this work, *errors* refer to all non-allelic sequences, *artifacts* refer to sequence errors that arise through predictable processes during the PCR amplification (i.e. single repeat plus and minus stutters) while *noise* are sequences that arise through processes which cannot easily be modeled statistically (e.g. base substitutions and complex stutters). The numeric values for noise and stutter allele calling thresholds are presented below in *Setting interpretation thresholds for allele calling*.

3. Results & discussion

3.1. Inter-laboratory sequencing quality

Sequencing quality metrics for each laboratory are shown in Table 2. Although elevated phasing and pre-phasing values were observed in some cases, the sequencing quality parameters were considered acceptable at all participating laboratories. The proportion of reads that could be mapped to the seven STR markers increased with the amount of DNA template, which was expected as less template typically leads to more primer dimers (Supplementary figure S 1). Between laboratories, the proportions varied between 42% and 74% for similar samples (Supplementary figure S 1), which likely reflects a variability in the efficiency of the two PCR steps in library preparation or the bead purification procedure. A large proportion of off-target reads is not necessarily a problem, as deeper sequencing can be employed to provide sufficient coverage. However, less non-target amplification results in more economical and predictable sequencing, decreasing the need of reruns. The difference in on-target proportions correlates with the cluster densities in Table 2, since a large proportion of small non-target PCR products gives an underestimation of molar quantity when estimating the library concentration by fluorogenic dyes, leading to a larger volume of the library being added to the sequencing flow cell. Still, the cluster densities of all runs resulted in a satisfactory number of reads, indicating that fluorometry, a cheap and simple technique, can be used for library quantification across laboratories. Less variation in cluster density may be achievable by qPCR quantification, although at the expense of higher labor and reagent costs [30]. Quantification by qPCR may also require serial dilutions to ensure a result within the working range of the method, adding another source of variation.

Table 2
Sequencing quality metrics at each laboratory. L1 values were reported only to one or two decimals. Lab L5 used an instrument with a patterned flow cell, hence no cluster density is given. Phasing values above 0.25 and pre-phasing above 0.15 are indicated by *.

Lab	Cluster density	Clusters passing filter	Read 1 Phasing	Read 1 Pre-phasing	Read 2 Phasing	Read 2 Pre-phasing
L1	1611	87.6	0.17	0.20	0.19	0.06
L2	1037	94.60	0.209	*0.285	*0.312	*0.315
L3	1615	89.45	0.163	*0.167	0.251	0.129
L4	1430	92.83	0.197	*0.215	0.194	0.038
L5	-	83.34	0.195	0.064	0.165	0.005

3.2. Effect of UMIs on sequencing variation between laboratories

When comparing results for 1 ng of 2800 M DNA, large differences were observed between the numbers of STR reads in replicates, even within a lab (Fig. 1A). Summarizing the reads into a consensus read for each UMI sequence evened out much of the variation (Fig. 1B). Use of consensus reads decreased the read number variation both between replicates within laboratories (mean coefficient of variation (CV) without UMIs 13%, with UMIs 3%), between laboratories (CV without UMIs 23%, with UMIs 12%) and between markers (CV without UMIs 39%, with UMIs 30%). In terms of balance between the markers, despite using different batches of primers, laboratories L2, L3 and L5 showed well-balanced numbers of consensus reads (Fig. 1C-D). Notably, D8S1179 was now well represented compared to before primer rebalancing [8]. Laboratories L1 and L5, which used primers mixed at the same occasion, had an overrepresentation of D3S1358 STR reads and consensus reads, while D8S1179 was underrepresented.

3.3. Error rates following consensus read generation and ML filtration

The error rates, i.e. the proportions of consensus reads not supporting the expected genotypes, were similar for single-source samples with 250 and 1000 pg, and those were thus combined when calculating error rates (Supplementary figure S 2). Comparing laboratories, the error rates for single-source samples were in the range of 2.5 – 3.0% among those using MiSeq instruments (L1 – L4) (Fig. 2A). Lab L5, with a NextSeq instrument, obtained a somewhat higher error rate at 3.6% (99%

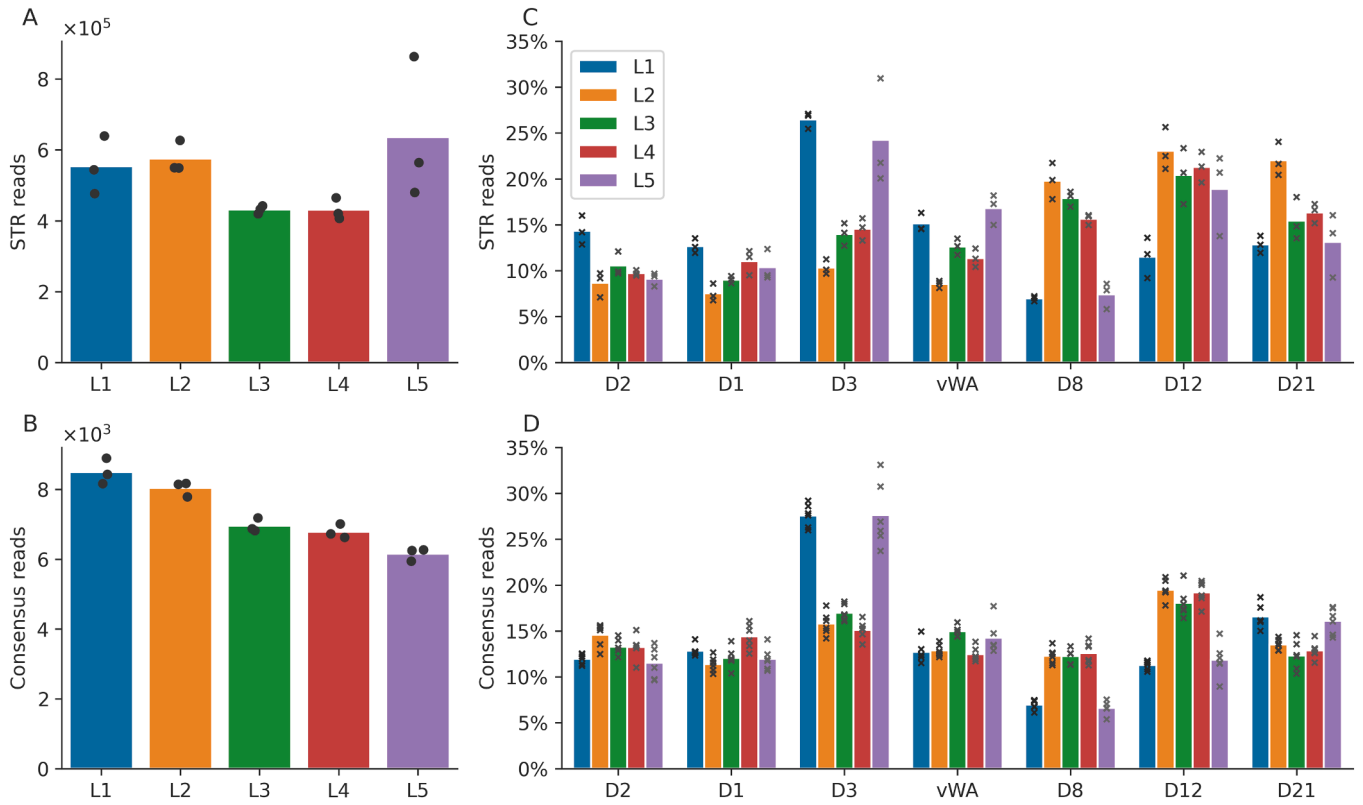

Fig. 1. Read counts for 1 ng 2800 M DNA. Locus designations according to Supplementary table S 1. A. Number of STR reads mapped to markers per reaction. B. Number of consensus reads per reaction. N.b: the y scale differs between A and B. C. Proportion of STR reads that were aligned to each marker. D. Proportion of consensus reads aligned to each marker. (n = 3 repeatability samples per lab).

confidence interval for difference: 0.3 – 1.3 percentage points). The error rates for each of the seven markers were similar to what was observed previously (Fig. 2B) [8].

Applying the ML neural network model that was developed in the original forensic SiMSen-Seq study yielded major improvements in error rate at all laboratories, regardless of sequencing instrument used (Fig. 2C-D) [8]. Notably, with ML filtered consensus reads, lab L5 had an error rate of 1.8%, which was now similar to the laboratories using MiSeq with a range of 1.3 – 1.7% (99% c.i.: −0.004 – 0.7 p.p.). This shows that the ML model is robust between instrument types and laboratories, and does not need to be retrained when the method is implemented at a new laboratory.

3.4. *Applicability of SiMSen-Seq STR analysis*

Known PCR inhibitors were included in library preparation to investigate the practical applicability of the forensic SiMSen-Seq STR method. Spiking samples with humic acid and hematin did not affect the sequencing library peak pattern in electrophoresis, except at the highest hematin concentration (400 μM) where no or very low peaks were present at the expected sizes (Supplementary figure S 3). When samples with the two highest concentrations of each inhibitor were sequenced, the consensus read count was found to be unaffected at both 8 and 16 ng/μL humic acid in the DNA sample (corresponding to 4.8 and 9.6 ng/μL in the barcoding PCR reactions). With hematin, moderate inhibition was observed at 200 μM and strong inhibition at 400 μM (120 and 240 μM in the reactions, respectively; Fig. 3A). It may be deduced that 200 μM hematin only affected the initial barcoding PCR where the UMI families are generated, since the library preparation yielded the expected electrophoresis pattern and a similar number of STR reads as uninhibited samples (540 000 ± 14 000 STR reads versus 579 000 ± 37 000 for controls (mean ± sd)) but the number of consensus reads (e.g. UMI families passing quality thresholds) was reduced by 45% (Fig. 3A). It is likely that the 2.5X dilution after barcoding combined with the four-fold higher polymerase concentration could overcome the inhibitory effect in the adaptor PCR. At 400 μM hematin, less than ten consensus reads were obtained for most markers, although D3S1358 and vWA appeared to be somewhat more tolerant than the others and showed all the expected alleles with 30 – 140 consensus reads in one replicate. It is notable that although the SiMSen-Seq protocol has not been specifically optimized for inhibitor tolerance, the method tolerated similar or up to ten times higher inhibitor concentrations than current commercially available STR sequencing kits [14,16,31].

The effect of lower sequencing depths was simulated to investigate how the number of consensus reads scales with the STR read count by randomly sampling STR reads at various ratios before marker alignment (Fig. 3B). When sequencing 32 reactions with a MiSeq standard flow cell, the read depth for uninhibited 1 ng reactions was approximately $5 \bullet 10^5$ STR reads per sample, or $7 \bullet 10^4$ STR reads per marker. It was found that the number of consensus reads decreased by a lower rate than the read depth sampling ratio. For example, decreasing the raw read depth by half only removed 16% of the consensus reads while theoretically allowing 64 reactions in a single MiSeq run.

While this protocol uses paired-end sequencing, other STR sequencing methods may sequence in the forward direction only which saves about one day in the workflow [16]. With our method, using only the forward reads for analysis still gave a major reduction of errors but resulted in a marginally higher final error rate across all consensus reads of 3.2%, compared to 2.9% for paired-ends (see Supplementary figure S 4 for per-lab and per-sample statistics). When considering implementation, a laboratory can make decisions about desirable read depth and whether paired-end sequencing should be applied based on the outcome of in-house validation and priorities. This decision should take into account the type of sequencing instrument used, types of samples

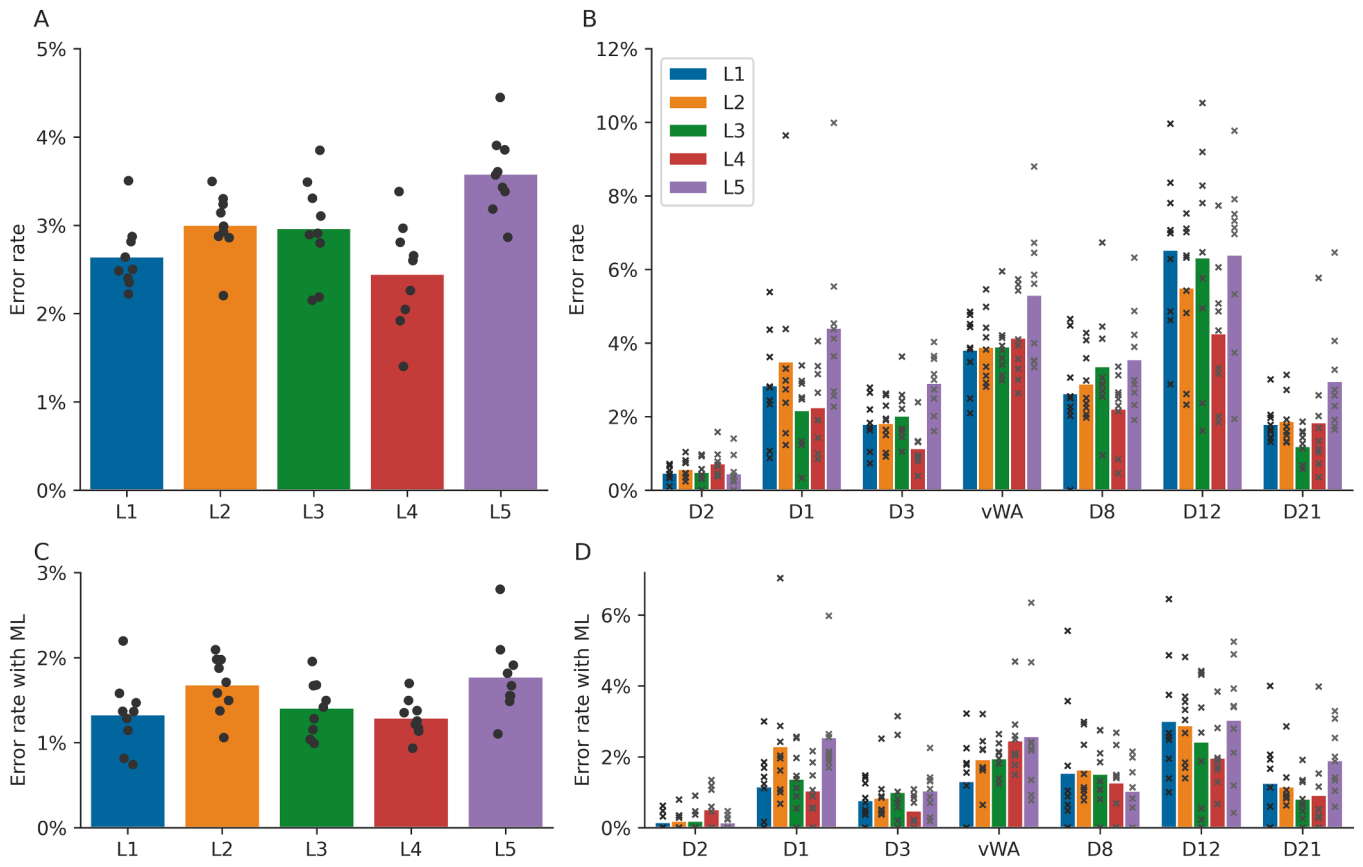

Fig. 2. Proportions of incorrect consensus reads (error rate), for single source samples with 250 or 1000 pg DNA. A. Per laboratory. B. Per marker. C. ML filtered consensus reads, per laboratory. D. ML filtered consensus reads, per marker. (n = 11 samples per lab).

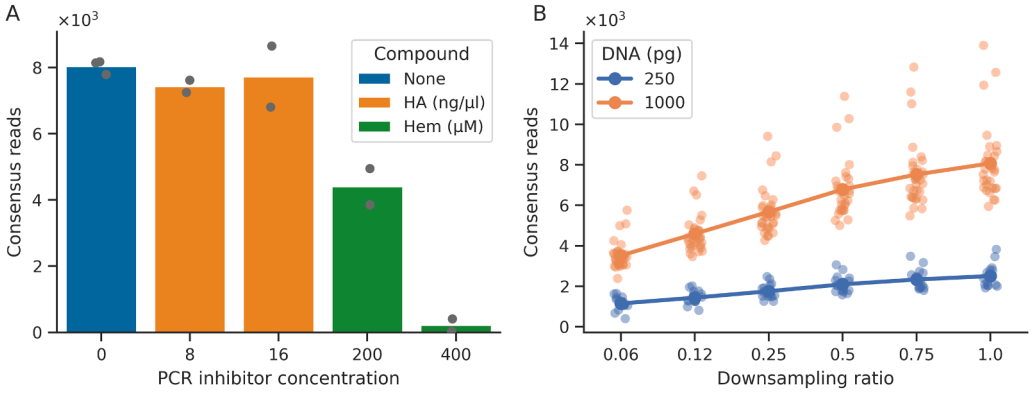

Fig. 3. Effects of PCR inhibitors and lowered sequencing depth on consensus read count. A. Humic acid (HA) and hematin (Hem) in reactions with 1 ng DNA. B. Number of consensus reads per marker when lower sequencing depth was simulated by randomly removing reads.

analyzed, the expected number of samples and any time constraints.

3.5. Setting interpretation thresholds for allele calling

For the allele calling thresholds to be as robust as possible, we wished to include the data set from the previous forensic SiMSen-Seq STR study by Sidstedt et al. [8]. Especially for the stutter rate calculations, it is desirable to include the variety of profiles in the earlier data set due to the dependence of stutter rate on allele length. The data set of Sidstedt et al. was acquired at one of the laboratories participating in the present study, using the library preparation and sequencing protocols described here, except for the two changes outlined in the Methods section. The total rates of noise reads were equivalent between the current and the previous study (0.35% versus 0.38% [8]), and the noise was similarly distributed with respect to the number of consensus reads observed for each noise sequence (Supplementary figure S 5 A). The observed stutter rates were also similar between the studies for all markers (Supplementary figure S 5B). Based on these observations, we expected the estimated thresholds to be valid for both datasets and decided to combine them to provide more robust estimates. Fixed thresholds for UMI family acceptance were used for estimation of interpretation thresholds and allele calling. ML filtering was not used here, as the non-transparent nature of the ML model makes the behavior of the analysis pipeline more difficult to interpret and understand in depth.

Results from a total of 94 single-source samples in both data sets were used to calculate stutter ratio thresholds. In total, 2.3% of all

consensus reads were minus or plus one stutter artifacts. The mean ratio of stutter to main allele reads for each marker was in the range of 0.3—3.3% for minus one stutter and 0.1—0.4% for plus one stutter (Supplementary table S 2). The ratios for individual observations are shown in Fig. 4A-B. Stuttering of multiple units and "zero" stutter (i.e. the simultaneous loss and gain of repeats) were rarely observed and, as they are too infrequent to be described statistically, are here treated as noise. Although stutter ratios are not normally distributed in STR sequencing [7], the mean plus two or three standard deviations are often used as thresholds for allele calling at stutter positions and are listed in Supplementary table S 2 [32].

The 0.35% of the consensus reads that were noise, i.e. erroneous sequences other than single unit minus or plus stutter, corresponded to on average 38 consensus reads per sample distributed over an average of 26 noise alleles. For 69% of the noise alleles only a single consensus read was observed, but some noise sequences with multiple consensus reads were seen in all laboratories and markers (Fig. 4C-D). The largest number of consensus reads for a sequence variant that was neither correct nor a stutter was 14. Inspection of the incorrect alleles with the highest number of consensus reads revealed that those were not random sequences, but rather related to true alleles by "zero" stuttering, minus or plus stutter by two repeat units, single-base substitutions or occasional minus stuttering in a 4-repeat stretch (Supplementary table S 3). Notably, sequences not related to the actual alleles present – true "noise" in the strict sense of the word – are very rarely observed after UMI correction. Again, the current dataset was combined with that of Sidstedt et al. to set allele calling thresholds for genotyping. The rate of incorrectly called alleles depending on the noise threshold is shown in Table 3. A threshold of five consensus reads, i.e. that five separate barcodes for a single allele are observed, resulted in a 5% probability of a false allele per marker (often termed *drop-in probability*). A threshold of five consensus reads can be compared to the mean number of consensus reads for heterozygotic alleles with 1 ng template, which was 558, suggesting a detection limit of around one percent for minor contributors, where about half of such alleles would be called.

Table 3
Numbers and proportions of called artifacts depending on noise threshold. Data combined from this work and the previous study [8]. (n = 94 samples).

Calling threshold	Called artifacts	Artifacts per sample	Artifacts per marker
14	1	0.01	0.00
11	3	0.03	0.00
10	5	0.05	0.01
9	6	0.06	0.01
8	9	0.10	0.01
7	10	0.11	0.02
6	20	0.21	0.03
5	36	0.38	0.05
4	86	0.91	0.13
3	182	1.94	0.28
2	446	4.74	0.68
1	1761	18.73	2.68

3.6. Allele calling for low template samples and mixtures

When the noise threshold of 5 consensus reads and stutter thresholds at mean plus three standard deviations were applied to single-source samples, all expected alleles were typed with 62 pg DNA template or more. With 31 pg, some dropouts were observed although allele calling was still acceptable at 93% (Fig. 5A, rightmost group). In two-person mixtures (Fig. 5A, first four groups), at least 90% of the minority

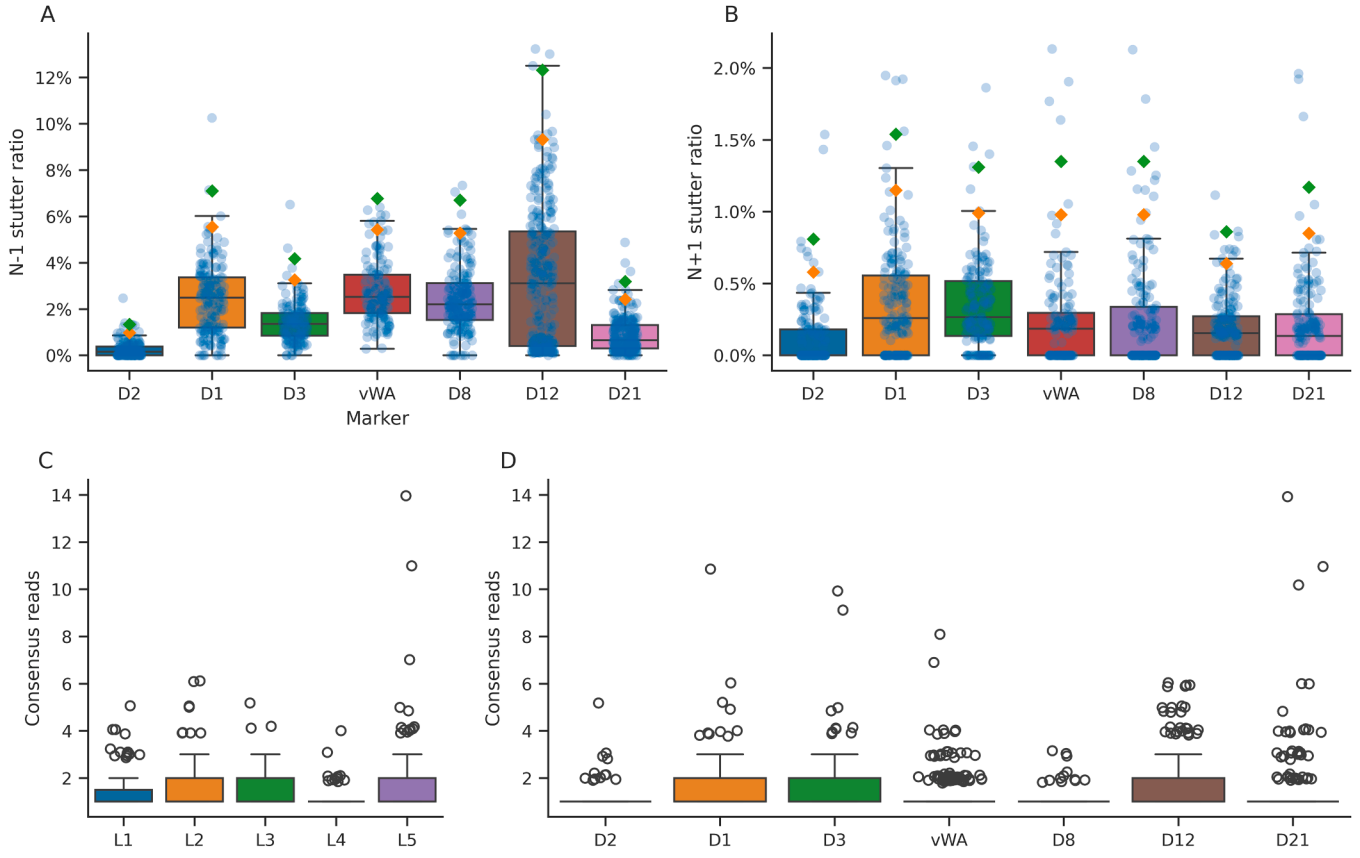

Fig. 4. Proportions of artifacts. A. Minus one stutter ratios per marker. B. Plus one stutter ratios per marker. Combined data from this work and the previous study [8]. Diamonds indicate mean plus two (orange) or three (green) standard deviations (n = 138–276 stutters per marker). C. Number of consensus reads for noise sequences, per lab. Jitter has been added to outliers. (n = 86–851 per lab) D. As C, per marker (n = 53–363 per marker).

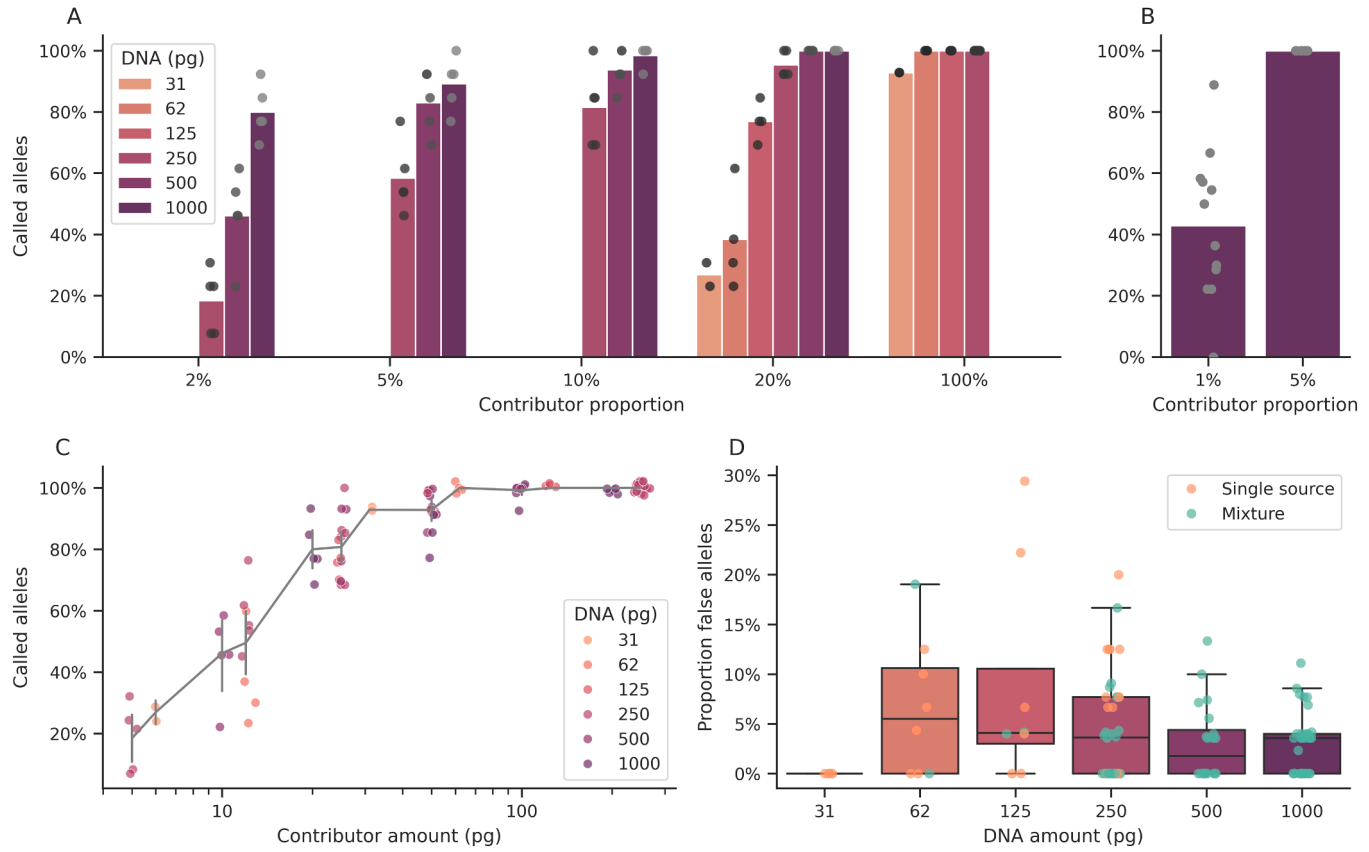

Fig. 5. Allele calling in mixtures. A. Proportion of unshared minority alleles called in 2-person mixtures, depending on total DNA amounts and minority proportions. (n = 5 per mixture) B. As A, but for 3- and 4-person mixtures with 1000 pg total DNA. C. Proportion of unshared minority alleles called in 2-person mixtures, depending on contributor amount. Color indicates total DNA amount. Jitter has been added to points for readability. Error bars indicate 95% confidence intervals. D. Proportion of false alleles (drop-in and stutter) to total number of called alleles (n = 2 261 typed alleles in total).

alleles were obtained whenever the contribution corresponded to 100 pg DNA or more, regardless of the total DNA amount (500 or 1000 pg, i.e. fourth or third groups). For contributors with 10–12 pg DNA, 48 ± 15% of unshared alleles (mean ± sd) passed the allele calling threshold (Fig. 5A, first to fourth groups). In the more complex mixtures with 1 ng DNA shown in Fig. 5B, full profiles for all 50 pg contributors were detected, compared to 42 ± 24% of the 10 pg contributor alleles. The latter result is similar to the theoretical expectation in the previous section, as approximately half of the 1% contributor alleles are called at this read depth.

When the calling of minority contributor alleles was compared to the DNA amount of the contributor, rather than the contributor fraction, the calling rate increased monotonously over a wide range of mixture proportions (Fig. 5C). This suggests that the allele calling performance depends more on the absolute contributor amount than the mixture proportion.

In total, 96.2% of the mixture alleles passing the allele calling thresholds were correct. The remaining 3.8% were artifactual; 2.4% in minus one stutter positions, 0.6% in plus one stutter positions, while the remaining 0.8% were in neither of these, and were thus considered noise. A higher proportion of incorrect alleles was observed for low DNA amounts (Fig. 5D), which was expected as the variability of stutter ratios increases with lowered read counts for the respective main alleles. When compared to, for example, the currently available ForenSeq Mainstay kit, the calling of minority alleles was more sensitive. Here, 90% of the unshared alleles were called for 50 pg DNA contributors, while with Mainstay about 50% of alleles were called for a similar sample [16].

4. Conclusions

The SiMSen-Seq STR assay was successfully applied with comparable results at five independent laboratories, using both MiSeq and NextSeq sequencing instruments. The application of UMIs decreased the variation between samples and laboratories and resulted in around 97% consensus reads supporting the correct genotypes. A slightly higher rate of erroneous consensus reads was observed for the laboratory using the NextSeq instrument, compared to the MiSeq laboratories. Differences were observed in proportions of on-target reads, which may be attributed to differences between PCR instruments or the bead purification, but these did not affect the quality of the sequencing results. Further improvement of the error rate was achieved using a machine learning model, which notably gave similar results on data from all laboratories and thus did not need to be retrained when circumstances such as primer batch or instrumentation changes. The method is relatively tolerant to model PCR inhibitors and is therefore expected to yield results also for typical casework samples. Although the current inhibitor tolerance is on par with or substantially better than current commercial MPS STR kits, there may be further room for improvement since no specific optimization to this end has yet been performed. Methods to improve inhibitor tolerance have previously been suggested [33–35] and could possibly be applied to the SiMSen-Seq method as well.

The primary purpose of a STR sequencing method with UMIs is to achieve higher sensitivity for minor contributor alleles in mixtures compared to present STR analysis methods. Here, we see that with reasonable allele calling thresholds, higher sensitivity for minor contributors in the 2–5% range can be expected compared to a current STR sequencing kit without UMIs [16]. In this work, simplistic thresholds for stutter and noise are used to demonstrate the performance of the

method. In forensic casework, it is likely preferable to model stutter ratios either during allele calling or within a probabilistic genotyping calculation [36–38]. The output score of an ML model can also be used as input for allele calling, instead of applying fixed family acceptance and allele calling thresholds [39].

The current multiplex is a proof-of-concept method including seven loci with different amplicon lengths and error profiles, representing the variation between the loci included in modern forensic STR assays. As SiMSen-Seq performed well with all of these, with UMIs yielding the largest improvements with the otherwise most error-prone markers, the method shows great promise for being expanded up to the 23–28 autosomal STR markers typical of modern forensic STR multiplexes. While methodologies for scaling of multiplexes have been described previously [40,41], care must be taken to minimize off-target amplification and to achieve a sufficiently balanced amplification to avoid markers dropping out in minor contributors. The ML model, if used, will also have to be retrained when new loci are added.

As of writing, STR sequencing multiplexes have been available for a decade but have only captured a miniscule proportion of the forensic DNA analysis market [2]. We show that by applying UMIs, the SiMSen-Seq method is a feasible path to improve the value added by STR sequencing, possibly improving the investigation of severe crime.

CRediT authorship contribution statement

Arvid H. Gynnå: Writing – original draft, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Maja Sidstedt:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Kevin M. Kiesler:** Investigation. **Adam Staadig:** Investigation. **Åsa Lindgren:** Investigation. **Firaol Tamiru Kebede:** Investigation. **Joakim Håkansson:** Resources, Funding acquisition. **Carolyn R. Steffen:** Investigation. **Tobias Österlund:** Software. **Yalda Bogestål:** Resources, Funding acquisition. **Andreas Tillmar:** Supervision. **Peter M. Vallone:** Supervision, Resources. **Anders Ståhlberg:** Supervision, Funding acquisition. **Johannes Hedman:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Ethics statement

All work performed at NIST has been reviewed and approved by the NIST Human Research Protections Office (MML−16–0080). All DNA samples used in this study were acquired as standard reference materials or commercial DNA material and were anonymous to the researchers. Hence, this study falls outside the scope of the Swedish Ethical Review Act and an ethical permit was not required for the work performed in Sweden.

Disclaimer

Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Commerce. Certain commercial software, equipment, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by NIST, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

Declaration of Competing Interest

Anders Ståhlberg is co-inventor of SiMSen-Seq that is patent protected (U.S. Serial No.:15/552,618). A. Ståhlberg declares stock ownership in Iscaff Pharma, SiMSen Diagnostics and Tulebovaasta. A. Ståhlberg is board member in Tulebovaasta.

Acknowledgements

This study was funded by Vinnova grant number 2020–01141 “Ultrakänsliga analyser för bättre hälsa och kriminalteknik (ULTRA-UDI)”. Anders Ståhlberg is funded by The Swedish Agency for Economic and Regional Growth (20370391), The Swedish Research Council (2021–1008), the Swedish state under the agreement between the Swedish government and the county councils, the ALF-agreement (1006009) and Region Västra Götaland Sweden.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fsigen.2026.103548.

References

[1] S. Iozzi, et al., Forensic genetics in NGS era: New frontiers for massively parallel typing, 12//, Forensic Science International Genetics Supplement Series 5 (2015) e418–e419, https://doi.org/10.1016/j.fsigss.2015.09.166.
[2] P. de Knijff, From next generation sequencing to now generation sequencing in forensics, Forensic Science International Genetics 38 (2019) 175–180, https://doi.org/10.1016/j.fsigen.2018.10.017.
[3] D. Ballard, J. Winkler-Galicki, J. Wesoły, Massive parallel sequencing in forensics: advantages, issues, technicalities, and prospects, 2020/07/01, Int. J. Leg. Med. 134 (4) (2020) 1291–1303, https://doi.org/10.1007/s00414-020-02294-0.
[4] C.C.G. Benschop, et al., Application of a probabilistic genotyping software to MPS mixture STR data is supported by similar trends in LRs compared with CE data, Forensic Science International Genetics 52 (2021) 102489, https://doi.org/10.1016/j.fsigen.2021.102489.
[5] M. Sidstedt, A.H. Gynnå, J. Hedman, R. Hedell, A comparsion of likelihood ratios between STR sequencing and capillary electrophoresis for complex mixtures, in: J. M. Butler (Ed.), 30th Congress of the International Society for Forensic Genetics, 2024, Universidade de Santiago de Compostela, Santiago de Compostela, 2025, pp. 453–459, https://doi.org/10.15304/cc.2025.1869.
[6] S.B. Vilsen, et al., Stutter analysis of complex STR MPS data, Forensic Sci. Int. Genet. 35 (2018) 107–112, https://doi.org/10.1016/j.fsigen.2018.04.003.
[7] M.M. Agudo, H. Aanes, A. Roseth, M. Albert, P. Gill, Ø. Bleka, A comprehensive characterization of MPS-STR stutter artefacts, 2022/09/01/, Forensic Science International Genetics 60 (2022) 102728, https://doi.org/10.1016/j.fsigen.2022.102728.
[8] M. Sidstedt, et al., Ultrasensitive sequencing of STR markers utilizing unique molecular identifiers and the SiMSen-Seq method, Forensic Science International Genetics 71 (2024), https://doi.org/10.1016/j.fsigen.2024.103047.
[9] J.M. Butler, H. Iyer, R. Press, M.K. Taylor, P.M. Vallone, and S. Willis, DNA Mixture Interpretation: A NIST Scientific Foundation Review, 2024.
[10] S. Filges, E. Yamada, A. Ståhlberg, T.E. Godfrey, Impact of Polymerase Fidelity on Background Error Rates in Next-Generation Sequencing with Unique Molecular Identifiers/Barcodes, 2019/03/05, Sci. Rep. 9 (1) (2019) 3503, https://doi.org/10.1038/s41598-019-39762-6.
[11] S.B. Seo, J. Ge, J.L. King, B. Budowle, Reduction of stutter ratios in short tandem repeat loci typing of low copy number DNA samples, 2014/01/01/, Forensic Science International Genetics 8 (1) (2014) 213–218, https://doi.org/10.1016/j.fsigen.2013.10.004.
[12] E. Yamanoi, M. Sakurada, Y. Ueno, Low stutter ratio by SuperFi polymerase, 2021/07/01/, Forensic Science International Reports 3 (2021) 100201, https://doi.org/10.1016/j.fsir.2021.100201.
[13] N.A. Courtney, et al., Engineered dna polymerase with reduced artifact formation, US18/595,339, U. S. Pat. Appl. (2024). US18/595,339.
[14] E.A.C. de Jong, M.H.J. Arts, K.J. van der Gaag, P.A.M. van Oers, J.P.G. Theelen, Developmental validation of the IDseek OmniSTR global autosomal STR profiling kit, Forensic Science International Genetics 76 (2025), https://doi.org/10.1016/j.fsigen.2025.103226.
[15] A.C. Jäger, et al., Developmental validation of the MiSeq FGx Forensic Genomics System for targeted next generation sequencing in forensic DNA casework and database laboratories, 5//, Forensic Science International Genetics 28 (2017) 52–70, https://doi.org/10.1016/j.fsigen.2017.01.011.
[16] K.M. Stephens, et al., Developmental validation of the ForenSeq MainstAY kit, MiSeq FGx sequencing system and ForenSeq Universal Analysis Software, Forensic Science International Genetics 64 (2023), https://doi.org/10.1016/j.fsigen.2023.102851.
[17] K. Elwick, P. Rydzak, J.M. Robertson, Evaluation of Library Preparation Workflows and Applications to Different Sample Types Using the PowerSeq® 46GY System with Massively Parallel Sequencing, Genes 14 (5) (2023) 977. ⟨https://www.mdpi.com/2073-4425/14/5/977⟩ ([Online]. Available).
[18] V. Sharma, E. Wurmbach, Systematic evaluation of the Precision ID GlobalFiler™ NGS STR panel v2 using single-source samples of various quantity and quality and mixed DNA samples, 2024/03/01/, Forensic Science International Genetics 69 (2024) 102995, https://doi.org/10.1016/j.fsigen.2023.102995.

[19] M.M. Agudo, et al., A comparison of likelihood ratios calculated from surface DNA mixtures using MPS and CE Technologies, Forensic Science International Genetics 73 (2024), https://doi.org/10.1016/j.fsigen.2024.103111.
[20] Q. Peng, R. Vijaya Satya, M. Lewis, P. Randad, Y. Wang, Reducing amplification artifacts in high multiplex amplicon sequencing by using molecular barcodes, 2015/08/07, BMC Genom. 16 (1) (2015) 589, https://doi.org/10.1186/s12864-015-1806-8.
[21] T. Kivioja, et al., Counting absolute numbers of molecules using unique molecular identifiers, 2012/01/01, Nat. Methods 9 (1) (2012) 72–74, https://doi.org/10.1038/nmeth.1778.
[22] S. Filges, P. Mouhanna, A. Ståhlberg, Digital Quantification of Chemical Oligonucleotide Synthesis Errors, Clin. Chem. 67 (10) (2021) 1384–1394, https://doi.org/10.1093/clinchem/hvab136.
[23] A. Staadig, J. Hedman, and A. Tillmar, Applying Unique Molecular Indices with an Extensive All-in-One Forensic SNP Panel for Improved Genotype Accuracy and Sensitivity, *Genes*, vol. 14, no. 4, doi: 10.3390/genes14040818.
[24] Y.L. Kwon, K.J. Shin, Unique molecular identifier-based amplicon sequencing of microhaplotypes for background noise mitigation, 2024/09/01/, Forensic Science International Genetics 72 (2024) 103096, https://doi.org/10.1016/j.fsigen.2024.103096.
[25] Q. Zhu, et al., Investigation into the genotyping performance of a unique molecular identifier based microhaplotypes MPS panel in complex DNA mixture, 2025/03/01/, Forensic Science International Genetics 76 (2025) 103236, https://doi.org/10.1016/j.fsigen.2025.103236.
[26] A.E. Woerner, S. Mandape, J.L. King, M. Muenzler, B. Crysup, B. Budowle, Reducing noise and stutter in short tandem repeat loci with unique molecular identifiers, Forensic Science International Genetics 51 (2021), https://doi.org/10.1016/j.fsigen.2020.102459.
[27] A. Ståhlberg, P.M. Krzyzanowski, J.B. Jackson, M. Egyud, L. Stein, T.E. Godfrey, Simple, multiplexed, PCR-based barcoding of DNA enables sensitive mutation detection in liquid biopsies using sequencing, e105-e105, Nucleic Acids Res. 44 (11) (2016), https://doi.org/10.1093/nar/gkw224.
[28] C.R. Steffen, E.L. Romsos, K.M. Kiesler, L.A. Borsuk, K.B. Gettings, P.M. Vallone, Make it "SNPPY" - Updates to SRM 2391d: PCR-Based DNA Profiling Standard, 2022/12/01/, Forensic Science International Genetics Supplement Series 8 (2022) 9–11, https://doi.org/10.1016/j.fsigss.2022.09.004.
[29] K.B. Gettings. *Forensic DNA Open Dataset*, National Institute of Standards and Technology, doi: 10.18434/M32157.
[30] C. Hussing, M.L. Kampmann, H.S. Mogensen, C. Børsting, N. Morling, Comparison of techniques for quantification of next-generation sequencing libraries, Forensic Science International Genetics Supplement Series 5 (2015) e276–e278, https://doi.org/10.1016/j.fsigss.2015.09.110.
[31] M. Sidstedt, C.R. Steffen, K.M. Kiesler, P.M. Vallone, P. Rådström, J. Hedman, The impact of common PCR inhibitors on forensic MPS analysis, Forensic Science International Genetics 40 (2019) 182–191, https://doi.org/10.1016/j.fsigen.2019.03.001.
[32] *Establishing robust thresholds and filters for the ForenSeq MainstAY Kit.* (2021). Verogen.
[33] P. Rådström, C. Löfström, M. Lövenklev, R. Knutsson, P. Wolffs, Strategies for Overcoming PCR Inhibition, in: Dieffenbach, Dveksler (Eds.), PCR primer, *2nd edition*, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY, USA, 2003, p. 12 (ch).
[34] M. Sidstedt, P. Rådström, J. Hedman, PCR inhibition in qPCR, dPCR and MPS—mechanisms and solutions, 2020/04/01, Anal. Bioanal. Chem. 412 (9) (2020) 2009–2023, https://doi.org/10.1007/s00216-020-02490-2.
[35] C.A. Kreader, Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein, 1996/03/01, Appl. Environ. Microbiol. 62 (3) (1996) 1102–1106, https://doi.org/10.1128/aem.62.3.1102-1106.1996.
[36] Ø. Bleka, R. Just, M.M. Agudo, P. Gill, MPSproto: An extension of EuroForMix to evaluate MPS-STR mixtures, Forensic Science International Genetics 61 (2022), https://doi.org/10.1016/j.fsigen.2022.102781.
[37] Ø. Bleka, G. Storvik, P. Gill, EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts, 2016/03/01/, Forensic Science International Genetics 21 (2016) 35–44, https://doi.org/10.1016/j.fsigen.2015.11.008.
[38] J. Hoogenboom, K.J. van der Gaag, R.H. de Leeuw, T. Sijen, P. de Knijff, J.F. J. Laros, FDSTools: A software package for analysis of massively parallel sequencing data with the ability to recognise and correct STR stutter and other PCR or sequencing noise, Forensic Science International Genetics 27 (2017) 27–40, https://doi.org/10.1016/j.fsigen.2016.11.007.
[39] B. Crysup, S. Mandape, J.L. King, M. Muenzler, K.B. Kapema, and A.E. Woerner, Using Unique Molecular Identifiers to Improve Allele Calling in Low-Template Mixtures, *Forensic Science International: Genetics*, doi: 10.1016/j.fsigen.2022.102807.
[40] J.M. Butler, Constructing STR Multiplex Assays, in: A. Carracedo (Ed.), Forensic DNA Typing Protocols, Humana Press, Totowa, NJ, 2005, pp. 53–65.
[41] A. Ståhlberg, P.M. Krzyzanowski, M. Egyud, S. Filges, L. Stein, T.E. Godfrey, Simple multiplexed PCR-based barcoding of DNA for ultrasensitive mutation detection by next-generation sequencing, 2017/04/01, Nat. Protoc. 12 (4) (2017) 664–682, https://doi.org/10.1038/nprot.2017.006.