

TECHNICAL NOTE

Criminalistics

Design and optimization of a 16S microbial qPCR multiplex for the presumptive identification of feces, saliva, vaginal and menstrual secretions

Carolyn Lewis PhD^{1,2}  | Sarah J. Seashols-Williams PhD² 

¹Integrative Life Sciences Doctoral Program, Virginia Commonwealth University, Richmond, Virginia, USA

²Department of Forensic Science, Virginia Commonwealth University, Richmond, Virginia, USA

Correspondence

Carolyn Lewis, PhD, Integrative Life Sciences Doctoral Program, Virginia Commonwealth University, Box 842030, 1000 West Cary Street, Richmond, VA 23284-2030, USA.

Email: lewisca4@alumni.vcu.edu

Funding information

This project was supported by a Graduate Research Fellowship (Award No. 2020-R2-CX-0031) awarded to CAL by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice.

Abstract

Molecular methods for body fluid identification have been extensively researched in the forensic community over the last decade, mostly focusing on RNA-based methods. Microbial DNA analysis has long been used for forensic applications, such as postmortem interval estimations, but only recently has it been applied to body fluid identification. High-throughput sequencing of the 16S ribosomal RNA gene by previous research groups revealed that microbial signatures and abundances vary across human body fluids at the genus and/or species taxonomic level. Since quantitative PCR is still the current technique used in forensic DNA analysis, the purpose of this study was to design a qPCR multiplex targeting the 16S gene of *Bacteroides uniformis*, *Streptococcus salivarius*, and *Lactobacillus crispatus* that can distinguish between feces, saliva, and vaginal/menstrual secretions, respectively. Primers and probes were designed at the species level because these bacteria are highly abundant within their respective fluid. The validated 16S triplex was evaluated in DNA extracts from thirty donors of each body fluid. A classification regression tree model resulted in 96.5% classification accuracy of the population data, which demonstrates the ability of this 16S triplex to presumptively identify these fluids with high confidence at the quantification step of the forensic workflow using minimal input volume of DNA extracted from evidentiary samples.

KEYWORDS

16S ribosomal RNA gene, *Bacteroides uniformis*, classification regression tree model, forensic serology, *Lactobacillus crispatus*, presumptive body fluid identification, qPCR multiplex validation, quantitative polymerase chain reaction and qPCR, *Streptococcus salivarius*

Highlights

- A microbial qPCR multiplex was developed to classify forensically relevant body fluids.
- Three microbial species are used to distinguish feces, saliva, and vaginal/menstrual secretions.
- A classification regression tree model resulted in a 96.5% classification accuracy.
- The method could presumptively detect body fluids during DNA quantification.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Journal of Forensic Sciences* published by Wiley Periodicals LLC on behalf of American Academy of Forensic Sciences.

1 | INTRODUCTION

Body fluid identification (BFID) is the first step of the forensic DNA analysis workflow that can play a crucial role in story corroborations of suspects, victims, and/or witnesses. It can be useful for investigative leads and/or crime scene reconstruction. Equally important, it allows a DNA analyst to determine the best location to swab or cut to obtain a DNA profile from an evidentiary sample [1,2]. Most methods currently used in forensic serology rely on enzymatic-based tests that result in a color change that is interpreted and recorded by an analyst. Although these serological methods have been utilized for decades, there are well-documented flaws associated with each [3–6]. Therefore, there has been extensive research among the forensic community that addresses molecular-based BFID methods, such as messenger RNA or microRNA analysis, DNA methylation, and microbial DNA analysis [7–11].

There are numerous studies that focus on microbiome-based BFID, but mostly rely on 16S rRNA gene sequencing, as this is a gold standard in microbiome analysis [12,13]. The 16S ribosomal subunit is specific to prokaryotes and is highly conserved within the same genus and species but has several hypervariable regions, all of which allow for taxonomical microbial classification [12]. An important consideration for forensic applications is the high cost and complex workflow of high-throughput sequencing (HTS) methods. As of now, HTS is not standard in forensic DNA analysis due to cost, sample preparation time, hands-on training requirements, and complicated back-end bioinformatic analyses. To better align with the current DNA analysis workflow, real-time (q)PCR methods for bacterial BFID have been proposed; however, most of these amplify multiple bacterial species to classify a single body fluid [14–16]. This could be problematic when developing a qPCR multiplex that can identify multiple body fluids due to the number of optical filters in qPCR instruments limiting how many sequences can be multiplexed in a single well. This approach would require the use of multiple reactions or wells, which would increase reagent cost and sample consumption compared with a single-well assay.

The most successful microbiome-based BFID studies focus on body fluids with high bacterial content, such as vaginal fluid, feces, and saliva [13,15–22]. One reported limitation is the inability to differentiate menstrual blood from vaginal secretions using microbial signatures, which could be useful information in a sexual assault case [21]. A major disadvantage is that some forensically relevant

body fluids, particularly blood and semen, are difficult to characterize using microbial signatures because low bacterial cell counts often result in poor DNA yields, a problem when considering the often-compromised nature of forensic evidence [13]. Several reports address these concerns by incorporating other types of molecular markers, such as messenger RNA; however, multiplexing challenges still apply to these integrated assays [17,18,23].

The purpose of this research was to design a qPCR multiplex targeting the 16S gene of three microbial species that are highly abundant in the respective body fluids—*Bacteroides uniformis* for feces, *Streptococcus salivarius* for saliva, and *Lactobacillus crispatus* for female intimate samples (vaginal fluid/menstrual blood). The main objective was to provide proof of concept that a single-well microbial qPCR assay can presumptively identify more than one forensically relevant body fluid; therefore, differentiating between female intimate samples and identifying blood and semen were not primary goals of this study.

2 | MATERIALS AND METHODS

2.1 | Primer and probe design

Primers and probes were designed using default parameters in Beacon Designer 8 (Premier Biosoft, Palo Alto, CA). The same primer sequences from the SYBR Green mode were used to design the probes in TaqMan mode, which allowed for primer specificity testing before ordering/testing the hydrolysis probe. Primers and probes were ordered from Integrated DNA Technologies (IDT, Coralville, IA). Dual-labeled hydrolysis probes were labeled with internal quenchers, HPLC-purified and normalized to 100 μ M in TE buffer. Sequence information for qPCR primers, probes, and respective targets are listed in Table 1.

2.2 | Sample collection and DNA extraction

Venous blood, menstrual blood, semen, saliva, feces, and vaginal samples were collected from 30 volunteers according to VCU's Institutional Review Board–approved protocol for human subjects' research (HM20002931). Menstrual blood, feces, saliva, and vaginal secretions were collected onto polyester swabs, while semen was

TABLE 1 Microbial target and sequence information utilized in the 16S qPCR triplex. HPLC-purified hydrolysis probes include 5' reporters, internal, and 3' quenchers

Microbial Target [Accession No.]	Forward Primer 5'-3'	Reverse Primer 5'-3'	Hydrolysis Probe 5'-3'	Amplicon Length (bp)
<i>Lactobacillus crispatus</i> [MN744551]	CAGCAGTAGGAATCTTC	CTGTTGATTACCGTCAA	/ATTO550N/ACCTCTATC/TAO/ CTCTTACCAACAACA/IAbRQSp/	145
<i>Bacteroides uniformis</i> [AP019724.1]	TAGCGGTGAAATGCTTAG	CATCGTTTACTGTGTGGA	/6FAM/CACGAAGAA/ZEN/ CTCCGATTGCGAAG/IAbkFQ/	136
<i>Streptococcus salivarius</i> [CP015282.1]	ATGCGTAGATATATGGAGG	CTACCAGGGTATCTAATCC	/SUN/CGAGCCTCA/ZEN/ GCGTCAGTTACA/IAbkFQ/	108

collected in sterile plastic containers and aliquoted onto polyester swabs (50 μ l). Venous blood was deposited onto polyester swabs by pricking the donor's sterilized finger with a finger lancet. Swabs were placed in swab boxes and dried in at room temperature for at least 24 h prior to DNA extraction. The donations were collected between 2017 and 2019 for a biological sample registry, so they had been stored at room temperature for a comparable amount of time as they might in a forensic evidence room. Other considerations when selecting samples from the registry were antibiotic usage, equal number of male/female donors (when applicable), various ethnicities and ages, different days of the menstrual cycle, and at least seven-day postcoital for intimate fluids (semen, vaginal fluid, and menstrual blood).

Whole swabs were extracted using the QIAamp® DNA Investigator Kit (Qiagen, Valencia, CA) following the "Isolation of Total DNA from Body Fluid Stains" protocol from the manufacturer. Reagent blanks were included in each batch of extractions to assess potential contamination. During cell lysis, no carrier RNA was added, and 20 μ l of 1 M DTT was added to semen samples. After the 10-min incubation at 70°C, residual liquid was collected by placing the swab in a DNA IQ™ spin basket (Promega, Madison, WI), returning it to the tube containing the lysate, and centrifuging at 3350 g for 5 min. Once the spin basket and swab were discarded, DNA purification was performed according to protocol without modification. All samples and reagent blanks were eluted in 50 μ l of ATE buffer (Qiagen), and total DNA was quantified using the NanoDrop™2000 (Thermo Fisher Scientific) before storage at -80°C (data not shown).

2.3 | qPCR

The designed multiplex is technically a qualitative-PCR assay since the goal is not to quantify any microbial DNA in the sample, which would require analyzing known quantity standards alongside the questioned samples. However, synthetic DNA standards were used to validate the multiplex and could be used as positive controls or for quantification, if desired; therefore, the MIQE guidelines for qPCR reports were followed throughout the project [24].

Initially, each microbial target was evaluated as a single-plex using gBlocks® Gene Fragments (IDT) as the qPCR standards. The gBlocks® were chosen as standards because there were no commercially available mock microbial community standards that included all three bacterial species of interest and because gBlocks® are manufactured (i.e., undergo extra quality control measures) for the purpose of validating qPCR assays. Sequences for each gBlock® standard are listed in Table S1. Each gBlock® was resuspended in TE buffer at 10 ng/ μ l and incubated at 50°C for 20 min, per manufacturer's instructions. Ten-fold serial dilutions were prepared at the concentration range of 5 pg/ μ l–0.05 fg/ μ l, aliquoted, and stored at -20°C. To validate the multiplex, the three gBlocks® sequences were pooled together at a concentration of 1 ng/ μ l before making ten-fold serial dilutions to the optimized concentration range (5 pg/ μ l–0.05 fg/ μ l).

2.4 | SYBR green assay

To ensure amplification of a single product, each primer set was evaluated first via a SYBR Green assay with melt curve analysis to ensure a single amplified product. Standards and no-template controls were analyzed in triplicate using 6.25 μ l of PerfeCTa SYBR Green SuperMix (2X) (VWR, Radnor, PA), 3.75 μ l of nuclease-free water, 0.25 μ l (10 μ M) of forward and reverse primers (Table 1), and 2 μ l of standard for a 12.5 μ l total reaction volume. Thermal cycling was conducted on the Applied Biosystems™ QuantStudio 6 Flex Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA) with PCR parameters set to 95°C for 3 min followed by 35 cycles of 95°C for 15 s, 58°C for 45 s, and 72°C for 30 s. With instrument default melt curve parameters. Raw data were analyzed in QuantStudio Software v1.3 (Thermo Fisher Scientific) with the baseline set at 1–6 cycles and threshold set to 0.02 for all targets.

2.5 | Hydrolysis probe-based assay

Once a single amplification product was confirmed in the SYBR Green assay, each target was individually evaluated in a probe-based assay before multiplexing. After multiplex validation using gBlock® standards, the assay was tested using body fluid samples.

Each single-plex qPCR reaction consisted of 5 μ l of PrimeTime® Gene Expression Master Mix (2X) (IDT), 2.5 μ l of nuclease-free water, 0.5 μ l of 20X assay mix, and 2 μ l of standard for a total reaction volume of 10 μ l. The 20X assay mixes for all targets were optimized to final determined concentrations of 400 nM for each primer and 200 nM for the probe. For the triplex reactions, 5 μ l of PrimeTime® Gene Expression Master Mix (2X), 1.5 μ l of nuclease-free water, 0.5 μ l of *L. crispatus* 20X assay mix, 0.5 μ l of *B. uniformis* 20X assay mix, 0.5 μ l of *S. salivarius* 20X assay mix, and 2 μ l of sample or standard were combined for a 10 μ l of total reaction volume. Thermal cycling parameters on the QuantStudio 6 Flex Real-Time PCR instrument were set to 95°C for 3 min followed by 40 cycles of 95°C for 5 s and 60°C for 30 s in fast cycling mode. All six standards were analyzed in duplicate with no-template controls on each plate. To conserve DNA extract, single technical replicates were run when testing body fluids; however, various extracts were tested multiple times throughout the validation to ensure repeatable results from body fluid samples. In QuantStudio Software v1.3, autobaseline and cycle threshold of 0.06 were applied initially during assay optimization, but the threshold was lowered to 0.04 for all targets during final validation and when evaluating body fluid specificity.

2.6 | Data analysis

During assay optimization, standard curve metrics such as slope, Y-intercept, and PCR efficiency (>90%) were analyzed for all targets, and technical validation was complete when the single-plex and multiplex quantification cycle (Cq) values for each standard

were within one cycle of one another. Raw data were exported from QuantStudio Software v1.3 and input into Microsoft Excel to calculate averages between technical replicates, replace a value of 40 for any “undetermined” result, and omit eight outliers that were greater than three standard deviations from the mean (yellow highlight in Table S2). Subsequent data analysis was performed in R version 4.1.1 (R foundation, Vienna, Austria). The data were randomly split into training and validation sets (70% and 30%, respectively) using the `sample()` function, and a classification regression tree (CART) model was created using the `rpart` package for each iterative analysis (15 analyses were conducted). Confusion matrices and classification accuracy percentages are reported for only one iteration of the validation dataset analysis. All data are reported as averages with standard deviations, where applicable.

2.7 | Linear range of classification

The goal of the linear range of classification study was to determine the lowest DNA concentration at which the qPCR assay will accurately classify as the correct body fluid. Total DNA was quantified using the NanoDrop™2000 (Thermo Fisher Scientific) and averaged among the 30 donors of each body fluid in Microsoft Excel. Five donors of each body fluid were selected based on having DNA concentrations close to the calculated averages (blood = 4.13 ng/μl, menstrual blood = 65.19 ng/μl, semen = 22.04 ng/μl, feces = 117.16 ng/μl, vaginal fluid = 57.34 ng/μl, saliva = 19.04 ng/μl). Each DNA extract was serially diluted ten-fold until the dilution concentration was less than the lower limit of the linear dynamic range (0.05 fg/μl), and then, qPCR was performed using the validated 16S triplex methods mentioned above. Data were analyzed using the trained CART model; however, since the confusion matrix output does not show which samples are misclassifying, the analysis was performed manually in Microsoft Excel for each sample.

3 | RESULTS AND DISCUSSION

3.1 | Assay validation

In both SYBR Green and probe-based assays, there was no amplification detected in any of the negative controls, including extraction reagent blanks. Single peaks were observed in the SYBR

Green melt curve analysis, which verified that there was only one PCR product for each primer pair (data not shown). The same primer sequences were then used for the probe-based assay, in which various primer/probe concentrations and 5' reporter dyes were evaluated during optimization. The final 16S triplex primer and probe concentrations for all three microbial targets were 400 nM (forward and reverse) and 200 nM, respectively. All reported data are representative of the 16S triplex at these concentrations. The 5' reporter dyes were chosen based on having similar excitation and emission wavelengths as dyes that are in use in commercial STR multiplex kits, for example, the ATTO550 and SUN dyes (IDT) emit in the same filters as ABY and VIC (Thermo Fisher Scientific), respectively. These choices were designed to ensure that qPCR instruments would already be calibrated for the requisite probe emission spectra, which would ease implementation in forensic laboratories.

The slopes, amplification efficiencies, and R^2 values of the standard curves (Table 2) were all within the recommended ranges for qPCR assays [25]. The reported standard curve data in Table 2 are averaged across six experiments, demonstrating repeatability and reproducibility of the assay. The linear dynamic range was determined to be 5 pg/μl–0.05 fg/μl (Table S3 in Appendix S1). We acknowledge that the lower limit of this range is not the lowest limit of detection since each microbial target amplifies at approximately 28 cycles; however, lower concentrations were not tested to minimize reagent consumption. Additionally, we felt that six standards were sufficient to validate assay performance because the overall purpose is to obtain a raw Cq value rather than to quantify any microbial DNA from a sample. Although it is possible to quantify DNA of each microbial species using this assay, further research expanding the lower limit of detection would be required to quantify microbial DNA for forensic applications.

An important consideration during assay validation was equivalent amplification of the microbial DNA from forensic body fluid samples in a single-plex assay when compared to the multiplex. To address this concern, multiple DNA extracts were tested for each target species both in single-plex and multiplex reactions and verified that Cq values were within one cycle of each other (data not shown).

3.2 | Evaluation of body fluid specificity

When evaluating raw data (Table S2), *S. salivarius* was detected in all body fluid samples at higher/similar abundances (lower Cq values)

TABLE 2 Standard curve metrics across six experiments during the 16S triplex assay validation

Microbial Target	Slope	Y-intercept	R^2	% Efficiency
<i>Lactobacillus crispatus</i>	-3.525 ± 0.05	2.719 ± 0.58	0.997 ± 0.002	92.2 ± 1.78
<i>Bacteroides uniformis</i>	-3.491 ± 0.03	2.433 ± 0.48	0.998 ± 0.002	93.4 ± 1.00
<i>Streptococcus salivarius</i>	-3.524 ± 0.03	2.549 ± 0.44	0.998 ± 0.002	92.2 ± 0.85

in saliva and menstrual blood compared with vaginal fluid and feces. As expected, *B. uniformis* and *L. crispatus* were most abundant in feces and vaginal/menstrual secretions, respectively. Of the three microbial species, detection of *L. crispatus* was the most variable, especially in vaginal fluid. One possible explanation for such high variability in vaginal fluid is PCR inhibition due to over-input of template DNA (*L. crispatus* was undetected in nine vaginal samples). This is further discussed in the linear range of classification section. Another possible explanation is that only one *Lactobacillus* species was included in the assay compared with other microbial qPCR studies that target more than one species for accurate vaginal fluid identification [14,16,22]. Incorporating more markers into the assay would account for the diversity of the vaginal microbiome across females, but it also would increase reagent cost and consumption of DNA extract, which is something we wanted to avoid when designing the assay. An alternative solution could be to utilize the same approach as Doi et al. and design primers flanking a conserved region of 16S at the genus level of four *Lactobacillus* species [14].

3.3 | Classification regression tree analysis

It should first be noted that although blood and semen samples were evaluated in this study, the goal was not to differentiate between them using this assay, as they have low bacterial DNA yields [13] and were expected to have nearly undetectable values for all microbial species. For this reason, blood and semen were grouped together as “Bld/SF” in the dataset and were used to eliminate either fluid as the biological source in an unknown sample. Two regression trees were created—one with vaginal fluid (VF) and menstrual blood (MB) as separate fluids (Figure S1) and one as a combined group (VF/MB) (Figure 1). While it was not the original intent of this study to differentiate between vaginal/menstrual secretions, both were evaluated out of curiosity to see how each model would perform. The individual model resulted in an overall classification accuracy of 84% with 54.5% of MB samples misclassified. Four of the MB samples were classified as VF while two were classified as saliva (Table S4). Differentiation between VF and MB using the individual CART model was dependent on *S. salivarius* Cq value cutoff of 18.86 (Figure S1);

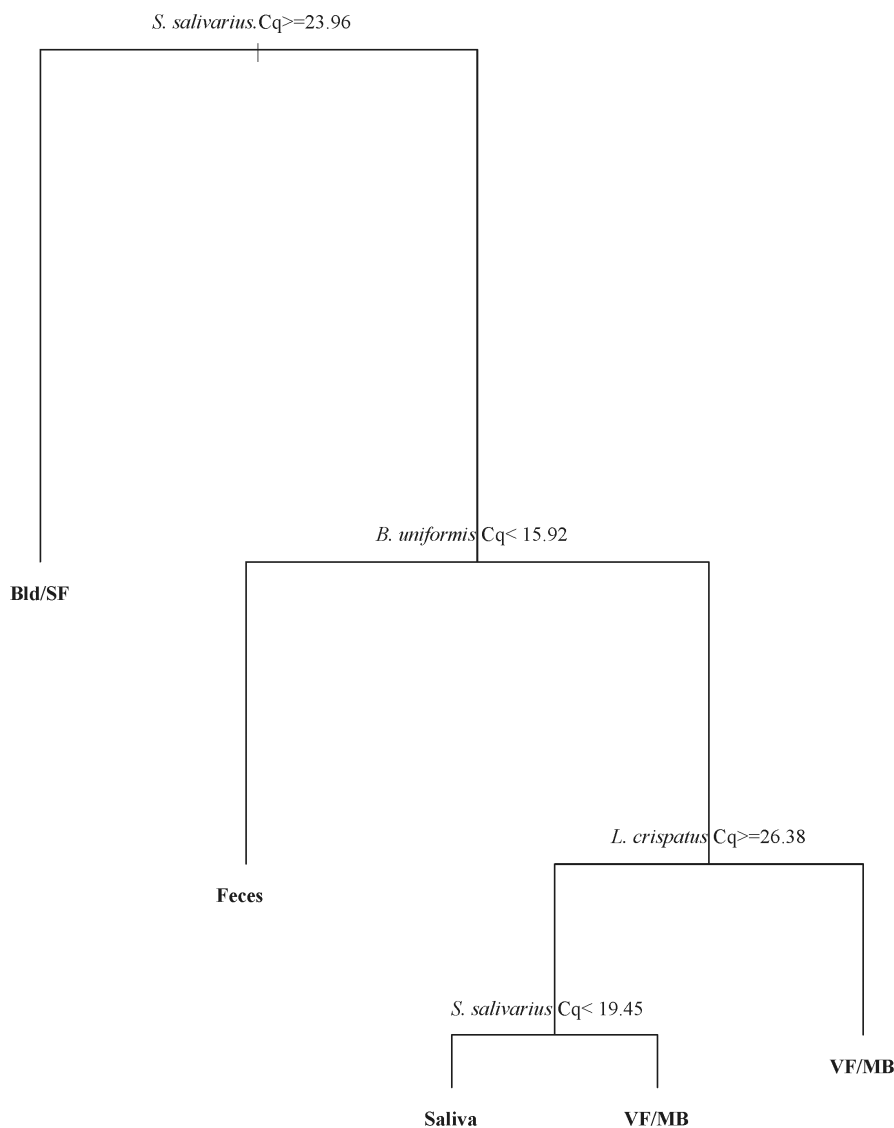


FIGURE 1 Classification Regression Tree (CART) model for classifying forensically relevant body fluids using a microbial 16S qPCR triplex when vaginal and menstrual secretions are grouped as female intimate samples ($n = 30$ donors of each body fluid, Bld/SF, blood/semenal fluid; VF/MB, vaginal fluid/menstrual blood)

however, a similar *S. salivarius* cutoff ($C_q = 19.2$) was used to distinguish between saliva and VF, which likely explains the observed misclassifications between VF, MB, and saliva (Table S4).

When female intimate sample data were combined, the tree plot (Figure 1) looked similar to that in Figure S1; however, the overall classification accuracy increased to 96.5% with only two samples (one saliva and one VF/MB) misclassified (Table 3). This demonstrated that combining female intimate samples in a dataset can increase classification accuracy. Furthermore, it supports the reported claim that it is difficult to differentiate vaginal fluid from menstrual blood using microbial signatures since both fluids originate from the same body cavity and can contain similar bacterial compositions at any point during the menstrual cycle [17,18,26]. This of course is somewhat appropriate given casework scenarios; however, as the project is still developmental in nature, additional future work will be required when classifying mixtures and distinguishing venous blood from menstrual secretions within a sample of mixed origin.

Importantly, 100% of fecal samples was correctly classified regardless of VF/MB grouping, and there were no misclassifications involving blood/semen samples. Saliva misclassifications were observed in both CART models, which could be due to higher-than-expected *S. salivarius* detection in other body fluids thus negatively impacting its anticipated saliva specificity. Another possible explanation is that *S. salivarius* primers and probe were designed at the species level, and differentiation among *Streptococcus* species in saliva has been reportedly more difficult using 16S compared with other genes [19,27]. If this is true, there could be other *Streptococcus* species amplifying, which would result in lower C_q values. This could possibly explain why *S. salivarius* was detected in all donors in every body fluid sample, whereas *L. crispatus* and *B. uniformis* were undetected in at least one donor of all body fluids (Table S2). These results are contrary to what has been reported in the literature, which state that *S. salivarius* was not detected in other body fluids; however, one group did not examine feces [28], and different methodologies were used, such as loop-mediated isothermal amplification (LAMP), reverse transcription LAMP, or direct PCR combined with immunochromatographic strip [28–30]. Importantly, none of these studies amplified 16S, which supports the previous statement that the 16S rRNA gene may not be the best target for saliva identification, especially at the species taxonomic level. Since saliva is

TABLE 3 Confusion matrix classifying body fluids using the 16S triplex in a trained Classification Regression Tree (CART) model

		Bld/SF	Feces	Saliva	VF/MB
	Bld/SF	23	0	0	0
	Feces	0	8	0	0
Actual	Saliva	0	0	6	1
	VF/MB	0	0	1	17

Note: A 96.5% overall classification rate was achieved when grouping female intimate samples together (VF/MB). Bold numbers indicate correct classifications (Bld/SF, blood/semenal fluid; VF/MB, vaginal fluid/menstrual blood) Predicted.

commonly present on crime scene evidence, incorporating a more specific saliva marker into the proposed qPCR multiplex may be useful for forensic casework implementation.

3.4 | Linear range of classification

The goal of this study was to determine at which ten-fold dilution of DNA extract the body fluid will classify correctly using the grouped female intimate CART model. All saliva samples could only be accurately classified when the DNA extract was input into the qPCR reaction (Table S5). The lowest DNA concentrations quantified via UV-spectrophotometry were observed in saliva compared with vaginal/menstrual secretions and feces (data not shown); therefore, it was expected that ten-fold dilutions of saliva extracts would not yield correct classification results. It should be noted that, unless otherwise stated, any dilution that was correctly classified was also correctly classified in the DNA extract; for example, vaginal fluid and feces were correctly classified in DNA extracts of all five donors but only in the first dilution (D1) for three donors (Table S5).

There were no fluids that classified correctly beyond the second dilution (D2), except for MB, which was still accurately classified in the second dilution for four out of five donors (Table S5). There was 1-MB sample that could only be classified in the DNA extract; however, *L. crispatus* was never detected in that individual throughout the project. This could be due to a lower abundance (or absence) of that *Lactobacillus* species in the vaginal microbiome of that particular donor, which could be a result from a clinical infection, such as bacterial vaginosis. Alternatively, that individual may naturally have decreased *L. crispatus* abundance during their menstrual cycle. These are reasons why large population sizes of various demographics should be evaluated for any microbial assay, so that intraperson and interperson variability can be assessed and so that false positive/negative rates can be described.

An interesting observation in this study was that in two vaginal samples, *L. crispatus* was undetected in the DNA extract yet was detected in subsequent dilutions, which supports the previous hypothesis that inputting too much DNA template into the reaction may lead to PCR inhibition. This only affected classification accuracy in one of the VF samples, where classification was only achieved in the first dilution but not in the DNA extract. Of note, the proposed CART analysis method evaluates raw C_q values, so it is expected that dilution of a DNA extract will equally dilute all DNA in the sample; therefore, a different analysis method or data normalization may be useful for compromised evidence samples.

4 | CONCLUSIONS

In this study, a microbial 16S qPCR multiplex was proposed, which can correctly classify vaginal fluid, menstrual blood, feces, and saliva with 84% overall accuracy through classification regression tree analysis. When female intimate samples were grouped together in

the CART model, overall accuracy increased to 96.5%. All observed misclassifications involved saliva samples, which indicates that different saliva-specific markers or redesign of the *Streptococcus* primers may increase the assay's ability to correctly classify saliva. Importantly, high accuracy was achieved using forensically relevant dried samples of appropriate volumes that had been stored at room temperature for an extended period of time, which suggest that bacterial signatures remain stable enough to characterize dried body fluids via qPCR. The assay was developed for use on a qPCR instrument with five optical filters, and only four are used in the multiplex (three for microbial targets and one for passive reference dye); therefore, it is possible to include one more primer set and remain a single-well assay. An internal positive control should be considered whether synthetic standards are not run alongside unknown samples. Alternatively, the incorporation of a human specific primer set would allow for human DNA quantification and confirm that the sample is of human origin. In this case, the method would not require any additional steps in the current DNA workflow and would consume minimal volume (1–2 µl) of DNA extract, which would allow for easy implementation into forensic laboratories.

This study demonstrated proof of concept that presumptive identification of multiple body fluids can be achieved using a single-well qPCR assay that does not rely on amplification of multiple microbial species per body fluid for accurate classification of feces, saliva, and vaginal/menstrual secretions. This is of importance because there are no current in-use serology tests that identify vaginal/menstrual secretions or feces, both of which could provide useful information during sexual assault investigations. Future research should address the incorporation of nonmicrobial biomarkers for blood and semen, since there are two of the most common fluids found on crime scene evidence. Body fluid mixtures, compromised or degraded samples, differences in microbial signatures across and within populations, and other forensic considerations, such as antibiotic usage, species specificity, and substrate comparisons, should also be addressed in future work as full characterization for developmental validation purposes.

ACKNOWLEDGMENTS

Dr. Paul Brooks for assistance with statistical modeling.

ORCID

Carolyn Lewis  <https://orcid.org/0000-0002-8251-1095>

Sarah J. Seashols-Williams  <https://orcid.org/0000-0003-0093-8261>

REFERENCES

- Butler JM. Sample collection, storage, and characterization. In: Butler JM, editor. Fundamentals of forensic DNA typing. Burlington, MA: Academic Press; 2010. p. 79–97. <https://doi.org/10.1016/B978-0-12-374999-4.00004-7>
- Gaensslen R, editor. Sourcebook in forensic serology, immunology, and biochemistry unit IX. Washington, DC: U.S. Department of Justice and National Institute of Justice; 1983. p. 1–130.
- Sakurada K, Watanabe K, Akutsu T. Current methods for body fluid identification related to sexual crime: Focusing on saliva, semen, and vaginal fluid. *Diagnostics*. 2020;10(9):693. <https://doi.org/10.3390/diagnostics10090693>
- Harbison S, Fleming R. Forensic body fluid identification: State of the art. *Res Rep Forensic Med Sci*. 2016;6:11–23. <https://doi.org/10.2147/rrfms.s57994>
- An JH, Shin KJ, Yang WI, Lee HY. Body fluid identification in forensics. *BMB Rep*. 2012;45(10):545–53. <https://doi.org/10.5483/BMBRep.2012.45.10.206>
- Tobe SS, Watson N, Daéid NN. Evaluation of six presumptive tests for blood, their specificity, sensitivity, and effect on high molecular-weight DNA. *J Forensic Sci*. 2007;52(1):102–9. <https://doi.org/10.1111/j.1556-4029.2006.00324.x>
- Sijen T. Molecular approaches for forensic cell type identification: On mRNA, miRNA, DNA methylation and microbial markers. *Forensic Sci Int Genet*. 2015;18:21–32. <https://doi.org/10.1016/j.fsigen.2014.11.015>
- Afolabi OA, Roeder AD, Iyengar A, Hadi S. Evaluation of genetic markers for forensic identification of human body fluids. *Forensic Sci Int Genet Suppl Ser*. 2017;6:e241–3. <https://doi.org/10.1016/J.FSIGSS.2017.09.098>
- Lee HY, An JH, Jung SE, Oh YN, Lee EY, Choi A, et al. Genome-wide methylation profiling and a multiplex construction for the identification of body fluids using epigenetic markers. *Forensic Sci Int Genet*. 2015;17:17–24. <https://doi.org/10.1016/J.FSIGEN.2015.03.002>
- Courts C, Madea B. Micro-RNA – A potential for forensic science? *Forensic Sci Int*. 2010;203:106–11. <https://doi.org/10.1016/j.forsciint.2010.07.002>
- Fachet C, Quarino L, Karnas KJ. High resolution melt curve analysis based on methylation status for human semen identification. *Forensic Sci Med Pathol*. 2017;13:86–91. <https://doi.org/10.1007/s12024-016-9825-6>
- Woo PCY, Lau SKP, Teng JLL, Tse H, Yuen KY. Then and now: Use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clin Microbiol Infect*. 2008;14:908–34. <https://doi.org/10.1111/J.1469-0691.2008.02070.X>
- Hanssen EN, Avershina E, Rudi K, Gill P, Snipen L. Body fluid prediction from microbial patterns for forensic application. *Forensic Sci Int Genet*. 2017;30:10–7. <https://doi.org/10.1016/j.fsigen.2017.05.009>
- Doi M, Gamo S, Okiura T, Nishimukai H, Asano M. A simple identification method for vaginal secretions using relative quantification of *Lactobacillus* DNA. *Forensic Sci Int Genet*. 2014;12:93–9. <https://doi.org/10.1016/j.fsigen.2014.05.005>
- Akutsu T, Motani H, Watanabe K, Iwase H, Sakurada K. Detection of bacterial 16S ribosomal RNA genes for forensic identification of vaginal fluid. *Leg Med*. 2012;14:160–2. <https://doi.org/10.1016/j.legalmed.2012.01.005>
- Giampaoli S, Berti A, Valeriani F, Gianfranceschi G, Piccolella A, Buggiotti L, et al. Molecular identification of vaginal fluid by microbial signature. *Forensic Sci Int Genet*. 2012;6:559–64. <https://doi.org/10.1016/j.fsigen.2012.01.005>
- Choi A, Shin KJ, Yang WI, Lee HY. Body fluid identification by integrated analysis of DNA methylation and body fluid-specific microbial DNA. *Int J Leg Med*. 2014;128:33–41. <https://doi.org/10.1007/s00414-013-0918-4>
- Fleming RI, Harbison S. The use of bacteria for the identification of vaginal secretions. *Forensic Sci Int Genet*. 2010;4:311–5. <https://doi.org/10.1016/J.FSIGEN.2009.11.008>
- Jung JY, Yoon HK, An S, Lee JW, Ahn ER, Kim YJ, et al. Rapid oral bacteria detection based on real-time PCR for the forensic identification of saliva. *Sci Rep*. 2018;8:10852. <https://doi.org/10.1038/s41598-018-29264-2>
- Miyamoto Y, Itoh K. Design of cluster-specific 16S rDNA oligonucleotide probes to identify bacteria of the *Bacteroides* subgroup

- harbored in human feces. *FEMS Microbiol Lett.* 1999;177:143–9. [https://doi.org/10.1016/S0378-1097\(99\)00291-8](https://doi.org/10.1016/S0378-1097(99)00291-8)
21. Dobay A, Haas C, Fucile G, Downey N, Morrison HG, Kratzer A, et al. Microbiome-based body fluid identification of samples exposed to indoor conditions. *Forensic Sci Int Genet.* 2019;40:105–13. <https://doi.org/10.1016/j.fsigen.2019.02.010>
 22. Zou KN, Ren LJ, Ping Y, Ma K, Li H, Cao Y, et al. Identification of vaginal fluid, saliva, and feces using microbial signatures in a Han Chinese population. *J Forensic Leg Med.* 2016;43:126–31. <https://doi.org/10.1016/J.JFLM.2016.08.003>
 23. Jakubowska J, Maciejewska A, Pawłowski R, Bielawski KP. mRNA profiling for vaginal fluid and menstrual blood identification. *Forensic Sci Int Genet.* 2013;7:272–8. <https://doi.org/10.1016/J.FSIGEN.2012.11.005>
 24. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin Chem.* 2009;55:611–22. <https://doi.org/10.1373/clinchem.2008.112797>
 25. Johnson G, Nolan T, Bustin SA. Real-time quantitative PCR, pathogen detection and MIQE. In: Wilks M, editor. *PCR detection of microbial pathogens.* 2nd ed. Hertfordshire, UK: Humana Press; 2013. p. 1–16.
 26. Chaban B, Links MG, Jayaprakash TP, Wagner EC, Bourque DK, Lohn Z, et al. Characterization of the vaginal microbiota of healthy Canadian women through the menstrual cycle. *Microbiome.* 2014;2:23. <https://doi.org/10.1186/2049-2618-2-23>
 27. Leake SL, Pagni M, Falquet L, Taroni F, Greub G. The salivary microbiome for differentiating individuals: proof of principle. *Microbes Infect.* 2016;18:399–405. <https://doi.org/10.1016/j.micinf.2016.03.011>
 28. Nakanishi H, Ohmori T, Hara M, Takada A, Shojo H, Adachi N, et al. A simple identification method of saliva by detecting streptococcus salivarius using loop-mediated isothermal amplification. *J Forensic Sci.* 2011;56(Suppl 1):S158–61. <https://doi.org/10.1111/j.1556-4029.2010.01579.x>
 29. Tsai LC, Su CW, Lee JCI, Lu HS, Chen HC, Lin YC, et al. The detection and identification of saliva in forensic samples by RT-LAMP. *Forensic Sci Med Pathol.* 2018;14:469–77. <https://doi.org/10.1007/s12024-018-0008-5>
 30. Won Lee J, Yeon Jung J, Lim SK. Simple and rapid identification of saliva by detection of oral streptococci using direct polymerase chain reaction combined with an immunochromatographic strip. *Forensic Sci Int Genet.* 2018;33:155–60. <https://doi.org/10.1016/j.fsigen.2017.12.011>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Lewis C, Seashols-Williams SJ. Design and optimization of a 16S microbial qPCR multiplex for the presumptive identification of feces, saliva, vaginal and menstrual secretions. *J Forensic Sci.* 2022;67:1660–1667. <https://doi.org/10.1111/1556-4029.15029>