RBMO

ELSEVIER

ARTICLE

16S rRNA long-read nanopore sequencing is feasible and reliable for endometrial microbiome analysis

BIOGRAPHY

During her PhD in Hamburg, Anna Oberle studied immunological and therapeutical mechanisms of cancer. During her studies, Anna visited laboratories in Boston and Washington, DC, and published several studies in high-ranking journals. Since 2018, Anna has been developing novel genetic approaches for reproductive health at the 'Wunschbaby Institut Feichtinger', Vienna.

Anna Oberle[1,#], Lara Urban[2,3,#], Stefanie Falch-Leis[4], Chiara Ennemoser[4], Yoko Nagai[5], Kyota Ashikawa[5], Patricia A. Ulm[1], Markus Hengstschläger[1], Michael Feichtinger[1,6,*]

KEY MESSAGE

Long-read *16S rRNA* gene sequencing using nanopore technology delivers comparable results to short-read *16S rRNA* sequencing. We show its potential for in-situ analysis of the endometrial microbiome, which could be widely applied owing to the cost efficiency and portable character of the nanopore sequencing technology.

ABSTRACT

Research question: Full-length *16S rRNA* gene sequencing using nanopore technology is a fast alternative to conventional short-read *16S rRNA* gene sequencing with low initial investment costs that has been used for various microbiome studies but has not yet been investigated as an alternative approach for endometrial microbiome analysis. Is in-situ *16S rRNA* gene long-read sequencing using portable nanopore sequencing technology feasible and reliable for endometrial microbiome analysis?

Design: A prospective experimental study based on 33 patients seeking infertility treatment between January and October 2019. A *16S rRNA* gene long-read nanopore sequencing protocol for analysing endometrial microbiome samples was established, including negative controls for contamination evaluation and positive controls for bias evaluation. Contamination caused by kit and exterior sources was identified and excluded from the analysis. Endometrial samples from 33 infertile patients were sequenced using the optimized long-read nanopore sequencing protocol and compared with conventional short-read sequencing carried out by external laboratories.

Results: Of the 33 endometrial patient samples, 23 successfully amplified (69.7%) and their microbiome was assessed using nanopore sequencing. Of those 23 samples, 14 (60.9%) were *Lactobacillus*-dominated (>80% of reads mapping to *Lactobacillus*), with 10 samples resulting in more than 90% *Lactobacillus* reads. Our long-read nanopore sequencing revealed results similar to two conventional short-read sequencing approaches and to long-read sequencing validation carried out in external laboratories.

Conclusion: In this pilot study, *16S rRNA* gene long-read nanopore sequencing was established to analyse the endometrial microbiome *in situ* that could be widely applied owing to its cost efficiency and portable character.

[1] Wunschbaby Institut Feichtinger, Lainzerstrasse 6, Vienna 1130, Austria
[2] European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK
[3] Department of Anatomy, University of Otago, PO Box 56, Dunedin 9054, New Zealand
[4] Tyrolpath, Pathologielabor Dr. Obrist Dr. Brunhuber, Hauptplatz 4, Zams 6511, Austria
[5] Varinos, Inc., Dai 2 Gotanda Fujikoshi Bldg., 6F 5-23-1 Higashigotanda, Shinagawa-ku Tokyo, Japan
[6] Department of Oncology - Pathology, Karolinska Institutet, Karolinska vägen, A2:07171 64 Solna Stockholm, Sweden
[#] Contributed equally.

KEYWORDS

Endometrial microbiome
IVF
Long-read sequencing
Nanopore sequencing
Transportable MinION sequencing

*Corresponding author. E-mail address: michael.feichtinger@wunschbaby.at (M. Feichtinger). https://doi.org/10.1016/j.rbmo.2021.03.016 1472-6483/
Declaration: The authors report no financial or commercial conflicts of interest.

INTRODUCTION

The concept of the sterile womb has recently been challenged by multiple investigators (*Chen et al., 2017*; *Moreno and Simon 2018*; *Winters et al., 2019*). Even more so, the microbial colonization of the endometrium has been suggested to affect implantation and pregnancy success (*Moore et al., 2000*; *Moreno et al., 2016*). *Lactobacillus* dominance of the endometrial microbiome has been the focus of several publications and was shown to be significantly associated with pregnancy outcome after IVF attempts (*Moreno et al., 2016*). It was further found to be underrepresented in women seeking fertility treatment (*Kyono et al., 2018*; *2019*). Additionally, endometrial microbiome sequencing might enhance the diagnostic accuracy in patients with repeated implantation failure caused by chronic endometritis (*Moreno et al., 2018*).

Currently, short-read sequencing of one or multiple hypervariable regions, e.g. the V3/V4 region, of the bacterial *16S rRNA* gene, is most commonly applied to investigate bacterial composition. Full-length sequencing of the *16S rRNA* gene has been shown to facilitate microbiome characterization by providing a deeper level of taxonomic resolution (*Wagner et al., 2016*; *Johnson et al., 2019*). Oxford Nanopore Technologies' (Oxford, UK) MinION has enabled real-time sequencing of long reads through nanopores as a result of continuous electrical sensing of sequential nucleotides along single-stranded DNA. Short-read sequencing technologies like Illumina are limited to a read length of a few hundred base pairs (bp), which can only cover parts of the *16S rRNA* gene; long-read sequencing technologies, however, like Oxford Nanopore Technologies or Pacific Bioscience can sequence reads longer than 2 Mbp (*Nakano et al., 2017*; *Jain et al., 2018*). For targeted *16S rRNA* microbiome sequencing, a read length of about 1500 bp is sufficient to cover the whole *16S rRNA* gene, including all hypervariable regions. Although long-read sequencing approaches used to suffer from high native error rates (*Karst et al., 2020*), sequencing accuracy has recently increased and has been shown to be adequate for microbial diversity studies (*Urban et al., 2021*).

Comparisons of short-read *16S rRNA* gene sequencing and long-read nanopore sequencing applied to cervical and gut samples have previously shown that both approaches delivered similar microbiome profiles, with nanopore sequencing significantly reducing the time investment (*Shin et al., 2016*; *Quan et al., 2019*). In addition, Oxford Nanopore Technologies' portable sequencing device, the MinION, does not require high initial investment costs and can easily be used *in situ*, e.g. in gynaecological clinics or health institutes.

Compared with other human tissues, the endometrium has a relatively low-biomass microbiome (*Moreno and Simon, 2018*; *Molina et al., 2021*). Because of this low biomass and high risk of amplification bias from contaminating bacteria, the assignment of uterine microbiota is challenging (*O'Callaghan et al., 2020*). Additionally, several studies that have investigated the uterine microbiome have failed to include sufficient experimental controls to validate their sequencing results (*O'Callaghan et al., 2020*; *Molina et al., 2021*). Therefore, an optimal sequencing framework that allows for reliable and cost-efficient assessment of the endometrial microbiome remains to be designed.

In the present study, the feasibility of *16S rRNA* gene long-read nanopore sequencing for endometrial microbiome analysis is investigated; a comparison is made with conventional short-read *16S rRNA* sequencing techniques as well as external nanopore sequencing validations; and its potential as a fast, inexpensive in-situ diagnostics tool in gynaecological clinics is explored.

MATERIALS AND METHODS

Study design and sampling

The endometrial microbiome was assessed using long-read nanopore sequencing from women seeking infertility treatment at the Wunschbaby Institut Feichtinger, Vienna, Austria. Samples were obtained from leftover endometrial tissue from routine endometrial scratching. All 33 patients included in the study were healthy, non-smoking women with primary (n = 16) and secondary infertility (n = 17). Women who had taken antibiotics or undergone intrauterine manipulation, e.g. embryo transfer or hysteroscopy, less than 1 month before the intervention were excluded from the study.

After placing a vaginal speculum, the cervix was visualized and the vaginal wall and cervix were thoroughly disinfected using octenidine dihydrochloride and phenoxyethanol (Octenisept, Schuelke and Mayr GmbH, Norderstedt, Germany). An endometrial suction curette (Probet) (Gynétics, Lommel, Belgium) was then inserted into the uterine cavity without touching the vaginal wall. To avoid vaginal or cervical contamination, a middle fraction of the endometrial tissue was directly pressed from the inside of the suction curette into a sterile 2-ml Eppendorf tube without touching the tube with the wall of the suction curette. The tissue was freshly frozen at –20°C within 1 h after sampling and stored for a maximum of 2 weeks at –20°C before DNA extraction.

In women with regular menstrual cycles, endometrial scratching was carried out from cycle day 20 to 23.

All participants included in this study provided written informed consent. The study was approved by the Ethics Committee of the Medical University of Vienna on 23 July 2019 (ID: EK-1181/2019). The authors report no conflict of interest.

DNA extraction

The performance of two DNA extraction protocols was compared using the ZymoBIOMICS Microbial Community Standard (Zymo Research, Irvine, CA, USA). First, DNA from the bacterial standard was extracted with the DNeasy Blood and Tissue kit (QIAGEN, Hilden, Germany) using the protocol for animal tissue extraction, including the pretreatment for difficult-to-lyse bacteria by using pathogen lysis tubes according to the manufacturer's instructions. Second, DNA from the bacterial standard was extracted using PureLink™ Microbiome DNA Purification Kit (Thermo Fisher Scientific, Carlsbad, CA, USA) according to the protocol. Quality and quantity of extracted DNA was analysed on an E-Gel NGS 0.8% Agarose and with the Qubit™ 4 Fluorometer using the Qubit™ dsDNA BR Assay Kit (all Thermo Fisher Scientific, Carlsbad, CA, USA).

For subsequent DNA extraction of each of our endometrial sample batches, a negative control (deionized water) was included in the entire workflow to determine contaminants from external

sources or from the kits used (the so-called 'kitome') (*Salter et al., 2014*; *Kim et al., 2017*; *de Goffau et al., 2018*).

Polymerase chain reaction amplification, library preparation and sequencing

To evaluate polymerase chain reaction (PCR) and sequencing bias for relevant gram-positive and gram-negative bacterial species, ZymoBIOMICS Microbial Community DNA Standard (Zymo Research, Irvine, CA, USA) was included as positive control in the *16S rRNA* gene nanopore sequencing workflow. The 16S Barcoding Kit SQK-RAB204 (Oxford Nanopore Technologies, Oxford, UK) was used for PCR amplification and library preparation according to the manufacturer's instructions: the kit-specific primer sequences 27F and 1492R was used for *16S rRNA* gene amplification (*Nossa et al., 2010*). This primer pair amplifies almost the entire *16S rRNA* gene, including all hypervariable regions V1 to V9, and results in DNA fragments of about 1500 bp. Primers in the 16S Barcoding Kit additionally contain a RAP adapter sequence and a barcode-specific sequence. Twelve different barcodes were used for sample discrimination per sequencing runs. All primer sequences are presented in Supplementary Table 1. For PCR input, the maximum amount of extracted DNA was used (10 µl) and 40 PCR cycles were used for amplification according to the manufacturer's instructions (initial denaturation: 60 s, 95°C; PCR cycles: 20 s, 95°C, 30 s, 55°C, 120 s, 65°C; final extension: 5 min, 65°C). A total of 10 µl of PCR product was used to check amplification success on an agarose gel. The PCR product was purified using 0.6x SPRIselect magnetic beads (Beckman Coulter, USA) and quantity was analysed with the Qubit™ 4 Fluorometer using the Qubit™ 1 x dsDNA HS Assay Kit. In case of insufficient amount of amplified DNA, another PCR with 20 amplification cycles was carried out and the PCR product was processed as described. Purified samples and controls were then pooled to be equimolar (7.5 fmol/µl). The FLO-MIN106 sequencing flow cells were primed using the Flow Cell Priming Kit EXP-FLP001 (Oxford Nanopore Technologies, Oxford, UK) and loaded with the pooled library according to the manufacturer's instructions. In total, four different flow cells were used in six sequencing runs, loading between three and 12 samples per run. Flow cells were washed after the first run according to the manufacturer's instructions (Flow Cell Wash Kit EXP-WSH003) (Oxford Nanopore Technologies, Oxford, UK) to be re-used once more for sequencing. The MinION sequencing device was used for sequencing for about 6 h.

Data analysis

Guppy, provided by the MinKNOW platform (Oxford Nanopore Technologies, oxford, UK), was used for live basecalling (*Wick et al., 2019*). DNA sequencing data after basecalling is publicly available at the European Nucleotide Archive (ENA) (accession number: PRJEB38794).

The *16S rRNA* workflow provided by the EPI2ME platform (Oxford Nanopore Technologies; https://epi2me.nanoporetech.com) was used for demultiplexing and bacterial taxonomy assignment to the taxonomic genus level, based on the National Center for Biotechnology Information (NCBI) *16S rRNA* bacterial database (*O'Leary et al., 2016*). The sequencing reads were filtered to a minimum Q-score of 9 and classified to bacterial genus level. After contamination analysis based on the negative controls (see 'Contamination analysis below'), potential bacterial contaminant was removed from the endometrial sequencing datasets. Samples with less than 10,000 classified bacterial reads, more than 90% contamination reads, or both, were considered as very low-biomass samples and excluded from downstream analyses. Bacterial genera accounting for less than 1% frequency in the respective sample were removed from the analysis.

Contamination analysis

To remove exogenous bacterial DNA contamination, the R package *microDecon* was used, which uses taxonomic proportions in negative controls to systematically identify and remove contaminating reads from the biological samples (*McKnight et al., 2019*). The function *decon* with default parameter settings was applied to the entire dataset of internal and external endometrial samples and negative controls.

To compare the present decontamination approach with other approaches, prevalence-based contaminant identification as implemented by the R package *decontam* (*Davis et al., 2018*) was also used. This approach has previously been highlighted as the best-performing computational contaminant identification tool if the microbiome under investigation is widely unknown (*Karstens et al., 2019*); however, it only allows for the removal of entire taxa instead of contaminating reads, and it has since been outperformed by *McKnight et al. (2019)*. The frequency-based approach and the prevalence approach were run separately on our samples and six negative control samples, and on the Varinos external validation and its two negative control samples. A stringent probability threshold of $P < 0.5$ was used to identify contaminants.

External validation

Two independent external validations using different sequencing approaches were carried out. Eleven endometrial DNA samples were analysed in an external laboratory (Tyrolpath Pathologielabor Zams, Austria) using different PCR and library preparation conditions. Samples were sequenced using both nanopore (Oxford Nanopore Technologies, Oxford, UK) and Illumina short-read sequencing (MiniSeq) (Illumina, Inc., San Diego, CA, USA). External validation of seven additional endometrial DNA samples and two negative control samples was carried out by another external laboratory (Varinos, Inc., Tokyo, Japan) specialized in endometrial microbiome sequencing using Illumina short-read sequencing technology (*Kyono et al., 2018*; *2019*). After sequencing, basecalling and bacterial taxonomy assignment (see 'Contamination analysis'), potential bacterial contaminants were removed from all sequencing datasets. Samples with less than 1000 bacterial reads were removed from the analysis. Data were subsequently analysed in Python, using the SciPy package for statistical calculations such as Spearman correlation coefficients (*Virtanen et al., 2020*).

External validation Nanopore: Tyrolpath

Bacterial *16S rRNA* gene amplification was carried out using the following six primer pairs: 8F x 534R, 343F x 798R, 517F x 926R, 784F x 1114R, 917F x 1407R and 1099F x 1541R. Primer sequences are listed in Supplementary Table 1. These primer pairs cover all hypervariable regions V1–V9 of the *16S rRNA* gene (*Nossa et al., 2010*) and have been

pooled into one bacterial primer mix and 45 PCR cycles underwent amplification. The Rapid PCR Barcoding Kit SQK-RPB004 (Oxford Nanopore Technologies, Oxford, UK) was used in accordance with the associated protocol provided by the manufacturer. After the amplicons were cleaned, the eluted sample was quantified with the Qubit™ 2.0 fluorometer using the Qubit™ dsDNA HS Assay Kit. All barcoded samples were pooled to a total of 50–100 fmol. After that, the flow cell was primed using the Flow Cell Priming Kit EXP-FLP001 (Oxford Nanopore Technologies, Oxford, UK) and loaded with the pooled library according to the manufacturer's instructions. Albacore software (Albacore v1.1; available to Oxford Nanopore Technologies customers via their community site: https://community.nanoporetech.com) was used for basecalling. Bacterial taxonomy assignment was carried out using the 'What's In My Pod' (WIMP) workflow from EPI2ME software (Oxford Nanopore Technologies, Oxford, UK), based on the NCBI RefSeq database (*O'Leary et al., 2016*).

External validation Illumina: Tyrolpath

Bacterial *16S rRNA* gene amplification was carried out using the same six primer pairs as for the nanopore library preparation at Tyrolpath laboratory: 8F x 534R, 343F x 798R, 517F x 926R, 784F x 1114R, 917F x 1407R and 1099F x 1541R. The Nextera XT DNA Library Preparation Kit FC-131-1096 (Illumina, Inc., San Diego, CA, USA) was used in accordance with the associated reference guide. After DNA fragmentation, the index primers of the Nextera XT Index Kit v2 Set A FC-131-2001 (Illumina, Inc., San Diego, CA, USA) were attached to the respective samples in a PCR reaction. The amplified libraries were cleaned, normalized and pooled in accordance with the manufacturer's protocol. Before sequencing, the pooled library was denatured and diluted according to 'Protocol B: bead-based normalization method' in the 'Denature and dilute libraries guide' of the MiniSeq System. The diluted library was denatured, loaded onto a reagent cartridge and sequenced on the MiniSeq sequencing device (Illumina, Inc., San Diego, CA, USA). The BaseSpace Sequence Hub (Illumina, Inc., San Diego, CA, USA) was used for basecalling and data analysis. The Kraken Metagenomics application (v2.0.1) (Illumina, Inc., San Diego, CA, USA) was used for bacterial taxonomy assignment.

External validation Illumina: Varinos

The primers 27Fmod and 338R were used for *16S rRNA* gene amplification, using targeting the V1–V2 region and including a 5′ Illumina Nextera XT overhang sequence (*Kim et al., 2013*; *Walters et al., 2016*) (Supplementary Table 1). To reduce bacterial contamination from polymerase reagent, yeast-made Taq polymerase (Mitsui Chemicals, Chiba, Japan) was used for amplification. Amplicon mixtures were purified by using Agencourt AMPure XP (Beckman Coulter, Brea CA, USA). The purified PCR samples were multiplexed by using a dual-index approach with the Nextera XT Index kit v2 (Illumina, Inc., San Diego, CA, USA). KAPA HiFi HotStart ReadyMix (Kapa Biosystems, Wilmington, MA, USA) was used for indexing PCR. The final library was paired-end sequenced at 2 × 251 bp using a MiSeq Reagent Kit v.3 on the Illumina MiSeq platform (Illumina, Inc., San Diego, CA, USA).

The adapter sequences were removed using Trimmomatic-0.38 (*Bolger et al., 2014*), and the paired-end reads were joined by using EA-Utils fastq-join (*Aronesty, 2013*). The primer binding region and low-quality reads (Q score <25, read length <250 bp or >400 bp) were removed from the sequences by using prinseq-lite-0.20.4 (*Schmieder and Edwards, 2011*). Operational taxonomic units were created by pick_otus.py of QIIME 1.9.1 (*Caporaso et al., 2010*) (de-novo operational taxonomic units picking using uclust *[Edgar, 2010]*, sequence similarity threshold = 99.5%). To assign species level taxonomy, a homology search using BLAST was carried out on the representative sequence of each OTU based on the SILVA132 (*Quast et al., 2013*) and STIRRUPS (*Fettweis et al., 2012*) databases. Sequences with less than 99% homology with the database and sequences with less than 95% alignment were estimated as chimeras and excluded from the analysis.

RESULTS

Bias evaluation

To minimize DNA extraction bias, two different DNA extraction methods were compared using the ZymoBIOMICS Microbial Community Standard (see Materials and methods section). Both methods introduced some bias to the expected bacterial composition (FIGURE 1). The PureLink™ Microbiome DNA Purification Kit ('DNA extraction a') was used for DNA extraction and was selected for all subsequent extractions as it represented one of the standard bacterial taxa that has been shown to be of relevance to endometrial research (*Moreno and Simon, 2018*), *Lactobacillus*, with less bias than the DNeasy Blood and Tissue kit ('DNA extraction b') (FIGURE 1).

To evaluate PCR and sequencing bias for gram-positive and gram-negative bacteria, ZymoBIOMICS Microbial Community DNA Standard was included as positive control during amplification, library preparation and sequencing (see Materials and methods section). All eight bacterial species included in the community standard were successfully amplified, sequenced and identified in the computational analysis workflow. *Escherichia* was slightly under-represented, whereas *Staphylococcus* and *Bacillus* were over-represented. *Lactobacillus* was represented quite accurately, with a sequencing bias variation approximately equal to 25% (FIGURE 1, Supplementary Table 2 and Supplementary Table 3).

To evaluate detection of additional relevant bacterial genera, small amounts (0.1% and 0.3%) of isolated bacterial DNA (*Neisseria* and *Streptococcus*) were spiked into the DNA standard and positively tested vaginal swabs were sequenced (*Mycoplasma*). All three bacterial species were successfully identified with the established nanopore sequencing workflow and were absent in the respective control samples of the spike-in experiment (data not shown).

Bacterial contamination

Bacterial taxa present in the laboratory environment, DNA extraction, amplification and sequencing library preparation kits can seriously alter the results of sequencing-based microbiome analysis, especially in low-biomass samples such as human tissue specimens (*Salter et al., 2014*; *de Goffau et al., 2018*). Therefore, negative control samples were included in the entire workflow to determine which bacterial taxa are part of the kitome (*Kim et al., 2017*) or occur because of other contaminating sources. Deionized water was used for DNA extraction instead of endometrial tissue, and this sample was used for PCR amplification, purification and was sequenced alongside the endometrial samples. The most abundant

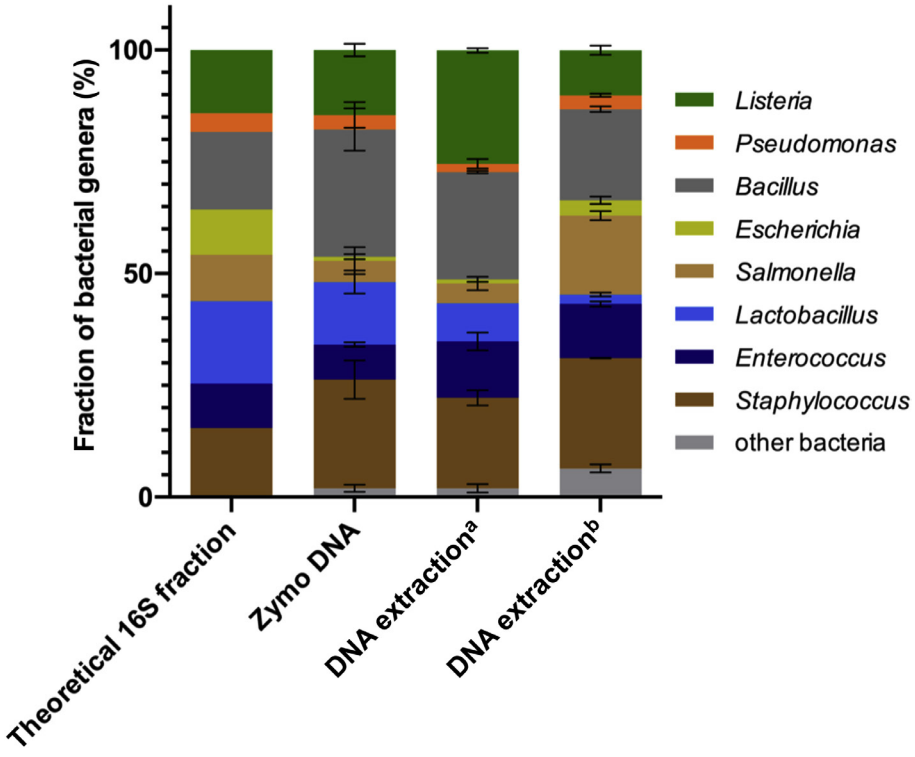

FIGURE 1 Evaluation of DNA extraction, polymerase chain reaction and sequencing bias for different gram-positive and gram-negative bacterial species. Sequencing of ZymoBIOMICS Microbial Community DNA Standards and ZymoBIOMICS Microbial Community Standards with different DNA extraction methods and comparison with expected 16S fraction. Mean and SD of all bacterial genera are shown from experiments based on Zymo Microbial Community DNA Standards (n = 3) and Zymo Microbial Community Standards after DNA extraction (using Thermo Fisher Scientific [Carlsbad, CA, USA] or QIAGEN [Hilden, Germany], n = 2 per DNA extraction protocol). [a]Thermo Fisher Scientific; [b]QIAGEN.

bacteria identified in six negative controls (one negative control per sequencing batch/flow cell run) are presented in FIGURE 2. Similar bacterial genera were present in the six negative controls, with the most abundant taxa comprising *Moraxella*, *Staphylococcus*, *Burkholderia*, *Streptococcus*, *Actinetobacter, Methylobacterium, Chryseobacterium, Luteibacter* and *Cutibacterium* (FIGURE 2).

To statistically determine the most probable contaminants, two different decontamination tools were compared: *microDecon* (*McKnight et al., 2019*) and *decontam* (*Davis et al., 2018*) (see Materials and methods section). The prevalence-based contaminant identification tool, *decontam,* was used. In total, 18 bacterial genera were identified as potential contaminants: *Moraxella, Staphylococcus, Burkholderia, Streptococcus, Acinetobacter, Methylobacterium, Cutibacterium, Pseudomonas, Acidovorax, Gemella, Psychrobacter, Bacillus* (identified from internal negative control sequencing) and *Acinetobacter, Pseudomonas, Acidovorax, Escherichia, Ralstonia,* and *Sphingomonas* (identified from Varinos negative control sequencing). As the *decontam* R package does not (partially) remove bacterial genera in the endometrial samples, the identified bacterial genera have to be removed entirely from all datasets. Frequency-based decontamination as implemented by *decontam* did not result in any identification of contaminants.

The frequency-based decontamination tool *microDecon (McKnight et al., 2019)* uses taxonomic proportions in negative controls to systematically identify and remove contaminating reads from the biological samples. Here, the following bacterial taxa were identified as contaminants: *Burkholderia* (removed entirely in all endometrial samples) and *Moraxella, Staphylococcus, Streptococcus, Acinetobacter, Methylobacterium, Cutibacterium, Pseudomonas, Lactobacillus, Acidovorax, Gemella, Bacillus, Enterococcus, Escherichia, Finegoldia, Ralstonia* and *Sphingomonas* (removed partially in some endometrial samples).

Most of the bacteria found in the negative control samples are widely known as laboratory and kit contamination specimens (*Salter et al., 2014*; *Kim et al., 2017*; *de Goffau et al., 2018*). All bacteria that surpass the 0.5% frequency threshold in the number of reads per sample were included in the final analysis. The read counts per taxon and sample before and after decontamination with *microDecon* are presented in Supplementary Table 5 and Supplementary Table 6, respectively.

Endometrial microbiome sequencing

A total of 33 endometrial biopsy samples from patients seeking fertility treatment were processed by the *16S rRNA* nanopore sequencing workflow after DNA extraction according to the PureLink™ protocol (Methods). Mean patients' body mass index was 22.65 (SD 3.55) and mean age 35.72 (SD 4.36) years.

Samples were considered as negative (no amplification) if after the second PCR amplification no *16S rRNA* gene amplification band was visible, if sequencing identified less than 10,000 bacterial reads, if sequencing revealed major composition of potential bacterial

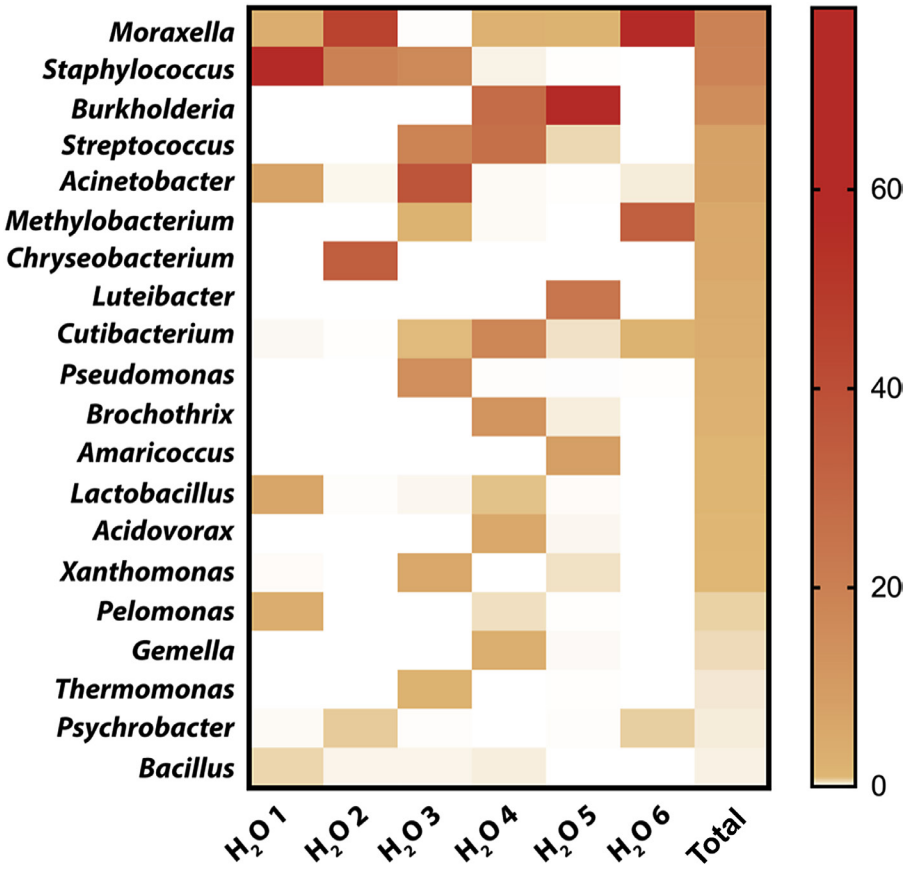

FIGURE 2 Heatmap of bacterial occurrence in negative controls. Deionized water samples (H_2O) were used for DNA extraction for each sequencing batch instead of endometrial tissue and these samples were carried along the entire workflow for PCR amplification, library preparation and sequencing together with the endometrial samples (se Materials and methods section) to determine candidate bacterial contaminants. Bacterial genera present with more than 0.1% in each negative control sample are shown. The bar indicates the colour code for the frequency in per cent in each sample.

contaminants (>90% of sequencing reads assigned to contaminants), or both. Of the 33 samples, 23 samples were successfully amplified and sequenced (69.7%). The other 10 samples (30.3%) did not show *16S rRNA* gene amplification (n = 8) or poor sequencing output with less than 10,000 bacterial reads remaining or more than 90% classified as potential contaminants (n = 2, identifiers 13 and 20). These samples were considered very low-biomass samples and excluded from downstream analyses (The PCR approach, DNA concentration of purified sequencing libraries after amplification, total number of classified bacterial reads, and number of bacterial reads after decontamination with *microDecon* are presented in Supplementary Table 4). The mean number of reads assigned to bacterial genera before decontamination (n =74,257) was reduced to 72,682 after decontamination with *microDecon* (ranging from 16,461 [identifier 23] to 182,006 [identifier 33]) (Supplementary Table 4).

The most abundant bacterial genus across all endometrial samples was *Lactobacillus*. From the 23 successfully sequenced endometrial biopsy samples, 14 (60.9%) were *Lactobacillus*-dominated, with more than 80% of all reads identified as *Lactobacillus* genus (FIGURE 3). Of these, 10 samples showed more than 90% *Lactobacillus* reads. The few endometrial samples that were not dominated by *Lactobacillus* showed different patterns of bacterial composition, including *Moraxella, Staphylococcus, Streptococcus, Kushneria* and *Haemophilus* (FIGURE 3).

External validation of endometrial microbiome sequencing

Altogether, 11 DNA samples were analysed by the internal nanopore sequencing workflow and by external nanopore and Illumina short-read sequencing in the Tyrolpath laboratory (see Materials and methods section; identifiers 1–11). Three of the 11 samples were successfully amplified and sequenced using Tyrolpath Illumina sequencing (27.3%, identifiers 1, 2 and 8), five using Tyrolpath nanopore sequencing (45.5%, identifiers 1, 2, 3, 8 and 11) and eight using our internal nanopore sequencing (72.7%, identifiers 1, 2, 3, 5, 6, 7, 8 and 9). Samples 3 and 11 from the Tyrolpath nanopore sequencing workflow were removed from the analyses as they identified less than 1000 bacterial reads. The average number of classified bacterial reads of the retained Tyrolpath validations after decontamination with *microDecon* was 12,315 (ranging from 3332 [identifier 1, Nanopore] to 48,483 [identifier 2, Illumina]) (Supplementary Table 6).

The results for the three endometrial sample identifiers (1, 2 and 8) were then compared; these were successfully sequenced by the internal approach and the Tyrolpath validations (FIGURE 4 and Supplementary Figure 1). The internal identifier 1 analysis identified relative abundances

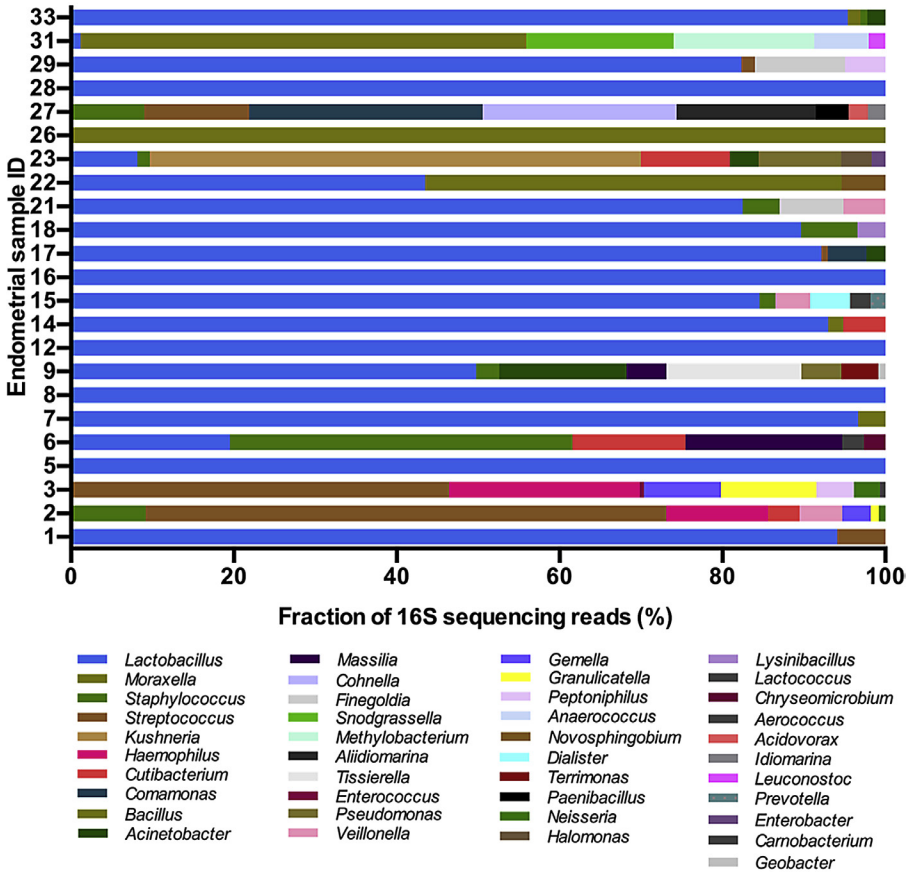

FIGURE 3 Analysis of endometrial biopsy samples. *16S rRNA* nanopore sequencing results of endometrial biopsy samples after successful *16S rRNA* amplification and sequencing. Bacterial genera identified as potential contaminants by prevalence-based contamination analysis were excluded from the endometrial sequencing results.

of 94% of *Lactobacillus* and 6% of *Streptococcus*. The Tyrolpath nanopore approach concordantly identified these two genera as most prevalent bacteria at 22% and 31%, respectively. The Tyrolpath Illumina analysis identified *Streptococcus* (36%), *Propionibacterium* (26%) and *Lactobacillus* (8%) as the most prevalent genera (FIGURE 4). Correlation analyses (Spearman correlation coefficient) (Supplementary Figure 1) still resulted in significant correlations between the internal and the respective Tyrolpath approaches. In the case of identifier 2, all three approaches confirmed *Streptococcus* as the most abundant bacteria (FIGURE 4). Although the Tyrolpath Illumina analysis also indicated a substantial presence of *Propionibacterium* (38%), a strong concordance was generally observed between our internal and external results (Supplementary Figure 1). All three sequencing approaches further confirmed *Lactobacillus* dominance (>90%) in sample identifier 8 (FIGURE 4 and Supplementary Figure 1).

Altogether, seven DNA samples were analysed by the internal nanopore sequencing workflow and by external Illumina short-read sequencing carried out by Varinos (see Materials and methods section) (identifiers 18, 22, 23, 24, 25, 28 and 33). Two of these seven samples (identifiers 24 and 25) did not show amplification in our internal nanopore approach. Varinos sequencing of identifier 24, however, detected bacteria, mainly *Cutibacterium* (35%) and *Ralstonia* (27%). For identifier 25, more than 83% of the Varinos sequencing reads were assigned to contaminating bacterial taxa (Supplementary Table 5 and Supplementary Table 6).

The mean number of classified bacterial reads after decontamination with *microDecon* across all retained Varinos samples was 14,335 (ranging from 1197 [ID 25] to 27,658 [ID 33]) (Supplementary Table 6).

The microbial composition of the internal and the external validation approach was then compared (FIGURE 4 and Supplementary Figure 1). According to both sequencing approaches, samples 18, 28 and 33 showed *Lactobacillus* dominance (>90%), and samples 22 and 23 showed non-*Lactobacillus* dominated microbiomes (FIGURE 4 and Supplementary Figure 1). Samples 22 and 23, however, showed a higher proportion of *Lactobacillus* in the internal nanopore sequencing compared with Varinos sequencing (sample 22: 43% versus 7%; sample 23: 8% versus 2.5%, respectively) (FIGURE 4 and Supplementary Table 6). In Sample 22, a substantial proportion of *Cutibacterium* was found (58%) in the Varinos Illumina sequencing, whereas our internal nanopore sequencing mainly found *Moraxella* bacterial reads (51%) alongside the *Lactobacillus* reads (FIGURE 4). In Sample 23 our results pointed towards dominance of *Kushneria* (60%) and *Cutibacterium* (11%), whereas Varinos Illumina sequencing showed *Cutibacterium* (35%), *Ralstonia* (22%) and *Staphylococcus* (11%) to be the most abundant bacterial genera (FIGURE 4 and Supplementary Table 6).

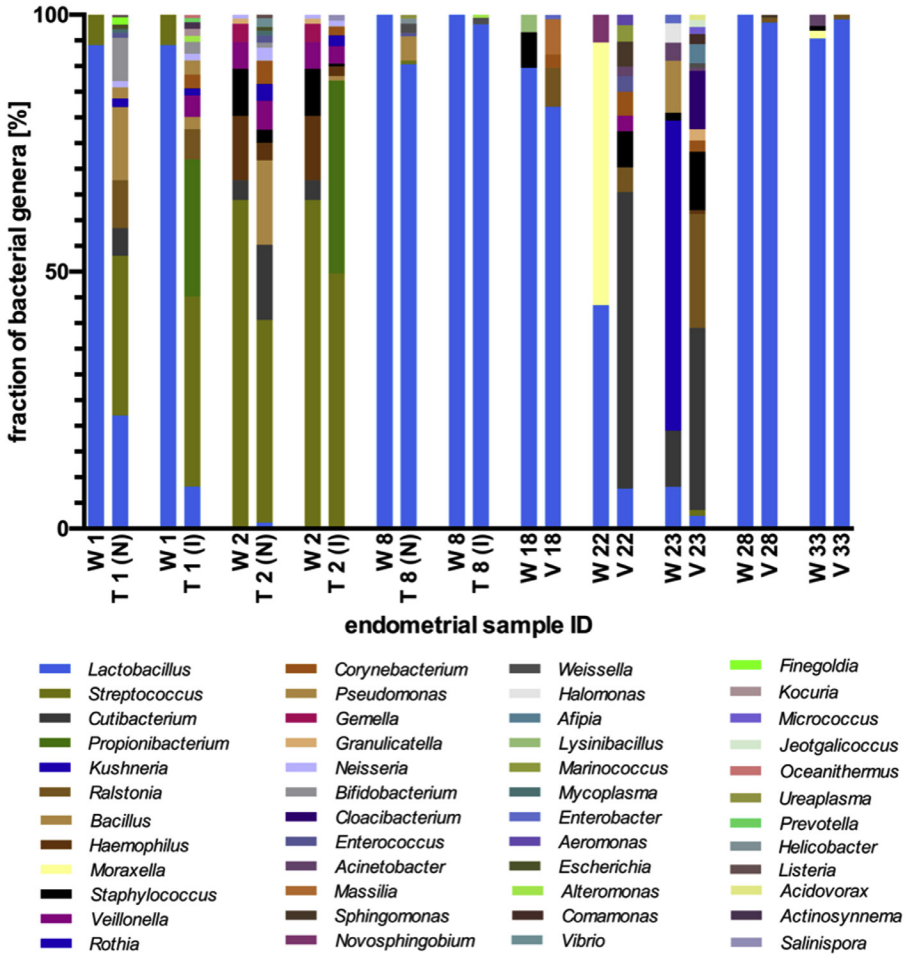

FIGURE 4 External validation of nanopore sequencing results. Comparison of endometrial microbiome sequencing results with nanopore and Illumina sequencing in external laboratories. I, Illumina sequencing; N, nanopore sequencing; T, Tyrolpath laboratory, V, Varinos laboratory, W, Wunschbaby Institut Feichtinger laboratory (internal sequencing).

DISCUSSION

In the present study, we show that *16S rRNA* gene long-read nanopore sequencing has the potential to enable accurate assessments of the endometrial microbiome, and could serve as a fast, inexpensive and portable diagnostics workflow. In addition to providing in-depth analyses of reproducibility, contamination and quantification, we show that nanopore sequencing delivers similar results to conventional short-read sequencing platforms.

Culture-independent microbiome analysis using next-generation sequencing technology is highly effective in assessing microbiome composition, including difficult-to-culture microbiota and extremely low abundant bacteria (*Moreno et al., 2016*; *Wee et al., 2018*). Short-read sequencing of one or multiple hypervariable regions of the *16S rRNA* gene used to be the standard technique for microbiome analysis. Recent studies on cervical and gut microbiota, however, have shown high concordance between long-read nanopore sequencing and conventional short-read sequencing (*Shin et al., 2016*; *Quan et al., 2019*). As the microbial load in the uterine cavity is much lower than in many other human tissues, the assessment of uterine microbiome imposes several additional challenges. Most importantly, environmental and laboratory contamination can critically alter sequencing results (*Salter et al., 2014*; *de Goffau et al., 2018*; *Stinson et al., 2019*; *Molina et al., 2021*). Therefore, rigorous technical evaluation of sequencing workflows is crucial to enable the identification of background noise (*O'Callaghan et al., 2020*; *Molina et al., 2021*). In the relatively novel research field of uterine microbiome analysis, many studies have not included any extensive evaluation of technical validity (*Haahr et al., 2019*; *O'Callaghan et al., 2020*; *Molina et al., 2021*). To address this lack of technical evaluation, we included negative and positive controls in our study to systematically evaluate bacterial contaminants and sequencing bias control.

With the use of standard bacterial communities and spike-in experiments, we show that taxa across various bacterial phyla, including species relevant for endometrial research, could be identified correctly by our nanopore sequencing workflow. Especially in low bacterial biomass specimens, such as endometrial samples, DNA extraction can introduce erroneous representation of microbial communities (*Velasquez-Mejia et al., 2018*; *Bjerre et al., 2019*). We, therefore, evaluated two different commercially available and commonly used DNA extraction protocols to minimize bias in the analysed endometrial microbial community and showed that the bias in bacterial detection can be partially

mitigated by optimizing DNA extraction methodology. The Zymo Microbial Community DNA Standard was used to analyse PCR amplification bias, with up to 40 PCR cycles run; additional PCR amplification, as carried out in some endometrial samples, might introduce additional bias (*O'Callaghan et al., 2020*).

Sequencing results of 33 endometrial samples and six negative controls, together with the application of stringent computational decontamination, allowed us to identify the bacterial contamination of our samples. Previously, a study on the mid-endometrial microbial profile of hysterectomy specimens did not detect *16S rRNA* signal exceeding that of negative controls in 40% of its samples (*Winters et al., 2019*). In our case, 10 samples (30.3%) could not be interpreted, but mostly (n = 8) because of the absence of any *16S rRNA* gene amplification; only two samples had to be excluded owing to a low number of bacterial reads (<10,000) and contamination exceeding 90% of the sequencing reads. Therefore, endometrial and low-biomass microbiome sequencing projects in general rely on the inclusion of negative controls along the entire workflow to avoid incorrect microbiome descriptions. For future studies, we recommend including quantification of total *16S rRNA* gene with quantitative PCR before library amplification in the workflow to determine bacterial load and identify very low biomass samples before sequencing.

The evaluation of our negative samples revealed a characteristic kitome contamination (*Salter et al., 2014*; *de Goffau et al., 2018*) (**FIGURE 2**). The correct identification and elimination of these contaminants is important for a correct interpretation of the sequencing results, especially in the case of low-biomass specimens like endometrial samples. We evaluated two different tools for identifying and eliminating contamination, *microDecon* and *decontam* (*Davis et al., 2018*; *McKnight et al., 2019*), with the first one performing better in a standardized comparison (*McKnight et al., 2019*). Both methods identified overlapping bacterial taxa as contaminants; however, the decontaminated results differed, mainly because of *microDecon*'s ability to only partially removing bacterial taxa. We recommend the use of *microDecon*, because, with *decontam (Davis et al., 2018)*, all reads assigned to a potential contaminating bacterial taxon must be removed across samples. The removal of the entire bacterial taxa from the sequencing results can, however, manifest in misleading results and clinical misinterpretation, especially for bacterial genera potentially causing pathologic conditions. Importantly, in our case, the genera *Streptococcus*, *Staphylococcus* and *Burkholderia* were identified by *decontam* as bacterial contaminants and, therefore, excluded from all downstream analyses by this decontamination tool. Their presence in the endometrium, however, has previously been found to be associated with chronic endometritis (*Streptococcus*, *Staphylococcus (Moreno et al., 2018*; *Ma et al., 2020*) and recurrent implantation failure (*Burkholderia) (Kitaya et al., 2019*), respectively. With the use of the tool *microDecon, Staphylococcus* and *Streptococcus* were only partially removed from some of the endometrial samples, rendering the identification of these potentially pathogenic bacteria feasible despite contamination.

To computationally assess proportions of contamination per taxon and sample, other computational tools, such as SourceTracker (*Knights et al., 2011*), require extensive prior knowledge about the microbiome under investigation (*Karstens et al., 2019*), which is not available for endometrial tissue. This shows that the differentiation between relevant bacterial taxa and potential contaminants remains a challenging problem, and clinical diagnosis will have to take the associated uncertainty into account.

We found dominance of *Lactobacillus* in 14 out of our 23 endometrial samples (60.9%) with more than 80% of all reads identified as *Lactobacillus* genus (**FIGURE 3**). Although a frequency of 90% *Lactobacillus* was previously used to discriminate between *Lactobacillus*-dominant and other endometrial samples (*Moreno et al., 2016*; *Kyono et al., 2018*; *2019*), a recent study (*Kyono et al., 2019*) observed comparable pregnancy rates of women with over 80% of *Lactobacillus* and those with over 90% of *Lactobacillus*. Even at a 90% threshold, 10 (43.5%) of our samples could be classified as *Lactobacillus*-dominant. These proportions of *Lactobacillus* are comparable to other studies analysing the endometrial microbiome after transcervical sampling of patients with infertility (*Moreno et al., 2016*; *Kyono et al., 2018*; *2019*).

This dominance of *Lactobacillus* in the endometrial cavity observed by us and others has, however, been challenged by studies using hysteroscopy for sampling instead of transcervical catheter to avoid contamination during sampling. These studies found lower frequencies of *Lactobacillus* bacteria in the endometrial cavity (*Chen et al., 2017*; *Winters et al., 2019*), highlighting another challenge of endometrial microbiome analysis, sample extraction. In the present study, endometrial samples were obtained by transcervical catheter; however, contamination risk was reduced by using endometrial suction curettes, which allow for sampling inside the straw, and by careful sampling without touching the vagina. Several studies have shown that the cervical canal is an overall safe route for sampling without substantial contamination (*Chen et al., 2017*; *Moreno and Simon, 2018*, *Wee et al., 2018*). Contamination with *Lactobacillus* outside of the endometrium, however, might still have taken place, and we propose that future studies, including a direct comparison between hysteroscopy and transcervical catheter, will have to assess the possible extent of intra-individual *Lactobacillus* caused by different sampling approaches. In addition to *Lactobacillus*, we found *Moraxella, Staphylococcus, Streptococcus* and *Kushneria* to be dominant in some endometrial samples.

The concordance between the external validation sequencing results and our internal nanopore sequencing results was relatively high (**FIGURE 4** and Supplementary Figure 1). This takes into account that both external validation cohorts were analysed by completely different workflows in different sequencing laboratories (including different primers, PCR protocols, sequencing technologies, computational analysis tools and bacterial databases). Similar concordance rates across validation sequencing results have previously been described for the cervical microbiome (*Quan et al., 2019*).

In the first external validation based on nanopore and Illumina short-read sequencing (Tyrolpath), all three samples that passed external quality control showed similar bacterial composition, including relative amounts of the

Lactobacillus genus (Supplementary Figure 1). In one sample (identifier 1), we discovered substantially lower fractions of *Lactobacillus* in the validation datasets (FIGURE 4). To follow up on this, we ascertained the assignment of our sequencing reads to the *Lactobacillus* genus by using an alternative classification tool and an independent database, which have been shown to perform well when classifying nanopore-sequenced *16S rRNA* reads (*Urban et al., 2021*), and we aligned our reads using Minimap2 (*Li, 2018*) to the *16S rRNA* SILVA132 database (*Quast et al., 2013*). This confirmed the dominance of *Lactobacillus* in our data (data not shown). The difference in *Lactobacillus* proportions might, therefore, be attributed to different *16S rRNA* sequencing primers and PCR protocols. The same applies to a slight variation in bacterial composition of sample identifier 2 between our dataset and the external datasets (FIGURE 4 and Supplementary Figure 1), with the external validations pointing towards the presence of additional bacterial taxa at low frequency. Alternatively, this variation might be due to the usage of different databases: the external validation bacterial taxonomy assignment was based on the NCBI RefSeq database, whereas we used the NCBI bacterial 16S database.

In the second external validation based on Illumina short-read sequencing provided by an established endometrial microbiome laboratory (Varinos), three out of five samples showed very similar bacterial composition (FIGURE 4 and Supplementary Figure 1). The two non-*Lactobacillus* dominant samples (identifiers 22 and 23) were significantly different when the external results were compared with our internal results. Both samples showed *Cutibacterium* as the most dominant genus in the external Illumina sequencing (58% and 35%, respectively), whereas our internal analysis classified *Moraxella* (51%) and *Lactobacillus* (43.5%) as dominant genera in sample 22 and *Kushneria* (60%) and *Cutibacterium* (11%) as dominant genera in sample 23. *Cutibacterium* has been associated with the endometrial microbiome (*Riganelli et al., 2020*); however, it was also recently described as a contaminant (*Gschwind et al., 2020*).

Altogether, the external validations suggest high reproducibility of *Lactobacillus* dominance. Workflow variations such as different sequencing primers, bacterial databases and classification tools, however, can result in differences in bacterial composition. We are, however, confident that ongoing efforts by us and others, e.g. *Urban et al., (2021)*, to standardize experimental and computational approaches, will shortly improve the comparability across platforms.

Several studies have previously associated endometrial microbiome profiles with clinical outcomes (*Moreno et al., 2016*; *Kyono et al., 2018*; *2019*). Our exploratory pilot study did not have the statistical power to provide correlations between endometrial microbiome and pregnancy outcome owing to the low sample size. We propose, however, that future large-scale studies investigating the relationship between bacterial composition and clinical outcome can rely on in-situ *16S rRNA* nanopore sequencing of the endometrial microbiome.

In conclusion, our findings support the applicability of long-read *16S rRNA* gene nanopore sequencing for endometrial microbiome analyses. To the best of our knowledge, we present the first comparison of different sequencing platforms for endometrial microbiome analysis that simultaneously include a detailed evaluation of bias and contamination. We, therefore, propose the future potential of nanopore sequencing, specifically of the portable MinION sequencing device, in combination with the described workflow as a comprehensive, efficient and inexpensive approach for assessing endometrial microbiomes and *Lactobacillus* dominance *in situ*. We further expect that, as nanopore sequencing quality continues to increase through refined pore chemistries, basecalling algorithms and consensus sequencing workflows (*Urban et al., 2021*), bacterial taxonomic classifications are likely to improve and advance opportunities for bacterial species and even strain discovery. We, therefore, reason that the low initial investment costs and the portability of the sequencing device might allow for such analyses to become increasingly accessible to gynaecologic clinics and public health institutes around the world.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.rbmo.2021.03.016.

REFERENCES

Aronesty, E. **Comparison of sequencing utility programs.** The Open Bioinformatics Journal 2013; 7: 1–8

Bjerre, R.D., Hugerth, L.W., Boulund, F., Seifert, M., Johansen, J.D., Engstrand, L. **Effects of sampling strategy and DNA extraction on human skin microbiome investigations.** Scientific reports 2019; 9: 17287

Bolger, A.M., Lohse, M., Usadel, B. **Trimmomatic: A flexible trimmer for illumina sequence data.** Bioinformatics 2014; 30: 2114–2120

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., Mcdonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R. **Qiime allows analysis of high-throughput community sequencing data.** Nat. Methods 2010; 7: 335–336

Chen, C., Song, X., Wei, W., Zhong, H., Dai, J., Lan, Z., Li, F., Yu, X., Feng, Q., Wang, Z., Xie, H., Chen, X., Zeng, C., Wen, B., Zeng, L., Du, H., Tang, H., Xu, C., Xia, Y., Xia, H., Yang, H., Wang, J., Wang, J., Madsen, L., Brix, S., Kristiansen, K., Xu, X., Li, J., Wu, R., Jia, H. **The microbiota continuum along the female reproductive tract and its relation to uterine-related diseases.** Nature communications 2017; 8: 875

Davis, N.M., Proctor, D.M., Holmes, S.P., Relman, D.A., Callahan, B.J. **Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data.** Microbiome 2018; 6: 226

De Goffau, M.C., Lager, S., Salter, S.J., Wagner, J., Kronbichler, A., Charnock-Jones, D.S., Peacock, S.J., Smith, G.C.S., Parkhill, J. **Recognizing the reagent microbiome.** Nat. Microbiol. 2018; 3: 851–853

Edgar, R.C. **Search and clustering orders of magnitude faster than blast.** Bioinformatics 2010; 26: 2460–2461

Fettweis, J.M., Serrano, M.G., Sheth, N.U., Mayer, C.M., Glascock, A.L., Brooks, J.P., Jefferson, K.K., Vaginal Microbiome, C., Buck, G.A. **Species-level classification of the vaginal microbiome.** BMC genomics 2012; 13: S17

Gschwind, R., Fournier, T., Kennedy, S., Tsatsaris, V., Cordier, A.G., Barbut, F., Butel, M.J., Wydau-Dematteis, S. **Evidence for contamination as the origin for bacteria found in human placenta rather than a microbiota.** PLoS One 2020; 15e0237232

Haahr, T., Humaidan, P., Jensen, J.S. **Non-transparent and insufficient descriptions of non-validated microbiome methods and related reproductive outcome results should be interpreted with caution.** Hum. Reprod. 2019; 34: 2083–2084

Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A., Tyson, J.R., Beggs, A.D., Dilthey, A.T., Fiddes, I.T., Malla, S., Marriott, H., Nieto, T., O'grady, J., Olsen, H.E., Pedersen, B.S., Rhie, A., Richardson, H., Quinlan, A.R., Snutch, T.P., Tee, L., Paten, B., Phillippy, A.M., Simpson, J.T., Loman, N.J., Loose, M. **Nanopore sequencing and assembly of a human genome with ultra-long reads.** Nat. Biotechnol. 2018; 36: 338–345

Johnson, J.S., Spakowicz, D.J., Hong, B.Y., Petersen, L.M., Demkowicz, P., Chen, L.,

Leopold, S.R., Hanson, B.M., Agresta, H.O., Gerstein, M., Sodergren, E., Weinstock, G.M. **Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis.** Nature communications 2019; 10: 5029

Karst, S.M., Ziels, R.M., Kirkegaard, R.H., Sørensen, E.A., Mcdonald, D., Zhu, Q., Knight, R., Albertsen, M. **Enabling high-accuracy long-read amplicon sequences using unique molecular identifiers with nanopore or pacbio sequencing.** bioRxiv 2020645903

Karstens, L., Asquith, M., Davin, S., Fair, D., Gregory, W.T., Wolfe, A.J., Braun, J., Mcweeney, S. **Controlling for contaminants in low-biomass 16S rRNA gene sequencing experiments.** mSystems 2019: 4

Kim, D., Hofstaedter, C.E., Zhao, C., Mattei, L., Tanes, C., Clarke, E., Lauder, A., Sherrill-Mix, S., Chehoud, C., Kelsen, J., Conrad, M., Collman, R.G., Baldassano, R., Bushman, F.D., Bittinger, K. **Optimizing methods and dodging pitfalls in microbiome research.** Microbiome 2017; 5: 52

Kim, S.W., Suda, W., Kim, S., Oshima, K., Fukuda, S., Ohno, H., Morita, H., Hattori, M. **Robustness of gut microbiota of healthy adults in response to probiotic intervention revealed by high-throughput pyrosequencing.** DNA Res. 2013; 20: 241–253

Kitaya, K., Nagai, Y., Arai, W., Sakuraba, Y., Ishikawa, T. **Characterization of microbiota in endometrial fluid and vaginal secretions in infertile women with repeated implantation failure.** Mediators Inflamm. 2019; 20194893437

Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., Bushman, F.D., Knight, R., Kelley, S.T. **Bayesian community-wide culture-independent microbial source tracking.** Nat. Methods 2011; 8: 761–763

Kyono, K., Hashimoto, T., Kikuchi, S., Nagai, Y., Sakuraba, Y. **A pilot study and case reports on endometrial microbiota and pregnancy outcome: An analysis using 16S rRNA gene sequencing among ivf patients, and trial therapeutic intervention for dysbiotic endometrium.** Reprod. Med. Biol. 2019; 18: 72–82

Kyono, K., Hashimoto, T., Nagai, Y., Sakuraba, Y. **Analysis of endometrial microbiota by 16s ribosomal rna gene sequencing among infertile patients: A single-center pilot study.** Reprod. Med. Biol. 2018; 17: 297–306

Li, H. **Minimap2: Pairwise alignment for nucleotide sequences.** Bioinformatics 2018; 34: 3094–3100

Ma, B., France, M.T., Crabtree, J., Holm, J.B., Humphrys, M.S., Brotman, R.M., Ravel, J. **A comprehensive non-redundant gene catalog reveals extensive within-community intraspecies diversity in the human vagina.** Nature communications 2020; 11: 940

Mcknight, D.T., Huerlimann, R., Bower, D.S., Schwarzkopf, L., Alford, R.A., Zenger, K.R. **Microdecon: A highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies.** Environmental DNA 2019; 1: 14–25

Molina, N.M., Sola-Leyva, A., Haahr, T., Aghajanova, L., Laudanski, P., Castilla, J.A., Altmae, S. **Analysing endometrial microbiome: Methodological considerations and recommendations for good practice.** Hum. Reprod. 2021

Moore, D.E., Soules, M.R., Klein, N.A., Fujimoto, V.Y., Agnew, K.J., Eschenbach, D.A. **Bacteria in the transfer catheter tip influence the live-birth rate after in vitro fertilization.** Fertil. Steril. 2000; 74: 1118–1124

Moreno, I., Cicinelli, E., Garcia-Grau, I., Gonzalez-Monfort, M., Bau, D., Vilella, F., De Ziegler, D., Resta, L., Valbuena, D., Simon, C. **The diagnosis of chronic endometritis in infertile asymptomatic women: A comparative study of histology, microbial cultures, hysteroscopy, and molecular microbiology.** Am. J. Obstet. Gynecol. 2018; 218

Moreno, I., Codoner, F.M., Vilella, F., Valbuena, D., Martinez-Blanch, J.F., Jimenez-Almazan, J., Alonso, R., Alama, P., Remohi, J., Pellicer, A., Ramon, D., Simon, C. **Evidence that the endometrial microbiota has an effect on implantation success or failure.** Am. J. Obstet. Gynecol. 2016; 215: 684–703

Moreno, I., Simon, C. **Relevance of assessing the uterine microbiota in infertility.** Fertil. Steril. 2018; 110: 337–343

Nakano, K., Shiroma, A., Shimoji, M., Tamotsu, H., Ashimine, N., Ohki, S., Shinzato, M., Minami, M., Nakanishi, T., Teruya, K., Satou, K., Hirano, T. **Advantages of genome sequencing by long-read sequencer using smrt technology in medical area.** Hum. Cell. 2017; 30: 149–161

Nossa, C.W., Oberdorf, W.E., Yang, L., Aas, J.A., Paster, B.J., Desantis, T.Z., Brodie, E.L., Malamud, D., Poles, M.A., Pei, Z. **Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome.** World J. Gastroenterol 2010; 16: 4135–4144

O'callaghan, J.L., Turner, R., Dekker Nitert, M., Barrett, H.L., Clifton, V., Pelzer, E.S. **Re-assessing microbiomes in the low-biomass reproductive niche.** BJOG: an international journal of obstetrics and gynaecology 2020; 127: 147–158

O'leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., Mcveigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C.M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V.S., Kodali, V.K., Li, W., Maglott, D., Masterson, P., Mcgarvey, K.M., Murphy, M.R., O'neill, K., Pujar, S., Rangwala, S.H., Rausch, D., Riddick, L.D., Schoch, C., Shkeda, A., Storz, S.S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R.E., Vatsan, A.R., Wallin, C., Webb, D., Wu, W., Landrum, M.J., Kimchi, A., Tatusova, T., Dicuccio, M., Kitts, P., Murphy, T.D., Pruitt, K.D. **Reference sequence (refseq) database at ncbi: Current status, taxonomic expansion, and functional annotation.** Nucleic Acids Res. 2016; 44: D733–D745

Quan, L., Dong, R., Yang, W., Chen, L., Lang, J., Liu, J., Song, Y., Ma, S., Yang, J., Wang, W., Meng, B., Tian, G. **Simultaneous detection and comprehensive analysis of hpv and microbiome status of a cervical liquid-based cytology sample using nanopore minion sequencing.** Scientific reports 2019; 9: 19337

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glockner, F.O. **The silva ribosomal rna gene database project: Improved data processing and web-based tools.** Nucleic Acids Res. 2013; 41: D590–D596

Riganelli, L., Iebba, V., Piccioni, M., Illuminati, I., Bonfiglio, G., Neroni, B., Calvo, L., Gagliardi, A., Levrero, M., Merlino, L., Mariani, M., Capri, O., Pietrangeli, D., Schippa, S., Guerrieri, F. **Structural variations of vaginal and endometrial microbiota: Hints on female infertility.** Front Cell Infect. Microbiol. 2020; 10: 350

Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J., Walker, A.W. **Reagent and laboratory contamination can critically impact sequence-based microbiome analyses.** BMC Biol. 2014; 12: 87

Schmieder, R., Edwards, R. **Quality control and preprocessing of metagenomic datasets.** Bioinformatics 2011; 27: 863–864

Shin, J., Lee, S., Go, M.J., Lee, S.Y., Kim, S.C., Lee, C.H., Cho, B.K. **Analysis of the mouse gut microbiome using full-length 16S rRNA amplicon sequencing.** Scientific reports 2016; 6: 29681

Stinson, L.F., Keelan, J.A., Payne, M.S. **Identification and removal of contaminating microbial DNA from pcr reagents: Impact on low-biomass microbiome analyses.** Lett. Appl. Microbiol. 2019; 68: 2–8

Urban, L., Holzer, A., Baronas, J.J., Hall, M.B., Braeuninger-Weimer, P., Scherm, M.J., Kunz, D.J., Perera, S.N., Martin-Herranz, D.E., Tipper, E.T., Salter, S.J., Stammnitz, M.R. **Freshwater monitoring by nanopore sequencing.** Elife 2021: 10

Velasquez-Mejia, E.P., De La Cuesta-Zuluaga, J., Escobar, J.S. **Impact of DNA extraction, sample dilution, and reagent contamination on 16S rRNA gene sequencing of human feces.** Appl. Microbiol. Biotechnol. 2018; 102: 403–411

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Van Der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, I., Feng, Y., Moore, E.W., Vanderplas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., Van Mulbregt, P., Scipy, C. **Scipy 1.0: Fundamental algorithms for scientific computing in python.** Nat. Methods 2020; 17: 261–272

Wagner, J., Coupland, P., Browne, H.P., Lawley, T.D., Francis, S.C., Parkhill, J. **Evaluation of pacbio sequencing for full-length bacterial 16S rRNA gene classification.** BMC Microbiol. 2016; 16: 274

Walters, W., Hyde, E.R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., Gilbert, J.A., Jansson, J.K., Caporaso, J.G., Fuhrman, J.A., Apprill, A., Knight, R. **Improved bacterial 16S rRNA gene (v4 and v4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys.** mSystems 2016: 1

Wee, B.A., Thomas, M., Sweeney, E.L., Frentiu, F.D., Samios, M., Ravel, J., Gajer, P., Myers, G., Timms, P., Allan, J.A., Huston, W.M. **A retrospective pilot study to determine whether the reproductive tract microbiota differs between women with a history of infertility and fertile women.** The Australian & New Zealand journal of obstetrics & gynaecology 2018; 58: 341–348

Wick, R.R., Judd, L.M., Holt, K.E. **Performance of neural network basecalling tools for oxford nanopore sequencing.** Genome Biol. 2019; 20: 129

Winters, A.D., Romero, R., Gervasi, M.T., Gomez-Lopez, N., Tran, M.R., Garcia-Flores, V., Pacora, P., Jung, E., Hassan, S.S., Hsu, C.D., Theis, K.R. **Does the endometrial cavity have a molecular microbial signature?** Scientific reports 2019; 9: 9905

Received 27 November 2020; received in revised form 1 March 2021; accepted 3 March 2021.