

## ORIGINAL ARTICLE OPEN ACCESS

# Evaluating Sequencing Strategies for Endometrial Microbiome Profiling in Endometrial Cancer: A Comparative Study of Short- and Long-Read 16S rRNA Approaches

Sophia Bebelman<sup>1,2</sup>  | Anastasiia Artuyants<sup>1,2</sup>  | Bianca Nijmeijer<sup>2</sup>  | Sandra Fitzgerald<sup>2</sup>  | Claire Henry<sup>3</sup>  | Cherie Blenkiron<sup>2</sup> 

<sup>1</sup>Auckland Cancer Society Research Centre, University of Auckland, Auckland, New Zealand | <sup>2</sup>Department of Molecular Medicine and Pathology, University of Auckland, Auckland, New Zealand | <sup>3</sup>Department of Surgery and Anaesthesia, University of Otago, Wellington, New Zealand

**Correspondence:** Cherie Blenkiron ([c.blenkiron@auckland.ac.nz](mailto:c.blenkiron@auckland.ac.nz))

**Received:** 9 September 2025 | **Revised:** 4 March 2026 | **Accepted:** 27 March 2026

**Keywords:** 16S rRNA gene sequencing | bacterial community | endometrial cancer | uterine microbiome

## ABSTRACT

**Background:** Endometrial cancer (EC) is the most common gynaecological malignancy globally, with rising incidence and notable disparities in outcomes. In New Zealand, EC rates have increased significantly, particularly among Māori and Pacific women, who face higher risks of advanced disease and poorer outcomes. Microbial dysbiosis has been implicated in EC pathogenesis, but characterising the uterine microbiome is challenging due to low microbial biomass and high contamination risk.

**Aims:** This study aimed to pilot a protocol that could inform the preparation of a larger cohort trial. Short-read Illumina MiSeq and long-read Oxford Nanopore Technologies (ONT) 16S rRNA gene sequencing were investigated to profile the uterine microbiome in people with EC.

**Methods and Results:** Uterine and vaginal swabs were analysed to assess platform performance in terms of DNA recovery, sequencing success, diversity metrics, and taxonomic resolution. The impact of sample freezing or immediate lysis prior to DNA extraction was also evaluated. ONT sequencing provided enhanced species-level resolution and improved detection of low-abundance taxa but showed variable performance in low-yield samples. Freezing prior to cell DNA extraction modestly increased bacterial 16S copy numbers and improved community consistency. Contamination was a problem across both platforms, particularly in low-biomass samples, but can be minimised during data analysis.

**Conclusion:** This study provides practical guidance for sequencing platform selection and sample handling in uterine microbiome research. Our findings support future efforts to elucidate microbial contributions to EC pathogenesis and highlight the importance of rigorous contamination control. Importantly, this is the first presentation of a New Zealand cohort and contributes valuable data from an underrepresented population and informs future research in diverse clinical settings.

## 1 | Introduction

Endometrial cancer (EC) is the most common gynaecological malignancy in high-income countries, with incidence rates rising globally [1]. It accounts for approximately 95% of

uterine cancers and is typically diagnosed in postmenopausal women [2]. While early-stage EC is often treatable, advanced or recurrent disease presents a significant clinical challenge. Strikingly, in New Zealand (Aotearoa), EC incidence has increased by nearly 60% over the past decade, with Māori and

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.

© 2026 The Author(s). *Cancer Reports* published by Wiley Periodicals LLC.

Pacific women disproportionately affected [3]. These populations experience higher rates of aggressive disease, delayed diagnosis, and poorer outcomes, highlighting the need for improved understanding of EC pathogenesis and more inclusive and global research efforts.

Advances in high-throughput sequencing have transformed our understanding of the human microbiome and its role in immunity, metabolism and disease [4]. In oncology, microbial dysbiosis has been linked to colorectal, cervical and breast cancers, with growing interest in its potential contribution to EC pathogenesis [5–7].

The uterus, once considered a sterile environment, is now recognised to host a unique and dynamic microbial community, albeit in relatively lower abundance than adjacent reproductive tract regions. Emerging evidence suggests that microbial imbalance may influence local inflammation, immune modulation, and carcinogenesis [8–11]. However, research into the uterine microbiome remains limited, largely due to technical challenges [10, 12]. A key challenge lies in the low microbial biomass of endometrial samples, which complicates accurate taxonomic classification and increases susceptibility to contamination from reagents, handling, or adjacent anatomical sites [13].

16S rRNA gene sequencing is widely used as an affordable approach for bacterial community profiling, but platform selection influences resolution and reliability [14, 15]. Short-read Illumina MiSeq sequencing offers high accuracy and robust bioinformatics support, but is limited by its short read length, which restricts genus- and species-level classification [16–19]. In contrast, long-read Oxford Nanopore Technologies (ONT) sequencing enables full-length 16S profiling, potentially improving taxonomic resolution and detection of low-abundance taxa [16, 19]. An advantage of Nanopore technology is its potential to generate sequence data from low-input samples, making it suitable for studies with limited biological material [20].

Despite these technological advances, few studies have systematically compared these sequencing platforms for low-biomass uterine samples, and none have focused on EC patients in New Zealand—a geographically isolated population with distinct demographic and clinical characteristics. Moreover, methodological variables such as sample preservation and DNA extraction protocols may further influence microbial community profiles, yet these factors are rarely evaluated in low-biomass environments.

In this study, we conducted a direct comparison of Illumina MiSeq and ONT 16S rRNA sequencing on uterine swabs from EC patients. Our objectives were: (1) to evaluate platform performance in terms of DNA recovery, sequencing success, diversity metrics, and taxonomic resolution; and (2) to assess the feasibility of uterine sampling and identify methodological considerations for future microbiome research in EC. By addressing both technical and population-specific considerations, this work will inform future uterine microbiome research and support efforts to elucidate microbial contributions to EC pathogenesis.

## 2 | Methods

### 2.1 | Ethical Approval and Patient Recruitment

A total of 31 participants were recruited from Te Whatu Ora Wellington Hospital for the collection of uterine swabs and genetic material analysis. Ethics approval was obtained from the New Zealand Health and Disability Ethics Committee (HDEC 15/CEN143) and the University of Otago (H20/002 and H25/0421). Eligible participants were aged 18 years or older and undergoing hysterectomy for any stage of endometrial cancer. Exclusion criteria included pregnancy, breastfeeding, and recent antibiotic use.

In a separate cohort, vaginal swabs were collected from three individuals undergoing surgery for suspected endometriosis at Te Whatu Ora Wellington Hospital. Ethics approval for this study was granted by the Central Health and Disability Ethics Committee (HDEC 2022 EXP 12616). Participants were aged between 16 and 45 years, with exclusions for menstruation at the time of sampling and known cervical abnormalities. All participants provided written informed consent to donate tissues and associated clinical data for the purposes of research.

### 2.2 | Sample Collection and Processing

Following surgical removal of the uterus, four swabs were collected per participant—two from the tumour site and two from adjacent non-tumorous tissue. In cases where the tumour size limited visible delineation of the tissues, fewer swabs were collected.

To identify the optimal sample collection method for downstream analysis, four swab and storage protocols were evaluated, varying in swab type, buffer composition, and storage temperature: Method A: Copan FLOQ (Copan Diagnostics) swabs stored at room temperature (RT) in DNA/RNA Shield lysis buffer (Zymo Research) for up to 4 weeks; Method B: Zymo DNA/RNA Shield collection swabs stored at RT; Method C: FLOQ swabs collected in Tris-EDTA (TE) buffer (Thermo Fisher Scientific), centrifuged, and the pellet resuspended in DNA/RNA Shield (RT); Method D: FLOQ swabs stored in TE buffer at  $-80^{\circ}\text{C}$  (for 1–2 months), thawed, centrifuged, and the pellet resuspended in lysis buffer prior to processing.

Vaginal swabs were collected by the operating surgeon or nurse following anaesthesia and prior to surgical preparation. The collection medium consisted of 1× phosphate-buffered saline (PBS; Thermo Fisher Scientific) supplemented with 25 mM trehalose dihydrate (Sigma-Aldrich) and 25 mM HEPES (Thermo Fisher Scientific). Swabs were obtained by inserting a FLOQ swab into the vagina and rotating for approximately 30 s, then placed into a tube containing 1 mL of collection medium. Samples were visually inspected, and any showing signs of blood contamination were excluded.

All swabs were preserved in their respective buffers and transported to the University of Auckland for subsequent analysis.

## 2.3 | DNA Extraction and Quantification

Genomic DNA was extracted using a ZymoBIOMICS DNA Microprep Kit (Zymo Research), with minor modifications to enhance microbial lysis. Briefly, 1 mL of swab/cell lysate (in DNA/RNA shield) was transferred to BashingBead Lysis Tubes and subjected to mechanical disruption using a Disruptor Genie bead beater (Scientific Industries) at maximum speed for 20 min. Lysates were centrifuged at 10000×g for 1 min, and 400 µL of supernatant was mixed with DNA Binding Buffer and applied to Zymo-Spin IC Columns. Columns were washed with DNA Wash Buffers 1 and 2, and DNA was eluted in 20 µL of DNase/RNase-Free Water. To remove potential inhibitors, eluates were purified using the Zymo-Spin II-µHRC Filter and ZymoBIOMICS HRC Prep Solution by centrifugation at 16000×g for 3 min, and stored at -20°C until analysis. DNA concentrations were measured using the Qubit dsDNA High Sensitivity (HS) Assay Kit (Thermo Fisher Scientific), following the manufacturer's protocol and measured using the Qubit 4 Fluorometer.

## 2.4 | Droplet Digital PCR (ddPCR)

To assess the ratio of human to bacterial DNA, droplet digital PCR (ddPCR) was performed using the QX200 Droplet Digital PCR System (Bio-Rad). Dual assays targeting the human *NRAS* gene and the bacterial 16S rRNA gene were used.

Each 23 µL ddPCR reaction contained 1× ddPCR Supermix for Probes (no dUTP) (Bio-Rad), 0.43 µM each of *NRAS* forward (5'-TGGTGAAACCTGTTTGTGG-3') and reverse (5'-ATTGGTCTCTCATGGCACTG-3') primers, 0.43 µM each of universal bacterial forward (5'-TGGAGCATGTGGTTTAATTCGA-3') and reverse (5'-TGCGGGACTTAACCCAACA-3') primers, 0.21 µM of a FAM-labelled 16S probe (5'-6-FAM-CACGAGCTGACGACARCCATGCA-ZEN-Iowa Black FQ 3'), 0.21 µM of a HEX-labelled *NRAS* probe (5'-HEX-CTGGATACA-ZEN-GCTGGACAAGAAG-IABkFQ 3'), and 5 ng of DNA template (0.22 ng/µL final concentration).

Droplets were generated using a Bio-Rad Automated Droplet Generator and then amplified using a Bio-Rad C1000 Touch Thermal Cycler (Bio-Rad) under the following conditions: enzyme activation at 95°C for 10 min; 40 cycles of 94°C for 30 s and 59°C for 1 min; followed by enzyme deactivation at 98°C for 10 min and read using a QX200 Droplet Reader. Absolute quantification was performed using QuantaSoft Analysis Software (v1.7.4.0917). Positive controls included *NRAS*-only (human cells) and *E. coli*-only samples; a no-template control (NTC) was also included.

## 2.5 | 16S rRNA Gene Amplicon Sequencing (Illumina MiSeq)

DNA samples were submitted to Auckland Genomics for 16S rRNA gene amplicon sequencing. DNA concentrations were normalised to 5 ng/µL prior to library preparation. The ZymoBIOMICS Microbial Community DNA Standard (Zymo Research) was used as a positive control, and molecular-grade water served as a negative control.

Amplicon libraries targeting the V3-V5 hypervariable regions were prepared using the Illumina protocol with modifications. The following primers were used: 357F: 5'-TCGTCGGCAGCGT CAGATGTGTATAAGAGACAGCTACGGGAGGCAGCAG-3' and 926R: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGCCGTCAATTCMTTTRAGT-3'.

PCR amplification was performed using Platinum SuperFi II PCR Master Mix (Thermo Fisher Scientific) for 35 cycles, using 2.5 µL (12.5 ng) of DNA with the following thermal cycling conditions: an initial denaturation step at 95°C for 3 min, followed by 35 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s and extension at 72°C for 30 s. A final extension was carried out at 72°C for 5 min. Amplicons were purified using AMPure XP magnetic beads and eluted in 20 µL of nuclease-free water.

A second indexing PCR was performed using custom Nextera XT index primers, followed by another bead cleanup and elution in 40 µL. Indexed libraries were pooled equimolarly, assessed for fragment size using a Bioanalyzer, normalised to 4 nM, spiked with ~24% PhiX, and sequenced on an Illumina MiSeq using a 600-cycle v3 kit (2×300 bp).

A total of 14744455 read pairs passed the filter, generating approximately 7.58 Gbp of data, with an average of ~150 000 paired-end reads per sample, providing sufficient depth for microbial community analysis.

## 2.6 | Full-Length 16S rRNA Gene Sequencing (Oxford Nanopore Technologies)

Genomic DNA samples were submitted to Auckland Genomics, University of Auckland, for full-length 16S rRNA gene sequencing using Oxford Nanopore Technologies (ONT). DNA concentrations were normalised to 2.5 ng/µL, and 10 µL (25 ng) was used for the PCR.

### 2.6.1 | First-Round PCR Amplification

Full-length 16S rRNA gene regions were amplified using ONT-adapted universal primers targeting positions 27F to 1492R, with overhang sequences compatible with downstream barcoding: ONT\_27F: 5'-TTTCTGTTGGTGCTGATATTGCAGAGTT TGATCCTGGCTCAG-3' and ONT\_1492R: 5'-ACTTGCCTGT CGCTCTATCTTCTACGGYTACCTTGTTACGACTT-3'.

Amplification was performed using Platinum SuperFi II PCR Master Mix (Thermo Fisher Scientific) with the following thermal cycling conditions: an initial denaturation step at 95°C for 1 min, followed by 35 cycles of denaturation at 95°C for 20 s, annealing at 55°C for 30 s, and extension at 65°C for 2 min. A final extension was carried out at 65°C for 5 min.

Positive controls included the ZymoBIOMICS microbial community DNA standard (Zymo Research), and molecular-grade nuclease-free water served as the negative control. Amplicons were purified using a 1:1 ratio of AMPure XP beads and eluted

in 20 µL of nuclease-free water. Post-cleanup, amplicons were quantified and re-normalised to 2.5 ng/µL.

## 2.6.2 | Indexing PCR and Library Preparation

Barcoding was performed using LongAmp Hot Start Taq 2X Master Mix (New England Biolabs) and an EXP-PBC096 barcoding kit from ONT. Reactions were set up in 25 µL total volume, with the following thermal cycling conditions: an initial denaturation at 95°C for 3 min, followed by 12 cycles of denaturation at 95°C for 15 s, annealing at 62°C for 15 s, and extension at 65°C for 45 s. A final extension step was performed at 65°C for 5 min.

Barcoded samples were pooled at equimolar concentrations, purified using AMPure XP beads, and eluted in 40 µL of nuclease-free water. Libraries were prepared using an SQK-LSK114 ligation sequencing kit (ONT), following the manufacturer's instructions.

## 2.6.3 | Sequencing and Basecalling

Sequencing was performed on a PromethION FLO-PRO114M flow cell. During sequencing, barcode trimming was disabled, and adapter trimming was enabled. Basecalling was conducted in two stages using Dorado (v0.8.0): Real-time basecalling during the run using the 'high accuracy' model and Post-run basecalling using the 'super high accuracy' model to improve read quality.

## 2.7 | Data Analysis

### 2.7.1 | Illumina MiSeq Data Processing

Illumina reads were processed using the QIIME2 (2024.2 amplicon distribution) pipeline via the New Zealand eScience Infrastructure [21]. Only forward reads were used (covering 16S rRNA section V3), trimmed to 200 bp, and denoised using DADA2 [22]. A phylogenetic tree was constructed using the align-to-tree-mafft-fasttree workflow from the q2-phylogeny plugin.

Taxonomic classification was performed using a pre-trained naïve Bayes classifier via the q2-feature-classifier plugin [23]. The classifier was trained on full-length 16S sequences from the SILVA 138 OTU database at 99% similarity [24–27]. For comparison, taxonomy was also assigned using the NCBI database [28], with a classifier generated using RESCRIPt, filtered for bacterial sequences with a minimum read length of 800 bp [25].

### 2.7.2 | Oxford Nanopore Technologies (ONT) Data Processing

ONT long-read data were processed using the Epi2me 16S workflow, also run via the New Zealand eScience Infrastructure. Taxonomic classification was performed to the genus level using minimap2 and the SILVA 138.1 database, with default settings unless otherwise specified.

Minor differences in taxonomic group naming between SILVA database versions used for Illumina and ONT were corrected to match the most recent version prior to comparative analysis.

### 2.7.3 | Diversity Analysis in R

Both Illumina and ONT datasets were further analysed using R (version 4.3.2) [29]. For alpha rarefaction, genus-level data were converted into phyloseq objects, and rarefaction curves were generated using the rarecurve function from the vegan package (v2.6–10) [30].

For diversity analysis, datasets were converted into phyloseq objects and rarefied to a sequencing depth determined by alpha rarefaction analysis (Figures S2 and S12) [18, 31]. Alpha diversity metrics (observed richness, Shannon, and Simpson indices) were calculated using estimate\_richness from phyloseq, and results were aggregated using dplyr (v1.1.4) [32]. Wilcoxon tests were used for pairwise comparisons, and Kruskal–Wallis tests followed by Dunn's tests with Benjamini–Hochberg correction were applied for multi-group comparisons using rstatix (v0.7.2) [33]. Visualisations were generated using ggplot2 (v3.5.2) [34].

Beta diversity was assessed using Bray–Curtis dissimilarity and principal coordinate analysis (PCoA) via phyloseq (v1.46.0). PERMANOVA testing was performed using adonis2 from vegan, and pairwise PERMANOVA was conducted using pairwise.adonis2 from the pairwiseAdonis package (v0.4.1), with false discovery rate (FDR) controlled via the Benjamini–Hochberg method [35].

### 2.7.4 | Taxonomic Composition and PCA in R

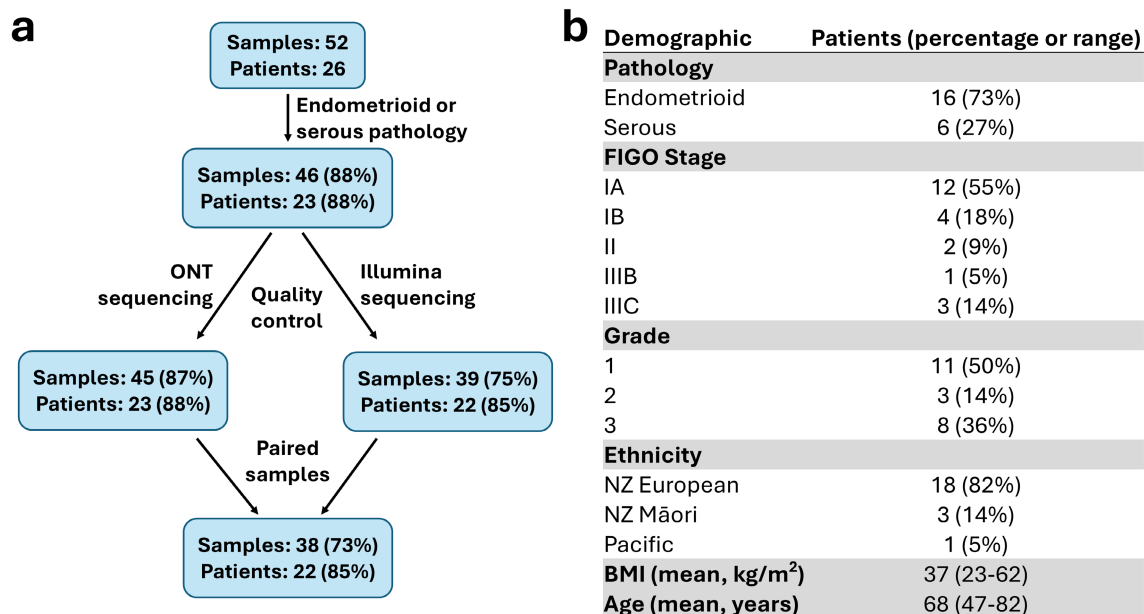
For taxonomic bar plots, genus-level data were filtered to exclude unassigned reads. Where genus-level classification was unavailable, higher taxonomic ranks were used. The top 29 genera were selected based on read counts, and the remaining taxa were grouped under 'Other'. Relative abundance was calculated per sample and visualised using ggplot2.

For principal component analysis (PCA), genus-level data were Hellinger-transformed using decostand from vegan, and taxa with zero variance were removed. Dimensionality reduction was performed using prcomp, and the percentage of variance explained by each principal component was calculated. The first two components were visualised using ggplot2.

## 3 | Results

### 3.1 | DNA Yield and Sequencing Success

To evaluate sequencing platform performance in profiling the uterine microbiome, endometrial tumour swabs were collected from 26 patients undergoing hysterectomy for endometrial cancer (Figure 1a,b). This enabled direct sampling of the uterine cavity, minimising contamination from the lower genital tract. From each patient, two swabs were taken from the tumour site and two from adjacent non-cancerous endometrial tissue. In total,



**FIGURE 1** | Sample and patient retention through processing and quality control. (a) Flow diagram showing the number and percentage of samples and patients retained after each stage of sequencing data processing and quality control. (b) Summary of demographic characteristics of patients whose samples passed all processing and quality control steps. Ethnicity is self-reported. Pathology and demographics were taken from clinical notes. Age and BMI are at surgery.

112 samples were obtained, representing an initial cohort that included 3 Māori, 2 Pacific peoples, and the remainder of European (Pākehā) descent. Comparative analysis of the Illumina data revealed no significant differences in read count, microbial diversity, or taxonomic composition between tumour and adjacent tissue swabs (data not shown) in line with a prior study [10]. Therefore, 1–2 swabs per patient from the tumour site only were retained for platform comparison to reduce redundancy.

Three patients were excluded due to tumour histology inconsistent with the study's focus on serous and endometrioid subtypes. Among the remaining 23 patients, 73% ( $n = 17$ ) presented with endometrioid carcinoma, consistent with its higher prevalence in the general population (Figure 1b). This subtype is frequently associated with metabolic comorbidities and tends to occur in younger women [36]. This was reflected in the cohort's average BMI of 37 kg/m<sup>2</sup> and the presence of five women under 60 years of age (Figure 1b).

To confirm the presence of bacterial DNA and assess sample suitability for sequencing, droplet digital PCR (ddPCR) was performed prior to library preparation (Table S1). The average human-to-bacterial DNA ratio was 166:1, consistent with the low microbial biomass expected in uterine samples. For comparison, high biomass vaginal swabs had an average human-to-bacteria ratio of 1:5.5. Despite the dominance of host DNA in uterine swabs, bacterial DNA was detectable in all samples.

A total of 46 swabs were sequenced using both ONT and Illumina platforms and distinct data analysis pipelines. Samples were excluded if they failed to meet quality thresholds defined by Epi2me (ONT) or QIIME2 (Illumina), including a minimum of 10 000 high-quality reads. ONT sequencing yielded higher read counts (mean = 677 000 reads) compared to

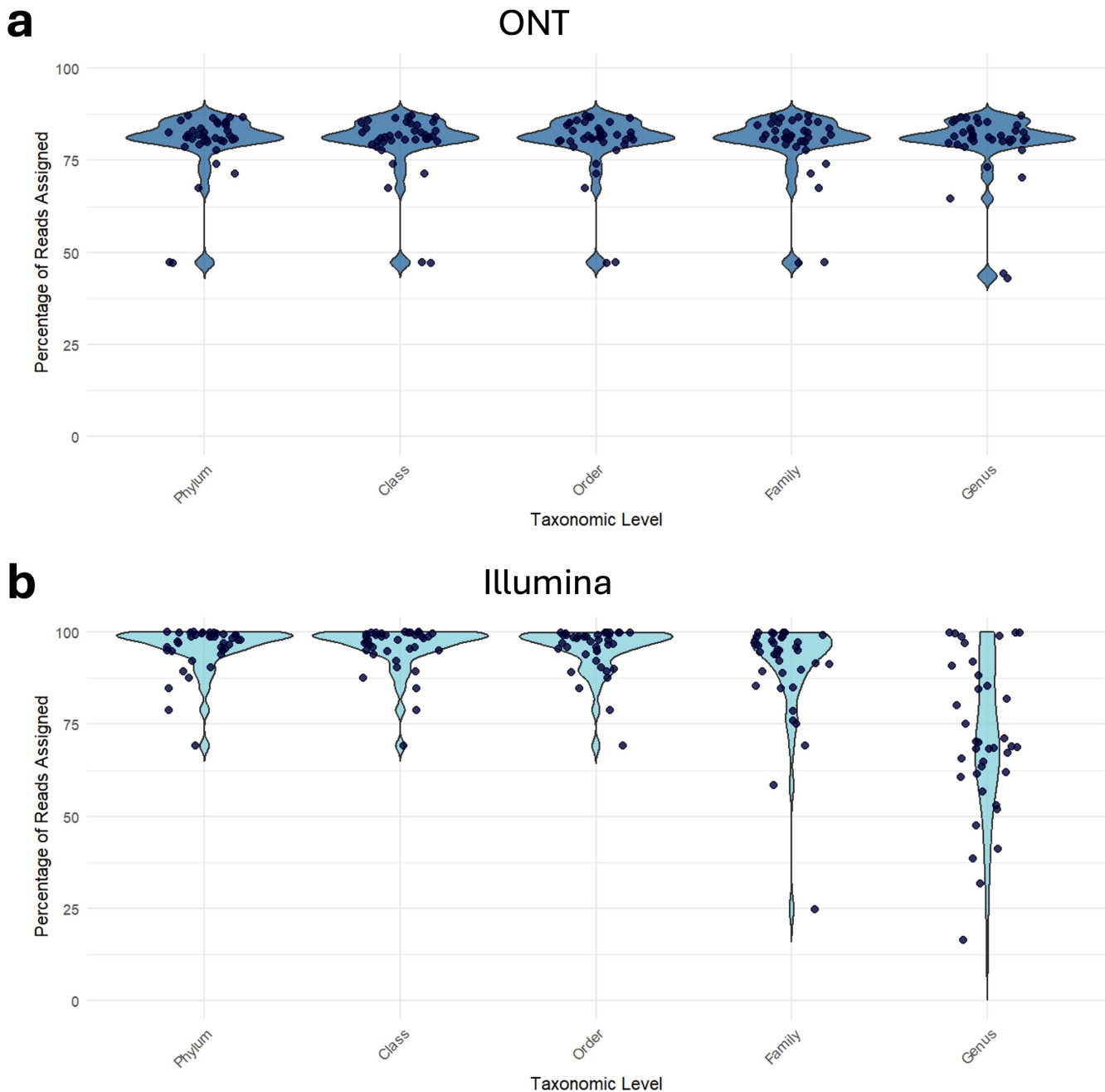
Illumina (mean = 110 000 reads). The success rate for Illumina (75%) was lower than for ONT (87%). Only samples that passed QC for both platforms were included in downstream comparative analyses to ensure consistency, leaving 38 samples from 22 patients.

## 3.2 | Taxonomic Resolution and Read Depth

To assess taxonomic assignment performance between sequencing platforms, we evaluated two widely used reference databases: SILVA and NCBI. SILVA is a highly curated database optimised for accurate classification at higher taxonomic levels (e.g., phylum to genus), but it is not recommended for species-level classification when using Epi2me due to limited resolution. In contrast, the NCBI database supports species-level assignments in both Epi2me and QIIME2 pipelines, offering finer taxonomic granularity [21, 25].

Initial analyses were conducted using both databases to compare taxonomic assignment, diversity metrics, and overall microbial profiling (Figures S1, S3 and S7). While NCBI enabled species-level resolution, SILVA demonstrated superior consistency and accuracy at higher taxonomic ranks and outperformed NCBI when benchmarked against a mock microbial community. Given the known limitations of 16S rRNA gene sequencing for species-level identification [19], and the greater reliability of SILVA at the genus level, all downstream comparative analyses between ONT and Illumina platforms were conducted using the SILVA database, with taxonomic resolution restricted to the genus level.

Taxonomic assignment efficiency was quantified across hierarchical levels for both sequencing platforms (Figure 2). ONT

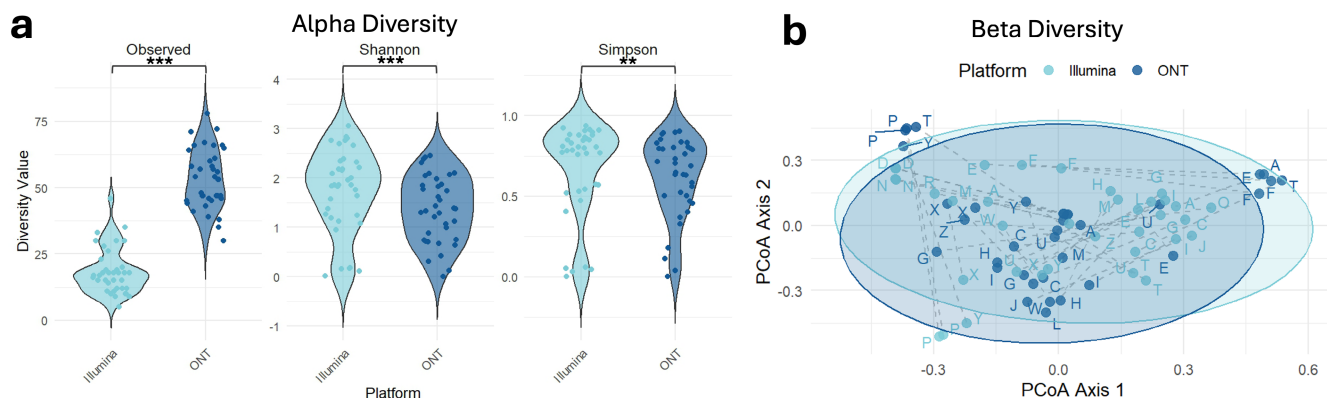


**FIGURE 2** | Taxonomic assignment efficiency between sequencing platforms. Violin plots showing the percentage of reads assigned at each taxonomic level for samples sequenced using (a) ONT and (b) Illumina platforms. Each dot represents an individual swab sample. Reads classified as unknown, unassigned, eukaryotic, or uncultured were excluded from the assigned read count to focus on confidently classified microbial taxa.

sequencing assigned approximately 80% of reads consistently from phylum to genus level, reflecting the advantage of long-read sequencing by spanning multiple variable regions of the 16S gene. Illumina sequencing, which targets a shorter hypervariable region, showed higher assignment rates at broader taxonomic levels—95% of reads were classified from phylum to order. However, assignment rates declined at finer levels, with around 90% of reads classified at the family level and only 70% at the genus level. This drop-off in resolution is consistent with the limitations of short-read sequencing, which may not capture sufficient sequence variation for confident genus-level classification.

### 3.3 | Alpha and Beta Diversity: Genus Richness, Evenness, and Community Composition

To ensure comparability across sequencing platforms, alpha rarefaction was performed to normalise read depth and reduce bias in diversity estimates. Due to substantial differences in sequencing output, ONT samples yielded significantly higher read counts than Illumina; platform-specific rarefaction thresholds were applied: 140 000 reads for ONT and 10 000 reads for Illumina, balancing retention of microbial diversity with quality control (Figure S2).



**FIGURE 3** | Comparison of alpha and beta diversity between Illumina and ONT sequencing platforms at the genus level. (a) Alpha Diversity Comparison: Violin plots of Observed, Shannon, and Simpson indices at the genus level, comparing Illumina (light blue) and ONT (dark blue) sequencing platforms. Each dot represents an individual sample. Samples were rarefied to 10000 reads (Illumina) or 140000 reads (ONT), excluding unassigned reads. Statistical significance was assessed using paired Wilcoxon rank-sum tests (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , adjusted for multiple comparisons). (b) Beta Diversity (Bray–Curtis PCoA): Principal coordinate analysis of Bray–Curtis dissimilarity at the genus level, comparing Illumina and ONT platforms. Each dot represents a sample, labelled by the patient letter. Dotted lines connect matched samples sequenced on both platforms. Ellipses indicate 95% confidence intervals. Samples were rarefied as above, excluding unassigned reads. PERMANOVA yielded  $F = 10.273$ ,  $R^2 = 0.128$ ,  $p = 0.001$  (999 permutations).

Alpha diversity metrics revealed platform-specific differences in microbial richness and evenness. ONT samples exhibited significantly higher genus richness, with a mean of 53.3 observed genera per sample, compared to 18.4 genera in Illumina samples (Wilcoxon test,  $p = 0.00000521$ ; Figure 3a). This suggests that ONT's long-read sequencing, combined with high depth, enables broader detection of microbial taxa. When each method was rarefied to 10000 reads, to account for the 14× higher read depth in the ONT data, this significant difference was lost (Figure S8).

Interestingly, Shannon diversity indices, which account for both richness and evenness, were slightly higher in Illumina samples (mean = 1.74) than ONT (mean = 1.37,  $p = 0.000467$ ). This indicates that although ONT detected more genera, Illumina reads were more evenly distributed across taxa, a difference that was retained when the methods were rarefied to 10000 reads (Figure S8). Simpson indices, which emphasise dominant taxa and community evenness, were also slightly higher in Illumina samples (mean = 0.673 for Illumina, 0.619 for ONT,  $p = 0.00850$ ), suggesting greater evenness among the most abundant genera.

Beta diversity analysis was conducted to assess differences in microbial community composition between platforms. Bray–Curtis dissimilarity revealed substantial overlap in community profiles, yet the sequencing platform accounted for 12.8% of the observed variance, which was statistically significant ( $p = 0.001$ ; Figure 3b). While many paired samples clustered closely, platform-specific effects were evident in some cases, highlighting the influence of sequencing technology on community structure.

### 3.4 | Community Structure Differences and Taxonomic Resolution

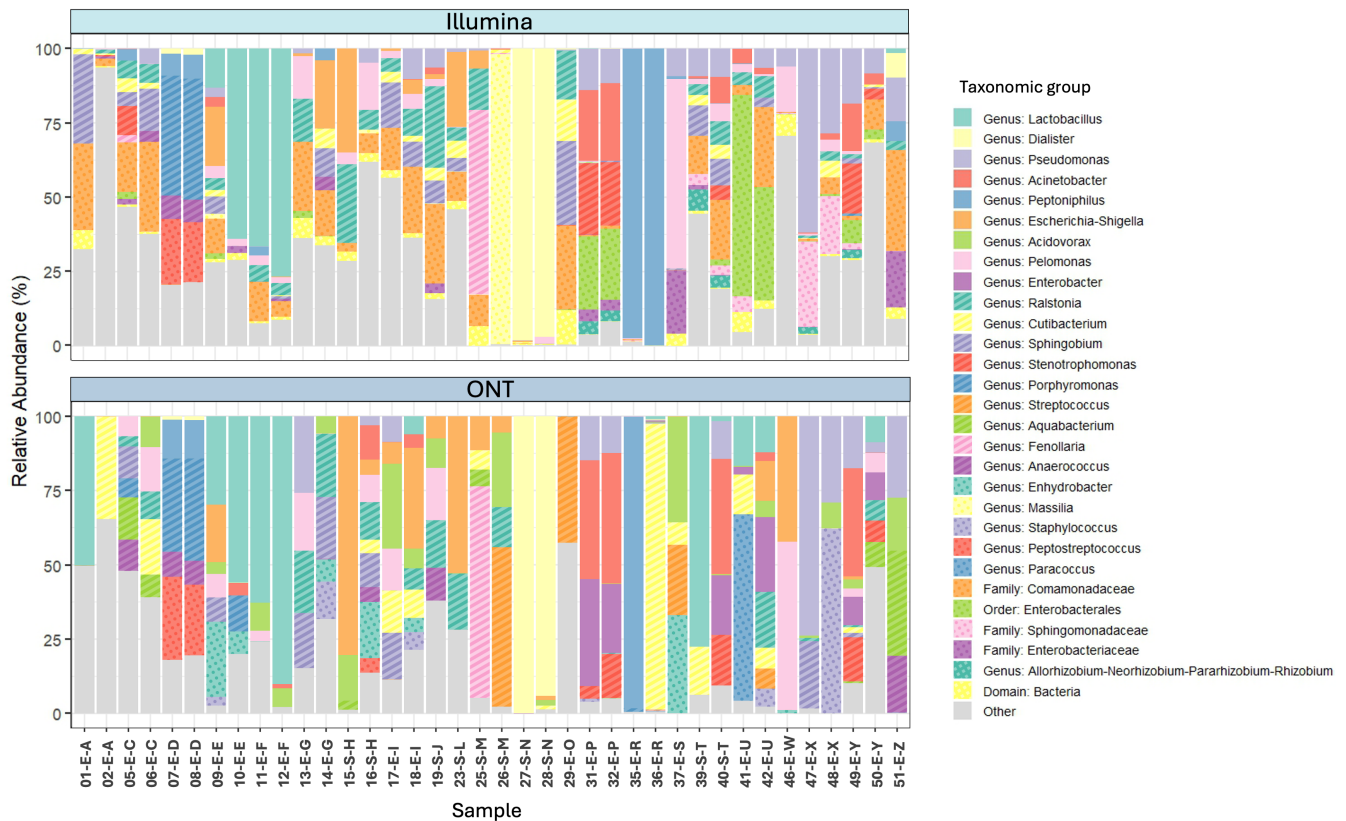
Overall, there was strong concordance in taxonomic profiles between samples sequenced using Illumina and ONT platforms

(Figure 4 and Figure S6). Microbial community composition was more influenced by inter-patient variability than by sequencing platform, with replicate swabs from the same patient showing high similarity regardless of the technology used. This suggests that both platforms reliably capture the dominant microbial signatures present in individual samples.

Despite overall concordance, differences in taxonomic resolution between platforms were evident. As previously noted (Figure 2), ONT assigned a greater proportion of reads at the genus level compared to Illumina. This discrepancy was particularly pronounced in taxa where Illumina reads were frequently classified only at the family or order level due to limitations in short-read resolution.

An example of this is the *Comamonadaceae* family, where ONT sequencing identified 23 distinct genera, while Illumina detected only 4. Notably, 79% of *Comamonadaceae* reads in Illumina samples were assigned only at the family level, with no genus-level classification. In contrast, ONT assigned just 0.003% of reads to the family level, with the majority classified into specific genera: *Acidovorax* (40%), *Pelomonas* (36%), and *Aquabacterium* (20%). *Acidovorax* was not detected at all by Illumina, and *Pelomonas* and *Aquabacterium* were underrepresented (16% and 3%, respectively).

These findings highlight the ability of ONT methodology to provide a more granular view of microbial community structure, particularly in low-biomass environments where taxonomic resolution is critical. The enhanced genus-level classification offered by ONT may be especially valuable for identifying subtle microbial shifts or rare taxa that could be biologically relevant but are missed or unclassified by short-read platforms. This has important implications for future studies aiming to link microbial signatures with clinical outcomes or disease mechanisms in endometrial cancer.



**FIGURE 4 |** Relative abundance of taxonomic groups across samples. Bar plots showing the relative abundance of the 29 most abundant taxonomic groups across all samples. Taxonomic classification is presented at the genus level, or at a higher level when genus-level assignment was not possible. Unassigned reads were excluded from the analysis, and less abundant groups were combined into a single category labelled ‘Other’. Samples are represented by sample number, histological subtype (E, endometrioid; S, serous) and patient letter.

### 3.5 | Detection of Potential Contaminants

To benchmark diversity metrics and validate sequencing performance, high-biomass vaginal swabs were included as positive controls. This group included 1 Māori and 2 of European (Pākehā) descent. Although variation is seen, the vaginal microbiome is typically dominated by *Lactobacillus* spp., which play a key role in maintaining mucosal health and preventing infection [37]. Compared to uterine samples, vaginal swabs exhibited lower genus richness, lower Shannon and Simpson indices, and greater compositional similarity across samples (Figure S4a). Beta diversity analysis confirmed that vaginal samples formed a distinct cluster, clearly separated from uterine samples, with PCA further supporting this distinction (Figures S4b and S5).

To assess potential contamination, negative controls were included at two critical stages of the workflow: during DNA extraction and processing (processing control) and during library preparation and sequencing (no-template control). The Illumina negative controls yielded insufficient reads for meaningful analysis and were excluded from downstream comparisons.

In contrast, both ONT negative controls generated substantial read counts (506 090 and 250 240 reads, compared with 14 689–1 249 523 reads for ONT uterine swabs), which

permitted detailed taxonomic profiling. Principal component analysis indicated that the no-template control clustered with several uterine samples (Figure S9). Taxonomic profiles of the controls showed distinct signatures: *Enhydrobacter*, *Cutibacterium* and *Acinetobacter*, taxa typically linked to skin or environmental sources, were identified in the sequencing control, while the processing control was dominated by *Lactobacillus*, consistent with possible carryover from vaginal samples processed in parallel. Because these taxa were also observed in some uterine samples, the control profiles were used to guide interpretation of low-abundance signals and to ensure confidence in distinguishing sample-specific from background contributions.

These findings highlight the need to distinguish true biological signals from background contamination in environments with low microbial load. With the implementation of appropriate controls, statistical tools can aid in contaminant identification and removal during analysis [38, 39]. Despite challenges with contamination, ONT sequencing demonstrated clear advantages for microbiome profiling in the low-biomass uterine swabs, providing greater sequencing depth, broader taxonomic coverage, and a higher proportion of reads classified at the genus level, enabling a more detailed and nuanced view of the microbial communities present in the uterine environment.

### 3.6 | Sample Collection and Storage Method Optimisation

We next evaluated sample collection and storage methods to optimise DNA yield and microbial recovery. The swabs used for platform comparison were collected using Method B (Figure 5), which involved immediate lysis of swab contents and storage at room temperature to simplify shipping. While this convenient method was sufficient for sequencing, it consistently yielded low bacterial proportions (Figures S10 and S11) despite having slightly higher DNA yields than other methods, limiting downstream analysis.

To improve sample quality, we tested flocked swabs, which are used clinically and designed to enhance sample uptake and release, particularly in low-biomass environments. In addition to swab type, we assessed two alternative processing approaches. First, we lysed a cell pellet in DNA/RNA shield with ambient storage immediately after collection, which was designed to concentrate microbial DNA and reduce host DNA contamination (Method C). We also assessed freezing the swab immediately after collection in TE buffer, followed by thawing and pellet isolation prior to DNA extraction (Method D). This method was informed by a study suggesting that freezing can preserve microbial integrity and improve DNA recovery in low-biomass samples [40].

Four endometrial tumour swabs from each of five patients were processed using Methods A–D and evaluated using the same quality control criteria applied to the initial cohort. DNA yield and human-to-bacteria ratios, as quantified by ddPCR, were somewhat consistent across all methods (Figures S10 and S11). However, bacterial DNA copy numbers varied between methods and patients (Figure 6 and Figure S11). Methods A and B showed PCR inhibition at the highest 50 ng DNA input. Overall,

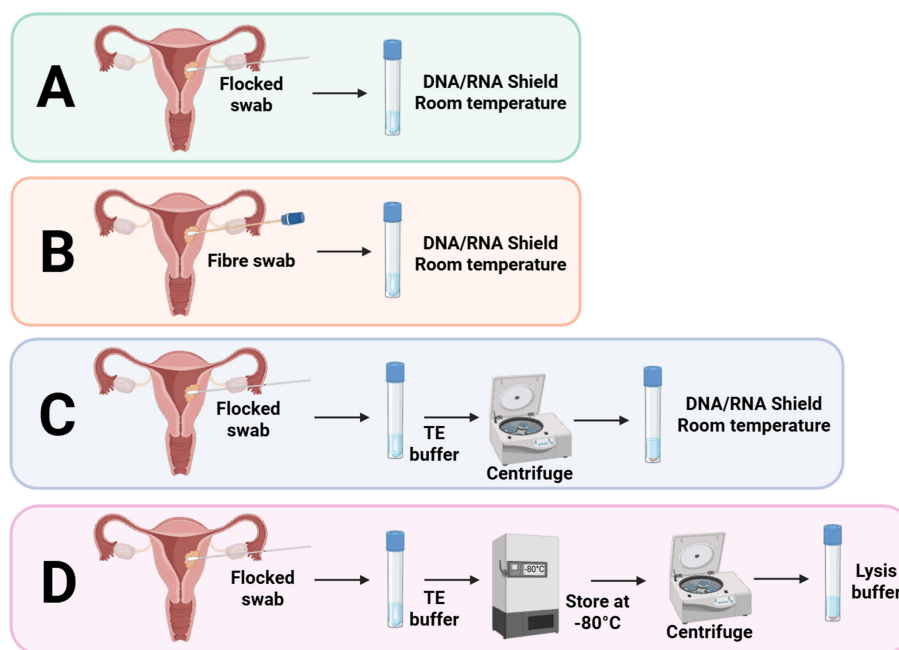
Method D had a higher bacterial copy number, despite lower overall yields of DNA, and no evidence of PCR inhibition at higher input quantities.

To assess the impact of these methods on microbial diversity, all samples were rarefied to 130 000 reads to normalise sequencing depth (Figure S12). Alpha diversity analysis revealed no statistically significant differences between methods in terms of observed genus richness, Shannon, or Simpson indices (Figure 7a). However, Methods C and D showed greater variability in the number of genera identified, suggesting that these approaches may be more sensitive to patient-specific microbial profiles.

Beta diversity analysis using Bray–Curtis dissimilarity showed substantial overlap across all four methods (Figure 7b). Nevertheless, Methods A and B exhibited greater variability between samples, while Method D produced more consistent community profiles across patients. A pairwise PERMANOVA test revealed a statistically significant difference between Methods C and D, with processing method accounting for 22% of the observed variation ( $p = 0.028$ ). Despite these differences, inter-patient variability was the dominant source of diversity, indicating that the choice of processing method had a minimal impact on overall microbial community structure (Figure S13).

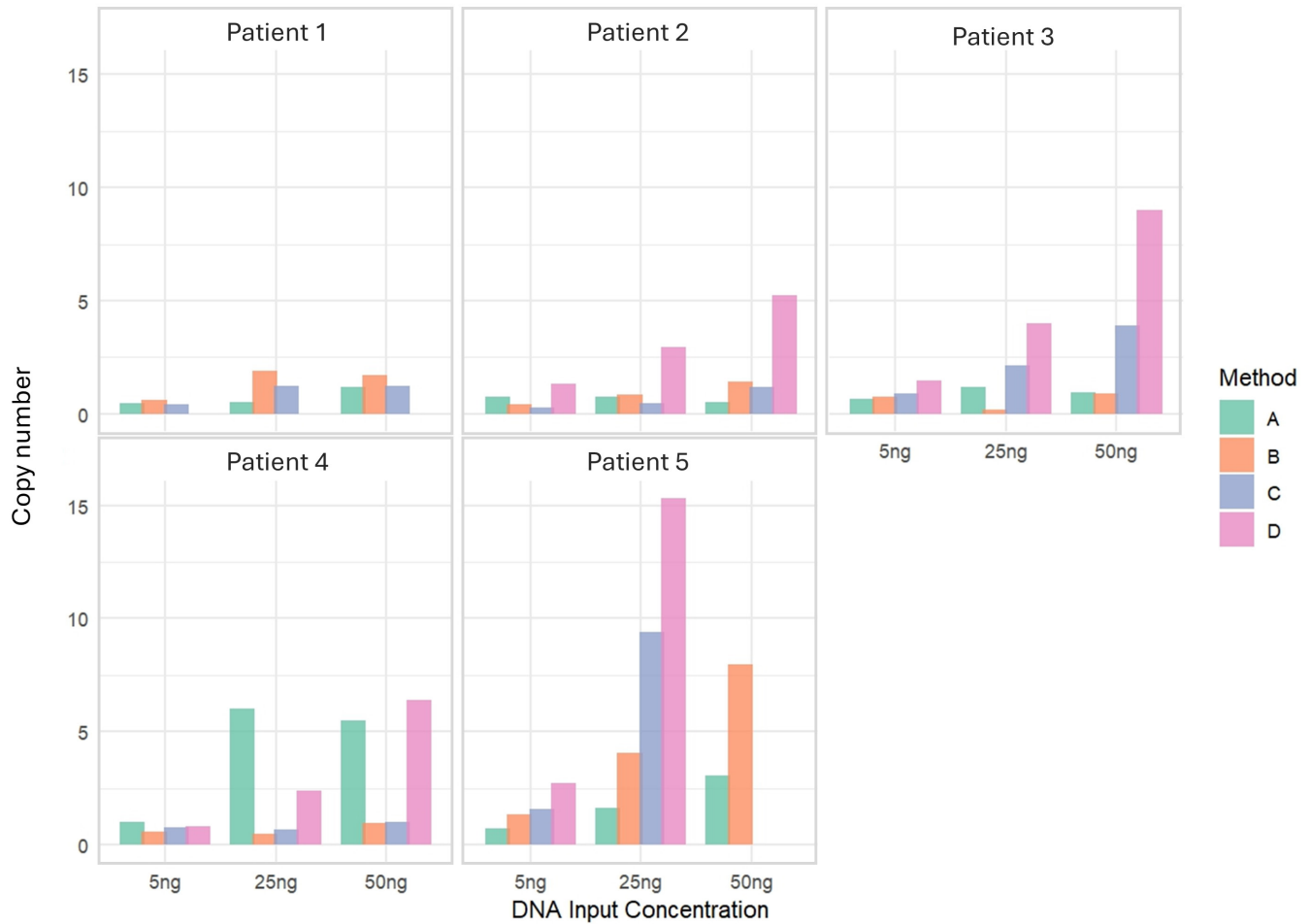
### 3.7 | Taxonomic Composition by Processing Method

To assess the impact of sample processing on microbial community composition, taxonomic profiles were compared across the four methods (A–D) for five patients (Figure 8). While overall

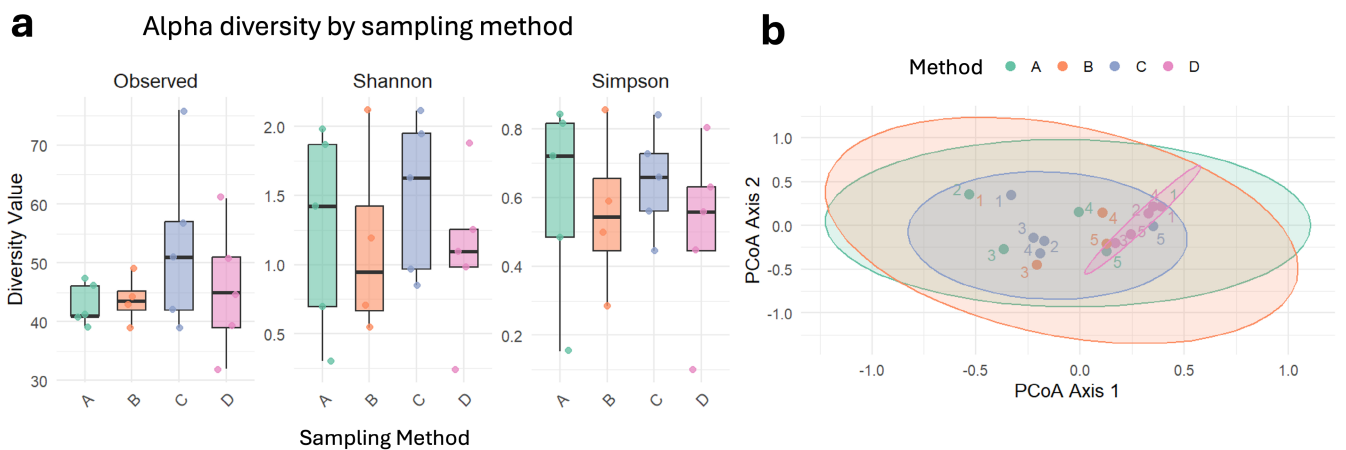


**FIGURE 5** | Overview of sample processing methods for uterine swab microbiome analysis. Illustration summarising the four sample processing methods (A–D) compared in this study for uterine swab microbiome analysis. Each method varies in its approach to sample handling, DNA extraction, and preparation. Figure created using BioRender.

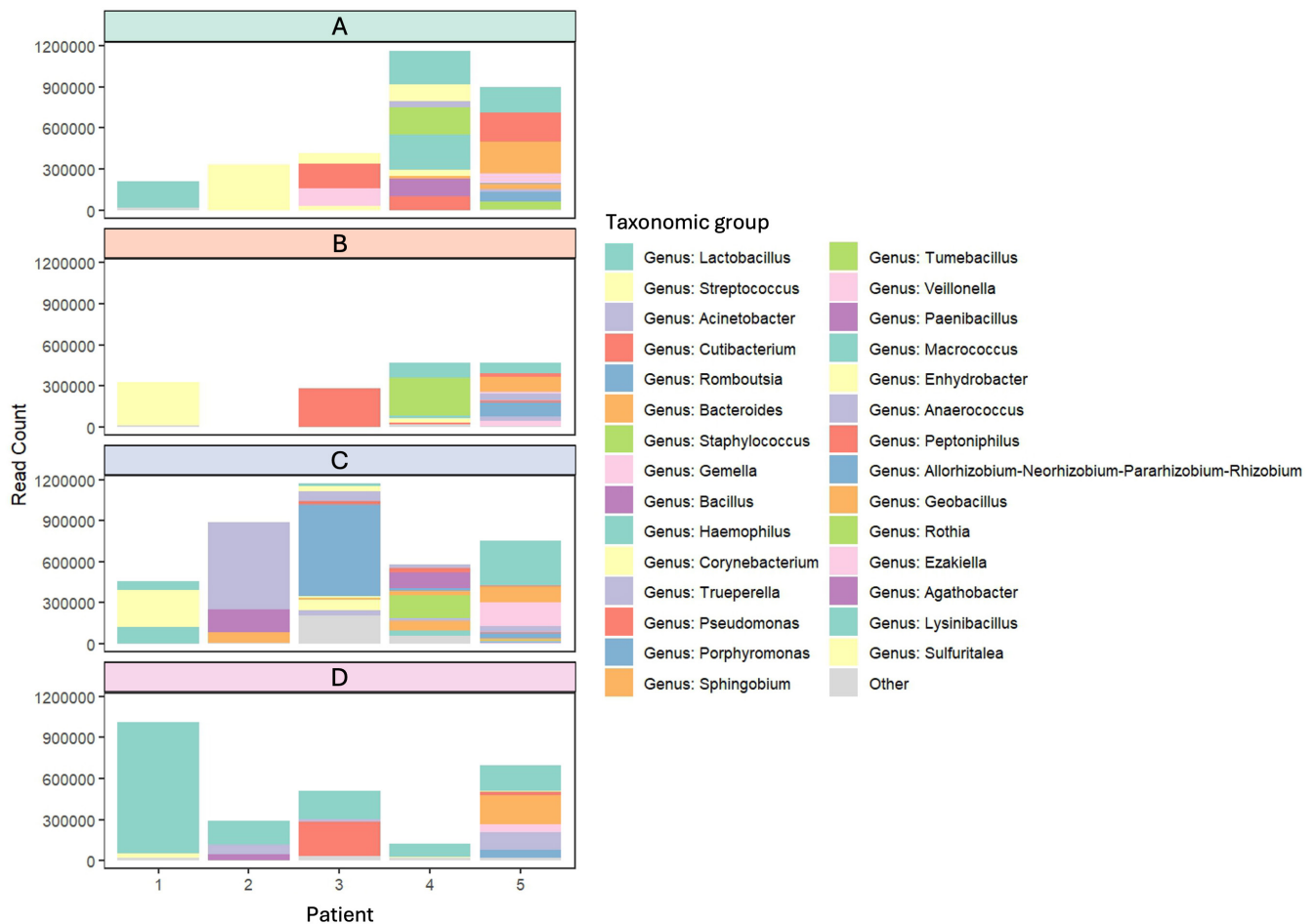
## DNA copy number by method and DNA concentration



**FIGURE 6** | Bacterial 16S rRNA gene copy number across patients and processing methods. Bar plots showing bacterial 16S rRNA gene copy number across samples from patients 1–5, each collected using four processing methods (A–D). For each sample, three DNA input concentrations (5, 25 and 50 ng) were tested using ddPCR. Each bar represents an individual sample. Missing bars are due to insufficient sample for ddPCR to be carried out.



**FIGURE 7** | Microbial diversity across processing methods. (a) Box plots showing alpha diversity metrics—Observed, Shannon, and Simpson indices, based on genus-level data across four processing methods (A–D). Each dot represents an individual sample, with colours corresponding to: Method A: Green; Method B: Orange; Method C: Blue; Method D: Pink. Samples were rarefied to 130000 reads, and unassigned reads were excluded. A Kruskal–Wallis test revealed no statistically significant differences in alpha diversity between processing methods. (b) Principal Coordinate Analysis (PCoA) plots based on Bray–Curtis dissimilarity, comparing microbial community composition across the same four processing methods. Each dot represents an individual sample, labelled by patient number (1–5), and ellipses represent the 95% confidence interval for each method group. Data was generated using ONT. Samples were rarefied as above, and unassigned reads were excluded. Pairwise PERMANOVA tests indicated a significant difference between Method C and Method D ( $R^2 = 0.218$ ,  $p = 0.035$ ).



**FIGURE 8** | Relative abundance of taxonomic groups by patient and processing method. Bar plots showing the relative abundance of the 29 most abundant taxonomic groups within each sample. Taxonomic classification is presented at the genus level, or at a higher level when genus-level assignment was not possible. Unassigned reads were excluded from the analysis, and less abundant taxa were grouped under the category labelled ‘Other’. Samples represent five patients (1–5), each processed using four different methods (A–D). The sample from patient 2, processing method B, was excluded due to low read count.

diversity metrics showed minimal differences between methods, taxonomic assignment revealed method-specific variation in the dominant genera detected.

In Patient 1, Methods A and D consistently identified *Lactobacillus* as the dominant genus, whereas Methods B and C primarily detected *Streptococcus*. In Patient 2, Method A identified *Corynebacterium* and *Streptococcus*, while Methods C and D detected *Acinetobacter*, with Method D also identifying *Lactobacillus*. Patient 3 showed further variation: Methods A, B, and D identified high levels of *Cutibacterium*, whereas Method C was dominated by *Romboutsia*. In contrast, Patient 5 demonstrated strong consistency across all four methods, with each identifying the same dominant genera, noting that this patient overall had a higher bacterial proportion in all swabs.

These findings suggest that while inter-patient variability is the primary driver of microbial composition, processing method can influence taxonomic resolution and detection sensitivity, particularly for low-abundance or method-sensitive taxa. Notably, Method D consistently identified *Lactobacillus* across multiple patients, a pattern not observed with other methods, indicating

that freezing prior to processing may enhance recovery of certain genera.

Overall, visible differences were observed between the four methods in terms of dominant taxa. However, Method D demonstrated a slightly higher bacterial read count and greater consistency in detecting key genera such as *Lactobacillus*. Given the sensitivity of low-biomass microbiome analyses to collection and processing variability, standardisation of methodology is essential. Taken together, Method D, which incorporates freezing prior to processing, was deemed to be optimal for low-biomass uterine swab analysis. It consistently yielded higher bacterial DNA concentrations, produced more stable community profiles, and maintained taxonomic resolution, making it well-suited for future studies of the uterine microbiome.

## 4 | Discussion

This study provides foundational insights into the uterine microbiome of patients with endometrial cancer in New Zealand,

addressing a significant gap in global microbiome research. Australasia remains markedly underrepresented in human microbiome studies, with most data originating from North America, Europe, and parts of Asia. Given that microbiome composition is influenced by geography, lifestyle, and ethnicity [41], findings from international cohorts cannot be directly extrapolated to New Zealand. This underscores the importance of region-specific investigations to uncover potentially unique microbial signatures relevant to disease pathogenesis and biomarker discovery.

#### 4.1 | Sequencing Platform Performance

Our comparison of Oxford Nanopore Technologies (ONT) and Illumina sequencing platforms revealed notable differences in microbial profiles, driven largely by the distinct 16S rRNA regions targeted by each platform. Prior studies have shown that different regions of the 16S gene preferentially detect specific taxa [19, 42]. For example, the V1–V2 regions offer high sensitivity and specificity, while the V3–V4 regions better capture overall diversity. Combinations including the V3 region have demonstrated improved taxonomic resolution [19, 43].

In our study, ONT sequencing of the full-length 16S rRNA gene (V1–V9) revealed greater richness and enabled more precise taxonomic assignments at lower levels (e.g., genus) compared to Illumina, which targeted only the V3 region. This limited region coverage likely contributed to Illumina's reduced ability to assign reads at finer taxonomic levels [44]. This enhanced resolution is particularly valuable in cancer research, where detecting low-abundance or method-sensitive taxa may reveal rare microbial signatures linked to tumour subtype, immune infiltration, or treatment response.

Additionally, the sequencing and bioinformatic pipelines differ significantly between platforms. ONT produced a higher read count, and alpha rarefaction analysis indicated that microbial diversity was not fully captured. In contrast, Illumina curves plateaued before 50 000 reads, suggesting saturation and lower overall richness. Although rarefying reads to equal depths is common practice, doing so here would have compromised the richness captured by ONT. Previous studies have reported mixed findings regarding platform performance, with some favouring Illumina and others ONT [45–47].

Taxonomic assignment criteria also varied. Illumina assigned nearly all reads at the order level, while ONT achieved only ~80% assignment at the phylum level. These differences stem from platform-specific error profiles and assignment algorithms. Illumina uses a naïve Bayes classifier trained on the full 16S gene with a 99% similarity threshold, whereas ONT employs minimap2 with 95% similarity and 90% coverage thresholds, parameters chosen to accommodate ONT's higher error rate [15]. Despite these adjustments, ONT's lower assignment rates at higher taxonomic levels likely reflect residual error and limited correction capabilities [48]. However, ongoing improvements in ONT technology and bioinformatics pipelines continue to enhance accuracy [15, 48].

Overall, ONT provided deeper resolution at lower taxonomic levels and a more comprehensive view of microbial communities

in our uterine samples, consistent with other studies in low-biomass environments [45–47, 49].

#### 4.2 | Species-Level Limitations and Taxonomic Ambiguity

Species-level identification using 16S rRNA sequencing remains challenging, particularly when using the SILVA database, which lacks robust species-level curation [26, 50]. Accordingly, both Epi2me and QIIME2 advise against using SILVA for species-level assignments. Moreover, the 16S gene alone cannot reliably distinguish closely related species, such as those within *Shigella* and *Escherichia* [51]. This limitation affects our ability to identify both clinically relevant species and exclude potential contaminants.

For instance, genera such as *Streptococcus*, *Staphylococcus*, and *Cutibacterium*, previously reported in uterine microbiome studies [52–55], include species commonly found on skin, in soil, or in water. Without species-level resolution, it is difficult to determine whether these represent true uterine residents or contaminants. Similarly, *Porphyromonas somerae* has been implicated in endometrial cancer pathogenesis due to its potential role in promoting inflammation and epithelial disruption [8, 56], but while *Porphyromonas* was detected in most samples, we cannot confirm the presence of *P. somerae* specifically, limiting our ability to explore its association with tumour biology in our cohort without application of additional targeted methodology.

To overcome these limitations, future studies could incorporate whole metagenome shotgun sequencing or metatranscriptomics, which allow for species-level identification and functional profiling, but are more expensive and require more complex analysis pipelines than 16S rRNA analysis. These approaches could enable more precise characterisation of microbial communities and their metabolic or immunological contributions to EC. However, they require substantial sequencing depth and rigorous contamination control, particularly critical in low-biomass environments like the uterus. Statistical decontamination tools such as Decontam may also help distinguish true biological signals from artefactual noise, improving confidence in species-level associations with cancer outcomes [38].

#### 4.3 | Uterine Microbiome Composition

The composition of the uterine microbiome in patients with and without endometrial cancer remains a subject of active investigation and debate, with evidence of global differences that are driven by methodological and cohort-specific factors, or a combination of both. While early studies suggested a dominance of *Lactobacillus* species in the healthy uterus, more recent research has revealed a more diverse and context-dependent microbial landscape, particularly in malignancy.

In our study, *Lactobacillus* was dominant in some samples, consistent with findings from [11, 57, 58], who reported *Lactobacillus crispatus* depletion in EC patients and its anti-proliferative effects on endometrial organoids. However, other samples showed enrichment of taxa such as *Acinetobacter*, *Pseudomonas* and

*Comamonadaceae*, aligning with Winters, Romero [59] findings in EC patients. These genera are often associated with environmental niches and may reflect either true colonisation or contamination, warranting expansion of global uterine datasets to answer this ongoing question.

Beyond these genera, our dataset also detected *Bacteroides*, *Prevotella*, and *Shigella*, which have been previously linked to EC and other gynaecological conditions [9, 10]. *Prevotella* and *Porphyromonas*, in particular, have emerged as recurrent taxa in EC-associated microbiomes. *Porphyromonas somerae*, for example, has been shown to invade endometrial cells and exhibit pathogenic behaviour stimulated by oestrogen exposure, a key risk factor in EC.

Other studies have identified *Anaerococcus*, *Peptoniphilus* and *Buttiarella* as enriched in EC patients, with microbial swapping between vaginal and rectal sites potentially contributing to uterine colonisation [60]. These taxa may influence inflammation, immune modulation, and hormone metabolism, processes central to EC pathogenesis. Moreover, the vaginal microbiome has been shown to correlate with tumour grade and histology, with distinct community state types predictive of low-grade versus high-grade disease [61].

Taken together, our findings are broadly consistent with previous reports, but also highlight the complexity and variability of the uterine microbiome in EC. The presence of both protective (*Lactobacillus*) and potentially pathogenic (*Porphyromonas*, *Prevotella*) taxa suggests that microbial balance may influence disease trajectory. However, distinguishing true residents from contaminants remains a challenge, particularly in low-biomass environments.

To fully characterise the endometrial cancer microbiome in New Zealand and beyond, future studies will require larger, ethnically diverse cohorts, longitudinal sampling, and integration of functional data (e.g., metagenomics, metabolomics). Such efforts could uncover microbial biomarkers for early detection, stratify patients by risk and inform microbiome-targeted interventions in EC care.

#### 4.4 | Contamination Risks and Control Strategies

Contamination remains one of the most significant challenges in low-biomass microbiome studies, where even small amounts of exogenous bacterial DNA can dominate sequencing results, masking true microbial signals and potentially leading to false associations with disease [12]. Erroneous findings can also arise during analysis by mis-annotation of bacterial sequences [62]. This is particularly problematic in cancer research, where microbial signatures are increasingly being explored as diagnostic or prognostic biomarkers. Misidentifying contaminants as disease-associated taxa could mislead biomarker discovery efforts and compromise translational applications, underscoring the importance of rigorous contamination control throughout the experimental workflow [63, 64].

Contamination sources are diverse and include sampling instruments, DNA extraction kits, sequencing reagents, and laboratory

environments. Commercial DNA extraction kits carry their own ‘kit microbiome’, which can introduce consistent but misleading microbial profiles [12, 65]. In the context of endometrial cancer, where microbial loads are low and the clinical implications of microbial shifts are still being defined, even minor contamination can have outsized effects on interpretation. Therefore, a robust contamination testing strategy must incorporate collection-site swabs, DNA extraction blanks, and no-template sequencing controls, each tailored to the specific reagents and protocols used at every stage [39].

Our study incorporated negative controls, including DNA extraction blanks and no-template sequencing controls, to monitor potential contamination at key stages of the ONT workflow. Taxa typically associated with human skin (*Cutibacterium*, *Enhydrobacter*), water (*Acinetobacter*), and vaginal flora (*Lactobacillus*) were occasionally detected in these controls. These observations informed our criteria for interpreting microbial profiles, helping to distinguish genuine signals from possible background contributions [63]. To improve confidence in microbial signals and support robust biomarker discovery, future studies would benefit from broader and stage-specific controls, alongside computational decontamination tools and refined protocols to minimise background interference.

#### 4.5 | Sample Processing Methods and Reproducibility

Variations in sample collection and storage methods contribute to subtle differences in microbial profiles. These could stem from differences in contamination risk during collection and processing, including the use of distinct swab types, storage conditions, laboratory equipment, and environmental exposure [12]. Additionally, methodological differences, such as using the cell pellet (samples C and D) versus the whole sample (A and B), and the impact of freezing, clearly influenced microbial composition in our study.

Freezing can disrupt bacterial cell walls, particularly in Gram-positive species, enhancing DNA extraction efficiency. In faecal microbiome studies, this has been associated with an increased *Firmicutes*-to-*Bacteroidetes* ratio [66, 67]. However, no such trend was observed in our data. While the differences between the four collection methods were not unexpected, they highlight the importance of consistency in DNA collection and storage protocols when comparing across experiments or studies [68].

With relevance to endometrial cancer, reproducibility in microbiome profiling is essential for identifying microbial biomarkers that could inform diagnosis, prognosis, or therapeutic response. Methodological inconsistencies can obscure biologically meaningful patterns and hinder cross-study comparisons. Standardising protocols for low-biomass sample handling will be critical for advancing microbiome research in EC and enabling robust meta-analyses across diverse cohorts.

This pilot study lays the groundwork for future investigations into the uterine microbiome in endometrial cancer, particularly

within underrepresented populations such as those in New Zealand. While we did not aim to fully characterise the microbial landscape associated with endometrial cancer, we successfully established a reproducible protocol for uterine sample collection and processing that improves bacterial DNA recovery and sequencing depth. Our comparative analysis identified ONT as the preferred platform for low-biomass environments, offering superior genus-level resolution and broader taxonomic coverage. Importantly, we demonstrated the necessity of implementing contamination controls throughout the workflow, from sample collection to sequencing, to ensure accurate microbial profiling. This is especially critical in cancer research, where microbial signals may be subtle yet biologically significant.

Ultimately, a clearer understanding of the uterine microbiome may inform diagnostic or therapeutic strategies for endometrial cancer. Microbial signatures could serve as biomarkers for early detection, stratify patients by tumour subtype or risk, or even guide microbiome-targeted interventions. Future research should explore longitudinal sampling, integrate host transcriptomic or immunologic data, and assess microbial function to better understand host–microbiome interactions in endometrial cancer.

#### Author Contributions

Conceptualisation: Cherie Blenkiron, Anastasiia Artuyants and Claire Henry. Methodology: Cherie Blenkiron and Anastasiia Artuyants. Investigation: Anastasiia Artuyants, Sophia Bebelman and Bianca Nijmeijer. Sample and clinical data collection: Claire Henry. Software: Sophia Bebelman. Formal analysis: Sophia Bebelman. Resources: Claire Henry, Sandra Fitzgerald, Anastasiia Artuyants and Cherie Blenkiron. Data curation: Sophia Bebelman. Writing – original Draft: Sophia Bebelman, Anastasiia Artuyants and Cherie Blenkiron. Visualisation: Sophia Bebelman. Supervision: Anastasiia Artuyants and Cherie Blenkiron. Funding acquisition: Anastasiia Artuyants and Cherie Blenkiron.

#### Acknowledgements

This work was supported by the Cancer Research Trust New Zealand, the Cancer Society New Zealand, and the Maurice and Phyllis Paykel Trust. We are deeply grateful to the women who generously participated in this biobanking study and made this research possible. The authors also acknowledge the use of the New Zealand eScience Infrastructure (NeSI) high-performance computing facilities, consulting support, and training services as part of this research. New Zealand's national facilities are provided by NeSI and funded jointly by NeSI's collaborator institutions and through the Ministry of Business, Innovation & Employment's Research Infrastructure programme ([https://www.nesi.org.nz/?utm\\_source=chatgpt.com](https://www.nesi.org.nz/?utm_source=chatgpt.com)). We also thank the team at Auckland Genomics, University of Auckland, for their expertise and assistance with PCR and sequencing. Open access publishing facilitated by The University of Auckland, as part of the Wiley - The University of Auckland agreement via the Council of Australasian University Librarians

#### Funding

This work was supported by the Cancer Research Trust New Zealand, the Cancer Society of New Zealand and the Maurice and Phyllis Paykel Trust. The funders had no involvement in study design, completion, or in the decision to publish this work.

#### Disclosure

AI was not used to generate text or images in this work.

#### Ethics Statement

Ethics approval was obtained from the New Zealand Health and Disability Ethics Committee (HDEC 15/CEN143 and HDEC 2022 EXP 12616) and the University of Otago (H20/002 and H25/0421).

#### Conflicts of Interest

The authors declare no conflicts of interest.

#### Data Availability Statement

The data that support the findings of this study are openly available in the NCBI Sequence Read Archive (SRA) at <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1313862>, reference number PRJNA1313862, and will be available following an embargo from the date of publication. The data are scheduled for public release on 27 September 2026.

#### References

1. H. Sung, J. Ferlay, R. L. Siegel, et al., “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” *CA: A Cancer Journal for Clinicians* 71, no. 3 (2021): 209–249.
2. E. J. Crosbie, S. J. Kitson, J. N. McAlpine, A. Mukhopadhyay, M. E. Powell, and N. Singh, “Endometrial Cancer,” *Lancet* 399, no. 10333 (2022): 1412–1428.
3. O. W. Scott, S. Tin Tin, S. M. Bigby, and J. M. Elwood, “Rapid Increase in Endometrial Cancer Incidence and Ethnic Differences in New Zealand,” *Cancer Causes & Control* 30, no. 2 (2019): 121–127.
4. G. A. Ogunrinola, J. O. Oyewale, O. O. Oshamika, and G. I. Olasehinde, “The Human Microbiome and Its Impacts on Health,” *International Journal of Microbiology* 2020 (2020): 8045646.
5. M. Garvey, “Intestinal Dysbiosis: Microbial Imbalance Impacts on Colorectal Cancer Initiation, Progression and Disease Mitigation,” *Bio-medicine* 12, no. 4 (2024): 740.
6. K. Sekaran, R. P. Varghese, M. Gopikrishnan, et al., “Unraveling the Dysbiosis of Vaginal Microbiome to Understand Cervical Cancer Disease Etiology-An Explainable AI Approach,” *Genes* 14, no. 4 (2023): 936.
7. E. M. Abdelqader, W. S. Mahmoud, H. M. Gebreel, M. M. Kamel, and M. Abu-Elghait, “Correlation Between Gut Microbiota Dysbiosis, Metabolic Syndrome and Breast Cancer,” *Scientific Reports* 15, no. 1 (2025): 6652.
8. T. A. Crooks, J. D. Madison, D. M. Walsh, et al., “*Porphyromonas somerae* Invasion of Endometrial Cancer Cells,” *Frontiers in Microbiology* 12 (2021): 674835.
9. C. Li, Y. Gu, Q. He, et al., “Integrated Analysis of Microbiome and Transcriptome Data Reveals the Interplay Between Commensal Bacteria and Fibrin Degradation in Endometrial Cancer,” *Frontiers in Cellular and Infection Microbiology* 11 (2021): 748558.
10. M. R. Walther-António, J. Chen, F. Multinu, et al., “Potential Contribution of the Uterine Microbiome in the Development of Endometrial Cancer,” *Genome Medicine* 8, no. 1 (2016): 122.
11. Y. Zhao, Y. Liao, G. Xu, and Y. Wang, “Endometrial Microbiota Alteration in Female Patients With Endometrial Polyps Based on 16S rRNA Gene Sequencing Analysis,” *Frontiers in Cellular and Infection Microbiology* 14 (2024): 1351329.
12. S. J. Salter, M. J. Cox, E. M. Turek, et al., “Reagent and Laboratory Contamination Can Critically Impact Sequence-Based Microbiome Analyses,” *BMC Biology* 12 (2014): 87.

13. C. A. Selway, R. Eisenhofer, and L. S. Weyrich, "Microbiome Applications for Pathology: Challenges of Low Microbial Biomass Samples During Diagnostic Testing," *Journal of Pathology. Clinical Research* 6, no. 2 (2020): 97–106.
14. J. S. Johnson, D. J. Spakowicz, B.-Y. Hong, et al., "Evaluation of 16S rRNA Gene Sequencing for Species and Strain-Level Microbiome Analysis," *Nature Communications* 10, no. 1 (2019): 5029.
15. A. Esberg, N. Fries, S. Haworth, and I. Johansson, "Saliva Microbiome Profiling by Full-Genome 16S rRNA Oxford Nanopore Technology Versus Illumina MiSeq Sequencing," *Npj Biofilms and Microbiomes* 10, no. 1 (2024): 149.
16. B. M. Stevens, T. B. Creed, C. L. Reardon, and D. K. Manter, "Comparison of Oxford Nanopore Technologies and Illumina MiSeq Sequencing With Mock Communities and Agricultural Soil," *Scientific Reports* 13, no. 1 (2023): 9323.
17. S. Bejaoui, S. H. Nielsen, A. Rasmussen, et al., "Comparison of Illumina and Oxford Nanopore Sequencing Data Quality for *Clostridioides Difficile* Genome Analysis and Their Application for Epidemiological Surveillance," *BMC Genomics* 26, no. 1 (2025): 92.
18. B. Callahan, K. Sankaran, J. Fukuyama, P. McMurdie, and S. Holmes, "Bioconductor Workflow for Microbiome Data Analysis: From Raw Reads to Community Analyses [Version 1; Peer Review: 3 Approved]," *F1000Research* 5 (2016): 1492.
19. A. Sirichoat, N. Sankuntaw, C. Engchanil, et al., "Comparison of Different Hypervariable Regions of 16S rRNA for Taxonomic Profiling of Vaginal Microbiota Using Next-Generation Sequencing," *Archives of Microbiology* 203, no. 3 (2021): 1159–1166.
20. X. Zhou and K. Faust, "A High-Throughput and Time-Efficient Nanopore Full-Length 16S rRNA Gene Sequencing Protocol for Synthetic Microbial Communities," *Methods* 240 (2025): 14–20.
21. E. Bolyen, J. R. Rideout, M. R. Dillon, et al., "Reproducible, Interactive, Scalable and Extensible Microbiome Data Science Using QIIME 2," *Nature Biotechnology* 37, no. 8 (2019): 852–857.
22. B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. Johnson, and S. P. Holmes, "DADA2: High-Resolution Sample Inference From Illumina Amplicon Data," *Nature Methods* 13, no. 7 (2016): 581–583.
23. N. A. Bokulich, B. D. Kaehler, J. R. Rideout, et al., "Optimizing Taxonomic Classification of Marker-Gene Amplicon Sequences With QIIME 2's q2-Feature-Classifer Plugin," *Microbiome* 6, no. 1 (2018): 90.
24. C. Quast, E. Pruesse, P. Yilmaz, et al., "The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools," *Nucleic Acids Research* 41, no. Database issue (2013): D590–D596.
25. M. S. Robeson, II, D. R. O'Rourke, B. D. Kaehler, et al., "REScripT: Reproducible Sequence Taxonomy Reference Database Management," *PLoS Computational Biology* 17, no. 11 (2021): e1009581.
26. P. Yilmaz, L. W. Parfrey, P. Yarza, et al., "The SILVA and "All-Species Living Tree Project (LTP)" Taxonomic Frameworks," *Nucleic Acids Research* 42, no. Database issue (2014): D643–D648.
27. F. O. Glöckner, P. Yilmaz, C. Quast, et al., "25 Years of Serving the Community With Ribosomal RNA Gene Reference Databases and Tools," *Journal of Biotechnology* 261 (2017): 169–176.
28. BioProject [Internet], *Bacterial 16S Ribosomal RNA RefSeq Targeted Loci Project* (National Library of Medicine (US), National Center for Biotechnology Information, 2008), <https://www.ncbi.nlm.nih.gov/bioproject/33175>.
29. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2023), <https://www.R-project.org/>.
30. J. Oksanen, G. L. Simpson, F. G. Blanchet, et al., *vegan: Community Ecology Package* [R package version 2.6–10], 2025, <https://CRAN.R-project.org/package=vegan>.
31. P. J. McMurdie and S. Holmes, "Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data," *PLoS One* 8, no. 4 (2013): e61217.
32. H. Wickham, R. François, L. Henry, K. Müller, and D. Vaughan, *dplyr: A Grammar of Data Manipulation* [R package version 1.4], 2025, <https://dplyr.tidyverse.org>.
33. A. Kassambara, *rstatix: Pipe-Friendly Framework for Basic Statistical Tests* [R package version 0.7.2], 2023, <https://CRAN.R-project.org/package=rstatix>.
34. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016).
35. P. Martinez Arbizu, *pairwiseAdonis: Pairwise Multilevel Comparison Using Adonis 2017* [R Package Version 0.4.1].
36. M. A. Onstad, R. E. Schmandt, and K. H. Lu, "Addressing the Role of Obesity in Endometrial Cancer Risk, Prevention, and Treatment," *Journal of Clinical Oncology* 34, no. 35 (2016): 4225–4230.
37. W. J. Y. Chee, S. Y. Chew, and L. T. L. Than, "Vaginal Microbiota and the Potential of Lactobacillus Derivatives in Maintaining Vaginal Health," *Microbial Cell Factories* 19, no. 1 (2020): 203.
38. N. M. Davis, D. M. Proctor, S. P. Holmes, D. A. Relman, and B. J. Callahan, "Simple Statistical Identification and Removal of Contaminant Sequences in Marker-Gene and Metagenomics Data," *Microbiome* 6, no. 1 (2018): 226.
39. N. Fierer, P. M. Leung, R. Lappan, et al., "Guidelines for Preventing and Reporting Contamination in Low-Biomass Microbiome Studies," *Nature Microbiology* 10, no. 7 (2025): 1570–1580.
40. A. E. Asangba, J. Chen, K. M. Goergen, et al., "Diagnostic and Prognostic Potential of the Microbiome in Ovarian Cancer Treatment Response," *Scientific Reports* 13, no. 1 (2023): 730.
41. R. J. Abdill, S. P. Graham, V. Rubineti, et al., "Integration of 168,000 Samples Reveals Global Patterns of the Human Gut Microbiome," *Cell* 188, no. 4 (2025): 1100–18.e17.
42. R. López-Aladid, L. Fernández-Barat, V. Alcaraz-Serrano, et al., "Determining the Most Accurate 16S rRNA Hypervariable Region for Taxonomic Identification From Respiratory Samples," *Scientific Reports* 13, no. 1 (2023): 3974.
43. G. Fuks, M. Elgart, A. Amir, et al., "Combining 16S rRNA Gene Variable Regions Enables High-Resolution Microbial Community Profiling," *Microbiome* 6, no. 1 (2018): 17.
44. T. Stoeck, S. N. Katzenmeier, H.-W. Breiner, and V. Rubel, "Nanopore Duplex Sequencing as an Alternative to Illumina MiSeq Sequencing for eDNA-Based Biomonitoring of Coastal Aquaculture Impacts," *Metabarcoding and Metagenomics* 8 (2024): 8.
45. M. Szoboszlai, L. Schramm, D. Pinzauti, J. Scerri, A. Sandionigi, and M. Biazio, "Nanopore Is Preferable Over Illumina for 16S Amplicon Sequencing of the Gut Microbiota When Species-Level Taxonomic Classification, Accurate Estimation of Richness, or Focus on Rare Taxa Is Required," *Microorganisms* 11, no. 3 (2023).
46. I. Biada, M. A. Santacreu, O. González-Recio, and N. Ibáñez-Escriche, "Comparative Analysis of Illumina, PacBio, and Nanopore for 16S rRNA Gene Sequencing of Rabbit's Gut Microbiota," *Frontiers in Microbiomes* 4 (2025).
47. K. Yeo, J. Connell, G. Bouras, et al., "A Comparison Between Full-Length 16S rRNA Oxford Nanopore Sequencing and Illumina V3-V4 16S rRNA Sequencing in Head and Neck Cancer Tissues," *Archives of Microbiology* 206, no. 6 (2024): 248.
48. K. D. Curry, Q. Wang, M. G. Nute, et al., "Emu: Species-Level Microbial Community Profiling of Full-Length 16S rRNA Oxford Nanopore Sequencing Data," *Nature Methods* 19, no. 7 (2022): 845–853.
49. C. Tort-Miró, L. Lorenzo-Rebenaque, L. Montoro-Dasi, et al., "Nanopore Versus Illumina to Study the Gut Bacterial Diversity of Sows

and Piglets Between Farms With High and Low Health Status,” *BMC Veterinary Research* 21, no. 1 (2025): 246.

50. Y.-P. Hsieh, Y.-M. Hung, M.-H. Tsai, L.-C. Lai, and E. Y. Chuang, “16S-ITGDB: An Integrated Database for Improving Species Classification of Prokaryotic 16S Ribosomal RNA Sequences,” *Frontiers in Bioinformatics* 2 (2022).

51. N. K. Devanga Ragupathi, D. P. Muthurandhi Sethuvel, F. Y. Inbanathan, and B. Veeraghavan, “Accurate Differentiation of *Escherichia Coli* and *Shigella* Serogroups: Challenges and Strategies,” *New Microbes New Infections* 21 (2018): 58–62.

52. C. Leoni, O. Ceci, C. Manzari, et al., “Human Endometrial Microbiota at Term of Normal Pregnancies,” *Genes (Basel)* 10, no. 12 (2019): 971.

53. S. M. Miles, B. L. Hardy, and D. S. Merrell, “Investigation of the Microbiota of the Reproductive Tract in Women Undergoing a Total Hysterectomy and Bilateral Salpingo-Oophorectomy,” *Fertility and Sterility* 107, no. 3 (2017): 813–20.e1.

54. I. Moreno, I. Garcia-Grau, D. Perez-Villaroya, et al., “Endometrial Microbiota Composition Is Associated With Reproductive Outcome in Infertile Patients,” *Microbiome* 10, no. 1 (2022): 1.

55. L. Riganelli, V. Iebba, M. Piccioni, et al., “Structural Variations of Vaginal and Endometrial Microbiota: Hints on Female Infertility,” *Frontiers in Cellular and Infection Microbiology* 10 (2020): 350.

56. D. M. Walsh, A. N. Hokenstad, J. Chen, et al., “Postmenopause as a Key Factor in the Composition of the Endometrial Cancer Microbiome (ECbiome),” *Scientific Reports* 9, no. 1 (2019): 19213.

57. Q. Chen, X. Zhang, Q. Hu, W. Zhang, Y. Xie, and W. Wei, “The Alteration of Intrauterine Microbiota in Chronic Endometritis Patients Based on 16S rRNA Sequencing Analysis,” *Annals of Clinical Microbiology and Antimicrobials* 22, no. 1 (2023): 4.

58. J. Liang, M. Li, L. Zhang, et al., “Analysis of the Microbiota Composition in the Genital Tract of Infertile Patients With Chronic Endometritis or Endometrial Polyps,” *Frontiers in Cellular and Infection Microbiology* 13 (2023): 1125640.

59. A. D. Winters, R. Romero, M. T. Gervasi, et al., “Does the Endometrial Cavity Have a Molecular Microbial Signature?,” *Scientific Reports* 9, no. 1 (2019): 9905.

60. N. R. Jimenez, C. R. Herman, P. Łaniewski, et al., “Navigating Complexities of Polymorphic Microbiomes in Endometrial Cancer,” *Npj Biofilms and Microbiomes* 11, no. 1 (2025): 85.

61. H. Hakimjavadi, S. H. George, M. Taub, et al., “The Vaginal Microbiome Is Associated With Endometrial Cancer Grade and Histology,” *Cancer Research Communications* 2, no. 6 (2022): 447–455.

62. A. Gihawi, Y. Ge, J. Lu, et al., “Major Data Analysis Errors Invalidate Cancer Microbiome Findings,” *MBio* 14, no. 5 (2023): e01607-23.

63. I. N. Olomu, L. C. Pena-Cortes, R. A. Long, et al., “Elimination of “Kitome” and “Splashome” Contamination Results in Lack of Detection of a Unique Placental Microbiome,” *BMC Microbiology* 20, no. 1 (2020): 157.

64. K. Aagaard, J. Ma, K. M. Antony, R. Ganu, J. Petrosino, and J. Versalovic, “The Placenta Harbors a Unique Microbiome,” *Science Translational Medicine* 6, no. 237 (2014): 237ra65.

65. Q. R. Ducarmon, B. V. H. Hornung, A. R. Geelen, E. J. Kuijper, and R. D. Zwartink, “Toward Standards in Clinical Microbiota Studies: Comparison of Three DNA Extraction Methods and Two Bioinformatic Pipelines,” *mSystems* 5, no. 1 (2020): e00547-19.

66. F. Fouhy, J. Deane, M. C. Rea, et al., “The Effects of Freezing on Faecal Microbiota as Determined Using MiSeq Sequencing and Culture-Based Investigations,” *PLoS One* 10, no. 3 (2015): e0119355.

67. M. I. Bahl, A. Bergström, and T. R. Licht, “Freezing Fecal Samples Prior to DNA Extraction Affects the Firmicutes to Bacteroidetes Ratio

Determined by Downstream Quantitative PCR Analysis,” *FEMS Microbiology Letters* 329, no. 2 (2012): 193–197.

68. S. V. Jenkins, K. B. Vang, A. Gies, et al., “Sample Storage Conditions Induce Post-Collection Biases in Microbiome Profiles,” *BMC Microbiology* 18, no. 1 (2018): 227.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figure S1:** Taxonomic assignment of sequencing reads across platforms and databases. Violin plots showing the percentage of reads assigned at each taxonomic level for samples sequenced using ONT and Illumina platforms. Taxonomic classification was performed using both the NCBI and SILVA databases. Each dot represents an individual sample ( $n = 38$ ). Reads classified as unknown, unassigned, eukaryotic, or uncultured were excluded from the assigned read count. **Figure S2:** Alpha rarefaction analysis of Illumina and ONT sequencing data. Alpha rarefaction plots showing genus-level microbial diversity for (a) Illumina and (b) Oxford Nanopore Technologies datasets. The dotted line indicates the read depth used for diversity analysis: 10000 reads for Illumina and 140000 reads for ONT. **Figure S3:** Comparison of alpha and beta diversity metrics by taxonomic database. (a) Violin plots showing alpha diversity metrics—Observed, Shannon, and Simpson indices, based on genus-level data. Taxonomy was assigned using either the NCBI (light blue) or SILVA (dark blue) databases. Each dot represents an individual sample. Samples were rarefied to 10000 reads for Illumina and 140000 reads for ONT; unassigned reads were excluded. Paired Wilcoxon tests revealed increased observed diversity in the ONT samples when using SILVA compared to ncbi (adjusted  $p = 0.0000431$ ). There were no other statistically significant differences in diversity metrics between databases within the same sequencing platform. (b) Principal Coordinate Analysis (PCoA) plots based on Bray–Curtis dissimilarity, comparing genus-level taxonomic assignments using NCBI (light blue) and SILVA (dark blue). Each dot represents an individual sample, with dotted lines connecting the same sample classified by both databases. Ellipses represent the 95% confidence interval for each database group. Samples were rarefied as above, and unassigned reads were excluded. PERMANOVA tests indicated significant differences in community composition between databases: ONT:  $F = 8.19$ ,  $R^2 = 0.102$ ,  $p = 0.001$  (999 permutations); Illumina:  $F = 11.2$ ,  $R^2 = 0.138$ ,  $p = 0.001$ . **Figure S4:** Comparison of microbial diversity by sample site. (a) Violin plots showing alpha diversity metrics—Observed, Shannon, and Simpson indices, based on genus-level data, comparing samples from the uterus (blue,  $n = 38$ ) and vagina (red,  $n = 3$ ). Each dot represents an individual sample. Samples were rarefied to 140000 reads, and unassigned reads were excluded. Statistical significance was assessed using the Wilcoxon test:  $p < 0.05$  (\*),  $p < 0.01$  (\*\*). (b) Principal Coordinate Analysis (PCoA) plots based on Bray–Curtis dissimilarity, comparing microbial communities between uterus (blue,  $n = 38$ ) and vagina (red,  $n = 3$ ) swabs at the genus level. Each dot represents an individual sample, labelled by the patient letter. The ellipse indicates the 95% confidence interval for uterus samples. Samples were rarefied as above, and unassigned reads were excluded. PERMANOVA test results:  $F = 3.08$ ,  $R^2 = 0.07491$ ,  $p = 0.003$  (999 permutations). **Figure S5:** Principal component analysis of ONT samples by sample site. Principal Component Analysis (PCA) plot based on Hellinger-transformed genus-level data from ONT sequencing. Samples were derived from either the uterus (blue,  $n = 38$ ) or vagina (red,  $n = 3$ ). Uterus samples are labelled by the patient letter. **Figure S6:** Taxonomic composition of samples based on read abundance. Bar plots showing the read counts of the 29 most abundant taxonomic groups within each sample. Taxonomic classification was performed at the genus level, or at a higher taxonomic level when genus-level assignment was not possible. Taxa with lower abundance were grouped into a single category labelled ‘Other’ to simplify visualisation. **Figure S7:** Comparison of taxonomic composition across sequencing platforms and databases. Bar plots showing the read counts of the 29 most abundant taxonomic groups within each sample. Samples sequenced using ONT are displayed on the left, and those sequenced using Illumina on the right. Taxonomic

classification was performed using both the SILVA and NCBI databases. For each sample, the read count (top panels) and relative abundance (bottom panels) of taxonomic groups are shown. Taxa with lower abundance were grouped into a single category labelled 'Other'. Taxonomic groups are presented at the genus level, or at a higher level when genus-level assignment was not possible. Unassigned reads were excluded from the analysis. **Figure S8:** Comparison of alpha and beta diversity between Illumina and ONT sequencing platforms at the genus level. (a) Alpha Diversity Comparison: Violin plots of Observed, Shannon, and Simpson indices at the genus level, comparing Illumina (light blue) and ONT (dark blue) sequencing platforms. Each dot represents an individual sample. Samples were rarefied to 10000 reads, excluding unassigned reads. Statistical significance was assessed using paired Wilcoxon rank-sum tests ( $*p < 0.05$ ,  $**p < 0.01$ , adjusted for multiple comparisons). (b) Beta Diversity (Bray-Curtis PCoA): Principal coordinate analysis of Bray-Curtis dissimilarity at the genus level, comparing Illumina and ONT platforms. Each dot represents a sample, labelled by the patient letter. Dotted lines connect matched samples sequenced on both platforms. Ellipses indicate 95% confidence intervals. Samples were rarefied as above, excluding unassigned reads. PERMANOVA yielded  $F = 10.057$ ,  $R^2 = 0.123$ ,  $p = 0.001$  (999 permutations). **Figure S9:** Hellinger-transformed PCA of uterus samples and negative controls. Principal Component Analysis (PCA) plot based on Hellinger-transformed genus-level data from ONT sequencing. The plot compares uterus samples collected for platform comparison (blue,  $n = 38$ ) with two types of negative controls: no template control (green) added during the sequencing process and processing control (purple) added prior to DNA isolation and quantification. Uterus samples are labelled by patient letter. **Figure S10:** Comparisons of endometrial swab DNA output between different collection and storage methods (A–D). (a) The ratio of bacteria to human DNA copies using ddPCR. (b) The total DNA concentration in each sample. **Figure S11:** Summary of the average number of reads assigned to genus (from ONT data) and copies of bacterial DNA isolated using processing and storage methods A–D. **Figure S12:** Alpha rarefaction analysis of ONT samples by collection and storage method. Alpha rarefaction plots showing genus-level microbial diversity in ONT-sequenced samples used to compare different collection and storage methods. The dotted line at 130000 reads indicates the read depth used for diversity analysis. **Figure S13:** Microbial diversity across patients using ONT sequencing. (a) Box plots showing alpha diversity metrics—Observed, Shannon, and Simpson indices, based on genus-level data for samples collected from patients 1–5. Each dot represents an individual sample. Samples were rarefied to 130000 reads, and unassigned reads were excluded. Statistical significance was assessed using the Kruskal-Wallis test, followed by Dunn's post hoc test.  $p < 0.05$  (\*) adjusted for multiple tests. (b) Principal Coordinate Analysis (PCoA) plots based on Bray-Curtis dissimilarity, comparing microbial community composition across patients 1–5 at the genus level. Each dot represents an individual sample, labelled by the method used for sample processing and storage (A–D). Ellipses represent the 95% confidence interval for each patient group, excluding patient 2, which had too few samples for ellipse calculation. Samples were rarefied as above, and unassigned reads were excluded. Pairwise PERMANOVA tests revealed significant differences between: Patients 1 and 5 ( $R^2 = 0.398$ ,  $p = 0.029$ ), Patients 2 and 5 ( $R^2 = 0.410$ ,  $p = 0.035$ ), Patients 3 and 5 ( $R^2 = 0.430$ ,  $p = 0.029$ ), Patients 4 and 5 ( $R^2 = 0.310$ ,  $p = 0.028$ ). **Table S1:** Quality control of ONT and Illumina sequencing samples. Sample codes follow the format: sample number—pathology—patient code. Pathology types are categorised as: E: Endometrioid; S: Serous; U: Unknown; N: Non-cancerous (vaginal swabs from patients with endometriosis). Patient codes: A–Z: Paired samples tested on both ONT and Illumina; 1–5: Samples used to compare collection methods using ONT only. These include an additional letter (A–D) indicating the sample collection and processing method. Samples were considered passed if they yielded more than 10000 reads of acceptable quality. Failed samples were excluded from downstream analysis.