scientific reports

Check for updates

OPEN

Profiling of vaginal microbial communities in Chilean women via self-sampling and nanopore sequencing

Barbara Oliva-Arancibia[1,2], Pablo Villanueva[1,2] & Juan A. Ugalde[1]✉

The vaginal microbiota is a dynamic ecosystem that plays a central role in women's reproductive health, with *Lactobacillus* species typically dominating in healthy individuals. Disruptions to this microbiota can lead to dysbiosis, increasing susceptibility to infections such as bacterial vaginosis and contributing to adverse reproductive outcomes. Despite growing global interest in vaginal microbiome research, studies in Latin American populations remain limited, particularly in Chile. In this exploratory study, we aimed to characterize the vaginal microbiota of Chilean women using a combination of self-sampling and full-length 16S rRNA gene sequencing with Oxford Nanopore Technologies (ONT). We implemented a customized bioinformatic pipeline using the Emu classifier to achieve species-level taxonomic resolution. Our analysis revealed that most participants had microbiota dominated by *Lactobacillus* species, while a subset exhibited diverse microbial communities consistent with Community State Type (CST) IV. Alpha and beta diversity metrics supported these classifications and demonstrated that species-level resolution enhances the ability to distinguish CSTs. These findings provide the first species-level characterization of vaginal microbiota in Chilean women and demonstrate the utility of ONT sequencing and self-sampling in underrepresented populations. This work contributes essential baseline data for the region and highlights the importance of inclusive microbial profiling in advancing personalized approaches to women's health.

The vaginal microbiota is a dynamic and complex ecosystem composed of diverse microbial species that directly influence women's overall and reproductive health. This environment, dominated by members of the *Lactobacillus* genus, plays a crucial role in protecting the epithelial mucosa of the vaginal tract by maintaining an acidic environment (pH 3.3–4.5), which inhibits the growth of pathogens and prevents viral infections[1,2]. The balance between commensal and pathogenic species is essential to maintain a healthy state, preventing infections and other gynecological complications.

Lactobacillus species metabolize glycogen from vaginal epithelial cells, promoting their proliferation and the production of lactic acid, which is essential for maintaining an acidic environment[1]. Furthermore, these bacterial species secrete antimicrobial compounds, including hydrogen peroxide and bacteriocins, which further enhance vaginal defense mechanisms[3]. The most frequently reported *Lactobacillus* species include *L. crispatus*, *L. gasseri*, *L. jensenii,* and *L. iners*[4].

Disruption of a normal healthy microbiome can lead to a dysbiotic state, increasing the risk for conditions such as bacterial vaginosis (BV)[5], vulvovaginal candidiasis[6], and sexually transmitted infections, such as Chlamydia and HIV[7,8]. BV is the most common condition, which is characterized by decreased *Lactobacillus* abundance and an increase in anaerobic bacteria, such as *Gardnerella vaginalis*, *Fannyhessea vaginae*, and *Prevotella* species[9]. While symptoms include foul discharge and itching, many women are asymptomatic. In addition, BV raises the risk of serious infections and pregnancy complications[10].

Traditionally, diagnosis of vaginal dysbiosis depends on methods like Amsel's criteria, which assess clinical symptoms such as high vaginal pH and microscopy[11], and the Nugent method, which uses Gram staining to evaluate specific bacterial morphotypes[12]. However, these techniques have their limitations as they rely on human interpretation and do not provide an adequate bacterial identification[13].

Recent advances in sequencing technologies have allowed us to overcome these limitations. 16S rRNA gene sequencing is a powerful tool for studying microbial communities, as it enables bacterial identification without

[1]Center for Bioinformatics and Integrative Biology, Facultad de Ciencias de la Vida, Universidad Andrés Bello, Republica 330, 8370186 Santiago, Chile. [2]Barbara Oliva-Arancibia and Pablo Villanueva contributed equally to this work. ✉email: juan@ugalde.bio

the need for cultures[14]. However, most studies have been conducted using short-read sequencing platforms that do not span the entire 16S rRNA gene, limiting species-level resolution[15]. In the vaginal microbiome, studies commonly target specific hypervariable regions of the 16S rRNA, particularly the V3–V4 region, which provides genus-level resolution but often falls short in discriminating closely related species (e.g., within the *Lactobacillus* genus)[16,17]. The introduction of long-read sequencing platforms, such as Oxford Nanopore (ONT), has opened new possibilities for studying the vaginal microbiota by completely sequencing the full 16S rRNA gene, thus enabling species-level identification[18–20].

Community State Types (CSTs) are used to group vaginal microbiome profiles based on their bacterial composition. Originally defined by Ravel et al.[23] through clustering and later refined by France et al.[24] using a nearest-centroid method. More recently, it has been revised to emphasize the dynamic and functional aspects of these community states[25]. CSTs I, II, III, and V are dominated by *Lactobacillus* species: *L. crispatus* (CST I), *L. gasseri* (CST II), *L. iners* (CST III), and *L. jensenii* (CST V). CST I, dominated by *L. crispatus*, is associated with a healthy vaginal environment due to its high lactic acid production and ability to inhibit pathogen colonization[26]. *L. iners*, dominant in CST III, has a context-dependent role found in both healthy and dysbiotic states, raising questions about its protective function[9]. CST IV lacks a dominant *Lactobacillus* species and shows higher diversity of anaerobic bacteria, such as *Gardnerella*, *Prevotella*, and *Megasphaera*[10,27–29].

Although widely used, CSTs have key limitations: They reduce the vaginal microbiota's dynamic nature into static categories, potentially missing individual and temporal variability[25]. Their classification relies on sequencing resolution and database accuracy, which may overlook low-abundance but clinically relevant taxa[30]. Additionally, CSTs are descriptive and do not reflect microbial functions or host interactions, so their biological relevance should be interpreted cautiously and supported by functional and clinical data.

Despite advancements in microbial profiling technologies, most vaginal microbiota studies focus on North American and European populations, leaving gaps in knowledge about other groups, like Latin Americans. In Chile, research on the vaginal microbiota mainly relies on standard methods like the Amsel criteria. A survey of 100 women at a family health center, most of them asymptomatic, found that 32% of them had bacterial vaginosis (BV)[31]. Another study with 101 women using wet mount and Gram stain reported 46.5% of women with vaginal dysbiosis, including 16.8% with BV and 11.9% with vulvovaginal candidiasis[32]. These studies highlight the high prevalence of vaginal health conditions in Chilean women but are limited by their reliance on conventional diagnostic methods. There is an urgent need for advanced microbial profiling to better understand the vaginal microbiota in Chilean women and improve diagnosis, prevention, and treatment. Such research will help to identify region-specific microbial profiles and health risks, impacting public health. Only recently, novel initiatives in the region have examined the association between the vaginal microbiota and HPV in Latina women[33].

To address the need for more comprehensive profiling of vaginal microbiota in Chilean women, this study presents the development and application of a self-sampling method for vaginal swabs, along with a customized bioinformatic pipeline for estimating bacterial relative abundance. Using full-length 16S rRNA sequencing on the ONT platform, we achieved species-level taxonomic profiles, estimated species abundance, and classified samples into CSTs based on the microbiome composition. These demonstrate the potential of self-sampling and long-read sequencing to enhance our understanding of vaginal health, particularly in underrepresented populations. In doing so, we help address a geographic and demographic gap in vaginal microbiome research and provide a reference point for future studies in Latin America.

Methods

Implementation of a self-sampling method for vaginal samples

Criteria and study cohort

This study was approved by the Bioethics Committee of the Faculty of Life Sciences, Universidad Andrés Bello (Ethics Approval ID 009/2023). All the methods were performed following relevant guidelines and regulations. Twenty participants were recruited based on predefined criteria. Chilean women between the ages of 22 and 36 who were able to provide informed consent were eligible for inclusion. Participants were excluded if they were menstruating at the time of sampling, pregnant, or had used antibiotics in the past 30 days. Additionally, all participants completed a detailed questionnaire covering demographic information, employment status, health habits, dietary patterns, hygiene practices, antibiotic use, and sexual history (Supplementary Table S1).

Vaginal self-sampling protocol

The collection procedure was explained in detail to participants to ensure the consistency and reliability of the samples. Each participant was individually instructed to wash their hands before the procedure and perform an external vaginal cleansing using toilet paper moistened with water, wiping from the vulva towards the anus to prevent cross-contamination. For self-collection, participants were instructed to use sterile swabs in the DNA/RNA Shield Collection Tube with Swab (Zymo Research, California, USA). Participants were asked to open the sterile container and hold the swab using the marked indentation. Using their free hand, they were instructed to gently separate the skin folds around the vaginal opening and insert the swab approximately 5 cm into the vaginal canal. They were then instructed to rotate the swab 10 times to ensure adequate sample collection and carefully withdraw it, avoiding contact with external surfaces. Immediately after collection, they were asked to place the swab into the Shield buffer tube and store it at 4 °C until further processing.

DNA extraction and 16S rRNA gene sequencing

Genomic DNA extraction and library preparation

DNA was extracted from the self-collected vaginal samples using the Quick-DNA Miniprep Plus Kit (Zymo Research, California, USA), following the manufacturer's instructions. After a 2-min incubation at room

temperature, DNA was eluted in 50 µL of nuclease-free water. Concentration was measured using fluorometry (Qubit 4.0, Thermo Fisher Scientific, USA), while purity was determined by A260/A280 ratio using a Synergy spectrophotometer (Agilent, Biotek, California, USA). Libraries were prepared using the 16S Barcoding Kit (SQK-16S024, Oxford Nanopore Technologies, UK) to amplify the full-length 16S rRNA gene (1500 bp). Amplification was performed with genomic DNA (30 ng), universal primers 27F: 5′-AGAGAGTTTGATCMTGGCTCAG-3′ and 1492R: 5′-CGGTTACCTTGTTACGACTT-3′, which are flanked by additional sequences (barcodes) that differ for each sample, and LongAmp Hot Start Taq Master Mix (New England Biolabs, USA). PCR conditions are detailed in Supplementary Table S2.

Each sample was transferred to a 1.5 ml DNA LoBind Axygen tube and purified individually before pooling. Magnetic beads (Genosur, Chile) were resuspended by vortexing, and 30 µl were added to each sample and mixed by pipetting. After a 5-min incubation on an orbital shaker, two washes were performed with 80% ethanol, carefully removing the supernatant without disturbing the pellet. The beads were then resuspended in 10 µl of 10 mM Tris–HCl, 50 mM NaCl, and incubated for 2 min at room temperature. A total of 10 µl of eluate was recovered, and the beads were discarded.

Libraries were quantified using fluorometry (Qubit 4.0, Thermo Fisher Scientific, USA). Barcoded libraries were pooled to a final concentration of 70 fmoles in 10 µl of 10 mM Tris–HCl, 50 mM NaCl (pH 8). Next, 1 µl of Rapid Adapter (RAP) was added, mixed gently, and incubated for 5 min at room temperature. The prepared and purified libraries were incubated on ice until loading into the flow cell.

16S rRNA gene sequencing

The prepared libraries were mixed with 15 µl of the Sequencing Buffer (SB II), 10 µl Loading Beads (LB II), and nuclease-free water. The final adjusted sample was loaded into the flow cell R.9.4.1 (FLO-FLG001) and sequenced on the MinION Mk1B device (ONT, UK), which was connected to a desktop computer (Ryzen 9, 32 GB RAM, RTX 3060). In this study, 10 samples were loaded at a time, and each sequencing session lasted about 24 h. DNA sequencing data were acquired in POD5 format using the MINKNOW (version 24.06.16) software.

Bioinformatic analysis

Sequenced data processing

Dorado (version 7.4.14) was used for basecalling and demultiplexing with the following settings: High accuracy model 450 bps, Barcode Kit SQK-PBK004, and trim_barcodes = on. The resulting sequence data were generated as FASTQ files. These files underwent length filtering using fastp[34], with trimming and quality filtering options disabled. Reads were retained if their lengths ranged between 1400 and 1700 base pairs (–length_required 1400–length_limit 1700) for the expected full 16S rRNA gene length.

Vaginal mock community generation and analysis

A vaginal mock community was simulated for database comparisons using the genome sequences of some of the most common taxa reported in vaginal communities. Genome sequences were downloaded from the NCBI genome database[35], and details of each assembly are provided in Table 1. All genomes were concatenated and used as a reference genome for Grinder (version 0.5.3)[36] to simulate 16S amplicons. Standard full-length 16S primer sequences were applied for the forward_reverse parameter (27F: 5′-AGAGTTTGATYMTGGCTCAG-3′; 1492R: 5′-TACGGYTACCTTGTTACGACTT-3′), with 1000 reads specified for the total_reads parameter. The resulting error-prone reads were then processed using CuRe-Sim-LoRM[37] to introduce ONT-associated errors, using the tool's default parameters.

The obtained reads were analyzed using Emu (version 3.4.5)[38] with three databases. These included the default Emu database (version 3.4.5), which integrates rrnDB version 5.6[39] and NCBI 16S RefSeq[40]; the pre-built Silva database (version 138.1)[21]; and a custom GTDB database adapted for Emu. The custom GTDB database was constructed using GTDB release 220[22] and the Emu build-database option. Taxonomic assignments were performed using the Emu abundance function for each database. The relative abundance output tables were subsequently analyzed and compared against expected relative abundance values using R (version 4.2.2).

Organism scientific name	assembly accession	Size
Ureaplasma parvum ATCC 27815	GCA_000019345.1	751,679
Prevotella bivia DSM 20514	GCA_000262545.1	2,521,238
Lactobacillus gasseri 2016	GCA_000439915.2	1,892,784
Streptococcus agalactiae	GCA_001552035.1	2,079,123
Lactobacillus jensenii	GCA_001936235.1	1,672,949
Gardnerella vaginalis	GCA_002861965.1	1,675,681
Anaerococcus hydrogenalis	GCA_002871895.1	1,908,330
Lactobacillus iners	GCA_009556455.1	1,363,099
Lactobacillus crispatus	GCA_009769205.1	2,039,404
Bifidobacterium breve	GCA_024665435.1	2,336,700
Megasphaera paucivorans	GCA_900103535.1	2,912,276

Table 1. Genome sequences used for vaginal 16S mock community.

Taxonomic profiling and data analysis

Filtered reads were analyzed per sample to determine the taxonomic relative abundance of 16S rRNA gene sequences using Emu (version 3.4.5)[15]. The analysis was conducted using the Emu abundance function, with all parameters set to default. The default Emu database (version 3.4.5, rrnDB[39], and NCBI 16S RefSeq[40]) was chosen based on its performance. The relative abundance output tables were imported into R and converted into phyloseq objects using the *phyloseq* package (version 1.42.0)[41]. Taxonomic data were aggregated at the genus or species level using the tax_glom() function in phyloseq. Filtered and processed data were visualized using the *ggplot2* package (version 3.5.1)[42]. Alpha diversity and beta diversity (Bray–Curtis PCoA distance) for each sample were calculated using the plot_richness() and ordinate() functions in the *phyloseq* package. Beta diversity differences among CST groups were tested using PERMANOVA with the *adonis2()* function (999 permutations) and evaluated for homogeneity of group dispersions using PERMDISP with the *betadisper()* and *permutest()* functions of the vegan package.

Results

In silico validation of 16S rRNA gene metataxonomic analysis

Vaginal mock community and database selection

To select the most suitable database for this dataset, a comparison of different 16S rRNA gene taxonomic databases was conducted using an in silico-generated vaginal mock community. Following information from the literature[23], the mock community was designed to represent the most common bacterial taxa found in vaginal samples across all CSTs. A total of 1,000 in silico 16S rRNA gene amplicons were simulated using Grinder (version 0.5.3)[36] with its in silico PCR option, employing genomic DNA sequences from the selected taxa and standard full-length 16S rRNA primer sequences. Of the 11 taxa chosen initially, only 7 were successfully amplified under these simulation conditions; amplification failed for the strains *Gardnerella vaginalis*, *Ureaplasma parvum* ATCC 27815, *Bifidobacterium breve*, and *Megasphaera paucivorans*. The estimated relative abundances of the amplicons for each taxon are summarized in Table 2.

To mimic the error profile of Oxford Nanopore Technologies (ONT) long-read sequencing, we introduced errors into the in silico simulated amplicons using CuReSim-LoRM[37], a simulator that generates realistic ONT-like read errors based on empirical models. The resulting reads constituted the final vaginal mock community. Considering the benchmarking performance of the Emu tool[15], this method was selected for 16S taxonomic assignment. Three 16S rRNA databases were evaluated with this tool using the generated mock community: the Emu database, the SILVA database version available for Emu, and a GTDB database version adapted for Emu. Relative abundances obtained for the mock community at the genus level for each database are shown in Table 3 and graphically represented in Fig. 1. The estimated relative abundances from all databases were close to the expected values, with minor deviations attributable to simulated sequencing errors or taxonomic misassignments. Among the three databases, the Emu database demonstrated the highest concordance with expected abundances, showing relative abundances closest to the expected values, the lowest mean absolute error (0.0116), and misassigning only two genera.

The same analysis was conducted at the species level for the three databases, with a summarized version of the results presented in Table 4. In all cases, the relative abundances closely align with the expected values; however, certain species are either underrepresented or overrepresented. For instance, *Lactobacillus crispatus* is consistently overrepresented across all databases, while *Prevotella bivia* is underrepresented. The SILVA database showed a significant limitation, with five taxa being assigned only at the genus level rather than resolving to the species level, a drawback not observed with the other databases. In contrast, the GTDB and Emu databases demonstrated better species-level resolution but exhibited higher misassigned species. Specifically, the GTDB database misassigned 14 species, while the Emu database misassigned 11 species, with the latter showing lower relative abundances for these misassignments. Considering the importance of resolving taxa at the species level for subsequent analyses and minimizing taxonomic assignment errors in data comparable to the generated mock community, the Emu database was selected for further analyses.

Given the clinical relevance of *G. vaginalis*, the in silico PCR simulation was repeated using a previously described forward degenerate primer (10F:5′-RGTTYGATYCTGGCTCAG-3′) known to amplify this species[43]. Using the modified primer, the 16S rRNA gene of nine strains—including *G. vaginalis*—was successfully amplified and correctly assigned by Emu, confirming that the lack of detection in the initial analysis was due

Taxa	Amplified seq_id
Prevotella bivia	NZ_JH660658.1
Prevotella bivia	NZ_JH660660.1
Lactobacillus crispatus	NZ_CP039266.1
Lactobacillus jensenii	NZ_CP018809.1
Streptococcus agalactiae	NZ_CP012480.1
Anaerococcus hydrogenalis	NZ_PNHP01000002.1
Lactobacillus iners	NZ_CP045664.1
Lactobacillus gasseri	NZ_CP090409.1

Table 2. Species present in *in-silico* mock community. All in silico amplicons are in an equal relative abundance of 12.5% in the mock community.

Genus	Expected	Silva	GTDB	EMU
Anaerococcus	0.125	0.140	0.141	0.140
Lactobacillus	0.5	0.488	0.488	0.494
Prevotella	0.25	0.233	0.232	0.232
Streptococcus	0.125	0.134	0.134	0.131
Geobacillus	–	0.001	0	0
Limosilactobacillus	–	0.002	0	0
UCG-004	–	0.001	0	0
Akkermansia	–	0	0.001	0
Bifidobacterium	–	0	0.002	0
Ligilactobacillus	–	0	0.001	0
Bacillus	–	0	0	0.001
Lactiplantibacillus	–	0	0	0.001
Total miss assigned	–	0.004	0.004	0.002

Table 3. Relative genus abundances of in silico vaginal mock community estimated with different databases.

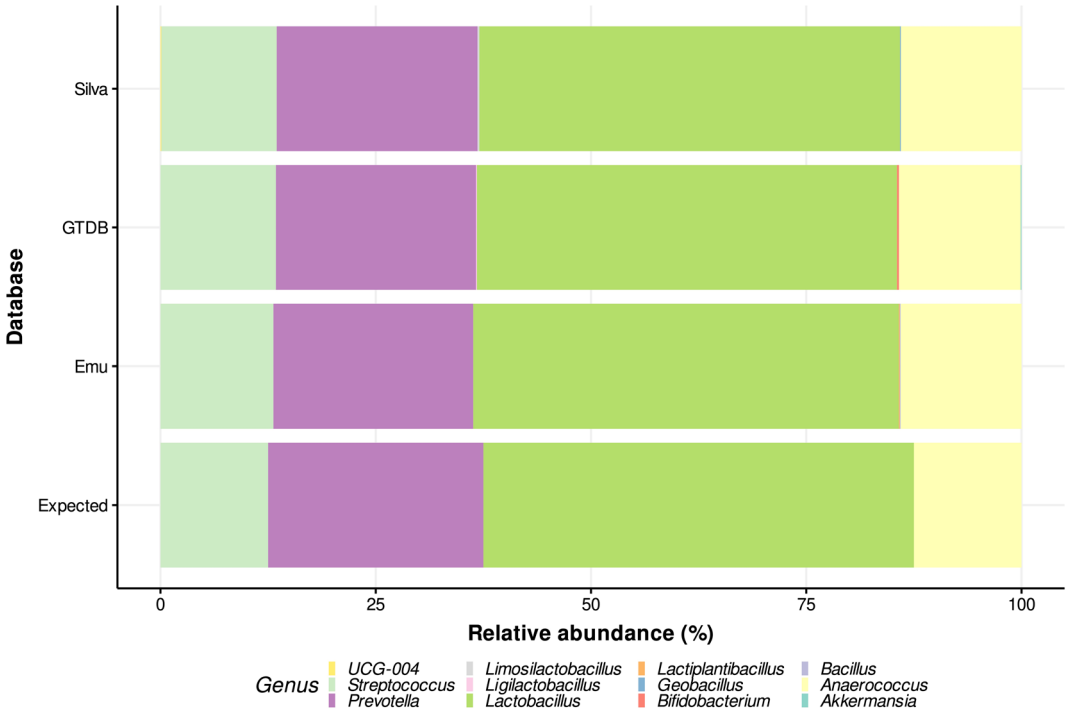

Fig. 1. Relative abundance plot of the in silico vaginal mock community comparing the three databases used. Bars represent relative abundances estimated in the mock community with different databases using the Emu tool.

to primer specificity rather than the absence of target sequences. These results are presented in Supplementary Fig. 1 and Supplementary Table S3.

ONT sequencing and 16S rRNA gene metataxonomic analysis of self-collected vaginal samples

Characteristics of participants

In this study, 20 Chilean women between the ages of 22 and 36 years (median age 32 years) consented to vaginal self-sampling. Each participant responded to an interview related to her lifestyle and health history. All participants declared themselves sexually active. Among them, 6 out of 20 reported recurrent infections at least thrice a year. In addition, 25% of the participants reported using antibiotics regularly. Regarding hygienic practices, responses varied, but mainly water use, and 6 participants reported daily use of intimate hygiene products. During the self-sampling procedure, only one participant experienced discomfort when inserting the swab.

Specie	Expected	Silva	GTDB	Emu
Prevotella bivia	0.25	0.229	0.230	0.229
Lactobacillus crispatus	0.125	0.131	0.137	0.137
Lactobacillus jensenii	0.125	0.120	0.126	0.129
Streptococcus agalactiae	0.125	0.126	0.132	0.122
Anaerococcus hydrogenalis	0.125	0.140	0.141	0.139
Lactobacillus iners	0.125	0.112	0.117	0.118
Lactobacillus gasseri	0.125	0.098	0.087	0.105
Only genus level abundance	–	0.034	0.000	0.000
N of miss assigned species	–	5	14	11
Miss assigned abundance	–	0.008	0.029	0.021
Mean absolute error	–	0.0126	0.0146	0.0116

Table 4. Relative species abundances of in silico vaginal mock community estimated with different databases.

Sample	Reads before filtering	Reads after filtering	filtered%
M1	23,343	21,698	7.05
M2	20,455	19,116	6.55
M3	27,351	25,735	5.91
M4	34,603	32,543	5.95
M5	40,313	37,919	5.94
M6	11,000	10,246	6.85
M7	8,000	7,444	6.95
M8	5,000	4,596	8.08
M9	9,000	8,359	7.12
M10	5,000	4,663	6.74
M11	38,000	35,394	6.86
M12	9,000	8,360	7.11
M13	6,000	5,542	7.63
M14	10,000	9,372	6.28
M15	14,000	13,003	7.12
M16	13,000	12,142	6.60
M17	12,263	11,060	9.81
M18	15,757	13,760	12.67
M19	17,096	15,261	10.73
M20	42,993	37,830	12.01

Table 5. Reads for each vaginal sample after sequencing and length filtering.

MinION-based 16S rRNA sequencing and taxonomic profiling of vaginal microbiota

To characterize the bacterial composition of vaginal samples, DNA extraction was performed, yielding concentrations ranging from 30 to 223 ng/µl, except for one sample with an insufficient concentration (< 0.2 ng/µl). Library preparation and purification were carried out for all extracted samples, which were successfully sequenced using the Oxford Nanopore Technologies (ONT) MinION platform. Sequencing and basecalling results yielded an average read length of 1.6 Kb. For taxonomic profiling, the 20 barcoded samples were length-filtered using fastp[34], retaining only reads between 1400 and 1700 base pairs, corresponding to the estimated full-length 16S rRNA gene. The number of reads obtained per sample and read counts after filtering are summarized in Table 5. On average, 7.7% (SD = 1.9%) of reads per sample were retained after filtering.

Filtered reads from each sample were taxonomically profiled using Emu[15] and the Emu 16S rRNA database. Notably, there were no unassigned reads across all samples, indicating that the Emu pipeline successfully classified all 16S rRNA genes present. The relative abundances obtained were processed in R as a phyloseq object for downstream analysis and visualization. The estimated abundances per sample at the genus level are presented in Fig. 2, with the complete dataset available in Supplementary Table S4. As expected, the genus *Lactobacillus* was the most abundant across samples, with 14 out of 20 samples showing a relative abundance above 75%, 12 exceeding 90%, and 5 surpassing 99%.

The second most abundant classification was the "Others" category, encompassing less represented genera that did not rank among the top 10. This pattern suggests that in samples dominated by *Lactobacillus*, the less abundant genera are a mix of low-abundance taxa without a consistent secondary dominant genus. Conversely, in the six samples where *Lactobacillus* relative abundance fell below 50%, the second most abundant, and in

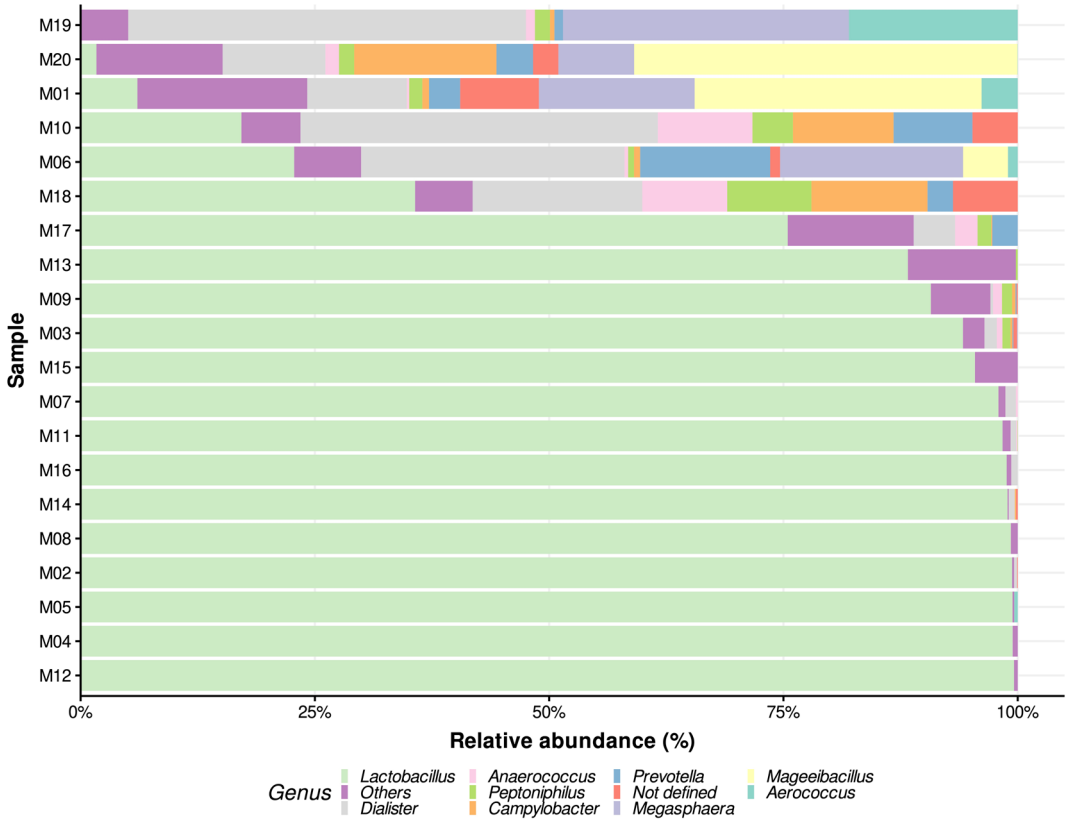

Fig. 2. Bar plot of the top ten most abundant genera across all analyzed vaginal 16S samples. The bars represent relative abundances estimated by Emu for each sequenced sample, displaying only the top ten genera. Remaining genera are grouped as "Others." The x-axis indicates the percentage of estimated abundances, while the y-axis lists samples ordered ascendingly by *Lactobacillus* abundance. The "Not defined" category corresponds to a particular taxon (*Peptostreptococcaceae* bacterium oral taxon 929) that lacks an assigned genus in the database used for analysis.

some cases, the dominant genus was *Dialister*. This genus and others evenly represented in these samples align with CST IV composition, characterized by the absence of a dominant genus[44].

Relative abundances were analyzed at the species level, with results presented in Fig. 3, while the complete dataset is available in Supplementary Table S5. Among the *Lactobacillus* species, *L. crispatus*, *L. iners*, and *L. gaseri* were the most relatively abundant, correlating strongly CST I, II, and III, respectively[23]. These species exhibited relative abundances often exceeding 90% in dominant samples, such as M07, M12, and M14. Notably, 7 out of the 20 samples exhibited a relative abundance of *L. crispatus* above 80%, aligning with healthy vaginal microbiota profiles. Interestingly, only one sample (M11) showed a dominant abundance of *L. jensenii* (98%), which is characteristic of CST V. It is also noteworthy that in cases where *L. iners* was the most abundant species (M05, M09, M13), its relative abundance never exceeded 80%, and the second most abundant species was also a *Lactobacillus* species.

Samples with reduced *Lactobacillus* dominance, such as M01 and M19, displayed more diverse microbial communities, including significant proportions of *Mageeibacillus indolicus* (30.6% in M01) and *Megasphaera elsdenii* (23.2% in M19). The "Others" category, representing less abundant taxa, contributed up to 30–40% in the more diverse samples (M01, M06, M10, M18, M19, M20), indicating substantial microbial complexity. Samples M19 and M20, where *Lactobacillus* species were minimal, exhibited a marked increase in diversity, with dominance by genera such as *Mageeibacillus* and *Megasphaera*. This shift toward increased diversity and reduced *Lactobacillus* dominance is consistent with CST IV profiles.

Considering that the species-level analysis yielded results consistent with previously reported key taxa and relative abundances characteristic of each CST, we performed a preliminary classification of the samples into groups based on these features. Samples were classified according to their dominant species, defined as the species representing more than 80% of the total relative abundance. The results of this preliminary classification are presented in Table 6.

Diversity analysis of vaginal microbiome samples

Genus diversity analysis correlates with a potential dysbiosis state

We conducted a diversity analysis at the genus level to evaluate the information provided at this taxonomic resolution and compare it with species-level resolution. This comparison aims to highlight the advantages of

Fig. 3. Bar plot of the top ten most abundant species across all analyzed vaginal 16S samples. The bars represent the species relative abundances estimated by Emu for each sequenced sample, displaying only the top ten most abundant across all samples. Remaining species are grouped as "Others." The x-axis indicates the percentage of estimated abundances, while the y-axis lists samples ordered ascendingly by species from the *Lactobacillus* genus abundance.

Sample	Most abundant specie	Abundance (%)	Group	Notes
M01	*Mageeibacillus indolicus*	30.60%	IV	Low *Lactobacillus*, diverse anaerobes
M02	*L. crispatus*	88.70%	I	*L. crispatus* dominant
M03	*L. crispatus*	88.10%	I	*L. crispatus* dominant
M04	*L. gasseri*	99.20%	II	*L. gasseri* dominant
M05	*L. iners*	66.90%	III	*L. iners* mainly dominant (*L. crispatus* second most abundant)
M06	Diverse (others)	28.30%	IV	Low *Lactobacillus*, diverse anaerobes
M07	*L. crispatus*	95.80%	I	*L. crispatus* dominant
M08	*L. crispatus*	98.10%	I	*L. crispatus* dominant
M09	*L. iners*	77.20%	III	*L. iners* mainly dominant (*L. jensenii* second most abundant)
M10	Diverse (others)	40%	IV	Low *Lactobacillus*, diverse anaerobes
M11	*L. jensenii*	98%	V	*L. jensenii* dominant
M12	*L. crispatus*	97.90%	I	*L. crispatus* dominant
M13	*L. iners*	53.10%	III	*L. iners* mainly dominant (*L. crispatus* second most abundant)
M14	*L. crispatus*	93.90%	I	*L. crispatus* dominant
M15	*L. crispatus*	93.00%	I	*L. crispatus* dominant
M16	*L. crispatus*	71.20%	I	*L. crispatus* mainly dominant (*L. jensenii* second most abundant)
M17	*L. gasseri*	75.30%	II	*L. gasseri* mainly dominant (Others second most abundant)
M18	Diverse (others)	40%	IV	Low *Lactobacillus*, diverse anaerobes
M19	*Mageeibacillus indolicus*	35.00%	IV	Low *Lactobacillus*, diverse anaerobes
M20	*Mageeibacillus indolicus*	40.90%	IV	Low *Lactobacillus*, diverse anaerobes

Table 6. Classification of samples into groups based on relative abundances of *Lactobacillus* species.

full-length 16S rRNA gene sequencing. Alpha diversity, measured using the Shannon and Simpson indices, was calculated for each sample at the genus level. The results are presented in Supplementary Fig. 2, with statistical comparisons of the diversity indices across previously determined groups shown in Supplementary Table S6. As expected, samples classified in group IV exhibited the highest diversity among all groups. Although the sample size is limited, group I had the lowest mean diversity, suggesting that *Lactobacillus* predominantly dominates this state with minimal presence of other bacterial genera. Given that group IV is consistent with CST IV, which is often associated with a potential dysbiotic state, its significantly higher mean diversity suggests that alpha diversity at the genus level may serve as an indicator of dysbiosis when considered independently.

A similar analysis was performed for beta diversity at the genus level using Bray–Curtis distance, with results shown in Supplementary Fig. 3. Principal component analysis results are provided in Supplementary Tables S7 and S8. Similar to the alpha diversity analysis, all samples clustered according to their previously determined groups. These results indicate that, at the genus level, it is not possible to distinguish between the different *Lactobacillus*-dominated groups. Still, samples corresponding to the diverse group IV can be separated. In Fig. 5, the principal component on the X-axis explains 75.2% of the variance, demonstrating that this type of analysis at the genus level provides a strong statistical representation. The loading analysis further indicates that the primary contributors to this separation are the relative abundances of *Lactobacillus* and various other bacterial genera.

Species-level diversity analysis allows community state classification

To further assess taxonomic resolution and validate the CST classification of our samples, we first conducted a beta diversity analysis at the species level using Bray–Curtis distance (Fig. 4). Additional results are provided in Supplementary Tables S9 and S10. The results show that samples clustered according to the preliminary groups and their correlative Community State Types (CSTs), supporting the classification based on dominant species and confirming concordance with previously defined CST structures. Statistical grouping analysis showed that the PERMANOVA test was significant ($p = 0.001$, $R2 = 0.79$), indicating strong differences among CSTs. However, PERMDISP was also significant ($p = 0.001$), suggesting that part of the observed separation may be attributed to differences in group dispersion, an effect likely amplified by the relatively small number of samples per CST. To complement these findings, we also analyzed alpha diversity grouped by CSTs (Fig. 5; Supplementary Table S11).

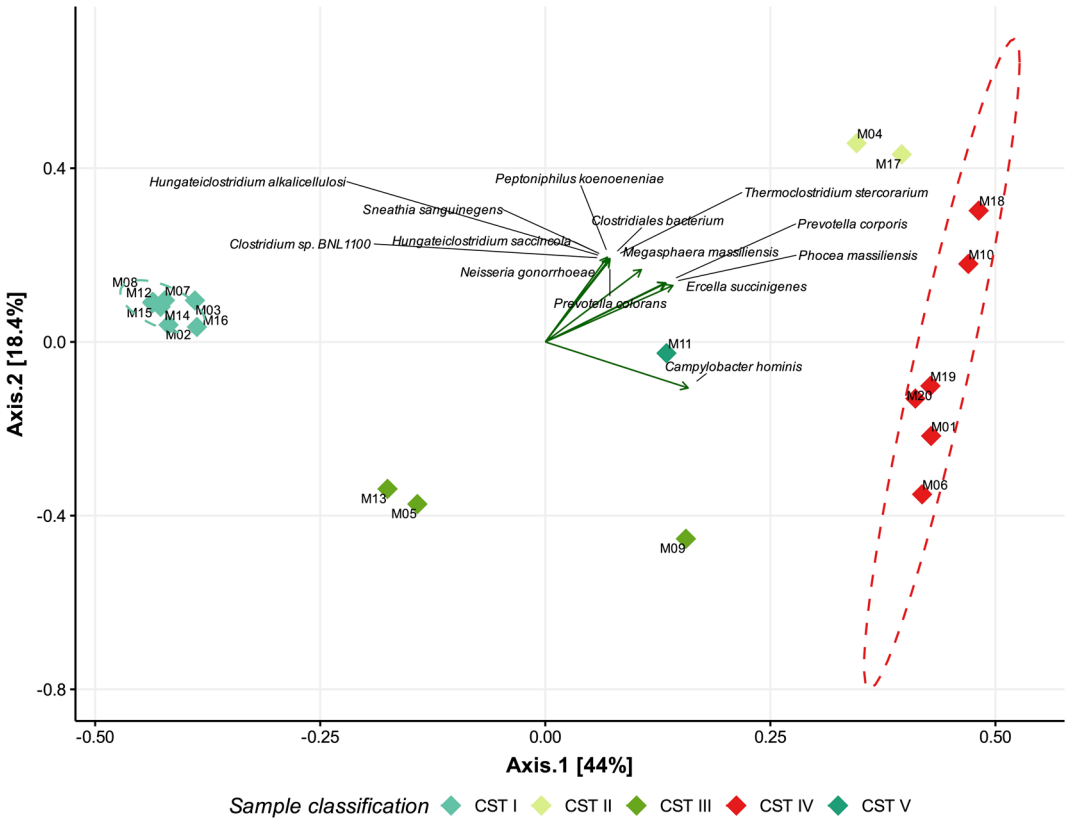

Fig. 4. Beta diversity analysis by PCoA based on Bray–Curtis distance at the species level. The graph represents the Bray–Curtis distances of vaginal samples analyzed at the species level. Each sample is color-coded according to its previously determined community state (CST). Ellipses were added to group samples by CST; however, due to limited sample size, only CST I and CST VI are displayed. Principal Component Analysis (PCA) loadings are represented as arrows labeled with the corresponding species, showing only the 15 most contributing taxa. The x-axis represents the first principal component, while the y-axis represents the second principal component.

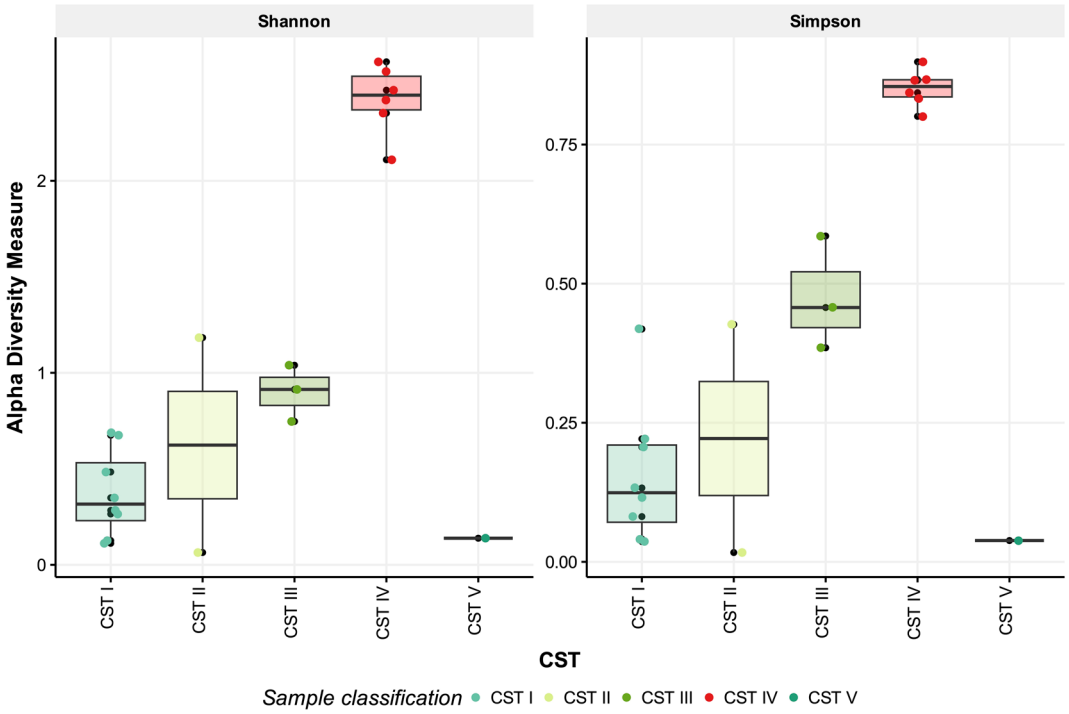

Fig. 5. Alpha diversity analysis at the species level in vaginal samples. Boxplots represent species-level diversity in vaginal samples, grouped according to community state types (CSTs). CSTs include: CST I (*L. crispatus*-dominant), CST II (*L. gasseri*-dominant), CST III (*L. iners*-dominant), CST IV (high bacterial diversity), and CST V (*L. jensenii*-dominant). The x-axis displays the different CSTs, where each point represents an individual sample. The y-axis indicates alpha diversity, measured using Shannon and Simpson indices.

CST IV exhibited the highest diversity values, with an average Shannon index of 2.42 and Simpson index of 0.851, indicating a more heterogeneous bacterial composition. In contrast, CSTs dominated by *Lactobacillus* species (I, II, III, and V) displayed lower diversity values, consistent with a more uniform microbiota. These results support the use of alpha diversity as a quantitative marker of potentially dysbiotic states and further highlight the added value of species-level analysis in distinguishing CSTs. Unlike the genus-level analysis, species-level beta diversity enabled a clearer separation between all CSTs, reflecting a finer resolution of microbial composition. The first principal component (X-axis) explained 44% of the variance, indicating a more complex diversity distribution at the species level. Notably, CST IV showed the greatest dispersion, consistent with its highly diverse and heterogeneous microbial profile. This variability suggests possible sub-classifications within CST IV, potentially representing distinct dysbiotic states or ecological shifts. The loading analysis (highlighting the 15 most influential taxa) revealed that several species, including *Neisseria gonorrhoeae*, *Prevotella corporis*, and *Campylobacter hominis*, contribute significantly to CST IV differentiation, contrasting with taxa predominant in *Lactobacillus*-dominated CSTs.

Discussion

The implementation of ONT-based 16SrRNA gene sequencing demonstrated its potential for accurately characterizing the vaginal microbiota at both the genus and species levels, as highlighted by previous studies using MinION technology for rapid microbiota analysis in clinical settings[18]. The in silico mock community validation confirmed the accuracy of different databases for taxonomic classification. Among the databases evaluated, Emu exhibited the highest accuracy[38], with minimal taxonomic misassignments compared to SILVA and GTDB. This reinforces the importance of selecting an appropriate reference database for taxonomic profiling, particularly in microbial communities where accurate species-level classification is critical for ecological and clinical interpretations.

The processing of vaginal samples, specifically in DNA extraction, yielded variable concentrations, with one sample presenting an insufficient amount for downstream analysis, which underscores the importance of careful sample handling and optimization in low-biomass environments. The sequencing results demonstrated an average read length of 1.6 kb, ensuring near-full-length 16S rRNA coverage, which enhanced taxonomic resolution and allowed for a more accurate characterization of the microbial communities. This result is consistent with previous studies demonstrating that nanopore sequencing of complete 16S rRNA gene amplicons from human intestinal and vaginal microbiota provides further information to determine the bacterial species[18,45,46].

Our findings are consistent with established CST classifications, demonstrating that *Lactobacillus*-dominated communities (CST I, II, III and V) were prevalent in the sampled cohort. The significant differences in community composition among our CST groupings, as detected by PERMANOVA, support the validity of

the CST classification in our dataset. However, the significant PERMDISP results indicate that some of this separation may be exacerbated by the small sample size, emphasizing the need for future studies with larger cohorts to confirm these patterns. In particular, we observed that 14 of 20 Chilean women had an abundance of *Lactobacillus* greater than 75%, with *L. crispatus* as the most predominant species. This finding supports previous studies that highlight the protective role of *L. crispatus* in vaginal health by promoting lactic acid production and maintaining a low vaginal pH, which contributes to preventing urogenital infections and bacterial vaginosis (BV)[26,47]. Moreover, cohort studies from the United States, Europe and Asia have consistently reported a high prevalence of *L. crispatus* in women with a healthy vaginal microbiota, reinforcing its role in vaginal homeostasis and infection resistance[9,48–50].

Conversely, *Lactobacillus iners* was found in moderate abundance in CST III, indicating its role as a transitional species within the vaginal microbiota. Due to its ability to colonize the vaginal niche following microbial disturbances, *L. iners* may provide limited protection against vaginal dysbiosis. However, it has also been implicated in increased susceptibility to BV, sexually transmitted infections (STIs), and adverse pregnancy outcomes, suggesting a dual role as both a symbiont and an opportunistic pathogen[9,51]. Notably, studies have suggested that *L. iners* can persist in dysbiotic environments and facilitate microbial shifts that may lead to recurrent infections[52].

In contrast, 6 of 20 Chilean women exhibited microbial communities similar to CST IV, characterized by reduced *Lactobacillus* dominance, increased bacterial diversity, and enrichment of *Dialister*, *Mageeibacillus* and *Megasphaera*, taxa associated with dysbiotic conditions such as bacterial vaginosis (BV)[27]. Previous studies have linked CST IV with increased susceptibility to vaginal inflammation and adverse gynecological outcomes, with *Dialister* and *Megasphaera* contributing to BV pathogenesis through the production of biogenic amines, which can alter vaginal homeostasis and promote microbial imbalances[27,53].

Interestingly, *Gardnerella vaginalis*, a hallmark taxon of CST IV and BV in Western cohorts, was not detected in our dataset. This was an expected outcome, as in silico PCR simulations using our selected primers (27F/1492R) showed a failure to amplify *G. vaginalis*, a result corroborated by previous studies noting mismatches in these primer-binding sites that hinder its amplification[54–56]. When the simulation was repeated using a previously validated *G. vaginalis*-specific 16S rRNA primer, successful amplification was achieved, confirming the presence of a primer-related bias. This limitation underscores an important technical consideration when using full-length 16S rRNA gene sequencing and highlights the need for optimized primer sets or complementary molecular methods to ensure the accurate detection of BV-associated taxa such as *G. vaginalis*.

In addition, the prevalence of CST IV varies among ethnicities and geographic regions, with a higher presence observed in women of African descent populations[57]. This variation may be influenced by genetic, environmental, and behavioral factors, which warrant further investigation in diverse populations to better understand the determinants of vaginal microbiota composition and stability.

Our study provides the first metataxonomic level characterization of vaginal microbiota in Chilean women, offering a foundation for future research on regional microbiome variations and their implications for women's health. The ability of 16S rRNA gene ONT sequencing to resolve species-level taxonomic assignments with high accuracy offers promising applications for diagnostics and microbiome-based therapeutic strategies. However, it does not allow for functional characterization. In this context, shotgun metagenomic sequencing represents a complementary approach, as it provides deeper insights into strain-level diversity, functional genes, and other microbial components such as viruses and fungi. Incorporating shotgun metagenomics in future studies could contribute to a more comprehensive understanding of the vaginal ecosystem[58]. Additionally, larger cohorts and longitudinal sampling will be essential to explore temporal microbiota dynamics and their relationship with health outcomes in this underrepresented population.

This study successfully characterized the vaginal microbial communities of 20 female residents in Chile using 16S rRNA gene sequencing with Oxford Nanopore Technologies (ONT). Our findings provide valuable insights into the taxonomic composition and Community State Types (CSTs) of vaginal microbiota[23], highlighting the applicability of long-read sequencing for vaginal microbial profiling. To our knowledge, this is the first study of its kind conducted in Chile, filling a regional knowledge gap and establishing a baseline for Latin American research. Using CSTs helped assess if microbial patterns in this cohort align with global community structures. Results suggest general concordance, indicating that CST classifications can be relevant beyond the populations in which they were initially defined. Given its exploratory nature and limited sample size, our results underscore the need for future research incorporating longitudinal data, functional analyses, and clinical metadata to evaluate the adequacy of current CST frameworks in diverse populations.

Data availability

All the data used in this study are available from the cited literature and corresponding websites. All the sequencing data obtained in this study were deposited in the NCBI SRA repository under BioProject code PRJNA1232418.

Code availability

All the data used in this study are available from the cited literature and corresponding websites. All the sequencing data obtained in this study were deposited in the NCBI SRA repository under BioProject code PRJNA1232418.

Received: 21 March 2025; Accepted: 9 October 2025
Published online: 17 November 2025

References

1. Pendharkar, S., Skafte-Holm, A., Simsek, G. & Haahr, T. Lactobacilli and their probiotic effects in the vagina of reproductive age women. *Microorganisms*. **11**, (2023).
2. Navarro, S. et al. Glycogen availability and pH variation in a medium simulating vaginal fluid influence the growth of vaginal *Lactobacillus* species and *Gardnerella vaginalis*. *BMC Microbiol.* **23**, 186 (2023).
3. Wu, S., Hugerth, L. W., Schuppe-Koistinen, I. & Du, J. The right bug in the right place: opportunities for bacterial vaginosis treatment. *NPJ Biofilms Microbiomes* **8**, 34 (2022).
4. Hou, K. et al. Microbiota in health and diseases. *Signal Transduct. Target. Ther.* **7**, 135 (2022).
5. Workowski, K. A. *et al.* Sexually transmitted infections treatment guidelines, 2021. *Centers for Disease Control and Prevention.* https://www.cdc.gov/std/treatment-guidelines/STI-Guidelines-2021.pdf (2021).
6. McKloud, E. *et al.* Recurrent vulvovaginal Candidiasis: A dynamic interkingdom biofilm disease of *Candida* and *Lactobacillus*. *mSystems.* **6**, e0062221 (2021).
7. Schwecht, I., Nazli, A., Gill, B. & Kaushic, C. Lactic acid enhances vaginal epithelial barrier integrity and ameliorates inflammatory effects of dysbiotic short chain fatty acids and HIV-1. *Sci. Rep.* **13**, 20065 (2023).
8. Chen, H. et al. Alterations of vaginal Microbiota in women with infertility and *Chlamydia trachomatis* infection. *Front. Cell. Infect. Microbiol.* **11**, 698840 (2021).
9. Zheng, N., Guo, R., Wang, J., Zhou, W. & Ling, Z. Contribution of *Lactobacillus iners* to vaginal health and diseases: A systematic review. *Front. Cell. Infect. Microbiol.* **11**, 792787 (2021).
10. Mohanty, T., Doke, P. P. & Khuroo, S. R. Effect of bacterial vaginosis on preterm birth: a meta-analysis. *Arch. Gynecol. Obstet.* **308**, 1247–1255 (2023).
11. Nonspecific vaginitis. Diagnostic criteria and microbial and epidemiologic associations. *Am. J. Med.* **74**, A28 (1983).
12. Nugent, R. P., Krohn, M. A. & Hillier, S. L. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. *J. Clin. Microbiol.* **29**, 297–301 (1991).
13. Redelinghuys, M. J., Geldenhuys, J., Jung, H. & Kock, M. M. Bacterial vaginosis: Current diagnostic avenues and future opportunities. *Front. Cell. Infect. Microbiol.* **10**, 354 (2020).
14. Heikema, A. P. et al. Comparison of Illumina versus nanopore 16S rRNA gene sequencing of the human nasal Microbiota. *Genes* **11**, 1105 (2020).
15. Curry, K. D. *et al.* Emu: Species-level microbial community profiling for full-length nanopore 16S reads. (2021) https://doi.org/10.1101/2021.05.02.442339.
16. Davidson, I. M., Nikbakht, E., Haupt, L. M., Ashton, K. J. & Dunn, P. J. Methodological approaches in 16S sequencing of female reproductive tract in fertility patients: a review. *J. Assist. Reprod. Genet.* **42**, 15–37 (2025).
17. Qing, W. *et al.* Species-level resolution for the vaginal microbiota with short amplicons. *mSystems* **9**, e0103923 (2024).
18. Komiya, S. et al. MinION, a portable long-read sequencer, enables rapid vaginal microbiota analysis in a clinical setting. *BMC Med. Genomics* **15**, 68 (2022).
19. Wang, Y., Zhao, Y., Bollas, A., Wang, Y. & Au, K. F. Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.* **39**, 1348–1365 (2021).
20. Ferneyhough, B. et al. A highly accurate nanopore-based sequencing workflow for culture and PCR-free microbial metagenomic profiling of urogenital samples. *BMC Urol.* **25**, 41 (2025).
21. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
22. Parks, D. H. et al. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* **50**, D785–D794 (2022).
23. Ravel, J. et al. Vaginal microbiome of reproductive-age women. *Proc. Natl. Acad. Sci. U. S. A.* **108**(Suppl 1), 4680–4687 (2011).
24. France, M. T. et al. VALENCIA: a nearest centroid classification method for vaginal microbial communities based on composition. *Microbiome* **8**, 166 (2020).
25. Lebeer, S. et al. A citizen-science-enabled catalogue of the vaginal microbiome and associated factors. *Nat. Microbiol.* **8**, 2183–2195 (2023).
26. Tachedjian, G., Aldunate, M., Bradshaw, C. S. & Cone, R. A. The role of lactic acid production by probiotic *Lactobacillus* species in vaginal health. *Res. Microbiol.* **168**, 782–792 (2017).
27. Molina, M. A., Biswas, S., Núñez-Samudio, V. & Landires, I. Targeting *Megasphaera* species to promote cervicovaginal health. *Trends Microbiol.* **32**, 628–630 (2024).
28. Dong, W. et al. Characteristics of vaginal Microbiota of women of reproductive age with infections. *Microorganisms* **12**, 1030 (2024).
29. Baud, A. et al. Microbial diversity in the vaginal microbiota and its link to pregnancy outcomes. *Sci. Rep.* **13**, 9061 (2023).
30. Callahan, B. J., McMurdie, P. J. & Holmes, S. P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* **11**, 2639–2643 (2017).
31. Lillo G, E., Lizama I, S., Medel C, J. & Martínez T, M. A. Diagnóstico de vaginosis bacteriana en un consultorio de planificación familiar de la Región Metropolitana, Chile. *Rev. Chilena Infectol.* **27**, (2010).
32. Villaseca, R. *et al.* Vaginal infections in a family health clinic in the metropolitan region, chile Infecciones vaginales en un Centro de Salud Familiar de la Región Metropolitana, Chile. *Rev. Chilena Infectol.* 41–47 (2015).
33. Mancilla, V., Jimenez, N. R., Bishop, N. S., Flores, M. & Herbst-Kralovetz, M. M. The vaginal Microbiota, human Papillomavirus infection, and cervical carcinogenesis: A systematic review in the Latina population. *J. Epidemiol. Glob. Health* **14**, 480–497 (2024).
34. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
35. Pruitt, K. D., Tatusova, T., Brown, G. R. & Maglott, D. R. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res* **40**, D130–D135 (2012).
36. Angly, F. E., Willner, D., Rohwer, F., Hugenholtz, P. & Tyson, G. W. Grinder: a versatile amplicon and shotgun sequence simulator. *Nucleic Acids Res.* **40**, e94 (2012).
37. Mesloub, Y., Beury, D., Vandermeeren, F. & Caboche, S. CuReSim-LoRM: A tool to simulate metabarcoding long reads. *Int. J. Mol. Sci.* **24**, 14005 (2023).
38. Curry, K. D. et al. Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. *Nat. Methods* **19**, 845–853 (2022).
39. Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R. K. & Schmidt, T. M. rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res* **43**, D593–D598 (2015).
40. O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733–D745 (2016).
41. McMurdie, P. J. & Holmes, S. Phyloseq: a bioconductor package for handling and analysis of high-throughput phylogenetic sequence data. *Pac. Symp. Biocomput.* 235–246 (2012).
42. Villanueva, R. A. M. & Chen, Z. J. ggplot2: Elegant graphics for data analysis (2nd ed.). *Measurement.* **17**, 160–167 (2019).
43. Verhelst, R. et al. Cloning of 16S rRNA genes amplified from normal and disturbed vaginal microflora suggests a strong association between *Atopobium vaginae*, *Gardnerella vaginalis* and bacterial vaginosis. *BMC Microbiol.* **4**, 16 (2004).
44. Ma, Z. S. & Li, L. Quantifying the human vaginal community state types (CSTs) with the species specificity index. *PeerJ* **5**, e3366 (2017).

45. Matsuo, Y. et al. Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION™ nanopore sequencing confers species-level resolution. *BMC Microbiol.* **21**, 35 (2021).
46. Lüth, T. et al. Improving analysis of the vaginal microbiota of women undergoing assisted reproduction using nanopore sequencing. *J. Assist. Reprod. Genet.* **39**, 2659–2667 (2022).
47. Qi, F. et al. Orally administrated *Lactobacillus gasseri* TM13 and *Lactobacillus crispatus* LG55 can restore the vaginal health of patients recovering from bacterial vaginosis. *Front. Immunol.* **14**, 1125239 (2023).
48. Zhou, X. et al. Differences in the composition of vaginal microbial communities found in healthy Caucasian and black women. *ISME J.* **1**, 121–133 (2007).
49. Borgdorff, H. et al. The association between ethnicity and vaginal microbiota composition in Amsterdam, the Netherlands. *PLoS ONE* **12**, e0181135 (2017).
50. Liu, Y. et al. Effectiveness of vaginal probiotics *Lactobacillus crispatus* chen-01 in women with high-risk HPV infection: a prospective controlled pilot study. *Aging (Albany NY)* **16**, 11446–11459 (2024).
51. Kadogami, D. et al. Impact of *Lactobacillus* in the uterine microbiota on *in vitro* fertilization outcomes. *J. Reprod. Immunol.* **160**, 104138 (2023).
52. Sabbatini, S. et al. *Lactobacillus iners* cell-free supernatant enhances biofilm formation and hyphal/pseudohyphal growth by *Candida albicans* vaginal isolates. *Microorganisms* **9**, 2577 (2021).
53. Nelson, T. M. et al. Vaginal biogenic amines: biomarkers of bacterial vaginosis or precursors to vaginal dysbiosis?. *Front. Physiol.* **6**, 253 (2015).
54. Johnson, J. S. et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* **10**, 5029 (2019).
55. Pereira-Marques, J. et al. Impact of host DNA and sequencing depth on the taxonomic resolution of whole metagenome sequencing for microbiome analysis. *Front. Microbiol.* **10**, 1277 (2019).
56. Frank, J. A. et al. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl. Environ. Microbiol.* **74**, 2461–2470 (2008).
57. Fettweis, J. M. et al. The vaginal microbiome and preterm birth. *Nat. Med.* **25**, 1012–1021 (2019).
58. Feehily, C. et al. Shotgun sequencing of the vaginal microbiome reveals both a species and functional potential signature of preterm birth. *NPJ Biofilms Microbiomes* **6**, 50 (2020).

Acknowledgements

We thank Dr. Carlos Blondel and the Blondel Lab for their support, including access to laboratory equipment and guidance in molecular biology. We also thank the BioNanotechnology & Microbiology Lab for providing additional laboratory resources. We would also like to thank the 20 volunteers who generously participated in this study by providing samples. This study was partially supported by the Agencia Nacional de Investigación y Desarrollo (ANID) of Chile through various grants: Fondecyt Regular 1221209 and Anillo ATE220061 to J.A.U, and National Doctorate Scholarship 21240354 to P.V.

Author contributions

B.O.-A. conceived and designed the study. B.O.-A. was responsible for participant selection and sample collection. DNA extraction and sequencing were performed by B.O.-A., with contributions from P.V. P.V. conducted the bioinformatic analyses and data visualization. J.A.U. provided conceptual guidance and supervised the entire research process. B.O.-A. and P.V. drafted the initial manuscript, which was subsequently reviewed and refined by J.A.U. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-23837-8.

Correspondence and requests for materials should be addressed to J.A.U.