

RESEARCH ARTICLE

Open Access

Novel high-resolution targeted sequencing of the cervicovaginal microbiome



Karolina M. Andralojc^{1,2†} , Mariano A. Molina^{1,3†}, Mengjie Qiu⁴, Bram Spruijtenburg¹, Menno Rasing², Bernard Pater⁴, Martijn A. Huynen⁴, Bas E. Dutilh^{4,5}, Thomas H. A. Ederveen⁴, Duaa Elmelik², Albert G. Siebers⁶, Diede Loopik⁷, Ruud L. M. Bekkers^{8,9}, William P. J. Leenders^{2,10*†} and Willem J. G. Melchers^{1*†}

Abstract

Background: The cervicovaginal microbiome (CVM) plays a significant role in women's cervical health and disease. Microbial alterations at the species level and characteristic community state types (CST) have been associated with acquisition and persistence of high-risk human papillomavirus (hrHPV) infections that may result in progression of cervical lesions to malignancy. Current sequencing methods, especially most commonly used multiplex 16S rRNA gene sequencing, struggle to fully clarify these changes because they generally fail to provide sufficient taxonomic resolution to adequately perform species-level associative studies. To improve CVM species designation, we designed a novel sequencing tool targeting microbes at the species taxonomic rank and examined its potential for profiling the CVM.

Results: We introduce an accessible and practical circular probe-based RNA sequencing (CiRNAseq) technology with the potential to profile and quantify the CVM. In vitro and in silico validations demonstrate that CiRNAseq can distinctively detect species in a mock mixed microbial environment, with the output data reflecting its ability to estimate microbes' abundance. Moreover, compared to 16S rRNA gene sequencing, CiRNAseq provides equivalent results but with improved sequencing sensitivity. Analyses of a cohort of cervical smears from hrHPV-negative women versus hrHPV-positive women with high-grade cervical intraepithelial neoplasia confirmed known differences in CST occurring in the CVM of women with hrHPV-induced lesions. The technique also revealed variations in microbial diversity and abundance in the CVM of hrHPV-positive women when compared to hrHPV-negative women.

Conclusions: CiRNAseq is a promising tool for studying the interplay between the CVM and hrHPV in cervical carcinogenesis. This technology could provide a better understanding of cervicovaginal CST and microbial species during health and disease, prompting the discovery of biomarkers, additional to hrHPV, that can help detect high-grade cervical lesions.

Keywords: Cervicovaginal microbiome, hrHPV, Targeted sequencing, smMIP, CiRNAseq, CST, CIN

* Correspondence: william.leenders@radboudumc.nl; willem.melchers@radboudumc.nl

[†]Karolina M. Andralojc and Mariano A. Molina contributed equally to this work.

William P. J. Leenders and Willem J. G. Melchers shared last authorship.

²Department of Biochemistry, Radboud Institute for Molecular Life Sciences, 6525, GA, Nijmegen, The Netherlands

¹Department of Medical Microbiology, Radboud University Medical Center, 6500, HB, Nijmegen, The Netherlands

Full list of author information is available at the end of the article

Background

High-risk human papillomavirus (hrHPV)-induced cervical cancer affects more than half a million women every year [1]. Although the oncogenic role of hrHPV is clear in this process, only a minority of hrHPV infections lead to cervical lesions, and ultimately, cancer. Hence, there is a need to better understand hrHPV-induced



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

alterations in the cervicovaginal environment that contribute to cancer development. Accordingly, recent efforts have focused on the host immune response and the cervicovaginal microbiome (CVM) [2–4]. The latter has a significant role in women's cervical health and disease [5]. Throughout women's lives, the CVM can change during the menstrual cycle, pregnancy, or after sexual activities [6–10]. Such microbiome changes also occur in pathogenic conditions like bacterial vaginosis (BV), Candidiasis, and viral infections [11–13]. Interestingly, the composition of the CVM also depends on microbial dominancy and diversity, creating characteristic community state types (CST) that could be either dominant for *Lactobacillus* species (CST I, II, III, and V) or diverse for other bacterial species (CST IV). Variations in the CVM have been widely described in relation to hrHPV infections, with CST IV significantly associated with high-grade cervical lesions and cancer [4, 14, 15]. Furthermore, recent investigations have determined that these microbiome alterations not only occur at the genus level but also at the species level, suggesting that specific microbial species and CST are associated with progressive or regressive behavior of cervical lesions and could act as biomarkers for the disease [16–18]. Nevertheless, studying the CVM and elucidating its function currently relies mostly on short length 16S rRNA gene sequencing (16S rRNA-seq), which struggles to distinguish microbes at this taxonomic rank [19–22].

Microbiome profiling using 16S rRNA-seq is based on the sequence analysis of hypervariable regions (VRs) in ribosomal 16S rRNA genes for microbe identification [23, 24]. PCR amplicons covering two VRs (e.g., V1-V2, V3-V4) are generated with degenerate primer sets and subjected to next-generation sequencing. The technique results in bacterial identification typically providing for family- or genus-level taxonomy, while species identification is achieved for a limited number of genera [25, 26]. Moreover, several studies have observed bias in microbiome profiling with 16S rRNA-seq due to variability in the selection of primers and VRs for amplification and sequencing [27–29]. Since changes in the CVM also take place at the species level, it is essential to develop detection methods with higher resolution and specificity. To this end, Pinna et al. suggest that increasing the number of analyzed VRs may improve the taxonomic resolution in microbiome profiling [30].

Circular probe-based RNA sequencing (CiRNAseq) using single-molecule molecular inversion probes (smMIPs) has proven to be a useful tool for cancer research [31–34] and hrHPV expression studies [35]. smMIPs can be designed to target any nucleic acid sequence and thus could be applied to recognize multiple VRs and to identify diverse microbes such as bacteria, fungi, and viruses simultaneously. Likewise, by targeting

and combining multiple VRs for microbiome profiling, CiRNAseq carries a potential to perform high-resolution sequencing with high specificity and sensitivity [30]. Besides being customizable for its targets, the addition of a unique molecule identifier (UMI) to a smMIP corrects for PCR amplification bias, making the counting of amplified smMIPs possible, which could be valuable for absolute microbiome quantification [31, 35]. Because CiRNAseq uses barcode technology, it can handle hundreds of samples in one sequencing run, making the technique cost-effective. Furthermore, it requires fewer specialized skills for data analyses and interpretation than other sequencing methods such as 16S rRNA-seq, making it a handy and accessible technology [36, 37].

We describe here the characteristics and potential of a CVM-specific CiRNAseq assay. We validate the technique's resolution, specificity, and performance in vitro with mock samples, and profile the CVM of a cohort of cervical smears from women with and without hrHPV-associated cervical abnormalities.

Methods

Study participants and samples

For this study, a total of 102 cervical smears in Preserv-Cyt were collected from women participating in the Dutch population-based cervical cancer screening program, which were received and processed at Radboudumc (Nijmegen, the Netherlands). Women participating in the cervical cancer screening program were informed that residual material could be used for anonymous research and had the opportunity to opt-out. Only residual material from women who did not opt-out was included. The histological follow-up outcomes were obtained from the nationwide network and registry of histo- and cytopathology in the Netherlands (PALGA; Houten, the Netherlands). hrHPV identification was performed as previously described [35]. All methods were performed following the institutional guidelines for using human samples. One set of ten hrHPV-positive smears was used for the comparative analyses with 16S rRNA-seq. DNA from these samples was isolated from 1 ml of residual material using DNA and Viral Small volume kit (Roche, cat. no. 6543588001) and subjected to CiRNAseq. The cohort of the remaining 92 cervical smears consisted of 46 hrHPV-positive samples of women with confirmed high-grade cervical intraepithelial neoplasia (CIN2+) and 46 hrHPV DNA-negative smears. Five milliliters of each cervical cell suspension was centrifuged for 5 min at 2500×g, and the pellet dissolved in 1 ml of Trizol reagent (Thermo Scientific). RNA was isolated through standard procedures and dissolved in 20 µl nuclease-free water. We routinely processed a maximum of 2 µg of RNA for

DNase treatment and cDNA generation, using Super-scriptII (Thermo) as previously described [35].

smMIP design and targeted sequencing

We compiled a list of 434 previously identified microbes that have been recognized as significant in the cervicovaginal environment by recent literature and the Human Vaginal Microbiome Project (Additional file 1) [38, 39]. The genome sequences were initially retrieved from the National Center for Biotechnology Information (NCBI [40]) using Biomart [41]. Sequences from small ribosomal subunit (SSU) and large ribosomal subunit (LSU) rRNA genes were selected and extracted using Biopython [42] and BEDTools [43], respectively [23, 44]. smMIPs against SSU and LSU rRNA genes were designed in MIPgen [36]. We selected smMIPs with homologous hybridization arms and dissimilar regions of interest (ROIs) and included a random octanucleotide UMI in the smMIP backbone. Next, we compared the selected ROI sequences with the corresponding rRNA sequences within the SILVA rRNA database. Only sequences that were 100% identical over the full length with this database were regarded as fit for annotation [45]. Subsequently, MegaBLAST and the Burrows-Wheeler Aligner (BWA) were combined to validate in silico the specificity of smMIPs in discriminating species [46, 47]. Thereafter, a greedy algorithm was implemented to validate the potential of a smMIP in identifying as many species at once as possible based on ROIs sequences. This validation resulted in the selection of 30 smMIPs targeting the 434 microbes and pathogens (Additional file 2). All smMIPs were validated on a dataset composed of genomes and annotations from species isolated from cervical smears (Fig. 1). Then, to standardize species detection and reduce the chance of false-positive annotation, we considered only species that were identified with two or more reactive smMIPs (on average six probes per species). This filtering resulted in the final selection of our targets consisting of 107 genera and 321 species that represent our cervicovaginal microbiome panel (CVMP), including bacteria, fungi, and parasites (Additional file 3). CiRNAseq was performed as previously described [31, 35]. For the sequencing of individual species, 10 ng of microbial DNA was analyzed. Analyses of cervicovaginal samples were performed on ~ 50 ng of cDNA/DNA generated according to standard protocols (see the “Study participants and samples” section). Following capture hybridization and probe circularization and purification, circularized probes were subjected to PCR with barcoded Illumina primers. After purification of the correct-size amplicons, quality control, and quantification as previously described [35], a 4-nM library was sequenced on the Illumina Nextseq500 platform (Illumina, San Diego, CA) at the Radboudumc sequencing facility.

CiRNAseq output analysis

Reads were mapped against reference ROIs within our CVMP using the SeqNext module of JSI Sequence Pilot version 4.2.2 build 502 (JSI Medical Systems, Ettenheim, Germany). The settings for read processing were a minimum of 50% matching bases, a maximum of 15% mismatches, and a minimum of 50% consecutive bases without a mismatch between them; for read assigning, the threshold was a minimum of 95% of identical bases with the ROIs. All identical PCR products were reduced to one consensus read (unique read counts, URC) using the UMI. We set an arbitrary threshold of at least 1000 URC from all smMIPs combined in an individual sample, below which we considered an output non-interpretable [48]. For microbial annotation, species with two reactive smMIPs were annotated when 100% of the specific set of smMIPs had URC. Species with three or more reactive smMIPs were annotated when more than 50% of their specific set of smMIPs had URC using a custom R script. For analyses where isolates from our CVMP were not considered, the URC for each isolate were summed to represent the bacterium at the species level. To define relative abundances, microbial species URC was divided by the total URC of all microbes annotated in the sample. For establishing microbial diversity, URC was turned to 1 and 0, indicating the presence or absence of microbes, respectively.

16S rRNA gene amplification and sequencing

Residual material from ten hrHPV-positive cervical smears in PreservCyt solution, randomly obtained from the Dutch population-based cervical cancer screening program with approval from the regional institutional review board and the National Institute for Public Health and Environment (No. 2014-1295), was initially pelleted by centrifugation. Pellets were suspended in 1 ml DNA/RNA shield buffer (Zymo, cat. no. R1104). DNA was extracted according to standard protocols and processed by BaseClear B.V. (Leiden, the Netherlands) for microbiome profiling using the primers 357F (5'-CCTACGGGAGGCAG CAG-3') and 802RV2 (5'-TACNVGGGTATCTAAKCC-3') that target the V3 and V4 variable regions of the 16S rRNA gene [29]. PCR protocol was as follows: 2 m 95 °C hot start; 35 cycles of 20 s 95 °C, 10 s 61 °C, 15 s 70 °C; 10 m 70 °C. The libraries were barcoded, multiplexed, and sequenced on an Illumina MiSeq machine with paired-end 300 cycles protocol and indexing by BaseClear [49]. Illumina sequencing data were quality checked and demultiplexed by BaseClear standards, and FASTQ files were generated.

16S rRNA gene sequencing data analysis

From the FASTQ files, forward and reverse reads were pairwise assembled with PEAR (v0.9.10 [50]) in default

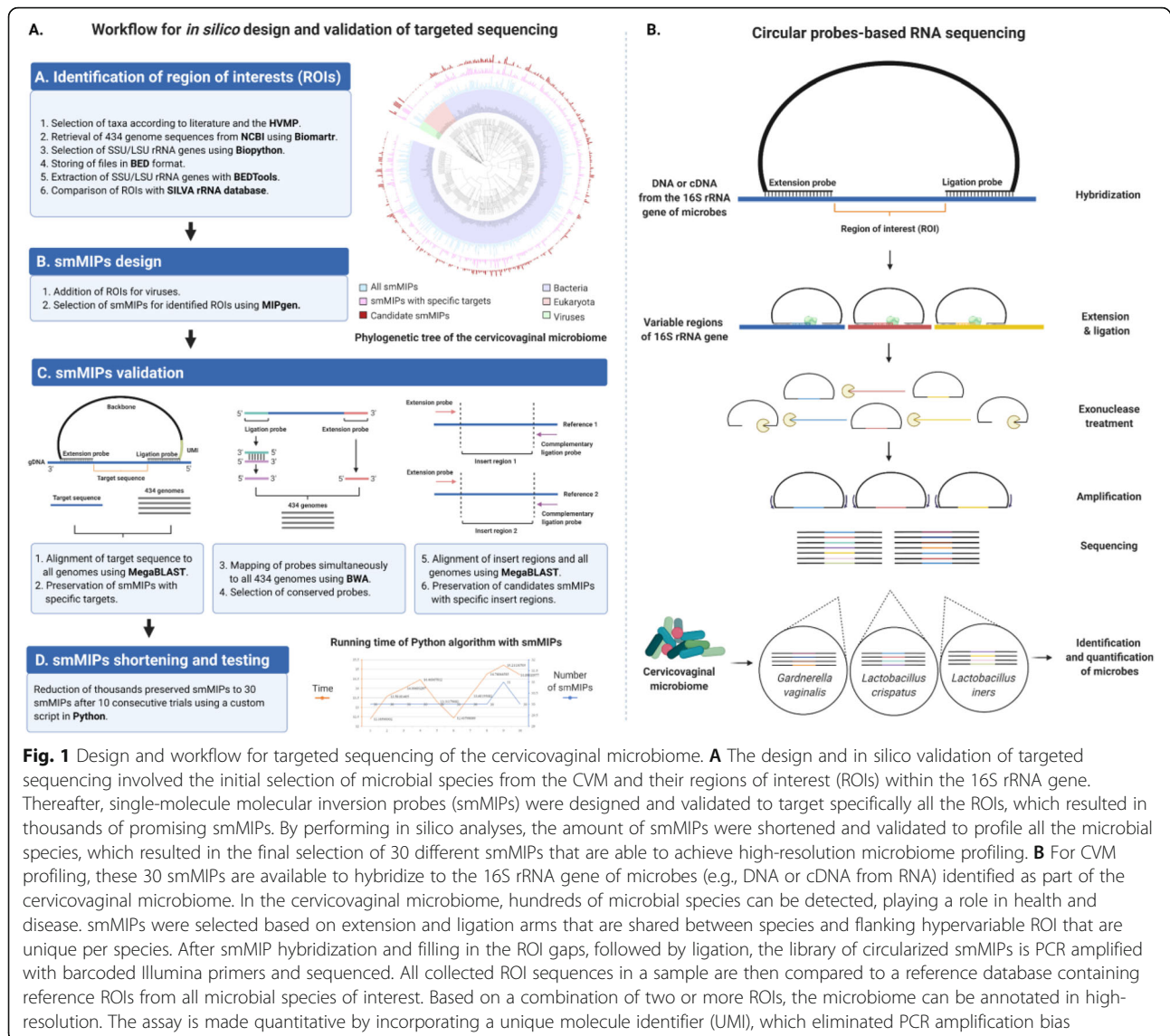


Fig. 1 Design and workflow for targeted sequencing of the cervicovaginal microbiome. **A** The design and *in silico* validation of targeted sequencing involved the initial selection of microbial species from the CVM and their regions of interest (ROIs) within the 16S rRNA gene. Thereafter, single-molecule molecular inversion probes (smMIPs) were designed and validated to target specifically all the ROIs, which resulted in thousands of promising smMIPs. By performing *in silico* analyses, the amount of smMIPs were shortened and validated to profile all the microbial species, which resulted in the final selection of 30 different smMIPs that are able to achieve high-resolution microbiome profiling. **B** For CVM profiling, these 30 smMIPs are available to hybridize to the 16S rRNA gene of microbes (e.g., DNA or cDNA from RNA) identified as part of the cervicovaginal microbiome. In the cervicovaginal microbiome, hundreds of microbial species can be detected, playing a role in health and disease. smMIPs were selected based on extension and ligation arms that are shared between species and flanking hypervariable ROI that are unique per species. After smMIP hybridization and filling in the ROI gaps, followed by ligation, the library of circularized smMIPs is PCR amplified with barcoded Illumina primers and sequenced. All collected ROI sequences in a sample are then compared to a reference database containing reference ROIs from all microbial species of interest. Based on a combination of two or more ROIs, the microbiome can be annotated in high-resolution. The assay is made quantitative by incorporating a unique molecule identifier (UMI), which eliminated PCR amplification bias

settings. For the generation of the 16S-derived taxa-to-sample compositional matrix, a customized Python workflow based on Quantitative Insights Into Microbial Ecology (QIIME v2.0 [51]) was adopted (<http://qiime.org>). Relative abundances per sample were calculated with QIIME2 default settings, where the reads per taxon were divided by the total number of bacterial reads for that sample.

In vitro validation of sequencing targets

To test *in vitro* the specificity and resolution of CiRNA-seq, we used 12 bacterial species listed in Additional file 4: Supplementary Table 1, obtained from the Medical Microbiology Department, Radboudumc, Nijmegen, the Netherlands. Bacteria were grown in appropriate culture media. Following growth, their genomic DNA was extracted using DNA and Viral Small volume kit (Roche,

cat. no. 6543588001). PCR and Sanger sequencing was performed to validate species identification. Water was used as the negative control. For CiRNA-seq, we prepared a concentration of 1.5 ng/ μ L from each microbes' DNA in a final volume of 40 μ L.

In silico validation of sequencing microbial species

For testing the specificity of CiRNA-seq towards the species *Anaerococcus vaginalis*, *Anaerococcus tetradius*, *Peptostreptococcus anaerobius*, *Gardnerella vaginalis*, *Bifidobacterium longum*, and *Prevotella buccalis* *in silico*, we downloaded the 16S rRNA gene of the species *A. vaginalis* (D14146), *A. tetradius* (D14142), *P. anaerobius* (D14150), *G. vaginalis* (M58744), *Bifidobacterium longum sp. suillum* (AB924532), and *P. buccalis* (AB547676) from the LPSN database (DSMZ, Germany). SnapGene® Viewer 5.3.2 (Insightful Science; snapgene.com).

com) was used to verify the hybridization of the ROIs that are targeted by the smMIPs to detect the species and the hybridization of the primers 357F and 802RV2 to the V3-V4 regions. Clustal Omega v1.2.4 [52] was used to align ROIs sequences from each compared species. Pairwise sequence similarities and phylogenetic analyses were calculated using the method recommended by Meier-Kolthoff et al. [53] for the 16S rRNA gene via the GGDC web server [54] available at <https://ggdc.dsmz.de/>

In vitro validation of RNA testing and quantification

To assess the capacity of CiRNAseq to quantify and analyze microbial RNA, an *Escherichia coli* (*E. coli*; ATCC 25922) culture in stationary phase was inoculated at 5% in BHI medium and incubated at 37 °C on a shaking platform at 100 rpm for 48 h. Optical density (OD₆₃₀) was measured every hour, and 1-ml aliquots were taken after each measurement, pelleted, and stored for nucleic acid isolation. After 26.5 h of culture, an aliquot was taken for autoclaving. A second aliquot was treated with 0.75 ml of cefoxitin (1 mg/ml), followed by further growth for an additional 20 h (Additional file 5: Supplementary Table 2). Nucleic acids were isolated from all aliquots using the MagNA Pure kit (Roche, cat. no. 03730964001). RNA concentrations (ng/ml) were measured using NanoDrop 2000 (Thermo Scientific). After treatment with DNAase, RNA was processed to cDNA for CiRNAseq analysis.

Statistical analyses

Analyses with our CVMP were performed using ClustVis [55]. For the clustering analysis (Fig. 5A), the settings were as follows: clustering distance for columns: Canberra [56, 57]; clustering method: Ward (unsquared distances); row scaling: Pareto scaling [58]. Canberra distance normalizes the absolute difference in abundance of each taxon, allowing comparison of minor taxa. A shorter Canberra distance indicates greater similarity.

Linear discriminant analysis (LDA) effect size was performed using the LEfSe tool [59]. LEfSe combines standard tests for statistical significance (Kruskal-Wallis test and pairwise Wilcoxon test) with LDA for feature selection. Alpha value for the factorial Kruskal-Wallis test was 0.05. Threshold on the logarithmic LDA score for discriminative features was 2.0 [59].

Microsoft Excel 2016® and GraphPad Prism v9.0.0 (GraphPad Software, Inc., USA) were used to analyze datasets and determine species richness, Shannon's diversity index, and Pearson's *r* correlations. The statistical significance of differences in microbial richness, diversity, and relative abundance was calculated using GraphPad with the Mann-Whitney test to obtain the *p*-value.

Significant differences between groups are denoted by * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, or **** $p < 0.0001$.

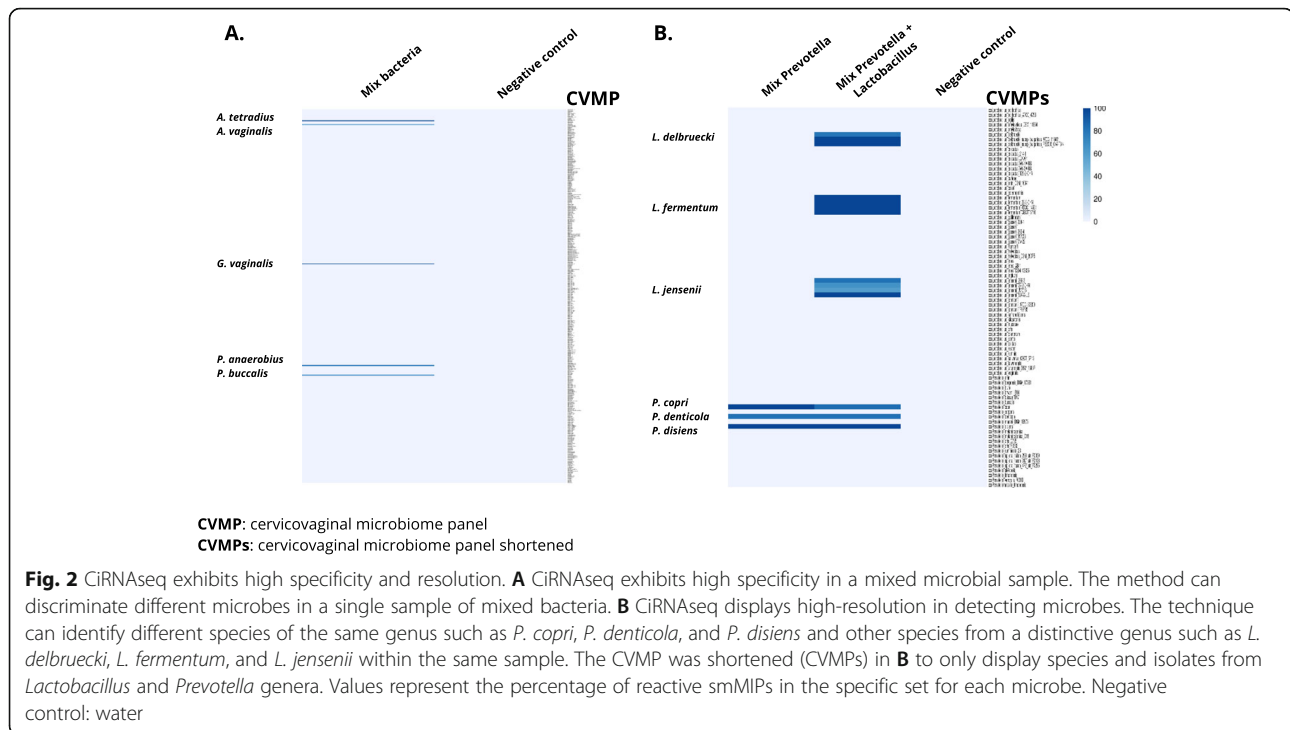
Results

CiRNAseq is a sequencing tool for high-resolution microbiome profiling that uses smMIPs to target multiple VRs of the 16S rRNA gene characterizing the composition, abundance, and diversity of the CVM. Following the identification of relevant microbes in the cervicovaginal niche and their regions of interest (ROIs) (Fig. 1A), we designed probes with homologous hybridization arms with high specificity for ribosomal RNA, that flank heterologous ROIs (Fig. 1A). With bioinformatic analyses (see the "Methods" section), we validated thousands of smMIPs to select only 30 that combined can detect 107 genera and 321 species within the CVM (Fig. 1A). The selected probes were subjected to an in silico validation to ensure accurate microbiome profiling (Fig. 1A) [38, 39]. By comparing these ROIs with a reference database, this method assigns URC to microbes of interest. Because we require that at least two different ROIs must be detected in a microbe, the CiRNAseq pipeline ensures a robust species-level annotation of the microbiome (Fig. 1B).

We performed in vitro and in silico validations to demonstrate the potential of CiRNAseq for high throughput sequencing of the microbiome and compared this new method to 16S rRNA-seq. We designed a dedicated CiRNAseq test to study the CVM in smears from hrHPV-negative women and women with hrHPV-associated CIN2+ lesions. We also validated the specificity, resolution, reproducibility, targeting (DNA/RNA), and quantification abilities of the technology in profiling the CVM.

CiRNAseq exhibits high specificity and resolution

To validate the specificity of CiRNAseq in a mixed microbial environment, we first tested the technique by analyzing a defined mixture of genomic DNA from *Anaerococcus tetradius*, *Anaerococcus vaginalis*, *Gardnerella vaginalis*, *Peptostreptococcus anaerobius*, and *Prevotella buccalis*, which are typical for the CVM (Fig. 2A, Additional file 4: Supplementary Table 1, Additional files 7 and 8). Water was used as a negative control. CiRNAseq correctly identified the five input species based on sequence comparison with the reference ROIs. In the negative control, the technique did not yield any reads (Fig. 2A). In silico analyses using the 16S rRNA gene of these exact species and their ROIs (Additional file 6: Supplementary Figure 1, and Additional file 9) further confirm that the strict technique's requirements for species annotation (see the "Methods" section) facilitate an accurate discrimination of microbes in a mixed microbial sample with high specificity.



Subsequently, we assessed the technique's resolution in detecting microbes at the species level (Fig. 2B, Additional file 4: Supplementary Tables 1, 7 and Additional file 10). To this end, we prepared a mixed microbial sample consisting of genomic DNA from three species of *Prevotella* (*Prevotella copri*, *Prevotella denticola*, and *Prevotella disiens*) and added these to a second mixed sample containing DNA from three *Lactobacillus* species (*Lactobacillus delbrueckii*, *Lactobacillus fermentum*, and *Lactobacillus jensenii*). All of these species are commonly found in the CVM. As represented in Fig. 2B, CiRNAseq correctly identified all individual species in all samples. Thus, CiRNAseq is able to distinguish microbes at the species level for this specific mixed microbial sample, showing potential for high-resolution sequencing of the CVM.

CiRNAseq RNA quantification capacity mirrors bacterial growth and activity

In natural niches such as the CVM, DNA is a very stable molecule, while RNA is rapidly degraded. Therefore, whereas DNA sequencing can reveal the presence of genomic DNA of bacterial species in a sample, RNA sequencing gives information on the activity of such species by identifying which genomic regions are transcribed to RNA [60]. To evaluate the CiRNAseq potential in quantifying active microbes at the RNA level, we examined how the growth of *E. coli*, a species that can be found in the CVM [61, 62], is reflected in the number of unique read counts (URC) obtained from

RNA sequencing. Following the growth of a pure culture of *E. coli* for 48 h through OD measurement every hour, we selected nine-time points where the *E. coli* culture was sampled for RNA isolation, including the bacterial lag, exponential, and stationary phases (Fig. 3A, in orange dots). We also selected two samples that were either autoclaved or treated with an antibiotic (Additional file 5: Supplementary Table 2). Samples were taken in duplicate and subjected to CiRNAseq to test reproducibility (Additional file 11: Supplementary Figure 2A). The mean number of URC achieved in these replicates for the lag and exponential phases is shown in Fig. 3B (green line, first seven-time points) and Additional file 11: Supplementary Figure 2B. When comparing the OD of *E. coli* culture to the mean of URC obtained from sequencing, we found that the values were significantly correlated, particularly from the lag to the exponential phase ($p = 0.0286$) (Fig. 3B). Samples taken from the stationary growth phase had lower URCs, indicating lower ribosomal activity in bacteria from the stationary phase than bacteria from the exponential growth phase.

We also analyzed the RNA concentrations of each aliquot taken for sequencing and compared them to the OD and URC, as shown in Fig. 3C, D, and Additional file 11: Supplementary Figure 2C. Here we noticed that the isolated total RNA matched the OD of *E. coli* growth phases (Fig. 3C). Furthermore, we observed that the RNA levels of the samples taken from the stationary phase (time points six and seven) were higher than those from the exponential phase (Fig. 3D), reflecting the

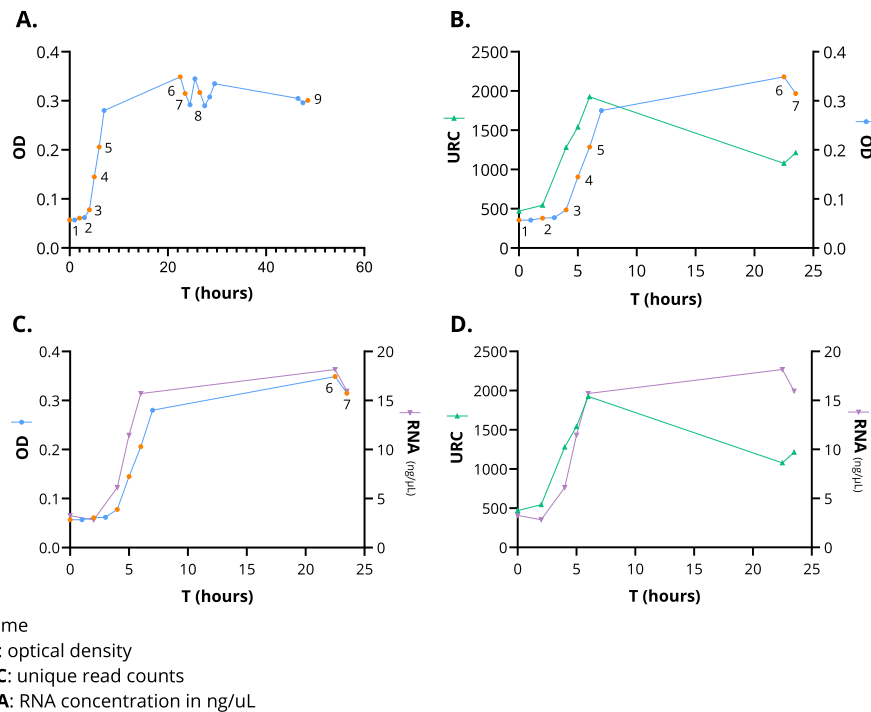


Fig. 3 CiRNAseq RNA quantification capacity mirrors bacterial growth and activity. **A** The OD obtained from monitoring *E. coli* growth for 48 h reveals the bacterial growth phases. The nine orange time points indicate the phases from when samples were taken for sequencing analyses. **B** *E. coli* URC correlated with the OD, particularly from the lag to the exponential phases. Samples taken in the stationary phase had lower URC than the last measurement within the exponential phase. **C** RNA concentrations of the samples taken for sequencing are parallel to the OD, indicating that low URC found in time points six and seven may reflect the measurement of ribosomal activity. **D** RNA concentrations also match the URC obtained from sequencing

accessible RNA for sequencing. As expected, we did not find any URC after autoclaving the sample taken in time point eight, even though the OD and RNA concentration measured previous to autoclaving was similar to the growth phase (Additional file 11: Supplementary Figure 2C). Similarly, the sample treated with cefoxitin (antibiotic) had a low number of URC, suggesting inhibition of ribosomal activities (Additional file 11: Supplementary Figure 2C). Thus, CiRNAseq can quantify microbes' RNA, mirroring translational activity and growth.

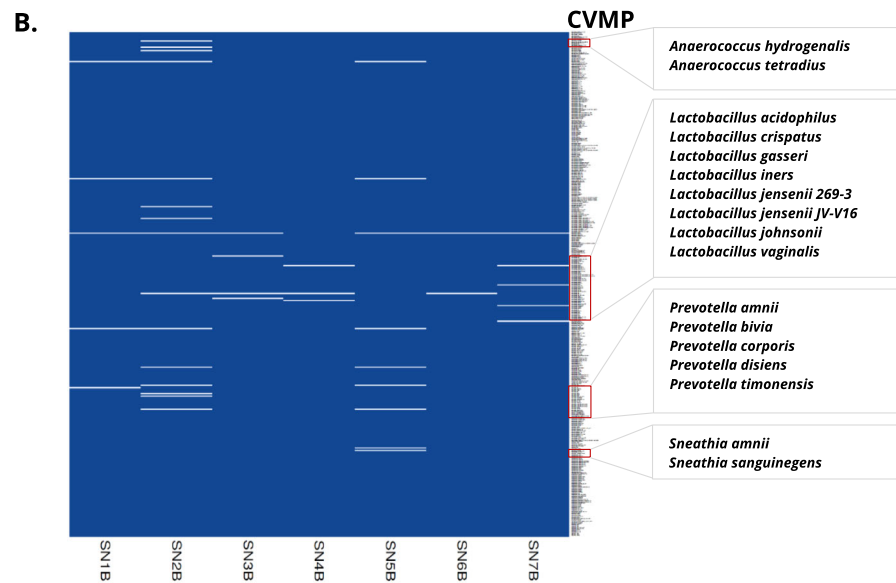
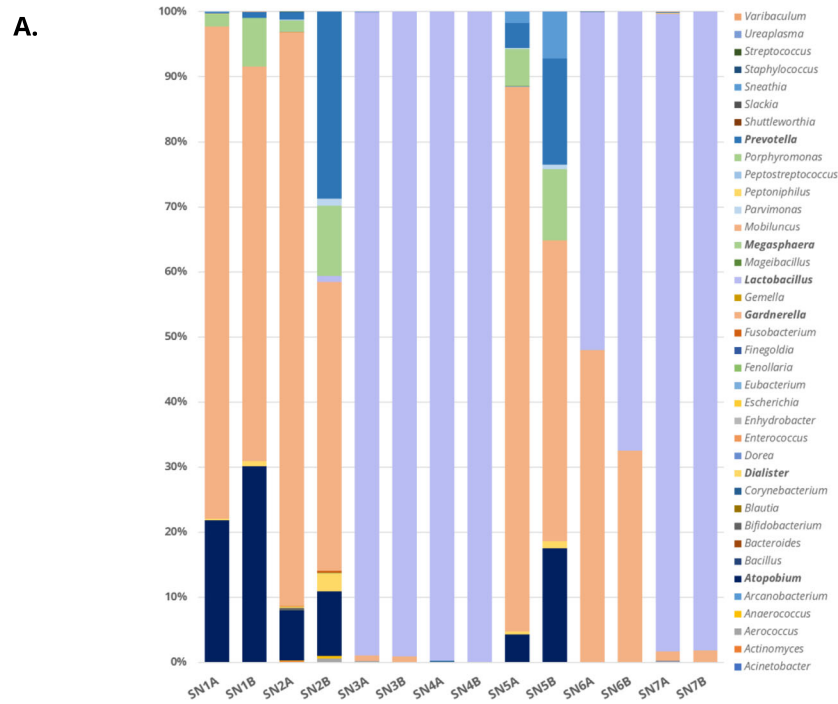
CiRNAseq provides genus-level microbiome profiling as 16S rRNA-seq but offers improved taxonomic resolution

Given that the gold-standard sequencing method for profiling the microbiome is 16S rRNA-seq, we compared both sequencing methodologies. First, we did an in silico comparison analyses between both techniques based on the 16S rRNA sequences of two different species within the family *Bifidobacteriaceae*. Species from the genera *Gardnerella* and *Bifidobacterium* can be found in the CVM and studies have described that 16S rRNA-seq struggles differentiating their species due to limited variability of the used VRs for microbiome profiling [8, 28,

29, 63]. As shown in Additional file 12: Supplementary Figure 3, the V3/4 regions of the species *Bifidobacterium longum* and *Gardnerella vaginalis* have > 90% similarity. Our designed smMIPs also target V5, V6, and V9 regions that have < 45% homology and therefore can discriminate the species with higher confidence (Additional file 5).

Next, we randomly selected ten hrHPV-positive smears, which were simultaneously profiled using CiRNAseq and 16S rRNA-seq at the DNA level. Two out of ten samples had low reads (< 2500 reads) with 16S rRNA-seq compared to the rest of the samples (> 80,000 reads) and were excluded from the analyses. One additional sample had < 1000 URC with CiRNAseq and was also excluded from the study. In the remaining seven samples, we determined the relative microbes' abundances. Following 16S rRNA-seq, we focused our analyses on 38 genera that were profiled by 16S rRNA-seq and were also available for microbiome profiling using CiRNAseq (Fig. 4 and Additional files 3, 13 – 15). Microbes with relative abundances ≤ 0.07% were considered non-present in the samples.

The seven remaining samples sequenced with 16S rRNA-seq (SN-A) and CiRNAseq (SN-B) were analyzed,



CVMP: cervicovaginal microbiome panel

Fig. 4 CiRNAseq provides genus-level microbiome profiling as 16S rRNA-seq but offers improved taxonomic resolution. **A** 16S rRNA-seq (SN-A) and CiRNAseq (SN-B) possess similar sequencing capacity when differentiating 31 of the 38 genera analyzed. The methods gave the same results with respect to quantifying the genera *Lactobacillus*, *Gardnerella*, *Atopobium*, and *Megasphaera*. The microbial composition of samples A and B is similar when analyzed using the two sequencing techniques. Microbial species and isolates URC were summed to show the results at the genus level. **B** CiRNAseq allows detecting bacteria at high-resolution. The technique suggested 24 different bacterial species, including two species of *Anaerococcus*, seven *Lactobacillus* species, five species of *Prevotella*, and two species of *Sneathia*. For **A**, values represent the relative abundances of each microbe in the sample. Bacterial species isolates within our CVMP were considered for display of **B**

as shown in Fig. 4. Here, we first observed that the relative abundances are highly similar using both techniques (Fig. 4A), suggesting that CiRNAseq and 16S rRNA-seq have a comparable efficiency in microbial identification and quantification. This finding can be easily observed in samples 3, 4, 6, and 7 (A and B), where both techniques detected *Lactobacillus* with equivalent relative abundances ($r = 0.9605$, $p = 0.0006$, Fig. 4A and Additional file 16: Supplementary Figure 4A). Likewise, both methods yielded similar relative abundances for *Gardnerella* ($r = 0.9384$, $p = 0.0018$, Fig. 4A and Additional file 16: Supplementary Figure 4B), *Atopobium* ($r = 0.9255$, $p = 0.0028$, Fig. 4A and Additional file 16: Supplementary Figure 4C), and *Megasphaera* ($r = 0.8344$, $p = 0.0196$, Fig. 4A and Additional file 16: Supplementary Figure 4D). Still, 16S rRNA-seq yielded a lower relative abundance than CiRNAseq for the genera *Dialister*, *Parvimonas*, *Prevotella*, and *Sneathia* (Fig. 4A).

Both techniques profiled *Gardnerella*, *Atopobium*, *Aerococcus*, *Dialister*, *Lactobacillus*, *Megasphaera*, *Parvimonas*, *Prevotella*, and *Sneathia*. CiRNAseq also detected *Anaerococcus* and *Fusobacterium* in higher relative abundances than 16S rRNA-seq (Fig. 4A).

Genera *Actinomyces*, *Bifidobacterium*, *Corynebacterium*, *Peptoniphilus*, and *Ureaplasma* were detected by 16S rRNA-seq (relative abundances between 0.10 and 0.40%), but not by CiRNAseq (Fig. 4A). From the 24 genera that 16S rRNA-seq yielded $\leq 0.07\%$ in relative abundances, CiRNAseq was concordant in 22 (91%). In general, 16S rRNA-seq and CiRNAseq were concordant in 31 out of the 38 genera analyzed (81%), illustrating the technique's specificity and sensitivity at the genus level.

To further investigate the species resolution of CiRNAseq in the CVM, we also analyzed samples SN1B to SN7B at this taxonomy level, as shown in Fig. 4B, Table 1, and Additional file 17. In total, we observed 24 different species from our CVMP. We were able to detect two species of *Anaerococcus*, seven species of *Lactobacillus*, five species of *Prevotella*, and two species of *Sneathia* (Fig. 4B, Table 1). When considering the classification of cervicovaginal CST [5], CiRNAseq also allows the characterization of CST IV (SN1B, SN2B, and SN5B; high diversity), CST III (SN3B and SN6B; *L. iners* dominance), and CST I (SN4B and SN7B; *L. crispatus* dominance), which also resembled the composition defined

Table 1 Species-level identification using circular probe-based RNA sequencing

Bacterial species	SN1B	SN2B	SN3B	SN4B	SN5B	SN6B	SN7B
<i>Aerococcus christensenii</i>		.					
<i>Anaerococcus hydrogenalis</i>		.					
<i>Anaerococcus tetradius</i>		.					
<i>Atopobium vaginae</i>	.	.			.		
<i>Dialister microaerophilus</i>	.	.			.		
<i>Fenollaria massiliensis</i>		.					
<i>Fusobacterium nucleatum</i>		.					
<i>Gardnerella vaginalis</i>
<i>Lactobacillus acidophilus</i>			.				
<i>Lactobacillus crispatus</i>				.			.
<i>Lactobacillus gasseri</i>							.
<i>Lactobacillus iners</i>		
<i>Lactobacillus jensenii</i>			.	.			
<i>Lactobacillus johnsonii</i>							.
<i>Lactobacillus vaginalis</i>							.
<i>Megasphaera genomosp type 1</i>	.	.			.		
<i>Parvimonas micra</i>		.			.		
<i>Prevotella amnii</i>		.			.		
<i>Prevotella bivia</i>	.						
<i>Prevotella corporis</i>		.					
<i>Prevotella disiens</i>		.					
<i>Prevotella timonensis</i>		.			.		
<i>Sneathia amnii</i>					.		
<i>Sneathia sanguinegens</i>					.		

by 16S rRNA-seq (Fig. 4A, B). Therefore, these CiRNA-seq results suggest the ability to identifying bacteria at the species level and in microbial communities with high specificity in the complex CVM niche.

CiRNAseq: CVM changes in women with hrHPV-induced lesions

Several studies suggest that accurate detection of microbial species in the CVM may be relevant for predicting the progression of hrHPV-induced precancerous cervical lesions and cancer [15, 64–66]. To investigate this, we applied CiRNAseq to RNA isolated from cervical smears of hrHPV-negative women (considered healthy, $n = 46$) and women with hrHPV-positive high-grade cervical intraepithelial neoplasia (CIN2+, $n = 46$) (Additional files 18 and 19).

Unsupervised clustering analysis using URC from each microbial species in individual samples of our cohort is shown to generate three clusters (Fig. 5A). The clusters represented the well-known community state types (CST) [5]. Cluster 1 consisted of 18 samples, of which 72.2% were hrHPV negative, and was characterized by a CST I that is dominated by *L. crispatus*. Additional *Lactobacillus* species such as *L. iners*, *L. jensenii*, *L. ultunensis*, and *L. acidophilus* were also common (Fig. 5A). With a Fisher's exact test, CST I showed a small association to hrHPV-negative women ($p = 0.0639$) when compared to hrHPV-positive women in cluster 1.

Cluster 2 consisted of 27 samples, of which 20 (74%) were from women with hrHPV-induced high-grade lesions. These women had a CVM consistent with CST IV, characterized by depletion of *Lactobacillus* species and colonization of mainly anaerobic bacteria such as *M. genomosp type 1*, *G. vaginalis*, *S. amnii*, *S. sanguinegens*, *D. micraerophilus*, and *A. vaginae*. With a Fisher's exact test, CST IV exhibited a significant association to hrHPV-positive women ($p = 0.0055$) when compared to hrHPV-negative women in cluster 2. The third cluster (3) contained 47 samples, of which 26 (55.3%) were hrHPV negative and 21 (44.7%) had hrHPV-induced lesions. Women's CVM in cluster 3 were still dominant for *Lactobacillus* species, and their microbial composition was consistent with other CST such as II (dominance for *L. gasseri*), III (dominance for *L. iners*), and V (dominance for *L. jensenii*) (Fig. 5A).

We also tested our cohort of 92 samples through a principal component analysis (PCA). We determined PC1 and PC2, representing 32.7% and 12.6% of our cohort, respectively (Fig. 5B). Here, we observed a minor separation of samples corresponding to both hrHPV-negative and hrHPV-positive women with some overlap. After analyzing the loading score of PC1 (Additional file 20), we found that anaerobic bacteria such as *M. genomosp type 1* and *G. vaginalis* showed the higher

correlation with PC1, suggesting an association of particular bacterial species with hrHPV status (Fig. 5B).

Although we observed a particular change in the CVM of samples within clusters 1 and 2, the microbiome composition was ambiguous in cluster 3, possibly due to the presence of different CST in this cluster (Fig. 5A). To further evaluate the microbial composition of our cohort, we performed a supervised average analysis comparing the CVM of hrHPV-negative ($n = 46$) and hrHPV-positive ($n = 46$) women (Additional file 21: Supplementary Figure 5). This analysis showed that hrHPV-negative women were typically colonized with *L. acidophilus*, *L. crispatus*, *L. jensenii*, *L. psittaci*, *L. ultunensis*, and *L. vaginalis*. In contrast, hrHPV-positive with high-grade lesions women possessed a more diverse microbiome with anaerobic bacteria such as *A. vaginae*, *D. micraerophilus*, *G. vaginalis*, *S. amnii*, and *S. sanguinegens*. Interestingly, *L. iners* was also present in hrHPV-positive women. Other bacteria found in hrHPV-positive women included *Prevotella* species such as *P. amnii*, *P. buccalis*, and *P. timonensis* (Additional file 21: Supplementary Figure 5). To confirm these observations, we performed a linear discriminant analysis (LDA) effect size (LEfSe) [59] modeling comparing microbiome composition and relative abundance between hrHPV-negative ($n = 45$, an outlier was excluded from this analysis) and hrHPV-positive samples ($n = 46$) (Fig. 5C). In the hrHPV-positive group, this analysis showed higher levels for *G. vaginalis*, *M. genomosp type 1*, *S. amnii*, *S. sanguinegens*, *P. anaerobius*, *D. micraerophilus*, *A. vaginae*, *P. amnii*, and *P. buccalis* ($p < 0.05$) (Fig. 5C and Additional file 22: Supplementary Figure 6A – 6I). In contrast, in the hrHPV-negative group, this analysis determined an over-representation of *L. acidophilus* ($p < 0.05$) (Fig. 5C and Additional file 22: Supplementary Figure 6J). Thus, the alteration of the CVM due to hrHPV infection is characterized by the change from a healthy *Lactobacillus* microbiota to an anaerobic-diverse microbiota that can be explored using CiRNAseq.

CiRNAseq profiling reveals alterations in the CVM

To further show the significance of CiRNAseq in studying CVM alterations, we examined the two clusters enriched for CST I (1) and CST IV (2) from the analysis described in Fig. 5A. We also assessed the difference in microbial richness, diversity, and relative abundance for *L. iners* in our cohort's two main groups: hrHPV-negative women versus hrHPV-positive women with CIN2+.

The clusters enriched for CST I and IV had 18 and 27 samples, respectively. The CVM from these two clusters seemingly varied in microbial diversity (Fig. 6A). CST I, containing mostly hrHPV-negative women, had a shallow microbial diversity characterized by *Lactobacillus*

species like *L. acidophilus*, *L. crispatus*, *L. iners*, *L. jensenii*, and *L. ultunensis*. Therefore, CST I was diverse at the species level but less diverse at the genus level (Fig. 6A). In contrast, within CST IV, consisting of mainly hrHPV-positive women, such *Lactobacillus* species were depleted, and only *L. iners* continued to be present (Fig. 6A), as described in previous analyses (Fig. 5 and Additional file 21: Supplementary Figure 5). Moreover, CST IV had a highly diverse microbiome characterized by *A. vaginae*, *D. micraerophilus*, *G. vaginalis*, *L. iners*, *M. genomosp type 1*, *P. timonensis*, *S. amnii*, *S. sanguinegens*, and other bacteria as detailed in Fig. 6A. To quantify this observation, we calculated species richness and alpha-diversity, which confirmed that hrHPV-negative women had a less rich (mean of 4.2 microbes) and diverse (mean of 1.22) microbiome when compared to hrHPV-positive women, mean of 6.6 for richness and 1.60 for alpha-diversity ($p < 0.05$) (Fig. 6B, C). In conclusion, CiRNAseq let us determine that, besides a CVM change upon hrHPV infections, there is an alteration of the microbial diversity.

Given that *L. iners* colonize both hrHPV-negative and hrHPV-positive women [64, 67] but did not show a strong association to hrHPV status in our LefSe analysis, we assessed the bacterium abundance independently. To this purpose, we examined our cohort of 92 cervical samples and selected samples for which CiRNAseq identified *L. iners*. Accordingly, we included 25 hrHPV-negative samples and 34 hrHPV-positive samples for this analysis. Following the estimation of relative abundances within the samples, we calculated the mean and significance of the differences, as observed in Fig. 6D. Here, we noticed that *L. iners* had a higher relative abundance in hrHPV-negative women (mean 19.3) when compared to hrHPV-positive women (mean 11.9, $p < 0.05$), suggesting that even though it is present in the diverse microbiome of hrHPV-positive women, the abundance of this specie decreases upon infection (Fig. 6D).

Discussion

16S rRNA gene sequencing is the most widely employed method for microbiota analysis, which have transformed community microbiological studies of the CVM [23, 28, 44, 68]. In this study, we introduce a novel targeted sequencing method with sufficient resolution and specificity to enable the profiling of cervicovaginal microbiota with similar performance to 16S rRNA-seq, but with additional advantages such as very high-throughput profiling, high taxonomic resolution, and improved sequencing sensitivity. Using CiRNAseq, we show that hrHPV-positive women with high-grade cervical intraepithelial neoplasia acquire a characteristic CST IV microbiome as observed by earlier 16S rRNA-seq studies.

CiRNAseq achieves improved sensitivity for microbiome profiling, which is a result of the underlying smMIP technique in which the same molecule is amplified multiple times over in a circular fashion. Our findings detailing the identification and quantification of genera such as *Lactobacillus*, *Gardnerella*, *Atopobium*, and *Megasphaera* with equivalent results to 16S rRNA-seq corroborate our technique's specificity at the genus level. Nonetheless, since CiRNAseq uses two and more VRs for microbiome profiling and can target both SSU and LSU for identifying some species, its resolution increases to the species taxonomy rank, but further studies on the level of classification confidence at species resolution are warranted. Using CiRNAseq we demonstrated that in fact such genera corresponded to specific species such as *L. crispatus*, *L. iners*, *L. jensenii*, *G. vaginalis*, *A. vaginae*, or *M. genomosp type 1* which are extremely relevant for women's cervical health and disease [4, 8, 69, 70]. Thus, our technology confirms recent studies highlighting the advantage of targeting and combining multiple VRs to improve the resolution of microbiome profiling [30, 71].

CiRNAseq showed that the CVM of women changes from a healthy-dominated *Lactobacillus* microbiome (CST I) to an anaerobic-diverse microbiome (CST IV) upon persistent hrHPV infection. Whether hrHPV infections result in a different CST or a CST IV microbiome is more permissive for hrHPV persistence is still unknown. Changes in vaginal pH have been associated with the microbial composition, particularly with depletion of *Lactobacillus* species and the enhancement of facultative anaerobic bacteria such as *G. vaginalis*, *D. micraerophilus*, *A. vaginae*, *Megasphaera* spp., and *Prevotella* spp. (CST IV) [5, 72, 73]. Interestingly, Mitra et al. recently described that CST IV is highly associated with hrHPV-induced high-grade cervical intraepithelial neoplasia [15]. Since we also observe this association in our cohort of samples, it corroborates and validates the findings obtained with CiRNAseq. On the other hand, the role of individual species in the alteration of the microbiome still remains unclear. Recent studies suggest that *G. vaginalis* drives the vaginal dysbiosis in hrHPV-infected women and exhibits an immunosuppressive role in the vagina, which could explain the higher abundance of *G. vaginalis* in hrHPV-positive women described in our study [2, 64]. Therefore, identifying individual species within the CVM may elucidate the roles of particular bacteria in the microbiome and provide alternative treatment strategies to prevent disease [74]. Furthermore, understanding the CVM change at this taxonomic rank may lead to identifying microbiome profiles that could act as predictive biomarkers for women at risk of developing cervical cancer [15, 16, 18, 63, 75]. Additional studies with a larger cohort of samples are needed to

clarify whether the species or CST described in the current study possess such function and explain how they would associate with the effect of hrHPV infections.

Using CiRNAseq, we also determined that hrHPV-positive women with CIN2+ have a more diverse CVM than hrHPV-negative women. In addition, our data suggest that several *Lactobacillus* species colonize hrHPV-negative women, and thus, they are more diverse at the species level than at the genus level, as previously reported [76]. Interestingly, we observed that *L. acidophilus* was highly abundant in hrHPV-negative women considered healthy, which could be attributed to the vaginal acidic conditions suitable for growth and its well-known antimicrobial activities [77, 78]. Additionally, we found higher levels of *L. iners* within the CVM of hrHPV-positive women. This finding is in line with previous research reporting that *L. iners* may not be as protective as other *Lactobacillus* species because particular *L. iners* strains have been associated with vaginal dysbiosis [3, 14]. Some studies suggest that D-lactate, produced by *L. crispatus* and not *L. iners*, enhances the trapping of HIV in the cervicovaginal mucus [11, 79]. By this mechanism, *L. crispatus*, but not *L. iners*, could also protect the basal epithelium from infection with hrHPV. Furthermore, the lower abundance of *L. iners* in smears from women with hrHPV-induced high-grade lesions could also be attributed to changes in the vaginal pH and a decline in the metabolic activities of *L. iners* [70, 73]. As far as we know, this is the first study to report a higher abundance of *L. acidophilus* in hrHPV-negative women and a lower abundance of *L. iners* in hrHPV-positive women with high-grade cervical intraepithelial neoplasia. Further studies are needed to investigate how the relative abundances of both *L. acidophilus* and *L. iners* species associate to hrHPV-induced malignancy [80].

The strengths of this study and CiRNAseq technique are the improvements to CVM profiling by targeting and combining multiple VRs and achieving higher taxonomic resolution. Potential limitations are our cohort size and the absence of testing CiRNAseq with well-defined and a high-complexity mock microbial community samples as controls [81]. Additionally, due to the use of a reference database with known microbial species, the method could have missed species that are not currently identified or suffered mutations [82, 83]. Moreover, designed probes may not hybridize to the template that carries a mismatch (e.g., diversity within species), which is why we have on average a set of six smMIPs to target and annotate one specific microbe. Worth mentioning, even though our set of 30 smMIPs can target 434 microbes, microbial species that use one smMIP for identification (e.g., viruses) were not included in our final CVMP neither annotated in this study. However,

since one of the benefits is the straightforward design of smMIPs, we could add species-specific smMIPs to our pool to broaden the list of detectable species.

Conclusions

In summary, CiRNAseq is a highly promising technology with the resolution and specificity for high-throughput sequencing, which makes it an interesting tool for uncovering the role of the CVM in health and disease. This study analyzed two outermost groups: hrHPV-negative women with no cervical lesions and hrHPV-positive women with associated high-grade cervical lesions. An obvious question is how the CVM behaves in hrHPV-positive women with no cervical lesions, low-grade cervical lesions, high-grade cervical lesions, and cancer, and such studies are on the way. Moreover, if there are “protective” or “pathogenic” species or CST associated with particular outcomes of hrHPV infections is still unresolved [84]. Future studies using CiRNAseq should allow us to evaluate the CST shift and the consequent alteration of the microbial diversity and high-resolution composition. Whether hrHPV virus or microbial species drive the CST shift is an exciting question to solve in the next studies. Nevertheless, since the microbiome depends on several factors such as hrHPV, the hrHPV genotype, the vaginal environment, and the host immune system, it is plausible that it may be more than one feature driving these microbial changes. Notably, CiRNAseq not only accomplishes quantitative microbiome profiling on the species level, but also achieves detection of viral RNAs and host gene expression products, which may allow investigations of host-microbiome interactions in a single test [35]. Overall, our work indicates that by distinguishing bacteria in high-resolution using CiRNAseq, we could further understand the association of the CVM and hrHPV infections and elucidate their potential role on cervical lesions' progression to cancer.

Abbreviations

16S rRNA-seq: 16S rRNA gene sequencing; BED: Browser Extensible Data; BWA: Burrows-Wheeler Aligner; CCSP: Cervical cancer screening program; CIN: Cervical intraepithelial neoplasia; CiRNAseq: Circular probes-based RNA sequencing; CST: Community state types; CVM: Cervicovaginal microbiome; CVMP: Cervicovaginal microbiome panel; CVMPs: Cervicovaginal microbiome panel shortened; gDNA: Genomic DNA; HVMP: Human Vaginal Microbiome Project; hrHPV: High-risk human papillomavirus; LSU: Large ribosomal subunit; NCBI: National Center for Biotechnology Information; OD: Optical density; ROIs: Region of interests; rRNA: Ribosomal RNA; smMIPs: Single-molecule molecular inversion probes; SSU: Small ribosomal subunit; UMI: Unique molecular identifier; URC: Unique read counts; VRs: Hypervariable regions

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-021-01204-z>.

Additional file 1. List of validated microbes with identifiers (.xlsx). List of 434 microbes selected for the in silico design and validation of smMIPs.

Additional file 2. List of smMIPs and sequences used for CiRNAseq profiling of the cervicovaginal microbiome (.pdf). List of designed 30 smMIPs with ligation and extension arms that target 434 relevant microbes in the cervicovaginal microbiome at the DNA and RNA levels using CiRNAseq.

Additional file 3. Cervicovaginal microbiome panel (.xlsx). First sheet. CVMP: list of 321 microbes profiled by CiRNAseq. Second sheet. CVMP – with isolates: list of 321 microbes including bacteria isolates. Third sheet. Common genera with 16S rRNA-seq; list of 38 genera analyzed to compare 16S rRNA-seq and CiRNAseq.

Additional file 4. Supplementary Table 1. Bacterial species used for in vitro experiments. MMB: Department of Medical Microbiology, Radboudumc.

Additional file 5 Supplementary Table 2. *E. coli* growth experiment. T: time; TP: time point of analyses; OD: optical density; RNAc: RNA concentration in ng/μL; URC: unique read counts; R1: replicate 1; R2: replicate 2; URCm: mean of URC from R1 and R2.

Additional file 6 Supplementary Figure 1. In silico validation of CiRNAseq specificity towards cervicovaginal microbial species. CiRNAseq targets multiple 16S rRNA gene VRs e.g. V5 – V9 for microbiome profiling. For reads assigning and species annotation, the method has a threshold of 95% of similarity between sequences and reference ROIs (A). Phylogenetic analyses of the 16S RNA gene for the species *Anaerococcus vaginalis*, *Anaerococcus tetradius*, *Peptostreptococcus anaerobius*, *Gardnerella vaginalis*, and *Prevotella buccalis* shows the similarity between the 16S rRNA genomes (B). Alignment analyses of the ROIs from the closest related species to the least related species, according to (B), demonstrate the specificity of CiRNAseq. For sequencing *A. vaginalis* and *A. tetradius* the technique uses the same set of five smMIPs, but two out of five ROIs exhibit <95% sequence similarity and thus do not fulfill the threshold for reads assigning (C). For *A. tetradius* and *P. anaerobius*, the technique uses the same four smMIPs, with their respective ROIs showing <90% sequence similarity (D). For *P. anaerobius* and *G. vaginalis*, the technique uses the same two smMIPs, with their ROIs having <85% sequence similarity (E). For *G. vaginalis* and *P. buccalis*, CiRNAseq employs the same two smMIPs, with their ROIs holding <75% sequence similarity (F). Marks (*) indicate smMIPs and ROIs that are dissimilar per bacterium and therefore were not included in the analyses. Unique smMIPs within the set per species increase the specificity and sensitivity of CiRNAseq for CVM profiling.

Additional file 7 Dataset File 1. In vitro experiment (.xlsx). First sheet. Part A: mixed bacteria and negative control. Second sheet. Part B: mixed *Prevotella spp.*, *Lactobacillus spp.*, and negative control. Values represent the percentage (%) of reactive smMIPs in each species.

Additional file 8. Raw data File 1. In vitro experiment A (.xlsx). Raw data from SeqNext following CiRNAseq for the in vitro experiment, part A. Values represent URC.

Additional file 9. In silico alignment analyses (.pdf). Results obtained with Clustal Omega v1.2.4 for Supplementary Figures 1 and 3.

Additional file 10. Raw data File 2. In vitro experiment B (.xlsx). Raw data from SeqNext following CiRNAseq for the in vitro experiment, part B. Values represent URC.

Additional file 11 Supplementary Figure 2. *E. coli* growth experiment. A Two replicates of the samples subjected to CiRNAseq shows the reproducibility of the technique by the number of unique read counts (URC) obtained in each replicate. B Mean of the replicates' URC. C OD and RNA concentrations analyzed in time points five (growth), eight (autoclaved), and nine (antibiotic) were comparable with each other. However, as expected, the *E. coli* growth sample from time point eight had no URC after autoclaving, while the sample treated with cefoxitin had low URC, suggesting inhibition of bacterial metabolic activities. OD and RNA concentrations were measured before autoclaving and antibiotic treatment. T: time in hours; OD: optical density; URC: unique read counts; ng/μL: RNA concentration. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$; NS, not significant.

Additional file 12 Supplementary Figure 3. CiRNAseq specificity towards *Bifidobacterium* and *Gardnerella* species. 16S rRNA-seq targets two variable

regions (VRs) of the 16S rRNA gene using a forward and a reverse primer (e.g., V3 and V4). Alternatively, CiRNAseq targets five VRs of the 16S sub-unit using five singular smMIPs to differentiate species of *B. longum* and *G. vaginalis*. There is a high percentage of similarity (>90%) when comparing the V3-V4 regions of *B. longum* and *G. vaginalis*, which, if it is not appropriately amplified, could result in misidentification. In contrast, the CiRNAseq ROIs for both species have different levels of identity. Two out of five ROIs also possess a high percentage of similarity (>90%), with both amplifying the V7 and V8 VRs, and needed to identify these microbes at the family level. The rest remaining three out of five ROIs share less than 45% of similarity, which endorses the resolution and specificity of CiRNAseq in detecting both species. The color red represents similarity in sequences, while the color black represents no similarity.

Additional file 13. Raw data File 3. 16S rRNA-seq (.txt). Raw data from QIIME following microbiome profiling of ten cervical smears using 16S rRNA-seq. Corresponding IDs (manuscript ID = raw data ID): SN1A = KA1, SN2A = S2, SN3A = S5, SN4A = KA6, SN5A = S7, SN6A = S8, SN7A = S9.

Additional file 14. Raw data File 4. CiRNAseq (.xlsx). Raw data from SeqNext following microbiome profiling of eight cervical smears using CiRNAseq. The first seven sheets represent the seven samples included in the final analyses. The last sheet shows the sample with <1000 URC and excluded from the calculations. Values represent URC.

Additional file 15. Dataset File 2. 16S rRNA-seq vs CiRNAseq (.csv). Relative abundances for the 38 genera analyzed in seven cervical smears profiled with both 16S rRNA-seq and CiRNAseq.

Additional file 16 Supplementary Figure 4. Pearson's correlation for *Lactobacillus*, *Gardnerella*, *Atopobium* and *Megasphaera*. Pearson's positive correlation obtained from comparing the detection of *Lactobacillus* (A) ($r = 0.9605$, $p = 0.0006$), *Gardnerella* (B) ($r = 0.9384$, $p = 0.0018$), *Atopobium* (C) ($r = 0.9255$, $p = 0.0028$), and *Megasphaera* ($r = 0.8344$, $p = 0.0196$) using both CiRNAseq and 16S rRNA-seq corroborates the specificity and sensitivity of CiRNAseq.

Additional file 17. Dataset File 3. CiRNAseq resolution (.csv). List of all species detected with CiRNAseq in the same seven cervical smears that were initially compared to 16S rRNA-seq. Species annotation is indicated by 1. No annotation is indicated by 0.

Additional file 18. Raw data File 5. hrHPV cohort (.xlsx). Raw data from SeqNext following microbiome profiling of 92 cervical smears using CiRNAseq. Values represent URC.

Additional file 19. Dataset File 4. hrHPV cohort (.csv). Cohort of 46 hrHPV negative samples and 46 hrHPV positive samples (CIN2+), for which the microbiomes were profiled using CiRNAseq. Species annotation was performed with a custom R script. The CVMP includes bacterial isolates. Values represent URC.

Additional file 20. Loading scores from PCA analysis (.csv). PCA loading scores showing that anaerobic bacteria have the higher correlation with PC1.

Additional file 21 Supplementary Figure 5. Average analysis of hrHPV cohort. Average analysis of hrHPV negative versus hrHPV positive with CIN2+ demonstrates a strong association of absence and presence of particular microbes. *L. acidophilus*, *L. jensenii* and *L. crispatus* were highly present in hrHPV negative women, but the microbiome seemingly changes to a diverse-anaerobic microbiota in hrHPV positive women with CIN2+.

Additional file 22 Supplementary Figure 6. LefSe analysis: relative abundances association with hrHPV status. Relative abundance counts of *G. vaginalis* (A), *M. genomosp type 1* (B), *S. amnii* (C), *S. sanguinegens* (D), *P. anaerobius* (E), *D. microaerophilus* (F), *A. vaginae* (G), *P. amnii* (H), and *P. buccalis* (I) were found significantly over-represented in hrHPV positive women whereas *Lactobacillus acidophilus* (J) was enriched in hrHPV negative women.

Acknowledgements

We thank Judith Kuijpers, Danny Quint, Jeanine Boot, and Corina van den Heuvel for their support to this work. We also thank Prof. Janneke van de Wiggert for helping in identifying the list of cervicovaginal species. BioRender.com was used to design figures for the manuscript.

Authors' contributions

KA, WM, and WL conceptualized the study. MQ designed and validated in silico the smMIPs, supervised by MH and BED. DE performed the CIRMseq. KA and BS performed the in vitro experiments. BP performed the bioinformatic analyses for hrHPV status. TE did the bioinformatic analyses for 16S rRNA-seq. MM performed the data analyses for the manuscript, supervised by KA, WM, and WL. MM drafted the manuscript which was revised by all authors (KA, MQ, MR, BP, BS, MH, BED, TE, DE, AS, DL, RB, WL, and WM). All authors approved the manuscript and contributed to the final version for publication.

Funding

This work was supported by a research grant obtained from the Ruby and Rose Foundation. MM was further supported by a scholarship from Secretaría Nacional de Ciencia, Tecnología e Innovación de Panamá (SENACYT, #270-2019-136). BED was supported by the European Research Council (ERC) Consolidator grant 865694: DiversiPHI.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files or are available from public repositories. The sequence read data generated in this study are available at EMBL in the European Nucleotide Archive, project PRJEB45937 [85].

Declarations**Ethics approval and consent to participate**

The regional institutional review board and the National Institute for Public Health and Environment granted approval before the start of the study (No. 2014-1295).

Consent for publication

Not applicable.

Competing interests

WL is CSO and shareholder of Predica Diagnostics. The other authors declare no competing interests.

Author details

¹Department of Medical Microbiology, Radboud University Medical Center, 6500, HB, Nijmegen, The Netherlands. ²Department of Biochemistry, Radboud Institute for Molecular Life Sciences, 6525, GA, Nijmegen, The Netherlands. ³Department of Medical Microbiology, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands. ⁴Center for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, 6525, GA, Nijmegen, The Netherlands. ⁵Theoretical Biology and Bioinformatics, Science for Life, Utrecht University, Padualaan 8, 3584, CH, Utrecht, The Netherlands. ⁶Department of Pathology, Radboud University Medical Center, 6500, HB, Nijmegen, The Netherlands. ⁷Department of Obstetrics and Gynecology, Radboud University Medical Center, 6500, HB, Nijmegen, The Netherlands. ⁸Department of Obstetrics and Gynecology, Catharina Hospital, 5602, ZA, Eindhoven, The Netherlands. ⁹GROW, School for Oncology & Developmental Biology, Maastricht University Medical Centre, 6200, MD, Maastricht, The Netherlands. ¹⁰Predica Diagnostics, Toernooiveld 1, 6525, GA, Nijmegen, The Netherlands.

Received: 1 June 2021 Accepted: 2 December 2021

Published online: 16 December 2021

References

1. Ferlay J. Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer; 2018. Available from: <https://gco.iarc.fr/today>. Accessed 23 Nov 2020
2. Murphy K, Mitchell CM. The interplay of host immunity, environment and the risk of bacterial vaginosis and associated reproductive health outcomes. *J Infect Dis.* 2016;214(suppl_1):S29–35.
3. Brotman RM, Shardell MD, Gajer P, Tracy JK, Zenilman JM, Ravel J, et al. Interplay between the temporal dynamics of the vaginal microbiota and human papillomavirus detection. *J Infect Dis.* 2014;210(11):1723–33. <https://doi.org/10.1093/infdis/jiu330>.
4. Moscicki A-B, Shi B, Huang H, Barnard E, Li H. Cervical-vaginal microbiome and associated cytokine profiles in a prospective study of HPV 16 acquisition, persistence, and clearance. *Front Cell Infect Microbiol.* 2020;10:528.
5. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, et al. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci.* 2011;108(Supplement 1):4680.
6. Chaban B, Links MG, Jayaprakash TP, Wagner EC, Bourque DK, Lohn Z, et al. Characterization of the vaginal microbiota of healthy Canadian women through the menstrual cycle. *Microbiome.* 2014;2(1):23. <https://doi.org/10.1186/2049-2618-2-23>.
7. MacIntyre DA, Chandiramani M, Lee YS, Kindinger L, Smith A, Angelopoulos N, et al. The vaginal microbiome during pregnancy and the postpartum period in a European population. *Sci Rep.* 2015;5(1):8988. <https://doi.org/10.1038/srep08988>.
8. Vodstrcil LA, Twin J, Garland SM, Fairley CK, Hocking JS, Law MG, et al. The influence of sexual activity on the vaginal microbiota and Gardnerella vaginalis clade diversity in young women. *PLoS One.* 2017;12(2):e0171856. <https://doi.org/10.1371/journal.pone.0171856>.
9. Koedooder R, Singer M, Schoenmakers S, Savelkoul PHM, Morré SA, de Jonge JD, et al. The vaginal microbiome as a predictor for outcome of in vitro fertilization with or without intracytoplasmic sperm injection: a prospective study. *Hum Reprod.* 2019;34(6):1042–54. <https://doi.org/10.1093/humrep/dez065>.
10. Gajer P, Brotman RM, Bai G, Sakamoto J, Schütte UME, Zhong X, et al. Temporal dynamics of the human vaginal microbiota. *Sci Transl Med.* 2012;4(132):132ra52.
11. Nunn KL, Wang Y-Y, Harit D, Humphrys MS, Ma B, Cone R, et al. Enhanced trapping of HIV-1 by human cervicovaginal mucus is associated with Lactobacillus crispatus-dominant microbiota. *mBio.* 2015;6(5):e01084-15.
12. Onderdonk AB, Delaney ML, Fichorova RN. The human microbiome during bacterial vaginosis. *Clin Microbiol Rev.* 2016;29(2):223–38. <https://doi.org/10.1128/CMR.00075-15>.
13. Liu M-B, Xu S-R, He Y, Deng G-H, Sheng H-F, Huang X-M, et al. Diverse vaginal microbiomes in reproductive-age women with vulvovaginal candidiasis. *PLoS One.* 2013;8(11):e79812. <https://doi.org/10.1371/journal.pone.0079812>.
14. Mitra A, MacIntyre DA, Marchesi JR, Lee YS, Bennett PR, Kyrgiou M. The vaginal microbiota, human papillomavirus infection and cervical intraepithelial neoplasia: what do we know and where are we going next. *Microbiome.* 2016;4(1):58. <https://doi.org/10.1186/s40168-016-0203-0>.
15. Mitra A, MacIntyre DA, Lee YS, Smith A, Marchesi JR, Lehne B, et al. Cervical intraepithelial neoplasia disease progression is associated with increased vaginal microbiome diversity. *Sci Rep.* 2015;5(1):16865. <https://doi.org/10.1038/srep16865>.
16. Molina MA, Carosi Diatricch L, Castany Quintana M, Melchers WJG, Andralojc KM. Cervical cancer risk profiling: molecular biomarkers predicting the outcome of hrHPV infection. *Expert Rev Mol Diagn.* 2020;20(11):1–22. <https://doi.org/10.1080/14737159.2020.1835472>.
17. Kyrgiou M, Mitra A, Moscicki A-B. Does the vaginal microbiota play a role in the development of cervical cancer. *Transl Res.* 2017;179:168–82. <https://doi.org/10.1016/j.trsl.2016.07.004>.
18. Mitra A, MacIntyre DA, Ntrisos G, Smith A, Tsilidis KK, Marchesi JR, et al. The vaginal microbiota associates with the regression of untreated cervical intraepithelial neoplasia 2 lesions. *Nat Commun.* 2020;11(1):1999. <https://doi.org/10.1038/s41467-020-15856-y>.
19. Bik EM, Bird SW, Bustamante JP, Leon LE, Nieto PA, Addae K, et al. A novel sequencing-based vaginal health assay combining self-sampling, HPV detection and genotyping, STI detection, and vaginal microbiome analysis. *PLoS One.* 2019;14(5):e0215945. <https://doi.org/10.1371/journal.pone.0215945>.
20. Yang Q, Wang Y, Wei X, Zhu J, Wang X, Xie X, et al. The alterations of vaginal microbiome in HPV16 infection as identified by shotgun metagenomic sequencing. *Front Cell Infect Microbiol.* 2020;10:286. <https://doi.org/10.3389/fcimb.2020.00286>.
21. Berman HL, McLaren MR, Callahan BJ. Understanding and interpreting community sequencing measurements of the vaginal microbiome. *BJOG.* 2020;127(2):139–46. <https://doi.org/10.1111/1471-0528.15978>.
22. Hong KH, Hong SK, Cho SI, Ra E, Han KH, Kang SB, et al. Analysis of the vaginal microbiome by next-generation sequencing and evaluation of its performance as a clinical diagnostic tool in vaginitis. *Ann Lab Med.* 2016;36(5):441–9. <https://doi.org/10.3343/alm.2016.36.5.441>.

23. Clarridge JE. Impact of 16 s rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev.* 2004;17(4):840–62. <https://doi.org/10.1128/CMR.17.4.840-862.2004>.
24. White T, Bruns T, Lee S, Taylor J, Innis M, Gelfand D, et al. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols: Guide Methods Appl.* 1990;18:315–22.
25. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 2012;6(8):1621–4. <https://doi.org/10.1038/ismej.2012.8>.
26. Earl JP, Adappa ND, Krol J, Bhat AS, Balashov S, Ehrlich RL, et al. Species-level bacterial community profiling of the healthy sinonasal microbiome using Pacific Biosciences sequencing of full-length 16S rRNA genes. *Microbiome.* 2018;6(1):190. <https://doi.org/10.1186/s40168-018-0569-2>.
27. Hugerth LW, Pereira M, Zha Y, Seifert M, Kaldhusdal V, Boulund F, et al. Assessment of in vitro and in silico protocols for sequence-based characterization of the human vaginal microbiome. *mSphere.* 2020;5(6):e00448–20. <https://doi.org/10.1128/mSphere.01253-20>.
28. Graspentner S, Loeper N, Künzel S, Baines JF, Rupp J. Selection of validated hypervariable regions is crucial in 16S-based microbiota studies of the female genital tract. *Sci Rep.* 2018;8(1):9678. <https://doi.org/10.1038/s41598-018-27757-8>.
29. Zeeuwen PLJM, Boekhorst J, Ederveen THA, Kleerebezem M, Schalkwijk J, van Hijum SAFT, et al. Reply to Meisel et al. *J Invest Dermatol.* 2017;137(4):961–2. <https://doi.org/10.1016/j.jid.2016.11.013>.
30. Pinna NK, Dutta A, Monzooral Haque M, Mande SS. Can targeting non-contiguous V-regions with paired-end sequencing improve 16 s rRNA-based taxonomic resolution of microbiomes?: an in silico evaluation. *Front Genet.* 2019;10:653. <https://doi.org/10.3389/fgene.2019.00653>.
31. de Bitter T, van de Water C, van den Heuvel C, Zeelen C, Eijkelenboom A, Tops B, et al. Profiling of the metabolic transcriptome via single molecule molecular inversion probes. *Sci Rep.* 2017;7(1):11402. <https://doi.org/10.1038/s41598-017-11035-0>.
32. van den Heuvel CNAM, van Ewijk A, Zeelen C, de Bitter T, Huynen M, Mulders P, et al. Molecular profiling of druggable targets in clear cell renal cell carcinoma through targeted RNA sequencing. *Front Oncol.* 2019;9:117. <https://doi.org/10.3389/fonc.2019.00117>.
33. van den Heuvel CNAM, Das AI, de Bitter T, Simmer F, Wurdinger T, Molina-Vila MA, et al. Quantification and localization of oncogenic receptor tyrosine kinase variant transcripts using molecular inversion probes. *Sci Rep.* 2018; 8(1):7072. <https://doi.org/10.1038/s41598-018-25328-5>.
34. Lenting K, van den Heuvel CNAM, van Ewijk A, Elmelik D, de Boer R, Tindall E, et al. Mapping actionable pathways and mutations in brain tumours using targeted RNA next generation sequencing. *Acta Neuropathol Commun.* 2019;7(1):185. <https://doi.org/10.1186/s40478-019-0826-z>.
35. van den Heuvel CNAM, Loopik DL, Ebisch RMF, Elmelik D, Andralojc KM, Huynen M, et al. RNA-based high-risk HPV genotyping and identification of high-risk HPV transcriptional activity in cervical tissues. *Mod Pathol.* 2020; 33(4):748–57. <https://doi.org/10.1038/s41379-019-0369-7>.
36. Boyle EA, O’Roak BJ, Martin BK, Kumar A, Shendure J. MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. *Bioinformatics.* 2014;30(18):2670–2. <https://doi.org/10.1093/bioinformatics/btu353>.
37. Eijkelenboom A, Kamping EJ, Kastner-van Raaij AW, Hendriks-Cornelissen SJ, Neveling K, Kuiper RP, et al. Reliable next-generation sequencing of formalin-fixed, paraffin-embedded tissue using single molecule tags. *J Mol Diagn.* 2016;18(6):851–63. <https://doi.org/10.1016/j.jmoldx.2016.06.010>.
38. Martin DH, Marrazzo JM. The vaginal microbiome: current understanding and future directions. *J Infect Dis.* 2016;214(suppl_1):S36–41.
39. Fettweis J, Alves J, Borzelleca J, Brooks J, Friedline C, Gao Y, et al. The vaginal microbiome: disease, genetics and the environment. *Nat Precedings.* 2010:1. <https://doi.org/10.1038/npre.2010.5150.1>.
40. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2016;44(D1):D7–D19. <https://doi.org/10.1093/nar/gkv1290>.
41. Drost H-G, Paszkowski J. Biomartr: genomic data retrieval with R. *Bioinformatics.* 2017;33(8):1216–7. <https://doi.org/10.1093/bioinformatics/btw821>.
42. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics.* 2009;25(11):1422–3. <https://doi.org/10.1093/bioinformatics/btp163>.
43. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26(6):841–2. <https://doi.org/10.1093/bioinformatics/btq033>.
44. Ludwig W, Schleifer KH. Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. *FEMS Microbiol Rev.* 1994;15(2-3):155–73. <https://doi.org/10.1111/j.1574-6976.1994.tb00132.x>.
45. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, et al. The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. *Nucleic Acids Res.* 2014;42(D1):D643–D8. <https://doi.org/10.1093/nar/gkt1209>.
46. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schäffer AA. Database indexing for production MegaBLAST searches. *Bioinformatics.* 2008;24(16):1757–64. <https://doi.org/10.1093/bioinformatics/btn322>.
47. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* 2009;25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
48. Hamady M, Knight R. Microbial community profiling for human microbiome projects: tools, techniques, and challenges. *Genome Res.* 2009;19(7):1141–52. <https://doi.org/10.1101/gr.085464.108>.
49. Ederveen THA, Smits JPH, Hajo K, van Schalkwijk S, Kouwenhoven TA, Lukovac S, et al. A generic workflow for Single Locus Sequence Typing (SLST) design and subspecies characterization of microbiota. *Sci Rep.* 2019; 9(1):19834. <https://doi.org/10.1038/s41598-019-56065-y>.
50. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate illumina Paired-End reAd mergeR. *Bioinformatics.* 2014;30(5):614–20. <https://doi.org/10.1093/bioinformatics/btt593>.
51. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.* 2010;7(5):335–6. <https://doi.org/10.1038/nmeth.f.303>.
52. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 2011;7(1):539. <https://doi.org/10.1038/msb.2011.75>.
53. Meier-Kolthoff JP, Göker M, Spröer C, Klenk H-P. When should a DDH experiment be mandatory in microbial taxonomy. *Arch Microbiol.* 2013; 195(6):413–8. <https://doi.org/10.1007/s00203-013-0888-4>.
54. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics.* 2013;14(1):60. <https://doi.org/10.1186/1471-2105-14-60>.
55. Metsalu T, Vilo J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res.* 2015;43(W1):W566–W70. <https://doi.org/10.1093/nar/gkv468>.
56. Dhakan DB, Maji A, Sharma AK, Saxena R, Pulikkan J, Grace T, et al. The unique composition of Indian gut microbiome, gene catalogue, and associated fecal metabolome deciphered using multi-omics approaches. *Gigascience.* 2019;8(3):giz004.
57. Watanabe H, Nakamura I, Mizutani S, Kurokawa Y, Mori H, Kurokawa K, et al. Minor taxa in human skin microbiome contribute to the personal identification. *PLoS One.* 2018;13(7):e0199947. <https://doi.org/10.1371/journal.pone.0199947>.
58. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome.* 2017;5(1):27. <https://doi.org/10.1186/s40168-017-0237-y>.
59. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. *Genome Biol.* 2011; 12(6):R60. <https://doi.org/10.1186/gb-2011-12-6-r60>.
60. Blazewicz SJ, Barnard RL, Daly RA, Firestone MK. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *ISME J.* 2013;7(11):2061–8. <https://doi.org/10.1038/ismej.2013.102>.
61. Ghartey JP, Smith BC, Chen Z, Buckley N, Lo Y, Ratner AJ, et al. *Lactobacillus crispatus* dominant vaginal microbiome is associated with inhibitory activity of female genital tract secretions against *Escherichia coli*. *PLoS One.* 2014; 9(5):e96659. <https://doi.org/10.1371/journal.pone.0096659>.
62. Cools P. The role of *Escherichia coli* in reproductive health: state of the art. *Res Microbiol.* 2017;168(9):892–901. <https://doi.org/10.1016/j.resmic.2017.02.002>.
63. Curty G, Costa RL, Siqueira JD, Meyrelles AI, Machado ES, Soares EA, et al. Analysis of the cervical microbiome and potential biomarkers from

- postpartum HIV-positive women displaying cervical intraepithelial lesions. *Sci Rep.* 2017;7(1):17364. <https://doi.org/10.1038/s41598-017-17351-9>.
64. Usyk M, Zolnik CP, Castle PE, Porras C, Herrero R, Gradissimo A, et al. Cervicovaginal microbiome and natural history of HPV in a longitudinal study. *PLoS Pathol.* 2020;16(3):e1008376. <https://doi.org/10.1371/journal.ppat.1008376>.
 65. Dareng EO, Ma B, Famooto AO, Adebamowo SN, Offiong RA, Olaniyan O, et al. Prevalent high-risk HPV infection and vaginal microbiota in Nigerian women. *Epidemiol Infect.* 2016;144(1):123–37. <https://doi.org/10.1017/S0950268815000965>.
 66. Chen Y, Hong Z, Wang W, Gu L, Gao H, Qiu L, et al. Association between the vaginal microbiome and high-risk human papillomavirus infection in pregnant Chinese women. *BMC Infect Dis.* 2019;19(1):677. <https://doi.org/10.1186/s12879-019-4279-6>.
 67. Kwasniewski W, Wolun-Cholewa M, Kotarski J, Warchol W, Kuzma D, Kwasniewska A, et al. Microbiota dysbiosis is associated with HPV-induced cervical carcinogenesis. *Oncol Lett.* 2018;16(6):7035–47. <https://doi.org/10.3892/ol.2018.9509>.
 68. Shin J, Lee S, Go M-J, Lee SY, Kim SC, Lee C-H, et al. Analysis of the mouse gut microbiome using full-length 16S rRNA amplicon sequencing. *Sci Rep.* 2016;6(1):29681. <https://doi.org/10.1038/srep29681>.
 69. van der Veer C, Hertzberger RY, Bruisten SM, Tytgat HLP, Swanenburg J, de Kat Angelino-Bart A, et al. Comparative genomics of human *Lactobacillus crispatus* isolates reveals genes for glycosylation and glycogen degradation: implications for in vivo dominance of the vaginal microbiota. *Microbiome.* 2019;7(1):49. <https://doi.org/10.1186/s40168-019-0667-9>.
 70. Borgdorff H, Armstrong SD, Tytgat HLP, Xia D, Ndayisaba GF, Wastling JM, et al. Unique insights in the cervicovaginal *Lactobacillus iners* and *L. crispatus* proteomes and their associations with microbiota dysbiosis. *PLoS One.* 2016;11(3):e0150767.
 71. Fuks G, Elgart M, Amir A, Zeisel A, Turnbaugh PJ, Soen Y, et al. Combining 16S rRNA gene variable regions enables high-resolution microbial community profiling. *Microbiome.* 2018;6(1):17. <https://doi.org/10.1186/s40168-017-0396-x>.
 72. Nelson TM, Borgogna J-LC, Brotman RM, Ravel J, Walk ST, Yeoman CJ. Vaginal biogenic amines: biomarkers of bacterial vaginosis or precursors to vaginal dysbiosis. *Front Physiol.* 2015;6:253.
 73. Clarke MA, Rodriguez AC, Gage JC, Herrero R, Hildesheim A, Wacholder S, et al. A large, population-based study of age-related associations between vaginal pH and human papillomavirus infection. *BMC Infect Dis.* 2012;12(1):33. <https://doi.org/10.1186/1471-2334-12-33>.
 74. Lavitola G, Della Corte L, De Rosa N, Nappi C, Bifulco G. Effects on vaginal microbiota restoration and cervical epithelialization in positive HPV patients undergoing vaginal treatment with carboxy-methyl-beta-glucan. *Biomed Res Int.* 2020;2020:5476389. <https://doi.org/10.1155/2020/5476389>.
 75. Chao X, Sun T, Wang S, Tan X, Fan Q, Shi H, et al. Research of the potential biomarkers in vaginal microbiome for persistent high-risk human papillomavirus infection. *Ann Transl Med.* 2020;8(4):100. <https://doi.org/10.21037/atm.2019.12.115>.
 76. Witkin SS, Linhares IM. Why do lactobacilli dominate the human vaginal microbiota. *BJOG.* 2017;124(4):606–11. <https://doi.org/10.1111/1471-0528.14390>.
 77. Chee WJY, Chew SY, Than LTL. Vaginal microbiota and the potential of *Lactobacillus* derivatives in maintaining vaginal health. *Microb Cell Fact.* 2020;19(1):203. <https://doi.org/10.1186/s12934-020-01464-4>.
 78. Satpute SK, Mone NS, Das P, Banat IM, Banpurkar AG. Inhibition of pathogenic bacterial biofilms on PDMS based implants by *L. acidophilus* derived biosurfactant. *BMC Microbiol.* 2019;19(1):39.
 79. Reimers LL, Mehta SD, Massad LS, Burk RD, Xie X, Ravel J, et al. The cervicovaginal microbiota and its associations with human papillomavirus detection in HIV-infected and HIV-uninfected women. *J Infect Dis.* 2016; 214(9):1361–9. <https://doi.org/10.1093/infdis/jiw374>.
 80. Petrova MI, Reid G, Vaneechoutte M, Lebeer S. *Lactobacillus iners*: friend or foe. *Trends Microbiol.* 2017;25(3):182–91. <https://doi.org/10.1016/j.tim.2016.11.007>.
 81. Hornung BVH, Zwittink RD, Kuijper EJ. Issues and current standards of controls in microbiome research. *FEMS Microbiol Ecol.* 2019;95(5):fiz045.
 82. Johnson JS, Spakowicz DJ, Hong B-Y, Petersen LM, Demkowicz P, Chen L, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun.* 2019;10(1):5029. <https://doi.org/10.1038/s41467-019-13036-1>.
 83. Thomas AM, Segata N. Multiple levels of the unknown in microbiome research. *BMC Biol.* 2019;17(1):48. <https://doi.org/10.1186/s12915-019-0667-z>.
 84. Ravel J, Brotman RM. Translating the vaginal microbiome: gaps and challenges. *Genome Med.* 2016;8(1):35. <https://doi.org/10.1186/s13073-016-0291-2>.
 85. Andralojc KM, Molina MA, et al. Novel high-resolution targeted sequencing of the cervicovaginal microbiome. European Nucleotide Archive. Hinxton, United Kingdom: EMBL; 2021. Available from: <https://www.ebi.ac.uk/ena/browser/view/PRJEB45937>. Accessed 27 Nov 2021.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

