MinION™, a portable long-read sequencer, enables rapid vaginal microbiota analysis in a clinical setting

Shinnosuke Komiya[1,2], Yoshiyuki Matsuo[3], So Nakagawa[4], Yoshiharu Morimoto[2], Kirill Kryukov[5], Hidetaka Okada[1], and Kiichi Hirota[3]

1 Department of Obstetrics and Gynecology, Kansai Medical University Graduate School of Medicine, Osaka, Japan

2 HORAC Grand Front Osaka Clinic, Osaka, Japan

3 Department of Human Stress Response Science, Institute of Biomedical Science, Kansai Medical University, Osaka, Japan

4 Department of Molecular Life Science, Tokai University School of Medicine, Kanagawa, Japan

5 Department of Genomics and Evolutionary Biology, National Institute of Genetics, Shizuoka, Japan

Correspondence:

Yoshiyuki Matsuo

E-mail: ysmatsuo-kyt@umin.ac.jp

Keywords:

16S rRNA; bacterial vaginosis; long-read sequencer; MinION™; nanopore

Abbreviations:

BV, bacterial vaginosis; CST, community state type; PCoA, principal coordinate analysis; PCR, polymerase chain reaction; rRNA, ribosomal RNA.

28 **Running heading:**

29 MinION™ enables rapid vaginal microbiota analysis.

30

Abstract

It has been suggested that the local microbiota in the reproductive organs is relevant to women's health and may also affect pregnancy outcomes. Analysis of partial 16S ribosomal RNA (rRNA) gene sequences generated by short-read sequencers has been used to identify vaginal and endometrial microbiota, but it requires a long time to obtain the results, making it unsuitable for the rapid analysis of small samples in a clinical context. We demonstrated a simple workflow using the nanopore sequencer MinION™ that allows high-resolution and rapid differentiation of vaginal microbiota. Vaginal samples collected from 18 participants were subjected to DNA extraction and full-length 16S rRNA gene sequencing with MinION™. The principal coordinate analysis showed no differences in the bacterial compositions regardless of the sample collection method. The vaginal microbiota results could be reported within 2 days of specimen receipt. Although bacterial vaginosis (BV) was not diagnosed by the Nugent score in any cases, groups with both healthy and BV-like vaginal microbiota were clearly characterized by MinION™ sequencing. We conclude that full-length 16S rRNA gene sequencing analysis with MinION™ provides a rapid means for identifying vaginal bacteria with higher resolution. Species-level profiling of human vaginal microbiota by MinION™ sequencing can allow the analysis of associations with conditions such as genital infections, endometritis, and threatened miscarriage.

Introduction

Bacterial vaginosis (BV) is one of the most common gynecological disorders among women of reproductive age [1]. It is caused by disruption of the vaginal microbiota due to various factors such as psychological stress, poor physical condition, menstrual cycle, and pregnancy. It has relatively mild subjective symptoms, such as abnormal vaginal discharge (odor, color, amount of discharge) and itching, but there are cases of unrecognized infection and recurrent infections without significant symptoms [1]. BV has been clinically reported to cause adverse reproductive outcomes, including sexually transmitted infections, genitourinary viral infections such as HIV-1/2, HPV, and HSV-2 [2], and preterm delivery [3].

Currently, Amsel's criteria based on clinical outcomes [4] and the Nugent score using Gram stain findings [5] are used as diagnostic criteria for BV. These methods are easy to use, but since they are based on the subjective opinion of the physician, the diagnosis largely depends on the skill of the examiner [6]. Bacterial culture of vaginal discharge is also a standard method to investigate the vaginal microbiota, but some bacteria are difficult to culture [6]. Moreover, in routine gynecological practice, vaginal discharge is not sampled unless patients have symptoms suggestive of vaginosis. Therefore, accurately assessing the vaginal microbiota has been difficult with these regular clinical methods. Due to the lack of a reliable methodology, the association of BV with other vaginal disorders has not been fully elucidated. Until recently, a more detailed analysis of vaginal microbiota in asymptomatic women was considered excessive and might not be clinically relevant. Given that BV has been implicated in women's health outcomes in various ways, it is crucial to develop a rapid and accurate means to diagnose BV with higher sensitivity. The increased risk of various infections, infertility, miscarriage and preterm birth due to BV has been shown to be impacted by the species that make up the vaginal microbiota [1-3]. In this context, accurate representation of the vaginal microbiota will allow for more appropriate medical treatment. If abnormal vaginal

78 microbiota is identified with persistent cervical lesions, the dysbiosis should be treated
79 even in the absence of symptoms. Detection of asymptomatic BV will potentially be
80 beneficial for the diagnosis and prevention of gynecological diseases.
81 16S ribosomal RNA (rRNA) gene sequencing analysis using next-generation sequencers
82 can be used to identify bacteria without culture in a manner not dependent on subjective
83 assessment by the examiner. Previous studies have shown that the vaginal microbiota of
84 healthy women is classified into community state type (CST) I to V [7]. CST is defined
85 by the dominant bacteria as follows: CST I: *Lactobacillus crispatus*, CST II:
86 *Lactobacillus gasseri*, CST III: *Lactobacillus iners*, CST IV: lack of *Lactobacillus* spp.,
87 and CST V: *Lactobacillus jensenni*. These narrow definitions mean that undiagnosed
88 cases of vaginal dysbiosis may be present. This new classification of the vaginal
89 microbiota has been clarified by performing sequencing analysis on cases with no
90 clinical symptoms. Because a deeper and more accurate understanding of the vaginal
91 microbiota may provide new clinical insights, we should demonstrate that 16S rRNA
92 analysis is quick and easy to perform in a broad range of contexts in clinical practice.
93 MinION™ detects the signal of a DNA nucleotide that passes through a nanopore
94 arranged on a flow cell [8]. This long-read sequencing technology can be used to
95 identify bacterial microbiota with the whole region of the 16S rRNA. MinION™ is
96 smaller, lighter, and easier to deploy than conventional massively parallel sequencing
97 platforms, and it also offers the advantage of high-resolution microbiota analysis with
98 long-read sequencing [9-11]. We constructed an in-hospital vaginal microbiota-
99 analyzing workflow using MinION™. This is the first report to demonstrate that
100 MinION™ can provide rapid patient feedback regarding vaginal microbiota analysis.
101

102 **Materials and methods**

103 **Ethics statement**

104 The study adhered to the Declaration of Helsinki throughout the protocol, and written

105 informed consent was obtained from all participants. This study is a cross-sectional
106 observational trial approved by the institutional review board of the medical corporation
107 Sankeikai (Approval Number: 2019-38). After the approval, the study protocol was
108 submitted to the UMIN-CTR clinical trial registry site
109 (https://www.umin.ac.jp/ctr/index.htm) (trial registration number: UMIN000038032),
110 which met the criteria of the International Committee of Medical Journal Editors
111 (ICMJE).
112
113 **Recruitment of participants**
114 From April 2019 to May 2020, participants who met all of the following enrollment
115 criteria were recruited: 1) primary infertility patients attending HORAC Grand Front
116 Osaka Clinic in Osaka, Japan; 2) premenopausal Japanese women who were not
117 pregnant; 3) no clinical symptoms strongly suggestive of BV; 4) scheduled hormone
118 replacement embryo transfer cycle with frozen-thawed blastocyst; and 5) those who
119 provided written consent. The exclusion criteria were as follows: 1) known history of
120 HIV-1/2, hepatitis B/C, syphilis, or genital chlamydia infection; 2) known history of
121 diabetes; 3) known history of cervical/vaginal surgery; 4) known history of intrauterine
122 device use; and 5) known history of antibiotic, steroid, or vaginal suppository use within
123 2 weeks. In addition to collecting vaginal samples, details of age, body mass index, and
124 basal hormone levels (anti-Mullerian hormone, luteinizing hormone, and follicle-
125 stimulating hormone on days 2–4 of the menstrual cycle), medical history, and infertility
126 treatment history were collected from the medical records. All vaginal samples were
127 collected on the day of embryo transfer.
128
129 **Nugent score**
130 Vaginal smears were analyzed by two pathologists affiliated with an external laboratory,
131 and the Nugent score was calculated based on microscopic findings: the numbers of

132 *Lactobacillus* (scored as 0 to 4), *Gardnerella* (scored as 0 to 4), and *Mobiluncus* (scored
133 as 0 to 2). A total score of 0 to 3 was diagnosed as representative of healthy vaginal
134 microbiota, 4 to 6 as an intermediate group, and 7 or more as BV [5]. Inappropriate
135 samples with defective specimen collection were diagnosed as indeterminate. If BV was
136 diagnosed based on the Nugent score, our policy was to treat it using antibiotics.
137

138 **Vaginal sample collection method**

139 Vaginal lavage samples were collected from 18 participants; from four participants,
140 additional swab samples were collected with written consent. When both lavage and
141 swab samples were collected, the swab samples were collected first. For the lavage
142 method, the inside of the vagina was washed with 10 ml of sterile saline, and at least 2
143 ml was collected using a sterile syringe. Collected lavage samples were stored at –30 °C
144 until DNA extraction. For the swab method, OMNIgene vaginal kit (OMR-130; DNA
145 Genotek Inc., Ottawa, Canada) was used to infiltrate vaginal mucus from the
146 uterovaginal region, and the swab tip was stored in the preservative solution of the
147 sampling kit. Swab samples were stored at room temperature until DNA extraction in
148 accordance with the manufacturer's instructions.
149

150 **Sequencing sample preparation**

151 DNA was extracted from 22 human vaginal samples using the QIAamp UCP Pathogen
152 Mini Kit (QIAGEN, Venlo, Netherlands) following the manufacturer's instructions.
153 Briefly, samples were subjected to mechanical cell lysis by bead-beating [12], and DNA
154 was isolated using silica membrane-based spin columns. After extraction, the DNA
155 content was measured using a NanoDrop® 1000 Spectrophotometer (Thermo Fisher
156 Scientific, MA, USA), and the concentrations of extracted DNA from the swab and
157 lavage samples were 6.1–32.3 ng/µl and 53.1–1092.4 ng/µl, respectively.
158

159 **MinION™ sequencing**

160 As described in our previous report [12], a slightly modified version of four-primer
161 polymerase chain reaction (PCR) with rapid adapter attachment was performed
162 (https://dx.doi.org/10.17504/protocols.io.bs9knh4w). For amplification of the V1–9
163 region of the 16S rRNA gene, a forward primer (S-D-Bact-0008-c-S-20) with the
164 anchor sequence 5'-
165 TTTCTGTTGGTGCTGATATTGCAGRGTTYGATYMTGGCTCAG-3' and a reverse
166 primer with the anchor sequence 5'-
167 ACTTGCCTGTCGCTCTATCTTCCGGYTACCTTGTTACGACTT-3' were used as
168 inner primers. For amplification of the V3–4 region, 341F with the anchor sequence 5'-
169 TTTCTGTTGGTGCTGATATTGCCCTACGGGNGGCWGCAG-3' and 806R with the
170 anchor sequence 5'-
171 ACTTGCCTGTCGCTCTATCTTCGGACTACHVGGGTWTCTAAT-3' were used as
172 inner primers. PCR amplification of the 16S rRNA gene was performed using the
173 KAPA2G™ Robust HotStart ReadyMix PCR Kit (Kapa Biosystems, MA, USA) with an
174 inner primer pair (50 nM for each) and the outer primer mixture (10 nM) of PCR
175 Barcoding Kit (SQK-PBK004; Oxford Nanopore Technologies, Oxford, UK) in a total
176 volume of 25 µl. Each sample was assigned an individual barcode, which was used to
177 identify the PCR product. Amplification was performed under the following PCR
178 conditions: initial denaturation at 95 °C for 3 min; five cycles of 15 s at 95 °C, 15 s at
179 55 °C, and 30 s at 72 °C; and 30 cycles of 15 s at 95 °C, 15 s at 62 °C, and 30 s at
180 72 °C; followed by a final extension at 72 °C for 1 min. The amplification products
181 were confirmed using 1% agarose gel electrophoresis (1 × TAE buffer) to be
182 approximately 1600 base pairs (V1–9 region) and 400 base pairs (V3–4 region), which
183 are the expected amplification sizes. After confirmation, the amplified DNA was
184 purified using AMPure® XP (Beckman Coulter, CA, USA). A total of 100 ng (10 µl) of
185 purified DNA was incubated with 1 µl of Rapid Adapter at room temperature for 5 min.

The concentrations of swab samples and lavage samples were 8.3–34 ng/µl and 3.58–61 ng/µl, respectively, as determined using a Quantus™ Fluorometer (Promega, WI, USA). The prepared DNA library (total of 11 µl) was mixed with 34 µl of Sequencing Buffer, 25.5 µl of Loading Beads, and 4.5 µl of water. The final adjusted sample was loaded into the flow cell R.9.4.1 (FLO-MIN106; Oxford Nanopore Technologies Ltd.) and attached to the MinION™ Mk1B, which was connected to a personal computer. In this study, four to six samples were loaded at one time and each sequencing session lasted about 90 min. DNA sequencing data were acquired in FAST5 format using MINKNOW software ver. 1.11.5 (Oxford Nanopore Technologies Ltd.).

Bioinformatic analysis workflow

We performed the bioinformatic analysis using a pipeline of multiple programs built in accordance with our previous reports [9,11]. An overview is given below.

1) Computer: Apple iMac 27-inch, Late 2015 (OS, macOS 10.14.6; CPU, 3.3 GHz Intel Core i5-6600; memory, 16 GB)
2) GUPPY software ver. 3.1.5 (Oxford Nanopore Technologies Ltd.): base calling, conversion of FAST5 files to FASTQ files with quality control based on quality score > 7.
3) SeqKit software ver. 0.10.0 [13]: extraction of a read length of 1,300 to 1,950 bp, based on the 16S rRNA length registered in the SILVA rRNA database ver. 132 (https://www.arb-silva.de/) [14].
4) TANTAN program ver. 18 [15]: removing simple repeat sequences.
5) Minimap2 program ver. 2.14 [16]: eliminating human genome information (Human Genome Assembly GRCh38, https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/) and matching each genome read to the 5,850 representative bacterial genome sequences (Table S1) stored in the GenomeSync database (http://genomesync.org) [11,12].

6) In-house Perl scripts (Genome Search Toolkit, http://kirill-kryukov.com/study/tools/gstk/): selecting the species with the highest Minimap2 score and determining the taxa based on the NCBI taxonomy database.
7) Krona software version 2.7 [17]: visualizing the frequency of detected bacterial species in a given sample.

Statistical analysis

Principal coordinate analysis was performed and the results were visualized with R version 4.0.2.

Results

Characteristics of participants

Our study consisted of 18 Japanese, ranging from 30 to 43 years of age, with a median age of 36.5. All participants had primary infertility with normal menstrual cycles, along with no underlying medical conditions (Table 1).

Identification of vaginal bacteria

The composition of the vaginal microbiota was investigated by 16S rRNA gene amplicon sequencing on the MinION™ platform (Figure 1). The V1–9 MinION™ sequencing for almost 90 min yielded 15,836–119,745 reads at first and then filtered 10,688–103,212 reads satisfying the conditions of 1,300–1,950 bp and QC > 7 (details in Table 2). The reads were mapped against the GenomeSync reference database for taxonomic assignment. The results of Nugent scoring and microbiota analysis of the filtered 3,000 reads using MinION™ are shown in Table S2. Based on the Nugent score criteria, four out of 18 cases were in the intermediate group, but none of them were diagnosed as BV. The MinION™ sequencing data were also analyzed by an alternative bioinformatics tool. The FASTQ 16S workflow produced almost similar taxonomic

240 profiles (Table S3). This result confirmed the validity of our method for taxonomic
241 classification and the bacterial compositions were comparable regardless of the program
242 and database used.
243
244 **Comparison of the results of swab and lavage samples**
245 For subjects 1–4, lavage and swab samples were collected and bacterial DNA extraction
246 and 16S rRNA analysis were performed to determine the effect of the sample collection
247 method on the results of the microbiota analysis. After filtering, 3,000 reads were
248 randomly extracted in eight samples (1-4_lavage and 1-4_swab), the results of which
249 are shown in Figure 2, with almost the same identification with the two sampling
250 methods. We performed principal coordinate analysis (PCoA) to assess the equivalence
251 of the lavage and swab sampling methods, the results of which are shown in Figure 3.
252 Regardless of the sample collection method, the results of the analysis of the same
253 subjects involved almost the same coordinates.
254
255 **Comparison of the results of sequence read number**
256 To assess the effect of read count on the results of the MinION™ sequencing, three
257 samples that showed characteristic bacterial microbiota [5_lavage: *Lactobacillus iners*
258 dominant (>99%), 6_lavage: *Lactobacillus crispatus* dominant (>99%), 8_lavage: BV
259 like] were selected and their taxonomic profiles were compared after filtering for 3,000
260 reads (random sampling) and 10,000 reads (random sampling) (Figure 4). In all
261 specimens, the results of vaginal microbiota analysis in MinION™ were similar for
262 different numbers of reads.
263
264 **Comparison of the results of V1–9 and V3–4 16S rRNA gene sequencing**
265 For the taxonomic classification of bacteria, we compared the resolution of long-read
266 (V1–9) and short-read (V3–4) 16S amplicon sequencing for six samples (sample

numbers 5, 6, 9, 10, 11, and 14). Of these, for sample numbers 5, 9, and 10, *Lactobacillus iners* was confirmed to be present at a rate of more than 99% (Group I), while for sample numbers 6, 11, and 14, *Lactobacillus crispatus* was confirmed to be present at more than 99% (Group C) in the V1–9 16S rRNA analysis. The V3–4 region was amplified by four-primer PCR from the bacterial DNA of six samples and sequenced with MinION™. From each sequencing result, 3,000 reads were randomly extracted and the results of the analysis of the V1–9 region were compared with those of the V3–4 region (Figure 5). In Group I, *Lactobacillus iners* was classified as being present at a rate of more than 98% in the V3–4 S16 rRNA analysis, which was comparable to the results obtained by the V1-9 sequencing (Figure 5A). In Group C, the V3-4 sequence alignment resulted in ambiguous identification of *Lactobacillus* species (Figure 5B). All samples showed *Lactobacillus crispatus*, *Lactobacillus acidophilus*, *Lactobacillus helveticus*, *Lactobacillus amylovorus*, *Lactobacillus kefiranofaciens*, *Lactobacillus hamsteri*, *Lactobacillus gallinarum*, *Lactobacillus kalixensi*s, and *Lactobacillus psittaci* in the V3–4 S16 rRNA analysis as potentially existing species, which were found in similar proportions across all samples in Group C.

Discussion

The vaginal microbiota is characterized by particularly low diversity compared with the human gut harboring a complex microbial community. [18]. Conventional methods, such as Amsel's criteria, the Nugent score, and bacterial culture, are not suitable for an accurate bacterial identification. Next-generation sequencing has revolutionized the profiling of the bacterial microbiota. In short, the combination of metagenomic sequencing and bioinformatic technology has made it possible to more accurately assess the bacterial microbiota without the influence of uncertainties such as examiner subjectivity or the capturability of each bacterium [6]. The MinION™ platform allows a single flow cell to be used repeatedly by assigning individual barcodes. MinION™

allows data generation by immediately reading the amplicon sequences that pass through the nanopores placed on the substrate of the flow cell. The real-time data generation allows the sequencing to end when the data reach the target amount.

In particular, analysis of the vaginal microbiota is a field in which MinION™'s capabilities can be maximized because of the low complexity of this microbiota under healthy conditions. However, to date, no studies have examined clinical vaginal metagenomic samples with MinION™ technology. In the present study, the number of bacterial species detected with MinION™ ranged from 1 to 22 (mean value 3.3), confirming the low diversity of the vaginal microflora (Table S2). Although no cases were diagnosed as BV by the Nugent scoring method, the 16S rRNA gene analysis using MinION™ revealed the presence of BV-causing bacteria such as *Streptococcus*, *Atopobium*, *Prevotella*, and *Gardnerella* [7] in the group categorized as intermediate based on Nugent's criteria. For the most part, the 16S rRNA gene analysis appeared to correlate with the Nugent scoring to diagnose BV. However, there was an inconsistency between the two methods in one case (14_lavage), where it was categorized as intermediate, even though only *Lactobacillus* was identified by the MinION™ sequencing. While Nugent scoring is a subjective evaluation and depends on the acumen of the examiner, the molecular method targeting the 16S rRNA gene could serve as a practical and reliable measure for the diagnosis of BV. In vaginal microbiota analysis, it is important to be able to profile *Lactobacillus* spp. and the causative organisms of BV at the species level. The Nugent score is calculated based on the number of bacterial morphotypes and therefore has limitations for discriminating bacterial taxa at the species level. Meanwhile, the sequence-based assessment of the vaginal microbiota can allow us to obtain an accurate representation of the bacterial composition. It has been reported that species-specific characteristics of *Lactobacillus* spp. are critical for the stability and maintenance of human vaginal microbiota [19,20]. In particular, *Lactobacillus iners*-dominated community has been more associated with

321 vaginal dysbiosis, suggesting its relevance in the pathophysiology of gynecological
322 diseases.
323 Previous studies on short-read sequencing have reported that the sample collection
324 method had no effect on the results [21]. We collected swab and lavage samples from
325 the same cases and determined the species composition of vaginal communities. (Figure
326 2). When a single bacterial taxon comprised 98%–99% of the total bacteria (sample
327 numbers 1 and 4), there was no change in the composition ratio depending on the
328 sample collection method. When the vaginal microbiota was composed of more than
329 two bacteria (sample numbers 2 and 3), the bacterial species comprising the microbiota
330 were found to be the same, although there were slight differences in the proportions
331 present. Thus, we concluded that the effect of the sample collection method was
332 negligible. In contrast to the characterization of highly complex bacterial communities
333 in the human gut [12], the analysis of vaginal microbiota of lower complexity would
334 require a much smaller number of reads. A calculation using the median filtered reads
335 predicted that 3,000 reads would be accumulated in a sequence of approximately 8 min,
336 indicating that the time required for vaginal microbiota analysis using MinION™ could
337 be further reduced.
338 The conventional parallel-type short-read sequencer cannot yield reads covering the full
339 length of the 16S rRNA gene, and the partial 16S rRNA gene sequencing has been
340 insufficient to identify bacterial taxa at the species level. Benchmarking with the
341 conventional sequencing method (Illumina MiSeq™ technology) demonstrated that the
342 full-length 16S gene sequencing by MinION™ gives a better resolution for bacterial
343 identification [12]. To clarify the usefulness of long-read sequencing in vaginal
344 microbiota testing, we picked up *Lactobacillus iners*-dominant (> 99% in V1–9
345 analysis) samples (Group I: sample numbers 5, 9, and 10) and *Lactobacillus crispatus*-
346 dominant (>99% in V1–9 analysis) samples (Group C: sample numbers 6, 11, and 14)
347 and compared the results of 16S rRNA analyses of the V1–9 and V3–4 regions (Figure

348 5). The bioinformatic pipeline used a method where each amplicon sequence was
349 directly compared with the database, scored for similarity (Minimap2 score) according
350 to a proprietary algorithm, and then applied to the bacterial species with the highest
351 score. Hence, if there were multiple species with the same score, we treated the reads as
352 potentially existing in all of the possible species. The reads assigned to "Others" in this
353 study could be (1) bacteria with low abundance, or (2) bacteria that could not be
354 assigned to a single species due to competition between multiple species. In Group I
355 (Figure 5A), *Lactobacillus iners* was well recognized regardless of the choice of the
356 regions to be sequenced, but the numbers of reads assigned to "Others" increased in the
357 V3–4 analysis. In Group C (Figure 5B), V3–4 analysis showed that *Lactobacillus*
358 *crispatus* accounted for about 50% of the reads and *Lactobacillus acidophilus*,
359 *Lactobacillus helveticus*, *Lactobacillus amylovorus*, *Lactobacillus kefiranofaciens*,
360 *Lactobacillus hamsteri*, and *Lactobacillus gallinarum* were classified in similar
361 proportions in all three specimens, which had not been found in V1–9 analysis. Due to
362 the lower discriminatory power of the V3-4 region, the considerable number of V3-4
363 reads could not be assigned to one species and evenly allocated to multiple taxonomic
364 bins represented as potentially existing species. The numbers of reads assigned to
365 "Others" were greater in the V3-4 data set as compared to the V1-9 data set. This
366 suggested that assigning bacteria using the V1–9 region is more informative for
367 determining the appropriate bacterial species than using only the V3–4 region. Similar
368 to our previous report [12], the long-read sequencing with MinION™ showed high
369 identification accuracy compared with the short-read sequencing under the same
370 bioinformatic pipeline.
371 In terms of the usefulness of the MinION™ sequencer in gynecological practice, it is
372 interesting that we were able to show the detailed condition of vaginal dysbiosis in a
373 short period of time, which was not noted by the Nugent score. It has been suggested
374 that vaginal dysbiosis may have a connection with the incidence of sexually transmitted

infections [22] and genital viral infections [23,24] or cervical lesions [25]. The full-length 16S rRNA analysis using MinION™ could be a promising option for bacterial identification in a reasonable time frame for diagnostic purposes.

Having used MinION™ in our vaginal microbiota analysis workflow, we have seen four key benefits. 1) Convenience: The unit is extremely light, small, and portable, and sequencing can occur via a USB connection to an in-clinic PC. 2) Speed: The average time between specimen collection and result disclosure is 2 days. 3) Economy: Analysis of a small number of samples can be performed without increasing the unit cost of the test, by using individual barcodes. 4) Functionality: Although MinION™ has been reported to have a slightly higher error rate, the technical issues are being resolved and even the intestinal microbiota can now be identified [12]. In particular, the analysis of vaginal microbiota characterized by low diversity, as in the present study, allowed for comparable species-level identification [26]. Although this study focused on the vaginal microbiota, a similar workflow could be applied to many clinical areas in the future, and the benefits of MinION™ could make it easier for clinicians to successfully perform bacterial metagenome analysis. In addition, future large-scale microbiota studies could lead to new clinical findings.

Conclusions

In conclusion, our validation shows that the MinION™ long-read sequencer provides a low-cost, rapid workflow for identifying vaginal microbiota with higher resolution in a clinical setting. Detecting vaginal microbiota at the species level has the potential to identify risks to women's health, as well as facilitating large-scale clinical studies in any medical field.

Data accessibility

Sequence data from this article have been deposited in the DDBJ DRA database (www.ddbj.nig.ac.jp/dra/index-e.html) under accession numbers DRR244979–DRR245006.

Author contributions

Conceptualization, SK, YMa, YMo, HO, and KH; methodology, SK, YMa, SN, KK, and KH; software, SN and KK; validation, SK, YMa, and KH; investigation, SK and YMa; data curation, SK and YMa; writing—original draft preparation, SK; writing—review and editing, YMa, SN, YMo, KK, HO, and KH; visualization, SK and YMa; supervision, YMa and KH; funding acquisition, YMa, KK, and KH. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by Japan Society for the Promotion of Science KAKENHI Grant Numbers JP19K09339 (to YMa), JP17H07123 (to KK), JP20K06612 (to KK), and the Branding Program as a World-leading Research University on Intractable Immune and Allergic Diseases supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Acknowledgments

We are grateful to Tadashi Imanishi (Tokai University School of Medicine) for his helpful comments. We would like to thank Edanz (https://en-author-services.edanzgroup.com/ac) for editing the English text of a draft of this manuscript.

Conflicts of interest

The authors declare no conflict of interest.

428

429 **Supplementary information**

430 Supplementary Table S1: Representative bacterial genomes stored in the GenomeSync
431 database.

432 Supplementary Table S2: Details of identified species with V1-9 16S rRNA analysis
433 using MinION™ by minimap2 best hit score (>0.5%).

434 Supplementary Table S3: Datils of identified taxa with V1-9 16S rRNA analysis using
435 MinION by FASTQ 16S workflow.

436

437 **References**

438 1 Thomason JL, Gelbart SM, Scaglione NJ (1991) Bacterial vaginosis: current review
439 with indications for asymptomatic therapy. *Am J Obstet Gynecol* **165**: 1210-1217,
440 doi: 10.1016/s0002-9378(12)90729-2.

441 2 Nardis C, Mosca L, Mastromarino P (2013) Vaginal microbiota and viral sexually
442 transmitted diseases. *Ann Ig* **25**: 443-456, doi: 10.7416/ai.2013.1946.

443 3 Larsson PG, Bergstrom M, Forsum U, Jacobsson B, Strand A, Wolner-Hanssen P
444 (2005) Bacterial vaginosis. Transmission, role in genital tract infection and
445 pregnancy outcome: an enigma. *APMIS* **113**: 233-245, doi: 10.1111/j.1600-
446 0463.2005.apm_01.x.

447 4 Amsel R, Totten PA, Spiegel CA, Chen KC, Eschenbach D, Holmes KK (1983)
448 Nonspecific vaginitis. Diagnostic criteria and microbial and epidemiologic
449 associations. *Am J Med* **74**: 14-22, doi: 10.1016/0002-9343(83)91112-9.

450 5 Nugent RP, Krohn MA, Hillier SL (1991) Reliability of diagnosing bacterial
451 vaginosis is improved by a standardized method of gram stain interpretation. *J Clin*
452 *Microbiol* **29**: 297-301, doi: 10.1128/JCM.29.2.297-301.1991.

453 6 van den Munckhof EHA, van Sitter RL, Boers KE, Lamont RF, te Witt R, Cessie S,
454 Knetsch CW, van Doorn LJ, Quint WGV, Molijn A, van Hall MAL (2019)
455 Comparison of Amsel criteria, Nugent score, culture and two CE-IVD marked
456 quantitative real-time PCRs with microbiota analysis for the diagnosis of bacterial
457 vaginosis. *Eur J Clin Microbiol Infect Dis* **38**: 959-966, doi: 10.1007/s10096-019-
458 03538-7.

459 7 Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, Karlebach S,
460 Gorle R, Russell J, Tacket CO, Brotman RM, Davis CC, Ault K, Peralta L, Forney
461 LJ (2011) Vaginal microbiome of reproductive-age women. *PNAS* **108 Suppl 1**:
462 4680-4687, doi: 10.1073/pnas.1002611107.

463 8 Mikheyev AS, Tin MM (2014) A first look at the Oxford Nanopore MinION

464 sequencer. *Mol Ecol Resour* **14**: 1097-1102, doi: 10.1111/1755-0998.12324.
465 9 Mitsuhashi S, Kryukov K, Nakagawa S, Takeuchi JS, Shiraishi Y, Asano K,
466 Imanishi T (2017) A portable system for rapid bacterial composition analysis using a
467 nanopore-based sequencer and laptop computer. *Sci Rep* **7**: 5657, doi:
468 10.1038/s41598-017-05772-5.
469 10 Kai S, Matsuo Y, Nakagawa S, Kryukov K, Matsukawa S, Tanaka H, Iwai T,
470 Imanishi T, Hirota K (2019) Rapid bacterial identification by direct PCR
471 amplificatiaon of 16S rRNA genes using the MinION™ nanopore sequencer. *FEBS*
472 *Open Bio* **9**: 548-557, doi: 10.1002/2211-5463.12590.
473 11 Nakagawa S, Inoue S, Kryukov K, Yamagishi J, Ohno A, Hayashida K, Nakazwe R,
474 Kalumbi M, Mwenya D, Asami N, Sugimoto C, Mutengo MM, Imanishi T (2019)
475 Rapid sequencing-based diagnosis of infectious bacterial species from meningitis
476 patients in Zambia. *Clin Transl Immunology* **8**, e01087, doi: 10.1002/cti2.1087.
477 12 Matsuo Y, Komiya S, Yasumizu Y, Yasuoka Y, Mizushima K, Takagi T, Krykov K,
478 Fukuda A, Morimoto Y, Naito Y, Okada H, Bono H, Nakagawa S, Hirota K (2021)
479 Full-length 16S rRNA gene amplicon analysis of human gut microbiota using
480 MinION nanopore sequencing confers species-level resolution. *BMC Microbiol* **21**:
481 35, possidoi: 10.1186/s12866-021-02094-5.
482 13 Shen W, Le S, Li Y, Hu F (2016) SeqKit: A Cross-Platform and Ultrafast Toolkit for
483 FASTA/Q File Manipulation. *PLoS One* **11**: e0163962, doi:
484 10.1371/journal.pone.0163962.
485 14 Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO
486 (2013) The SILVA ribosomal RNA gene database project: improved data processing
487 and web-based tools. *Nucl Acids Res* **41**: D590-D596, doi: 10.1093/nar/gks1219.
488 15 Frith MC (2011) A new repeat-masking method enables specific detection of
489 homologous sequences. *Nucleic Acids Res* **39**: e23, doi: 10.1093/nar/gkq1212.
490 16 Li H (2018) Minimap2: pairwise alignment for nucleotide sequences.

491 *Bioinformatics* **34**: 3094-3100, doi: 10.1093/bioinformatics/bty191.
492 17 Ondov BD, Bergman NH, Phillippy AM (2011) Interactive metagenomic
493 visualization in a Web browser. *BMC Bioinformatics* **12**: 385, doi: 10.1186/1471-
494 2105-12-385.
495 18 Human Microbiome Project Consortium (2012) Structure, function and diversity of
496 the healthy human microbiome. *Nature* **486**: 207-214, doi: 10.1038/nature11234.
497 19 Zheng N, Guo R, Yao Y, Jin M, Cheng Y, Ling Z (2019) Lactobacillus iners Is
498 Associated with Vaginal Dysbiosis in Healthy Pregnant Women: A Preliminary
499 Study. *Biomed Res Int* **article ID 6070734**, doi: 10.1155/2019/6079734
500 20 Petrova MI, Reid G, Vaneechoutte M, Lebeer S (2017) Lactobacillus iners: Friend
501 or Foe? *Trends Microbiol* **25**: 182-191, doi: 10.1016/j.tim.2016.11.007
502 21 Virtanen S, Kalliala I, Nieminen P, Salonen A (2017) Comparative analysis of
503 vaginal microbiota sampling using 16S rRNA gene analysis. *PLoS One* **12**:
504 e0181477, doi: 10.1371/journal.pone.0181477.
505 22 Shipitsyna E, Khusnutdinova T, Budilovskaya O, Krysanova A, Shalepo K,
506 Savicheva A, Unemo M (2020) Bacterial vaginosis-associated vaginal microbiota is
507 an age-independent risk factor for Chlamydia trachomatis, Mycoplasma genitalium
508 and Trichomonas vaginalis infections in low-risk women, St. Petersburg, Russia.
509 *Eur J Clin Microbiol Infect Dis* **39**: 1221-1230, doi: 10.1007/s10096-020-03831-w.
510 23 Borgogna JC, Shardell MD, Santori EK, Nelson TM, Rath JM, Glover ED, Ravel J,
511 Gravitt PE, Yeoman CJ, Brotman RM (2020) The vaginal metabolome and
512 microbiota of cervical HPV-positive and HPV-negative women: a cross-sectional
513 analysis. *BJOG* **127**: 182-192, doi: 10.1111/1471-0528.15981.
514 24 Hoang T, Toler E, DeLong K, Mafunda NA, Bloom SM, Zierden HC, Moench TR,
515 Coleman JS, Hanes J, Kwon DS, Lai SK, Cone RA, Ensign LM (2020) The
516 cervicovaginal mucus barrier to HIV-1 is diminished in bacterial vaginosis. *PLoS*
517 *Pathog* **16**: e1008236, doi: 10.1371/journal.ppat.1008236.

518 25 Dahoud W, Michael CW, Gokozan H, Nakanishi AK, Harbhajanka A (2019)
519 Association of Bacterial Vaginosis and Human Papilloma Virus Infection with
520 Cervical Squamous Intraepithelial Lesions. *Am J Clin Pathol* **152**: 185-189, doi:
521 10.1093/ajcp/aqz021.
522 26 Kilianski A, Haas JL, Corriveau EJ, Liem AT, Willis KL, Kadavy DR, Rosenzweig
523 CN, Minot SS (2015) Bacterial and viral identification and differentiation by
524 amplicon sequencing on the MinION nanopore sequencer. *Gigascience* **4**: 12, doi:
525 10.1186/s13742-015-0051-z.
526

527 **Table 1.** Participants' background.

Number	Age (y.o.)	AMH (ng/mL)	LH (mIU/mL)	FSH (mIU/mL)	BMI (kg/m^2)
1	36	0.47	7.2	6.8	18.8
2	36	1.15	3.7	5.0	25.6
3	42	0.23	8.3	11.3	21.0
4	30	9.16	6.8	7.2	26.0
5	43	1.67	2.8	5.5	20.8
6	33	3.11	3.2	5.5	25.0
7	35	5.89	5.1	7.9	20.9
8	35	4.98	2.6	4.9	18.8
9	43	2.26	1.8	6.0	19.4
10	39	0.59	6.5	6.3	22.4
11	37	2.34	8.9	6.0	18.6
12	34	4.52	5.2	5.8	20.3
13	35	2.62	8.8	8.2	23.0
14	37	3.16	2.0	3.8	22.1
15	40	3.69	8.2	7.6	20.4
16	38	2.27	6.4	8.4	18.1
17	35	3.66	6.9	7.8	20.4
18	38	0.44	3.5	6.0	19.1

528 AMH: anti-Mullerian hormone, LH: luteinizing hormone, FSH: follicle-stimulating
529 hormone, BMI: body mass index.
530

531 **Table 2.** Statistics of MinION™ V1–9 sequencing data.

Sample	Pass reads				Filtered reads	
	No. of reads	Min (bp)	Avg (bp)	Max (bp)	No. of reads	Avg (bp)
1_lavage_V1-9	115,957	128	1,434.3	4,377	91,449 (78.9%)	1,633.5
1_swab_V1-9	63,547	155	1,078.0	4,193	31,462 (49.5%)	1,619.3
2_lavage_V1-9	33,721	158	1,507.2	4,486	28,718 (85.2%)	1,607.0
2_swab_V1-9	53,584	145	1,203.6	3,746	33,294 (62.1%)	1,613.6
3_lavage_V1-9	28,071	154	1,396.3	4,875	20,958 (74.7%)	1,613.8
3_swab_V1-9	23,053	162	1,521.0	3,990	20,073 (87.1%)	1,614.5
4_lavage_V1-9	30,603	155	1,502.9	4,977	25,756 (84.2%)	1,639.5
4_swab_V1-9	37,597	161	1,103.2	3,383	19,588 (52.1%)	1,644.9
5_lavage_V1-9	15,836	147	1,346.1	3,393	10,688 (67.5%)	1,632.8
6_lavage_V1-9	38,020	144	1,517.2	3,787	32,255 (84.8%)	1,642.4
7_lavage_V1-9	61,432	158	1,218.8	5,035	37,018 (60.3%)	1,631.4
8_lavage_V1-9	52,071	111	1,257.5	3,344	33,906 (65.1%)	1,603.2
9_lavage_V1-9	119,745	125	1,530.5	4,664	103,212 (86.2%)	1,649.4
10_lavage_V1-9	102,333	150	1,496.5	4,980	84,028 (82.1%)	1,650.2
11_lavage_V1-9	61,756	121	1,541.3	3,519	54,116 (87.6%)	1,647.8
12_lavage_V1-9	101,755	120	1,433.9	4,759	77,756 (76.4%)	1,635.8
13_lavage_V1-9	71,041	134	1,437.5	4,828	56,601 (79.7%)	1,599.2
14_lavage_V1-9	42,629	159	1,367.7	4,981	31,447 (73.8%)	1,629.0
15_lavage_V1-9	52,908	172	1,452.9	3,333	42,910 (81.1%)	1,630.4
16_lavage_V1-9	52,230	153	1,480.3	4,345	43,549 (83.4%)	1,633.0
17_lavage_V1-9	38,219	154	1,478.3	3,453	31,757 (83.1%)	1,625.3
18_lavage_V1-9	28,886	185	1,464.7	3,328	23,816 (82.4%)	1,631.8

532 Min: minimum read length, Avg: average read length, Max: maximum read length.

533

534 **Figure legends**

535 **Figure 1.** Workflow of 16S rRNA amplicon sequencing with the MinION™ platform
536 and bioinformatic analysis. After collecting the vaginal sample, it should be stored in a
537 freezer at –30°C until DNA extraction begins. Some swab collection kits should be
538 stored at room temperature, in accordance with the manufacturer's instructions.
539 Sequencing libraries are generated by a four-primer PCR-based strategy; in the initial
540 stages of PCR, the 16S rRNA gene is amplified with an inner primer pair. The PCR
541 product is amplified with the outer primers and targeted to introduce identical barcode
542 and tag sequences at both ends, allowing for the attachment of adapter molecules in a
543 one-step reaction. The library is then loaded into a MinION™ connected to a personal
544 computer. With our experimental method, sequencing runtime of 10 min is sufficient for
545 3,000 reads to be obtained. A final report of the microbiota can be presented within 2
546 days of initiating the analysis.

547

548 **Figure 2.** Taxonomic profiles comparing the different sampling method (lavage and
549 swab) results of V1–9 16S rRNA MinION™ sequencing with 3,000 randomly sampled
550 reads after filtration. In the analytical algorithm, we assigned the bacterial name that
551 showed the highest Minimap2 score for each read; bacteria with an assignment of less
552 than 1% were included in "Others."

553

554 **Figure 3.** PCoA of vaginal microbiota by different sampling techniques. The
555 coordinates of swab samples and lavage samples were nearly equivalent and did not
556 show any effect of the vaginal specimen collection method in the 3,000 filtered reads of
557 V1–9 MinION™ sequencing.

558

559 **Figure 4.** Taxonomic profiles comparing the results obtained for 3,000 and 10,000
560 filtered reads of V1–9 MinION™ sequencing. In the analytical algorithm, we assigned

561 the bacterial name that showed the highest Minimap2 score for each read; bacteria with
562 an assignment of less than 1% were included in "Others."
563
564 **Figure 5.** Taxonomic profiles comparing the results of V3–4 and V1–9 16S rRNA
565 sequences using MinION™ in 3,000 filtered reads. (A) In three cases, *Lactobacillus*
566 *iners* was shown to constitute more than 99% of the bacteria by MinION™ analysis
567 with the V1–9 region (Group I). (B) In three cases, *Lactobacillus crispatus* was shown
568 to constitute more than 99% of the bacteria by MinION™ analysis with the V1–9 region
569 (Group C).

Figure 1

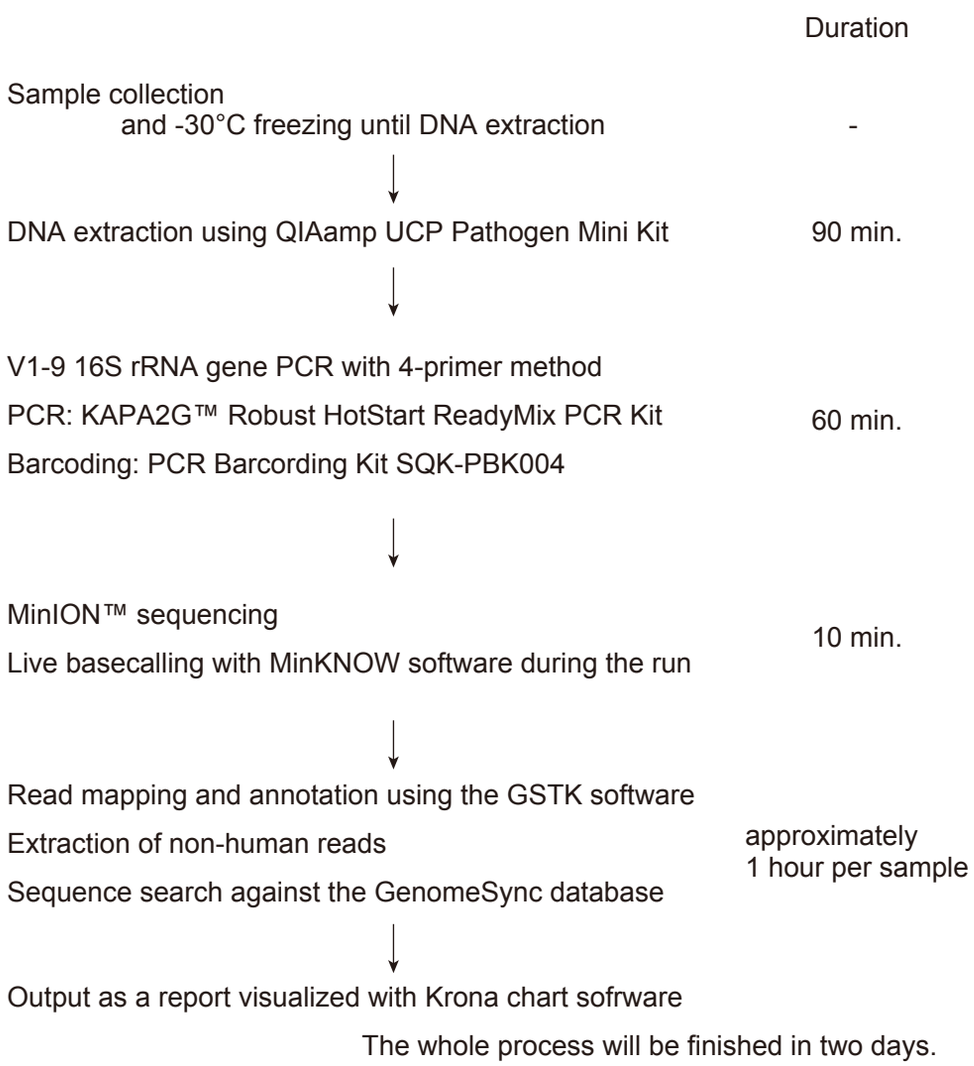

Duration
Sample collection
and -30°C freezing until DNA extraction
-
DNA extraction using QIAamp UCP Pathogen Mini Kit
90 min.
V1-9 16S rRNA gene PCR with 4-primer method
PCR: KAPA2G™ Robust HotStart ReadyMix PCR Kit
Barcoding: PCR Barcording Kit SQK-PBK004
60 min.
MinION™ sequencing
Live basecalling with MinKNOW software during the run
10 min.
Read mapping and annotation using the GSTK software
Extraction of non-human reads
Sequence search against the GenomeSync database
approximately
1 hour per sample
Output as a report visualized with Krona chart sofrware
The whole process will be finished in two days.

Figure 2

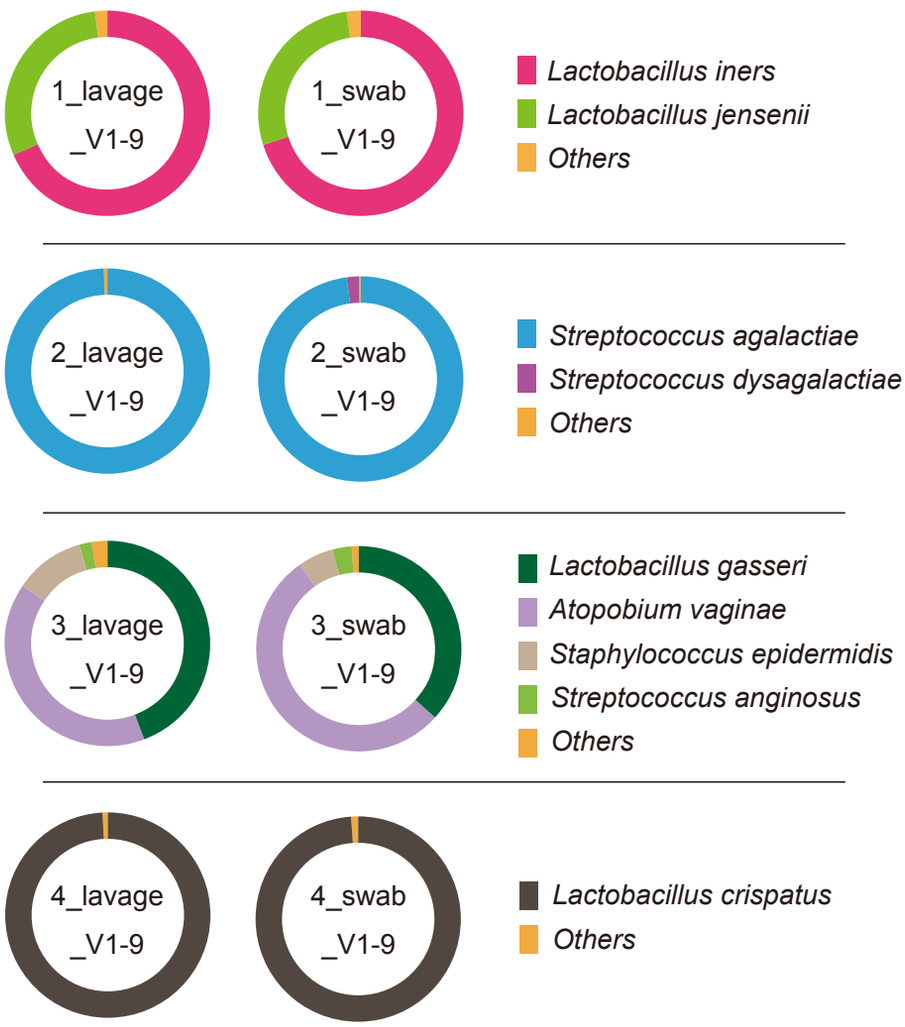

1_lavage
_V1-9
1_swab
_V1-9
Lactobacillus iners
Lactobacillus jensenii
Others
2_lavage
_V1-9
2_swab
_V1-9
Streptococcus agalactiae
Streptococcus dysagalactiae
Others
3_lavage
_V1-9
3_swab
_V1-9
Lactobacillus gasseri
Atopobium vaginae
Staphylococcus epidermidis
Streptococcus anginosus
Others
4_lavage
_V1-9
4_swab
_V1-9
Lactobacillus crispatus
Others

Figure 3

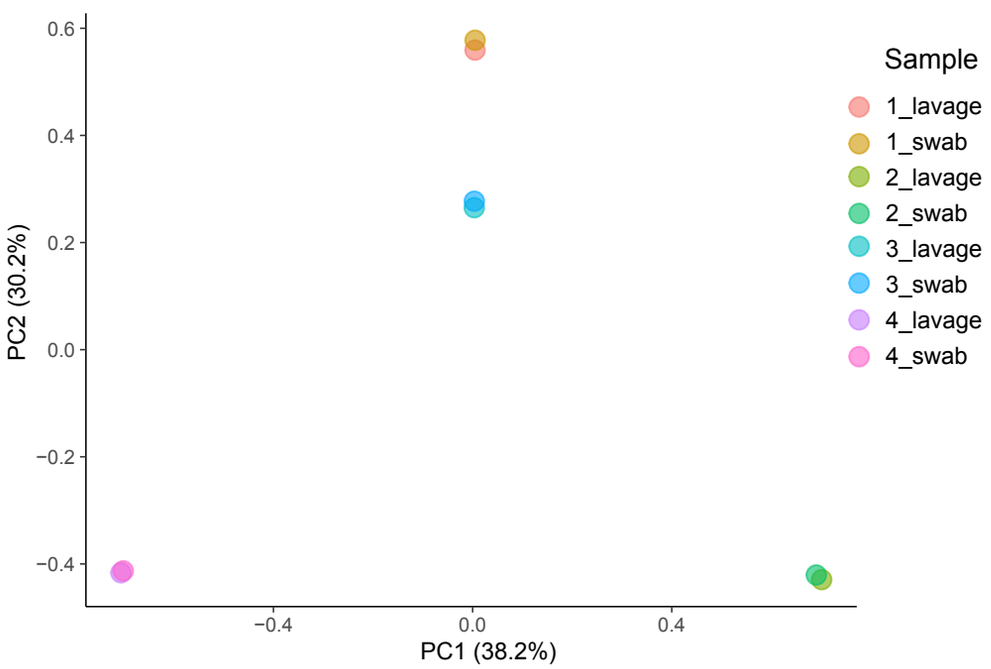

0.6
0.4
0.2
0.0
−0.2
−0.4
PC2 (30.2%)
−0.4
0.0
0.4
PC1 (38.2%)
Sample
1_lavage
1_swab
2_lavage
2_swab
3_lavage
3_swab
4_lavage
4_swab

Figure 4

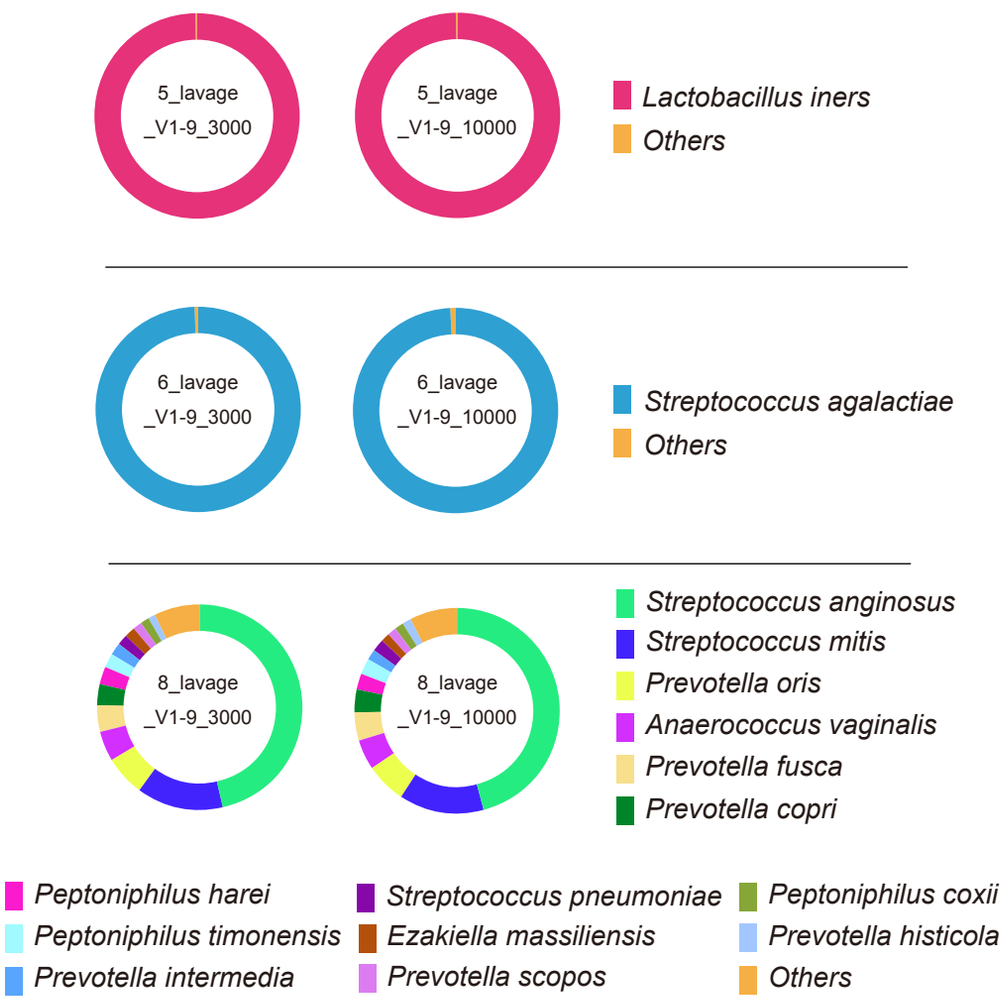
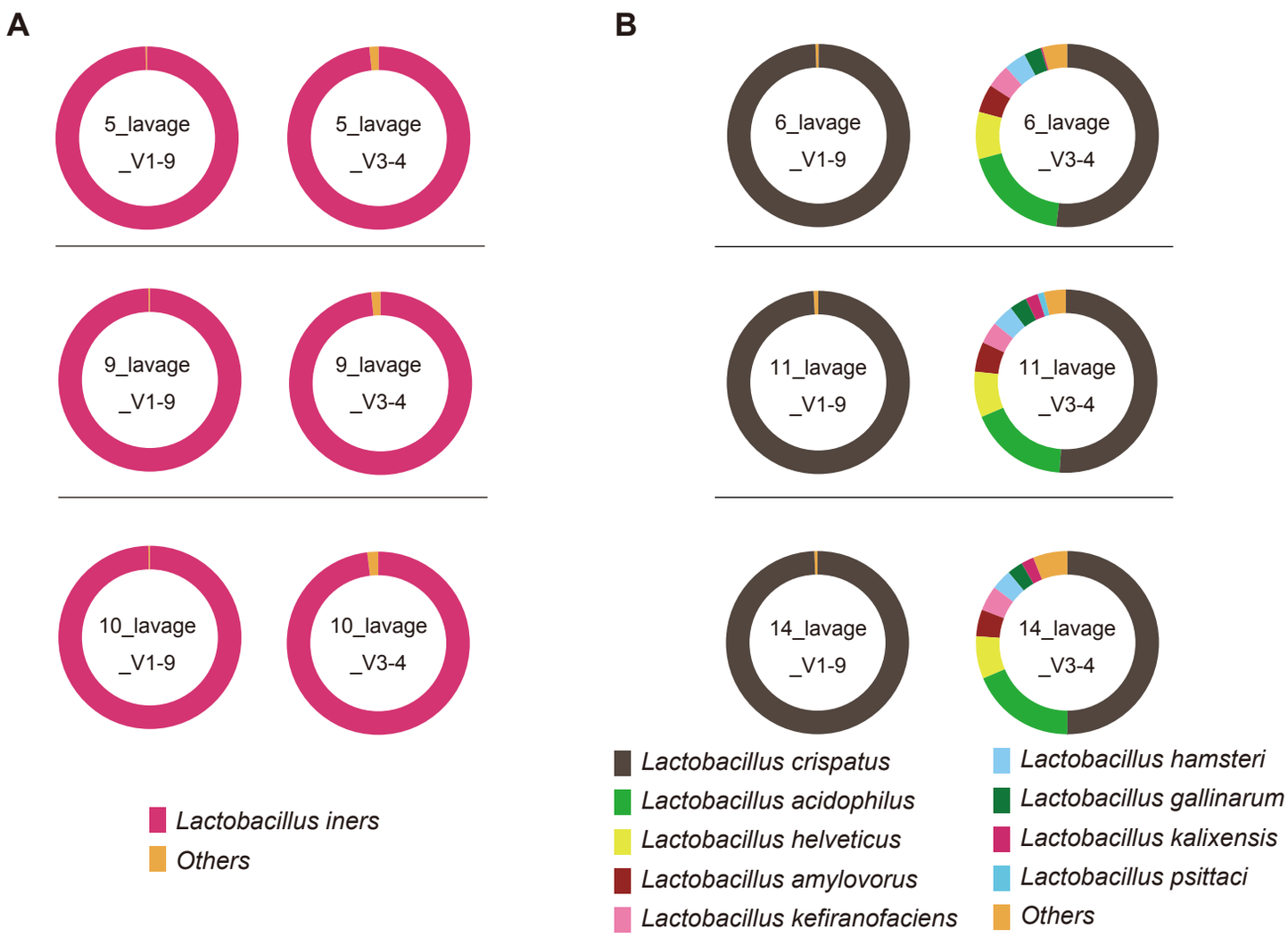

5_lavage
_V1-9_3000
5_lavage
_V1-9_10000
Lactobacillus iners
Others
6_lavage
_V1-9_3000
6_lavage
_V1-9_10000
Streptococcus agalactiae
Others
8_lavage
_V1-9_3000
8_lavage
_V1-9_10000
Streptococcus anginosus
Streptococcus mitis
Prevotella oris
Anaerococcus vaginalis
Prevotella fusca
Prevotella copri
Peptoniphilus harei
Streptococcus pneumoniae
Peptoniphilus coxii
Peptoniphilus timonensis
Ezakiella massiliensis
Prevotella histicola
Prevotella intermedia
Prevotella scopos
Others

Figure 5


A
5_lavage _V1-9
5_lavage _V3-4
9_lavage _V1-9
9_lavage _V3-4
10_lavage _V1-9
10_lavage _V3-4
Lactobacillus iners
Others
B
6_lavage _V1-9
6_lavage _V3-4
11_lavage _V1-9
11_lavage _V3-4
14_lavage _V1-9
14_lavage _V3-4
Lactobacillus crispatus
Lactobacillus acidophilus
Lactobacillus helveticus
Lactobacillus amylovorus
Lactobacillus kefiranofaciens
Lactobacillus hamsteri
Lactobacillus gallinarum
Lactobacillus kalixensis
Lactobacillus psittaci
Others