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

Running title: MinION™ enables rapid vaginal microbiota analysis

Shinnosuke Komiya[1,2], Yoshiyuki Matsuo[3,*], So Nakagawa[4], Yoshiharu Morimoto[2], Kirill Kryukov[5], Tadashi Imanishi[4], 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

ysmatsuo-kyt@umin.ac.jp

Department of Human Stress Response Science, Institute of Biomedical Science, Kansai Medical University, 2-5-1 Shin-machi, Hirakata, Osaka 573-1010, Japan

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.

Keywords

16S rRNA, bacterial vaginosis, next-generation sequencer, long-read sequencer, MinION™

Introduction

Bacterial vaginosis (BV) is one of the most common gynecological disorders among women of reproductive age. It is caused by disruption of the vaginal microbiota, and has relatively mild subjective symptoms, such as abnormal vaginal discharge (Thomason, et al., 1991). 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 (Nardis, et al., 2013), and preterm delivery (Larsson, et al., 2005). Currently, Amsel's criteria based on clinical outcomes (Amsel, et al., 1983) and the Nugent score using Gram stain findings (Nugent, et al., 1991) 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 (van den Munckhof, et al., 2019). Bacterial culture of vaginal discharge is also a standard method to investigate the vaginal microbiota, but some bacteria are difficult to culture (van den Munckhof, et al., 2019). 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.

16S ribosomal RNA (rRNA) gene sequencing analysis using next-generation sequencers can be used to identify bacteria without culture in a manner not dependent on subjective assessment by the examiner. Previous studies have shown that the vaginal microbiota of healthy women is classified into community state type (CST) I to V (Ravel, et al., 2011). CST is defined by the dominant bacteria as follows: CST I: *Lactobacillus crispatus*, CST II: *Lactobacillus gasseri*, CST Ⅲ: *Lactobacillus iners*, CST Ⅳ: lack of *Lactobacillus* spp., and CST V: *Lactobacillus jensenni*. These narrow definitions mean that undiagnosed cases of vaginal dysbiosis may be present. This new classification of the vaginal microbiota has been clarified by performing sequencing analysis on cases with no clinical symptoms. Because a deeper and more accurate

74 understanding of the vaginal microbiota may provide new clinical insights, we should
75 demonstrate that 16S rRNA analysis is quick and easy to perform in a broad range of
76 contexts in clinical practice.
77 MinION™ (Oxford Nanopore Technologies Ltd., Oxford, UK) detects the signal of a
78 DNA nucleotide that passes through a nanopore arranged on a flow cell (Mikheyev and
79 Tin, 2014). This long-read sequencing technology can be used to identify bacterial
80 microbiota with the whole region of the 16S rRNA. MinION™ is smaller, lighter, and
81 easier to deploy than conventional short-read sequencers, and it also offers the
82 advantage of high-resolution microbiota analysis with long-read sequencing
83 (Mitsuhashi, et al., 2017) (Kai, et al., 2019) (Nakagawa, et al., 2019). We constructed an
84 in-hospital vaginal microbiota-analyzing workflow using MinION™. This is the first
85 report to demonstrate that MinION™ can provide rapid patient feedback regarding
86 vaginal microbiota analysis.

Materials and Methods

Ethics statement

This study is a cross-sectional observational trial approved by the institutional review board of the medical corporation Sankeikai (Approval Number: 2019-38). After the approval, the study protocol was submitted to the UMIN-CTR clinical trial registry site (https://www.umin.ac.jp/ctr/index.htm) (trial registration number: UMIN000038032), which met the criteria of the International Committee of Medical Journal Editors (ICMJE).

Recruitment of participants

From April 2019 to May 2020, participants who met all of the following enrollment criteria were recruited: 1) primary infertility patients attending HORAC Grand Front Osaka Clinic in Osaka, Japan; 2) premenopausal Japanese women who were not pregnant; 3) no clinical symptoms strongly suggestive of BV; 4) scheduled hormone replacement embryo transfer cycle with frozen-thawed blastocyst; and 5) those who provided written consent. The exclusion criteria were as follows: 1) known history of HIV-1/2, hepatitis B/C, syphilis, or genital chlamydia infection; 2) known history of diabetes; 3) known history of cervical/vaginal surgery; 4) known history of intrauterine device use; and 5) known history of antibiotic, steroid, or vaginal suppository use within 2 weeks. In addition to collecting vaginal samples, details of age, body mass index, and basal hormone levels (anti-Mullerian hormone, luteinizing hormone, and follicle-stimulating hormone on days 2–4 of the menstrual cycle), medical history, and infertility treatment history were collected from the medical records. All vaginal samples were collected on the day of embryo transfer.

Nugent score

Vaginal smears were analyzed by pathologists affiliated with an external laboratory, and

the Nugent score was calculated based on microscopic findings: the numbers of *Lactobacillus* (scored as 0 to 4), *Gardnerella* (scored as 0 to 4), and *Mobiluncus* (scored as 0 to 2). A total score of 0 to 3 was diagnosed as representative of healthy vaginal microbiota, 4 to 6 as an intermediate group, and 7 or more as BV (Nugent, et al., 1991). Inappropriate samples with defective specimen collection were diagnosed as indeterminate. If BV was diagnosed based on the Nugent score, our policy was to treat it using antibiotics.

Vaginal sample collection method

Vaginal lavage samples were collected from 18 participants; from four participants, additional swab samples were collected with written consent. When both lavage and swab samples were collected, the swab samples were collected first. For the lavage method, the inside of the vagina was washed with 10 ml of sterile saline, and at least 2 ml was collected using a sterile syringe. Collected lavage samples were stored at –30 °C until DNA extraction. For the swab method, vaginal mucus from the uterovaginal region was collected using a packaged OMNIgene vaginal kit (OMR-130; DNA Genotek Inc., Ottawa, Canada). Swab samples were stored at room temperature until DNA extraction, in accordance with the manufacturer's instructions.

Sequencing sample preparation

DNA was extracted from 22 human vaginal samples using the QIAamp UCP Pathogen Mini Kit (QIAGEN, Venlo, Netherlands) by the bead beating and spin protocol, in accordance with the manufacturer's instructions. After extraction, the DNA content was measured using a NanoDrop® 1000 Spectrophotometer (Thermo Fisher Scientific, MA, USA), and the concentrations of extracted DNA from the swab and lavage samples were 6.1–32.3 ng/μl and 53.1–1092.4 ng/μl, respectively.

MinION™ sequencing

As described in our previous report (Matsuo, et al., 2020), a slightly modified version of four-primer polymerase chain reaction (PCR) with rapid adapter attachment was performed. For amplification of the V1-9 region of the 16S rRNA gene, a forward primer (S-D-Bact-0008-c-S-20) with the anchor sequence 5′-TTTCTGTTGGTGCTGATATTGCAGRGTTYGATYMTGGCTCAG-3′ and a reverse primer with the anchor sequence 5′-ACTTGCCTGTCGCTCTATCTTCCGGYTACCTTGTTACGACTT-3′ were used as inner primers. For amplification of the V3–4 region, 341F with the anchor sequence 5′-TTTCTGTTGGTGCTGATATTGCCCTACGGGNGGCWGCAG-3′ and 806R with the anchor sequence 5′-ACTTGCCTGTCGCTCTATCTTCGGACTACHVGGGTWTCTAAT-3′ were used as inner primers. PCR amplification of the 16S rRNA gene was performed using the KAPA2G™ Robust HotStart ReadyMix PCR Kit (Kapa Biosystems, MA, USA) with an inner primer pair (50 nM for each) and the outer primer mixture (10 nM) of PCR Barcoding Kit (SQK-PBK004; Oxford Nanopore Technologies, Oxford, UK) in a total volume of 25 µl. Each sample was assigned an individual barcode, which was used to identify the PCR product. Amplification was performed under the following PCR conditions: initial denaturation at 95 °C for 3 min; five cycles of 15 s at 95 °C, 15 s at 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 72 °C; followed by a final extension at 72 °C for 1 min. The amplification products were confirmed using 1% agarose gel electrophoresis (1 × TAE buffer) to be approximately 1600 base pairs (V1–9 region) and 400 base pairs (V3–4 region), which are the expected amplification sizes. After confirmation, the amplified DNA was purified using AMPure® XP (Beckman Coulter, CA, USA). A total of 100 ng (10 µl) of purified DNA was incubated with 1 µl of Rapid Adapter (Oxford Nanopore Technologies Ltd.) 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). Based on the manufacturer's instructions, 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 using MINKNOW software ver. 1.11.5 (Oxford Nanopore Technologies Ltd.), in accordance with the manufacturer's instructions.

Bioinformatic Analysis Workflow

We performed the bioinformatic analysis using a pipeline of multiple programs built in accordance with our previous reports (Mitsuhashi, et al., 2017) (Nakagawa, et al., 2019). 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 (Shen, et al., 2016): 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/).

4) TANTAN program ver. 18 (Frith, 2011): removing simple repeat sequences.

5) Minimap2 program ver. 2.14 (Li, 2018): eliminating human genome information (Human Genome Assembly GRCh38) and matching each genome read to the 5,850 representative bacterial genome sequences (Table S1) stored in the GenomeSync

195 database (http://genomesync.org).

196 6) In-house Perl scripts: selecting the species with the highest Minimap2 score and

197 determining the taxa based on the NCBI taxonomy database.

198 7) Krona software version 2.7 (Ondov, et al., 2011): visualizing the frequency of

199 detected bacterial species in a given sample.

200

201 **Statistical analysis**

202 Principal coordinate analysis was performed and the results were visualized with R

203 version 4.0.2.

204

205

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 I).

Identification of vaginal bacteria

The composition of the vaginal microbiota was investigated by 16S rRNA gene amplicon sequencing on the MinION™ platform (Fig. 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 II). The results of microbiota analysis of the filtered 3,000 reads using MinION™ are shown in Supplementary Table S2.

Comparison of the results of swab and lavage samples

For subjects 1–4, lavage and swab samples were collected and bacterial DNA extraction and 16S rRNA analysis were performed to determine the effect of the sample collection method on the results of the microbiota analysis. After filtering, 3,000 reads were randomly extracted in eight samples (1-4_lavage and 1-4_swab), the results of which are shown in Fig. 2, with almost the same identification with the two sampling methods. We performed principal coordinate analysis (PCoA) to assess the equivalence of the lavage and swab sampling methods, the results of which are shown in Fig. 3. Regardless of the sample collection method, the results of the analysis of the same subjects involved almost the same coordinates.

Comparison of the results of sequence read number

To assess the effect of read count on the results of the MinION™ sequencing, three

samples that showed characteristic bacterial microbiota [5_lavage: *Lactobacillus iners* dominant (>99%), 6_lavage: *Lactobacillus crispatus* dominant (>99%), 8_lavage: BV like] were selected and their taxonomic profiles were compared after filtering for 3,000 reads (random sampling) and 10,000 reads (random sampling) (Fig. 4). In all specimens, the results of vaginal microbiota analysis in MinION™ were similar for different numbers of reads.

Comparison of the results of V1–9 and V3–4 16S rRNA gene sequencing

For the taxonomic classification of bacteria, we compared the resolution of long-read (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 (Fig. 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 (Fig. 5a). In Group C, the V3-4 sequence alignment resulted in ambiguous identification of *Lactobacillus* species (Fig. 5b). All samples showed *Lactobacillus crispatus*, *Lactobacillus acidophilus*, *Lactobacillus helveticus*, *Lactobacillus amylovorus*, *Lactobacillus kefiranofaciens*, *Lactobacillus hamsteri*, *Lactobacillus gallinarum*, and *Lactobacillus kalixensis* 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 that in other parts of the body (Human Microbiome Project, 2012). Conventional methods, such as Amsel's criteria, the Nugent score, and bacterial culture, have enabled only rough assessment of the vaginal microbiota. Next-generation sequencing (NGS) 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 (van den Munckhof, et al., 2019). 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 vaginal microbiota analysis, it is important to be able to profile *Lactobacillus* spp. and the causative organisms of BV at the species level. Interestingly, no cases were diagnosed with BV by the Nugent score in this study, but the results of vaginal microbiota analysis using MinION™ to determine the profile by Minimap2 best-hit score identified not only healthy vaginal microbiota but also cases with BV-like microbiota (Supplementary Table S2).

In this study, 90-min MinION™ sequencing yielded 15,836 to 119,745 reads (Table II). To estimate the shortest sequencing time for accurately assessing constituents of the vaginal microbiota, the results of the analyses of 3,000, 10,000, and all reads were compared in the typical cases of healthy vaginal microbiota and BV. The results were

not affected by increasing the number of reads to 3,000 or more (Fig. 4), so we concluded that 3,000 reads were sufficient for vaginal microbiota analysis. A calculation using the median filtered reads predicted that 3,000 reads would be accumulated in a sequence of approximately 8 min, indicating that the time required for vaginal microbiota analysis using MinION™ could be further reduced.

Previous studies on short-read sequencing have reported that the sample collection method had no effect on the results (Virtanen, et al., 2017). Therefore, we collected swab and lavage samples from the same cases to determine whether the effect of the sample collection method was negligible in the case of long-read sequencing with our bioinformatic pipeline in 3,000 randomly sampled reads (Fig. 2). When a single bacterial taxon comprised 98%–99% of the total bacteria (sample numbers 1 and 4), there was no change in the composition ratio depending on the sample collection method. When the vaginal microbiota was composed of more than two bacteria (sample numbers 2 and 3), the bacterial species comprising the microbiota were found to be the same, although there were slight differences in the proportions present. PCoA was performed to visualize the influence of the sample collection method (Fig. 3). Regardless of the sampling method, each sample was located in approximately the same coordinates and we thus concluded that the sample collection method had no effect in MinION™ sequencing. As this is the first report of vaginal microbiota analysis using MinION™, we mainly selected lavage samples to ensure the recovery of more organisms, but we were able to show that swab samples were sufficient.

To clarify the usefulness of long-read sequencing, one of the properties of MinION™, we picked up *Lactobacillus iners*-dominant (> 99% in V1–9 analysis) samples (Group I: sample numbers 5, 9, and 10) and *Lactobacillus crispatus*-dominant (>99% in V1–9 analysis) samples (Group C: sample numbers 6, 11, and 14) and compared the results of 16S rRNA analyses of the V1–9 and V3–4 regions (Fig. 5). The bioinformatic pipeline used a method where each amplicon sequence was directly compared with the database,

scored for similarity (Minimap2 score) according to a proprietary algorithm, and then applied to the bacterial species with the highest score. Hence, if there were multiple species with the same score, we treated the reads as potentially existing in all of the possible species. The reads assigned to "Others" in this study could be (1) bacteria with low abundance, or (2) bacteria that could not be assigned to a single species due to competition between multiple species. In Group I (Fig. 5a), *Lactobacillus iners* was well recognized in the V3–4 16S rRNA analysis, but the numbers of reads assigned to "Others" increased in the V3–4 analysis with 41.9, 44.8, and 57.4 reads compared with 6.0, 5.0, and 6.0 reads in the V1–9 analysis, respectively. In Group C (Fig. 5b), V3–4 analysis showed that *Lactobacillus crispatus* accounted for about 50% of the reads and *Lactobacillus acidophilus*, *Lactobacillus helveticus*, *Lactobacillus amylovorus*, *Lactobacillus kefiranofaciens*, *Lactobacillus hamsteri*, and *Lactobacillus gallinarum* were classified in similar proportions in all three specimens, which had not been found in V1–9 analysis. The numbers of reads assigned to "Others" were 15.5, 25.5, and 36.0 reads in groups V1–9 and 130.2, 110.5, and 176.6 reads in groups V3–4, respectively. This suggested that assigning bacteria using the V1–9 region is more informative for determining the appropriate bacterial species than using only the V3–4 region. Similar to our previous report (Matsuo, et al., 2020), the long-read sequencing with MinION™ showed high identification accuracy compared with the short-read sequencing under the same bioinformatic pipeline.

In terms of the usefulness of the MinION™ sequencer in gynecological practice, it is interesting that we were able to show the detailed condition of vaginal dysbiosis in a short period of time, which was not noted by the Nugent score. It has been suggested that vaginal dysbiosis may have a connection with the incidence of sexually transmitted infections (Shipitsyna, et al., 2020) and genital viral infections (Borgogna, et al., 2020) (Hoang, et al., 2020) or cervical lesions (Dahoud, et al., 2019). It has also been shown that metagenomic analysis using NGS applications to assess vaginal microbiota can

provide more accurate results than conventional methods such as the Nugent score, Amsel's criteria, and bacterial cultures (van den Munckhof, et al., 2019). Despite the potential value of vaginal microbiota analysis in protecting women's health, the challenges of all currently used sequencing equipment, including their price and running costs, and the 2–3 weeks required to obtain results in outsourced testing (in Japan), have prevented widespread use of vaginal microbiota analysis.

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 (Matsuo, et al., 2020). In particular, the analysis of vaginal microbiota characterized by low diversity, as in the present study, allowed for comparable species-level identification (Kilianski, et al., 2015). 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.

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.

Authors' roles

S.K., Y.Ma., Y.Mo., H.O. and K.H. designed and supervised the study. S.K., Y.Mo. and H.O. contributed to sample collection. S.N., K.K. and T.I. built the bioinformatics pipeline. S.K., Y.Ma. and K.H. contributed to the experiments. S.K. analyzed the data. S.K. wrote the manuscript. Y.Ma., S.N., K.K., H.O. and K.H. contributed to editing the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We would like to thank Edanz (https://en-author-services.edanzgroup.com/ac) for editing the English text of a draft of this manuscript.

Funding

This work was supported by Japan Society for the Promotion of Science KAKENHI Grant Numbers JP19K09339 (to Y.Ma.), JP17H07123 (to K.K.), JP20K06612 (to K.K.), 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.

Conflict of interest

No authors have any conflicts of interest to declare.

Abbreviations

BV: Bacterial vaginosis, CST: community state type, NGS: Next-generation sequencing, PCoA: principal coordinate analysis, PCR: polymerase chain reaction, rRNA: ribosomal RNA

Data availability

392 Sequence data from this article have been deposited in the DDBJ DRA database
393 (www.ddbj.nig.ac.jp/dra/index-e.html) under accession numbers DRR244979–
394 DRR245006.
395
396 **Additional files**
397 **Supplementary Table S1.** Representative bacterial genomes stored in the GenomeSync
398 database.
399 **Supplementary Table S2.** Details of identified species by V1–9 16S rRNA analysis
400 using MinION™ by Minimap2 best-hit score (>1%).
401

References

Amsel R, Totten PA, Spiegel CA, Chen KC, Eschenbach D, Holmes KK. Nonspecific vaginitis. Diagnostic criteria and microbial and epidemiologic associations. *Am J Med* 1983; **74**: 14-22.

Borgogna JC, Shardell MD, Santori EK, Nelson TM, Rath JM, Glover ED, Ravel J, Gravitt PE, Yeoman CJ, Brotman RM. The vaginal metabolome and microbiota of cervical HPV-positive and HPV-negative women: a cross-sectional analysis. *BJOG* 2020; **127**: 182-192.

Dahoud W, Michael CW, Gokozan H, Nakanishi AK, Harbhajanka A. Association of Bacterial Vaginosis and Human Papilloma Virus Infection With Cervical Squamous Intraepithelial Lesions. *Am J Clin Pathol* 2019; **152**: 185-189.

Frith MC. A new repeat-masking method enables specific detection of homologous sequences. *Nucleic Acids Res* 2011; **39**: e23.

Hoang T, Toler E, DeLong K, Mafunda NA, Bloom SM, Zierden HC, Moench TR, Coleman JS, Hanes J, Kwon DS *et al.* The cervicovaginal mucus barrier to HIV-1 is diminished in bacterial vaginosis. *PLoS Pathog* 2020; **16**: e1008236.

Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. *Nature* 2012; **486**: 207-214.

Kai S, Matsuo Y, Nakagawa S, Kryukov K, Matsukawa S, Tanaka H, Iwai T, Imanishi T, Hirota K. Rapid bacterial identification by direct PCR amplificatiaon of 16S rRNA genes using the MinION™ nanopore sequencer. *FEBS Open Bio* 2019; **9**: 548-557.

Kilianski A, Haas JL, Corriveau EJ, Liem AT, Willis KL, Kadavy DR, Rosenzweig CN, Minot SS. Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer. *Gigascience* 2015; **4**: 12.

Larsson PG, Bergstrom M, Forsum U, Jacobsson B, Strand A, Wolner-Hanssen P. Bacterial vaginosis. Transmission, role in genital tract infection and pregnancy outcome: an enigma. *APMIS* 2005; **113**: 233-245.

Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018; **34**: 3094-3100.

Matsuo Y, Komiya S, Yasumizu Y, Yasuoka Y, Mizushima K, Takagi T, Krykov K, Fukuda A, Morimoto Y, Naito Y, Okada H, Bono H, Nakagawa S, Hirota K. Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION nanopore sequencing confers species-level resolution. *bioRxiv* 2020 https://doi.org/10.1101/2020.05.06.078147

Mikheyev AS, Tin MM. A first look at the Oxford Nanopore MinION sequencer. *Mol Ecol Resour* 2014; **14**: 1097-1102.

Mitsuhashi S, Kryukov K, Nakagawa S, Takeuchi JS, Shiraishi Y, Asano K, Imanishi T. A portable system for rapid bacterial composition analysis using a nanopore-based sequencer and laptop computer. *Sci Rep* 2017; **7**: 5657.

Nakagawa S, Inoue S, Kryukov K, Yamagishi J, Ohno A, Hayashida K, Nakazwe R, Kalumbi M, Mwenya D, Asami N *et al.* Rapid sequencing-based diagnosis of infectious bacterial species from meningitis patients in Zambia. *Clin Transl Immunology* 2019; **8**: e01087.

Nardis C, Mosca L, Mastromarino P. Vaginal microbiota and viral sexually transmitted diseases. *Ann Ig* 2013; **25**: 443-456.

Nugent RP, Krohn MA, Hillier SL. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. *J Clin Microbiol* 1991; **29**: 297-301.

Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* 2011; **12**: 385.

Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO *et al.* Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A* 2011; **108 Suppl 1**: 4680-4687.

Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q

456 File Manipulation. *PLoS One* 2016; **11**: e0163962.
457 Shipitsyna E, Khusnutdinova T, Budilovskaya O, Krysanova A, Shalepo K, Savicheva
458 A, Unemo M. Bacterial vaginosis-associated vaginal microbiota is an age-independent
459 risk factor for Chlamydia trachomatis, Mycoplasma genitalium and Trichomonas
460 vaginalis infections in low-risk women, St. Petersburg, Russia. *Eur J Clin Microbiol*
461 *Infect Dis* 2020; **39**: 1221-1230.
462 Thomason JL, Gelbart SM, Scaglione NJ. Bacterial vaginosis: current review with
463 indications for asymptomatic therapy. *Am J Obstet Gynecol* 1991; **165**: 1210-1217.
464 van den Munckhof EHA, van Sitter RL, Boers KE, Lamont RF, Te Witt R, le Cessie S,
465 Knetsch CW, van Doorn LJ, Quint WGV, Molijn A *et al.* Comparison of Amsel criteria,
466 Nugent score, culture and two CE-IVD marked quantitative real-time PCRs with
467 microbiota analysis for the diagnosis of bacterial vaginosis. *Eur J Clin Microbiol Infect*
468 *Dis* 2019; **38**: 959-966.
469 Virtanen S, Kalliala I, Nieminen P, Salonen A. Comparative analysis of vaginal
470 microbiota sampling using 16S rRNA gene analysis. *PLoS One* 2017; **12**: e0181477.
471

Tables

Table I Participants' background

Sample	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

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

475 **Table II** 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

476 Min: minimum read length, Avg: average read length, Max: maximum read length

477

Figure legends

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

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

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

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

505 bacterial name that showed the highest Minimap2 score for each read; bacteria with an
506 assignment of less than 1% were included in "Others."
507
508 **Fig. 5** Taxonomic profiles comparing the results of V3–4 and V1–9 16S rRNA sequences
509 using MinION™ in 3,000 filtered reads. **(a)** In three cases, *Lactobacillus iners* was shown
510 to constitute more than 99% of the bacteria by MinION™ analysis with the V1–9 region
511 (Group I). **(b)** In three cases, *Lactobacillus crispatus* was shown to constitute more than
512 99% of the bacteria by MinION™ analysis with the V1–9 region (Group C).
513
514

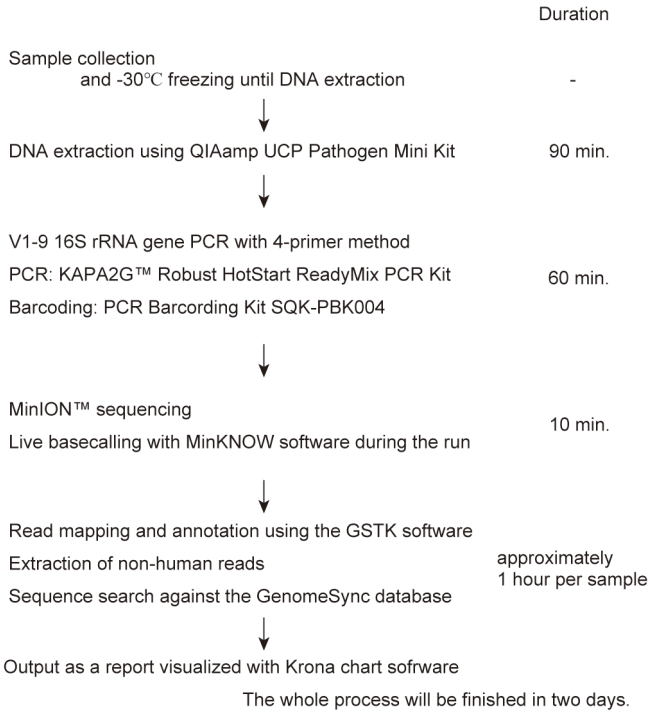

Figure 1

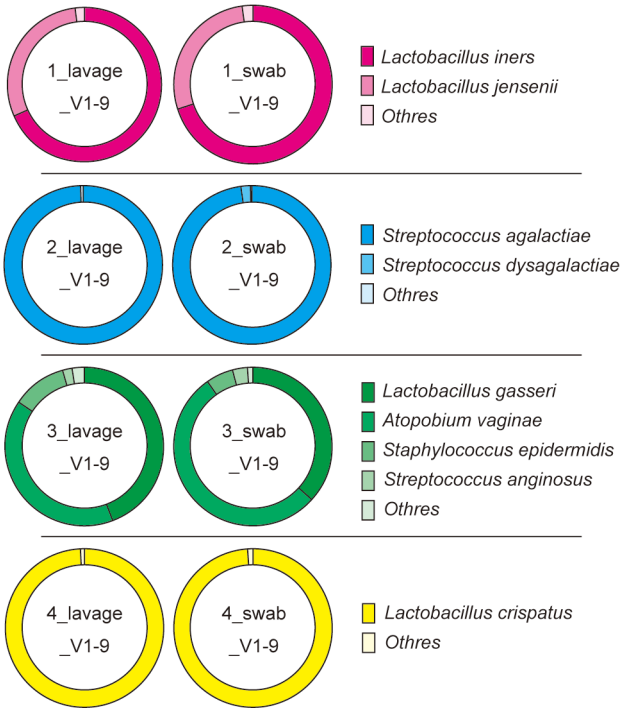

Figure 2

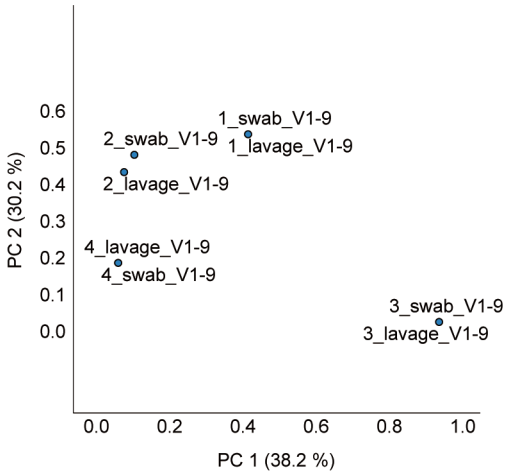

Figure 3

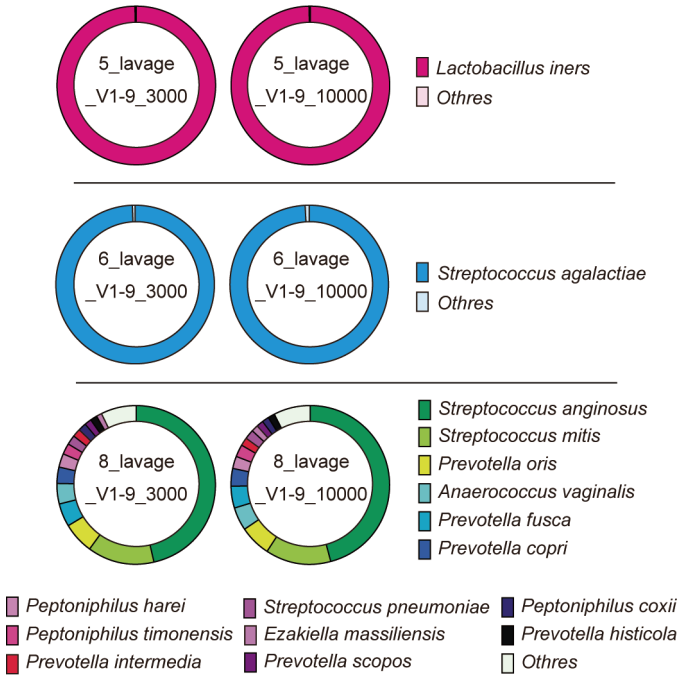

Figure 4

Fig. 5a

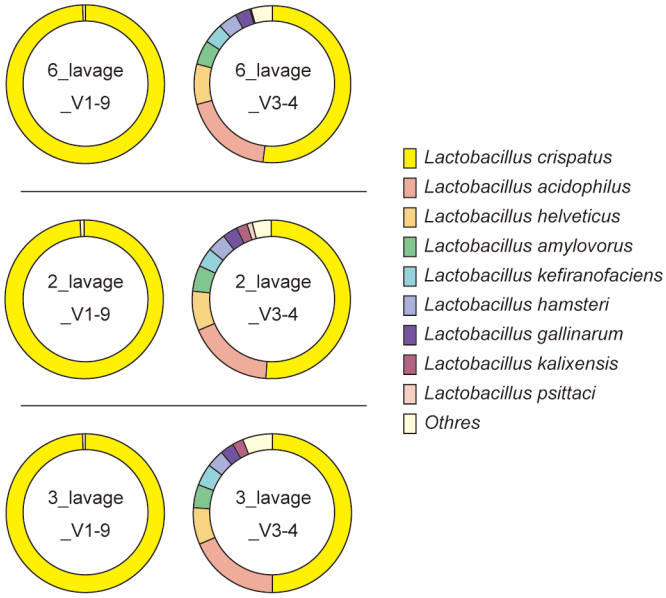

Fig. 5b