Evaluation of DNA extraction protocols from liquid-based cytology specimens for studying cervical microbiota

Takeo Shibata,[a,b] Mayumi Nakagawa,[a] Hannah N. Coleman,[a] Sarah M. Owens,[c] William W. Greenfield,[d] Toshiyuki Sasagawa,[b] Michael S. Robeson II[e]

[a]Department of Pathology, University of Arkansas for Medical Sciences, Little Rock, AR, USA

[b]Department of Obstetrics and Gynecology, Kanazawa Medical University, Uchinada, Ishikawa, Japan

[c]Biosciences Division, Argonne National Laboratory, Lemont, IL, USA

[d]Department of Obstetrics and Gynecology, University of Arkansas for Medical Sciences, Little Rock, AR, USA

[e]Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR, USA

Corresponding author: Michael S. Robeson II

Tel: 501-526-4242, Fax: 501-526-5964

Email: MRobeson@uams.edu

Abstract

Cervical microbiota (CM) are considered an important factor affecting the progression of cervical intraepithelial neoplasia (CIN) and are implicated in the persistence of human papillomavirus (HPV). Collection of liquid-based cytology (LBC) samples is routine for cervical cancer screening and HPV genotyping, and can be used for long-term cytological biobanking. Herein, we investigate the feasibility of leveraging LBC specimens for use in CM surveys by amplicon sequencing. As methodological differences in DNA extraction protocols can potentially bias the composition of microbiota, we set out to determine the performance of four commonly used DNA extraction kits (ZymoBIOMICS DNA Miniprep Kit; QIAamp PowerFecal Pro DNA Kit; QIAamp DNA Mini Kit; and IndiSpin Pathogen Kit) and their ability to capture the diversity of CM from LBC specimens. LBC specimens from 20 patients (stored for 716 ± 105 days) with cervical intraepithelial neoplasia (CIN) 2/3 or suspected CIN2/3 were each aliquoted for extraction by each of the four kits. We observed that, regardless of the extraction protocol used, all kits provided equivalent accessibility to the cervical microbiome, with some minor differences. For example, the ZymoBIOMICS kit appeared to differentially increase access of several more microbiota compared to the other kits. Potential kit contaminants were observed as well. Approximately 80% microbial genera were shared among all DNA extraction protocols. The variance of microbial composition per individual was larger than that of the DNA extraction protocol used. We also observed that HPV16 infection was significantly associated with community types that were not dominated by *Lactobacillus iners*.

Importance

Collection of LBC specimens is routine for cervical cancer screening and HPV genotyping, and can be used for long-term cytological biobanking. We demonstrated that LBC samples, which had been under prolonged storage prior to DNA extraction, were able to provide a robust assessment of the CM and its relationship to HPV status, regardless of the extraction kit used. Being able to retroactively access the CM from biobanked LBC samples, will allow researchers to better interrogate historical interactions between the CM and its relationship to CIN and HPV. This alone has the potential to bring CM research one-step closer to the clinical practice.

Keywords; cervical microbiota, DNA extraction, HPV, CIN, liquid-based cytology

Introduction

High-throughput sequencing (HTS) technology of 16S rRNA gene amplicon sequences has made it possible to better understand the relationships between cervicovaginal microbiota and human papillomavirus (HPV) infection (1) (2) (3) (4) (5) and HPV-related diseases (6) (7) (8) (9) (10). Cervicovaginal microbiota are considered to be an important factor affecting the progress of cervical intraepithelial neoplasia (CIN) (6) (7) (8) (9) and are implicated in the persistence of high-risk HPV (HR-HPV) (1) (2) and low-risk HPV (LR-HPV) (3). For example, the phyla *Actinobacteria* and *Fusobacteria* were enriched in HR-HPV positive environment (4) while the phyla *Actinobacteria*, *Proteobacteria*, and *Fusobacteria* in low-risk HPV (LR-HPV) (3). Additionally, *Lactobacillus iners*-dominant samples are associated with both HR-HPV and LR-HPV (5). Moreover, it has been shown that CIN risk was increased when the cervical microbes *Atopobium vaginae*, *Gardnerella vaginalis*, and *Lactobacillus iners* were present with HR-HPV (10). The cervicovaginal microbiome specified by *Lactobacillus*-dominant type or non-*Lactobacillus*-dominant type has been shown to interact with the immune system (7) (11). Inflammatory cytokines, such as Interleukin (IL)-1α and IL-18, were increased in non-*Lactobacillus*-dominant community types of reproductive-aged healthy women (11). In the analysis of patients with cervical cancer, non-*Lactobacillus*-dominant community types were positively associated with chemokines such as interferon gamma-induced protein 10 (IP-10) and soluble CD40-ligand activating dendric cells (DCs) (7). The metabolism of the cervicovaginal microbiome may be a substantial contributing factor to maternal health during pregnancy, although the mechanism is still unclear (12).

Little has been reported on the utility of liquid-based cytology (LBC) samples for use in cervical microbiome studies. Conventionally, microbiome sample collection methods entail the

use of swabs (13) or self-collection of vaginal discharge (14). To obtain a non-biased and broad range of cervical microbiota, DNA extraction should be optimized for a range of difficult-to-lyse-bacteria, *e.g*. *Firmicutes*, *Actinobacteria*, and *Lactobacillus* (13) (15) (16) (17) (18).

LBC samples are promising for cervicovaginal microbiome surveys, as they are an already established method of long-term cytological biobanking (19). In clinical practice, cervical cytology for cervical cancer screening or HPV genotyping is widely performed using a combination of cervical cytobrushes and LBC samples such as ThinPrep (HOLOGIC) or SurePath (BD). An LBC specimen can be used for not only cytological diagnosis but also additional diagnostic tests such as HPV, *Chlamydia*, *Neisseria gonorrhoeae*, and *Trichomonas* infection (20) (21) (22).

The ability to characterize microbial communities, as commonly assessed by 16S rRNA gene sequencing, can be biased as a result of methodological differences of cell lysis and DNA extraction protocols (23) (24) (25). Herein, we compare four different commercially available DNA extraction kits in an effort to assess their ability to characterize the cervical microbiota of LBC samples. Additionally, we examine the relationship between HPV infection and the composition of cervical microbiota.

Materials and Methods

Ethics

This study was approved by the Institutional Review Board at the University of Arkansas for Medical Sciences (IRB number 202790).

Sampling of cervical microbiome

LBC specimens were obtained from 20 patients enrolled in a Phase II clinical trial of an HPV therapeutic vaccine (NCT02481414). In order to be eligible, participants had to have high grade squamous intraepithelial lesions (HSILs) or cannot rule out HSILs in cervical cytology or CIN2/3 in cervical biopsy. Those who qualified for the study based on their cervical cytology underwent cervical biopsy, and they qualified for vaccination if the results were CIN2/3. The cervical cytology specimens in this current study were collected before the vaccination and reserved in the vial of the ThinPrep Pap Test (HOLOGIC) as described in Ravilla *et al.* 2019 (26). The storage period from sample collection to DNA extraction was 716 ± 105 days in this study.

HPV genotyping

HPV-DNA was detected by Linear Array HPV Genotyping Test (Roche Diagnostics) which can detect up to 37 HPV genotypes including 13 HR-HPV genotypes (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68) and 24 LR-HPV genotypes (6, 11, 26, 40, 42, 53, 54, 55, 61, 62, 64, 66, 67, 69, 70, 71, 72, 73, 81, 82, 83, 84, IS39, and CP6108) using ThinPrep solution (https://diagnostics.roche.com/global/en/products/params/linear-array-hpv-genotyping.html).

DNA extraction protocols

We selected four commercially available DNA extraction kits as the candidates for comparison: ZymoBIOMICS DNA Miniprep Kit (Zymo Research, D4300), QIAamp PowerFecal Pro DNA Kit (QIAGEN, 51804), QIAamp DNA Mini Kit (QIAGEN, 51304), and IndiSpin Pathogen Kit (Indical Bioscience, SPS4104). These kits have been successfully used in a variety of human cervical, vaginal, and gut microbiome surveys (10) (19) (27). We'll subsequently refer to each of these kits in abbreviated form as follows: ZymoBIOMICS, PowerFecalPro, QIAampMini, and IndiSpin. The protocols and any modifications are outlined in Table 1.

Each LBC sample was dispensed into four separate 2 mL sterile collection tubes (dispensed sample volume = 500 μL) to create four cohorts of 20 DNA extractions (Fig. 1). Each extraction cohort was processed through one of the four kits above. A total of 80 extractions (4 kits × 20 patients) were prepared for subsequent analyses. Applied sample volume of ThinPrep solution was 300 μL for ZymoBIOMICS, 300 μL for PowerFecalPro, and 200 μL for QIAampMini, and 300 μL for IndiSpin. The sample volume was standardized to 300 μL as long as the manufacturer's instructions allowed to do so. DNA extraction for all samples was performed by the same individual who practiced by performing multiple extractions for each kit before performing the actual DNA extraction on the samples analyzed in this study. Positive control was mock vaginal microbial communities composed of a mixture of genomic DNA from the American Type Culture Collection (ATCC MSA1007). Negative control was the ThinPrep preservation solution without the sample as blank extraction (28).

Measurement of DNA yield

DNA yield for each method was evaluated by spectrophotometer (Nanodrop One, Thermo Scientific). Analysis of the DNA yield from IndiSpin was omitted as nucleic acid is used as a carrier for this kit. The mean DNA yields per 100 μL ThinPrep sample volume were compared.

16S rRNA marker gene sequencing

Controls and the extracted DNA were sent to Argonne National Laboratory (IL, USA) for amplification and sequencing of the 16S rRNA gene on an Illumina MiSeq sequencing platform. Paired-end reads from libraries with ~250-bp inserts were generated for the V4 region using the barcoded primer set: 515FB: 5'-GTGYCAGCMGCCGCGGTAA-3' and 806RB: 5'-GGACTACNVGGGTWTCTAAT-3' (29) (30) (31) (32) (http://www.earthmicrobiome.org/protocols-and-standards/16s/). MiSeq Reagent Kit v2 (2 × 150 cycles, MS-102-2002) was used.

Sequence processing and analysis

Initial sequence processing and analyses were performed using QIIME 2 (33), any commands prefixed by `q2-` are QIIME 2 plugins. After demultiplexing of the paired-end reads by `q2-demux`, the imported sequence data was visually inspected via QIIME 2 View (https://view.qiime2.org), to determine the appropriate trimming and truncation parameters for generating Exact Sequence Variants (ESVs) (34) via `q2-dada2` (35). Hereafter, ESVs will be referred to as OTUs (Operational Taxonomic Units). The forward reads were trimmed at 15 bp and truncated at 150 bp; reverse reads were trimmed at 0 bp and truncated at 150 bp. The resulting OTUs were assigned taxonomy through `q2-feature-classifier classify-sklearn`, by using a pre-trained classifier for the amplicon region of interest (36). This enables

more robust taxonomic assignment of the OTUs (37). Taxonomy-based filtering was performed by using `q2-taxa filter-table` to remove any OTUs that were classified as "Chloroplast", "Mitochondria", "Eukaryota", "Unclassified" and those that did not have at least a Phylum-level classification. We then performed additional quality filtering via `q2-quality-control`, and only retained OTUs that had at least a 90% identity and 90% query alignment to the SILVA reference set (38). Then `q2-alignment` was used to generate a *de novo* alignment with MAFFT (39) which was subsequently masked by setting `max-gap-frequency 1 min-conservation 0.4`. Finally, q2-phylogeny was used to construct a midpoint-rooted phylogenetic tree using IQ-TREE (40) with automatic model selection using ModelFinder (41). Unless specified, subsequent analyses were performed after removing OTUs with a frequency of less than 0.0005% of the total data set (42).

Number of reads and OTUs before rarefying

Table 3 highlights the numbers of reads and OTUs among the DNA extraction protocols prior to rarefying the data. The reads and OTUs assigned to gram-positive and gram-negative was also shown. The number of "OTUs before rarefying" shown in Table 3 is distinguished from the "Observed OTUs" after rarefying in Fig. 3 for diversity analysis.

Microbiome analysis

To compare the taxonomic profiles among four types of DNA extraction methods (Fig. 1 & Table 1), the following analyses were performed; (I) bacterial microbiome composition, (II) detection of common and unique taxa, (III) alpha and beta diversity analysis, and (IV) identification of specific bacteria retained per DNA extraction method.

Microbiome composition

We generated the bar plot to exhibit bacterial microbiome composition per DNA extraction method at the family (Fig. 2A left) and genus (Fig. 2A right) taxonomic level. After all count data of taxonomy were converted to relative abundance, the top 10 abundant taxonomic groups in each family and genus level were plotted in colored bar plot (43) (44) (45). Variation of microbiome composition per DNA extraction method or per individual was assessed by the Adonis test (`q2-diversity adonis`) (46) (47).

Differentially accessible microbiota by DNA extraction protocol

We set out to determine which microbial taxonomic groups were differentially accessible across the sampling protocols by linear discriminant analysis (LDA) effect size (LEfSe) analyses (48). We further assessed the microbial taxa using jvenn (49) at family and genus level. The Venn diagram was created after removing OTUs with a frequency of less than 0.005% (42).

Alpha and beta diversity analyses with or without rarefying

Non-rarefying approaches to determine both alpha (within-sample) and beta (between-sample) diversity was assessed by Species richness using `q2-breakaway` (50) and Aitchison distance using `q2-deicode` (51). These were compared with rarefied data in which we applied Faith's Phylogenetic Diversity, Observed OTUs, Shannon's diversity index, Pielou's Evenness, Unweighted UniFrac distance, Weighted UniFrac distance, Jaccard distance, and Bray-Curtis distances via `q2-diversity` (33). In order to retain data from at least 15 of the 20 patients (*i.e.* 75%; four samples from each of the four DNA extraction methods), we set the sampling depth to

51,197 reads per sample. Overall our subsequence analysis consisted of 3,071,820 reads (27.6%, 3,071,820 / 11,149,582 reads). All diversity measurements in this study are listed in Table S1.

Community type and HPV status

In addition to the analysis above, we tested whether the samples clustered by microbiome composition were related to the patient's clinical and demographic characteristics such as, cervical biopsy diagnosis, race, and HPV16 status. HPV16 status has been reported to be associated with both racial differences as well as microbial community types (26) (52) (53) (54). We employed the Dirichlet Multinomial Mixtures (DMM) (55) model to determine the number of community types for bacterial cervical microbiome. Then, we clustered samples to the community type (9) (56). Since vaginal microbiota were reported to be clustered with different *Lactobacillus sp*. such as *L. crispatus*, *L. gasseri*, *L. iners*, or *L. jensenii* (16) (57), we also collapsed the taxonomy to the species level and performed a clustering analysis using "microbiome R package" (45). We then determined which bacterial taxa were differentially abundant among the patients with or without HPV16 via `q2-aldex2` (58) and LEfSe (48).

General statistical analysis

All data are presented as means ± standard deviation (SD). Comparisons were conducted with Fisher's exact test or Dunn's test with Benjamini-Hochberg-adjustment (59) or Wilcoxon test with Benjamini-Hochberg-adjustment or pairwise PERMANOVA when appropriate. A p value < 0.05 or a q value < 0.05 was considered statistically significant.

Data availability

MIMARKS compliant (60) DNA sequencing data are available via the Sequence Read Archive (SRA) at the National Center for Biotechnology Information (NCBI), under the BioProject Accession: PRJNA598197.

Results

Patients characteristics

The age of the patients was 31.4 ± 5.0 years. The distribution of race was 15% African American (n = 3) and 50% Caucasian (n = 10), and 35% Hispanic (n = 7). Cervical histology was 40% CIN2 (n = 8), 50% CIN3 (n = 10), and 10% benign (n = 2). HPV genotype were 50% HPV16 positive (n = 10), 10% HPV18 positive (n = 2), 90% HR-HPV positives (n = 18). Patient characteristics were summarized in Table 2.

DNA yield

DNA yield per 100 μL ThinPrep solution were 0.09 ± 0.06 μg in ZymoBIOMICS, 0.04 ± 0.01 μg in PowerFecalPro, and 0.21 ± 0.23 μg in QIAampMini. DNA yield was not calculated for IndiSpin, as Poly-A Carrier DNA was used. The DNA yield of PowerFecalPro was significantly lower than that of ZymoBIOMICS (adjusted p value < 0.001) and QIAampMini (adjusted p value < 0.001) based on Dunn's test with Benjamini-Hochberg-adjustment (Fig. S1).

Number of reads and OTUs before rarefying

We obtained a total of 11,149,582 reads for 80 DNA extractions. The 127,142 reads were produced from a positive control of mock sample and 1,773 reads from ThinPrep solution as the negative control. IndiSpin (168,349 ± 57,451 reads) produced a significantly higher number of reads compared to PowerFecalPro (115,610 ± 68,201 reads, p value = 0.020, Dunn's test with Benjamini-Hochberg-adjustment) as shown in Table 3. Approximately 90% of reads were assigned to gram-positive bacteria and about 10% of reads were assigned to gram-negative bacteria across all kits.

Prior to rarefying, the ZymoBIOMICS kit captured a greater representation of gram-negative bacterial OTUs (total 346, 17.3 ± 9.8) compared to PowerFecalPro (total 209, 10.5 ± 10.3, p value = 0.012, Dunn's test with Benjamini-Hochberg-adjustment, ratio of gram-negative bacteria: 41.9% vs 33.7%) as shown in Table 3. No significant differences in the number of OTUs before rarefying was detected for the entire bacterial community or gram-positive bacteria.

Microbiome composition per DNA extraction protocol

We analyzed whether differences in DNA extraction methods affect our ability to assess cervical microbiota composition. The patients can be identified by whether or not they displayed a *Lactobacillus*-dominant community type (Fig. 2A). Variation between individuals was a significantly greater influence on the observed microbial composition than was the method of DNA extraction (Fig. 2A).

The following top 10 abundant families are shown in Fig. 2A (left) and constituted approximately 95.7% of cervical bacteria in all kits (80 DNA extractions); *Lactobacillaceae* (58.9%), *Bifidobacteriaceae* (13.7%), *Veillonellaceae* (4.8%), *Prevotellaceae* (4.3%), *Family XI* (3.9%), *Atopobiaceae* (3.0%), *Leptotrichiaceae* (2.5%), *Streptococcaceae* (2.0%), *Lachnospiraceae* (1.6%). *Ruminococcaceae* (0.9%). The following top 10 abundant genera are shown in Fig. 2A (right) and constituted approximately 92% of cervical bacteria; *Lactobacillus* (58.9%), *Gardnerella* (13.6%), *Prevotella* (4.2%), *Megasphaera* (3.7%), *Atopobium* (3.0%), *Sneathia* (2.5%), *Streptococcus* (1.9%), *Parvimonas* (1.7%), *Shuttleworthia* (1.4%), and *Anaerococcus* (1.1%).

Shared and unique microbiota among DNA extraction protocols

All DNA extraction methods were generally commensurate with one another, there were 31 of 41 shared microbes at the family level (Fig. 2B left) and 45 of 57 shared microbes at the genus level (Fig. 2B right) among the DNA extraction protocols.

However, four gram-negative taxa were uniquely detected by ZymoBIOMICS and one taxon was uniquely detected by QIAampMini both at the genus level (Fig. 2B right). Of the uniquely detected ZymoBIOMICS OTUs, *Hydrogenophilus*, which was reported as enriched taxa in LR-HPV positive environment (3), was detected in 14 of the 80 DNA extractions, consisting of 2,488 reads (0.02% of all kit extractions). *Methylobacterium* was detected in 5 of the 80 DNA extractions (912 reads; 0.01%). A member of this genus, *Methylobacterium aerolatum*, has been reported to be more abundant in the endocervix than the vagina of healthy South African women (61). *Bacteroidetes*, which are often reported as enriched taxa in an HIV positive cervical environment (62), was detected in 12 of the 80 DNA extractions (1,028 reads; 0.01%). *Meiothermus* was detected in 9 of the 80 DNA extractions (882 reads; 0.01%). *Meiothermus* is not considered to reside within the human environment, and may be an extraction kit contaminant, as previously reported (63). A unique gram-positive taxa obtained from the QIAampMini, *Streptomyces*, which was reported to be detected from the cervicovaginal environment in the study of Kenyan women (64), was detected in 20 of 80 DNA extractions (6,862 reads; 0.06%). No unique taxa were detected in PowerFecalPro and IndiSpin.

Venn diagrams at family levels also exhibited that ZymoBIOMICS detected slightly more bacterial taxa (four unique taxa) as shown in Fig. 2B (left). These results showed that major bacteria were commonly detected among all extraction protocols, with only slightly more uniquely detected microbiota using ZymoBIOMICS.

Alpha and beta diversity

Significantly higher Species richness (`q2-breakaway`) was observed from the ZymoBIOMICS (56.1 ± 19.4) protocol compared to that of PowerFecalPro (43.2 ± 32.9, p = 0.025), QIAampMini (54.9 ± 29.8, not significant), and IndiSpin (63.6 ± 38.3, not significant) using Dunn's test with Benjamini-Hochberg-adjustment (Fig. 3). Similarly, Faith's Phylogenetic Diversity was observed to be higher with the ZymoBIOMICS protocol (6.6 ± 2.2), compared to PowerFecalPro (4.5 ± 1.9, p = 0.012), QIAampMini (5.0 ± 1.8, not significant), and IndiSpin (5.4 ± 1.7, not significant) using Dunn's test with Benjamini-Hochberg-adjustment (Fig. 3). The use of IndiSpin also resulted significantly higher alpha diversity than that of PowerFecalPro in an analysis of Species richness (p = 0.042, Dunn's test with Benjamini-Hochberg-adjustment). Non-phylogenetic alpha diversity metrics such as Observed OTUs, Shannon's diversity index, and Pielou's Evenness did not show differences among the four methods.

ZymoBIOMICS was able to significantly increase access to several taxonomic groups compared to the other DNA extraction methods. Additionally, as shown in Table 4, ZymoBIOMICS did capture a different microbial composition compared to other DNA extraction methods in the index of Unweighted UniFrac distances (PowerFecalPro: q = 0.002; QIAampMini: q = 0.002;and IndiSpin: q = 0.002) and in Jaccard distances (QIAampMini: q = 0.018 and IndiSpin: q = 0.033).

Differential accessibility of microbiota by DNA extraction protocol

LEfSe analysis identified taxonomic groups, defined with an LDA score of 2 or higher, for differential accessibility by extraction kit: 23 in ZymoBIOMICS, 0 in PowerFecalPro, 3 in QIAampMini, and 3 in IndiSpin (Fig. 4A). The following taxa were found to be highly

accessible (LDA score > 3) with the use of the ZymoBIOMICS kit: Phylum *Proteobacteria*, Class *Gammaproteobacteria*, Order *Betaproteobacteriales*, Family *Bacillaceae*, and Genus *Anoxybacillus*. Whereas the Order *Streptomycetales* was highly enriched with the use of the QIAampMini (LDA score > 3). As shown in the cladogram (Fig. 4B), despite the detection of a potential kit contaminant (*Meiothermus sp.* a member of the Phylum *Deinococcus-Thermus*), ZymoBIOMICS was able to increase the accessibility to additional microbiota compared to the other extraction protocols.

Microbial community type and HPV16

DMM detected two cervical microbial community types across all four DNA extraction protocols. Community type I was composed of the following: *Gardnerella sp.* (ZymoBIOMICS: 17.1%; PowerFecalPro: 20%; QIAampMini: 23%; IndiSpin: 20%), *Lactobacillus iners* (ZymoBIOMICS: 6.3%; PowerFecalPro: 5%; QIAampMini: 6%; IndiSpin: 5%), *Atopobium vaginae* (10) (ZymoBIOMICS: 3.5%; PowerFecalPro: 3%; QIAampMini: 4%; IndiSpin: 5%), *Clamydia trachomatis* (ZymoBIOMICS: 1.9%; PowerFecalPro: 2%; QIAampMini: 3%; IndiSpin: 2%), *Shuttleworthia sp.* (ZymoBIOMICS: 1.8%; PowerFecalPro: 2%; QIAampMini: 2%; IndiSpin: 2%). Some members of *Shuttleworthia* are considered to be bacterial vaginosis associated bacterium (BVAB) (65), further investigation is required to determine if this OTU is indeed a BVAB. We determined this community type “high diversity type”. Community type II was is dominated by *Lactobacillus iners* at 88%, 85%, 83%, and 85% respectively for ZymoBIOMICS, PowerFecalPro, QIAampMini, and IndiSpin.

The relationship between HPV16 infection and community type was observed to be significantly associated with community type I (HPV16 positive patients [n = 9], HPV16

negative patients [n = 1]) and not community type II (HPV16 positive patients [n = 1], HPV16 negative patients [n = 9], p = 0.001, Fisher's exact test) regardless of the DNA extraction kit used. In support of this result, analysis of differentially abundant microbiota using `q2-aldex` (Benjamini-Hochberg corrected p value of Wilcoxon test: $p < 0.001$, standardized distributional effect size: −1.2) revealed that *Lactobacillus iners* were differentially enriched in the cervical environment without HPV16. LEfSe analysis also detected that genus *Lactobacillus* were enriched in the cervical environment without HPV16 ($p < 0.001$, LDA score: −5.38). No significant differences were observed in the relationship between community type and HPV18 (p = 0.474, Fisher's exact test), HR-HPV (p = 0.474, Fisher's exact test), results of cervical biopsy (p = 0.554, Fisher's exact test), and race (African Americans vs not-African Americans: p = 1; Caucasian vs not-Caucasian: p = 0.656; Hispanic vs not-Hispanic: p = 0.350, Fisher's exact test).

Discussion

In this study, we evaluated the utility of LBC specimens for the collection and storage of cervical samples for microbiome surveys based on the 16S rRNA marker gene. We simultaneously compared the efficacy of several commonly used DNA extraction protocols on these samples in an effort to develop a standard operating procedure/protocol (SOP) for such work. We've also been able to show that there are two cervical microbial community types, which are associated with the dominance or non-dominance of *Lactobacillis iners*. Both community types were detected regardless of the DNA extraction protocol used.

This study evaluated the composition of microbiota across all DNA extraction methods. These findings document the importance of selecting DNA extraction methods in cervical microbiome studies from the LBC samples. All kits were commensurate in their ability to capture the microbial composition of each patient and the two observed cervical microbial community state types: making all of these protocols viable for discovering broad patterns of microbial diversity. However, we did observe that the ZymoBIOMICS protocol was better able to access additional cervical microbiota (Fig. 2B, 4A & B). Coincidentally, we detected potential DNA contamination only with the ZymoBIOMICS kit. The number of OTUs prior to rarefying revealed that the ZymoBIOMICS protocol detected more gram-negative OTUs than the PowerFecalPro (Table 3 & Fig. 2B). In particular, LEfSe analysis has shown that phylum *Proteobacteria* can be better detected with the ZymoBIOMICS kit (Fig. 4). Although rarefying microbiome data can be problematic (66), it can still provide robust and interpretable results for diversity analysis (67), we were able to observe commensurate findings with non-rarefying approaches such as `q2-breakaway` (50), `q2-deicode` (51), and LEfSe (48). Beta-diversity analysis via Unweighted UniFrac also revealed that ZymoBIOMICS was significantly different

from all other kits. There were no differences in non-phylogenic indices of alpha diversity with rarefying approaches. These findings lead us to surmise that phylogenetic indices may be more sensitive than the non-phylogenetic indices.

Although we hypothesized that the detection of difficult-to-lyse-bacteria (*e.g*. gram-positive bacteria) would vary by kit, we observed no significant differences (Table 3). As shown in Table 3, the number of reads of gram-positive and gram-negative bacteria also showed that there was no difference in the four kits. This is likely due to several modifications made to the extraction protocol as outlined in Table 1. That is, we added bead beating and mutanolysin to the QIAampMini protocol (68). We also modified the beating time of the ZymoBIOMICS kit down to 2 minutes from 10 minutes (the latter being recommended by the manufacturer) to minimize DNA shearing. We may use the extracted DNA from ZymoBIOMICS for long-read amplicon sequencing platforms such as PacBio (Pacific Biosciences of California, Inc) (69) or MinION (Oxford Nanopore Technologies) (70) (71). Excessive shearing can render these samples unusable for long-read sequencing. It is quite possible that we could have observed even more diversity with the ZymoBIOMICS kit for our amplicon survey if we conducted bead-beating for the full 10 minutes.

Community typing and detection of the differentially abundant microbiota revealed that *Lactobacillus iners* were more abundant in the cervical ecosystem without HPV16. These findings are congruent with those of Lee *et al*. (1) and Audirac-Chalifour *et al*. (72). Lee *et al*. reported that *Lactobacillus iners* were decreased in women with HPV positive (1). Also, the result that the proportion of *Lactobacillus iners* was higher in HPV-negative women compared to HPV-positive women (relative abundance 14.9% vs 2.1%) was reported by Audirac-Chalifour *et al* (72). Similarly, Tuominen *et al*. (18) reported that *Lactobacillus iners* were enriched in HPV

negative samples (relative abundance: 47.7%) compared to HPV positive samples (relative abundance: 18.6%, p value = 0.07) in the study of HPV positive-pregnant women (HPV16 positive rate: 15%). As established by the seminal study of Ranjeva *et al*. (73), a statistical model revealed that colonization of specific HPV type including multi HPV type infection depends on host-risk factors such as sexual behavior, race and ethnicity, and smoking. It is unclear whether the association between the cervical microbiome, host-specific traits, persistent infection of specific HPV types, such as HPV16, can be generalized and requires further investigation.

We focused on LBC samples as this is the recommended method of storage for cervical cytology (74). Here, we confirmed that LBC samples can be used for microbial community surveys by simply using the remaining LBC solution post HPV testing or cervical cytology. We used a sample volume of 200 or 300 μL ThinPrep solution in this study. HPV genotyping test using Linear Array HPV Genotyping Test (Roche Diagnostics) stably detects β-globin with a base length of 268 bp as a positive control. Therefore, using a similar sample volume as HPV genotyping (250 μL), it was expected that V4 (250 bp), which is near the base length of β-globin, would be PCR amplified. It has been pointed out by Ling *et al*. (75) that the cervical environment is of low microbial biomass. To control reagent DNA contamination and estimate the sample volume, DNA quantification by qPCR before sequencing is recommended (76). Mitra *et al* determined a sample volume of 500 μL for ThinPrep by qPCR in the microbiome study comparing sampling methods using cytobrush or swab from cervix (19). The average storage period from sample collection via LBC to DNA extraction was about two years in this study. Kim *et al*. reported that DNA from cervix stored in ThinPrep at room temperature or −80°C was stable for at least one year (77). Meanwhile, Castle *et al*. reported that β-globin DNA fragments of 268 bases or more were detected by PCR in 90 % (27 of 30 samples) of ThinPrep samples

stored for eight years at an uncontrolled ambient temperatures followed by a controlled ambient environment (10–26.7°C) (78). Low-temperature storage may allow the analysis of the short DNA fragments of the V4 region after even long-term storage, although further research is needed to confirm the optimal storage period in cervical microbiome studies using ThinPrep. SurePath LBC specimens are as widely used as ThinPrep, but the presence of formaldehyde within the SurePath preservation solution raises concerns about accessing enough DNA for analysis as compared to ThinPrep, which contains methanol (79) (80). It should also be noted that other storage solutions, *i.e*. those using guanidine thiocyanate have been reported for microbiome surveys of the cervix (81) and feces (82). A weakness of the current study is that we did not examine the reproducibility of our results as each sample was extracted using each kit once. However, the use of actual patient samples rather than mock samples is a strength of our approach.

In conclusion, regardless of the extraction protocol used, all kits provided equivalent accessibility to the cervical microbiome. All kits shared the ability to access 31 of 41 families and 45 of 57 genera (Fig. 2), approximately 90% of bacteria were gram-positive and 10% were gram-negative. Observed differences in microbial composition were due to the significant influence of the individual patient and not the extraction protocol. However, ZymoBIOMICS was observed to increase the accessibility of DNA from a greater range of microbiota compared to the other kits, in that the greatest number of significantly enriched taxa were identified (Fig. 3). This was not because of higher DNA yield nor ability to detect more gram-positive bacteria. Selection and characterization of an appropriate DNA extraction methods, for providing accurate census of cervical microbiota, and the human microbiome in general are important (23) (24) (25) (68) (77) (78). We have shown that the ability to characterize cervical microbiota from LBC

specimens is robust, even after prolonged storage. Our data also suggest that it is possible to reliably assess the relationship between HPV and the cervical microbiome, also supported by Kim *et al.* (77) and Castle *et al* (78). Even though we found all four extraction kits to be commensurate in their ability to broadly characterize the CM, this study lends support to the view that the selection of a DNA extraction kit depends on the questions asked of the data, and should be taken into account for any cervicovaginal microbiome and HPV research that leverages LBC specimens for use in clinical practice (15) (83).

Acknowledgments

We thank Togo Picture Gallery (http://togotv.dbcls.jp/pics.html) for stock images shown in Fig. 1.

This work was supported by the National Institutes of Health (R01CA143130, USA), Drs. Mae and Anderson Nettleship Endowed Chair of Oncologic Pathology (31005156, USA), and the Arkansas Biosciences Institute (the major component of the Tobacco Settlement Proceeds Act of 2000, G1-52249-01, USA).

M.N. designed and supervised this project. T.S. and M.S.R. conducted bioinformatics analysis and wrote paper. T.S., H.C., and M.N. created the protocol of DNA extraction. M.N., H.C., S.O., W.G., and T.S. provided important feedback. Samples in the clinical trial were collected by W.G. and his associates. DNA extraction was conducted by T.S. Sequencing of 16S RNA gene was conducted by S.O.

M.N. is one of the inventors named in the patents and patent applications for the HPV therapeutic vaccine PepCan. The remaining authors declare no conflicts of interest.

References

1. Lee JE, Lee S, Lee H, Song YM, Lee K, Han MJ, Sung J, Ko G. 2013. Association of the vaginal microbiota with human papillomavirus infection in a Korean twin cohort. PLoS One 8:e63514.
2. Huang X, Li C, Li F, Zhao J, Wan X, Wang K. 2018. Cervicovaginal microbiota composition correlates with the acquisition of high-risk human papillomavirus types. Int J Cancer 143:621-634.
3. Zhou Y, Wang L, Pei F, Ji M, Zhang F, Sun Y, Zhao Q, Hong Y, Wang X, Tian J, Wang Y. 2019. Patients With LR-HPV Infection Have a Distinct Vaginal Microbiota in Comparison With Healthy Controls. Front Cell Infect Microbiol 9:294.
4. Onywera H, Williamson AL, Mbulawa ZZA, Coetzee D, Meiring TL. 2019. The cervical microbiota in reproductive-age South African women with and without human papillomavirus infection. Papillomavirus Res 7:154-163.
5. Brotman RM, Shardell MD, Gajer P, Tracy JK, Zenilman JM, Ravel J, Gravitt PE. 2014. Interplay between the temporal dynamics of the vaginal microbiota and human papillomavirus detection. J Infect Dis 210:1723-33.
6. Godoy-Vitorino F, Romaguera J, Zhao C, Vargas-Robles D, Ortiz-Morales G, Vazquez-Sanchez F, Sanchez-Vazquez M, de la Garza-Casillas M, Martinez-Ferrer M, White JR, Bittinger K, Dominguez-Bello MG, Blaser MJ. 2018. Cervicovaginal Fungi and Bacteria Associated With Cervical Intraepithelial Neoplasia and High-Risk Human Papillomavirus Infections in a Hispanic Population. Front Microbiol 9:2533.
7. Łaniewski P, Barnes D, Goulder A, Cui H, Roe DJ, Chase DM, Herbst-Kralovetz MM. 2018. Linking cervicovaginal immune signatures, HPV and microbiota composition in cervical carcinogenesis in non-Hispanic and Hispanic women, Sci Rep, vol 8.
8. Mitra A, MacIntyre DA, Lee YS, Smith A, Marchesi JR, Lehne B, Bhatia R, Lyons D, Paraskevaidis E, Li JV, Holmes E, Nicholson JK, Bennett PR, Kyrgiou M. 2015. Cervical intraepithelial neoplasia disease progression is associated with increased vaginal microbiome diversity. Sci Rep 5:16865.
9. Piyathilake CJ, Ollberding NJ, Kumar R, Macaluso M, Alvarez RD, Morrow CD. 2016. Cervical Microbiota Associated with Higher Grade Cervical Intraepithelial Neoplasia in Women Infected with High-Risk Human Papillomaviruses. Cancer Prev Res (Phila) 9:357-66.
10. Oh HY, Kim BS, Seo SS, Kong JS, Lee JK, Park SY, Hong KM, Kim HK, Kim MK. 2015. The association of uterine cervical microbiota with an increased risk for cervical intraepithelial neoplasia in Korea. Clin Microbiol Infect 21:674 e1-9.
11. De Seta F, Campisciano G, Zanotta N, Ricci G, Comar M. 2019. The Vaginal Community State Types Microbiome-Immune Network as Key Factor for Bacterial Vaginosis and Aerobic Vaginitis. Front Microbiol 10:2451.
12. Oliver A, LaMere B, Weihe C, Wandro S, Lindsay KL, Wadhwa PD, Mills DA, Pride D, Fiehn O, Northen T, de Raad M, Li H, Martiny JBH, Lynch S, Whiteson K. 2019. Cervicovaginal microbiome composition drives metabolic profiles in healthy pregnancy. bioRxiv https://doi.org/10.1101/840520.

13. Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. Nature 486:207-14.
14. Bik EM, Bird SW, Bustamante JP, Leon LE, Nieto PA, Addae K, Alegria-Mera V, Bravo C, Bravo D, Cardenas JP, Carson GA, Caughey A, Covarrubias PC, Perez-Donoso J, Gass G, Gupta SL, Harman K, Hongo DMB, Jimenez JC, Kraal L, Melis-Arcos F, Morales EH, Morton A, Navas CF, Nunez H, Olivares E, Ordenes-Aenishanslins N, Ossandon FJ, Phan R, Pino R, Soto-Liebe K, Varas I, Vera-Wolf P, Walton NA, Almonacid DE, Goddard AD, Ugalde JA, Zneimer S, Richman J, Apte ZS. 2019. A novel sequencing-based vaginal health assay combining self-sampling, HPV detection and genotyping, STI detection, and vaginal microbiome analysis. PLoS One 14:e0215945.
15. Berman HL, McLaren MR, Callahan BJ. 2019. Understanding and Interpreting Community Sequencing Measurements of the Vaginal Microbiome. BJOG doi:10.1111/1471-0528.15978.
16. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, Brotman RM, Davis CC, Ault K, Peralta L, Forney LJ. 2011. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A 108 Suppl 1:4680-7.
17. Fettweis JM, Serrano MG, Brooks JP, Edwards DJ, Girerd PH, Parikh HI, Huang B, Arodz TJ, Edupuganti L, Glascock AL, Xu J, Jimenez NR, Vivadelli SC, Fong SS, Sheth NU, Jean S, Lee V, Bokhari YA, Lara AM, Mistry SD, Duckworth RA, 3rd, Bradley SP, Koparde VN, Orenda XV, Milton SH, Rozycki SK, Matveyev AV, Wright ML, Huzurbazar SV, Jackson EM, Smirnova E, Korlach J, Tsai YC, Dickinson MR, Brooks JL, Drake JI, Chaffin DO, Sexton AL, Gravett MG, Rubens CE, Wijesooriya NR, Hendricks-Munoz KD, Jefferson KK, Strauss JF, 3rd, Buck GA. 2019. The vaginal microbiome and preterm birth. Nat Med 25:1012-1021.
18. Tuominen H, Rautava S, Syrjanen S, Collado MC, Rautava J. 2018. HPV infection and bacterial microbiota in the placenta, uterine cervix and oral mucosa. Sci Rep 8:9787.
19. Mitra A, MacIntyre DA, Mahajan V, Lee YS, Smith A, Marchesi JR, Lyons D, Bennett PR, Kyrgiou M. 2017. Comparison of vaginal microbiota sampling techniques: cytobrush versus swab. Sci Rep 7:9802.
20. Bentz JS. 2005. Liquid-based cytology for cervical cancer screening. Expert Rev Mol Diagn 5:857-71.
21. Gibb RK, Martens MG. 2011. The impact of liquid-based cytology in decreasing the incidence of cervical cancer. Rev Obstet Gynecol 4:S2-S11.
22. Donders GG, Depuydt CE, Bogers JP, Vereecken AJ. 2013. Association of Trichomonas vaginalis and cytological abnormalities of the cervix in low risk women. PLoS One 8:e86266.
23. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, Tramontano M, Driessen M, Hercog R, Jung FE, Kultima JR, Hayward MR, Coelho LP, Allen-Vercoe E, Bertrand L, Blaut M, Brown JRM, Carton T, Cools-Portier S, Daigneault M, Derrien M, Druesne A, de Vos WM, Finlay BB, Flint HJ, Guarner F, Hattori M, Heilig H, Luna RA, van Hylckama Vlieg J, Junick J, Klymiuk I, Langella P, Le Chatelier E, Mai V, Manichanh C, Martin JC, Mery C, Morita H, O'Toole PW, Orvain C, Patil KR, Penders J, Persson S, Pons N, Popova M, Salonen A, Saulnier D, Scott KP, Singh B, et al. 2017. Towards standards for human fecal sample processing in metagenomic studies. Nat Biotechnol 35:1069-1076.

559 24. Stinson LF, Keelan JA, Payne MS. 2018. Comparison of Meconium DNA Extraction
560 Methods for Use in Microbiome Studies. Front Microbiol 9:270.
561 25. Teng F, Darveekaran Nair SS, Zhu P, Li S, Huang S, Li X, Xu J, Yang F. 2018. Impact of DNA
562 extraction method and targeted 16S-rRNA hypervariable region on oral microbiota
563 profiling. Sci Rep 8:16321.
564 26. Ravilla R, Coleman HN, Chow CE, Chan L, Fuhrman BJ, Greenfield WW, Robeson MS,
565 Iverson K, Spencer H, 3rd, Nakagawa M. 2019. Cervical Microbiome and Response to a
566 Human Papillomavirus Therapeutic Vaccine for Treating High-Grade Cervical Squamous
567 Intraepithelial Lesion. Integr Cancer Ther 18:1534735419893063.
568 27. Virtanen S, Kalliala I, Nieminen P, Salonen A. 2017. Comparative analysis of vaginal
569 microbiota sampling using 16S rRNA gene analysis. PLoS One 12:e0181477.
570 28. Kim D, Hofstaedter CE, Zhao C, Mattei L, Tanes C, Clarke E, Lauder A, Sherrill-Mix S,
571 Chehoud C, Kelsen J, Conrad M, Collman RG, Baldassano R, Bushman FD, Bittinger K.
572 2017. Optimizing methods and dodging pitfalls in microbiome research. Microbiome
573 5:52.
574 29. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A,
575 Gibbons SM, Ackermann G, Navas-Molina JA, Janssen S, Kopylova E, Vazquez-Baeza Y,
576 Gonzalez A, Morton JT, Mirarab S, Zech Xu Z, Jiang L, Haroon MF, Kanbar J, Zhu Q, Jin
577 Song S, Kosciolek T, Bokulich NA, Lefler J, Brislawn CJ, Humphrey G, Owens SM,
578 Hampton-Marcell J, Berg-Lyons D, McKenzie V, Fierer N, Fuhrman JA, Clauset A, Stevens
579 RL, Shade A, Pollard KS, Goodwin KD, Jansson JK, Gilbert JA, Knight R, Earth Microbiome
580 Project C. 2017. A communal catalogue reveals Earth's multiscale microbial diversity.
581 Nature 551:457-463.
582 30. Apprill A, McNally S, Parsons R, Weber L. 2015. Minor revision to V4 region SSU rRNA
583 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat Microb
584 Ecol 75:129-137.
585 31. Parada AE, Needham DM, Fuhrman JA. 2016. Every base matters: assessing small
586 subunit rRNA primers for marine microbiomes with mock communities, time series and
587 global field samples. Environ Microbiol 18:1403-14.
588 32. Walters W, Hyde ER, Berg-Lyons D, Ackermann G, Humphrey G, Parada A, Gilbert JA,
589 Jansson JK, Caporaso JG, Fuhrman JA, Apprill A, Knight R. 2016. Improved Bacterial 16S
590 rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers
591 for Microbial Community Surveys. mSystems 1.
592 33. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm
593 EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown
594 CT, Callahan BJ, Caraballo-Rodriguez AM, Chase J, Cope EK, Da Silva R, Diener C,
595 Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M,
596 Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B,
597 Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler
598 BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, et al. 2019. Reproducible,
599 interactive, scalable and extensible microbiome data science using QIIME 2. Nat
600 Biotechnol 37:852-857.
601 34. Callahan BJ, McMurdie PJ, Holmes SP. 2017. Exact sequence variants should replace
602 operational taxonomic units in marker-gene data analysis. ISME J 11:2639-2643.

603 35. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. 2016. DADA2:
604 High-resolution sample inference from Illumina amplicon data. Nat Methods 13:581-3.
605 36. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Gregory
606 Caporaso J. 2018. Optimizing taxonomic classification of marker-gene amplicon
607 sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6:90.
608 37. Werner JJ, Koren O, Hugenholtz P, DeSantis TZ, Walters WA, Caporaso JG, Angenent LT,
609 Knight R, Ley RE. 2012. Impact of training sets on classification of high-throughput
610 bacterial 16s rRNA gene surveys. ISME J 6:94-103.
611 38. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013.
612 The SILVA ribosomal RNA gene database project: improved data processing and web-
613 based tools. Nucleic Acids Res 41:D590-6.
614 39. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
615 improvements in performance and usability. Mol Biol Evol 30:772-80.
616 40. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective
617 stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol
618 32:268-74.
619 41. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017.
620 ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods
621 14:587-589.
622 42. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso
623 JG. 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon
624 sequencing. Nat Methods 10:57-9.
625 43. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive
626 analysis and graphics of microbiome census data. PLoS One 8:e61217.
627 44. Bisanz JE. 2018. qiime2R: Importing QIIME2 artifacts and associated data into R sessions.
628 https://github.com/jbisanz/qiime2R.
629 45. Lahti L, Shetty S. 2012-2019. microbiome R package. http://microbiome.github.io.
630 46. Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance.
631 Austral Ecol 26:32-46.
632 47. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara
633 RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2019. vegan: Community
634 Ecology Package. 2018. R package version 2.5-3. https://CRAN.R-
635 project.org/package=vegan.
636 48. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. 2011.
637 Metagenomic biomarker discovery and explanation. Genome Biol 12:R60.
638 49. Bardou P, Mariette J, Escudie F, Djemiel C, Klopp C. 2014. jvenn: an interactive Venn
639 diagram viewer. BMC Bioinformatics 15:293.
640 50. Willis A, Bunge J. 2015. Estimating diversity via frequency ratios. Biometrics 71:1042-9.
641 51. Martino C, Morton JT, Marotz CA, Thompson LR, Tripathi A, Knight R, Zengler K. 2019. A
642 Novel Sparse Compositional Technique Reveals Microbial Perturbations. mSystems 4.
643 52. Gao W, Weng J, Gao Y, Chen X. 2013. Comparison of the vaginal microbiota diversity of
644 women with and without human papillomavirus infection: a cross-sectional study. BMC
645 Infect Dis 13:271.

646 53. Montealegre JR, Peckham-Gregory EC, Marquez-Do D, Dillon L, Guillaud M, Adler-
647 Storthz K, Follen M, Scheurer ME. 2018. Racial/ethnic differences in HPV 16/18
648 genotypes and integration status among women with a history of cytological
649 abnormalities. Gynecol Oncol 148:357-362.
650 54. Xi LF, Kiviat NB, Hildesheim A, Galloway DA, Wheeler CM, Ho J, Koutsky LA. 2006.
651 Human papillomavirus type 16 and 18 variants: race-related distribution and persistence.
652 J Natl Cancer Inst 98:1045-52.
653 55. Morgan M. 2019. DirichletMultinomial: Dirichlet-Multinomial Mixture Model Machine
654 Learning for Microbiome Data.
655 http://bioconductor.org/packages/release/bioc/html/DirichletMultinomial.html.
656 56. Holmes I, Harris K, Quince C. 2012. Dirichlet multinomial mixtures: generative models
657 for microbial metagenomics. PLoS One 7:e30126.
658 57. DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, Sun CL,
659 Goltsman DS, Wong RJ, Shaw G, Stevenson DK, Holmes SP, Relman DA. 2015. Temporal
660 and spatial variation of the human microbiota during pregnancy. Proc Natl Acad Sci U S
661 A 112:11060-5.
662 58. Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB. 2013. ANOVA-like differential
663 expression (ALDEx) analysis for mixed population RNA-Seq. PLoS One 8:e67019.
664 59. Dinno A. 2017. dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums.
665 https://CRAN.R-project.org/package=dunn.test.
666 60. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, Gilbert JA, Karsch-
667 Mizrachi I, Johnston A, Cochrane G, Vaughan R, Hunter C, Park J, Morrison N, Rocca-
668 Serra P, Sterk P, Arumugam M, Bailey M, Baumgartner L, Birren BW, Blaser MJ, Bonazzi
669 V, Booth T, Bork P, Bushman FD, Buttigieg PL, Chain PS, Charlson E, Costello EK, Huot-
670 Creasy H, Dawyndt P, DeSantis T, Fierer N, Fuhrman JA, Gallery RE, Gevers D, Gibbs RA,
671 San Gil I, Gonzalez A, Gordon JI, Guralnick R, Hankeln W, Highlander S, Hugenholtz P,
672 Jansson J, Kau AL, Kelley ST, Kennedy J, Knights D, Koren O, et al. 2011. Minimum
673 information about a marker gene sequence (MIMARKS) and minimum information
674 about any (x) sequence (MIxS) specifications. Nat Biotechnol 29:415-20.
675 61. Balle C, Lennard K, Dabee S, Barnabas SL, Jaumdally SZ, Gasper MA, Maseko V, Mbulawa
676 ZZA, Williamson AL, Bekker LG, Lewis DA, Passmore JS, Jaspan HB. 2018. Endocervical
677 and vaginal microbiota in South African adolescents with asymptomatic Chlamydia
678 trachomatis infection. Sci Rep 8:11109.
679 62. Klein C, Gonzalez D, Samwel K, Kahesa C, Mwaiselage J, Aluthge N, Fernando S, West JT,
680 Wood C, Angeletti PC. 2019. Relationship between the Cervical Microbiome, HIV Status,
681 and Precancerous Lesions. MBio 10.
682 63. Glassing A, Dowd SE, Galandiuk S, Davis B, Chiodini RJ. 2016. Inherent bacterial DNA
683 contamination of extraction and sequencing reagents may affect interpretation of
684 microbiota in low bacterial biomass samples, Gut Pathog, vol 8.
685 64. Birse KD, Romas LM, Guthrie BL, Nilsson P, Bosire R, Kiarie J, Farquhar C, Broliden K,
686 Burgener AD. 2017. Genital Injury Signatures and Microbiome Alterations Associated
687 With Depot Medroxyprogesterone Acetate Usage and Intravaginal Drying Practices. J
688 Infect Dis 215:590-598.

689 65. Lennard K, Dabee S, Barnabas SL, Havyarimana E, Blakney A, Jaumdally SZ, Botha G,
690 Mkhize NN, Bekker LG, Lewis DA, Gray G, Mulder N, Passmore JS, Jaspan HB. 2018.
691 Microbial Composition Predicts Genital Tract Inflammation and Persistent Bacterial
692 Vaginosis in South African Adolescent Females. Infect Immun 86.
693 66. McMurdie PJ, Holmes S. 2014. Waste not, want not: why rarefying microbiome data is
694 inadmissible. PLoS Comput Biol 10:e1003531.
695 67. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, Lozupone C, Zaneveld JR,
696 Vazquez-Baeza Y, Birmingham A, Hyde ER, Knight R. 2017. Normalization and microbial
697 differential abundance strategies depend upon data characteristics. Microbiome 5:27.
698 68. Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ. 2012. Evaluation of methods for the
699 extraction and purification of DNA from the human microbiome. PLoS One 7:e33865.
700 69. Callahan BJ, Wong J, Heiner C, Oh S, Theriot CM, Gulati AS, McGill SK, Dougherty MK.
701 2019. High-throughput amplicon sequencing of the full-length 16S rRNA gene with
702 single-nucleotide resolution. Nucleic Acids Res 47:e103.
703 70. Calus ST, Ijaz UZ, Pinto AJ. 2018. NanoAmpli-Seq: a workflow for amplicon sequencing
704 for mixed microbial communities on the nanopore sequencing platform. Gigascience 7.
705 71. Wongsurawat T, Nakagawa M, Atiq O, Coleman HN, Jenjaroenpun P, Allred JI, Trammel
706 A, Puengrang P, Ussery DW, Nookaew I. 2019. An assessment of Oxford Nanopore
707 sequencing for human gut metagenome profiling: A pilot study of head and neck cancer
708 patients. J Microbiol Methods 166:105739.
709 72. Audirac-Chalifour A, Torres-Poveda K, Bahena-Roman M, Tellez-Sosa J, Martinez-
710 Barnetche J, Cortina-Ceballos B, Lopez-Estrada G, Delgado-Romero K, Burguete-Garcia
711 AI, Cantu D, Garcia-Carranca A, Madrid-Marina V. 2016. Cervical Microbiome and
712 Cytokine Profile at Various Stages of Cervical Cancer: A Pilot Study. PLoS One
713 11:e0153274.
714 73. Ranjeva SL, Mihaljevic JR, Joseph MB, Giuliano AR, Dwyer G. 2019. Untangling the
715 dynamics of persistence and colonization in microbial communities. ISME J
716 doi:10.1038/s41396-019-0488-7:1-13.
717 74. Linder J, Zahniser D. 1998. ThinPrep Papanicolaou testing to reduce false-negative
718 cervical cytology. Arch Pathol Lab Med 122:139-44.
719 75. Ling Z, Liu X, Chen X, Zhu H, Nelson KE, Xia Y, Li L, Xiang C. 2011. Diversity of
720 cervicovaginal microbiota associated with female lower genital tract infections. Microb
721 Ecol 61:704-14.
722 76. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J,
723 Loman NJ, Walker AW. 2014. Reagent and laboratory contamination can critically
724 impact sequence-based microbiome analyses. BMC Biol 12:87.
725 77. Kim Y, Choi KR, Chae MJ, Shin BK, Kim HK, Kim A, Kim BH. 2013. Stability of DNA, RNA,
726 cytomorphology, and immunoantigenicity in Residual ThinPrep Specimens. APMIS
727 121:1064-72.
728 78. Castle PE, Solomon D, Hildesheim A, Herrero R, Concepcion Bratti M, Sherman ME,
729 Cecilia Rodriguez A, Alfaro M, Hutchinson ML, Terence Dunn S, Kuypers J, Schiffman M.
730 2003. Stability of archived liquid-based cervical cytologic specimens. Cancer 99:89-96.

79. Rebolj M, Rask J, van Ballegooijen M, Kirschner B, Rozemeijer K, Bonde J, Rygaard C, Lynge E. 2015. Cervical histology after routine ThinPrep or SurePath liquid-based cytology and computer-assisted reading in Denmark. Br J Cancer 113:1259-74.
80. Naeem RC, Goldstein DY, Einstein MH, Ramos Rivera G, Schlesinger K, Khader SN, Suhrland M, Fox AS. 2017. SurePath Specimens Versus ThinPrep Specimen Types on the COBAS 4800 Platform: High-Risk HPV Status and Cytology Correlation in an Ethnically Diverse Bronx Population. Lab Med 48:207-213.
81. Ritu W, Enqi W, Zheng S, Wang J, Ling Y, Wang Y. 2019. Evaluation of the Associations Between Cervical Microbiota and HPV Infection, Clearance, and Persistence in Cytologically Normal Women. Cancer Prev Res (Phila) 12:43-56.
82. Hosomi K, Ohno H, Murakami H, Natsume-Kitatani Y, Tanisawa K, Hirata S, Suzuki H, Nagatake T, Nishino T, Mizuguchi K, Miyachi M, Kunisawa J. 2017. Method for preparing DNA from feces in guanidine thiocyanate solution affects 16S rRNA-based profiling of human microbiota diversity. Sci Rep 7:4339.
83. Sarangi AN, Goel A, Aggarwal R. 2019. Methods for Studying Gut Microbiota: A Primer for Physicians. J Clin Exp Hepatol 9:62-73.
84. Silhavy TJ, Kahne D, Walker S. 2010. The bacterial cell envelope. Cold Spring Harb Perspect Biol 2:a000414.

Table 1: Characteristics of four different DNA extraction protocols

Kit (Cat. No.)	Manufacturer	Sample volume	Enzyme	Beads	Beating	DNA carrier	Others
ZymoBIOMICS DNA Miniprep Kit (D4300)	Zymo Research	300 µL	No	Ceramic[a]	2 min[b]	No	[c]
QIAamp PowerFecal Pro DNA Kit (51804)	Qiagen	300 µL	No	Ceramic[d]	10 min[b]	No	[c]
QIAamp DNA Mini Kit (51304)	Qiagen	200 µL	Mutanolysin[e]	No	No	No	[c, f, g]
IndiSpin Pathogen Kit (SPS4104)	Indical Bioscience	300 µL	No	Ceramic[h]	10 min[b]	Yes	[c, i]

a: https://www.zymoresearch.com/pages/microbial-isolation. b: Disruptor Genie (USA Scientific, Inc.) was used under the maximum speed. c: Nuclease free water (85 µL) as DNA elution buffer was used. d: PowerBead Pro Tubes (https://www.qiagen.com/us/products/discovery-and-translational-research/lab-essentials/plastics/powerbead-tubes/#orderinginformation). e: Instead of lysozyme or lysostaphin, mutanolysin was used as per Yuan *et al*, 2012 (68). f: DNA Purification from Blood or Body Fluids; Protocols for Bacteria; Isolation of genomic DNA from gram-positive bacteria in QIAamp DNA Mini and Blood Mini Handbook fifth edition was referenced. g: Heating at 56°C for 30 min and 95°C for 15 min was performed. h: Pathogen Lysis Tubes S (https://www.qiagen.com/dk/shop/pcr/pathogen-lysis-tubes/). i: Pretreatment B2 as per QIAamp cador Pathogen Mini Handbook.

751

Table 2. Patients characteristics

Characteristics		Values
Number of patients, n		20
Total number of DNA extracts, n		80
Age, mean (SD)		31.4 (5.0)
Race	African American, n (%)	3 (15)
	Caucasian, n (%)	10 (50)
	Hispanic, n (%)	7 (35)
Cervical biopsy	CIN2, n (%)	8 (40)
	CIN3, n (%)	10 (50)
	Benign, n (%)	2 (10)
HPV typing	HPV positive, n (%)	19 (95)
	HPV16 positive, n (%)	10 (50)
	HPV18 positive, n (%)	2 (10)
	HPV16 or 18 positives, n (%)	10 (50)
	HR-HPV positives, n (%)	18 (90)

SD: standard deviation. CIN: cervical intraepithelial neoplasia. HR-HPV: high-risk HPV (HPV16 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68)

752

Table 3. Reads and OTUs before rarefying assigned to all, gram-, and gram-negative bacteria per DNA extraction protocols

Parameters	Community	Methods	Values	Ratio of GP or GN	p value
Number of reads (mean ± SD)	All	Zy	2,705,044 (135,252 ± 66,011)		[a]
		Pro	2,312,207 (115,610 ± 68,201)		
		QIA	2,765,343 (138,267 ± 49,781)		
		IN	3,366,988 (168,349 ± 57,451)		
	GP	Zy	2,430,380 (121,519 ± 56,209)	89.8%	NS
		Pro	2,116,458 (105,823 ± 57,590)	91.5%	
		QIA	2,503,578 (125,179 ± 46,073)	90.5%	
		IN	2,985,941 (149,297 ± 46,936)	88.7%	
	GN	Zy	274,664 (13,733 ± 29,162)	10.2%	NS
		Pro	195,749 (9,788 ± 23,070)	8.5%	
		QIA	261,765 (13,088 ± 22,638)	9.5%	
		IN	381,047 (19,052 ± 33,038)	11.3%	
Number of OTUs (mean ± SD)	All	Zy	825 (41.3 ± 16.8)		NS
		Pro	621 (31.1 ± 19.4)		
		QIA	778 (38.9 ± 22.4)		
		IN	792 (39.6 ± 22.7)		
	GP	Zy	479 (24.0 ± 9.2)	58.1%	NS
		Pro	412 (20.6 ± 12.7)	66.3%	
		QIA	513 (25.7 ± 13.7)	65.9%	
		IN	531 (26.6 ± 14.9)	67.0%	
	GN	Zy	346 (17.3 ± 9.8)	41.9%	[b]
		Pro	209 (10.5 ± 10.3)	33.7%	
		QIA	265 (13.3 ± 9.2)	34.1%	
		IN	261 (13.1 ± 8.3)	33.0%	

Community of gram-positive bacteria were defined as phylum *Actinobacteria* and *Firmicutes*, which are composed of thick peptidoglycan layers without outer membrane (84). Community of gram-negative bacteria was defined as a community of bacteria other than phylum *Actinobacteria* and *Firmicutes* in this study. a: I - P: 0.0199; I - Q: 0.1590; P - Q: 0.1436; I - Z: 0.1495; P - Z: 0.1712; and Q - Z: 0.4059. b: I - P: 0.2116; I - Q: 0.4837; P - Q: 0.1143; I - Z: 0.0938; P - Z: 0.0116; Q - Z: 0.1448. Dunn's test with Benjamini-Hochberg-adjustment were performed for comparison of the number of read and OTU by DNA extraction method. Zy: ZymoBIOMICS DNA Miniprep Kit, Pro: QIAamp PowerFecal Pro DNA Kit, QIA: QIAamp DNA Mini Kit, IN: IndiSpin Pathogen Kit. SD: standard deviation. All: all bacteria, GP: gram-positive bacteria, GN: gram-negative bacteria. NS: not significant.

753

Table 4. Beta diversity among DNA extraction methods

Index	Protocol	Protocols compared	p values	q values
Aitchison distance	ZymoBIOMICS	PowerFecalPro	NS	NS
		QIAampMini	NS	NS
		IndiSpin	NS	NS
	PowerFecalPro	QIAampMini	NS	NS
		IndiSpin	NS	NS
	QIAampMini	IndiSpin	NS	NS
Unweighted UniFrac distance	ZymoBIOMICS	PowerFecalPro	0.001	0.002
		QIAampMini	0.001	0.002
		IndiSpin	0.001	0.002
	PowerFecalPro	QIAampMini	NS	NS
		IndiSpin	0.015	0.023
	QIAampMini	IndiSpin	NS	NS
Weighted UniFrac distance	ZymoBIOMICS	PowerFecalPro	NS	NS
		QIAampMini	NS	NS
		IndiSpin	NS	NS
	PowerFecalPro	QIAampMini	NS	NS
		IndiSpin	NS	NS
	QIAampMini	IndiSpin	NS	NS
Jaccard distance	ZymoBIOMICS	PowerFecalPro	0.037	NS
		QIAampMini	0.003	0.018
		IndiSpin	0.011	0.033
	PowerFecalPro	QIAampMini	NS	NS
		IndiSpin	NS	NS
	QIAampMini	IndiSpin	NS	NS
Bray-Curtis distance	ZymoBIOMICS	PowerFecalPro	NS	NS

	QIAampMini	NS	NS
	IndiSpin	NS	NS
PowerFecalPro	QIAampMini	NS	NS
	IndiSpin	NS	NS
QIAampMini	IndiSpin	NS	NS

Pairwise PERMANOVA was tested for comparing beta diversity of DNA extraction method. NS: not significant.

754

Table S1. Diversity analysis in this study

No.	Parameter	Alpha or Beta diversity	Used data with/without rarefying	Input data with/without phylogenetic information	Plugin of QIIME 2
1	Species richness	Alpha	Not rarefied	Non-phylogenetic	`q2-breakaway` (50)
2	Faith's Phylogenetic Diversity	Alpha	Rarefied	Phylogenetic	`q2-diversity`
3	Observed OTUs	Alpha	Rarefied	Non-phylogenetic	`q2-diversity`
4	Shannon's diversity index	Alpha	Rarefied	Non-phylogenetic	`q2-diversity`
5	Pielou's Evenness	Alpha	Rarefied	Non-phylogenetic	`q2-diversity`
6	Aitchison distance	Beta	Not rarefied	Non-phylogenetic	`q2-deicode` (51)
7	Unweighted UniFrac distance	Beta	Rarefied	Phylogenetic	`q2-diversity`
8	Weighted UniFrac distance	Beta	Rarefied	Phylogenetic	`q2-diversity`
9	Jaccard distance	Beta	Rarefied	Non-phylogenetic	`q2-diversity`
10	Bray-Curtis distances	Beta	Rarefied	Non-phylogenetic	`q2-diversity`
11	Adonis	Beta	Rarefied	Non-phylogenetic	`q2-diversity adonis` (46) (47)

755

Figure legends

Fig. 1. Overview of the study design using the 16S rRNA gene to compare the DNA extraction protocol for cervical microbiota. (A) Liquid-based cytology (LBC) specimens from 20 patients with CIN2/3 or suspected CIN2/3. (B) A total of 80 DNA extractions were performed. (C) The four DNA extraction methods. (D) DNA of mock vaginal community as a positive control and preservation solution as a negative control. (E) Sequencing using Illumina MiSeq. (F) Analysis of the taxonomic profiles among the DNA extraction protocols. Images form Togo Picture Gallery (http://togotv.dbcls.jp/ja/pics.html) were used to create this figure.

Fig. 2. Taxonomic resolution among DNA extraction protocols. (A) Relative abundance of microbe at family level (left) and genus level (right) per DNA extraction method showed the pattern that variance of microbe composition per patient was higher than that per DNA extraction protocol. These pattern were confirmed by values of Adonis test (`q2-diversity adonis`); F.Model: 199.4, R2: 0.982, and p value: 0.001 for patients and F.Model: 2.9, R2: 0.003, and p value: 0.002 for DNA extraction (46) (47). After all count data of taxonomy were converted to relative abundance as shown in the y-axis, the top ten taxonomy at each family and genus level were plotted in colored bar plot and other relatively few taxonomies were not plotted. The 20 patients ID were described in the x-axis. (B) Venn diagrams showed that ZymoBIOMICS had four unique taxa at family (left) and genus (right) taxonomic level. Thirty-one of 41 families and 45 of 57 genera were detected with all DNA extraction protocols.

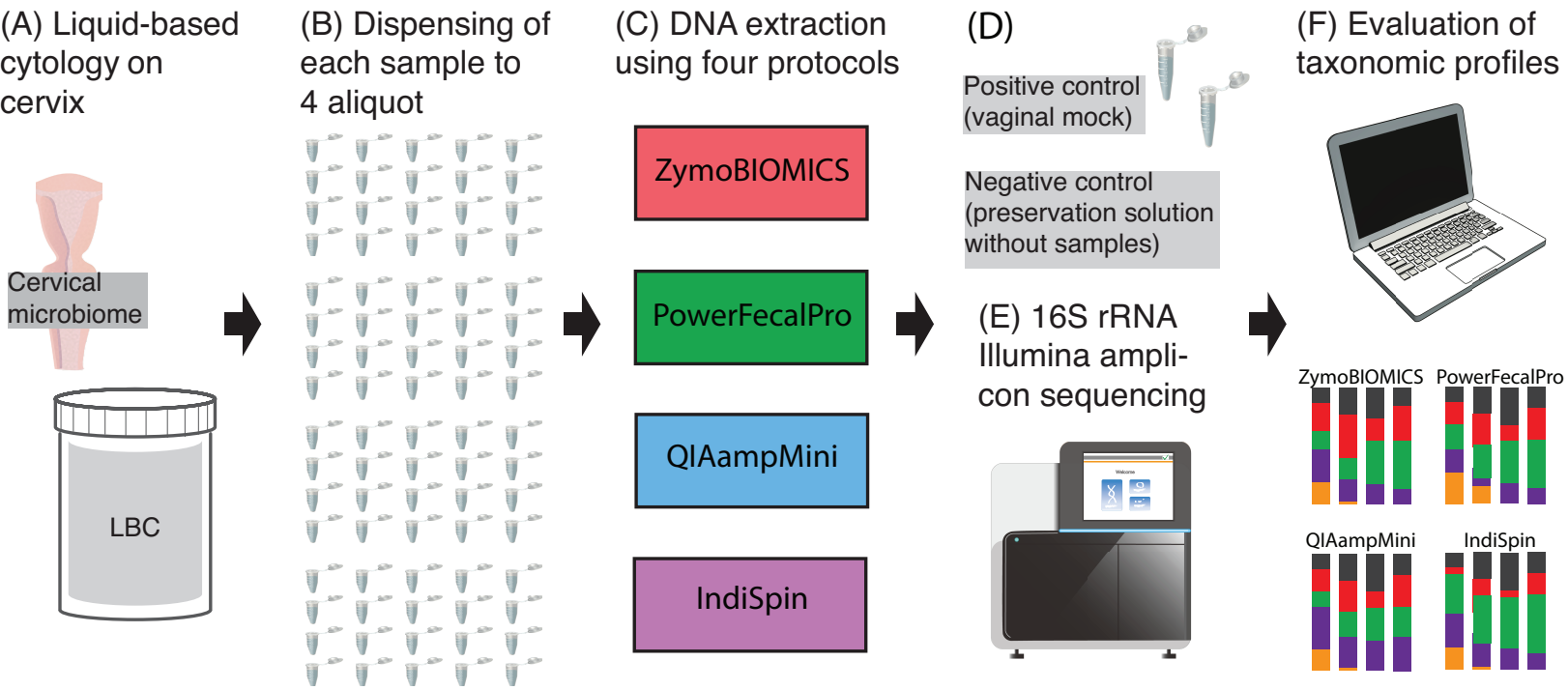
Fig. 3. Comparisons of alpha diversity between different DNA extraction protocols. The alpha diversity indices determined by Species richness and Phylogenetic diversity are

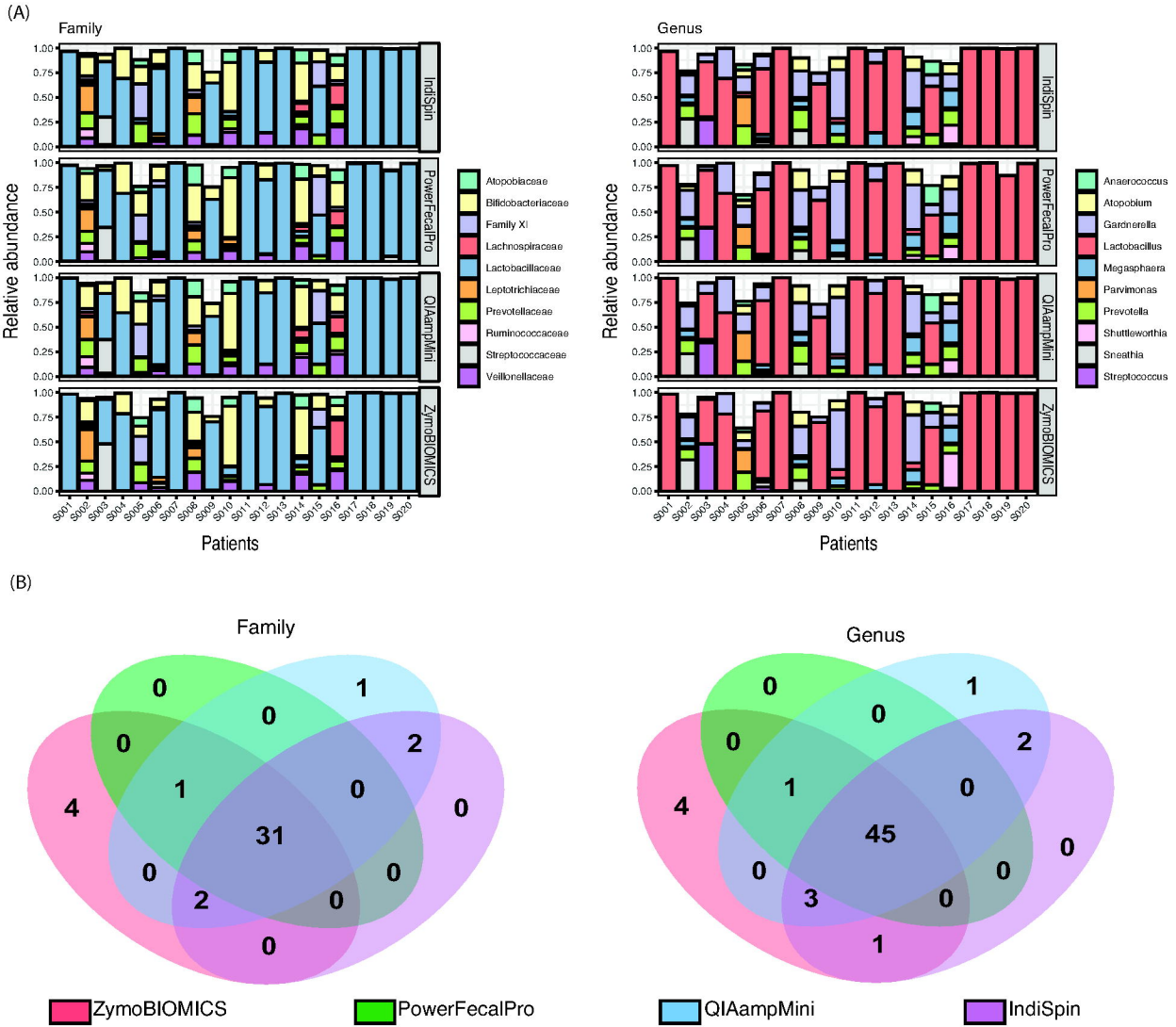
significantly higher with ZymoBIOMICS in comparison with PowerFecalPro ($p = 0.025$ and 0.012, respectively, Dunn's test with Benjamini-Hochberg-adjustment). IndiSpin also showed significantly higher diversity than that of PowerFecalPro using analysis of Species richness ($p = 0.042$, Dunn's test with Benjamini-Hochberg-adjustment). No significant differences were observed in other alpha diversity indexes such as observed OTUs, Shannon's diversity index, and Pielou's Evenness. Zy: ZymoBIOMICS DNA Miniprep Kit, Pro: QIAamp PowerFecal Pro DNA Kit, QIA: QIAamp DNA Mini Kit, IN: IndiSpin Pathogen Kit.

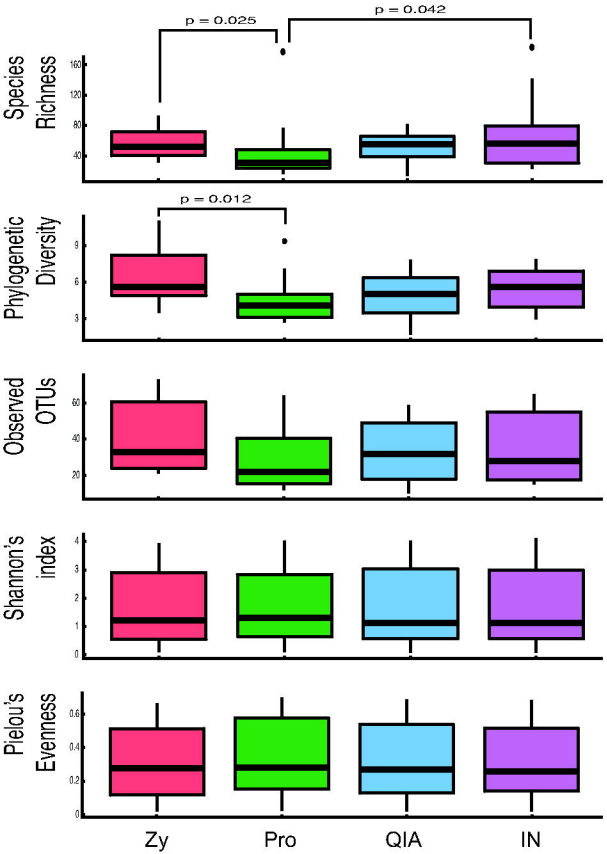
Fig. 4. Distinct detections of microbe among the DNA extraction protocols. (A) A bar graph showing 23 significantly enriched taxa with ZymoBIOMICS, 3 with QIAamp DNA Mini Kit, and 3 with IndiSpin Pathogen Kit determined by the linear discriminant analysis (LDA) effect size (LEfSe) analyses (48). (B) A taxonomic cladogram from the same LEfSe analyses showing that the significantly enriched microbiota in ZymoBIOMICS were composed of phylum *Proteobacteria*. Also note that *Meiothermus (a member of* the phylum *Deinococcus-Thermus)* is likely an extraction kit contaminant. Zy: ZymoBIOMICS DNA Miniprep Kit, Pro: QIAamp PowerFecal Pro DNA Kit, QIA: QIAamp DNA Mini Kit, IN: IndiSpin Pathogen Kit. g_: genus, f_: family, o_: order, c_: class, p_: phylum.

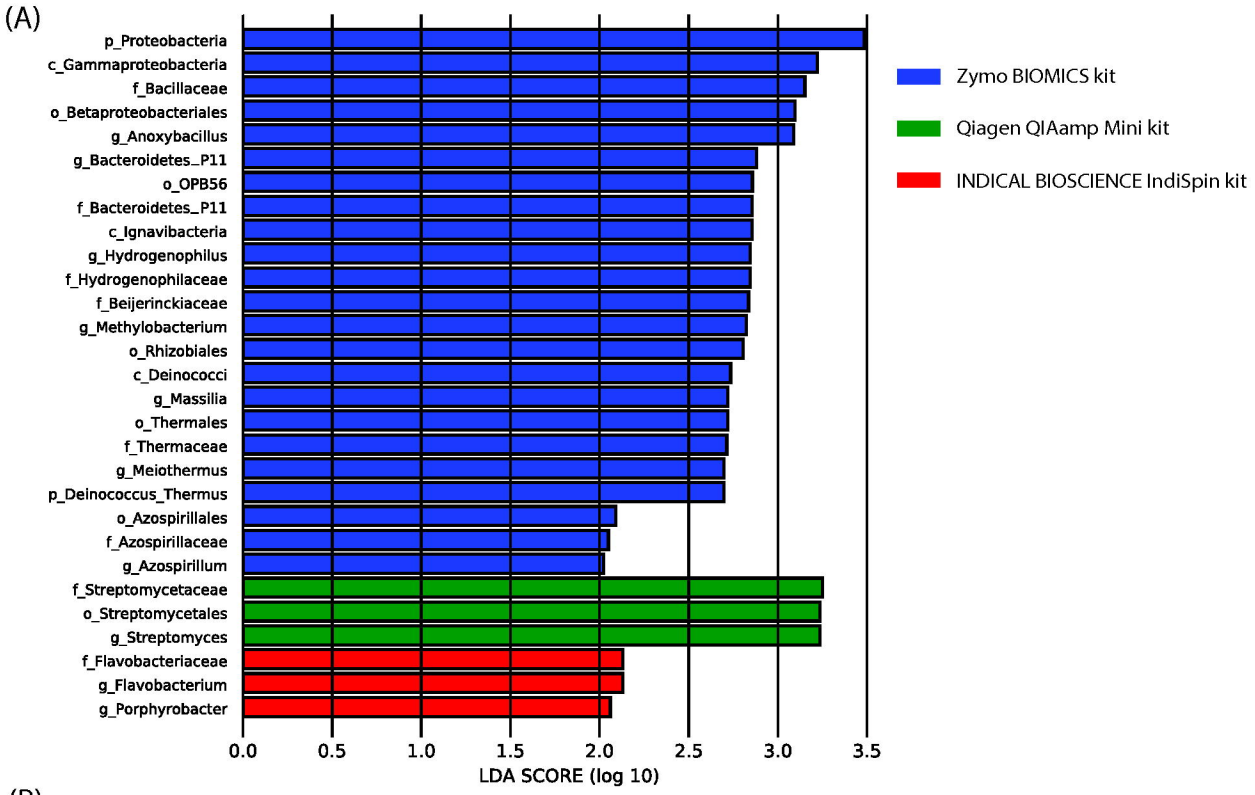
Fig. S1. Comparison of DNA yields by DNA extraction protocols. DNA yield of QIAampMini was significantly higher than that of PowerFecalPro ($p < 0.001$, Dunn's test with Benjamini-Hochberg-adjustment). Also, the DNA yield of ZymoBIOMICS was significantly higher than that of PowerFecalPro ($p < 0.001$, Dunn's test with Benjamini-Hochberg-adjustment). The amount of DNA was calculated based on the absorbance of nucleic acids measured by

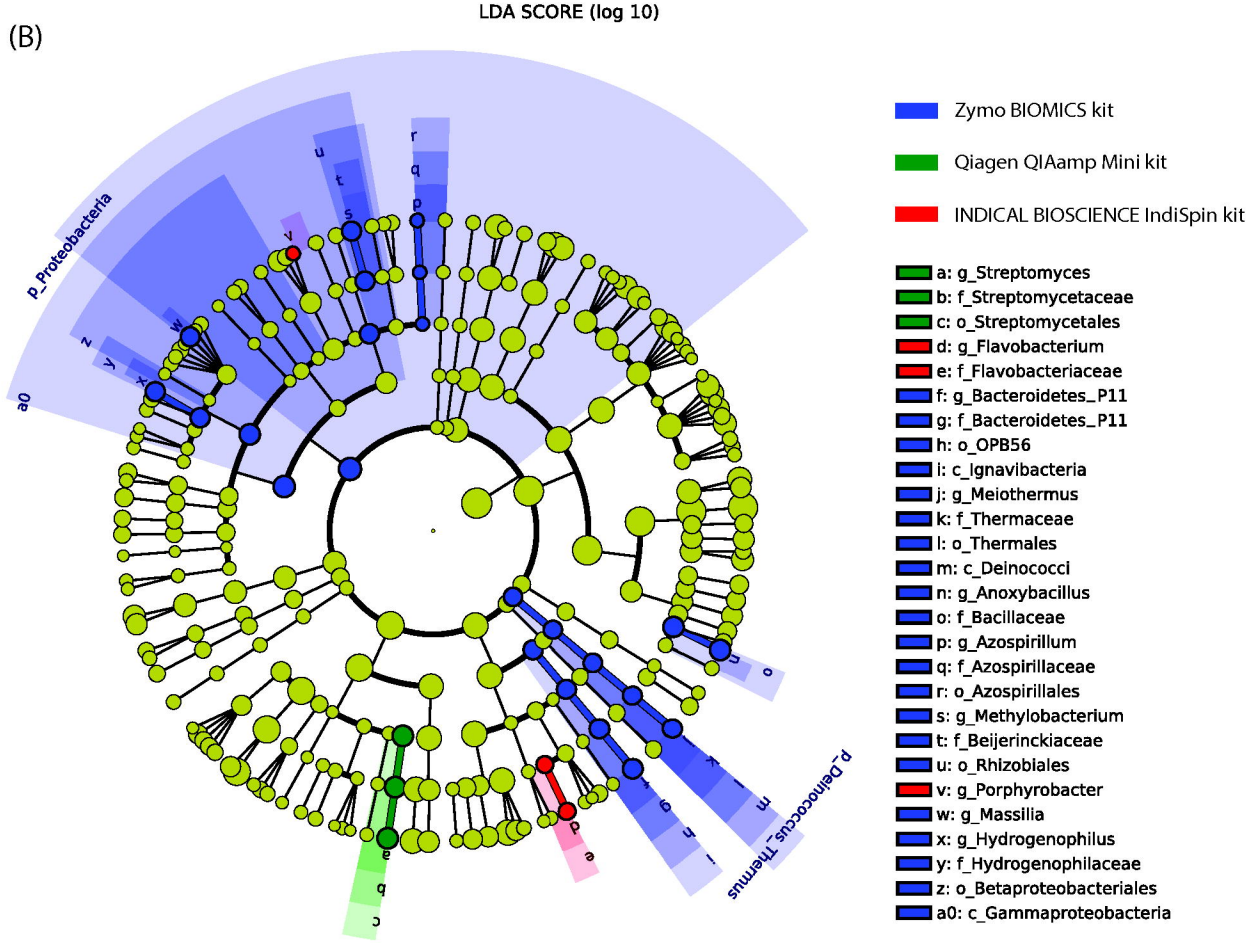
Nanodrop One. By the protocol recommended by the manufacturer, nucleic acid (Poly-A carrier) was used in IndiSpin. Therefore, IndiSpin was excluded from the analysis of DNA yield. The amount of DNA yield per 100 μL ThinPrep sample volume were compared. The bar graph shows the mean and standard deviation. Zy: ZymoBIOMICS DNA Miniprep Kit, Pro: QIAamp PowerFecal Pro DNA Kit, QIA: QIAamp DNA Mini Kit.


(A) Liquid-based cytology on cervix
Cervical microbiome
LBC
(B) Dispensing of each sample to 4 aliquot
(C) DNA extraction using four protocols
ZymoBIOMICS
PowerFecalPro
QIAampMini
IndiSpin
(D)
Positive control (vaginal mock)
Negative control (preservation solution without samples)
(E) 16S rRNA Illumina ampli-con sequencing
(F) Evaluation of taxonomic profiles
ZymoBIOMICS
PowerFecalPro
QIAampMini
IndiSpin


(A)
Family
Genus
Relative abundance
Patients
IndiSpin
PowerFecalPro
QIAampMini
ZymoBIOMICS
Atopobiaceae
Bifidobacteriaceae
Family XI
Lachnospiraceae
Lactobacillaceae
Leptotrichiaceae
Prevotellaceae
Ruminococcaceae
Streptococcaceae
Veillonellaceae
Anaerococcus
Atopobium
Gardnerella
Lactobacillus
Megasphaera
Parvimonas
Prevotella
Shuttleworthia
Sneathia
Streptococcus
(B)
Family
Genus
ZymoBIOMICS
PowerFecalPro
QIAampMini
IndiSpin


p = 0.025
p = 0.042
Species Richness
160
120
80
40
p = 0.012
Phylogenetic Diversity
9
6
3
Observed OTUs
60
40
20
Shannon's index
4
3
2
1
0
Pielou's Evenness
0.6
0.4
0.2
0
Zy
Pro
QIA
IN


(A)
p_Proteobacteria
c_Gammaproteobacteria
f_Bacillaceae
o_Betaproteobacteriales
g_Anoxybacillus
g_Bacteroidetes_P11
o_OPB56
f_Bacteroidetes_P11
c_Ignavibacteria
g_Hydrogenophilus
f_Hydrogenophilaceae
f_Beijerinckiaceae
g_Methylobacterium
o_Rhizobiales
c_Deinococci
g_Massilia
o_Thermales
f_Thermaceae
g_Meiothermus
p_Deinococcus_Thermus
o_Azospirillales
f_Azospirillaceae
g_Azospirillum
f_Streptomycetaceae
o_Streptomycetales
g_Streptomyces
f_Flavobacteriaceae
g_Flavobacterium
g_Porphyrobacter
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
LDA SCORE (log 10)
Zymo BIOMICS kit
Qiagen QIAamp Mini kit
INDICAL BIOSCIENCE IndiSpin kit


(B)
p_Proteobacteria
p_Deinococcus_Thermus
Zymo BIOMICS kit
Qiagen QIAamp Mini kit
INDICAL BIOSCIENCE IndiSpin kit
a: g_Streptomyces
b: f_Streptomycetaceae
c: o_Streptomycetales
d: g_Flavobacterium
e: f_Flavobacteriaceae
f: g_Bacteroidetes_P11
g: f_Bacteroidetes_P11
h: o_OPB56
i: c_Ignavibacteria
j: g_Meiothermus
k: f_Thermaceae
l: o_Thermales
m: c_Deinococci
n: g_Anoxybacillus
o: f_Bacillaceae
p: g_Azospirillum
q: f_Azospirillaceae
r: o_Azospirillales
s: g_Methylobacterium
t: f_Beijerinckiaceae
u: o_Rhizobiales
v: g_Porphyrobacter
w: g_Massilia
x: g_Hydrogenophilus
y: f_Hydrogenophilaceae
z: o_Betaproteobacteriales
a0: c_Gammaproteobacteria