

RESEARCH ARTICLE

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

Takeo Shibata^{1,2}, Mayumi Nakagawa¹, Hannah N. Coleman¹, Sarah M. Owens³, William W. Greenfield⁴, Toshiyuki Sasagawa², Michael S. Robeson, II^{5*}

1 Department of Pathology, University of Arkansas for Medical Sciences, Little Rock, AR, United States of America, **2** Department of Obstetrics and Gynecology, Kanazawa Medical University, Uchinada, Ishikawa, Japan, **3** Biosciences Division, Argonne National Laboratory, Lemont, IL, United States of America, **4** Department of Obstetrics and Gynecology, University of Arkansas for Medical Sciences, Little Rock, AR, United States of America, **5** Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR, United States of America

* MRobeson@uams.edu



OPEN ACCESS

Citation: Shibata T, Nakagawa M, Coleman HN, Owens SM, Greenfield WW, Sasagawa T, et al. (2021) Evaluation of DNA extraction protocols from liquid-based cytology specimens for studying cervical microbiota. PLoS ONE 16(8): e0237556. <https://doi.org/10.1371/journal.pone.0237556>

Editor: Brenda A. Wilson, University of Illinois Urbana-Champaign, UNITED STATES

Received: August 10, 2020

Accepted: August 12, 2021

Published: August 30, 2021

Copyright: © 2021 Shibata et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: MIMARKS compliant ([10.1038/nbt.1823](https://doi.org/10.1038/nbt.1823)) DNA sequencing data are available via the Sequence Read Archive (SRA) at the National Center for Biotechnology Information (NCBI), under the BioProject Accession: PRJNA598197.

Funding: This work was funded by the National Institutes of Health in the form of a grant (R01CA143130, USA), Drs. Mae and Anderson Nettleship Endowed Chair of Oncologic Pathology in the form of a grant (31005156, USA), and the

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. We sought to determine whether it is possible to access microbial DNA from LBC specimens, and compared the performance of four different extraction protocols: (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 CIN values of 2 or 3 were each aliquoted for each of the four kits. Loss of microbial diversity due to long-term LBC storage could not be assessed due to lack of fresh LBC samples. Comparisons with other types of cervical sampling were not performed. We observed that all DNA extraction kits provided equivalent accessibility to the cervical microbial DNA within stored LBC samples. Approximately 80% microbial genera were shared among all DNA extraction protocols. Potential kit contaminants were observed as well. Variation between individuals was a significantly greater influence on the observed microbial composition than was the method of DNA extraction. We also observed that HPV16 was significantly associated with community types that were not dominated by *Lactobacillus iners*.

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 [1–5] and HPV-related diseases [6–10]. Cervicovaginal microbiota are considered to be an important factor affecting the progress of cervical

Arkansas Biosciences Institute (the major component of the Tobacco Settlement Proceeds Act of 2000, G1-52249-01, USA) awarded to M.N.

Competing interests: The authors would like to declare the following patents/patent applications associated with this research: M.N. is one of the inventors named in the patents and patent applications for the HPV therapeutic vaccine PepCan. Patent issued: Human Papilloma Virus Therapeutic Vaccine Nakagawa, M. and Chang, B. S. Patent No. 9,974,849 issued on 5/22/2018 Patent application: Human Papilloma Virus Therapeutic Vaccine Nakagawa, M. and Chang, B. S. International Application (PCT/US14/60198) filed on 10/11/2014. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

intraepithelial neoplasia (CIN) [6–9] and are implicated in the persistence of high-risk HPV (HR-HPV) [1, 2] and low-risk HPV (LR-HPV) [3]. However, microbial signatures associated with either HR-HPV or LR HPV can vary depending on the population under study, e.g., the phyla *Actinobacteria* and *Fusobacteria* were found to be enriched in HR-HPV positive Chinese women [4] while another study observed these groups associated with low-risk HPV (LR-HPV) in South African women [3]. Additionally, *Lactobacillus iners*-dominant samples are associated with both HR-HPV and LR-HPV [5], often associated with moderate CIN risk [10]. Moreover, it has been shown that CIN risk was increased in patients with HR-HPV [10] when the cervical microbes *Atopobium vaginae*, *Gardnerella vaginalis*, and *Lactobacillus iners* were present in greater proportion compared to *L. crispatus*. The cervicovaginal microbiome is often described by the abundance of *Lactobacillus spp.*, i.e. the community is either referred to as a *Lactobacillus*-dominant type or non-*Lactobacillus*-dominant type, and can interact with the immune system in different ways [7, 11]. For example, inflammatory cytokines, such as Interleukin (IL)-1 α 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 dendritic 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]. Prior research, on the importance of the microbiome in cancer therapeutics via checkpoint inhibitors [13], along with our own work on the role of Cervical microbiota (CM) in vaccine response [14], suggests that the CM has a significant role to play in disease progression and therapeutic treatment. We continue our work here to further assess the use of liquid based cytology (LBC) samples in surveying microbial community DNA.

Little has been reported on the utility of LBC samples for use in cervical microbiome studies. Conventionally, microbiome sample collection methods entail the use of swabs [15] or self-collection of vaginal discharge [16]. 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* [15, 17–20].

LBC samples are promising for cervicovaginal microbiome surveys, as they are an already established method of long-term cytological biobanking [21]. 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 [22–24]. Despite the promising potential to use LBC samples for surveying cervicovaginal microbiota, it is known that DNA contained within LBC samples may degrade over prolonged storage times when kept at ambient or non-freezing temperatures [25, 26]. Although others have shown minimal DNA degradation of LBC samples stored at -80 °C [26], the ability to reliably access microbial DNA remains to be seen, and is the focus of our study.

Furthermore, the ability to characterize these microbiota, as commonly assessed by 16S rRNA gene sequencing, can be biased as a result of methodological differences of cell lysis and DNA extraction protocols [27–29]. Herein, we compare four different commercially available DNA extraction kits in an effort to assess their ability to extract and characterize any viable microbial DNA from stored LBC samples. Additionally, we examine the relationship between HPV infection and the composition of cervical microbiota still accessible after prolonged LBC storage.

Methods

Recruiting patients

Patients were participants of a single center randomized double blind Phase II clinical trial (NCT02481414) in which enrollees were assigned to receive an HPV therapeutic vaccine called PepCan or an adjuvant derived from *Candida albicans* (Candin[®], Nielsen BioSciences, San Diego, CA). Pre-injection liquid based cervical cytology (ThinPrep) samples from 20 consecutive enrollees who gave written informed consent between 3/21/2017 and 12/11/2017 were used for this study. Patients were recruited mainly through referrals from clinics from inside and outside of the medical center. Flyers and Google advertisements were also utilized. Inclusion Criteria: aged 18–50 years, had recent (≤ 60 days) Pap smear result consistent with high grade squamous intraepithelial lesion (HSIL) or “cannot rule out HSIL” or HSIL on colposcopy guided biopsy, untreated for HSIL or “Cannot rule out HSIL”, able to provide informed consent, willingness and able to comply with the requirements of the protocol. Exclusion Criteria: history of disease or treatment causing immunosuppression (e.g., cancer, HIV, organ transplant, autoimmune disease), being pregnant or attempting to be pregnant within the period of study participation, breast feeding or planning to breast feed within the period of study participation, allergy to Candida antigen, history of severe asthma requiring emergency room visit or hospitalization within the past 5 years, history of invasive squamous cell carcinoma of the cervix, history of having received PepCan. 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. All collected samples are representative of a larger population in gynecology clinics with abnormal Pap tests. If in the opinion of the Principal Investigator or other Investigators, it is not in the best interest of the patient to enter this study, the patient was excluded. Patients’ age, race, and ethnicity were recorded based on standard NIH requirements. All categories and definitions, e.g. ethnicity, age, etc., were based on NIH Guidelines.

Sampling of cervical microbiome

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 [14]. The specimens were frozen in an ultra-low temperature freezer (-80°C) on the day of collection. The storage period from sample collection to DNA extraction was 716 ± 105 days.

HPV genotyping

HPV-DNA was detected by Linear Array HPV Genotyping Test (Roche Diagnostics) which can detect up to 37 HPV genotypes using ThinPrep solution [30]. HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68 were defined as HR-HPV genotypes; and HPV6, 11, 40, 42, 54, 61, 62, 71, 72, 81, 83, 84, and CP6108 were defined as LR-HPV genotypes [31, 32].

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, 21, 33]. We’ll subsequently refer to each of these kits in abbreviated form as follows: ZymoBIOMICS, PowerFecalPro, QIAamp-Mini, and IndiSpin. The protocols and any modifications are outlined in Table 1.

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 μ L	No	Ceramic ^a	2 min ^b	No	c, j
QIAamp PowerFecal Pro DNA Kit (51804)	Qiagen	300 μ L	No	Ceramic ^d	10 min ^b	No	c, j
QIAamp DNA Mini Kit (51304)	Qiagen	200 μ L	Mutanolysin ^e	No	No	No	c, f, g, j
IndiSpin Pathogen Kit (SPS4104)	Indical Bioscience	300 μ L	No	Ceramic ^h	10 min ^b	Yes	c, i, j

a: [34].

b: Disruptor Genie (USA Scientific, Inc.) was used under the maximum speed.

c: Nuclease free water (85 μ L) as DNA elution buffer was used.

d: PowerBead Pro Tubes [35].

e: Instead of lysozyme or lysostaphin, mutanolysin was used as per Yuan *et al*, 2012 [36].

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 [37].

i: Pretreatment B2 as per QIAamp cador Pathogen Mini Handbook.

j: Samples were thawed at room temperature and immediately extracted at the same temperature following the respective kit instructions.

<https://doi.org/10.1371/journal.pone.0237556.t001>

Each LBC sample was dispensed into four separate 2 mL sterile collection tubes (dispensed sample volume = 500 μ L) 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 \times 20 patients) were prepared for subsequent analyses. Applied sample volume of ThinPrep solution was 300 μ L for ZymoBIOMICS, 300 μ L for PowerFecalPro, 200 μ L for QIAampMini, and 300 μ L for IndiSpin. The sample volume was standardized to 300 μ L 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

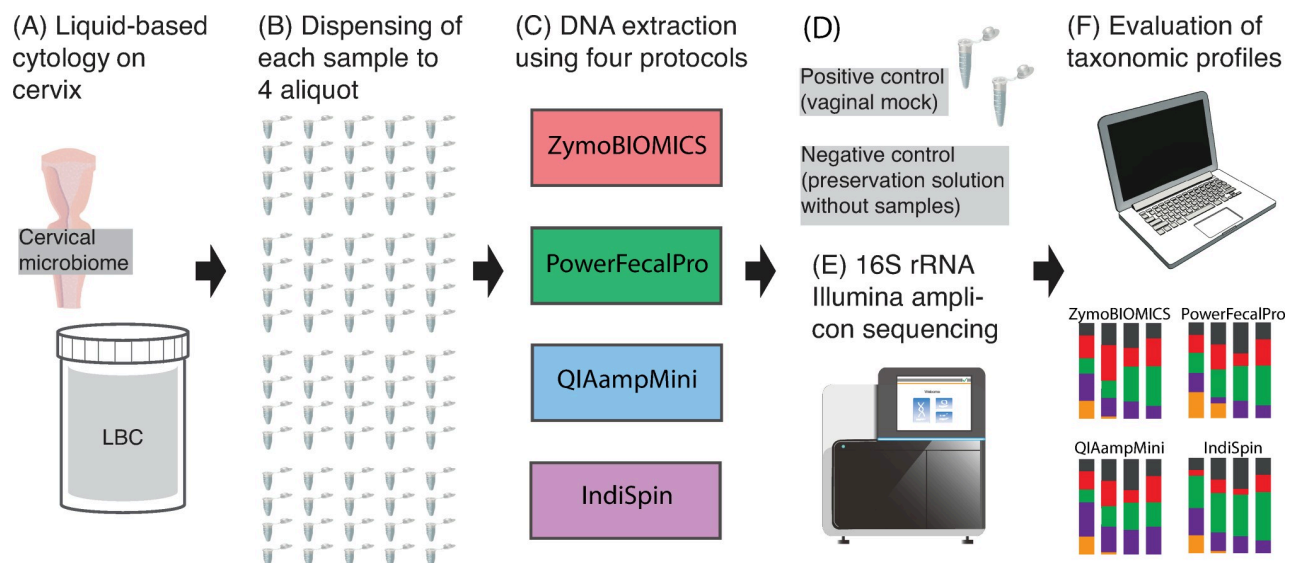


Fig 1. Overview of the study design using 16S rRNA gene to compare the DNA extraction protocol. (A) Liquid-based cytology (LBC) specimens from 20 patients with CIN2/3 or suspected CIN2/3. (B) A total of 80 DNA extractions were performed. (C) The four DNA extraction methods. (D) DNA of mock vaginal community 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 from Togo Picture Gallery [39] were used to create this figure.

<https://doi.org/10.1371/journal.pone.0237556.g001>

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 [38].

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 μ L 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 [40]. The same volume of DNA was used for each reaction, and then normalized at the PCR pooling step. This ensures that equal amounts of each amplified sample are added to the sequencing pool. Paired-end reads from libraries with ~250-bp inserts were generated for the V4 region using the barcoded primer set: 515FB: 5' -GTGYCAGCMGCCGCGTAA-3' and 806RB: 5' -GGACTACNVGGGTWTCTAAT-3' [41–45]. MiSeq Reagent Kit v2 (2 \times 150 cycles, MS-102-2002) was used.

Sequence processing and analysis

Initial sequence processing and analyses were performed using QIIME 2 [46], 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 [47], to determine the appropriate trimming and truncation parameters for generating Exact Sequence Variants (ESVs) [48] via `q2-dada2` [49]. ESVs will be referred to as Operational Taxonomic Units (OTUs). 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 [50]. This enables more robust taxonomic assignment of the OTUs [51]. 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 [52]. Then `q2-alignment` was used to generate a *de novo* alignment with MAFFT [53] 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 [54] with automatic model selection using ModelFinder [55]. Unless specified, subsequent analyses were performed after removing OTUs with a very low frequency of less than 0.0005% of the total data set in this case [56].

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.

Overall microbial composition was investigated at the family and genus 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 [57–59]. Variation of microbiome composition per DNA extraction method or per individual was assessed by the Adonis test (`q2-diversity adonis`) [60, 61].

We set out to determine which microbial taxonomic groups were differentially accessible across the sampling protocols by LEfSe analyses [62]. We further assessed the microbial taxa using `jvenn` [63] at family and genus level. The Venn diagram was created after removing OTUs with a frequency of less than 0.005% [56]. We used Scheffe test [64] to perform post-hoc analysis of the LEfSe output.

Analytical approaches (at the OTU-level) that do not require the rarefying of data, such as `q2-breakaway` [65] and Aitchison distance using `q2-deicode` [66] were used to determine both alpha (within-sample) and beta (between-sample) diversity respectively. These were compared with traditional methods, that often require the data to be rarefied. Here we applied the following traditional alpha and beta-diversity metrics: 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` [46]. In order to maintain a reasonable balance between sequencing depth and sample size, we determined that a rarefaction depth of 51,197 reads allowed us to retain data for all four kits for 15 of the 20 individual patients. Overall, our subsequence analysis consisted of 3,071,820 reads (27.6%, 3,071,820 / 11,149,582 reads). All diversity measurements used in this study are listed in [S1 Table](#).

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 [14, 67–69]. We employed the DMM [70] model to determine the number of community types for bacterial cervical microbiome. Then, we clustered samples to the community type [9, 71]. Since vaginal microbiota were reported to be clustered with different *Lactobacillus sp.* such as *L. crispatus*, *L. gasseri*, *L. iners*, or *L. jensenii* [18, 72], we also collapsed the taxonomy to the species level and performed a clustering analysis using “microbiome” R package [59]. We then determined which bacterial taxa were differentially abundant among the patients with or without HPV16 via `q2-aldex2` [73] and LEfSe [62].

General statistical analysis

All data are presented as means \pm standard deviation (SD). Comparisons were conducted with Fisher's exact test or Dunn's test with Benjamini-Hochberg-adjustment [74] 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. We did not control for confounding variables such as socioeconomic status, nutrition, environmental exposures, or similar factors.

Ethics approval and consent to participate

This study was approved by the Institutional Review Board at the University of Arkansas for Medical Sciences (IRB # 202790). No minors were enrolled in this study.

Consent for publication

Written informed consent for publication was obtained for all patients under IRB # 202790; NCT # NCT02481414; IND # 15173.

Availability of data and materials

MIMARKS compliant [75] 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

Patient characteristics

The age of the patients ($n = 20$) was 31.4 ± 5.0 years. The distribution of race was 15% African American ($n = 3$), 50% European descent ($n = 10$), and 35% Hispanic ($n = 7$). Cervical histology was 40% CIN2 ($n = 8$), 50% CIN3 ($n = 10$), and 10% benign ($n = 2$). HPV genotypes were 50% HPV16 positive ($n = 10$), 10% HPV18 positive ($n = 2$), 90% HR-HPV positives ($n = 18$), 45% LR-HPV positives ($n = 9$), and 75% multiple HPV positives ($n = 15$). Patient characteristics were summarized in Table 2.

DNA yield

DNA yield per 100 μL ThinPrep solution was $0.09 \pm 0.06 \mu\text{g}$ in ZymoBIOMICS, $0.04 \pm 0.01 \mu\text{g}$ in PowerFecalPro, and $0.21 \pm 0.23 \mu\text{g}$ 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 (S1 Fig).

Table 2. Patient 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)
	European descent, 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)
	LR-HPV positives, n (%)	9 (45)
	Multiple HPV infections	15 (75)

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). LR-HPV: low-risk HPV (6, 11, 40, 42, 54, 61, 62, 71, 72, 81, 83, 84, and CP 6108).

<https://doi.org/10.1371/journal.pone.0237556.t002>

Mock and negative controls

We observed that we were able to reasonably recover the expected taxa of our mock community positive control from the American Type Culture Collection (ATCC MSA1007). Each of the following taxa should have a relative abundance of ~16.7% of the total sample. It should be noted that factors, such as sample preparation and primer biases, can cause deviations from the expected mock community [76–78]. We observed: 7.184% *Gardnerella* spp., 14.807% *Lactobacillus jensenii* (40.185% *Lactobacillus* spp.), 16.530% *Mycoplasma hominis*, 14.311% *Prevotella bivia* (14.327% *Prevotella* spp.), and 21.429% *Streptococcus agalactiae* (21.449% *Streptococcus* spp.). A total of 127,193 reads were generated from the mock community control. Of which, 99.68% (126,783 / 127,193 reads) were from expected members of the mock community. For the negative control (ThinPrep preservation solution) only 1,791 reads were generated. 1,400 reads were from *Staphylococcus* spp., 323 reads were from *Micrococcus* spp., and 47 reads were *Lactobacillus* spp. The remaining 21 reads were spurious.

Number of reads and Operational Taxonomic Units (OTUs) before rarefying

We obtained a total of 11,149,582 reads for 80 DNA extractions. 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 & S4 Fig). Variation between patients was a significantly greater influence on the observed microbial composition than was the method of DNA extraction (F.Model: 199.4, R²: 0.982, and p value: 0.001 for patients vs F.Model: 2.9, R²: 0.003, and p value: 0.002 for DNA extraction, Adonis test, 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.

Table 3. Reads and OTUs before rarefying assigned to all, gram-positive, 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%	b	
		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%	cc	
		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)		dc
			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%	e	
		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%	f	
		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 was defined as phylum *Actinobacteria* and *Firmicutes*, which are composed of thick peptidoglycan layers without outer membrane [79]. Community of gram-negative bacteria was defined as a community of bacteria other than phylum *Actinobacteria* and *Firmicutes* in this study.

a: IN—PRO: 0.0199; IN—QIA: 0.1590; PRO—QIA: 0.1436; IN—ZY: 0.1495; PRO—ZY: 0.1712; and QIA—ZY: 0.4059.

b: IN—PRO: 0.0361; IN—QIA: 0.1435; PRO—QIA: 0.2210; IN—ZY: 0.1717; PRO—ZY: 0.2108; QIA—ZY: 0.4540.

c: IN—PRO: 0.5539; IN—QIA: 0.3862; PRO—QIA: 0.4572; IN—ZY: 0.5293; PRO—ZY: 0.9967; QIA—ZY: 0.7573.

d: IN—PRO: 0.3101; IN—QIA: 0.4539; PRO—QIA: 0.2514; IN—ZY: 0.2567; PRO—ZY: 0.1194; QIA—ZY: 0.2726.

e: IN—PRO: 0.2715; IN—QIA: 0.6809; PRO—QIA: 0.1148; IN—ZY: 0.5463; PRO—ZY: 0.1710; QIA—ZY: 0.4986.

f: IN—PRO: 0.2116; IN—QIA: 0.4837; PRO—QIA: 0.1143; IN—ZY: 0.0938; PRO—ZY: 0.0116; QIA—ZY: 0.1448. Dunn's test with Benjamini-Hochberg-adjustment were performed for comparison of the number of reads and OTUs 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.

<https://doi.org/10.1371/journal.pone.0237556.t003>

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, *Methylobacterium* was detected in 5 of the 80 DNA extractions, consisting of 912 reads: 0.01% of all kit extractions. 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 [80]. *Bacteroidetes*, which are often reported as enriched taxa in an HIV positive cervical environment [81], 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%) and *Hydrogenophilus* was detected in 14 of the 80 DNA extractions (2,488 reads, 0.02%). *Meiothermus* and *Hydrogenophilus* [82] are not considered to reside within the human environment, and are likely kit contaminants, as previously reported [83]. A unique gram-positive taxon

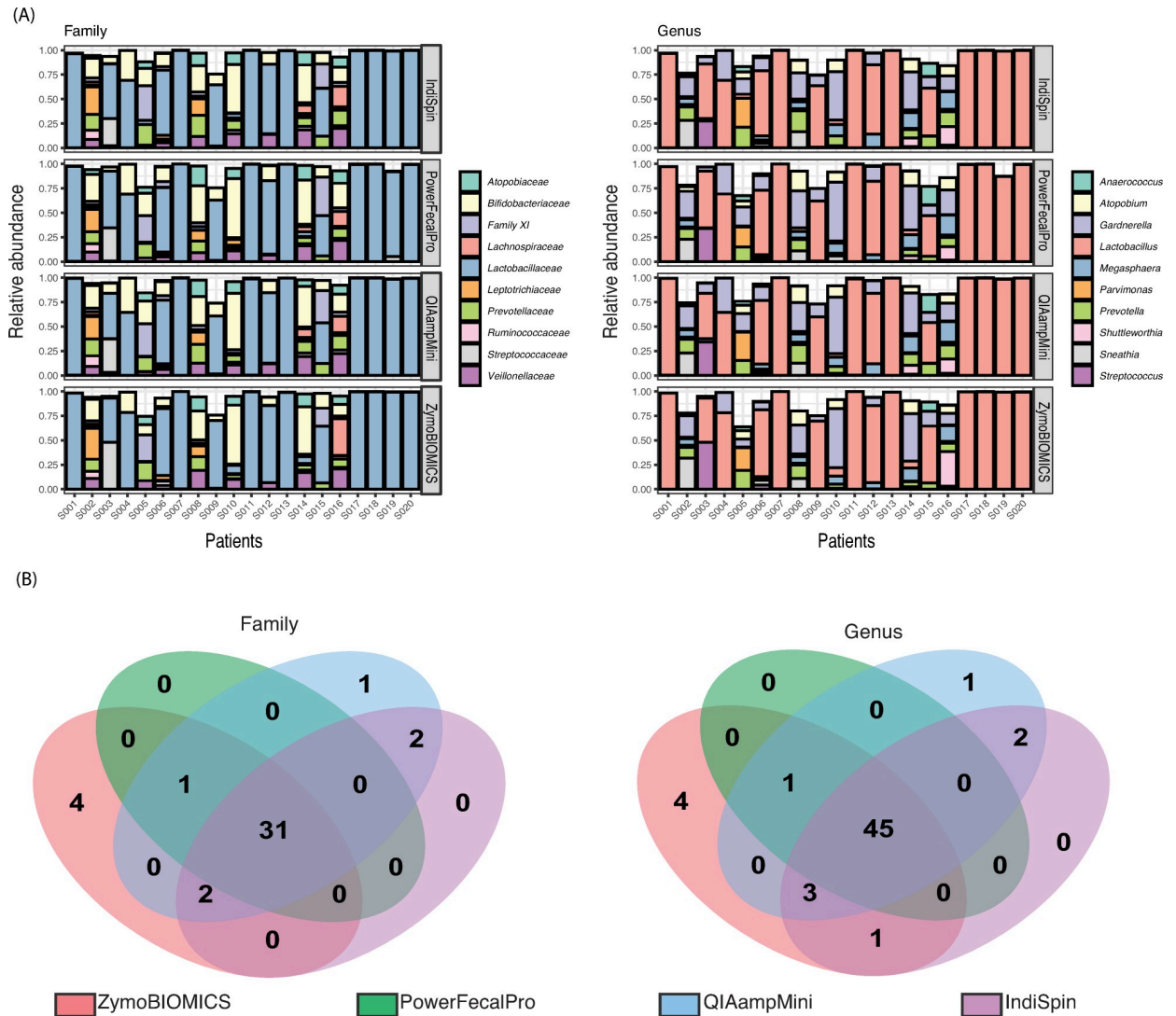


Fig 2. Taxonomic resolution among DNA extraction protocols. (A) Relative abundance of microbe at family level (left) and genus level (right) per DNA extraction method showed the pattern that variance of microbe composition per patient was higher than that per DNA extraction protocol. These patterns were confirmed by values of Adonis test (q_2 -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 [60, 61]. 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, considering only those OTUs with a frequency greater than 0.005% shown, revealed 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.

<https://doi.org/10.1371/journal.pone.0237556.g002>

obtained from the QIAampMini, *Streptomyces*, which was reported to be detected from the cervicovaginal environment in the study of Kenyan women [84], was detected in all 20 of the QIAampMini DNA extractions (6,862 reads; 0.06%). No unique taxa were detected in PowerFecalPro and IndiSpin. Although less than 0.005% of the total data set, two samples of IndiSpin also detected potential kit contaminant, *Tepidiphilus* (*Hydrogenophilaceae*).

Alpha and beta diversity

The observed alpha diversity was similar across all kits except for a few cases (Fig 3). Significantly higher species richness (q_2 -breakaway) was observed between the ZymoBIOMICS

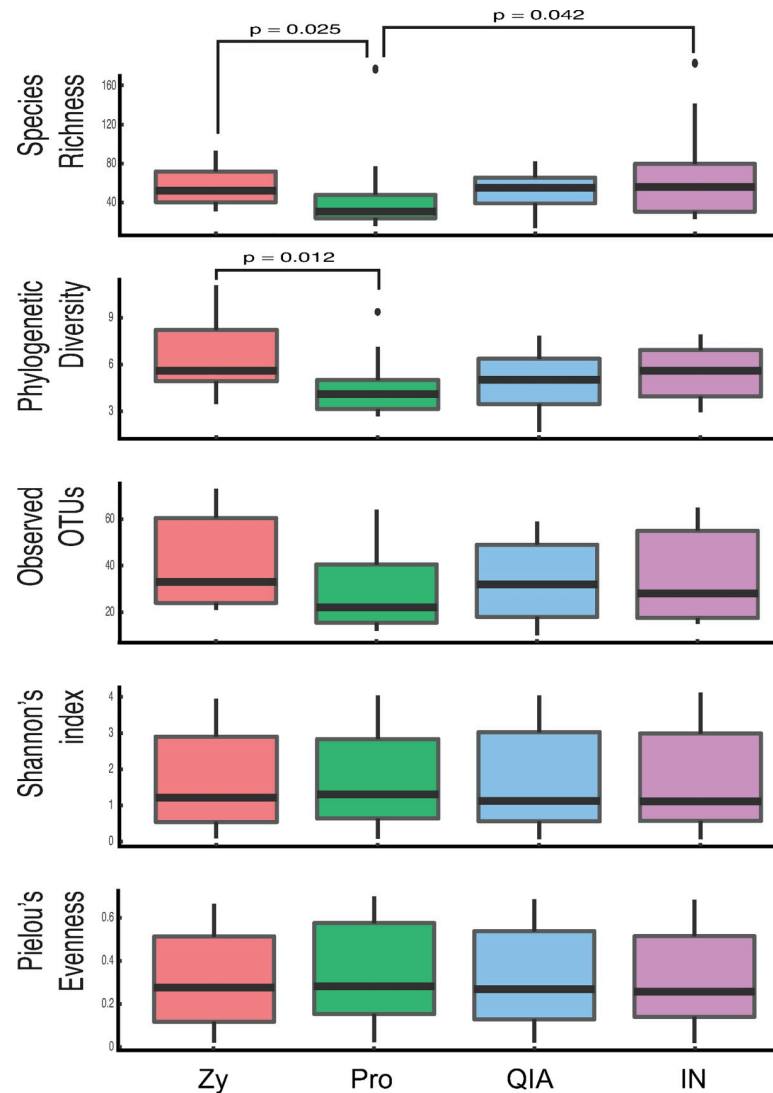


Fig 3. Comparisons of alpha diversity between different DNA extraction protocols. The alpha diversity indices determined by Species richness and Phylogenetic diversity are significantly higher with ZymoBIOMICS in comparison with PowerFecalPro ($p = 0.025$ and 0.012 , respectively, Dunn's test with Benjamini-Hochberg-adjustment). 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.

<https://doi.org/10.1371/journal.pone.0237556.g003>

(56.1 ± 19.4) protocol and that of PowerFecalPro (43.2 ± 32.9 , $p = 0.025$) (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 , $p = 0.012$). The use of IndiSpin also resulted significantly higher alpha diversity than that of PowerFecalPro in an analysis of Species richness ($p = 0.042$). 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.

Similar to the alpha diversity results above, no significant differences were observed with other metrics, including q_2 -deicode (Aichison distances). Only qualitative metrics such as Unweighted UniFrac and Jaccard distance, revealed significant differences in a few cases (Table 4 and S2 and S3 Figs). Most observed differences were between ZymoBIOMICS and

Table 4. Beta diversity among DNA extraction methods.

Index	Protocol	Protocols compared	p values	q values
Aitchison distance (DEICODE)	ZymoBIOMICS	PowerFecalPro	0.921	0.921
		QIAampMini	0.771	0.921
		IndiSpin	0.423	0.921
	PowerFecalPro	QIAampMini	0.700	0.921
		IndiSpin	0.434	0.921
		QIAampMini	0.854	0.921
Unweighted UniFrac distance	ZymoBIOMICS	PowerFecalPro	0.001	0.002
		QIAampMini	0.001	0.002
		IndiSpin	0.001	0.002
	PowerFecalPro	QIAampMini	0.090	0.108
		IndiSpin	0.015	0.023
	QIAampMini	IndiSpin	0.325	0.325
Weighted UniFrac distance	ZymoBIOMICS	PowerFecalPro	0.936	0.993
		QIAampMini	0.897	0.993
		IndiSpin	0.954	0.993
	PowerFecalPro	QIAampMini	0.982	0.993
		IndiSpin	0.959	0.993
	QIAampMini	IndiSpin	0.993	0.993
Jaccard distance	ZymoBIOMICS	PowerFecalPro	0.037	0.074
		QIAampMini	0.003	0.018
		IndiSpin	0.011	0.033
	PowerFecalPro	QIAampMini	0.184	0.276
		IndiSpin	0.526	0.526
	QIAampMini	IndiSpin	0.256	0.307
Bray-Curtis distance	ZymoBIOMICS	PowerFecalPro	0.995	0.999
		QIAampMini	0.990	0.999
		IndiSpin	0.996	0.999
	PowerFecalPro	QIAampMini	0.999	0.999
		IndiSpin	0.996	0.999
	QIAampMini	IndiSpin	0.999	0.999

Pairwise PERMANOVA was tested for comparing beta diversity of DNA extraction method. Significant values in bold.

<https://doi.org/10.1371/journal.pone.0237556.t004>

other DNA extraction methods with when qualitative metrics such as Unweighted UniFrac (PowerFecalPro: $q = 0.002$; QIAampMini: $q = 0.002$; and IndiSpin: $q = 0.002$) and Jaccard distances (QIAampMini: $q = 0.018$ and IndiSpin: $q = 0.033$) were used. With PowerFecalPro vs. IndiSpin in Unweighted UniFrac ($q = 0.023$), being the only other observed significant difference. All other metrics showed no significance differences with regard to beta diversity.

Differential accessibility of microbiota by DNA extraction protocol

Linear discriminant analysis (LDA) effect size (LEfSe) analysis [62], identified several taxonomic groups, defined with an LDA score of 2 or higher (*one-against-all*), 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 *Streptomyetales* was highly enriched with the use of the QIAampMini

(LDA score > 3). However, post-hoc analysis of the LEfSe output, using Scheffe test [64] revealed that only the contaminants *Meiothermus* and *Hydrogenophilus* were enriched with Zymo, and *Streptomyces* was enriched in QIA (Fig 4A). These results reveal minimal to no significant enrichment of specific microbiota across extraction kits.

Microbial community type and HPV16

Dirichlet Multinomial Mixtures (DMM) model [70] detected two cervical microbial community types across all four DNA extraction protocols (S4 Fig). 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%), *Clamychia 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) [85], 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 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 (S4A Fig). In support of this result, analysis of differentially abundant microbiota using q_2 -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, S4B Fig). 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), LR-HPV (p = 0.370, Fisher’s exact test), multiple HPV infections (p = 0.303, 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; European descent vs non-European descent: p = 0.656; Hispanic vs non-Hispanic: p = 0.350, Fisher’s exact test, S4A Fig).

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 *Lactobacillus iners* and HPV16 status (Fig 2A & S4A Fig). The relationship between community types and HPV16 was detected regardless of the DNA extraction protocol used.

This study evaluated the composition of microbiota accessible across all DNA extraction methods. 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. It should be noted, however, that a singular kit should be used through the entirety of a study to minimize any

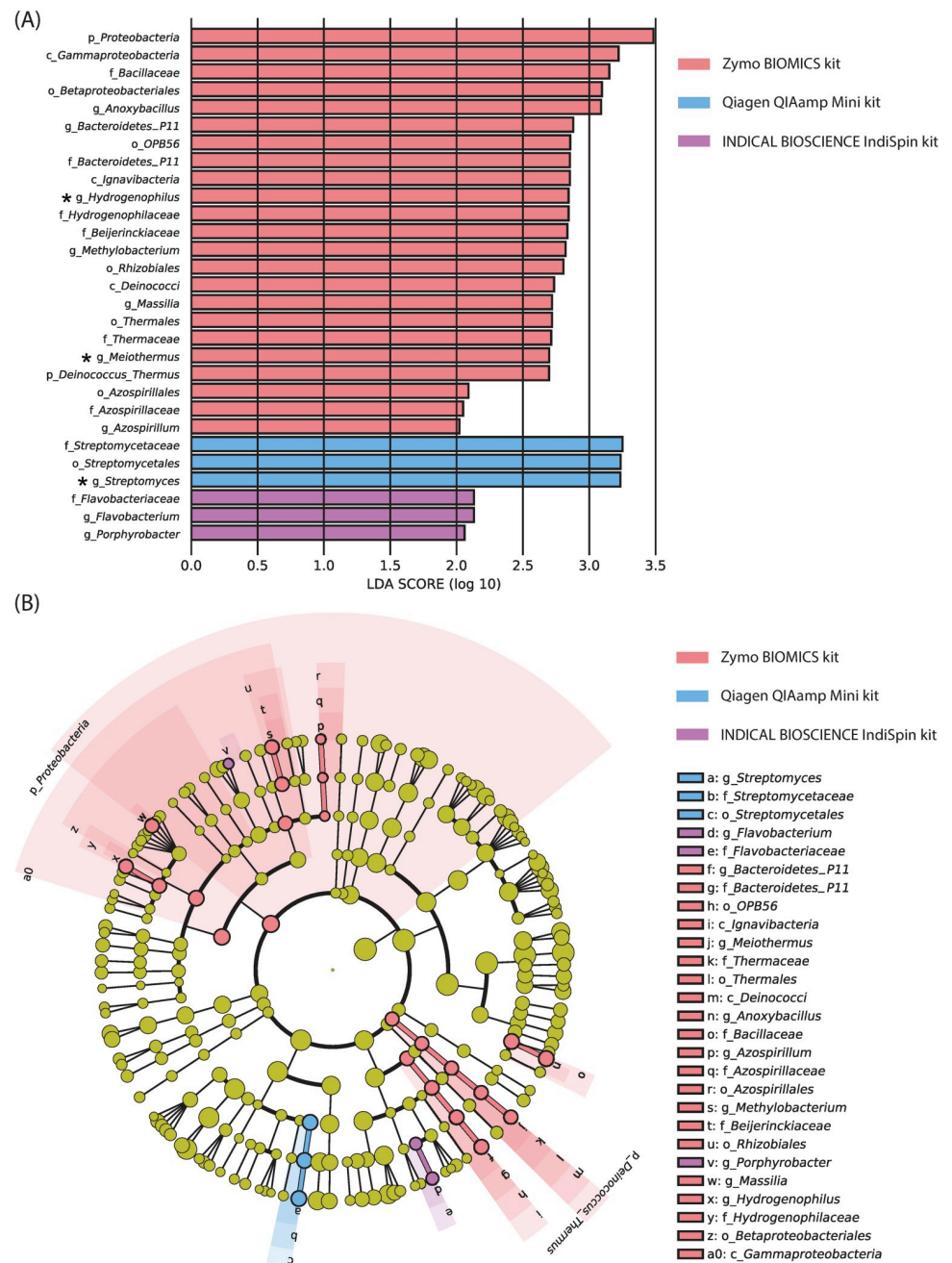


Fig 4. Distinct detections of microbes among the DNA extraction protocols. (A) A bar graph showing 23 significantly enriched taxa with ZymoBIOMICS, 3 with QIAamp DNA Mini Kit, and 3 with IndiSpin Pathogen Kit determined by the linear discriminant analysis (LDA) effect size (LEfSe) analyses [62]. Asterisks denote taxa of genus level that were significant after post-hoc significant testing with Scheffe. (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*) *Hydrogenophilaceae* (a member of the phylum *Proteobacteria*), and *Hydrogenophilus* (a member of the phylum *Proteobacteria*) are 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.

<https://doi.org/10.1371/journal.pone.0237556.g004>

subtle differences between samples, particularly when qualitative or richness-based diversity metrics are used. We detected potential DNA contamination with the ZymoBIOMICS and IndiSpin kits. The number of OTUs prior to rarefying revealed that the ZymoBIOMICS protocol detected more gram-negative OTUs than the PowerFecalPro (Fig 2B & Table 3). In particular, LEfSe analysis has shown that phylum *Proteobacteria* can be better detected with the ZymoBIOMICS kit (Fig 4). This signature was no longer observed after *post hoc* testing.

Although rarefying microbiome data can be problematic [86], it can still provide robust and interpretable results for diversity analysis [87], we were able to observe commensurate findings with non-rarefying approaches such as q_2 -breakaway [65], q_2 -deicode [66], and LEfSe [62]. Beta-diversity analysis via Unweighted UniFrac also revealed that ZymoBIOMICS was significantly different from all other kits (Table 4). There were no differences in non-phylogenetic indices of alpha diversity (Fig 3). These findings lead us to surmise that qualitative metrics are more sensitive to differences between extraction kits, while quantitative metrics were more sensitive to differences between subject (S2 & S3 Figs).

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). The number of reads of gram-positive and gram-negative bacteria also showed that there was no difference in the four kits (Table 3). 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 QIAamp-Mini protocol [36]. 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) [88] or MinION (Oxford Nanopore Technologies) [89, 90]. 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.

One limitation of our study is the lack of fresh LBC samples that would have enabled assessment the effects of prolonged storage on determining microbial community composition due to potential DNA degradation [25]. We think this may be unlikely, as our LBC samples were immediately frozen in -80°C , and DNA degradation within LBC samples stored at -80°C has been shown to be minimal [26]. However, the possibility that the observed microbial community composition may not be indicative of the community at the time of sampling remains. Despite this, our observations are commensurate with several prior studies in this area as outlined below. Community typing and detection of the differentially abundant microbiota revealed that *Lactobacillus iners* were more abundant in the cervical ecosystem without HPV16 (S4 Fig). These findings are congruent with those of, Usyk *et al.* [91], Lee *et al.* [1], and Audirac-Chalifour *et al.* [92]. Usyk *et al.*, reported that *L. iners* was associated with clearance of HR-HPV infections [91]. Lee *et al.* reported that *L. iners* were decreased in HPV positive women [1]. Also, the results indicated that the proportion of *L. 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* [92]. Similarly, Tuominen *et al.* [20] reported that *L. 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%) [93]. As established by the seminal study of Ranjeva *et al.* [94], a statistical model revealed that colonization of specific HPV types 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, and 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 [95]. We used a sample volume of 200 or 300 μ L ThinPrep solution in this study. The 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 μ L), 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.* [96] 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 [97]. Mitra *et al.* determined a sample volume of 500 μ L for ThinPrep by qPCR in the cervical microbiome study comparing sampling methods using cytobrush or swab [21]. 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 the cervix stored in ThinPrep at room temperature or -80°C was stable for at least one year [26]. 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 temperature followed by a controlled ambient environment (10 – 26.7°C) [25]. 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 [98, 99]. It should also be noted that other storage solutions, *i.e.*, those using guanidine thiocyanate have been reported for microbiome surveys of the cervix [100] and feces [101]. 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 as samples were limited in quantity, and we lacked fresh sample controls to assess the effects of prolonged storage to alter microbial community composition. Although several studies, have shown general stability and accessibility of DNA [26, 102, 103], there is potential for DNA degradation for samples not stored at low temperatures [25, 26]. However, the use of actual patient samples rather than mock samples is a strength of our approach.

Conclusions

In conclusion, regardless of the extraction protocol used, all kits provided equivalent broad accessibility to the cervical microbiome. Observed differences in microbial composition were due to the significant influence of the individual patient and not the extraction protocol. We have shown that the ability to characterize cervical microbiota from LBC specimens is possible, we were limited in our ability to directly assess if the observed microbial community composition would reflect that of a fresh sample. Despite this limitation, we were able to assess the relationship between HPV and the cervical microbiome, also supported by Kim *et al.* [26] and Castle *et al.* [25]. Cervical microbiome in patients with HPV16 or HPV18 which causes 70% of cervical cancers and CIN [104] warrants critical future study. Selection and characterization of appropriate DNA extraction methods are important for providing an accurate census of cervical microbiota and the human microbiome in general [25–29, 36]. Although we found all four extraction kits to be commensurate in their ability to broadly characterize the CM, one singular kit should be used throughout the entirety of a given study. 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 [17, 105].

Supporting information

S1 Fig. Comparison of DNA yields by DNA extraction protocols. DNA yield of QIAamp-Mini 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 μ L 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. (TIF)

S2 Fig. Phylogenetic beta-diversity. Weighted UniFrac (A & B) and Unweighted UniFrac (C & D), PCoA colored by subject ID (top row) and DNA extraction kit (bottom row). Weighted UniFrac clusters samples by subject whereas Unweighted UniFrac appears more sensitive to the type of DNA extraction kit. Data were rarefied to 51,197 reads per sample. (TIF)

S3 Fig. Deicode (robust Aitchison PCA) beta-diversity. Non-rarefaction-based analysis of beta-diversity. Samples are colored by individual subject ID (A) and DNA extraction kit (B). Samples predominately cluster by subject and not DNA extraction kit. (TIF)

S4 Fig. Community type and HPV 16 assessed by using 4 kits (A) Community types were classified into two types in all DNA extraction kits, mainly based on the percentage of *Lactobacillus iners*. HPV16 infection was negatively associated with the dominance of *L. iners* (community type I; $p = 0.001$, Fisher's exact test) regardless of DNA extraction method. Although, we observed slight variation in the abundance of microbiota across the extraction kits (even within the same individual patient), the ability to detect two community types was identical across all DNA extraction kits. No significant differences were observed in the relationship of other phenotypes of patients (HPV18, HR-HPV, LR-HPV, multiple HPV infections, Biopsy, and Race). The top 15 bacteria detected for each DNA extraction kit are shown. Samples were clustered by the Dirichlet component. Narrow columns show each sample and a broader column shows averages of samples. Rows show taxa at the species level. Dark or thin colors correspond to larger or smaller counts of OTUs, respectively. CT: community type. (B) LEfSe analysis, using combined data from all four kits detected a significant enrichment of 66 taxa in the cervical environment with HPV16 infection and 17 taxa without HPV16 infection. Genus *Lactobacillus* were enriched in the HPV16 negative patients ($p < 0.001$, LDA score: 5.38). Asterisks denote taxa that were significant after post-hoc significant testing with Scheffe test [64]. (TIF)

S1 Table. Diversity metrics used in this study. List of alpha and beta diversity metrics used and their respective QIIME 2 plugins. (DOCX)

Acknowledgments

We thank Togo Picture Gallery [39] for stock images shown in Fig 1.

Author Contributions

Conceptualization: Mayumi Nakagawa, Michael S. Robeson, II.

Data curation: Takeo Shibata, Michael S. Robeson, II.

Formal analysis: Takeo Shibata, Michael S. Robeson, II.

Funding acquisition: Mayumi Nakagawa.

Investigation: Takeo Shibata, Hannah N. Coleman, Sarah M. Owens, William W. Greenfield, Michael S. Robeson, II.

Methodology: Takeo Shibata, Hannah N. Coleman, Sarah M. Owens.

Resources: Mayumi Nakagawa, William W. Greenfield.

Supervision: Mayumi Nakagawa, Toshiyuki Sasagawa, Michael S. Robeson, II.

Writing – original draft: Takeo Shibata, Michael S. Robeson, II.

Writing – review & editing: Takeo Shibata, Mayumi Nakagawa, Hannah N. Coleman, Sarah M. Owens, William W. Greenfield, Toshiyuki Sasagawa, Michael S. Robeson, II.

References

1. Lee JE, Lee S, Lee H, Song YM, Lee K, Han MJ, et al. Association of the vaginal microbiota with human papillomavirus infection in a Korean twin cohort. *PLoS One*. 2013; 8(5):e63514. Epub 2013/05/30. <https://doi.org/10.1371/journal.pone.0063514> PMID: 23717441; PubMed Central PMCID: PMC3661536.
2. Huang X, Li C, Li F, Zhao J, Wan X, Wang K. Cervicovaginal microbiota composition correlates with the acquisition of high-risk human papillomavirus types. *Int J Cancer*. 2018; 143(3):621–34. Epub 2018/02/27. <https://doi.org/10.1002/ijc.31342> PMID: 29479697.
3. Zhou Y, Wang L, Pei F, Ji M, Zhang F, Sun Y, et al. Patients With LR-HPV Infection Have a Distinct Vaginal Microbiota in Comparison With Healthy Controls. *Front Cell Infect Microbiol*. 2019; 9:294. Epub 2019/09/27. <https://doi.org/10.3389/fcimb.2019.00294> PMID: 31555603; PubMed Central PMCID: PMC6722871.
4. Onywera H, Williamson AL, Mbulawa ZZA, Coetzee D, Meiring TL. The cervical microbiota in reproductive-age South African women with and without human papillomavirus infection. *Papillomavirus Res*. 2019; 7:154–63. Epub 2019/04/16. <https://doi.org/10.1016/j.pvr.2019.04.006> PMID: 30986570; PubMed Central PMCID: PMC6475661.
5. Brotman RM, Shardell MD, Gajer P, Tracy JK, Zenilman JM, Ravel J, et al. Interplay between the temporal dynamics of the vaginal microbiota and human papillomavirus detection. *J Infect Dis*. 2014; 210(11):1723–33. Epub 2014/06/20. <https://doi.org/10.1093/infdis/jiu330> PMID: 24943724; PubMed Central PMCID: PMC4296189.
6. Godoy-Vitorino F, Romaguera J, Zhao C, Vargas-Robles D, Ortiz-Morales G, Vazquez-Sanchez F, et al. Cervicovaginal Fungi and Bacteria Associated With Cervical Intraepithelial Neoplasia and High-Risk Human Papillomavirus Infections in a Hispanic Population. *Front Microbiol*. 2018; 9:2533. Epub 2018/11/09. <https://doi.org/10.3389/fmicb.2018.02533> PMID: 30405584; PubMed Central PMCID: PMC6208322.
7. Łaniewski P, Barnes D, Goulder A, Cui H, Roe DJ, Chase DM, et al. Linking cervicovaginal immune signatures, HPV and microbiota composition in cervical carcinogenesis in non-Hispanic and Hispanic women. *Sci Rep*. 2018. <https://doi.org/10.1038/s41598-018-25879-7> PMID: 29765068
8. Mitra A, MacIntyre DA, Lee YS, Smith A, Marchesi JR, Lehne B, et al. Cervical intraepithelial neoplasia disease progression is associated with increased vaginal microbiome diversity. *Sci Rep*. 2015; 5:16865. Epub 2015/11/18. <https://doi.org/10.1038/srep16865> PMID: 26574055; PubMed Central PMCID: PMC4648063.

9. Piyathilake CJ, Ollberding NJ, Kumar R, Macaluso M, Alvarez RD, Morrow CD. Cervical Microbiota Associated with Higher Grade Cervical Intraepithelial Neoplasia in Women Infected with High-Risk Human Papillomaviruses. *Cancer Prev Res (Phila)*. 2016; 9(5):357–66. Epub 2016/03/05. <https://doi.org/10.1158/1940-6207.CAPR-15-0350> PMID: 26935422; PubMed Central PMCID: PMC4869983.
10. Oh HY, Kim BS, Seo SS, Kong JS, Lee JK, Park SY, et al. The association of uterine cervical microbiota with an increased risk for cervical intraepithelial neoplasia in Korea. *Clin Microbiol Infect*. 2015; 21(7):674 e1-9. Epub 2015/03/11. <https://doi.org/10.1016/j.cmi.2015.02.026> PMID: 25752224.
11. De Seta F, Campisciano G, Zanotta N, Ricci G, Comar M. The Vaginal Community State Types Microbiome-Immune Network as Key Factor for Bacterial Vaginosis and Aerobic Vaginitis. *Front Microbiol*. 2019; 10:2451. Epub 2019/11/19. <https://doi.org/10.3389/fmicb.2019.02451> PMID: 31736898; PubMed Central PMCID: PMC6831638.
12. Oliver A, LaMere B, Weihe C, Wandro S, Lindsay KL, Wadhwa PD, et al. Cervicovaginal microbiome composition drives metabolic profiles in healthy pregnancy. *bioRxiv* <https://doi.org/10.1101/840520>. 2019.
13. Firwana B, Avaritt N, Shields B, Ravilla R, Makhoul I, Hutchins L, et al. Do checkpoint inhibitors rely on gut microbiota to fight cancer? *J Oncol Pharm Pract*. 2018; 24(6):468–72. <https://doi.org/10.1177/1078155217715903> PMID: 28625074
14. Ravilla R, Coleman HN, Chow CE, Chan L, Fuhrman BJ, Greenfield WW, et al. Cervical Microbiome and Response to a Human Papillomavirus Therapeutic Vaccine for Treating High-Grade Cervical Squamous Intraepithelial Lesion. *Integr Cancer Ther*. 2019; 18:1534735419893063. Epub 2019/12/14. <https://doi.org/10.1177/1534735419893063> PMID: 31833799; PubMed Central PMCID: PMC6913049.
15. Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012; 486(7402):207–14. Epub 2012/06/16. <https://doi.org/10.1038/nature11234> PMID: 22699609; PubMed Central PMCID: PMC3564958.
16. Bik EM, Bird SW, Bustamante JP, Leon LE, Nieto PA, Addae K, et al. A novel sequencing-based vaginal health assay combining self-sampling, HPV detection and genotyping, STI detection, and vaginal microbiome analysis. *PLoS One*. 2019; 14(5):e0215945. Epub 2019/05/03. <https://doi.org/10.1371/journal.pone.0215945> PMID: 31042762; PubMed Central PMCID: PMC6493738 have received stock options as well as other compensation. Some authors have patents pending in relation to this work: US Application No 15/198,818, Method and system for diagnostic testing, Application No 16/084,945, Method and system for microbiome-derived diagnostics and therapeutics for bacterial vaginosis, and Application No 16/115,542, Method and system for characterization for female reproductive system-related conditions associated with microorganisms. The data in this article were used in the development of a commercially available test product developed and marketed by uBiome. This does not alter our adherence to PLOS ONE policies on sharing data and materials.
17. Berman HL, McLaren MR, Callahan BJ. Understanding and interpreting community sequencing measurements of the vaginal microbiome. *BJOG*. 2020; 127(2):139–46. Epub 2019/10/10. <https://doi.org/10.1111/1471-0528.15978> PMID: 31597208.
18. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, et al. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A*. 2011; 108 Suppl 1:4680–7. Epub 2010/06/11. <https://doi.org/10.1073/pnas.1002611107> PMID: 20534435; PubMed Central PMCID: PMC3063603.
19. Fettweis JM, Serrano MG, Brooks JP, Edwards DJ, Girerd PH, Parikh HI, et al. The vaginal microbiome and preterm birth. *Nat Med*. 2019; 25(6):1012–21. Epub 2019/05/31. <https://doi.org/10.1038/s41591-019-0450-2> PMID: 31142849; PubMed Central PMCID: PMC6750801.
20. Tuominen H, Rautava S, Syrjanen S, Collado MC, Rautava J. HPV infection and bacterial microbiota in the placenta, uterine cervix and oral mucosa. *Sci Rep*. 2018; 8(1):9787. Epub 2018/06/30. <https://doi.org/10.1038/s41598-018-27980-3> PMID: 29955075; PubMed Central PMCID: PMC6023934.
21. Mitra A, MacIntyre DA, Mahajan V, Lee YS, Smith A, Marchesi JR, et al. Comparison of vaginal microbiota sampling techniques: cytobrush versus swab. *Sci Rep*. 2017; 7(1):9802. Epub 2017/08/31. <https://doi.org/10.1038/s41598-017-09844-4> PMID: 28852043; PubMed Central PMCID: PMC5575119.
22. Bentz JS. Liquid-based cytology for cervical cancer screening. *Expert Rev Mol Diagn*. 2005; 5(6):857–71. Epub 2005/11/01. <https://doi.org/10.1586/14737159.5.6.857> PMID: 16255628.
23. Gibb RK, Martens MG. The impact of liquid-based cytology in decreasing the incidence of cervical cancer. *Rev Obstet Gynecol*. 2011; 4(Suppl 1):S2–S11. Epub 2011/05/28. PMID: 21617785; PubMed Central PMCID: PMC3101960.
24. Donders GG, Depuydt CE, Bogers JP, Vereecken AJ. Association of *Trichomonas vaginalis* and cytological abnormalities of the cervix in low risk women. *PLoS One*. 2013; 8(12):e86266. Epub 2014/01/05. <https://doi.org/10.1371/journal.pone.0086266> PMID: 24386492; PubMed Central PMCID: PMC3875579.

25. Castle PE, Solomon D, Hildesheim A, Herrero R, Conception Bratti M, Sherman ME, et al. Stability of archived liquid-based cervical cytologic specimens. *Cancer*. 2003; 99(2):89–96. Epub 2003/04/22. <https://doi.org/10.1002/cncr.11058> PMID: 12704688.
26. Kim Y, Choi KR, Chae MJ, Shin BK, Kim HK, Kim A, et al. Stability of DNA, RNA, cytomorphology, and immunoantigenicity in Residual ThinPrep Specimens. *APMIS*. 2013; 121(11):1064–72. Epub 2013/04/10. <https://doi.org/10.1111/apm.12082> PMID: 23566220.
27. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, et al. Towards standards for human fecal sample processing in metagenomic studies. *Nat Biotechnol*. 2017; 35(11):1069–76. Epub 2017/10/03. <https://doi.org/10.1038/nbt.3960> PMID: 28967887.
28. Stinson LF, Keelan JA, Payne MS. Comparison of Meconium DNA Extraction Methods for Use in Microbiome Studies. *Front Microbiol*. 2018; 9:270. Epub 2018/03/09. <https://doi.org/10.3389/fmicb.2018.00270> PMID: 29515550; PubMed Central PMCID: PMC5826226.
29. Teng F, Darveekaran Nair SS, Zhu P, Li S, Huang S, Li X, et al. Impact of DNA extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling. *Sci Rep*. 2018; 8(1):16321. Epub 2018/11/07. <https://doi.org/10.1038/s41598-018-34294-x> PMID: 30397210; PubMed Central PMCID: PMC6218491.
30. Roche Molecular Diagnostics. LINEAR ARRAY® HPV Genotyping. <https://diagnostics.roche.com/global/en/products/params/linear-array-hpv-genotyping.html>. Accessed 12 Mar 2020.
31. de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H. Classification of papillomaviruses. *Virology*. 2004; 324(1):17–27. Epub 2004/06/09. <https://doi.org/10.1016/j.virol.2004.03.033> PMID: 15183049.
32. Munoz N, Bosch FX, de Sanjose S, Herrero R, Castellsague X, Shah KV, et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med*. 2003; 348(6):518–27. Epub 2003/02/07. <https://doi.org/10.1056/NEJMoa021641> PMID: 12571259.
33. Virtanen S, Kalliala I, Nieminen P, Salonen A. Comparative analysis of vaginal microbiota sampling using 16S rRNA gene analysis. *PLoS One*. 2017; 12(7):e0181477. Epub 2017/07/21. <https://doi.org/10.1371/journal.pone.0181477> PMID: 28723942; PubMed Central PMCID: PMC5517051.
34. Microbial Isolation | ZYMO RESEARCH. <https://www.zymoresearch.com/pages/microbial-isolation>. Accessed 12 Mar 2020.
35. PowerBead Tubes—QIAGEN Online Shop. <https://www.qiagen.com/us/products/discovery-and-translational-research/lab-essentials/plastics/powerbead-tubes/#orderinginformation>. Accessed 12 Mar 2020.
36. Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ. Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One*. 2012; 7(3):e33865. Epub 2012/03/30. <https://doi.org/10.1371/journal.pone.0033865> PMID: 22457796; PubMed Central PMCID: PMC3311548.
37. QIAGEN. Pathogen Lysis Tubes—QIAGEN. <https://www.qiagen.com/dk/shop/pcr/pathogen-lysis-tubes/>. Accessed 12 Mar 2020.
38. Kim D, Hofstaedter CE, Zhao C, Mattei L, Tanes C, Clarke E, et al. Optimizing methods and dodging pitfalls in microbiome research. *Microbiome*. 2017; 5(1):52. Epub 2017/05/10. <https://doi.org/10.1186/s40168-017-0267-5> PMID: 28476139; PubMed Central PMCID: PMC5420141.
39. Togo Picture Gallery. <http://togotv.dbcls.jp/pics.html>. Accessed 12 Mar 2020.
40. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A*. 2011; 108 Suppl 1(Supplement 1):4516–22. Epub 2010/06/11. <https://doi.org/10.1073/pnas.1000080107> PMID: 20534432; PubMed Central PMCID: PMC3063599.
41. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*. 2017; 551(7681):457–63. Epub 2017/11/02. <https://doi.org/10.1038/nature24621> PMID: 29088705; PubMed Central PMCID: PMC6192678.
42. Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat Microb Ecol*. 2015; 75(2):129–37. <https://doi.org/10.3354/ame01753> WOS:000357106200004.
43. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol*. 2016; 18(5):1403–14. Epub 2015/08/15. <https://doi.org/10.1111/1462-2920.13023> PMID: 26271760.
44. Walters W, Hyde ER, Berg-Lyons D, Ackermann G, Humphrey G, Parada A, et al. Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems*. 2016; 1(1). Epub 2016/11/09. <https://doi.org/10.1128/mSystems.00009-15> PMID: 27822518; PubMed Central PMCID: PMC5069754.

45. Earth Microbiome Project. 16S Illumina amplicon protocol. <http://www.earthmicrobiome.org/protocols-and-standards/16s/>. Accessed 12 Mar 2020.
46. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol.* 2019; 37(8):852–7. Epub 2019/07/26. <https://doi.org/10.1038/s41587-019-0209-9> PMID: 31341288; PubMed Central PMCID: PMC7015180.
47. QIIME 2 View. <https://view.qiime2.org/>. Accessed 12 Mar 2020.
48. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* 2017; 11(12):2639–43. Epub 2017/07/22. <https://doi.org/10.1038/ismej.2017.119> PMID: 28731476; PubMed Central PMCID: PMC5702726.
49. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods.* 2016; 13(7):581–3. Epub 2016/05/24. <https://doi.org/10.1038/nmeth.3869> PMID: 27214047; PubMed Central PMCID: PMC4927377.
50. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome.* 2018; 6(1):90. Epub 2018/05/19. <https://doi.org/10.1186/s40168-018-0470-z> PMID: 29773078; PubMed Central PMCID: PMC5956843.
51. Werner JJ, Koren O, Hugenholtz P, DeSantis TZ, Walters WA, Caporaso JG, et al. Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. *ISME J.* 2012; 6(1):94–103. Epub 2011/07/01. <https://doi.org/10.1038/ismej.2011.82> PMID: 21716311; PubMed Central PMCID: PMC3217155.
52. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 2013; 41(Database issue):D590–6. Epub 2012/11/30. <https://doi.org/10.1093/nar/gks1219> PMID: 23193283; PubMed Central PMCID: PMC3531112.
53. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013; 30(4):772–80. Epub 2013/01/19. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690; PubMed Central PMCID: PMC3603318.
54. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015; 32(1):268–74. Epub 2014/11/06. <https://doi.org/10.1093/molbev/msu300> PMID: 25371430; PubMed Central PMCID: PMC4271533.
55. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jeremiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017; 14(6):587–9. Epub 2017/05/10. <https://doi.org/10.1038/nmeth.4285> PMID: 28481363; PubMed Central PMCID: PMC5453245.
56. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods.* 2013; 10(1):57–9. Epub 2012/12/04. <https://doi.org/10.1038/nmeth.2276> PMID: 23202435; PubMed Central PMCID: PMC3531572.
57. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One.* 2013; 8(4):e61217. Epub 2013/05/01. <https://doi.org/10.1371/journal.pone.0061217> PMID: 23630581; PubMed Central PMCID: PMC3632530.
58. Bisanz JE. qiime2R: Importing QIIME2 artifacts and associated data into R sessions. <https://github.com/jbisanz/qiime2R>. Accessed 12 Mar 2020.
59. Lahti L, Shetty S. microbiome R package. <http://microbiome.github.io>. Accessed 12 Mar 2020.
60. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 2001; 26(1):32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x> WOS:000167002000004.
61. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlenn D, et al. vegan: Community Ecology Package. R package version 2.5–3. <https://CRAN.R-project.org/package=vegan>. Accessed 12 Mar 2020.
62. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. *Genome Biol.* 2011; 12(6):R60. Epub 2011/06/28. <https://doi.org/10.1186/gb-2011-12-6-r60> PMID: 21702898; PubMed Central PMCID: PMC3218848.
63. Bardou P, Mariette J, Escudie F, Djemiel C, Klopp C. jvrenn: an interactive Venn diagram viewer. *BMC Bioinformatics.* 2014; 15(1):293. Epub 2014/09/02. <https://doi.org/10.1186/1471-2105-15-293> PMID: 25176396; PubMed Central PMCID: PMC4261873.
64. Cao Y. microbiomeMarker: microbiome biomarker analysis. R package version 0.0.1.9000. <https://github.com/yiluheihei/microbiomeMarker>. Accessed 21 Nov 2020. <https://doi.org/10.5281/zenodo.3749415>

65. Willis A, Bunge J. Estimating diversity via frequency ratios. *Biometrics*. 2015; 71(4):1042–9. Epub 2015/06/04. <https://doi.org/10.1111/biom.12332> PMID: 26038228.
66. Martino C, Morton JT, Marotz CA, Thompson LR, Tripathi A, Knight R, et al. A Novel Sparse Compositional Technique Reveals Microbial Perturbations. *mSystems*. 2019; 4(1). Epub 2019/02/26. <https://doi.org/10.1128/mSystems.00016-19> PMID: 30801021; PubMed Central PMCID: PMC6372836.
67. Gao W, Weng J, Gao Y, Chen X. Comparison of the vaginal microbiota diversity of women with and without human papillomavirus infection: a cross-sectional study. *BMC Infect Dis*. 2013; 13(1):271. Epub 2013/06/14. <https://doi.org/10.1186/1471-2334-13-271> PMID: 23758857; PubMed Central PMCID: PMC3684509.
68. Montealegre JR, Peckham-Gregory EC, Marquez-Do D, Dillon L, Guillaud M, Adler-Storthz K, et al. Racial/ethnic differences in HPV 16/18 genotypes and integration status among women with a history of cytological abnormalities. *Gynecol Oncol*. 2018; 148(2):357–62. Epub 2017/12/26. <https://doi.org/10.1016/j.ygyno.2017.12.014> PMID: 29276057; PubMed Central PMCID: PMC5801201.
69. Xi LF, Kiviat NB, Hildesheim A, Galloway DA, Wheeler CM, Ho J, et al. Human papillomavirus type 16 and 18 variants: race-related distribution and persistence. *J Natl Cancer Inst*. 2006; 98(15):1045–52. Epub 2006/08/03. <https://doi.org/10.1093/jnci/djj297> PMID: 16882941.
70. Morgan M. DirichletMultinomial: Dirichlet-Multinomial Mixture Model Machine Learning for Microbiome Data. <http://bioconductor.org/packages/release/bioc/html/DirichletMultinomial.html>. Accessed 12 Mar 2020.
71. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS One*. 2012; 7(2):e30126. Epub 2012/02/10. <https://doi.org/10.1371/journal.pone.0030126> PMID: 22319561; PubMed Central PMCID: PMC3272020.
72. DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, et al. Temporal and spatial variation of the human microbiota during pregnancy. *Proc Natl Acad Sci U S A*. 2015; 112(35):11060–5. Epub 2015/08/19. <https://doi.org/10.1073/pnas.1502875112> PMID: 26283357; PubMed Central PMCID: PMC4568272.
73. Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB. ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-Seq. *PLoS One*. 2013; 8(7):e67019. Epub 2013/07/12. <https://doi.org/10.1371/journal.pone.0067019> PMID: 23843979; PubMed Central PMCID: PMC3699591.
74. Dinno A. dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums. <https://CRAN.R-project.org/package=dunn.test>. Accessed 12 Mar 2020.
75. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, et al. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIXS) specifications. *Nat Biotechnol*. 2011; 29(5):415–20. Epub 2011/05/10. <https://doi.org/10.1038/nbt.1823> PMID: 21552244; PubMed Central PMCID: PMC3367316.
76. Silverman JD, Bloom RJ, Jiang S, Durand HK, Mukherjee S, David LA. Measuring and mitigating PCR bias in microbiome data. *BioRxiv*. 2019:604025.
77. Laursen MF, Dalgaard MD, Bahl MI. Genomic GC-Content Affects the Accuracy of 16S rRNA Gene Sequencing Based Microbial Profiling due to PCR Bias. *Front Microbiol*. 2017; 8:1934. Epub 2017/10/21. <https://doi.org/10.3389/fmicb.2017.01934> PMID: 29051756; PubMed Central PMCID: PMC5633598.
78. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One*. 2011; 6(12):e27310. Epub 2011/12/24. <https://doi.org/10.1371/journal.pone.0027310> PMID: 22194782; PubMed Central PMCID: PMC3237409.
79. Silhavy TJ, Kahne D, Walker S. The bacterial cell envelope. *Cold Spring Harb Perspect Biol*. 2010; 2(5):a000414. Epub 2010/05/11. <https://doi.org/10.1101/cshperspect.a000414> PMID: 20452953; PubMed Central PMCID: PMC2857177.
80. Balle C, Lennard K, Dabee S, Barnabas SL, Jaumdally SZ, Gasper MA, et al. Endocervical and vaginal microbiota in South African adolescents with asymptomatic Chlamydia trachomatis infection. *Sci Rep*. 2018; 8(1):11109. Epub 2018/07/25. <https://doi.org/10.1038/s41598-018-29320-x> PMID: 30038262; PubMed Central PMCID: PMC6056523.
81. Klein C, Gonzalez D, Samwel K, Kahesa C, Mwaiselage J, Aluthge N, et al. Relationship between the Cervical Microbiome, HIV Status, and Precancerous Lesions. *mBio*. 2019; 10(1). Epub 2019/02/21. <https://doi.org/10.1128/mBio.02785-18> PMID: 30782659; PubMed Central PMCID: PMC6381280.
82. Hayashi NR, Ishida T, Yokota A, Kodama T, Igarashi Y. *Hydrogenophilus thermoluteolus* gen. nov., sp. nov., a thermophilic, facultatively chemolithoautotrophic, hydrogen-oxidizing bacterium. *Int J Syst Bacteriol*. 1999; 49 Pt 2:783–6. Epub 1999/05/13. <https://doi.org/10.1099/00207713-49-2-783> PMID: 10319503.

83. Glassing A, Dowd SE, Galandiuk S, Davis B, Chiodini RJ. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathog.* 2016. <https://doi.org/10.1186/s13099-016-0103-7> PMID: 27239228
84. Birse KD, Romas LM, Guthrie BL, Nilsson P, Bosire R, Kiarie J, et al. Genital Injury Signatures and Microbiome Alterations Associated With Depot Medroxyprogesterone Acetate Usage and Intravaginal Drying Practices. *J Infect Dis.* 2017; 215(4):590–8. Epub 2016/12/25. <https://doi.org/10.1093/infdis/jiw590> PMID: 28011908; PubMed Central PMCID: PMC5388302.
85. Lennard K, Dabee S, Barnabas SL, Havyarimana E, Blakney A, Jaumdally SZ, et al. Microbial Composition Predicts Genital Tract Inflammation and Persistent Bacterial Vaginosis in South African Adolescent Females. *Infect Immun.* 2018; 86(1). Epub 2017/10/19. <https://doi.org/10.1128/IAI.00410-17> PMID: 29038128; PubMed Central PMCID: PMC5736802.
86. McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol.* 2014; 10(4):e1003531. Epub 2014/04/05. <https://doi.org/10.1371/journal.pcbi.1003531> PMID: 24699258; PubMed Central PMCID: PMC3974642.
87. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome.* 2017; 5(1):27. Epub 2017/03/04. <https://doi.org/10.1186/s40168-017-0237-y> PMID: 28253908; PubMed Central PMCID: PMC5335496.
88. Callahan BJ, Wong J, Heiner C, Oh S, Theriot CM, Gulati AS, et al. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Res.* 2019; 47(18):e103. Epub 2019/07/04. <https://doi.org/10.1093/nar/gkz569> PMID: 31269198; PubMed Central PMCID: PMC6765137.
89. Calus ST, Ijaz UZ, Pinto AJ. NanoAmpli-Seq: a workflow for amplicon sequencing for mixed microbial communities on the nanopore sequencing platform. *Gigascience.* 2018; 7(12). Epub 2018/11/27. <https://doi.org/10.1093/gigascience/giy140> PMID: 30476081; PubMed Central PMCID: PMC6298384.
90. Wongsurawat T, Nakagawa M, Atiq O, Coleman HN, Jenjaroenpun P, Allred JI, et al. An assessment of Oxford Nanopore sequencing for human gut metagenome profiling: A pilot study of head and neck cancer patients. *J Microbiol Methods.* 2019; 166:105739. Epub 2019/10/19. <https://doi.org/10.1016/j.mimet.2019.105739> PMID: 31626891; PubMed Central PMCID: PMC6956648.
91. Usyk M, Zolnik CP, Castle PE, Porras C, Herrero R, Gradissimo A, et al. Cervicovaginal microbiome and natural history of HPV in a longitudinal study. *PLoS Pathog.* 2020; 16(3):e1008376. Epub 2020/03/28. <https://doi.org/10.1371/journal.ppat.1008376> PMID: 32214382; PubMed Central PMCID: PMC7098574 Schiller and Douglas R. Lowy report that they are named inventors on US Government-owned HPV vaccine patents that are licensed to GlaxoSmithKline and Merck and for which the National Cancer Institute receives licensing fees. They are entitled to limited royalties as specified by federal law.
92. Audirac-Chalifour A, Torres-Poveda K, Bahena-Roman M, Tellez-Sosa J, Martinez-Barnetche J, Cortina-Ceballos B, et al. Cervical Microbiome and Cytokine Profile at Various Stages of Cervical Cancer: A Pilot Study. *PLoS One.* 2016; 11(4):e0153274. Epub 2016/04/27. <https://doi.org/10.1371/journal.pone.0153274> PMID: 27115350; PubMed Central PMCID: PMC4846060.
93. Di Paola M, Sani C, Clemente AM, Iossa A, Perissi E, Castronovo G, et al. Characterization of cervicovaginal microbiota in women developing persistent high-risk Human Papillomavirus infection. *Sci Rep.* 2017; 7(1):10200. Epub 2017/09/02. <https://doi.org/10.1038/s41598-017-09842-6> PMID: 28860468; PubMed Central PMCID: PMC5579045.
94. Ranjeva SL, Mihaljevic JR, Joseph MB, Giuliano AR, Dwyer G. Untangling the dynamics of persistence and colonization in microbial communities. *ISME J.* 2019; 13(12):2998–3010. Epub 2019/08/25. <https://doi.org/10.1038/s41396-019-0488-7> PMID: 31444482; PubMed Central PMCID: PMC6863904.
95. Linder J, Zahniser D. ThinPrep Papanicolaou testing to reduce false-negative cervical cytology. *Arch Pathol Lab Med.* 1998; 122(2):139–44. Epub 1998/03/14. PMID: 9499356.
96. Ling Z, Liu X, Chen X, Zhu H, Nelson KE, Xia Y, et al. Diversity of cervicovaginal microbiota associated with female lower genital tract infections. *Microb Ecol.* 2011; 61(3):704–14. Epub 2011/02/03. <https://doi.org/10.1007/s00248-011-9813-z> PMID: 21287345.
97. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 2014; 12(1):87. Epub 2014/11/13. <https://doi.org/10.1186/s12915-014-0087-z> PMID: 25387460; PubMed Central PMCID: PMC4228153.
98. Rebolj M, Rask J, van Ballegooijen M, Kirschner B, Rozemeijer K, Bonde J, et al. Cervical histology after routine ThinPrep or SurePath liquid-based cytology and computer-assisted reading in Denmark. *Br J Cancer.* 2015; 113(9):1259–74. Epub 2015/10/09. <https://doi.org/10.1038/bjc.2015.339> PMID: 26448176; PubMed Central PMCID: PMC4815798.

99. Naeem RC, Goldstein DY, Einstein MH, Ramos Rivera G, Schlesinger K, Khader SN, et al. 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*. 2017; 48(3):207–13. Epub 2017/04/06. <https://doi.org/10.1093/labmed/lmx019> PMID: 28379422.
100. Ritu W, Enqi W, Zheng S, Wang J, Ling Y, Wang Y. Evaluation of the Associations Between Cervical Microbiota and HPV Infection, Clearance, and Persistence in Cytologically Normal Women. *Cancer Prev Res (Phila)*. 2019; 12(1):43–56. Epub 2018/11/23. <https://doi.org/10.1158/1940-6207.CAPR-18-0233> PMID: 30463989.
101. Hosomi K, Ohno H, Murakami H, Natsume-Kitatani Y, Tanisawa K, Hirata S, et al. Method for preparing DNA from feces in guanidine thiocyanate solution affects 16S rRNA-based profiling of human microbiota diversity. *Sci Rep*. 2017; 7(1):4339. Epub 2017/07/01. <https://doi.org/10.1038/s41598-017-04511-0> PMID: 28659635; PubMed Central PMCID: PMC5489508.
102. Akahane T, Yamaguchi T, Kato Y, Yokoyama S, Hamada T, Nishida Y, et al. Comprehensive validation of liquid-based cytology specimens for next-generation sequencing in cancer genome analysis. *PLoS One*. 2019; 14(6):e0217724. Epub 2019/06/15. <https://doi.org/10.1371/journal.pone.0217724> PMID: 31199826; PubMed Central PMCID: PMC6568385.
103. Cuschieri KS, Beattie G, Hassan S, Robertson K, Cubie H. Assessment of human papillomavirus mRNA detection over time in cervical specimens collected in liquid based cytology medium. *J Virol Methods*. 2005; 124(1–2):211–5. Epub 2005/01/25. <https://doi.org/10.1016/j.jviromet.2004.11.005> PMID: 15664071.
104. Human papillomavirus (HPV) and cervical cancer. [https://www.who.int/news-room/fact-sheets/detail/human-papillomavirus-\(hpv\)-and-cervical-cancer](https://www.who.int/news-room/fact-sheets/detail/human-papillomavirus-(hpv)-and-cervical-cancer). Accessed 12 Mar 2020.
105. Sarangi AN, Goel A, Aggarwal R. Methods for Studying Gut Microbiota: A Primer for Physicians. *J Clin Exp Hepatol*. 2019; 9(1):62–73. Epub 2019/02/19. <https://doi.org/10.1016/j.jceh.2018.04.016> PMID: 30774267; PubMed Central PMCID: PMC6363981.