bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

3

4 Takeo Shibata,[1,2] Mayumi Nakagawa,[1] Hannah N. Coleman,[1] Sarah M. Owens,[3] William W.
5 Greenfield,[4] Toshiyuki Sasagawa,[2] Michael S. Robeson II[5]

6

7 [1]Department of Pathology, University of Arkansas for Medical Sciences, Little Rock, AR, USA
8 [2]Department of Obstetrics and Gynecology, Kanazawa Medical University, Uchinada, Ishikawa,
9 Japan
10 [3]Biosciences Division, Argonne National Laboratory, Lemont, IL, USA
11 [4]Department of Obstetrics and Gynecology, University of Arkansas for Medical Sciences, Little
12 Rock, AR, USA
13 [5]Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little
14 Rock, AR, USA

15

16 * Corresponding author: Michael S. Robeson II
17 Email: MRobeson@uams.edu (M.S.R.)

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

18 # Abstract

19 Cervical microbiota (CM) are considered an important factor affecting the progression of
20 cervical intraepithelial neoplasia (CIN) and are implicated in the persistence of human
21 papillomavirus (HPV). Collection of liquid-based cytology (LBC) samples is routine for cervical
22 cancer screening and HPV genotyping and can be used for long-term cytological biobanking. We
23 sought to determine whether it is possible to access microbial DNA from LBC specimens, and
24 compared the performance of four different extraction protocols: (ZymoBIOMICS DNA
25 Miniprep Kit; QIAamp PowerFecal Pro DNA Kit; QIAamp DNA Mini Kit; and IndiSpin
26 Pathogen Kit) and their ability to capture the diversity of CM from LBC specimens. LBC
27 specimens from 20 patients (stored for 716 ± 105 days) with CIN values of 2 or 3 were each
28 aliquoted for each of the four kits. Loss of microbial diversity due to long-term LBC storage
29 could not be assessed due to lack of fresh LBC samples. Comparisons with other types of
30 cervical sampling were not performed. We observed that all DNA extraction kits provided
31 equivalent accessibility to the cervical microbial DNA within stored LBC samples.
32 Approximately 80% microbial genera were shared among all DNA extraction protocols.
33 Potential kit contaminants were observed as well. Variation between individuals was a
34 significantly greater influence on the observed microbial composition than was the method of
35 DNA extraction. We also observed that HPV16 was significantly associated with community
36 types that were not dominated by *Lactobacillus iners*.

37

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

39 # Introduction

40 High-throughput sequencing (HTS) technology of 16S rRNA gene amplicon sequences has made
41 it possible to better understand the relationships between cervicovaginal microbiota and human
42 papillomavirus (HPV) infection [1] [2] [3] [4] [5] and HPV-related diseases [6] [7] [8] [9] [10].
43 Cervicovaginal microbiota are considered to be an important factor affecting the progress of
44 cervical intraepithelial neoplasia (CIN) [6] [7] [8] [9] and are implicated in the persistence of
45 high-risk HPV (HR-HPV) [1] [2] and low-risk HPV (LR-HPV) [3]. However, microbial
46 signatures associated with either HR-HPV or LR HPV can vary depending on the population
47 under study, e.g., the phyla *Actinobacteria* and *Fusobacteria* were found to be enriched in HR-
48 HPV positive Chinese women [4] while another study observed these groups associated with
49 low-risk HPV (LR-HPV) in South African women [3]. Additionally, *Lactobacillus iners*-
50 dominant samples are associated with both HR-HPV and LR-HPV [5], often associated with
51 moderate CIN risk [10]. Moreover, it has been shown that CIN risk was increased in patients
52 with HR-HPV [10] when the cervical microbes *Atopobium vaginae*, *Gardnerella vaginalis*, and
53 *Lactobacillus iners* were present in greater proportion compared to *L. crispatus*. The
54 cervicovaginal microbiome is often described by the abundance of *Lactobacillus spp., i.e.* the
55 community is either referred to as a *Lactobacillus*-dominant type or non-*Lactobacillus*-dominant
56 type, and can interact with the immune system in different ways [7] [11]. For example,
57 inflammatory cytokines, such as Interleukin (IL)-1α and IL-18, were increased in non-
58 *Lactobacillus*-dominant community types of reproductive-aged healthy women [11]. In the
59 analysis of patients with cervical cancer, non-*Lactobacillus*-dominant community types were
60 positively associated with chemokines such as interferon gamma-induced protein 10 (IP-10) and
61 soluble CD40-ligand activating dendric cells (DCs) [7]. The metabolism of the cervicovaginal

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

62 microbiome may be a substantial contributing factor to maternal health during pregnancy,
63 although the mechanism is still unclear [12]. Prior research, on the importance of the microbiome
64 in cancer therapeutics via checkpoint inhibitors [13], along with our own work on the role of CM
65 in vaccine response [14], suggests that the CM has a significant role to play in disease
66 progression and therapeutic treatment. We continue our work here to further assess use liquid-
67 based cytology (LBC) samples to survey microbial community DNA.

68 Little has been reported on the utility of LBC samples for use in cervical microbiome
69 studies. Conventionally, microbiome sample collection methods entail the use of swabs [15] or
70 self-collection of vaginal discharge [16]. To obtain a non-biased and broad range of cervical
71 microbiota, DNA extraction should be optimized for a range of difficult-to-lyse-bacteria, *e.g.*
72 *Firmicutes*, *Actinobacteria*, and *Lactobacillus* [15] [17] [18] [19] [20].

73 LBC samples are promising for cervicovaginal microbiome surveys, as they are an
74 already established method of long-term cytological biobanking [21]. In clinical practice,
75 cervical cytology for cervical cancer screening or HPV genotyping is widely performed using a
76 combination of cervical cytobrushes and LBC samples such as ThinPrep (HOLOGIC) or
77 SurePath (BD). An LBC specimen can be used for not only cytological diagnosis but also
78 additional diagnostic tests such as HPV, *Chlamydia*, *Neisseria gonorrhoeae*, and *Trichomonas*
79 infection [22] [23] [24]. Despite the promising potential to use LBC samples for surveying
80 cervicovaginal microbiota, it is known that DNA contained within LBC samples may degrade
81 over prolonged storage times when kept at ambient or non-freezing temperatures [25] [26].
82 Although others have shown minimal DNA degradation of LBC samples stored at -80 □C [26],
83 the ability to reliably access microbial DNA remains to be seen, and is the focus of our study.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

84 Furthermore, the ability to characterize these microbiota, as commonly assessed by 16S
85 rRNA gene sequencing, can be biased as a result of methodological differences of cell lysis and
86 DNA extraction protocols [27] [28] [29]. Herein, we compare four different commercially
87 available DNA extraction kits in an effort to assess their ability to extract and characterize any
88 viable microbial DNA from stored LBC samples. Additionally, we examine the relationship
89 between HPV infection and the composition of cervical microbiota still accessible after
90 prolonged LBC storage.
91

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

92 # Methods

93 ## Recruiting patients

94 Patients were participants of a single center randomized double blind Phase II clinical trial
95 (NCT02481414) in which enrollees were assigned to receive an HPV therapeutic vaccine called
96 PepCan or an adjuvant derived from *Candida albicans* (Candin®, Nielsen BioSciences, San
97 Diego, CA). Pre-injection liquid based cervical cytology (ThinPrep) samples from 20
98 consecutive enrollees who gave written informed consent between 3/21/2017 and 12/11/2017
99 were used for this study. Patients were recruited mainly through referrals from clinics from
100 inside and outside of the medical center. Flyers and Google advertisements were also utilized.
101 Inclusion Criteria: aged 18–50 years, had recent (≤ 60 days) Pap smear result consistent with
102 HSIL or "cannot rule out HSIL" or HSIL on colposcopy guided biopsy, untreated for HSIL or
103 "Cannot rule out HSIL", able to provide informed consent, willingness and able to comply with
104 the requirements of the protocol. Exclusion Criteria: history of disease or treatment causing
105 immunosuppression (e.g., cancer, HIV, organ transplant, autoimmune disease), being pregnant or
106 attempting to be pregnant within the period of study participation, breast feeding or planning to
107 breast feed within the period of study participation, allergy to Candida antigen, history of severe
108 asthma requiring emergency room visit or hospitalization within the past 5 years, history of
109 invasive squamous cell carcinoma of the cervix, history of having received PepCan. Those who
110 qualified for the study based on their cervical cytology underwent cervical biopsy, and they
111 qualified for vaccination if the results were CIN2/3. All collected samples are representative
112 of a larger population in gynecology clinics with abnormal Pap tests. If in the opinion of the
113 Principal Investigator or other Investigators, it is not in the best interest of the patient to enter this
114 study, the patient was excluded. Patients' age, race, and ethnicity were recorded based on

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

115 standard NIH requirements. All categories and definitions, *e.g.* ethnicity, age, etc., were based on
116 NIH Guidelines.
117

118 Sampling of cervical microbiome

119 The cervical cytology specimens in this current study were collected before the vaccination and
120 reserved in the vial of the ThinPrep Pap Test (HOLOGIC) as described in Ravilla *et al*. 2019
121 [14]. The specimens were frozen in an ultra-low temperature freezer (-80□C) on the day of
122 collection. The storage period from sample collection to DNA extraction was 716 ± 105 days.
123

124 HPV genotyping

125 HPV-DNA was detected by Linear Array HPV Genotyping Test (Roche Diagnostics) which can
126 detect up to 37 HPV genotypes using ThinPrep solution [30]. HPV16, 18, 31, 33, 35, 39, 45, 51,
127 52, 56, 58, 59, and 68 were defined as HR-HPV genotypes; and HPV6, 11, 40, 42, 54, 61, 62, 71,
128 72, 81, 83, 84, and CP6108 were defined as LR-HPV genotypes [31] [32] .
129

130 DNA extraction protocols

131 We selected four commercially available DNA extraction kits as the candidates for comparison:
132 ZymoBIOMICS DNA Miniprep Kit (Zymo Research, D4300), QIAamp PowerFecal Pro DNA
133 Kit (QIAGEN, 51804), QIAamp DNA Mini Kit (QIAGEN, 51304), and IndiSpin Pathogen Kit
134 (Indical Bioscience, SPS4104). These kits have been successfully used in a variety of human
135 cervical, vaginal, and gut microbiome surveys [10] [21] [33]. We'll subsequently refer to each of

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

136 these kits in abbreviated form as follows: ZymoBIOMICS, PowerFecalPro, QIAampMini, and

137 IndiSpin. The protocols and any modifications are outlined in Table 1.

138

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, j]
QIAamp PowerFecal Pro DNA Kit (51804)	Qiagen	300 µL	No	Ceramic[d]	10 min[b]	No	[c, j]
QIAamp DNA Mini Kit (51304)	Qiagen	200 µL	Mutanolysin[e]	No	No	No	[c, f, g, j]
IndiSpin Pathogen Kit (SPS4104)	Indical Bioscience	300 µL	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 µL) 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.

139

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

140 Each LBC sample was dispensed into four separate 2 mL sterile collection tubes
141 (dispensed sample volume = 500 μL) to create four cohorts of 20 DNA extractions (Figure 1).
142 Each extraction cohort was processed through one of the four kits above. A total of 80
143 extractions (4 kits × 20 patients) were prepared for subsequent analyses. Applied sample volume
144 of ThinPrep solution was 300 μL for ZymoBIOMICS, 300 μL for PowerFecalPro, 200 μL for
145 QIAampMini, and 300 μL for IndiSpin. The sample volume was standardized to 300 μL as long
146 as the manufacturer's instructions allowed to do so. DNA extraction for all samples was
147 performed by the same individual who practiced by performing multiple extractions for each kit
148 before performing the actual DNA extraction on the samples analyzed in this study. Positive
149 control was mock vaginal microbial communities composed of a mixture of genomic DNA from
150 the American Type Culture Collection (ATCC MSA1007). Negative control was the ThinPrep
151 preservation solution without the sample as blank extraction [38].

152

153 **Figure 1. Overview of the study design using 16S rRNA gene to compare the DNA**
154 **extraction protocol.** (A) Liquid-based cytology (LBC) specimens from 20 patients with CIN2/3
155 or suspected CIN2/3. (B) A total of 80 DNA extractions were performed. (C) The four DNA
156 extraction methods. (D) DNA of mock vaginal community as a positive control and preservation
157 solution as a negative control. (E) Sequencing using Illumina MiSeq. (F) Analysis of the
158 taxonomic profiles among the DNA extraction protocols. Images form Togo Picture Gallery [39]
159 were used to create this figure.

160

161 Measurement of DNA yield

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

162 DNA yield for each method was evaluated by spectrophotometer (Nanodrop One, Thermo
163 Scientific). Analysis of the DNA yield from IndiSpin was omitted as nucleic acid is used as a
164 carrier for this kit. The mean DNA yields per 100 µL ThinPrep sample volume were compared.
165

166 16S rRNA marker gene sequencing

167 Controls and the extracted DNA were sent to Argonne National Laboratory (IL, USA) for
168 amplification and sequencing of the 16S rRNA gene on an Illumina MiSeq sequencing platform
169 [40]. The same volume of DNA was used for each reaction, and then normalized at the PCR
170 pooling step. This ensures that equal amounts of each amplified sample are added to the
171 sequencing pool. Paired-end reads from libraries with ~250-bp inserts were generated for the V4
172 region using the barcoded primer set: 515FB: 5'-GTGYCAGCMGCCGCGGTAA-3' and 806RB:
173 5'-GGACTACNVGGGTWTCTAAT-3' [41] [42] [43] [44] [45]. MiSeq Reagent Kit v2 (2×150
174 cycles, MS-102-2002) was used.
175

176 Sequence processing and analysis

177 Initial sequence processing and analyses were performed using QIIME 2 [46], any commands
178 prefixed by `q2-` are QIIME 2 plugins. After demultiplexing of the paired-end reads by `q2-`
179 `demux`, the imported sequence data was visually inspected via QIIME 2 View [47], to determine
180 the appropriate trimming and truncation parameters for generating Exact Sequence Variants
181 (ESVs) [48] via `q2-dada2` [49]. ESVs will be referred to as Operational Taxonomic Units
182 (OTUs). The forward reads were trimmed at 15 bp and truncated at 150 bp; reverse reads were
183 trimmed at 0 bp and truncated at 150 bp. The resulting OTUs were assigned taxonomy through
184 `q2-feature-classifier classify-sklearn`, by using a pre-trained classifier for the

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

185 amplicon region of interest [50]. This enables more robust taxonomic assignment of the OTUs
186 [51]. Taxonomy-based filtering was performed by using `q2-taxa filter-table` to remove
187 any OTUs that were classified as “Chloroplast”, “Mitochondria”, “Eukaryota”, “Unclassified”
188 and those that did not have at least a Phylum-level classification. We then performed additional
189 quality filtering via `q2-quality-control`, and only retained OTUs that had at least a 90%
190 identity and 90% query alignment to the SILVA reference set [52]. Then `q2-alignment` was
191 used to generate a *de novo* alignment with MAFFT [53] which was subsequently masked by
192 setting `max-gap-frequency 1 min-conservation 0.4`. Finally, `q2-phylogeny`
193 was used to construct a midpoint-rooted phylogenetic tree using IQ-TREE [54] with automatic
194 model selection using ModelFinder [55]. Unless specified, subsequent analyses were performed
195 after removing OTUs with a very low frequency of less than 0.0005% of the total data set in this
196 case [56].
197

198 Microbiome analysis

199 To compare the taxonomic profiles among four types of DNA extraction methods (Figure 1 &
200 Table 1), the following analyses were performed; (I) bacterial microbiome composition, (II)
201 detection of common and unique taxa, (III) alpha and beta diversity analysis, and (IV)
202 identification of specific bacteria retained per DNA extraction method.
203

204 Overall microbial composition was investigated at the family and genus taxonomic level. After
205 all count data of taxonomy were converted to relative abundance, the top 10 abundant taxonomic
206 groups in each family and genus level were plotted in colored bar plot [57] [58] [59]. Variation

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

207 of microbiome composition per DNA extraction method or per individual was assessed by the
208 Adonis test (`q2-diversity adonis`) [60] [61].
209
210 We set out to determine which microbial taxonomic groups were differentially accessible across
211 the sampling protocols by LEfSe analyses [62]. We further assessed the microbial taxa using
212 jvenn [63] at family and genus level. The Venn diagram was created after removing OTUs with a
213 frequency of less than 0.005% [56]. We used Scheffe test [64] to perform post-hoc analysis of
214 the LEfSe output.
215
216 Analytical approaches (at the OTU-level) that do not require the rarefying of data, such as `q2-`
217 `breakaway` [65] and Aitchison distance using `q2-deicode` [66] were used to determine both
218 alpha (within-sample) and beta (between-sample) diversity respectively. These were compared
219 with traditional methods, that often require the data to be rarefied. Here we applied the following
220 traditional alpha and beta-diversity metrics: Faith's Phylogenetic Diversity, Observed OTUs,
221 Shannon's diversity index, Pielou's Evenness, Unweighted UniFrac distance, Weighted UniFrac
222 distance, Jaccard distance, and Bray-Curtis distances via `q2-diversity` [46]. In order to
223 maintain a reasonable balance between sequencing depth and sample size, we determined that a
224 rarefaction depth of 51,197 reads allowed us to retain data for all four kits for 15 of the 20
225 individual patients. Overall, our subsequence analysis consisted of 3,071,820 reads (27.6%,
226 3,071,820 / 11,149,582 reads). All diversity measurements used in this study are listed in Table
227 S1.
228

229 ## Community type and HPV status

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

230 In addition to the analysis above, we tested whether the samples clustered by microbiome
231 composition were related to the patient's clinical and demographic characteristics such as,
232 cervical biopsy diagnosis, race, and HPV16 status. HPV16 status has been reported to be
233 associated with both racial differences as well as microbial community types [14] [67] [68] [69].
234 We employed the DMM [70] model to determine the number of community types for bacterial
235 cervical microbiome. Then, we clustered samples to the community type [9] [71]. Since vaginal
236 microbiota were reported to be clustered with different *Lactobacillus sp*. such as *L. crispatus*, *L.*
237 *gasseri*, *L. iners*, or *L. jensenii* [18] [72], we also collapsed the taxonomy to the species level and
238 performed a clustering analysis using "microbiome" R package [59]. We then determined which
239 bacterial taxa were differentially abundant among the patients with or without HPV16 via `q2-`
240 `aldex2` [73] and LEfSe [62].
241

242 General statistical analysis

243 All data are presented as means ± standard deviation (SD). Comparisons were conducted with
244 Fisher's exact test or Dunn's test with Benjamini-Hochberg-adjustment [74] or Wilcoxon test
245 with Benjamini-Hochberg-adjustment or pairwise PERMANOVA when appropriate. A p value <
246 0.05 or a q value < 0.05 was considered statistically significant. We did not control for
247 confounding variables such as socioeconomic status, nutrition, environmental exposures, or
248 similar factors.
249

250 Ethics approval and consent to participate

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

253

254 ## Consent for publication

255 Written informed consent for publication was obtained for all patients under IRB # 202790; NCT

256 # NCT02481414; IND # 15173.

257

258 ## Availability of data and materials

259 MIMARKS compliant [75] DNA sequencing data are available via the Sequence Read Archive

260 (SRA) at the National Center for Biotechnology Information (NCBI), under the BioProject

261 Accession: PRJNA598197.

262

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

263 # Results

264 ## Patient characteristics

265 The age of the patients (n = 20) was 31.4 ± 5.0 years. The distribution of race was 15% African
266 American (n = 3), 50% European descent (n = 10), and 35% Hispanic (n = 7). Cervical histology
267 was 40% CIN2 (n = 8), 50% CIN3 (n = 10), and 10% benign (n = 2). HPV genotypes were 50%
268 HPV16 positive (n = 10), 10% HPV18 positive (n = 2), 90% HR-HPV positives (n = 18), 45%
269 LR-HPV positives (n = 9), and 75% multiple HPV positives (n = 15). Patient characteristics were
270 summarized in Table 2.
271

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)

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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)

272

273 **DNA yield**

274 DNA yield per 100 μL ThinPrep solution was 0.09 ± 0.06 μg in ZymoBIOMICS, 0.04 ± 0.01 μg
275 in PowerFecalPro, and 0.21 ± 0.23 μg in QIAampMini. DNA yield was not calculated for
276 IndiSpin, as Poly-A Carrier DNA was used. The DNA yield of PowerFecalPro was significantly
277 lower than that of ZymoBIOMICS (adjusted p value < 0.001) and QIAampMini (adjusted p
278 value < 0.001) based on Dunn's test with Benjamini-Hochberg-adjustment (Figure S1).
279

280 **Mock and Negative Controls**

281 We observed that we were able to reasonably recover the expected taxa of our mock community
282 positive control from the American Type Culture Collection (ATCC MSA1007). Each of the
283 following taxa should have a relative abundance of ~16.7% of the total sample. It should be
284 noted that factors, such as sample preparation and primer biases, can cause deviations from the
285 expected mock community [76] [77] [78]. We observed: 7.184% *Gardnerella spp*., 14.807%
286 *Lactobacillus jensenii* (40.185% *Lactobacillus spp.*), 16.530% *Mycoplasma hominis*, 14.311%
287 *Prevotella bivia* (14.327 % *Prevotella spp.*), and 21.429% *Streptococcus agalactiae* (21.449%
288 *Streptococcus spp*.). A total of 127,193 reads were generated from the mock community control.
289 Of which, 99.68% (126,783 / 127,193 reads) were from expected members of the mock

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

290 community. For the negative control (ThinPrep preservation solution) only 1,791 reads were
291 generated. 1,400 reads were from *Staphylococcus spp.*, 323 reads were from *Micrococcus spp.*,
292 and 47 reads were *Lactobacillus spp.*. The remaining 21 reads were spurious.
293

294 Number of reads and Operational Taxonomic Units (OTUs) before 295 rarefying

296 We obtained a total of 11,149,582 reads for 80 DNA extractions. IndiSpin (168,349 ± 57,451
297 reads) produced a significantly higher number of reads compared to PowerFecalPro (115,610 ±
298 68,201 reads, p value = 0.020, Dunn's test with Benjamini-Hochberg-adjustment) as shown in
299 Table 3 Approximately 90% of reads were assigned to gram-positive bacteria and about 10% of
300 reads were assigned to gram-negative bacteria across all kits.
301

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	All	ZY	2,705,044 (135,252 ± 66,011)		[a]
(mean ± SD)		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%	

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

	GN	ZY	274,664 (13,733 ± 29,162)	10.2%	[c]
		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)		[d]
		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:

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

all bacteria, GP: gram-positive bacteria, GN: gram-negative bacteria.

302

303 Prior to rarefying, the ZymoBIOMICS kit captured a greater representation of gram-
304 negative bacterial OTUs (total 346, 17.3 ± 9.8) compared to PowerFecalPro (total 209, 10.5 ±
305 10.3, p value = 0.012, Dunn's test with Benjamini-Hochberg-adjustment, ratio of gram-negative
306 bacteria: 41.9% vs 33.7%) as shown in Table 3. No significant differences in the number of
307 OTUs before rarefying was detected for the entire bacterial community or gram-positive bacteria.
308

309 Microbiome composition per DNA extraction protocol

310 We analyzed whether differences in DNA extraction methods affect our ability to assess cervical
311 microbiota composition. The patients can be identified by whether or not they displayed a
312 *Lactobacillus*-dominant community type (Figures 2A & S4). Variation between patients was a
313 significantly greater influence on the observed microbial composition than was the method of
314 DNA extraction (F.Model: 199.4, R2: 0.982, and p value: 0.001 for patients vs F.Model: 2.9, R2:
315 0.003, and p value: 0.002 for DNA extraction, Adonis test, Figure 2A).
316

317 **Figure 2. Taxonomic resolution among DNA extraction protocols.** (A) Relative abundance of
318 microbe at family level (left) and genus level (right) per DNA extraction method showed the
319 pattern that variance of microbe composition per patient was higher than that per DNA extraction
320 protocol. These patterns were confirmed by values of Adonis test (`q2-diversity adonis`);
321 F.Model: 199.4, R2: 0.982, and p value: 0.001 for patients and F.Model: 2.9, R2: 0.003, and p
322 value: 0.002 for DNA extraction [60] [61]. After all count data of taxonomy were converted to
323 relative abundance as shown in the y-axis, the top ten taxonomy at each family and genus level

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

324 were plotted in colored bar plot and other relatively few taxonomies were not plotted. The 20
325 patients ID were described in the x-axis. (B) Venn diagrams, considering only those OTUs with
326 a frequency greater than 0.005% shown, revealed that ZymoBIOMICS had four unique taxa at
327 family (left) and genus (right) taxonomic level. Thirty-one of 41 families and 45 of 57 genera
328 were detected with all DNA extraction protocols.

329

330 The following top 10 abundant families are shown in Figure 2A (left) and constituted
331 approximately 95.7% of cervical bacteria in all kits (80 DNA extractions); *Lactobacillaceae*
332 (58.9%), *Bifidobacteriaceae* (13.7%), *Veillonellaceae* (4.8%), *Prevotellaceae* (4.3%), *Family XI*
333 (3.9%), *Atopobiaceae* (3.0%), *Leptotrichiaceae* (2.5%), *Streptococcaceae* (2.0%),
334 *Lachnospiraceae* (1.6%). *Ruminococcaceae* (0.9%). The following top 10 abundant genera are
335 shown in Figure 2A (right) and constituted approximately 92% of cervical bacteria;
336 *Lactobacillus* (58.9%), *Gardnerella* (13.6%), *Prevotella* (4.2%), *Megasphaera* (3.7%),
337 *Atopobium* (3.0%), *Sneathia* (2.5%), *Streptococcus* (1.9%), *Parvimonas* (1.7%), *Shuttleworthia*
338 (1.4%), and *Anaerococcus* (1.1%).

339

340 Shared and unique microbiota among DNA extraction protocols

341 All DNA extraction methods were generally commensurate with one another, there were
342 31 of 41 shared microbes at the family level (Figure 2B left) and 45 of 57 shared microbes at the
343 genus level (Figure 2B right) among the DNA extraction protocols.

344 However, four gram-negative taxa were uniquely detected by ZymoBIOMICS and one
345 taxon was uniquely detected by QIAampMini both at the genus level (Figure 2B right). Of the
346 uniquely detected ZymoBIOMICS OTUs, *Methylobacterium* was detected in 5 of the 80 DNA

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

347 extractions, consisting of 912 reads: 0.01% of all kit extractions. A member of this genus,
348 *Methylobacterium aerolatum*, has been reported to be more abundant in the endocervix than the
349 vagina of healthy South African women [80]. *Bacteroidetes*, which are often reported as
350 enriched taxa in an HIV positive cervical environment [81], was detected in 12 of the 80 DNA
351 extractions (1,028 reads; 0.01%). *Meiothermus* was detected in 9 of the 80 DNA extractions (882
352 reads; 0.01%) and *Hydrogenophilus* was detected in 14 of the 80 DNA extractions (2,488 reads,
353 0.02%). *Meiothermus* and *Hydrogenophilus* [82] are not considered to reside within the human
354 environment, and are likely kit contaminants, as previously reported [83]. A unique gram-
355 positive taxon obtained from the QIAampMini, *Streptomyces*, which was reported to be detected
356 from the cervicovaginal environment in the study of Kenyan women [84], was detected in all 20
357 of the QIAampMini DNA extractions (6,862 reads; 0.06%). No unique taxa were detected in
358 PowerFecalPro and IndiSpin. Although less than 0.005% of the total data set, two samples of
359 IndiSpin also detected potential kit contaminant, *Tepidiphilus* (*Hydrogenophilaceae*).
360

361 **Alpha and beta diversity**

362 The observed alpha diversity was similar across all kits except for a few cases (Figure 3).
363 Significantly higher species richness (`q2-breakaway`) was observed between the
364 ZymoBIOMICS (56.1 ± 19.4) protocol and that of PowerFecalPro (43.2 ± 32.9, $p = 0.025$)
365 (Figure 3). Similarly, Faith's Phylogenetic Diversity was observed to be higher with the
366 ZymoBIOMICS protocol (6.6 ± 2.2), compared to PowerFecalPro (4.5 ± 1.9, $p = 0.012$). The use
367 of IndiSpin also resulted significantly higher alpha diversity than that of PowerFecalPro in an
368 analysis of Species richness ($p = 0.042$). Non-phylogenetic alpha diversity metrics such as

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

369 Observed OTUs, Shannon's diversity index, and Pielou's Evenness did not show differences
370 among the four methods.
371
372 **Figure 3. Comparisons of alpha diversity between different DNA extraction protocols.** The
373 alpha diversity indices determined by Species richness and Phylogenetic diversity are
374 significantly higher with ZymoBIOMICS in comparison with PowerFecalPro ($p = 0.025$ and
375 0.012, respectively, Dunn's test with Benjamini-Hochberg-adjustment). IndiSpin also showed
376 significantly higher diversity than that of PowerFecalPro using analysis of Species richness ($p =$
377 0.042, Dunn's test with Benjamini-Hochberg-adjustment). No significant differences were
378 observed in other alpha diversity indexes such as observed OTUs, Shannon's diversity index,
379 and Pielou's Evenness. Zy: ZymoBIOMICS DNA Miniprep Kit, Pro: QIAamp PowerFecal Pro
380 DNA Kit, QIA: QIAamp DNA Mini Kit, IN: IndiSpin Pathogen Kit.
381
382 Similar to the alpha diversity results above, no significant differences were observed with
383 other metrics, including `q2-deicode` (Aichison distances). Only qualitative metrics such as
384 Unweighted UniFrac and Jaccard distance, revealed significant differences in a few cases (Table
385 4 Figure S2, & Figure S3). Most observed differences were between ZymoBIOMICS and other
386 DNA extraction methods with when qualitative metrics such as Unweighted UniFrac
387 (PowerFecalPro: $q = 0.002$; QIAampMini: $q = 0.002$; and IndiSpin: $q = 0.002$) and Jaccard
388 distances (QIAampMini: $q = 0.018$ and IndiSpin: $q = 0.033$) were used. With PowerFecalPro vs.
389 IndiSpin in Unweighted UniFrac ($q = 0.023$), being the only other observed significant
390 difference. All other metrics showed no significance differences with regard to beta diversity.
391

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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	IndiSpin	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

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

		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.

392

393 Differential accessibility of microbiota by DNA extraction protocol

394 Linear discriminant analysis (LDA) effect size (LEfSe) analysis [62], identified several
395 taxonomic groups, defined with an LDA score of 2 or higher *(one-against-all*), for differential
396 accessibility by extraction kit: 23 in ZymoBIOMICS, 0 in PowerFecalPro, 3 in QIAampMini,
397 and 3 in IndiSpin (Figure 4A). The following taxa were found to be highly accessible (LDA
398 score > 3) with the use of the ZymoBIOMICS kit: Phylum *Proteobacteria*, Class
399 *Gammaproteobacteria*, Order *Betaproteobacteriales*, Family *Bacillaceae*, and Genus
400 *Anoxybacillus*. Whereas the Order *Streptomycetales* was highly enriched with the use of the
401 QIAampMini (LDA score > 3). However, post-hoc analysis of the LEfSe output, using Scheffe
402 test [64] revealed that only the contaminants *Meiothermus* and *Hydrogenophilus* were enriched
403 with Zymo, and *Streptomyces* was enriched in QIA (Figure 4A). These results reveal minimal to
404 no significant enrichment of specific microbiota across extraction kits.

405

406 **Figure 4. Distinct detections of microbe among the DNA extraction protocols.** (A) A bar
407 graph showing 23 significantly enriched taxa with ZymoBIOMICS, 3 with QIAamp DNA Mini
408 Kit, and 3 with IndiSpin Pathogen Kit determined by the linear discriminant analysis (LDA)

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

409 effect size (LEfSe) analyses [62]. Asterisks denote taxa of genus level that were significant after
410 post-hoc significant testing with Scheffe. (B) A taxonomic cladogram from the same LEfSe
411 analyses showing that the significantly enriched microbiota in ZymoBIOMICS were composed
412 of phylum *Proteobacteria*. Also note that *Meiothermus* (a member of the phylum *Deinococcus-*
413 *Thermus*) *Hydrogenophilaceae* (a member of the phylum *Proteobacteria*), and *Hydrogenophilus*
414 (a member of the phylum *Proteobacteria*) are likely an extraction kit contaminant. Zy:
415 ZymoBIOMICS DNA Miniprep Kit, Pro: QIAamp PowerFecal Pro DNA Kit, QIA: QIAamp
416 DNA Mini Kit, IN: IndiSpin Pathogen Kit. g_: genus, f_: family, o_: order, c_: class, p_: phylum.
417

418 ## Microbial community type and HPV16

419 Dirichlet Multinomial Mixtures (DMM) model [70] detected two cervical microbial community
420 types across all four DNA extraction protocols (Figure S4). Community type I was composed of
421 the following: *Gardnerella sp.* (ZymoBIOMICS: 17.1%; PowerFecalPro: 20%; QIAampMini:
422 23%; IndiSpin: 20%), *Lactobacillus iners* (ZymoBIOMICS: 6.3%; PowerFecalPro: 5%;
423 QIAampMini: 6%; IndiSpin: 5%), *Atopobium vaginae* [10] (ZymoBIOMICS: 3.5%;
424 PowerFecalPro: 3%; QIAampMini: 4%; IndiSpin: 5%), *Clamydia trachomatis* (ZymoBIOMICS:
425 1.9%; PowerFecalPro: 2%; QIAampMini: 3%; IndiSpin: 2%), *Shuttleworthia sp.*
426 (ZymoBIOMICS: 1.8%; PowerFecalPro: 2%; QIAampMini: 2%; IndiSpin: 2%). Some members
427 of *Shuttleworthia* are considered to be bacterial vaginosis□associated bacterium (BVAB) [85],
428 further investigation is required to determine if this OTU is indeed a BVAB. We determined this
429 community type “high diversity type”. Community type II was is dominated by *Lactobacillus*
430 *iners* at 88%, 85%, 83%, and 85% respectively for ZymoBIOMICS, PowerFecalPro,
431 QIAampMini, and IndiSpin.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

432 The relationship between HPV16 infection and community type was observed to be
433 significantly associated with community type I (HPV16 positive patients [n = 9], HPV16
434 negative patients [n = 1]) and not community type II (HPV16 positive patients [n = 1], HPV16
435 negative patients [n = 9], $p = 0.001$, Fisher's exact test) regardless of the DNA extraction kit used
436 (Figure S4A). In support of this result, analysis of differentially abundant microbiota using `q2-`
437 `aldex` (Benjamini-Hochberg corrected p value of Wilcoxon test: $p < 0.001$, standardized
438 distributional effect size: −1.2) revealed that *Lactobacillus iners* were differentially enriched in
439 the cervical environment without HPV16. LEfSe analysis also detected that genus *Lactobacillus*
440 were enriched in the cervical environment without HPV16 ($p < 0.001$, LDA score: 5.38, Figure
441 S4B). No significant differences were observed in the relationship between community type and
442 HPV18 ($p = 0.474$, Fisher's exact test), HR-HPV ($p = 0.474$, Fisher's exact test), LR-HPV ($p =$
443 0.370, Fisher's exact test), multiple HPV infections ($p = 0.303$, Fisher's exact test), results of
444 cervical biopsy ($p = 0.554$, Fisher's exact test), and race (African Americans vs not-African
445 Americans: $p = 1$; European descent vs non-European descent: $p = 0.656$; Hispanic vs non-
446 Hispanic: $p = 0.350$, Fisher's exact test, Figure S4A).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

447 # Discussion

448 In this study, we evaluated the utility of LBC specimens for the collection and storage of cervical
449 samples for microbiome surveys based on the 16S rRNA marker gene. We simultaneously
450 compared the efficacy of several commonly used DNA extraction protocols on these samples in
451 an effort to develop a standard operating procedure/protocol (SOP) for such work. We've also
452 been able to show that there are two cervical microbial community types, which are associated
453 with the dominance or non-dominance of *Lactobacillis iners* and HPV16 status (Figures 2A &
454 S4A). The relationship between community types and HPV16 was detected regardless of the
455 DNA extraction protocol used.

456 This study evaluated the composition of microbiota accessible across all DNA extraction
457 methods. All kits were commensurate in their ability to capture the microbial composition of
458 each patient and the two observed cervical microbial community state types, making all of these
459 protocols viable for discovering broad patterns of microbial diversity. It should be noted,
460 however, that a singular kit should be used through the entirety of a study to minimize any subtle
461 differences between samples, particularly when qualitative or richness-based diversity metrics
462 are used. We detected potential DNA contamination with the ZymoBIOMICS and IndiSpin kits.
463 The number of OTUs prior to rarefying revealed that the ZymoBIOMICS protocol detected more
464 gram-negative OTUs than the PowerFecalPro (Figure 2B & Table 3). In particular, LEfSe
465 analysis has shown that phylum *Proteobacteria* can be better detected with the ZymoBIOMICS
466 kit (Figure 4). This signature was no longer observed after *post hoc* testing.

467 Although rarefying microbiome data can be problematic [86], it can still provide robust
468 and interpretable results for diversity analysis [87], we were able to observe commensurate
469 findings with non-rarefying approaches such as `q2-breakaway` [65], `q2-deicode` [66], and

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

470 LEfSe [62]. Beta-diversity analysis via Unweighted UniFrac also revealed that ZymoBIOMICS
471 was significantly different from all other kits (Table 4). There were no differences in non-
472 phylogenetic indices of alpha diversity (Figure 3). These findings lead us to surmise that
473 qualitative metrics are more sensitive to differences between extraction kits, while quantitative
474 metrics were more sensitive to differences between subject (Figures S2 & S3).

475 Although we hypothesized that the detection of difficult-to-lyse-bacteria (*e.g*. gram-
476 positive bacteria) would vary by kit, we observed no significant differences (Table 3). The
477 number of reads of gram-positive and gram-negative bacteria also showed that there was no
478 difference in the four kits (Table 3). This is likely due to several modifications made to the
479 extraction protocol as outlined in Table 1. That is, we added bead beating and mutanolysin to the
480 QIAampMini protocol [36]. We also modified the beating time of the ZymoBIOMICS kit down
481 to 2 minutes from 10 minutes (the latter being recommended by the manufacturer) to minimize
482 DNA shearing. We may use the extracted DNA from ZymoBIOMICS for long-read amplicon
483 sequencing platforms such as PacBio (Pacific Biosciences of California, Inc) [88] or MinION
484 (Oxford Nanopore Technologies) [89] [90]. Excessive shearing can render these samples
485 unusable for long-read sequencing. It is quite possible that we could have observed even more
486 diversity with the ZymoBIOMICS kit for our amplicon survey if we conducted bead-beating for
487 the full 10 minutes.

488 One limitation of our study is the lack of fresh LBC samples that would have enabled
489 assessment the effects of prolonged storage on determining microbial community composition
490 due to potential DNA degradation [25]. We think this may be unlikely, as our LBC samples
491 were immediately frozen in -80□C, and DNA degradation within LBC samples stored at -80□C
492 has been shown to be minimal [26]. However, the possibility that the observed microbial

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

493 community composition may not be indicative of the community at the time of sampling remains.
494 Despite this, our observations are commensurate with several prior studies in this area as
495 outlined below. Community typing and detection of the differentially abundant microbiota
496 revealed that *Lactobacillus iners* were more abundant in the cervical ecosystem without HPV16
497 (Figure S4). These findings are congruent with those of, Usyk *et al*. [91], Lee *et al*. [1], and
498 Audirac-Chalifour *et al*. [92]. Usyk *et al*., reported that *L. iners* was associated with clearance of
499 HR-HPV infections [91]. Lee *et al*. reported that *L. iners* were decreased in HPV positive women
500 [1]. Also, the results indicated that the proportion of *L. iners* was higher in HPV-negative women
501 compared to HPV-positive women (relative abundance 14.9% vs 2.1%) was reported by
502 Audirac-Chalifour *et al* [92]. Similarly, Tuominen *et al*. [20] reported that *L. iners* were enriched
503 in HPV negative samples (relative abundance: 47.7%) compared to HPV positive samples
504 (relative abundance: 18.6%, p value = 0.07) in the study of HPV positive-pregnant women
505 (HPV16 positive rate: 15%) [93]. As established by the seminal study of Ranjeva *et al*. [94], a
506 statistical model revealed that colonization of specific HPV types including multi-HPV type
507 infection depends on host-risk factors such as sexual behavior, race and ethnicity, and smoking.
508 It is unclear whether the association between the cervical microbiome, host-specific traits, and
509 persistent infection of specific HPV types, such as HPV16, can be generalized and requires
510 further investigation.

511 We focused on LBC samples as this is the recommended method of storage for cervical
512 cytology [95]. We used a sample volume of 200 or 300 μL ThinPrep solution in this study. The
513 Linear Array HPV Genotyping Test (Roche Diagnostics) stably detects β-globin with a base
514 length of 268 bp as a positive control. Therefore, using a similar sample volume as HPV
515 genotyping (250 μL), it was expected that V4 (250 bp), which is near the base length of β-globin,

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

516 would be PCR amplified. It has been pointed out by Ling *et al*. [96] that the cervical
517 environment is of low microbial biomass. To control reagent DNA contamination and estimate
518 the sample volume, DNA quantification by qPCR before sequencing is recommended [97]. Mitra
519 *et al* determined a sample volume of 500 μL for ThinPrep by qPCR in the cervical microbiome
520 study comparing sampling methods using cytobrush or swab [21]. The average storage period
521 from sample collection via LBC to DNA extraction was about two years in this study. Kim *et al*.
522 reported that DNA from the cervix stored in ThinPrep at room temperature or −80°C was stable
523 for at least one year [26]. Meanwhile, Castle *et al*. reported that β-globin DNA fragments of 268
524 bases or more were detected by PCR in 90 % (27 of 30 samples) of ThinPrep samples stored for
525 eight years at an uncontrolled ambient temperature followed by a controlled ambient
526 environment (10–26.7°C) [25]. Low-temperature storage may allow the analysis of the short
527 DNA fragments of the V4 region after even long-term storage, although further research is
528 needed to confirm the optimal storage period in cervical microbiome studies using ThinPrep.
529 SurePath LBC specimens are as widely used as ThinPrep, but the presence of formaldehyde
530 within the SurePath preservation solution raises concerns about accessing enough DNA for
531 analysis as compared to ThinPrep, which contains methanol [98] [99]. It should also be noted
532 that other storage solutions, *i.e*., those using guanidine thiocyanate have been reported for
533 microbiome surveys of the cervix [100] and feces [101]. A weakness of the current study is that
534 we did not examine the reproducibility of our results as each sample was extracted using each kit
535 once as samples were limited in quantity, and we lacked fresh sample controls to assess the
536 effects of prolonged storage to alter microbial community composition. Although several studies,
537 have shown general stability and accessibility of DNA [26] [102] [103], there is potential for

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

538 DNA degradation for samples not stored at low temperatures [25] [26]. However, the use of

539 actual patient samples rather than mock samples is a strength of our approach.

540

541

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

542 # Conclusions

543 In conclusion, regardless of the extraction protocol used, all kits provided equivalent broad
544 accessibility to the cervical microbiome. Observed differences in microbial composition were
545 due to the significant influence of the individual patient and not the extraction protocol. We have
546 shown that the ability to characterize cervical microbiota from LBC specimens is possible, we
547 were limited in our ability to directly assess if the observed microbial community composition
548 would reflect that of a fresh sample. Despite this limitation, we were able to assess the
549 relationship between HPV and the cervical microbiome, also supported by Kim *et al*. [26] and
550 Castle *et al* [25]. Cervical microbiome in patients with HPV16 or HPV18 which causes 70% of
551 cervical cancers and CIN [104] warrants critical future study. Selection and characterization of
552 appropriate DNA extraction methods are important for providing an accurate census of cervical
553 microbiota and the human microbiome in general [27] [28] [29] [36] [26] [25]. Although we
554 found all four extraction kits to be commensurate in their ability to broadly characterize the CM,
555 one singular kit should be used throughout the entirety of a given study. This study lends support
556 to the view that the selection of a DNA extraction kit depends on the questions asked of the data,
557 and should be taken into account for any cervicovaginal microbiome and HPV research that
558 leverages LBC specimens for use in clinical practice [17] [105].

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

559 # Acknowledgements

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

561

562 # Competing interests

563 M.N. is one of the inventors named in the patents and patent applications for the HPV
564 therapeutic vaccine PepCan. Patent issued: Human Papilloma Virus Therapeutic Vaccine
565 Nakagawa, M. and Chang, B.S. Patent No. 9,974,849 issued on 5/22/2018
566 Patent application: Human Papilloma Virus Therapeutic Vaccine Nakagawa, M. and Chang, B.S.
567 International Application (PCT/US14/60198) filed on 10/11/2014. The remaining authors declare
568 no conflicts of interest.

569

570 # Funding

571 This work was supported by the National Institutes of Health (R01CA143130, USA), Drs. Mae
572 and Anderson Nettleship Endowed Chair of Oncologic Pathology (31005156, USA), and the
573 Arkansas Biosciences Institute (the major component of the Tobacco Settlement Proceeds Act of
574 2000, AWD00052249, USA) warded to M.N.

575

576 # Authors' contributions

577 M.N. designed and supervised this project. Ta.S. and M.S.R. conducted bioinformatics analysis
578 and wrote paper. Ta.S., H.C., and M.N. created the protocol of DNA extraction. M.N., H.C., S.O.,
579 W.G., and To.S. provided important feedback. Samples in the clinical trial were collected by

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

580 W.G. and his associates. DNA extraction was conducted by Ta.S. Sequencing of 16S RNA gene

581 was conducted by S.O.

582

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

583 # References

584 1. Lee JE, Lee S, Lee H, Song YM, Lee K, Han MJ, et al. Association of the vaginal
585 microbiota with human papillomavirus infection in a Korean twin cohort. PLoS One.
586 2013;8(5):e63514. Epub 2013/05/30. doi: 10.1371/journal.pone.0063514. PubMed PMID:
587 23717441; PubMed Central PMCID: PMCPMC3661536.
588 2. Huang X, Li C, Li F, Zhao J, Wan X, Wang K. Cervicovaginal microbiota composition
589 correlates with the acquisition of high-risk human papillomavirus types. Int J Cancer.
590 2018;143(3):621-34. Epub 2018/02/27. doi: 10.1002/ijc.31342. PubMed PMID: 29479697.
591 3. Zhou Y, Wang L, Pei F, Ji M, Zhang F, Sun Y, et al. Patients With LR-HPV Infection
592 Have a Distinct Vaginal Microbiota in Comparison With Healthy Controls. Front Cell Infect
593 Microbiol. 2019;9:294. Epub 2019/09/27. doi: 10.3389/fcimb.2019.00294. PubMed PMID:
594 31555603; PubMed Central PMCID: PMCPMC6722871.
595 4. Onywera H, Williamson AL, Mbulawa ZZA, Coetzee D, Meiring TL. The cervical
596 microbiota in reproductive-age South African women with and without human papillomavirus
597 infection. Papillomavirus Res. 2019;7:154-63. Epub 2019/04/16. doi: 10.1016/j.pvr.2019.04.006.
598 PubMed PMID: 30986570; PubMed Central PMCID: PMCPMC6475661.
599 5. Brotman RM, Shardell MD, Gajer P, Tracy JK, Zenilman JM, Ravel J, et al. Interplay
600 between the temporal dynamics of the vaginal microbiota and human papillomavirus detection. J
601 Infect Dis. 2014;210(11):1723-33. Epub 2014/06/20. doi: 10.1093/infdis/jiu330. PubMed PMID:
602 24943724; PubMed Central PMCID: PMCPMC4296189.
603 6. Godoy-Vitorino F, Romaguera J, Zhao C, Vargas-Robles D, Ortiz-Morales G, Vazquez-
604 Sanchez F, et al. Cervicovaginal Fungi and Bacteria Associated With Cervical Intraepithelial
605 Neoplasia and High-Risk Human Papillomavirus Infections in a Hispanic Population. Front
606 Microbiol. 2018;9:2533. Epub 2018/11/09. doi: 10.3389/fmicb.2018.02533. PubMed PMID:
607 30405584; PubMed Central PMCID: PMCPMC6208322.
608 7. Łaniewski P, Barnes D, Goulder A, Cui H, Roe DJ, Chase DM, et al. Linking
609 cervicovaginal immune signatures, HPV and microbiota composition in cervical carcinogenesis
610 in non-Hispanic and Hispanic women. Sci Rep. 82018.
611 8. Mitra A, MacIntyre DA, Lee YS, Smith A, Marchesi JR, Lehne B, et al. Cervical
612 intraepithelial neoplasia disease progression is associated with increased vaginal microbiome
613 diversity. Sci Rep. 2015;5:16865. Epub 2015/11/18. doi: 10.1038/srep16865. PubMed PMID:
614 26574055; PubMed Central PMCID: PMCPMC4648063.
615 9. Piyathilake CJ, Ollberding NJ, Kumar R, Macaluso M, Alvarez RD, Morrow CD.
616 Cervical Microbiota Associated with Higher Grade Cervical Intraepithelial Neoplasia in Women
617 Infected with High-Risk Human Papillomaviruses. Cancer Prev Res (Phila). 2016;9(5):357-66.
618 Epub 2016/03/05. doi: 10.1158/1940-6207.CAPR-15-0350. PubMed PMID: 26935422; PubMed
619 Central PMCID: PMCPMC4869983.
620 10. Oh HY, Kim BS, Seo SS, Kong JS, Lee JK, Park SY, et al. The association of uterine
621 cervical microbiota with an increased risk for cervical intraepithelial neoplasia in Korea. Clin
622 Microbiol Infect. 2015;21(7):674 e1-9. Epub 2015/03/11. doi: 10.1016/j.cmi.2015.02.026.
623 PubMed PMID: 25752224.
624 11. De Seta F, Campisciano G, Zanotta N, Ricci G, Comar M. The Vaginal Community State
625 Types Microbiome-Immune Network as Key Factor for Bacterial Vaginosis and Aerobic

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

626 Vaginitis. Front Microbiol. 2019;10:2451. Epub 2019/11/19. doi: 10.3389/fmicb.2019.02451.
627 PubMed PMID: 31736898; PubMed Central PMCID: PMCPMC6831638.
628 12. Oliver A, LaMere B, Weihe C, Wandro S, Lindsay KL, Wadhwa PD, et al.
629 Cervicovaginal microbiome composition drives metabolic profiles in healthy pregnancy. bioRxiv
630 https://doi.org/10.1101/840520. 2019.
631 13. Firwana B, Avaritt N, Shields B, Ravilla R, Makhoul I, Hutchins L, et al. Do checkpoint
632 inhibitors rely on gut microbiota to fight cancer? J Oncol Pharm Pract. 2018;24(6):468-72.
633 14. Ravilla R, Coleman HN, Chow CE, Chan L, Fuhrman BJ, Greenfield WW, et al. Cervical
634 Microbiome and Response to a Human Papillomavirus Therapeutic Vaccine for Treating High-
635 Grade Cervical Squamous Intraepithelial Lesion. Integr Cancer Ther.
636 2019;18:1534735419893063. Epub 2019/12/14. doi: 10.1177/1534735419893063. PubMed
637 PMID: 31833799; PubMed Central PMCID: PMCPMC6913049.
638 15. Human Microbiome Project C. Structure, function and diversity of the healthy human
639 microbiome. Nature. 2012;486(7402):207-14. Epub 2012/06/16. doi: 10.1038/nature11234.
640 PubMed PMID: 22699609; PubMed Central PMCID: PMCPMC3564958.
641 16. Bik EM, Bird SW, Bustamante JP, Leon LE, Nieto PA, Addae K, et al. A novel
642 sequencing-based vaginal health assay combining self-sampling, HPV detection and genotyping,
643 STI detection, and vaginal microbiome analysis. PLoS One. 2019;14(5):e0215945. Epub
644 2019/05/03. doi: 10.1371/journal.pone.0215945. PubMed PMID: 31042762; PubMed Central
645 PMCID: PMCPMC6493738 have received stock options as well as other compensation. Some
646 authors have patents pending in relation to this work: US Application No 15/198,818, Method
647 and system for diagnostic testing, Application No 16/084,945, Method and system for
648 microbiome-derived diagnostics and therapeutics for bacterial vaginosis, and Application No
649 16/115,542, Method and system for characterization for female reproductive system-related
650 conditions associated with microorganisms. The data in this article were used in the development
651 of a commercially available test product developed and marketed by uBiome. This does not alter
652 our adherence to PLOS ONE policies on sharing data and materials.
653 17. Berman HL, McLaren MR, Callahan BJ. Understanding and interpreting community
654 sequencing measurements of the vaginal microbiome. BJOG. 2020;127(2):139-46. Epub
655 2019/10/10. doi: 10.1111/1471-0528.15978. PubMed PMID: 31597208.
656 18. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, et al. Vaginal
657 microbiome of reproductive-age women. Proc Natl Acad Sci U S A. 2011;108 Suppl 1:4680-7.
658 Epub 2010/06/11. doi: 10.1073/pnas.1002611107. PubMed PMID: 20534435; PubMed Central
659 PMCID: PMCPMC3063603.
660 19. Fettweis JM, Serrano MG, Brooks JP, Edwards DJ, Girerd PH, Parikh HI, et al. The
661 vaginal microbiome and preterm birth. Nat Med. 2019;25(6):1012-21. Epub 2019/05/31. doi:
662 10.1038/s41591-019-0450-2. PubMed PMID: 31142849; PubMed Central PMCID:
663 PMCPMC6750801.
664 20. Tuominen H, Rautava S, Syrjanen S, Collado MC, Rautava J. HPV infection and
665 bacterial microbiota in the placenta, uterine cervix and oral mucosa. Sci Rep. 2018;8(1):9787.
666 Epub 2018/06/30. doi: 10.1038/s41598-018-27980-3. PubMed PMID: 29955075; PubMed
667 Central PMCID: PMCPMC6023934.
668 21. Mitra A, MacIntyre DA, Mahajan V, Lee YS, Smith A, Marchesi JR, et al. Comparison
669 of vaginal microbiota sampling techniques: cytobrush versus swab. Sci Rep. 2017;7(1):9802.
670 Epub 2017/08/31. doi: 10.1038/s41598-017-09844-4. PubMed PMID: 28852043; PubMed
671 Central PMCID: PMCPMC5575119.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

672 22. Bentz JS. Liquid-based cytology for cervical cancer screening. Expert Rev Mol Diagn.
673 2005;5(6):857-71. Epub 2005/11/01. doi: 10.1586/14737159.5.6.857. PubMed PMID: 16255628.
674 23. Gibb RK, Martens MG. The impact of liquid-based cytology in decreasing the incidence
675 of cervical cancer. Rev Obstet Gynecol. 2011;4(Suppl 1):S2-S11. Epub 2011/05/28. PubMed
676 PMID: 21617785; PubMed Central PMCID: PMCPMC3101960.
677 24. Donders GG, Depuydt CE, Bogers JP, Vereecken AJ. Association of Trichomonas
678 vaginalis and cytological abnormalities of the cervix in low risk women. PLoS One.
679 2013;8(12):e86266. Epub 2014/01/05. doi: 10.1371/journal.pone.0086266. PubMed PMID:
680 24386492; PubMed Central PMCID: PMCPMC3875579.
681 25. Castle PE, Solomon D, Hildesheim A, Herrero R, Concepcion Bratti M, Sherman ME, et
682 al. Stability of archived liquid-based cervical cytologic specimens. Cancer. 2003;99(2):89-96.
683 Epub 2003/04/22. doi: 10.1002/cncr.11058. PubMed PMID: 12704688.
684 26. Kim Y, Choi KR, Chae MJ, Shin BK, Kim HK, Kim A, et al. Stability of DNA, RNA,
685 cytomorphology, and immunoantigenicity in Residual ThinPrep Specimens. APMIS.
686 2013;121(11):1064-72. Epub 2013/04/10. doi: 10.1111/apm.12082. PubMed PMID: 23566220.
687 27. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, et al. Towards
688 standards for human fecal sample processing in metagenomic studies. Nat Biotechnol.
689 2017;35(11):1069-76. Epub 2017/10/03. doi: 10.1038/nbt.3960. PubMed PMID: 28967887.
690 28. Stinson LF, Keelan JA, Payne MS. Comparison of Meconium DNA Extraction Methods
691 for Use in Microbiome Studies. Front Microbiol. 2018;9:270. Epub 2018/03/09. doi:
692 10.3389/fmicb.2018.00270. PubMed PMID: 29515550; PubMed Central PMCID:
693 PMCPMC5826226.
694 29. Teng F, Darveekaran Nair SS, Zhu P, Li S, Huang S, Li X, et al. Impact of DNA
695 extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling. Sci
696 Rep. 2018;8(1):16321. Epub 2018/11/07. doi: 10.1038/s41598-018-34294-x. PubMed PMID:
697 30397210; PubMed Central PMCID: PMCPMC6218491.
698 30. Roche Molecular Diagnostics. LINEAR ARRAY® HPV Genotyping.
699 https://diagnostics.roche.com/global/en/products/params/linear-array-hpv-genotyping.html.
700 Accessed 12 Mar 2020.
701 31. de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H. Classification of
702 papillomaviruses. Virology. 2004;324(1):17-27. Epub 2004/06/09. doi:
703 10.1016/j.virol.2004.03.033. PubMed PMID: 15183049.
704 32. Munoz N, Bosch FX, de Sanjose S, Herrero R, Castellsague X, Shah KV, et al.
705 Epidemiologic classification of human papillomavirus types associated with cervical cancer. N
706 Engl J Med. 2003;348(6):518-27. Epub 2003/02/07. doi: 10.1056/NEJMoa021641. PubMed
707 PMID: 12571259.
708 33. Virtanen S, Kalliala I, Nieminen P, Salonen A. Comparative analysis of vaginal
709 microbiota sampling using 16S rRNA gene analysis. PLoS One. 2017;12(7):e0181477. Epub
710 2017/07/21. doi: 10.1371/journal.pone.0181477. PubMed PMID: 28723942; PubMed Central
711 PMCID: PMCPMC5517051.
712 34. Microbial Isolation | ZYMO RESEARCH.
713 https://www.zymoresearch.com/pages/microbial-isolation. Accessed 12 Mar 2020.
714 35. PowerBead Tubes - QIAGEN Online Shop.
715 https://www.qiagen.com/us/products/discovery-and-translational-research/lab-
716 essentials/plastics/powerbead-tubes/#orderinginformation. Accessed 12 Mar 2020.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

717 36. Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ. Evaluation of methods for the
718 extraction and purification of DNA from the human microbiome. PLoS One. 2012;7(3):e33865.
719 Epub 2012/03/30. doi: 10.1371/journal.pone.0033865. PubMed PMID: 22457796; PubMed
720 Central PMCID: PMCPMC3311548.
721 37. QIAGEN. Pathogen Lysis Tubes - QIAGEN.
722 https://www.qiagen.com/dk/shop/pcr/pathogen-lysis-tubes/. Accessed 12 Mar 2020.
723 38. Kim D, Hofstaedter CE, Zhao C, Mattei L, Tanes C, Clarke E, et al. Optimizing methods
724 and dodging pitfalls in microbiome research. Microbiome. 2017;5(1):52. Epub 2017/05/10. doi:
725 10.1186/s40168-017-0267-5. PubMed PMID: 28476139; PubMed Central PMCID:
726 PMCPMC5420141.
727 39. Togo Picture Gallery. http://togotv.dbcls.jp/pics.html. Accessed 12 Mar 2020.
728 40. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et
729 al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc
730 Natl Acad Sci U S A. 2011;108 Suppl 1(Supplement 1):4516-22. Epub 2010/06/11. doi:
731 10.1073/pnas.1000080107. PubMed PMID: 20534432; PubMed Central PMCID:
732 PMCPMC3063599.
733 41. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A
734 communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017;551(7681):457-
735 63. Epub 2017/11/02. doi: 10.1038/nature24621. PubMed PMID: 29088705; PubMed Central
736 PMCID: PMCPMC6192678.
737 42. Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA
738 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat Microb Ecol.
739 2015;75(2):129-37. doi: 10.3354/ame01753. PubMed PMID: WOS:000357106200004.
740 43. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit
741 rRNA primers for marine microbiomes with mock communities, time series and global field
742 samples. Environ Microbiol. 2016;18(5):1403-14. Epub 2015/08/15. doi: 10.1111/1462-
743 2920.13023. PubMed PMID: 26271760.
744 44. Walters W, Hyde ER, Berg-Lyons D, Ackermann G, Humphrey G, Parada A, et al.
745 Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer
746 Marker Gene Primers for Microbial Community Surveys. mSystems. 2016;1(1). Epub
747 2016/11/09. doi: 10.1128/mSystems.00009-15. PubMed PMID: 27822518; PubMed Central
748 PMCID: PMCPMC5069754.
749 45. Earth Microbiome Project. 16S Illumina amplicon protocol.
750 http://www.earthmicrobiome.org/protocols-and-standards/16s/. Accessed 12 Mar 2020.
751 46. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al.
752 Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat
753 Biotechnol. 2019;37(8):852-7. Epub 2019/07/26. doi: 10.1038/s41587-019-0209-9. PubMed
754 PMID: 31341288; PubMed Central PMCID: PMCPMC7015180.
755 47. QIIME 2 View. https://view.qiime2.org/. Accessed 12 Mar 2020.
756 48. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace
757 operational taxonomic units in marker-gene data analysis. ISME J. 2017;11(12):2639-43. Epub
758 2017/07/22. doi: 10.1038/ismej.2017.119. PubMed PMID: 28731476; PubMed Central PMCID:
759 PMCPMC5702726.
760 49. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2:
761 High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581-3.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

762 Epub 2016/05/24. doi: 10.1038/nmeth.3869. PubMed PMID: 27214047; PubMed Central
763 PMCID: PMCPMC4927377.
764 50. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing
765 taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-
766 classifier plugin. Microbiome. 2018;6(1):90. Epub 2018/05/19. doi: 10.1186/s40168-018-0470-z.
767 PubMed PMID: 29773078; PubMed Central PMCID: PMCPMC5956843.
768 51. Werner JJ, Koren O, Hugenholtz P, DeSantis TZ, Walters WA, Caporaso JG, et al.
769 Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys.
770 ISME J. 2012;6(1):94-103. Epub 2011/07/01. doi: 10.1038/ismej.2011.82. PubMed PMID:
771 21716311; PubMed Central PMCID: PMCPMC3217155.
772 52. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA
773 ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic
774 Acids Res. 2013;41(Database issue):D590-6. Epub 2012/11/30. doi: 10.1093/nar/gks1219.
775 PubMed PMID: 23193283; PubMed Central PMCID: PMCPMC3531112.
776 53. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
777 improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-80. Epub
778 2013/01/19. doi: 10.1093/molbev/mst010. PubMed PMID: 23329690; PubMed Central PMCID:
779 PMCPMC3603318.
780 54. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective
781 stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol.
782 2015;32(1):268-74. Epub 2014/11/06. doi: 10.1093/molbev/msu300. PubMed PMID: 25371430;
783 PubMed Central PMCID: PMCPMC4271533.
784 55. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder:
785 fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587-9. Epub
786 2017/05/10. doi: 10.1038/nmeth.4285. PubMed PMID: 28481363; PubMed Central PMCID:
787 PMCPMC5453245.
788 56. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, et al. Quality-
789 filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods.
790 2013;10(1):57-9. Epub 2012/12/04. doi: 10.1038/nmeth.2276. PubMed PMID: 23202435;
791 PubMed Central PMCID: PMCPMC3531572.
792 57. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis
793 and graphics of microbiome census data. PLoS One. 2013;8(4):e61217. Epub 2013/05/01. doi:
794 10.1371/journal.pone.0061217. PubMed PMID: 23630581; PubMed Central PMCID:
795 PMCPMC3632530.
796 58. Bisanz JE. qiime2R: Importing QIIME2 artifacts and associated data into R sessions.
797 https://github.com/jbisanz/qiime2R. Accessed 12 Mar 2020.
798 59. Lahti L, Shetty S. microbiome R package. http://microbiome.github.io. Accessed 12 Mar
799 2020.
800 60. Anderson MJ. A new method for non-parametric multivariate analysis of variance.
801 Austral Ecol. 2001;26(1):32-46. doi: DOI 10.1111/j.1442-9993.2001.01070.pp.x. PubMed
802 PMID: WOS:000167002000004.
803 61. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan:
804 Community Ecology Package. R package version 2.5-3. https://CRAN.R-
805 project.org/package=vegan. Accessed 12 Mar 2020.
806 62. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic
807 biomarker discovery and explanation. Genome Biol. 2011;12(6):R60. Epub 2011/06/28. doi:

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

808 10.1186/gb-2011-12-6-r60. PubMed PMID: 21702898; PubMed Central PMCID:
809 PMCPMC3218848.
810 63. Bardou P, Mariette J, Escudie F, Djemiel C, Klopp C. jvenn: an interactive Venn diagram
811 viewer. BMC Bioinformatics. 2014;15(1):293. Epub 2014/09/02. doi: 10.1186/1471-2105-15-
812 293. PubMed PMID: 25176396; PubMed Central PMCID: PMCPMC4261873.
813 64. Cao Y. microbiomeMarker: microbiome biomarker analysis. R package version
814 0.0.1.9000. https://github.com/yiluheihei/microbiomeMarker. Accessed 21 Nov 2020. doi: DOI:
815 10.5281/zenodo.3749415.
816 65. Willis A, Bunge J. Estimating diversity via frequency ratios. Biometrics.
817 2015;71(4):1042-9. Epub 2015/06/04. doi: 10.1111/biom.12332. PubMed PMID: 26038228.
818 66. Martino C, Morton JT, Marotz CA, Thompson LR, Tripathi A, Knight R, et al. A Novel
819 Sparse Compositional Technique Reveals Microbial Perturbations. mSystems. 2019;4(1). Epub
820 2019/02/26. doi: 10.1128/mSystems.00016-19. PubMed PMID: 30801021; PubMed Central
821 PMCID: PMCPMC6372836.
822 67. Gao W, Weng J, Gao Y, Chen X. Comparison of the vaginal microbiota diversity of
823 women with and without human papillomavirus infection: a cross-sectional study. BMC Infect
824 Dis. 2013;13(1):271. Epub 2013/06/14. doi: 10.1186/1471-2334-13-271. PubMed PMID:
825 23758857; PubMed Central PMCID: PMCPMC3684509.
826 68. Montealegre JR, Peckham-Gregory EC, Marquez-Do D, Dillon L, Guillaud M, Adler-
827 Storthz K, et al. Racial/ethnic differences in HPV 16/18 genotypes and integration status among
828 women with a history of cytological abnormalities. Gynecol Oncol. 2018;148(2):357-62. Epub
829 2017/12/26. doi: 10.1016/j.ygyno.2017.12.014. PubMed PMID: 29276057; PubMed Central
830 PMCID: PMCPMC5801201.
831 69. Xi LF, Kiviat NB, Hildesheim A, Galloway DA, Wheeler CM, Ho J, et al. Human
832 papillomavirus type 16 and 18 variants: race-related distribution and persistence. J Natl Cancer
833 Inst. 2006;98(15):1045-52. Epub 2006/08/03. doi: 10.1093/jnci/djj297. PubMed PMID:
834 16882941.
835 70. Morgan M. DirichletMultinomial: Dirichlet-Multinomial Mixture Model Machine
836 Learning for Microbiome Data.
837 http://bioconductor.org/packages/release/bioc/html/DirichletMultinomial.html. Accessed 12 Mar
838 2020.
839 71. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for
840 microbial metagenomics. PLoS One. 2012;7(2):e30126. Epub 2012/02/10. doi:
841 10.1371/journal.pone.0030126. PubMed PMID: 22319561; PubMed Central PMCID:
842 PMCPMC3272020.
843 72. DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, et al.
844 Temporal and spatial variation of the human microbiota during pregnancy. Proc Natl Acad Sci U
845 S A. 2015;112(35):11060-5. Epub 2015/08/19. doi: 10.1073/pnas.1502875112. PubMed PMID:
846 26283357; PubMed Central PMCID: PMCPMC4568272.
847 73. Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB. ANOVA-like differential
848 expression (ALDEx) analysis for mixed population RNA-Seq. PLoS One. 2013;8(7):e67019.
849 Epub 2013/07/12. doi: 10.1371/journal.pone.0067019. PubMed PMID: 23843979; PubMed
850 Central PMCID: PMCPMC3699591.
851 74. Dinno A. dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums.
852 https://CRAN.R-project.org/package=dunn.test. Accessed 12 Mar 2020.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

853 75. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, et al. Minimum
854 information about a marker gene sequence (MIMARKS) and minimum information about any
855 (x) sequence (MIxS) specifications. Nat Biotechnol. 2011;29(5):415-20. Epub 2011/05/10. doi:
856 10.1038/nbt.1823. PubMed PMID: 21552244; PubMed Central PMCID: PMCPMC3367316.
857 76. Silverman JD, Bloom RJ, Jiang S, Durand HK, Mukherjee S, David LA. Measuring and
858 mitigating PCR bias in microbiome data. BioRxiv. 2019:604025.
859 77. Laursen MF, Dalgaard MD, Bahl MI. Genomic GC-Content Affects the Accuracy of 16S
860 rRNA Gene Sequencing Based Microbial Profiling due to PCR Bias. Front Microbiol.
861 2017;8:1934. Epub 2017/10/21. doi: 10.3389/fmicb.2017.01934. PubMed PMID: 29051756;
862 PubMed Central PMCID: PMCPMC5633598.
863 78. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and
864 sequencing artifacts on 16S rRNA-based studies. PLoS One. 2011;6(12):e27310. Epub
865 2011/12/24. doi: 10.1371/journal.pone.0027310. PubMed PMID: 22194782; PubMed Central
866 PMCID: PMCPMC3237409.
867 79. Silhavy TJ, Kahne D, Walker S. The bacterial cell envelope. Cold Spring Harb Perspect
868 Biol. 2010;2(5):a000414. Epub 2010/05/11. doi: 10.1101/cshperspect.a000414. PubMed PMID:
869 20452953; PubMed Central PMCID: PMCPMC2857177.
870 80. Balle C, Lennard K, Dabee S, Barnabas SL, Jaumdally SZ, Gasper MA, et al.
871 Endocervical and vaginal microbiota in South African adolescents with asymptomatic
872 Chlamydia trachomatis infection. Sci Rep. 2018;8(1):11109. Epub 2018/07/25. doi:
873 10.1038/s41598-018-29320-x. PubMed PMID: 30038262; PubMed Central PMCID:
874 PMCPMC6056523.
875 81. Klein C, Gonzalez D, Samwel K, Kahesa C, Mwaiselage J, Aluthge N, et al. Relationship
876 between the Cervical Microbiome, HIV Status, and Precancerous Lesions. mBio. 2019;10(1).
877 Epub 2019/02/21. doi: 10.1128/mBio.02785-18. PubMed PMID: 30782659; PubMed Central
878 PMCID: PMCPMC6381280.
879 82. Hayashi NR, Ishida T, Yokota A, Kodama T, Igarashi Y. Hydrogenophilus
880 thermoluteolus gen. nov., sp. nov., a thermophilic, facultatively chemolithoautotrophic,
881 hydrogen-oxidizing bacterium. Int J Syst Bacteriol. 1999;49 Pt 2:783-6. Epub 1999/05/13. doi:
882 10.1099/00207713-49-2-783. PubMed PMID: 10319503.
883 83. Glassing A, Dowd SE, Galandiuk S, Davis B, Chiodini RJ. Inherent bacterial DNA
884 contamination of extraction and sequencing reagents may affect interpretation of microbiota in
885 low bacterial biomass samples. Gut Pathog. 82016.
886 84. Birse KD, Romas LM, Guthrie BL, Nilsson P, Bosire R, Kiarie J, et al. Genital Injury
887 Signatures and Microbiome Alterations Associated With Depot Medroxyprogesterone Acetate
888 Usage and Intravaginal Drying Practices. J Infect Dis. 2017;215(4):590-8. Epub 2016/12/25. doi:
889 10.1093/infdis/jiw590. PubMed PMID: 28011908; PubMed Central PMCID: PMCPMC5388302.
890 85. Lennard K, Dabee S, Barnabas SL, Havyarimana E, Blakney A, Jaumdally SZ, et al.
891 Microbial Composition Predicts Genital Tract Inflammation and Persistent Bacterial Vaginosis
892 in South African Adolescent Females. Infect Immun. 2018;86(1). Epub 2017/10/19. doi:
893 10.1128/IAI.00410-17. PubMed PMID: 29038128; PubMed Central PMCID: PMCPMC5736802.
894 86. McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is
895 inadmissible. PLoS Comput Biol. 2014;10(4):e1003531. Epub 2014/04/05. doi:
896 10.1371/journal.pcbi.1003531. PubMed PMID: 24699258; PubMed Central PMCID:
897 PMCPMC3974642.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

898 87. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and
899 microbial differential abundance strategies depend upon data characteristics. Microbiome.
900 2017;5(1):27. Epub 2017/03/04. doi: 10.1186/s40168-017-0237-y. PubMed PMID: 28253908;
901 PubMed Central PMCID: PMCPMC5335496.
902 88. Callahan BJ, Wong J, Heiner C, Oh S, Theriot CM, Gulati AS, et al. High-throughput
903 amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution.
904 Nucleic Acids Res. 2019;47(18):e103. Epub 2019/07/04. doi: 10.1093/nar/gkz569. PubMed
905 PMID: 31269198; PubMed Central PMCID: PMCPMC6765137.
906 89. Calus ST, Ijaz UZ, Pinto AJ. NanoAmpli-Seq: a workflow for amplicon sequencing for
907 mixed microbial communities on the nanopore sequencing platform. Gigascience. 2018;7(12).
908 Epub 2018/11/27. doi: 10.1093/gigascience/giy140. PubMed PMID: 30476081; PubMed Central
909 PMCID: PMCPMC6298384.
910 90. Wongsurawat T, Nakagawa M, Atiq O, Coleman HN, Jenjaroenpun P, Allred JI, et al. An
911 assessment of Oxford Nanopore sequencing for human gut metagenome profiling: A pilot study
912 of head and neck cancer patients. J Microbiol Methods. 2019;166:105739. Epub 2019/10/19. doi:
913 10.1016/j.mimet.2019.105739. PubMed PMID: 31626891; PubMed Central PMCID:
914 PMCPMC6956648.
915 91. Usyk M, Zolnik CP, Castle PE, Porras C, Herrero R, Gradissimo A, et al. Cervicovaginal
916 microbiome and natural history of HPV in a longitudinal study. PLoS Pathog.
917 2020;16(3):e1008376. Epub 2020/03/28. doi: 10.1371/journal.ppat.1008376. PubMed PMID:
918 32214382; PubMed Central PMCID: PMCPMC7098574 Schiller and Douglas R. Lowy report
919 that they are named inventors on US Government-owned HPV vaccine patents that are licensed
920 to GlaxoSmithKline and Merck and for which the National Cancer Institute receives licensing
921 fees. They are entitled to limited royalties as specified by federal law.
922 92. Audirac-Chalifour A, Torres-Poveda K, Bahena-Roman M, Tellez-Sosa J, Martinez-
923 Barnetche J, Cortina-Ceballos B, et al. Cervical Microbiome and Cytokine Profile at Various
924 Stages of Cervical Cancer: A Pilot Study. PLoS One. 2016;11(4):e0153274. Epub 2016/04/27.
925 doi: 10.1371/journal.pone.0153274. PubMed PMID: 27115350; PubMed Central PMCID:
926 PMCPMC4846060.
927 93. Di Paola M, Sani C, Clemente AM, Iossa A, Perissi E, Castronovo G, et al.
928 Characterization of cervico-vaginal microbiota in women developing persistent high-risk Human
929 Papillomavirus infection. Sci Rep. 2017;7(1):10200. Epub 2017/09/02. doi: 10.1038/s41598-
930 017-09842-6. PubMed PMID: 28860468; PubMed Central PMCID: PMCPMC5579045.
931 94. Ranjeva SL, Mihaljevic JR, Joseph MB, Giuliano AR, Dwyer G. Untangling the
932 dynamics of persistence and colonization in microbial communities. ISME J. 2019;13(12):2998-
933 3010. Epub 2019/08/25. doi: 10.1038/s41396-019-0488-7. PubMed PMID: 31444482; PubMed
934 Central PMCID: PMCPMC6863904.
935 95. Linder J, Zahniser D. ThinPrep Papanicolaou testing to reduce false-negative cervical
936 cytology. Arch Pathol Lab Med. 1998;122(2):139-44. Epub 1998/03/14. PubMed PMID:
937 9499356.
938 96. Ling Z, Liu X, Chen X, Zhu H, Nelson KE, Xia Y, et al. Diversity of cervicovaginal
939 microbiota associated with female lower genital tract infections. Microb Ecol. 2011;61(3):704-14.
940 Epub 2011/02/03. doi: 10.1007/s00248-011-9813-z. PubMed PMID: 21287345.
941 97. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and
942 laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

943 2014;12(1):87. Epub 2014/11/13. doi: 10.1186/s12915-014-0087-z. PubMed PMID: 25387460;
944 PubMed Central PMCID: PMCPMC4228153.
945 98. Rebolj M, Rask J, van Ballegooijen M, Kirschner B, Rozemeijer K, Bonde J, et al.
946 Cervical histology after routine ThinPrep or SurePath liquid-based cytology and computer-
947 assisted reading in Denmark. Br J Cancer. 2015;113(9):1259-74. Epub 2015/10/09. doi:
948 10.1038/bjc.2015.339. PubMed PMID: 26448176; PubMed Central PMCID: PMCPMC4815798.
949 99. Naeem RC, Goldstein DY, Einstein MH, Ramos Rivera G, Schlesinger K, Khader SN, et
950 al. SurePath Specimens Versus ThinPrep Specimen Types on the COBAS 4800 Platform: High-
951 Risk HPV Status and Cytology Correlation in an Ethnically Diverse Bronx Population. Lab Med.
952 2017;48(3):207-13. Epub 2017/04/06. doi: 10.1093/labmed/lmx019. PubMed PMID: 28379422.
953 100. Ritu W, Enqi W, Zheng S, Wang J, Ling Y, Wang Y. Evaluation of the Associations
954 Between Cervical Microbiota and HPV Infection, Clearance, and Persistence in Cytologically
955 Normal Women. Cancer Prev Res (Phila). 2019;12(1):43-56. Epub 2018/11/23. doi:
956 10.1158/1940-6207.CAPR-18-0233. PubMed PMID: 30463989.
957 101. Hosomi K, Ohno H, Murakami H, Natsume-Kitatani Y, Tanisawa K, Hirata S, et al.
958 Method for preparing DNA from feces in guanidine thiocyanate solution affects 16S rRNA-
959 based profiling of human microbiota diversity. Sci Rep. 2017;7(1):4339. Epub 2017/07/01. doi:
960 10.1038/s41598-017-04511-0. PubMed PMID: 28659635; PubMed Central PMCID:
961 PMCPMC5489508.
962 102. Akahane T, Yamaguchi T, Kato Y, Yokoyama S, Hamada T, Nishida Y, et al.
963 Comprehensive validation of liquid-based cytology specimens for next-generation sequencing in
964 cancer genome analysis. PLoS One. 2019;14(6):e0217724. Epub 2019/06/15. doi:
965 10.1371/journal.pone.0217724. PubMed PMID: 31199826; PubMed Central PMCID:
966 PMCPMC6568385.
967 103. Cuschieri KS, Beattie G, Hassan S, Robertson K, Cubie H. Assessment of human
968 papillomavirus mRNA detection over time in cervical specimens collected in liquid based
969 cytology medium. J Virol Methods. 2005;124(1-2):211-5. Epub 2005/01/25. doi:
970 10.1016/j.jviromet.2004.11.005. PubMed PMID: 15664071.
971 104. Human papillomavirus (HPV) and cervical cancer. https://www.who.int/news-room/fact-
972 sheets/detail/human-papillomavirus-(hpv)-and-cervical-cancer. Accessed 12 Mar 2020.
973 105. Sarangi AN, Goel A, Aggarwal R. Methods for Studying Gut Microbiota: A Primer for
974 Physicians. J Clin Exp Hepatol. 2019;9(1):62-73. Epub 2019/02/19. doi:
975 10.1016/j.jceh.2018.04.016. PubMed PMID: 30774267; PubMed Central PMCID:
976 PMCPMC6363981.
977

978

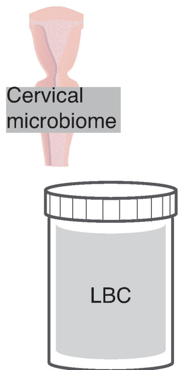
bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

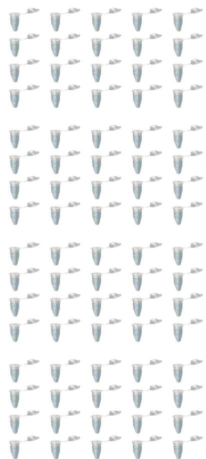
979 # Supporting information

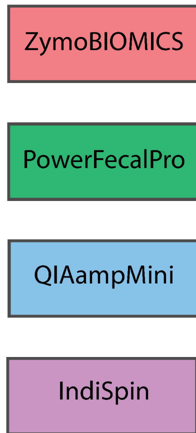
980 **Figure S1. Comparison of DNA yields by DNA extraction protocols.** DNA yield of
981 QIAampMini was significantly higher than that of PowerFecalPro ($p < 0.001$, Dunn's test with
982 Benjamini-Hochberg-adjustment). Also, the DNA yield of ZymoBIOMICS was significantly
983 higher than that of PowerFecalPro ($p < 0.001$, Dunn's test with Benjamini-Hochberg-adjustment).
984 The amount of DNA was calculated based on the absorbance of nucleic acids measured by
985 Nanodrop One. By the protocol recommended by the manufacturer, nucleic acid (Poly-A carrier)
986 was used in IndiSpin. Therefore, IndiSpin was excluded from the analysis of DNA yield. The
987 amount of DNA yield per 100 μL ThinPrep sample volume were compared. The bar graph shows
988 the mean and standard deviation. Zy: ZymoBIOMICS DNA Miniprep Kit, Pro: QIAamp
989 PowerFecal Pro DNA Kit, QIA: QIAamp DNA Mini Kit.
990
991 **Figure S2. Phylogenetic beta-diversity.** Weighted UniFrac (A & B) and Unweighted UniFrac
992 (C & D), PCoA colored by subject ID (top row) and DNA extraction kit (bottom row).
993 Weighted UniFrac clusters samples by subject whereas Unweighted UniFrac appears more
994 sensitive to the type of DNA extraction kit. Data were rarefied to 51,197 reads per sample.
995
996 **Figure S3. Deicode (Robust Aitchison PCA) beta-diversity.** Non-rarefaction-based analysis of
997 beta-diversity. Samples are colored by individual subject ID (A) and DNA extraction kit (B).
998 Samples predominately cluster by subject and not DNA extraction kit.
999
1000 **Figure S4. Community type and HPV 16 assessed by using 4 kits** (A) Community types were
1001 classified into two types in all DNA extraction kits, mainly based on the percentage of

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.27.921619; this version posted July 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1002 *Lactobacillus iners*. HPV16 infection was negatively associated with the dominance of *L. iners*
1003 (community type I; p = 0.001, Fisher's exact test) regardless of DNA extraction method.
1004 Although, we observed slight variation in the abundance of microbiota across the extraction kits
1005 (even within the same individual patient), the ability to detect two community types was
1006 identical across all DNA extraction kits. No significant differences were observed in the
1007 relationship of other phenotypes of patients (HPV18, HR-HPV, LR-HPV, multiple HPV
1008 infections, Biopsy, and Race). The top 15 bacteria detected for each DNA extraction kit are
1009 shown. Samples were clustered by the Dirichlet component. Narrow columns show each sample
1010 and a broader column shows averages of samples. Rows show taxa at the species level. Dark or
1011 thin colors correspond to larger or smaller counts of OTUs, respectively. CT: community type.
1012 (B) LEfSe analysis, using combined data from all four kits detected a significant enrichment of
1013 66 taxa in the cervical environment with HPV16 infection and 17 taxa without HPV16 infection.
1014 Genus *Lactobacillus* were enriched in the HPV16 negative patients (p < 0.001, LDA score: 5.38).
1015 Asterisks denote taxa that were significant after post-hoc significant testing with Scheffe test
1016 [64].
1017
1018


(A) Liquid-based cytology on cervix
Cervical microbiome
LBC

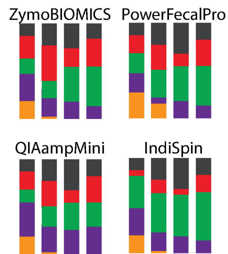

(B) Dispensing of each sample to 4 aliquot


(C) DNA extraction using four protocols
ZymoBIOMICS
PowerFecalPro
QIAampMini
IndiSpin

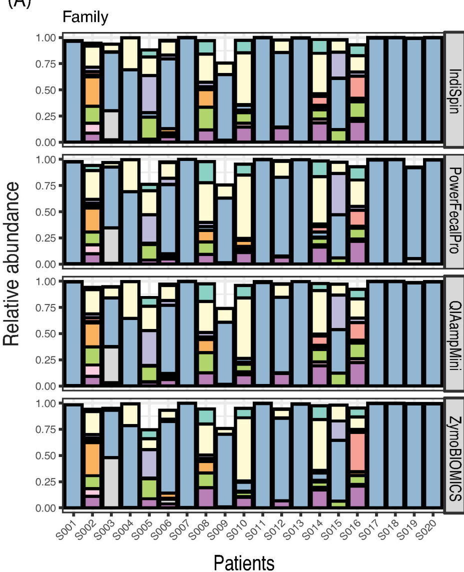
(D)
Positive control (vaginal mock)
Negative control (preservation solution without samples)

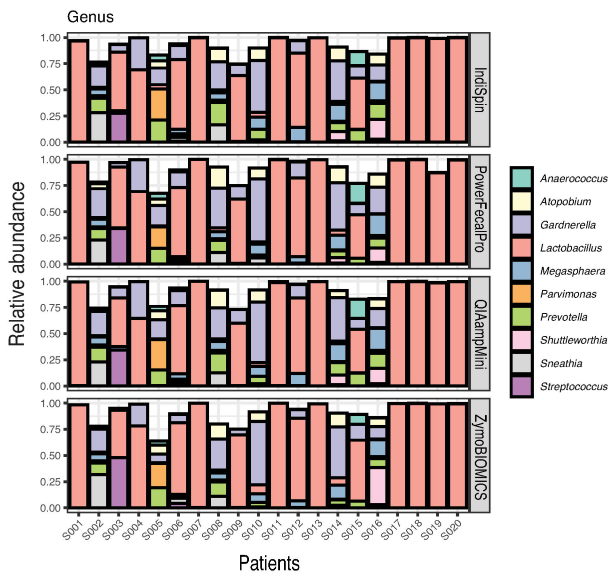
(E) 16S rRNA Illumina amplicon sequencing

(F) Evaluation of taxonomic profiles

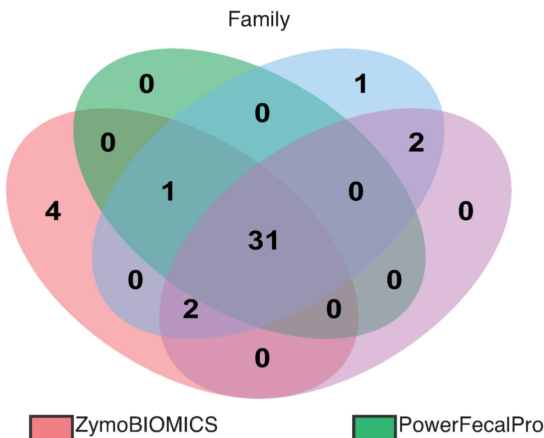

ZymoBIOMICS
PowerFecalPro
QIAampMini
IndiSpin

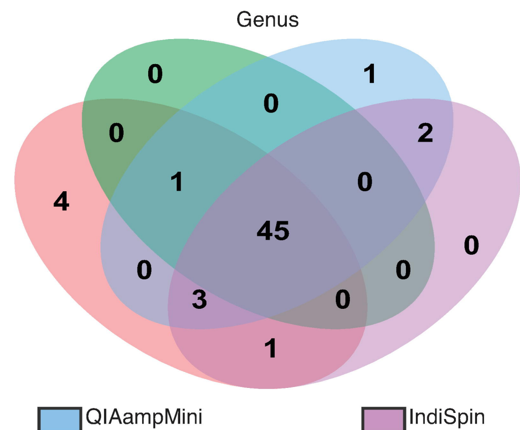
(A)



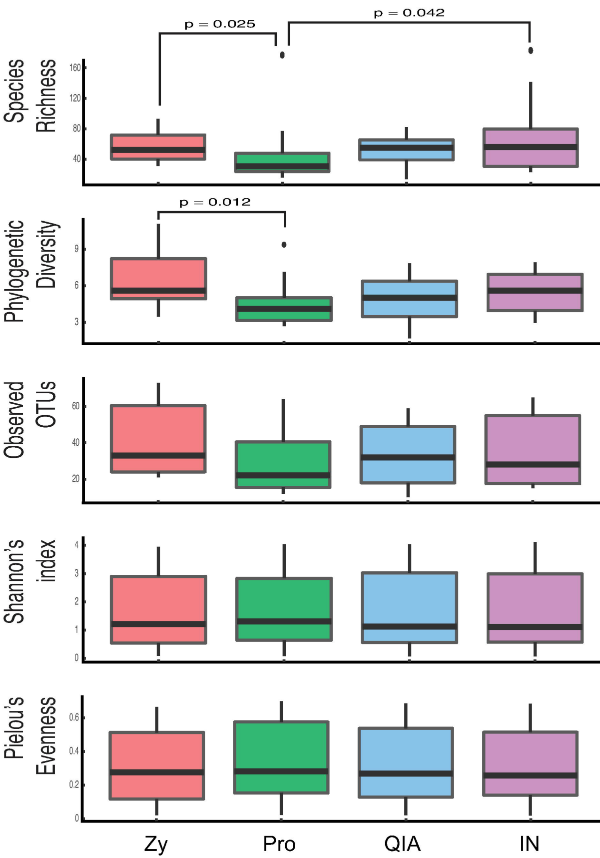


(B)





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