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

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

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

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

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

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

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

Takeo Shibata: TShibata@uams.edu; Mayumi Nakagawa: MNakagawa@uams.edu; Hannah N. Coleman: ColemanHannahN@uams.edu; Sarah M. Owens: sarah.owens@anl.gov; William W. Greenfield: GreenfieldWilliamW@uams.edu; Toshiyuki Sasagawa: tsasa@kanazawa-med.ac.jp; Michael S. Robeson II: MRobeson@uams.edu.

Corresponding author: Michael S. Robeson II

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

Email: MRobeson@uams.edu

Abstract

Background: Cervical microbiota (CM) are considered an important factor affecting the progression of cervical intraepithelial neoplasia (CIN) and are implicated in the persistence of human papillomavirus (HPV). Collection of liquid-based cytology (LBC) samples is routine for cervical cancer screening and HPV genotyping, and can be used for long-term cytological biobanking. Herein, we investigate the feasibility of leveraging LBC specimens for use in CM surveys by amplicon sequencing. As methodological differences in DNA extraction protocols can potentially bias the composition of microbiota, we set out to determine the performance of four commonly used DNA extraction kits (ZymoBIOMICS DNA Miniprep Kit; QIAamp PowerFecal Pro DNA Kit; QIAamp DNA Mini Kit; and IndiSpin Pathogen Kit) and their ability to capture the diversity of CM from LBC specimens.

Results: LBC specimens from 20 patients (stored for 716 ± 105 days) with cervical intraepithelial neoplasia (CIN) 2/3 or suspected CIN2/3 were each aliquoted for the four kits. We observed that, regardless of the extraction protocol used, all kits provided equivalent accessibility to the cervical microbiome, with some minor differences. For example, the ZymoBIOMICS kit appeared to differentially increase access of several microbiota compared to the other kits. Potential kit contaminants were observed as well. Approximately 80% microbial genera were shared among all DNA extraction protocols. The variance of microbial composition per individual was larger than that of the DNA extraction protocol used. We also observed that HPV16 infection was significantly associated with community types that were not dominated by *Lactobacillus iners*.

Conclusions: Collection of LBC specimens is routine for cervical cancer screening and HPV genotyping, and can be used for long-term cytological biobanking. We demonstrated that LBC

samples, which had been under prolonged storage prior to DNA extraction, were able to provide a robust assessment of the CM and its relationship to HPV status, regardless of the extraction kit used. Being able to retroactively access the CM from biobanked LBC samples, will allow researchers to better interrogate historical interactions between the CM and its relationship to CIN and HPV. This alone has the potential to bring CM research one-step closer to the clinical practice.

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

55 **Background**

56 High-throughput sequencing (HTS) technology of 16S rRNA gene amplicon sequences has made
57 it possible to better understand the relationships between cervicovaginal microbiota and human
58 papillomavirus (HPV) infection [1] [2] [3] [4] [5] and HPV-related diseases [6] [7] [8] [9] [10].
59 Cervicovaginal microbiota are considered to be an important factor affecting the progress of
60 cervical intraepithelial neoplasia (CIN) [6] [7] [8] [9] and are implicated in the persistence of
61 high-risk HPV (HR-HPV) [1] [2] and low-risk HPV (LR-HPV) [3]. For example, the phyla
62 *Actinobacteria* and *Fusobacteria* were enriched in HR-HPV positive environment [4] while the
63 phyla *Actinobacteria*, *Proteobacteria*, and *Fusobacteria* in low-risk HPV (LR-HPV) [3].
64 Additionally, *Lactobacillus iners*-dominant samples are associated with both HR-HPV and LR-
65 HPV [5]. Moreover, it has been shown that CIN risk was increased when the cervical microbes
66 *Atopobium vaginae*, *Gardnerella vaginalis*, and *Lactobacillus iners* were present with HR-HPV
67 [10]. The cervicovaginal microbiome specified by *Lactobacillus*-dominant type or non-
68 *Lactobacillus*-dominant type has been shown to interact with the immune system [7] [11].
69 Inflammatory cytokines, such as Interleukin (IL)-1α and IL-18, were increased in non-
70 *Lactobacillus*-dominant community types of reproductive-aged healthy women [11]. In the
71 analysis of patients with cervical cancer, non-*Lactobacillus*-dominant community types were
72 positively associated with chemokines such as interferon gamma-induced protein 10 (IP-10) and
73 soluble CD40-ligand activating dendric cells (DCs) [7]. The metabolism of the cervicovaginal
74 microbiome may be a substantial contributing factor to maternal health during pregnancy,
75 although the mechanism is still unclear [12].

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

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

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

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

94 **Results**

95 **Patient characteristics**

96 The age of the patients (n = 20) was 31.4 ± 5.0 years. The distribution of race was 15% African
97 American (n = 3), 50% Caucasian (n = 10), and 35% Hispanic (n = 7). Cervical histology was
98 40% CIN2 (n = 8), 50% CIN3 (n = 10), and 10% benign (n = 2). HPV genotypes were 50%
99 HPV16 positive (n = 10), 10% HPV18 positive (n = 2), and 90% HR-HPV positives (n = 18).
100 Patient characteristics were summarized in Table 2.

101

102 **DNA yield**

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

108

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

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

117 Prior to rarefying, the ZymoBIOMICS kit captured a greater representation of gram-
118 negative bacterial OTUs (total 346, 17.3 ± 9.8) compared to PowerFecalPro (total 209, 10.5 ±
119 10.3, p value = 0.012, Dunn's test with Benjamini-Hochberg-adjustment, ratio of gram-negative
120 bacteria: 41.9% vs 33.7%) as shown in Table 3. No significant differences in the number of
121 OTUs before rarefying was detected for the entire bacterial community or gram-positive
122 bacteria.
123

124 **Microbiome composition per DNA extraction protocol**

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

130 The following top 10 abundant families are shown in Figure 2A (left) and constituted
131 approximately 95.7% of cervical bacteria in all kits (80 DNA extractions); *Lactobacillaceae*
132 (58.9%), *Bifidobacteriaceae* (13.7%), *Veillonellaceae* (4.8%), *Prevotellaceae* (4.3%), *Family XI*
133 (3.9%), *Atopobiaceae* (3.0%), *Leptotrichiaceae* (2.5%), *Streptococcaceae* (2.0%),
134 *Lachnospiraceae* (1.6%). *Ruminococcaceae* (0.9%). The following top 10 abundant genera are
135 shown in Figure 2A (right) and constituted approximately 92% of cervical bacteria;
136 *Lactobacillus* (58.9%), *Gardnerella* (13.6%), *Prevotella* (4.2%), *Megasphaera* (3.7%),
137 *Atopobium* (3.0%), *Sneathia* (2.5%), *Streptococcus* (1.9%), *Parvimonas* (1.7%), *Shuttleworthia*
138 (1.4%), and *Anaerococcus* (1.1%).
139

140 **Shared and unique microbiota among DNA extraction protocols**

141 All DNA extraction methods were generally commensurate with one another, there were 31 of
142 41 shared microbes at the family level (Figure 2B left) and 45 of 57 shared microbes at the genus
143 level (Figure 2B right) among the DNA extraction protocols.

144 However, four gram-negative taxa were uniquely detected by ZymoBIOMICS and one
145 taxon was uniquely detected by QIAampMini both at the genus level (Figure 2B right). Of the
146 uniquely detected ZymoBIOMICS OTUs, *Methylobacterium* was detected in 5 of the 80 DNA
147 extractions, consisting of 912 reads; 0.01% of all kit extractions. A member of this genus,
148 *Methylobacterium aerolatum*, has been reported to be more abundant in the endocervix than the
149 vagina of healthy South African women [26]. *Bacteroidetes*, which are often reported as
150 enriched taxa in an HIV positive cervical environment [27], was detected in 12 of the 80 DNA
151 extractions (1,028 reads; 0.01%). *Meiothermus* was detected in 9 of the 80 DNA extractions (882
152 reads; 0.01%) and *Hydrogenophilus* was detected in 14 of the 80 DNA extractions (2,488 reads,
153 0.02%). *Meiothermus* and *Hydrogenophilus* [28] are not considered to reside within the human
154 environment, and are likely kit contaminants, as previously reported [29]. A unique gram-
155 positive taxa obtained from the QIAampMini, *Streptomyces*, which was reported to be detected
156 from the cervicovaginal environment in the study of Kenyan women [30], was detected in 20 of
157 80 DNA extractions (6,862 reads; 0.06%). No unique taxa were detected in PowerFecalPro and
158 IndiSpin. Although less than 0.005% of the total data set, two samples of IndiSpin also detected
159 potential kit contaminant, *Tepidiphilus* (*Hydrogenophilaceae*).

160 Venn diagrams at family levels also exhibited that ZymoBIOMICS detected slightly
161 more bacterial taxa (four unique taxa) as shown in Figure 2B (left). These results showed that

162 major bacteria were commonly detected among all extraction protocols, with only slightly more
163 uniquely detected microbiota using ZymoBIOMICS.
164
165 **Alpha and beta diversity**
166 Significantly higher Species richness (`q2-breakaway`) was observed from the
167 ZymoBIOMICS (56.1 ± 19.4) protocol compared to that of PowerFecalPro (43.2 ± 32.9, p =
168 0.025), QIAampMini (54.9 ± 29.8, not significant), and IndiSpin (63.6 ± 38.3, not significant)
169 using Dunn's test with Benjamini-Hochberg-adjustment (Figure 3). Similarly, Faith's
170 Phylogenetic Diversity was observed to be higher with the ZymoBIOMICS protocol (6.6 ± 2.2),
171 compared to PowerFecalPro (4.5 ± 1.9, p = 0.012), QIAampMini (5.0 ± 1.8, not significant), and
172 IndiSpin (5.4 ± 1.7, not significant) using Dunn's test with Benjamini-Hochberg-adjustment
173 (Figure 3). The use of IndiSpin also resulted significantly higher alpha diversity than that of
174 PowerFecalPro in an analysis of Species richness (p = 0.042, Dunn's test with Benjamini-
175 Hochberg-adjustment). Non-phylogenetic alpha diversity metrics such as Observed OTUs,
176 Shannon's diversity index, and Pielou's Evenness did not show differences among the four
177 methods.
178 ZymoBIOMICS was able to significantly increase access to several taxonomic groups
179 compared to the other DNA extraction methods. Additionally, as shown in Table 4,
180 ZymoBIOMICS did capture a different microbial composition compared to other DNA
181 extraction methods in the index of Unweighted UniFrac distances (PowerFecalPro: q = 0.002;
182 QIAampMini: q = 0.002; and IndiSpin: q = 0.002) and in Jaccard distances (QIAampMini: q =
183 0.018 and IndiSpin: q = 0.033).
184

185 **Differential accessibility of microbiota by DNA extraction protocol**

186 Linear discriminant analysis (LDA) effect size (LEfSe) analysis [31] identified
187 taxonomic groups, defined with an LDA score of 2 or higher, for differential accessibility by
188 extraction kit: 23 in ZymoBIOMICS, 0 in PowerFecalPro, 3 in QIAampMini, and 3 in IndiSpin
189 (Figure 4A). The following taxa were found to be highly accessible (LDA score > 3) with the use
190 of the ZymoBIOMICS kit: Phylum *Proteobacteria*, Class *Gammaproteobacteria*, Order
191 *Betaproteobacteriales*, Family *Bacillaceae*, and Genus *Anoxybacillus*. Whereas the Order
192 *Streptomycetales* was highly enriched with the use of the QIAampMini (LDA score > 3). As
193 shown in the cladogram (Figure 4B), despite the detection of a potential kit contaminants
194 (*Meiothermus, Hydrogenophilaceae, and Hydrogenophilus*), ZymoBIOMICS was able to
195 increase the accessibility to additional microbiota compared to the other extraction protocols.
196

197 **Microbial community type and HPV16**

198 Dirichlet Multinomial Mixtures (DMM) model [32] detected two cervical microbial community
199 types across all four DNA extraction protocols (Figure S2). Community type I was composed of
200 the following: *Gardnerella sp.* (ZymoBIOMICS: 17.1%; PowerFecalPro: 20%; QIAampMini:
201 23%; IndiSpin: 20%), *Lactobacillus iners* (ZymoBIOMICS: 6.3%; PowerFecalPro: 5%;
202 QIAampMini: 6%; IndiSpin: 5%), *Atopobium vaginae* [10] (ZymoBIOMICS: 3.5%;
203 PowerFecalPro: 3%; QIAampMini: 4%; IndiSpin: 5%), *Clamydia trachomatis* (ZymoBIOMICS:
204 1.9%; PowerFecalPro: 2%; QIAampMini: 3%; IndiSpin: 2%), *Shuttleworthia sp.*
205 (ZymoBIOMICS: 1.8%; PowerFecalPro: 2%; QIAampMini: 2%; IndiSpin: 2%). Some members
206 of *Shuttleworthia* are considered to be bacterial vaginosis-associated bacterium (BVAB) [33],
207 further investigation is required to determine if this OTU is indeed a BVAB. We determined this

community type “high diversity type”. Community type II was is dominated by *Lactobacillus iners* at 88%, 85%, 83%, and 85% respectively for ZymoBIOMICS, PowerFecalPro, QIAampMini, and IndiSpin.

The relationship between HPV16 infection and community type was observed to be significantly associated with community type I (HPV16 positive patients [n = 9], HPV16 negative patients [n = 1]) and not community type II (HPV16 positive patients [n = 1], HPV16 negative patients [n = 9], $p = 0.001$, Fisher's exact test) regardless of the DNA extraction kit used (Figure S2A). In support of this result, analysis of differentially abundant microbiota using `q2-aldex` (Benjamini-Hochberg corrected p value of Wilcoxon test: $p < 0.001$, standardized distributional effect size: -1.2) revealed that *Lactobacillus iners* were differentially enriched in the cervical environment without HPV16. LEfSe analysis also detected that genus *Lactobacillus* were enriched in the cervical environment without HPV16 ($p < 0.001$, LDA score: 5.38, Figure S2B). No significant differences were observed in the relationship between community type and HPV18 ($p = 0.474$, Fisher's exact test), HR-HPV ($p = 0.474$, Fisher's exact test), results of cervical biopsy ($p = 0.554$, Fisher's exact test), and race (African Americans vs not-African Americans: $p = 1$; Caucasian vs not-Caucasian: $p = 0.656$; Hispanic vs not-Hispanic: $p = 0.350$, Fisher's exact test, Figure S2A).

225 **Discussion**

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

234 This study evaluated the composition of microbiota across all DNA extraction methods.
235 These findings document the importance of selecting DNA extraction methods in cervical
236 microbiome studies from the LBC samples. All kits were commensurate in their ability to
237 capture the microbial composition of each patient and the two observed cervical microbial
238 community state types, making all of these protocols viable for discovering broad patterns of
239 microbial diversity. However, we did observe that the ZymoBIOMICS protocol was better able
240 to access additional cervical microbiota (Figure 2B, 4A & B). Coincidentally, we detected
241 potential DNA contamination with the ZymoBIOMICS and IndiSpin kits. The number of OTUs
242 prior to rarefying revealed that the ZymoBIOMICS protocol detected more gram-negative OTUs
243 than the PowerFecalPro (Table 3 & Figure 2B). In particular, LEfSe analysis has shown that
244 phylum *Proteobacteria* can be better detected with the ZymoBIOMICS kit (Figure 4).

245 Although rarefying microbiome data can be problematic [34], it can still provide robust
246 and interpretable results for diversity analysis [35], we were able to observe commensurate
247 findings with non-rarefying approaches such as `q2-breakaway` [36], `q2-deicode` [37], and

248 LEfSe [31]. Beta-diversity analysis via Unweighted UniFrac also revealed that ZymoBIOMICS
249 was significantly different from all other kits. There were no differences in non-phylogenic
250 indices of alpha diversity with rarefying approaches. These findings lead us to surmise that
251 phylogenetic indices may be more sensitive than the non-phylogenetic indices.

252 Although we hypothesized that the detection of difficult-to-lyse-bacteria (*e.g*. gram-
253 positive bacteria) would vary by kit, we observed no significant differences (Table 3). As shown
254 in Table 3, the number of reads of gram-positive and gram-negative bacteria also showed that
255 there was no difference in the four kits. This is likely due to several modifications made to the
256 extraction protocol as outlined in Table 1. That is, we added bead beating and mutanolysin to the
257 QIAampMini protocol [38]. We also modified the beating time of the ZymoBIOMICS kit down
258 to 2 minutes from 10 minutes (the latter being recommended by the manufacturer) to minimize
259 DNA shearing. We may use the extracted DNA from ZymoBIOMICS for long-read amplicon
260 sequencing platforms such as PacBio (Pacific Biosciences of California, Inc) [39] or MinION
261 (Oxford Nanopore Technologies) [40] [41]. Excessive shearing can render these samples
262 unusable for long-read sequencing. It is quite possible that we could have observed even more
263 diversity with the ZymoBIOMICS kit for our amplicon survey if we conducted bead-beating for
264 the full 10 minutes.

265 Community typing and detection of the differentially abundant microbiota revealed that
266 *Lactobacillus iners* were more abundant in the cervical ecosystem without HPV16. These
267 findings are congruent with those of, Usyk *et al*. [42], Lee *et al*. [1], and Audirac-Chalifour *et al*.
268 [43]. Usyk *et al*., reported that *L. iners* was associated with clearance of HR-HPV infections
269 [42]. Lee *et al*. reported that *L. iners* were decreased in HPV positive women [1]. Also, the
270 results indicated that the proportion of *L. iners* was higher in HPV-negative women compared to

271 HPV-positive women (relative abundance 14.9% vs 2.1%) was reported by Audirac-Chalifour *et*
272 *al* [43]. Similarly, Tuominen *et al*. [18] reported that *L. iners* were enriched in HPV negative
273 samples (relative abundance: 47.7%) compared to HPV positive samples (relative abundance:
274 18.6%, p value = 0.07) in the study of HPV positive-pregnant women (HPV16 positive rate:
275 15%) [44]. As established by the seminal study of Ranjeva *et al*. [45], a statistical model
276 revealed that colonization of specific HPV types including multi HPV type infection depends on
277 host-risk factors such as sexual behavior, race and ethnicity, and smoking. It is unclear whether
278 the association between the cervical microbiome, host-specific traits, and persistent infection of
279 specific HPV types, such as HPV16, can be generalized and requires further investigation.
280 We focused on LBC samples as this is the recommended method of storage for cervical
281 cytology [46]. Here, we confirmed that LBC samples can be used for microbial community
282 surveys by simply using the remaining LBC solution post HPV testing or cervical cytology. We
283 used a sample volume of 200 or 300 μL ThinPrep solution in this study. The Linear Array HPV
284 Genotyping Test (Roche Diagnostics) stably detects β-globin with a base length of 268 bp as a
285 positive control. Therefore, using a similar sample volume as HPV genotyping (250 μL), it was
286 expected that V4 (250 bp), which is near the base length of β-globin, would be PCR amplified. It
287 has been pointed out by Ling *et al*. [47] that the cervical environment is of low microbial
288 biomass. To control reagent DNA contamination and estimate the sample volume, DNA
289 quantification by qPCR before sequencing is recommended [48]. Mitra *et al* determined a sample
290 volume of 500 μL for ThinPrep by qPCR in the cervical microbiome study comparing sampling
291 methods using cytobrush or swab [19]. The average storage period from sample collection via
292 LBC to DNA extraction was about two years in this study. Kim *et al*. reported that DNA from
293 the cervix stored in ThinPrep at room temperature or −80°C was stable for at least one year [49].

294 Meanwhile, Castle *et al*. reported that β-globin DNA fragments of 268 bases or more were
295 detected by PCR in 90 % (27 of 30 samples) of ThinPrep samples stored for eight years at an
296 uncontrolled ambient temperature followed by a controlled ambient environment (10–26.7°C)
297 [50]. Low-temperature storage may allow the analysis of the short DNA fragments of the V4
298 region after even long-term storage, although further research is needed to confirm the optimal
299 storage period in cervical microbiome studies using ThinPrep. SurePath LBC specimens are as
300 widely used as ThinPrep, but the presence of formaldehyde within the SurePath preservation
301 solution raises concerns about accessing enough DNA for analysis as compared to ThinPrep,
302 which contains methanol [51] [52]. It should also be noted that other storage solutions, *i.e.* those
303 using guanidine thiocyanate have been reported for microbiome surveys of the cervix [53] and
304 feces [54]. A weakness of the current study is that we did not examine the reproducibility of our
305 results as each sample was extracted using each kit once. However, the use of actual patient
306 samples rather than mock samples is a strength of our approach.

307 **Conclusions**

308 In conclusion, regardless of the extraction protocol used, all kits provided equivalent broad
309 accessibility to the cervical microbiome. Observed differences in microbial composition were
310 due to the significant influence of the individual patient and not the extraction protocol.
311 However, ZymoBIOMICS was observed to increase the accessibility of DNA from a greater
312 range of microbiota compared to the other kits, in that the greatest number of significantly
313 enriched taxa were identified. This was not because of higher DNA yield nor ability to detect
314 more gram-positive bacteria. We have shown that the ability to characterize cervical microbiota
315 from LBC specimens is robust, even after prolonged storage. Our data also suggest that it is
316 possible to reliably assess the relationship between HPV and the cervical microbiome, also
317 supported by Kim *et al*. [49] and Castle *et al* [50]. Cervical microbiome in patients with HPV16
318 or HPV18 which causes 70% of cervical cancers and CIN [55] warrants critical future study.
319 Selection and characterization of appropriate DNA extraction methods are important for
320 providing an accurate census of cervical microbiota and the human microbiome in general [23]
321 [24] [25] [38] [49] [50]. Even though we found all four extraction kits to be commensurate in
322 their ability to broadly characterize the CM, this study lends support to the view that the
323 selection of a DNA extraction kit depends on the questions asked of the data, and should be
324 taken into account for any cervicovaginal microbiome and HPV research that leverages LBC
325 specimens for use in clinical practice [15] [56].

326 **Methods**

327 **Sampling of cervical microbiome**

328 LBC specimens were obtained from 20 patients enrolled in a Phase II clinical trial of an HPV
329 therapeutic vaccine (NCT02481414). In order to be eligible, participants had to have high grade
330 squamous intraepithelial lesions (HSILs) or cannot rule out HSILs in cervical cytology or
331 CIN2/3 in cervical biopsy. Those who qualified for the study based on their cervical cytology
332 underwent cervical biopsy, and they qualified for vaccination if the results were CIN2/3. The
333 cervical cytology specimens in this current study were collected before the vaccination and
334 reserved in the vial of the ThinPrep Pap Test (HOLOGIC) as described in Ravilla *et al*. 2019
335 [57]. The storage period from sample collection to DNA extraction was 716 ± 105 days in this
336 study.

337

338 **HPV genotyping**

339 HPV-DNA was detected by Linear Array HPV Genotyping Test (Roche Diagnostics) which can
340 detect up to 37 HPV genotypes including 13 HR-HPV genotypes (16, 18, 31, 33, 35, 39, 45, 51,
341 52, 56, 58, 59, and 68) and 24 LR-HPV genotypes (6, 11, 26, 40, 42, 53, 54, 55, 61, 62, 64, 66,
342 67, 69, 70, 71, 72, 73, 81, 82, 83, 84, IS39, and CP6108) using ThinPrep solution [58].

343

344 **DNA extraction protocols**

345 We selected four commercially available DNA extraction kits as the candidates for comparison:
346 ZymoBIOMICS DNA Miniprep Kit (Zymo Research, D4300), QIAamp PowerFecal Pro DNA
347 Kit (QIAGEN, 51804), QIAamp DNA Mini Kit (QIAGEN, 51304), and IndiSpin Pathogen Kit
348 (Indical Bioscience, SPS4104). These kits have been successfully used in a variety of human

349 cervical, vaginal, and gut microbiome surveys [10] [19] [59]. We'll subsequently refer to each of
350 these kits in abbreviated form as follows: ZymoBIOMICS, PowerFecalPro, QIAampMini, and
351 IndiSpin. The protocols and any modifications are outlined in Table 1.

352 Each LBC sample was dispensed into four separate 2 mL sterile collection tubes
353 (dispensed sample volume = 500 μL) to create four cohorts of 20 DNA extractions (Figure 1).
354 Each extraction cohort was processed through one of the four kits above. A total of 80
355 extractions (4 kits × 20 patients) were prepared for subsequent analyses. Applied sample volume
356 of ThinPrep solution was 300 μL for ZymoBIOMICS, 300 μL for PowerFecalPro, 200 μL for
357 QIAampMini, and 300 μL for IndiSpin. The sample volume was standardized to 300 μL as long
358 as the manufacturer's instructions allowed to do so. DNA extraction for all samples was
359 performed by the same individual who practiced by performing multiple extractions for each kit
360 before performing the actual DNA extraction on the samples analyzed in this study. Positive
361 control was mock vaginal microbial communities composed of a mixture of genomic DNA from
362 the American Type Culture Collection (ATCC MSA1007). Negative control was the ThinPrep
363 preservation solution without the sample as blank extraction [60].

364

365 **Measurement of DNA yield**

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

369

370 **16S rRNA marker gene sequencing**

371 Controls and the extracted DNA were sent to Argonne National Laboratory (IL, USA) for
372 amplification and sequencing of the 16S rRNA gene on an Illumina MiSeq sequencing platform.
373 Paired-end reads from libraries with ~250-bp inserts were generated for the V4 region using the
374 barcoded primer set: 515FB: 5'-GTGYCAGCMGCCGCGGTAA-3' and 806RB: 5'-
375 GGACTACNVGGGTWTCTAAT-3' [61] [62] [63] [64] [65]. MiSeq Reagent Kit v2 (2 × 150
376 cycles, MS-102-2002) was used.
377

378 **Sequence processing and analysis**

379 Initial sequence processing and analyses were performed using QIIME 2 [66], any commands
380 prefixed by `q2-` are QIIME 2 plugins. After demultiplexing of the paired-end reads by `q2-`
381 `demux`, the imported sequence data was visually inspected via QIIME 2 View [67], to determine
382 the appropriate trimming and truncation parameters for generating Exact Sequence Variants
383 (ESVs) [68] via `q2-dada2` [69]. ESVs will be referred to as Operational Taxonomic Units
384 (OTUs). The forward reads were trimmed at 15 bp and truncated at 150 bp; reverse reads were
385 trimmed at 0 bp and truncated at 150 bp. The resulting OTUs were assigned taxonomy through
386 `q2-feature-classifier classify-sklearn`, by using a pre-trained classifier for the
387 amplicon region of interest [70]. This enables more robust taxonomic assignment of the OTUs
388 [71]. Taxonomy-based filtering was performed by using `q2-taxa filter-table` to remove
389 any OTUs that were classified as "Chloroplast", "Mitochondria", "Eukaryota", "Unclassified"
390 and those that did not have at least a Phylum-level classification. We then performed additional
391 quality filtering via `q2-quality-control`, and only retained OTUs that had at least a 90%
392 identity and 90% query alignment to the SILVA reference set [72]. Then `q2-alignment` was
393 used to generate a *de novo* alignment with MAFFT [73] which was subsequently masked by

394 setting `max-gap-frequency 1 min-conservation 0.4`. Finally, `q2-phylogeny`
395 was used to construct a midpoint-rooted phylogenetic tree using IQ-TREE [74] with automatic
396 model selection using ModelFinder [75]. Unless specified, subsequent analyses were performed
397 after removing OTUs with a very low frequency [76], of less than 0.0005% of the total data set
398 in this case.
399
400 **Number of reads and OTUs before rarefying**
401 Table 3 highlights the numbers of reads and OTUs among the DNA extraction protocols prior to
402 rarefying the data. The reads and OTUs assigned to gram-positive and gram-negative are also
403 shown. The number of "OTUs before rarefying" shown in Table 3 is distinguished from the
404 "Observed OTUs" after rarefying in Figure 3 for diversity analysis.
405
406 **Microbiome analysis**
407 To compare the taxonomic profiles among four types of DNA extraction methods (Figure 1 &
408 Table 1), the following analyses were performed; (I) bacterial microbiome composition, (II)
409 detection of common and unique taxa, (III) alpha and beta diversity analysis, and (IV)
410 identification of specific bacteria retained per DNA extraction method.
411
412 **Microbiome composition**
413 We generated the bar plot to exhibit bacterial microbiome composition per DNA extraction
414 method at the family (Figure 2A left) and genus (Figure 2A right) taxonomic level. After all
415 count data of taxonomy were converted to relative abundance, the top 10 abundant taxonomic
416 groups in each family and genus level were plotted in colored bar plot [77] [78] [79]. Variation

417 of microbiome composition per DNA extraction method or per individual was assessed by the
418 Adonis test (`q2-diversity adonis`) [80] [81].
419
420 **Differentially accessible microbiota by DNA extraction protocol**
421 We set out to determine which microbial taxonomic groups were differentially accessible across
422 the sampling protocols by LEfSe analyses [31]. We further assessed the microbial taxa using
423 jvenn [82] at family and genus level. The Venn diagram was created after removing OTUs with a
424 frequency of less than 0.005% [76].
425
426 **Alpha and beta diversity analyses with or without rarefying**
427 Non-rarefying approaches to determine both alpha (within-sample) and beta (between-sample)
428 diversity was assessed by Species richness using `q2-breakaway` [36] and Aitchison distance
429 using `q2-deicode` [37]. These were compared with rarefied data in which we applied Faith's
430 Phylogenetic Diversity, Observed OTUs, Shannon's diversity index, Pielou's Evenness,
431 Unweighted UniFrac distance, Weighted UniFrac distance, Jaccard distance, and Bray-Curtis
432 distances via `q2-diversity` [66]. In order to retain data from at least 15 of the 20 patients
433 (*i.e.* 75%; four samples from each of the four DNA extraction methods), we set the sampling
434 depth to 51,197 reads per sample. Overall our subsequence analysis consisted of 3,071,820 reads
435 (27.6%, 3,071,820 / 11,149,582 reads). All diversity measurements in this study are listed in
436 Table 5.
437
438 **Community type and HPV status**

439 In addition to the analysis above, we tested whether the samples clustered by microbiome
440 composition were related to the patient's clinical and demographic characteristics such as,
441 cervical biopsy diagnosis, race, and HPV16 status. HPV16 status has been reported to be
442 associated with both racial differences as well as microbial community types [57] [83] [84] [85].
443 We employed the DMM [32] model to determine the number of community types for bacterial
444 cervical microbiome. Then, we clustered samples to the community type [9] [86]. Since vaginal
445 microbiota were reported to be clustered with different *Lactobacillus sp*. such as *L. crispatus*, *L.*
446 *gasseri*, *L. iners*, or *L. jensenii* [16] [87], we also collapsed the taxonomy to the species level and
447 performed a clustering analysis using "microbiome R package" [79]. We then determined which
448 bacterial taxa were differentially abundant among the patients with or without HPV16 via `q2-`
449 `aldex2` [88] and LEfSe [31].
450

451 **General statistical analysis**

452 All data are presented as means ± standard deviation (SD). Comparisons were conducted with
453 Fisher's exact test or Dunn's test with Benjamini-Hochberg-adjustment [89] or Wilcoxon test
454 with Benjamini-Hochberg-adjustment or pairwise PERMANOVA when appropriate. A p value <
455 0.05 or a q value < 0.05 was considered statistically significant.

456 **Declarations**

457 **Ethics approval and consent to participate**

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

460

461 **Consent for publication**

462 Written informed consent for publication was obtained for all patients.

463

464 **Availability of data and materials**

465 MIMARKS compliant [90] DNA sequencing data are available via the Sequence Read Archive
466 (SRA) at the National Center for Biotechnology Information (NCBI), under the BioProject
467 Accession: PRJNA598197.

468

469 **Competing interests**

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

472

473 **Funding**

474 This work was supported by the National Institutes of Health (R01CA143130, USA), Drs. Mae
475 and Anderson Nettleship Endowed Chair of Oncologic Pathology (31005156, USA), and the
476 Arkansas Biosciences Institute (the major component of the Tobacco Settlement Proceeds Act of
477 2000, G1-52249-01, USA).

478

479 **Authors' contributions**

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

485

486 **Acknowledgements**

487 We thank Togo Picture Gallery [91] for stock images shown in Figure 1.

488

489 **Authors' information**

490 **Affiliations**

491 Department of Pathology, University of Arkansas for Medical Sciences, Little Rock, AR, USA**:**
492 Takeo Shibata, Mayumi Nakagawa, & Hannah N. Coleman
493 Department of Obstetrics and Gynecology, Kanazawa Medical University, Uchinada, Ishikawa,
494 Japan: Takeo Shibata & Toshiyuki Sasagawa
495 Biosciences Division, Argonne National Laboratory, Lemont, IL, USA: Sarah M. Owens
496 Department of Obstetrics and Gynecology, University of Arkansas for Medical Sciences, Little
497 Rock, AR, USA: William W. Greenfield
498 Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little
499 Rock, AR, USA: Michael S. Robeson II

500

501 **Corresponding author**

Correspondence to Michael S. Robeson II

503 **References**

504 1. Lee JE, Lee S, Lee H, Song YM, Lee K, Han MJ, et al. Association of the vaginal
505 microbiota with human papillomavirus infection in a Korean twin cohort. PLoS One.
506 2013;8(5):e63514; doi: 10.1371/journal.pone.0063514.
507 2. Huang X, Li C, Li F, Zhao J, Wan X, Wang K. Cervicovaginal microbiota composition
508 correlates with the acquisition of high-risk human papillomavirus types. Int J Cancer.
509 2018;143(3):621-34; doi: 10.1002/ijc.31342.
510 3. Zhou Y, Wang L, Pei F, Ji M, Zhang F, Sun Y, et al. Patients With LR-HPV Infection
511 Have a Distinct Vaginal Microbiota in Comparison With Healthy Controls. Front Cell
512 Infect Microbiol. 2019;9:294; doi: 10.3389/fcimb.2019.00294.
513 4. Onywera H, Williamson AL, Mbulawa ZZA, Coetzee D, Meiring TL. The cervical
514 microbiota in reproductive-age South African women with and without human
515 papillomavirus infection. Papillomavirus Res. 2019;7:154-63; doi:
516 10.1016/j.pvr.2019.04.006.
517 5. Brotman RM, Shardell MD, Gajer P, Tracy JK, Zenilman JM, Ravel J, et al. Interplay
518 between the temporal dynamics of the vaginal microbiota and human papillomavirus
519 detection. J Infect Dis. 2014;210(11):1723-33; doi: 10.1093/infdis/jiu330.
520 6. Godoy-Vitorino F, Romaguera J, Zhao C, Vargas-Robles D, Ortiz-Morales G, Vazquez-
521 Sanchez F, et al. Cervicovaginal Fungi and Bacteria Associated With Cervical
522 Intraepithelial Neoplasia and High-Risk Human Papillomavirus Infections in a Hispanic
523 Population. Front Microbiol. 2018;9:2533; doi: 10.3389/fmicb.2018.02533.
524 7. Łaniewski P, Barnes D, Goulder A, Cui H, Roe DJ, Chase DM, et al. Linking
525 cervicovaginal immune signatures, HPV and microbiota composition in cervical
526 carcinogenesis in non-Hispanic and Hispanic women. In: Sci Rep. 2018.
527 8. Mitra A, MacIntyre DA, Lee YS, Smith A, Marchesi JR, Lehne B, et al. Cervical
528 intraepithelial neoplasia disease progression is associated with increased vaginal
529 microbiome diversity. Sci Rep. 2015;5:16865; doi: 10.1038/srep16865.
530 9. Piyathilake CJ, Ollberding NJ, Kumar R, Macaluso M, Alvarez RD, Morrow CD.
531 Cervical Microbiota Associated with Higher Grade Cervical Intraepithelial Neoplasia in
532 Women Infected with High-Risk Human Papillomaviruses. Cancer Prev Res (Phila).
533 2016;9(5):357-66; doi: 10.1158/1940-6207.CAPR-15-0350.
534 10. Oh HY, Kim BS, Seo SS, Kong JS, Lee JK, Park SY, et al. The association of uterine
535 cervical microbiota with an increased risk for cervical intraepithelial neoplasia in Korea.
536 Clin Microbiol Infect. 2015;21(7):674 e1-9; doi: 10.1016/j.cmi.2015.02.026.
537 11. De Seta F, Campisciano G, Zanotta N, Ricci G, Comar M. The Vaginal Community State
538 Types Microbiome-Immune Network as Key Factor for Bacterial Vaginosis and Aerobic
539 Vaginitis. Front Microbiol. 2019;10:2451; doi: 10.3389/fmicb.2019.02451.
540 12. Oliver A, LaMere B, Weihe C, Wandro S, Lindsay KL, Wadhwa PD, et al.
541 Cervicovaginal microbiome composition drives metabolic profiles in healthy pregnancy.
542 bioRxiv https://doi.org/10.1101/840520. 2019.
543 13. Human Microbiome Project Consortium. Structure, function and diversity of the healthy
544 human microbiome. Nature. 2012;486(7402):207-14; doi: 10.1038/nature11234.
545 14. Bik EM, Bird SW, Bustamante JP, Leon LE, Nieto PA, Addae K, et al. A novel
546 sequencing-based vaginal health assay combining self-sampling, HPV detection and

547 genotyping, STI detection, and vaginal microbiome analysis. PLoS One.
548 2019;14(5):e0215945; doi: 10.1371/journal.pone.0215945.
549 15. Berman HL, McLaren MR, Callahan BJ. Understanding and Interpreting Community
550 Sequencing Measurements of the Vaginal Microbiome. BJOG. 2019; doi: 10.1111/1471-
551 0528.15978.
552 16. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, et al. Vaginal
553 microbiome of reproductive-age women. Proc Natl Acad Sci U S A. 2011;108 Suppl
554 1:4680-7; doi: 10.1073/pnas.1002611107.
555 17. Fettweis JM, Serrano MG, Brooks JP, Edwards DJ, Girerd PH, Parikh HI, et al. The
556 vaginal microbiome and preterm birth. Nat Med. 2019;25(6):1012-21; doi:
557 10.1038/s41591-019-0450-2.
558 18. Tuominen H, Rautava S, Syrjanen S, Collado MC, Rautava J. HPV infection and
559 bacterial microbiota in the placenta, uterine cervix and oral mucosa. Sci Rep.
560 2018;8(1):9787; doi: 10.1038/s41598-018-27980-3.
561 19. Mitra A, MacIntyre DA, Mahajan V, Lee YS, Smith A, Marchesi JR, et al. Comparison
562 of vaginal microbiota sampling techniques: cytobrush versus swab. Sci Rep.
563 2017;7(1):9802; doi: 10.1038/s41598-017-09844-4.
564 20. Bentz JS. Liquid-based cytology for cervical cancer screening. Expert Rev Mol Diagn.
565 2005;5(6):857-71; doi: 10.1586/14737159.5.6.857.
566 21. Gibb RK, Martens MG. The impact of liquid-based cytology in decreasing the incidence
567 of cervical cancer. Rev Obstet Gynecol. 2011;4(Suppl 1):S2-S11.
568 22. Donders GG, Depuydt CE, Bogers JP, Vereecken AJ. Association of Trichomonas
569 vaginalis and cytological abnormalities of the cervix in low risk women. PLoS One.
570 2013;8(12):e86266; doi: 10.1371/journal.pone.0086266.
571 23. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, et al. Towards
572 standards for human fecal sample processing in metagenomic studies. Nat Biotechnol.
573 2017;35(11):1069-76; doi: 10.1038/nbt.3960.
574 24. Stinson LF, Keelan JA, Payne MS. Comparison of Meconium DNA Extraction Methods
575 for Use in Microbiome Studies. Front Microbiol. 2018;9:270; doi:
576 10.3389/fmicb.2018.00270.
577 25. Teng F, Darveekaran Nair SS, Zhu P, Li S, Huang S, Li X, et al. Impact of DNA
578 extraction method and targeted 16S-rRNA hypervariable region on oral microbiota
579 profiling. Sci Rep. 2018;8(1):16321; doi: 10.1038/s41598-018-34294-x.
580 26. Balle C, Lennard K, Dabee S, Barnabas SL, Jaumdally SZ, Gasper MA, et al.
581 Endocervical and vaginal microbiota in South African adolescents with asymptomatic
582 Chlamydia trachomatis infection. Sci Rep. 2018;8(1):11109; doi: 10.1038/s41598-018-
583 29320-x.
584 27. Klein C, Gonzalez D, Samwel K, Kahesa C, Mwaiselage J, Aluthge N, et al. Relationship
585 between the Cervical Microbiome, HIV Status, and Precancerous Lesions. MBio.
586 2019;10(1); doi: 10.1128/mBio.02785-18.
587 28. Hayashi NR, Ishida T, Yokota A, Kodama T, Igarashi Y. Hydrogenophilus
588 thermoluteolus gen. nov., sp. nov., a thermophilic, facultatively chemolithoautotrophic,
589 hydrogen-oxidizing bacterium. Int J Syst Bacteriol. 1999;49 Pt 2:783-6; doi:
590 10.1099/00207713-49-2-783.

591 29. Glassing A, Dowd SE, Galandiuk S, Davis B, Chiodini RJ. Inherent bacterial DNA
592 contamination of extraction and sequencing reagents may affect interpretation of
593 microbiota in low bacterial biomass samples. In: Gut Pathog. 2016.
594 30. Birse KD, Romas LM, Guthrie BL, Nilsson P, Bosire R, Kiarie J, et al. Genital Injury
595 Signatures and Microbiome Alterations Associated With Depot Medroxyprogesterone
596 Acetate Usage and Intravaginal Drying Practices. J Infect Dis. 2017;215(4):590-8; doi:
597 10.1093/infdis/jiw590.
598 31. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic
599 biomarker discovery and explanation. Genome Biol. 2011;12(6):R60; doi: 10.1186/gb-
600 2011-12-6-r60.
601 32. Morgan M. DirichletMultinomial: Dirichlet-Multinomial Mixture Model Machine
602 Learning for Microbiome Data.
603 http://bioconductor.org/packages/release/bioc/html/DirichletMultinomial.html. Accessed
604 12 Mar 2020.
605 33. Lennard K, Dabee S, Barnabas SL, Havyarimana E, Blakney A, Jaumdally SZ, et al.
606 Microbial Composition Predicts Genital Tract Inflammation and Persistent Bacterial
607 Vaginosis in South African Adolescent Females. Infect Immun. 2018;86(1); doi:
608 10.1128/IAI.00410-17.
609 34. McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is
610 inadmissible. PLoS Comput Biol. 2014;10(4):e1003531; doi:
611 10.1371/journal.pcbi.1003531.
612 35. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and
613 microbial differential abundance strategies depend upon data characteristics.
614 Microbiome. 2017;5(1):27; doi: 10.1186/s40168-017-0237-y.
615 36. Willis A, Bunge J. Estimating diversity via frequency ratios. Biometrics.
616 2015;71(4):1042-9; doi: 10.1111/biom.12332.
617 37. Martino C, Morton JT, Marotz CA, Thompson LR, Tripathi A, Knight R, et al. A Novel
618 Sparse Compositional Technique Reveals Microbial Perturbations. mSystems. 2019;4(1);
619 doi: 10.1128/mSystems.00016-19.
620 38. Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ. Evaluation of methods for the
621 extraction and purification of DNA from the human microbiome. PLoS One.
622 2012;7(3):e33865; doi: 10.1371/journal.pone.0033865.
623 39. Callahan BJ, Wong J, Heiner C, Oh S, Theriot CM, Gulati AS, et al. High-throughput
624 amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution.
625 Nucleic Acids Res. 2019;47(18):e103; doi: 10.1093/nar/gkz569.
626 40. Calus ST, Ijaz UZ, Pinto AJ. NanoAmpli-Seq: a workflow for amplicon sequencing for
627 mixed microbial communities on the nanopore sequencing platform. Gigascience.
628 2018;7(12); doi: 10.1093/gigascience/giy140.
629 41. Wongsurawat T, Nakagawa M, Atiq O, Coleman HN, Jenjaroenpun P, Allred JI, et al. An
630 assessment of Oxford Nanopore sequencing for human gut metagenome profiling: A pilot
631 study of head and neck cancer patients. J Microbiol Methods. 2019;166:105739; doi:
632 10.1016/j.mimet.2019.105739.
633 42. Usyk M, Zolnik CP, Castle PE, Porras C, Herrero R, Gradissimo A, et al. Cervicovaginal
634 microbiome and natural history of HPV in a longitudinal study. PLoS Pathog.
635 2020;16(3):e1008376; doi: 10.1371/journal.ppat.1008376.

636 43. Audirac-Chalifour A, Torres-Poveda K, Bahena-Roman M, Tellez-Sosa J, Martinez-
637 Barnetche J, Cortina-Ceballos B, et al. Cervical Microbiome and Cytokine Profile at
638 Various Stages of Cervical Cancer: A Pilot Study. PLoS One. 2016;11(4):e0153274; doi:
639 10.1371/journal.pone.0153274.
640 44. Di Paola M, Sani C, Clemente AM, Iossa A, Perissi E, Castronovo G, et al.
641 Characterization of cervico-vaginal microbiota in women developing persistent high-risk
642 Human Papillomavirus infection. Sci Rep. 2017;7(1):10200; doi: 10.1038/s41598-017-
643 09842-6.
644 45. Ranjeva SL, Mihaljevic JR, Joseph MB, Giuliano AR, Dwyer G. Untangling the
645 dynamics of persistence and colonization in microbial communities. ISME J. 2019:1-13;
646 doi: 10.1038/s41396-019-0488-7.
647 46. Linder J, Zahniser D. ThinPrep Papanicolaou testing to reduce false-negative cervical
648 cytology. Arch Pathol Lab Med. 1998;122(2):139-44.
649 47. Ling Z, Liu X, Chen X, Zhu H, Nelson KE, Xia Y, et al. Diversity of cervicovaginal
650 microbiota associated with female lower genital tract infections. Microb Ecol.
651 2011;61(3):704-14; doi: 10.1007/s00248-011-9813-z.
652 48. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and
653 laboratory contamination can critically impact sequence-based microbiome analyses.
654 BMC Biol. 2014;12(1):87; doi: 10.1186/s12915-014-0087-z.
655 49. Kim Y, Choi KR, Chae MJ, Shin BK, Kim HK, Kim A, et al. Stability of DNA, RNA,
656 cytomorphology, and immunoantigenicity in Residual ThinPrep Specimens. APMIS.
657 2013;121(11):1064-72; doi: 10.1111/apm.12082.
658 50. Castle PE, Solomon D, Hildesheim A, Herrero R, Concepcion Bratti M, Sherman ME, et
659 al. Stability of archived liquid-based cervical cytologic specimens. Cancer.
660 2003;99(2):89-96; doi: 10.1002/cncr.11058.
661 51. Rebolj M, Rask J, van Ballegooijen M, Kirschner B, Rozemeijer K, Bonde J, et al.
662 Cervical histology after routine ThinPrep or SurePath liquid-based cytology and
663 computer-assisted reading in Denmark. Br J Cancer. 2015;113(9):1259-74; doi:
664 10.1038/bjc.2015.339.
665 52. Naeem RC, Goldstein DY, Einstein MH, Ramos Rivera G, Schlesinger K, Khader SN, et
666 al. SurePath Specimens Versus ThinPrep Specimen Types on the COBAS 4800 Platform:
667 High-Risk HPV Status and Cytology Correlation in an Ethnically Diverse Bronx
668 Population. Lab Med. 2017;48(3):207-13; doi: 10.1093/labmed/lmx019.
669 53. Ritu W, Enqi W, Zheng S, Wang J, Ling Y, Wang Y. Evaluation of the Associations
670 Between Cervical Microbiota and HPV Infection, Clearance, and Persistence in
671 Cytologically Normal Women. Cancer Prev Res (Phila). 2019;12(1):43-56; doi:
672 10.1158/1940-6207.CAPR-18-0233.
673 54. Hosomi K, Ohno H, Murakami H, Natsume-Kitatani Y, Tanisawa K, Hirata S, et al.
674 Method for preparing DNA from feces in guanidine thiocyanate solution affects 16S
675 rRNA-based profiling of human microbiota diversity. Sci Rep. 2017;7(1):4339; doi:
676 10.1038/s41598-017-04511-0.
677 55. Human papillomavirus (HPV) and cervical cancer. https://www.who.int/news-room/fact-
678 sheets/detail/human-papillomavirus-(hpv)-and-cervical-cancer. Accessed 12 Mar 2020.
679 56. Sarangi AN, Goel A, Aggarwal R. Methods for Studying Gut Microbiota: A Primer for
680 Physicians. J Clin Exp Hepatol. 2019;9(1):62-73; doi: 10.1016/j.jceh.2018.04.016.

681 57. Ravilla R, Coleman HN, Chow CE, Chan L, Fuhrman BJ, Greenfield WW, et al. Cervical
682 Microbiome and Response to a Human Papillomavirus Therapeutic Vaccine for Treating
683 High-Grade Cervical Squamous Intraepithelial Lesion. Integr Cancer Ther.
684 2019;18:1534735419893063; doi: 10.1177/1534735419893063.
685 58. Roche Molecular Diagnostics. LINEAR ARRAY® HPV Genotyping.
686 https://diagnostics.roche.com/global/en/products/params/linear-array-hpv-
687 genotyping.html. Accessed 12 Mar 2020.
688 59. Virtanen S, Kalliala I, Nieminen P, Salonen A. Comparative analysis of vaginal
689 microbiota sampling using 16S rRNA gene analysis. PLoS One. 2017;12(7):e0181477;
690 doi: 10.1371/journal.pone.0181477.
691 60. Kim D, Hofstaedter CE, Zhao C, Mattei L, Tanes C, Clarke E, et al. Optimizing methods
692 and dodging pitfalls in microbiome research. Microbiome. 2017;5(1):52; doi:
693 10.1186/s40168-017-0267-5.
694 61. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A
695 communal catalogue reveals Earth's multiscale microbial diversity. Nature.
696 2017;551(7681):457-63; doi: 10.1038/nature24621.
697 62. Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA
698 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat Microb
699 Ecol. 2015;75(2):129-37; doi: 10.3354/ame01753.
700 63. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit
701 rRNA primers for marine microbiomes with mock communities, time series and global
702 field samples. Environ Microbiol. 2016;18(5):1403-14; doi: 10.1111/1462-2920.13023.
703 64. Walters W, Hyde ER, Berg-Lyons D, Ackermann G, Humphrey G, Parada A, et al.
704 Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed
705 Spacer Marker Gene Primers for Microbial Community Surveys. mSystems. 2016;1(1);
706 doi: 10.1128/mSystems.00009-15.
707 65. Earth Microbiome Project. 16S Illumina amplicon protocol.
708 http://www.earthmicrobiome.org/protocols-and-standards/16s/. Accessed 12 Mar 2020.
709 66. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al.
710 Reproducible, interactive, scalable and extensible microbiome data science using QIIME
711 2. Nat Biotechnol. 2019;37(8):852-7; doi: 10.1038/s41587-019-0209-9.
712 67. QIIME 2 View. https://view.qiime2.org/. Accessed 12 Mar 2020.
713 68. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace
714 operational taxonomic units in marker-gene data analysis. ISME J. 2017;11(12):2639-43;
715 doi: 10.1038/ismej.2017.119.
716 69. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2:
717 High-resolution sample inference from Illumina amplicon data. Nat Methods.
718 2016;13(7):581-3; doi: 10.1038/nmeth.3869.
719 70. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing
720 taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-
721 classifier plugin. Microbiome. 2018;6(1):90; doi: 10.1186/s40168-018-0470-z.
722 71. Werner JJ, Koren O, Hugenholtz P, DeSantis TZ, Walters WA, Caporaso JG, et al.
723 Impact of training sets on classification of high-throughput bacterial 16s rRNA gene
724 surveys. ISME J. 2012;6(1):94-103; doi: 10.1038/ismej.2011.82.

725 72. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA
726 ribosomal RNA gene database project: improved data processing and web-based tools.
727 Nucleic Acids Res. 2013;41(Database issue):D590-6; doi: 10.1093/nar/gks1219.
728 73. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
729 improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-80; doi:
730 10.1093/molbev/mst010.
731 74. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective
732 stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol.
733 2015;32(1):268-74; doi: 10.1093/molbev/msu300.
734 75. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder:
735 fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587-9;
736 doi: 10.1038/nmeth.4285.
737 76. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, et al. Quality-
738 filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat
739 Methods. 2013;10(1):57-9; doi: 10.1038/nmeth.2276.
740 77. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis
741 and graphics of microbiome census data. PLoS One. 2013;8(4):e61217; doi:
742 10.1371/journal.pone.0061217.
743 78. Bisanz JE. qiime2R: Importing QIIME2 artifacts and associated data into R sessions.
744 https://github.com/jbisanz/qiime2R. Accessed 12 Mar 2020.
745 79. Lahti L, Shetty S. microbiome R package. http://microbiome.github.io. Accessed 12 Mar
746 2020.
747 80. Anderson MJ. A new method for non-parametric multivariate analysis of variance.
748 Austral Ecol. 2001;26(1):32-46; doi: DOI 10.1111/j.1442-9993.2001.01070.pp.x.
749 81. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan:
750 Community Ecology Package. R package version 2.5-3. https://CRAN.R-
751 project.org/package=vegan. Accessed 12 Mar 2020.
752 82. Bardou P, Mariette J, Escudie F, Djemiel C, Klopp C. jvenn: an interactive Venn diagram
753 viewer. BMC Bioinformatics. 2014;15(1):293; doi: 10.1186/1471-2105-15-293.
754 83. Gao W, Weng J, Gao Y, Chen X. Comparison of the vaginal microbiota diversity of
755 women with and without human papillomavirus infection: a cross-sectional study. BMC
756 Infect Dis. 2013;13(1):271; doi: 10.1186/1471-2334-13-271.
757 84. Montealegre JR, Peckham-Gregory EC, Marquez-Do D, Dillon L, Guillaud M, Adler-
758 Storthz K, et al. Racial/ethnic differences in HPV 16/18 genotypes and integration status
759 among women with a history of cytological abnormalities. Gynecol Oncol.
760 2018;148(2):357-62; doi: 10.1016/j.ygyno.2017.12.014.
761 85. Xi LF, Kiviat NB, Hildesheim A, Galloway DA, Wheeler CM, Ho J, et al. Human
762 papillomavirus type 16 and 18 variants: race-related distribution and persistence. J Natl
763 Cancer Inst. 2006;98(15):1045-52; doi: 10.1093/jnci/djj297.
764 86. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for
765 microbial metagenomics. PLoS One. 2012;7(2):e30126; doi:
766 10.1371/journal.pone.0030126.
767 87. DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, et al.
768 Temporal and spatial variation of the human microbiota during pregnancy. Proc Natl
769 Acad Sci U S A. 2015;112(35):11060-5; doi: 10.1073/pnas.1502875112.

770 88. Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB. ANOVA-like differential
771 expression (ALDEx) analysis for mixed population RNA-Seq. PLoS One.
772 2013;8(7):e67019; doi: 10.1371/journal.pone.0067019.
773 89. Dinno A. dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums.
774 https://CRAN.R-project.org/package=dunn.test. Accessed 12 Mar 2020.
775 90. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, et al. Minimum
776 information about a marker gene sequence (MIMARKS) and minimum information about
777 any (x) sequence (MIxS) specifications. Nat Biotechnol. 2011;29(5):415-20; doi:
778 10.1038/nbt.1823.
779 91. Togo Picture Gallery. http://togotv.dbcls.jp/pics.html. Accessed 12 Mar 2020.
780 92. Microbial Isolation | ZYMO RESEARCH.
781 https://www.zymoresearch.com/pages/microbial-isolation. Accessed 12 Mar 2020.
782 93. PowerBead Tubes - QIAGEN Online Shop.
783 https://www.qiagen.com/us/products/discovery-and-translational-research/lab-
784 essentials/plastics/powerbead-tubes/#orderinginformation. Accessed 12 Mar 2020.
785 94. QIAGEN. Pathogen Lysis Tubes - QIAGEN.
786 https://www.qiagen.com/dk/shop/pcr/pathogen-lysis-tubes/. Accessed 12 Mar 2020.
787 95. Silhavy TJ, Kahne D, Walker S. The bacterial cell envelope. Cold Spring Harb Perspect
788 Biol. 2010;2(5):a000414; doi: 10.1101/cshperspect.a000414.
789

790

Table 1: Characteristics of four different DNA extraction protocols

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

a: [92]. 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 [93]. e: Instead of lysozyme or lysostaphin, mutanolysin was used as per Yuan *et al*, 2012 [38]. 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 [94]. i: Pretreatment B2 as per QIAamp cador Pathogen Mini Handbook.

791

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)
	Caucasian, n (%)	10 (50)
	Hispanic, n (%)	7 (35)
Cervical biopsy	CIN2, n (%)	8 (40)
	CIN3, n (%)	10 (50)
	Benign, n (%)	2 (10)
HPV typing	HPV positive, n (%)	19 (95)
	HPV16 positive, n (%)	10 (50)
	HPV18 positive, n (%)	2 (10)
	HPV16 or 18 positives, n (%)	10 (50)
	HR-HPV positives, n (%)	18 (90)

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

792

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

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

Community of gram-positive bacteria was defined as phylum *Actinobacteria* and *Firmicutes*, which are composed of thick peptidoglycan layers without outer membrane [95]. Community of gram-negative bacteria was defined as a community of bacteria other than phylum *Actinobacteria* and *Firmicutes* in this study. a: I - P: 0.0199; I - Q: 0.1590; P - Q: 0.1436; I - Z: 0.1495; P - Z: 0.1712; and Q - Z: 0.4059. b: I - P: 0.2116; I - Q: 0.4837; P - Q: 0.1143; I - Z: 0.0938; P - Z: 0.0116; Q - Z: 0.1448. Dunn's test with Benjamini-Hochberg-adjustment were performed for comparison of the number of reads and OTUs by DNA extraction method. Zy: ZymoBIOMICS DNA Miniprep Kit, Pro: QIAamp PowerFecal Pro DNA Kit, QIA: QIAamp DNA Mini Kit, IN: IndiSpin Pathogen Kit. SD: standard deviation. All: all bacteria, GP: gram-positive bacteria, GN: gram-negative bacteria. NS: not significant.

793

Table 4. Beta diversity among DNA extraction methods

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

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

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

794

Table 5. Diversity analysis in this study

No.	Parameter	Alpha or Beta diversity	Used data with/without rarefying	Input data with/without phylogenetic information	Plugin of QIIME 2
1	Species richness	Alpha	Not rarefied	Non-phylogenetic	`q2-breakaway` [36]
2	Faith's Phylogenetic Diversity	Alpha	Rarefied	Phylogenetic	`q2-diversity`
3	Observed OTUs	Alpha	Rarefied	Non-phylogenetic	`q2-diversity`
4	Shannon's diversity index	Alpha	Rarefied	Non-phylogenetic	`q2-diversity`
5	Pielou's Evenness	Alpha	Rarefied	Non-phylogenetic	`q2-diversity`
6	Aitchison distance	Beta	Not rarefied	Non-phylogenetic	`q2-deicode` [37]
7	Unweighted UniFrac distance	Beta	Rarefied	Phylogenetic	`q2-diversity`
8	Weighted UniFrac distance	Beta	Rarefied	Phylogenetic	`q2-diversity`
9	Jaccard distance	Beta	Rarefied	Non-phylogenetic	`q2-diversity`
10	Bray-Curtis distances	Beta	Rarefied	Non-phylogenetic	`q2-diversity`
11	Adonis	Beta	Rarefied	Non-phylogenetic	`q2-diversity adonis` [80] [81]

795

796 **Figure legends**

797 **Figure 1. Overview of the study design using 16S rRNA gene to compare the DNA**
798 **extraction protocol.** (A) Liquid-based cytology (LBC) specimens from 20 patients with CIN2/3
799 or suspected CIN2/3. (B) A total of 80 DNA extractions were performed. (C) The four DNA
800 extraction methods. (D) DNA of mock vaginal community as a positive control and preservation
801 solution as a negative control. (E) Sequencing using Illumina MiSeq. (F) Analysis of the
802 taxonomic profiles among the DNA extraction protocols. Images form Togo Picture Gallery [91]
803 were used to create this figure.

804

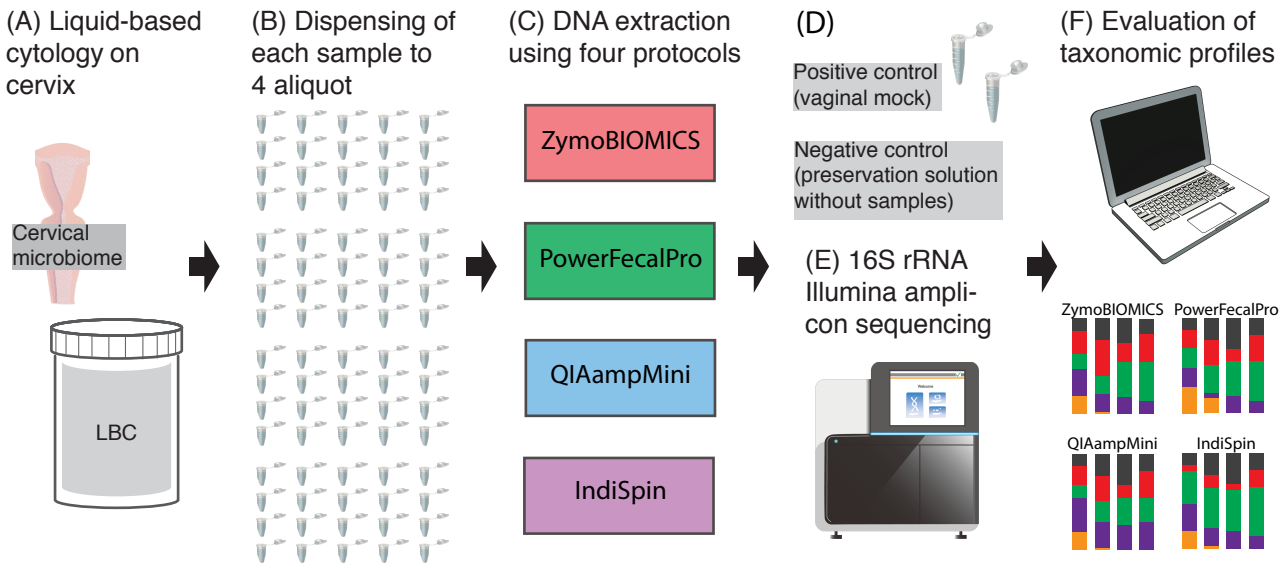
805 **Figure 2. Taxonomic resolution among DNA extraction protocols.** (A) Relative abundance of
806 microbe at family level (left) and genus level (right) per DNA extraction method showed the
807 pattern that variance of microbe composition per patient was higher than that per DNA extraction
808 protocol. These patterns were confirmed by values of Adonis test (`q2-diversity adonis`);
809 F.Model: 199.4, R2: 0.982, and p value: 0.001 for patients and F.Model: 2.9, R2: 0.003, and p
810 value: 0.002 for DNA extraction [80] [81]. After all count data of taxonomy were converted to
811 relative abundance as shown in the y-axis, the top ten taxonomy at each family and genus level
812 were plotted in colored bar plot and other relatively few taxonomies were not plotted. The 20
813 patients ID were described in the x-axis. (B) Venn diagrams, considering only those OTUs with
814 a frequency greater than 0.005% shown, revealed that ZymoBIOMICS had four unique taxa at
815 family (left) and genus (right) taxonomic level. Thirty-one of 41 families and 45 of 57 genera
816 were detected with all DNA extraction protocols.

817

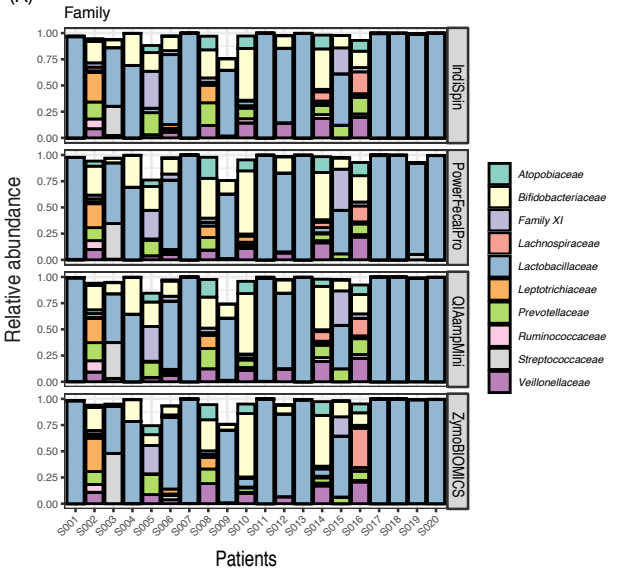
818 **Figure 3. Comparisons of alpha diversity between different DNA extraction protocols.** The
819 alpha diversity indices determined by Species richness and Phylogenetic diversity are
820 significantly higher with ZymoBIOMICS in comparison with PowerFecalPro (p = 0.025 and
821 0.012, respectively, Dunn's test with Benjamini-Hochberg-adjustment). IndiSpin also showed
822 significantly higher diversity than that of PowerFecalPro using analysis of Species richness (p =
823 0.042, Dunn's test with Benjamini-Hochberg-adjustment). No significant differences were
824 observed in other alpha diversity indexes such as observed OTUs, Shannon's diversity index,
825 and Pielou's Evenness. Zy: ZymoBIOMICS DNA Miniprep Kit, Pro: QIAamp PowerFecal Pro
826 DNA Kit, QIA: QIAamp DNA Mini Kit, IN: IndiSpin Pathogen Kit.
827
828 **Figure 4. Distinct detections of microbe among the DNA extraction protocols.** (A) A bar
829 graph showing 23 significantly enriched taxa with ZymoBIOMICS, 3 with QIAamp DNA Mini
830 Kit, and 3 with IndiSpin Pathogen Kit determined by the linear discriminant analysis (LDA)
831 effect size (LEfSe) analyses [31]. (B) A taxonomic cladogram from the same LEfSe analyses
832 showing that the significantly enriched microbiota in ZymoBIOMICS were composed of phylum
833 *Proteobacteria*. Also note that *Meiothermus* (a member of the phylum *Deinococcus-Thermus*)
834 *Hydrogenophilaceae* (a member of the phylum *Proteobacteria*), and *Hydrogenophilus* (a
835 member of the phylum *Proteobacteria*) are likely an extraction kit contaminant. Zy:
836 ZymoBIOMICS DNA Miniprep Kit, Pro: QIAamp PowerFecal Pro DNA Kit, QIA: QIAamp
837 DNA Mini Kit, IN: IndiSpin Pathogen Kit. g_: genus, f_: family, o_: order, c_: class, p_:
838 phylum.
839

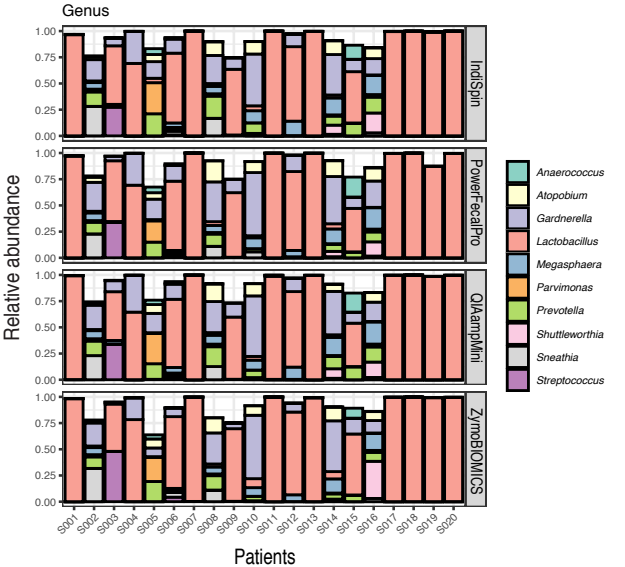
840 **Figure S1. Comparison of DNA yields by DNA extraction protocols.** DNA yield of
841 QIAampMini was significantly higher than that of PowerFecalPro ($p < 0.001$, Dunn's test with
842 Benjamini-Hochberg-adjustment). Also, the DNA yield of ZymoBIOMICS was significantly
843 higher than that of PowerFecalPro ($p < 0.001$, Dunn's test with Benjamini-Hochberg-
844 adjustment). The amount of DNA was calculated based on the absorbance of nucleic acids
845 measured by Nanodrop One. By the protocol recommended by the manufacturer, nucleic acid
846 (Poly-A carrier) was used in IndiSpin. Therefore, IndiSpin was excluded from the analysis of
847 DNA yield. The amount of DNA yield per 100 μL ThinPrep sample volume were compared. The
848 bar graph shows the mean and standard deviation. Zy: ZymoBIOMICS DNA Miniprep Kit, Pro:
849 QIAamp PowerFecal Pro DNA Kit, QIA: QIAamp DNA Mini Kit.
850
851 **Figure S2. Community type and HPV 16 assessed by using 4 kits** (A) Community types were
852 classified into two types in all DNA extraction kits, mainly based on the percentage of
853 *Lactobacillus iners*. HPV16 infection was negatively associated with the dominance of *L. iners*
854 (community type I; $p = 0.001$, Fisher's exact test) regardless of DNA extraction method.
855 Although, we observed slight variation in the abundance of microbiota across the extraction kits
856 (even within the same individual patient), the ability to detect two community types was
857 identical across all DNA extraction kits. No significant differences were observed in the
858 relationship of other phenotypes of patients (HPV18, HR-HPV, Biopsy, and Race). The top 15
859 bacteria detected for each DNA extraction kit are shown. Samples were clustered by the
860 Dirichlet component. Narrow columns show each sample and a broader column shows averages
861 of samples. Rows show taxa at the species level. Dark or thin colors correspond to larger or
862 smaller counts of OTUs, respectively. CT: community type. (B) LEfSe analysis using combined

863 data from all four kits detected a significant enrichment of 66 taxa in the cervical environment
864 with HPV16 infection and 17 taxa without HPV16 infection. Genus Lactobacillus were enriched
865 in the HPV16 negative patients ($p < 0.001$, LDA score: 5.38).

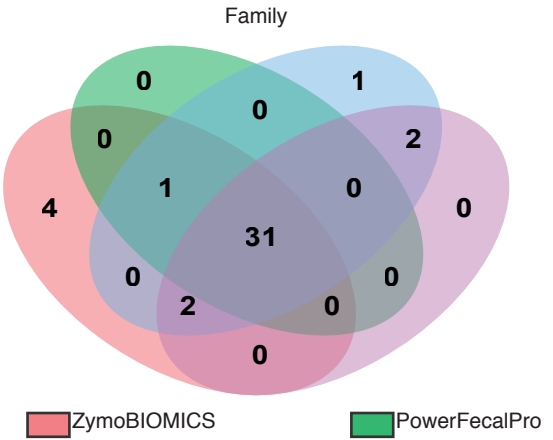

(A) Liquid-based cytology on cervix
Cervical microbiome
LBC
(B) Dispensing of each sample to 4 aliquot
(C) DNA extraction using four protocols
ZymoBIOMICS
PowerFecalPro
QIAampMini
IndiSpin
(D)
Positive control (vaginal mock)
Negative control (preservation solution without samples)
(E) 16S rRNA Illumina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

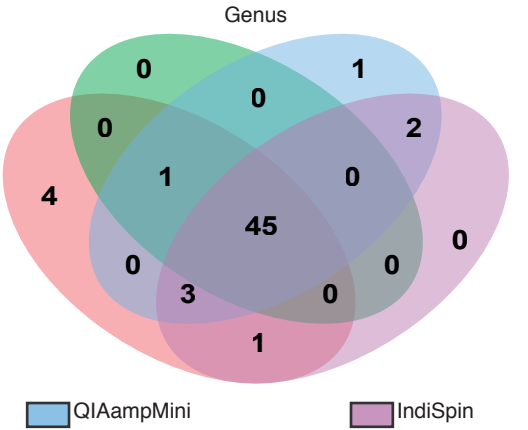
(A)

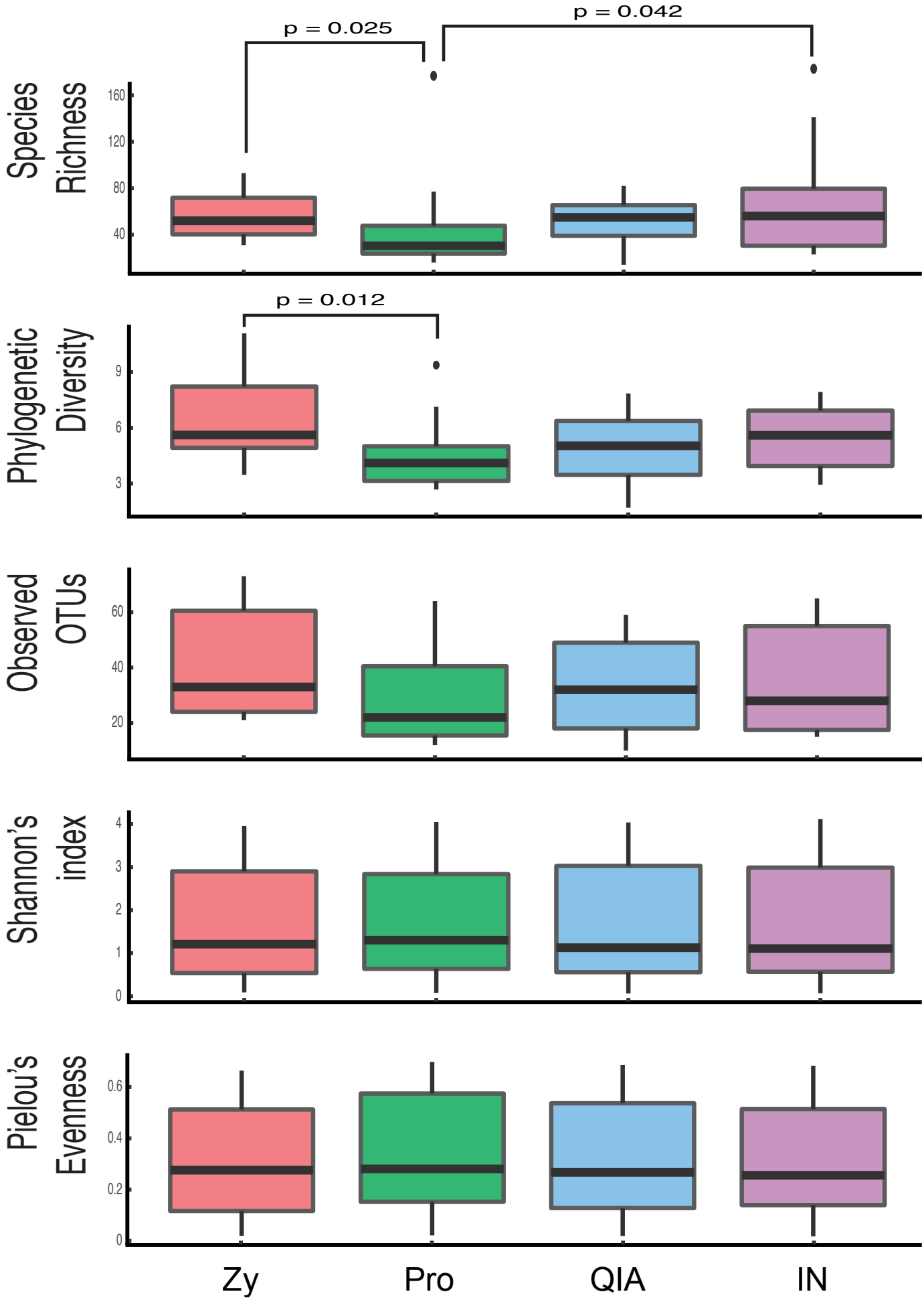

Family
Relative abundance
IndiSpin
PowerFecalPro
QIAampMini
ZymoBIOMICS
Patients
Atopobiaceae
Bifidobacteriaceae
Family XI
Lachnospiraceae
Lactobacillaceae
Leptotrichiaceae
Prevotellaceae
Ruminococcaceae
Streptococcaceae
Veillonellaceae

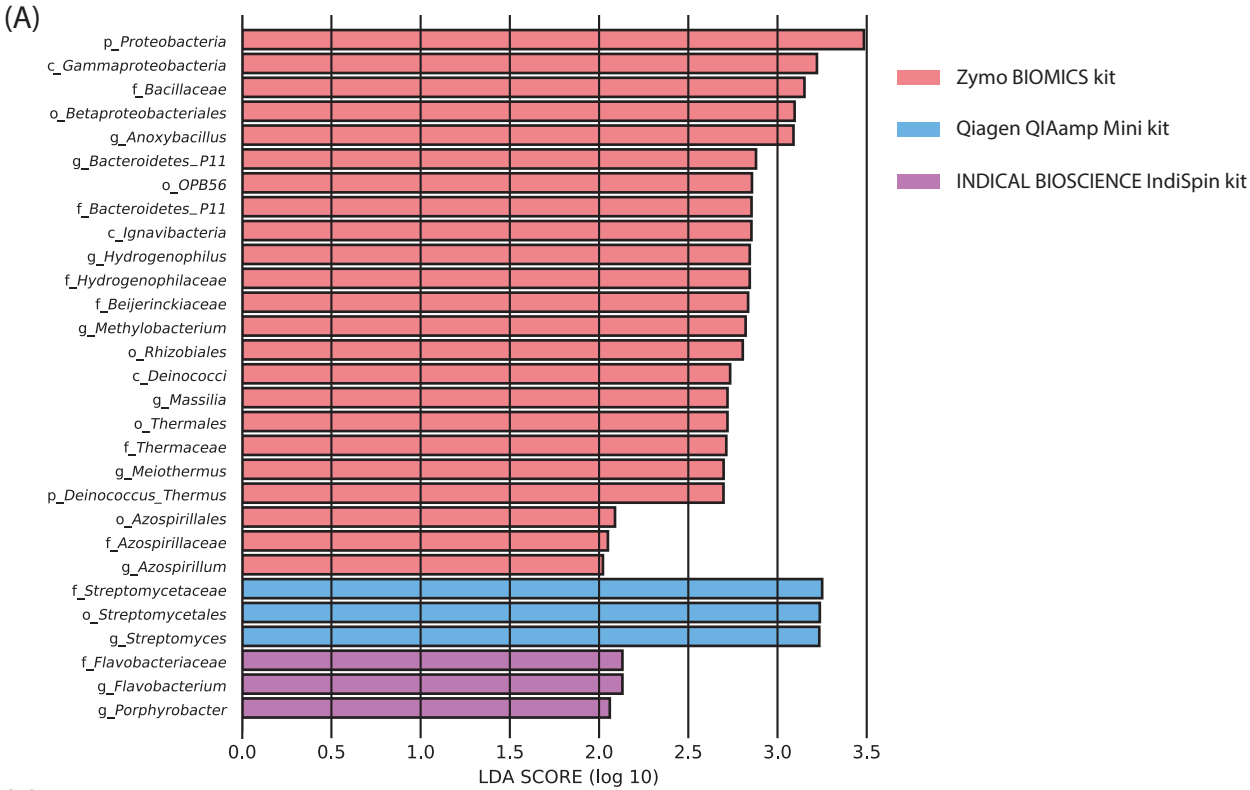

Genus
Relative abundance
IndiSpin
PowerFecalPro
QIAampMini
ZymoBIOMICS
Patients
Anaerococcus
Atopobium
Gardnerella
Lactobacillus
Megasphaera
Parvimonas
Prevotella
Shuttleworthia
Sneathia
Streptococcus

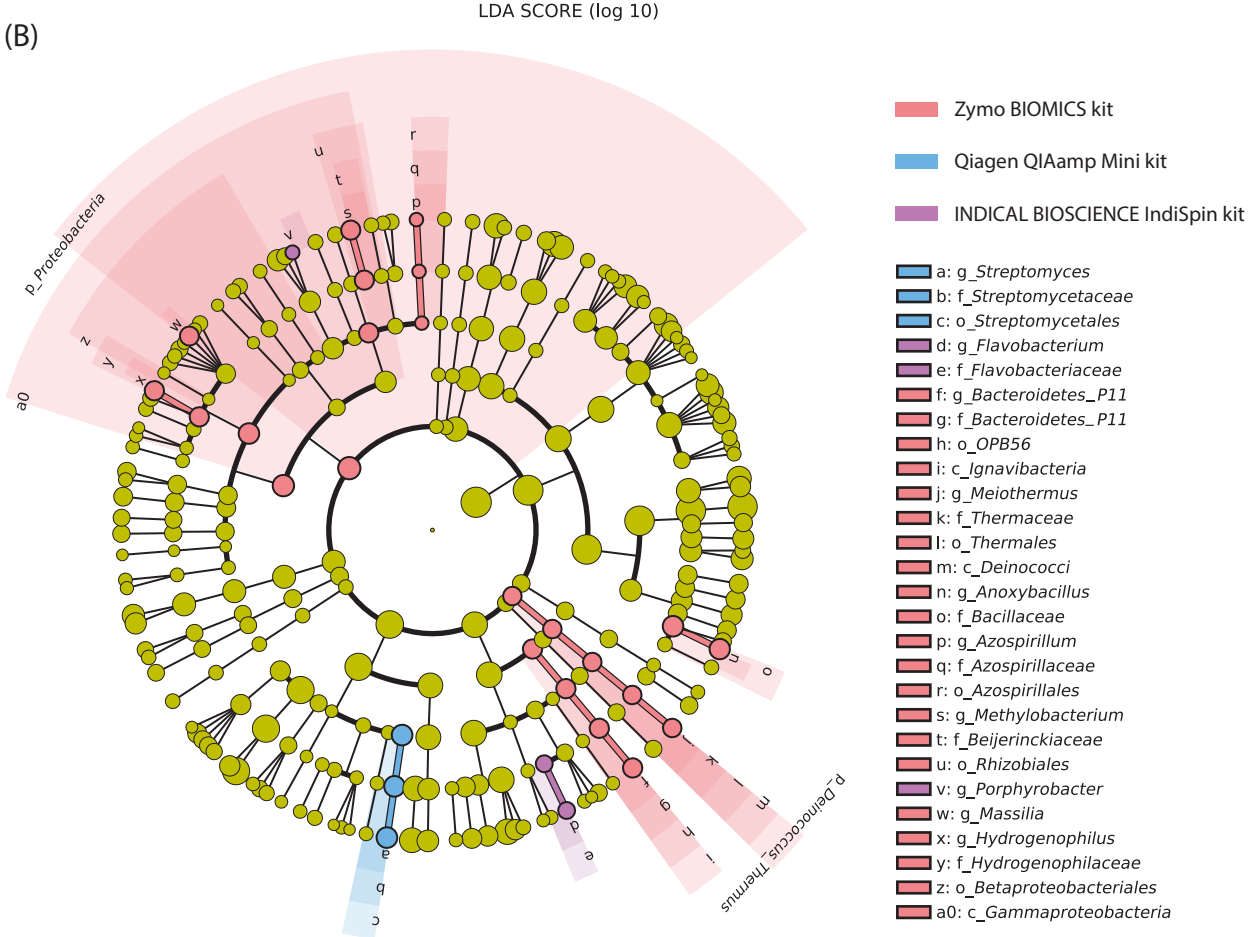
(B)


Family
ZymoBIOMICS
PowerFecalPro


Genus
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