Research Article

Evaluation of the Associations Between Cervical Microbiota and HPV Infection, Clearance, and Persistence in Cytologically Normal Women

Wu Ritu1, Wu Enqi1, Siriguleng Zheng2, Jiandong Wang3, Yaqin Ling1, and Yan Wang3

Abstract

The aim of this study was to investigate the associations between cervical microbiota and different human papillomavirus (HPV) infection statuses in cytologically normal women. The cervical microbiota of HPV-positive or -negative women with a normal cytologic diagnosis was characterized and compared using 16S rDNA-based high-throughput sequencing, and the differences in cervical microbiota associated with new acquisition, persistence, and clearances of HPV genotypes were analyzed via one-year follow-up. The results showed that the cervical microbial richness of HPV-positive women was lower than for HPV-negative women, and the difference was more significant in the postmenopausal group relative to the premenopausal group. Ureaplasma parvum and related taxa were associated with baseline HPV positivity, while Brochothrix, Diplorickettsia, Ezakiella, Faecalibacterium, and Fusobacterium genera and their related taxa and Pseudomonas aeruginosa were associated with baseline HPV negativity. For HPV-positive women, the baseline abundance of Actinomyces was negatively associated with new HPV infection, Alloprevotella tannerae, Prevotella nigrescens, and Prevotella oulorum; and Dialister invisus were positively associated with new HPV-type infection within the year of follow-up. Lactobacillus delbrueckii was found to be negatively associated with persistent HPV infection and 9 taxa belonging to Prevotella, Dialister, and Lachnospiraceae were found to be positively associated with persistence, and/or negatively associated with clearance of HPV types. We also observed 10 novel taxa associated with the clearance/persistence of HPV that had not been reported elsewhere. Those taxa associated with different infection statuses of HPV could be used as a biomarker to help predict the risk of developing persistent HPV infection.

Introduction

Cervical cancer is one of the most common cancers among women worldwide, and persistent infection with high-risk human papillomavirus (HPV) types has been confirmed to be the necessary factor for developing malignancies (1). It is known that HPV infection is very common in sexually active women, and the infecting HPVs can usually be spontaneously eliminated from individuals within 6 to 18 months. Only a small proportion of infected women retain the virus, and this could lead to the development of cervical intraepithelial neoplasia (CIN) and cervical carcinoma (2). The mechanism by which some individuals develop a persistent HPV infection that goes on to develop into clinically significant disease, however, remains largely unclear.

Emerging evidence shows that the human microbiome can mirror the host’s physiology and that it plays an important role in human health (3). Several cross-sectional studies addressing the association between the microbiota in the female reproductive tract and HPV infection and related diseases have been undertaken, and they clearly show that there are significant differences between HPV-negative and -positive women and between healthy women and women with HPV-related diseases with respect to microbial structure, diversity, and species composition.

Key Laboratory of Ethnomedicine (Minzu University of China), Ministry of Education, Beijing, China. 2Department of Information Technology, Polytechnic College, Beijing, China. 3Department of Gynecological Oncology, Beijing Obstetrics and Gynecology Hospital of Capital Medical University, Beijing, China.

Note: Supplementary data for this article are available at Cancer Prevention Research Online (http://cancerpreveas.aacrjournals.org/).

W. Ritu and W. Enqi contributed equally to this article.

Corresponding Authors: Yan Wang, Beijing Obstetrics and Gynecology Hospital of Capital Medical University, Beijing 100006, China. Phone: 186-1065-3059; Fax: 010-6512-1602; E-mail: wwy33196@sina.com; and Yaqin Ling, Key Laboratory of Ethnomedicine (Minzu University of China), Ministry of Education, Beijing 100081, P. R. China. Phone: 188-1022-2674; E-mail: lingyq@lzu.edu.cn

doi: 10.1158/1940-6207.CAPR-18-0233

©2018 American Association for Cancer Research.

www.aacrrjournals.org

Published OnlineFirst November 21, 2018; DOI: 10.1158/1940-6207.CAPR-18-0233
(4–6). Regarding the comparison between HPV-positive and -negative women, the studies in a community of Korean and Chinese women both found HPV-positive women had significantly higher microbial diversity than HPV-negative women (4, 5). Lee and colleagues also observed the HPV-positive women have significantly less Lactobacillus spp. than their counterparts, and identified Sneathia spp. to be a microbiological marker of HPV infection (4). In the comparison between healthy women and women with HPV-related diseases, the studies of Oh and colleagues (6) and Audirac-Chalifour and colleagues (7) conclude that women with CIN had a higher vaginal diversity than healthy controls. The presence of Anaerococcus vaginae, Gardnerella vaginalis, and L. iners in the absence of L. crispatus were identified to be the most high-risk combination for development of CIN. In addition, Brotman and coworkers performed a longitudinal study in a North American cohort of 32 sexually active, premenopausal women over the course of 16 weeks using a twice-weekly interval of evaluation to describe the temporal relationship between vaginal microbiota and HPV detection. Using hierarchical taxonomic clustering, the vaginal microbial profile of each woman was classified into a total of five community state types (CST). The authors found that the CST was associated with changes in HPV status and a low lactobacillus community with high proportions of the genera Atopobium was found to have the slowest rate of HPV clearance (8).

The aforementioned evidence indicates that the cervicovaginal microbiota plays a substantial role in the infection and clearance of HPV virus in the reproductive tract and constitutes a new biomarker reservoir to predict the persistence or regression of HPV. So far, however, only a few longitudinal studies concerning associations between cervicovaginal microbiota and HPV infection/clearance have been conducted, and they were of very limited sample size (9, 10).

To investigate the correlation between cervicovaginal microbiota and HPV infection status, particularly in asymptomatic women, we herein characterized and compared the composition of the cervical microbiota of 144 HPV-positive or -negative women with a normal cytologic diagnosis, using 16S rDNA-based high-throughput sequencing; and the differences in cervical microbiota associated with new acquisition, persistence, and clearances of HPV were analyzed throughout one year of follow-up.

**Materials and Methods**

**Sample collection and processing**

With the approval from the Institutional Review Board of the Department of Gynecologic Oncology of Beijing Obstetrics and Gynecology Hospital of Capital Medical University (Beijing, China), the subjects of this study were enrolled from women participating in health examinations in the Beijing Obstetrics and Gynecology Hospital (Beijing, China), from August to September 2016. Written and verbal consent were obtained from each participant. The exclusive criteria included current pregnancy, sexual intercourse or washout within 3 days, oral or vaginal usage of antibiotics within 2 weeks, previous history of cervical or other lower genital cancer, previous hysterectomy, or destructive therapy of the cervix. To eliminate the impact of the potential change in microbiota that is driven by cervical lesions, only women with normal cytologic results were included in this study. All of the participating women were asked to attend the 1-year follow-up visit for repeated HPV testing.

Risk factor information of lifestyle profile (such as age, cigarette smoking, parity, menstrual status, and use of hormonal/oral contraceptives) was collected with a questionnaire while waiting for the physical examination.

Three cervical swab specimens were collected from each participant by cyto-brush (Qiagen) during routine pelvic examination. The first specimens were used for cytologic diagnosis (SurePath Liquid System, BD Diagnostics TriPath); the second specimens were stored in standard transport medium (STM) for HPV testing; and the third specimens were stored in 4 mol/L guanidine thiocyanate solution for further microbiota analysis. The specimens for microbiota analysis were frozen at −20°C immediately, and then transferred to the laboratory within 24 hours and stored at −80°C before DNA extraction.

DNA was extracted from the residual STM-stored cervical specimens using the QiAmp Mini DNA Kit (Qiagen) in accordance with the manufacturer’s recommendations.

Total bacterial DNA was extracted from samples using the Power Soil DNA Isolation Kit (MO BIO Laboratories) according to the manufacture’s protocol. DNA quality and quantity were assessed using the ratios of 260 nm/280 nm and 260 nm/230 nm. The DNA was then stored at −80°C until further processing.

**HPV assay and genotyping**

Cervical swab specimens stored in STM were tested by Hybrid Capture 2 (HC2;Qiagen), using both the high-risk (HR) panel of 13 pooled types (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68) and the low-risk (LR) panel (types 6, 11, 42, 43, and 44). Samples were processed according to the manufacturer’s instructions and considered positive if they reached the threshold value of at least 1.0 pg/mL of HPV DNA. All of the specimens with positive HC2 (HR and/or LR HPV) results were further tested by Linear Array (Roche) for HPV genotyping.

**Microbiota sequencing**

The V4 region of the bacterial 16S rRNA gene was amplified with the common primer pair (forward primer,
5'-TGCCACGGCTCGGCCTAG-3'; reverse primer, 5'-GGATATACGGTGTTCTAAT-3') combined with adapter sequences and barcode sequences (11). PCR amplification was performed with a modified protocol in a total volume of 50 µL that contained 10 µL of buffer, 0.2 µL of Q5 High-Fidelity DNA Polymerase, 10 µL of High GC Enhancer, 1 µL of dNTP, and 10 µmol/L of each primer, and 60 ng of genomic DNA. Thermal cycling conditions were as follows: an initial denaturation at 95°C for 5 minutes, followed by 10 cycles at 98°C for 1 minute, 50°C for 1 minute, and 72°C for 1 minute, with a final primer template extension at 72°C for 7 minutes. The PCR products from the first-step PCR were purified with VAHTSTM DNA Clean Beads. A second round of PCR was then performed in a 40-µL reaction volume that contained 20 µL of 2 × Phusion HF MM, 8 µL of ddH2O, 10 µmol/L of each primer, and 10 µL of PCR products from the first step. Thermal cycling conditions were as follows: an initial denaturation at 98°C for 30 seconds, followed by 10 cycles at 98°C for 10 seconds, 65°C for 30 seconds, and 72°C for 30 seconds, with a final extension at 72°C for 5 minutes. Finally, all PCR products were quantified with Quant-iT dsDNA HS Reagent and pooled together. We performed high-throughput sequencing analysis of bacterial rRNA genes on the purified, pooled sample using the Illumina Hiseq 2500 platform (2 × 250 paired ends).

Bioinformatics analysis

The raw reads were demultiplexed and then trimmed, merged, and filtered by program Usearch9.0.2132_i86linux32 (available at https://www.drive5.com) following the UPARSE pipeline. We trimmed all reads to the position of the first base with a quality score <= 2, and discarded the sequences shorter than 64 after trimming. Paired reads with a number of expected errors > 1.00 were further filtered out during the filtering step. Sequences were dereplicated and clustered with a threshold of 97% similarity for picking operational taxonomic unit (OTU) representatives after Chimera checking. We then mapped all of the sequences back to the representative sequences resulting in the OTU table for all of the samples. OTU taxonomies (from Phylum to Genus) were determined using the RDP Classifier (available at https://pyro.cme.msu.edu/classifier/form.spr) with a confidence threshold of 80%; the species level taxonomies of the OTUs were further determined using EzBioCloud Identify Service (https://www.ezbiocloud.net/identify). OTU representative sequences were aligned and further filtered to create a phylogenetic tree using QIME1.90 pipeline (available at http://qiime.org/). The OTU table was randomly subsampled down to the size of the smallest sample to obtain equal sequencing depth. Finally, a total of 61,096 reads per sample were used for further analysis.

Alpha diversity, which represents the taxonomic diversity of microbial populations within one sample, was evaluated by Chao1, Observed_otus, PD_whole_tree, Shannon and Simpson indices using QIIME pipeline. Of them, the Observed_otus, Chao1, and PD_whole_tree values reflect the richness aspect of microorganisms in the sample. The Shannon and Simpson are diversity indices, combining the abundance and evenness of OTUs. Beta diversity represents the distance or dissimilarity between microbial composition among different samples. The Jaccard distance matrix, the Bray–Curtis distance matrix, the weighted Unifrac distance matrix, and the Janson–Shannon distance matrix were calculated by QIIME and used for evaluating the beta diversity of the bacterial communities.

Statistical analysis

All statistical analyses were performed using SPSS version 22 (IBM) and the statistical package R (R Foundation for Statistical Computing).

Categorical variables were presented as frequencies and percentages. We used χ² and Fisher exact tests to assess statistical associations between/among variables. Numerical variables were expressed as mean ± SD. We used one-way ANOVA (one-way analysis of variance) to compare the differences among different groups.

Spearman correlation analysis was performed for each of the alpha diversity indices (including Chao1, PD_whole_tree, observed_otus, and the Shannon and Simpson) to explore their correlations with taxa at various levels of taxonomy and other variables.

Principal coordinate analysis (PCoA) and Adonis tests db-RDA were performed on Jaccard distance matrix, the Bray–Curtis distance matrix, the weighted Unifrac distance matrix, the Janson–Shannon distance matrix, respectively, to investigate the differences of the beta diversity between different characteristic variables of the participants. The CST of the cervical microbiota was assigned by hierarchical clustering analysis based on the Jensen–Shannon distance matrix and Ward linkage described previously (12, 13).

The LEfSe difference analysis with LDA cutoff of ±2 and Wilcoxon rank-sum test were performed on abundance data to explore the taxa that were significantly different among the groups. We performed a pairwise Spearman correlation with a 2-tailed probability of t for each correlation.

Bayesian network

To assess the potential associations between/among different relevant variables, a Bayesian network was constructed using bnlearn R package, V.3.6.45 (available on http://www.bnlearn.com/), applying the Hill Climbing
algorithm, which is a greedy search algorithm based upon scoring metrics (14).

Results

Characteristics of participants

A total of 144 women were recruited to this study at baseline, and after 1 year 133 women attended the follow-up visit and were enrolled in the study. Of these women, 90 were HPV positive and 43 were HPV negative at baseline; 55 women who were positive at baseline became HPV negative after 1 year, and 1 woman who was HPV negative at baseline became positive at the follow-up visit.

Based upon the HPV genotyping results of 2 visits, 2 binary categorical variables were created for all participants: the baseline HPV positivity and the acquisition of new HPV genotypes at the follow-up visit. For the baseline HPV-positive participants, 2 more binary categorical variables were created according to whether any type of HPV was cleared in the follow-up visit, and whether a patient was positive for any type of HPV at both visits. The risk-factor characteristics available for the participants for each HPV status are summarized in Table 1. The detailed information on HPV genotyping before and after follow-up was shown in Supplementary Table S1.

The age of the women with no new HPV genotype acquisition within a year were significantly higher than their counterparts, while the proportion of "no parity" was significantly higher. There were no other significant differences among different groups of HPV statuses in terms of age, parity, delivery type, menstrual status, cigarette smoking, or use of hormonal contraceptives.

Overall assessment of intestinal microbiota

A total of 10,103K PE-reads of the 16S rDNA gene V4 region were generated from the 133 specimens, with an average of 75,965.3 \( \pm 1,168.9 \) (SD) reads for each specimen, ranging from 65,379 to 78,172. A total of 10,016 K high-quality PE-reads were obtained after trimming and filtering. In the OTU clustering process, a total of 2,085 OTUs were filtered, with a yield of 1,528 OTUs. After alignment of the OTU representative sequences using QIIME pipeline, a total of 1,494 OTUs were identified and annotated. Among these, 1,342 OTUs at the phylum level had an annotation reliability over 0.8 and covered 26 phyla; 1,283 OTUs at the class level had an annotation reliability over 0.8 and covered 48 classes; 1,220 OTUs at the order level had an annotation reliability over 0.8 and covered 85 orders; 1,113 OTUs at the family level had an annotation reliability over 0.8 and covered 183 families; and 792 OTUs at the genus level had an annotation reliability over 0.8 and covered 386 genera.

Overall assessment of intestinal microbiota

A total of 10,103K PE-reads of the 16S rDNA gene V4 region were generated from the 133 specimens, with an average of 75,965.3 \( \pm 1,168.9 \) (SD) reads for each specimen, ranging from 65,379 to 78,172. A total of 10,016 K high-quality PE-reads were obtained after trimming and filtering. In the OTU clustering process, a total of 2,085 OTUs were filtered, with a yield of 1,528 OTUs. After alignment of the OTU representative sequences using QIIME pipeline, a total of 1,494 OTUs were identified and annotated. Among these, 1,342 OTUs at the phylum level had an annotation reliability over 0.8 and covered 26 phyla; 1,283 OTUs at the class level had an annotation reliability over 0.8 and covered 48 classes; 1,220 OTUs at the order level had an annotation reliability over 0.8 and covered 85 orders; 1,113 OTUs at the family level had an annotation reliability over 0.8 and covered 183 families; and 792 OTUs at the genus level had an annotation reliability over 0.8 and covered 386 genera.

Table 1. Summary of the risk factor characteristics available for the participants in each HPV status

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Baseline HPV positive</th>
<th>New HPV-type acquisition</th>
<th>Any HPV-type cleared</th>
<th>Any HPV-type persisted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age range (mean ± SD)</td>
<td>27.0 (4.6 ± 8.9)</td>
<td>27.0 (4.6 ± 8.9)</td>
<td>27.0 (4.6 ± 8.9)</td>
<td>27.0 (4.6 ± 8.9)</td>
</tr>
<tr>
<td>Parity number</td>
<td>0</td>
<td>10 (7.5%)</td>
<td>11 (8.3%)</td>
<td>12 (9.0%)</td>
</tr>
<tr>
<td>Delivery type</td>
<td>Vaginal delivery</td>
<td>77 (57.9%)</td>
<td>67 (51.2%)</td>
<td>67 (51.2%)</td>
</tr>
<tr>
<td>Currently smoking</td>
<td>11 (8.3%)</td>
<td>7 (5.4%)</td>
<td>7 (5.4%)</td>
<td>7 (5.4%)</td>
</tr>
<tr>
<td>Hormonal contraception</td>
<td>11 (8.3%)</td>
<td>16 (12.5%)</td>
<td>16 (12.5%)</td>
<td>16 (12.5%)</td>
</tr>
<tr>
<td>NOTE: Bold text denotes significant p-value (P &lt; 0.05).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Firmicutes (60.9%), Actinobacteria (14.4%), Proteobacteria (9.9%), and Bacteroidetes (8.5%) were found to be the 4 most dominant taxa at the phylum level; as were Bacilli (57.5%), Actinobacteria (14.5%), Bacteroidia (8.4%), and Gammaproteobacteria (7.3%) at the class level; and Lactobacillales (57.0%), Bifidobacteriales (10.0%), Bacteroidales (8.4%), and Pseudomonadales (4.7%) at the order level. At the family level, Lactobacillaceae (53.8%), Bifidobacteriaceae (10.0%), and Prevotellaceae (8.0%) were the most dominant taxa, covering more than 10% of the overall reads. At the genus level, a total of 30 genera had an abundance of greater than 0.1%; and of these, Lactobacillus (53.8%), Prevotella (7.9%), Atopobium (4.0%), Pseudomonas (3.3%), Streptococcus (1.9%), and Actinobacter (1.1%) were the most dominant taxa covering more than 1% of the overall reads.

Comparison of the diversity of microbial communities among different HPV statuses

To compare the alpha diversity, the Chao1, PD-whole tree, observed-otu, and Shannon and Simpson indices were calculated after we randomly subsampled the OTU table down to 61,096 reads per sample, the size of the smallest sample to obtain equal sequencing depth. The 3 diversity indices regarding the richness (Chao1, PD-whole tree, and observed-otu indices) were all lower for the HPV-positive group compared with the HPV-negative women (Fig. 1a,b and c). Although the variabilities were high among samples, the difference for the Chao1 index was significant in postmenopausal women (Fig. 1a5). For the diversity indices regarding evenness, no differences were found for both Shannon and Simpson indices between HPV-positive and -negative women (Fig. 1d and e). However, after stratification by menstrual status, the evenness diversity of HPV-positive women was found to be higher than for the HPV-negative group in the premenopausal women (Fig. 1d and e); whereas an opposite trend was found in the postmenopausal group (Fig. 1d and 5). Similar tendency was observed for all 5 diversity indices relative to women with new HPV infections in a year and for women without new HPV infections, but the differences were not statistically significant. Among HPV-positive women, no significant differences were observed in the comparison of the cervical microbiota diversity between women with or without HPV persistence, or with or without HPV clearance within a year.

Comparison of beta diversity of microbial communities among different HPV statuses

To examine the bacterial community structure, 5 distance matrices were created on the basis of the 61,096 reads after we randomly subsampled the OTU table, the Jaccard distance matrix, the Bray-Curtis distance matrix, the Janson-Shannon distance matrix, the unweighted Unifrac distance matrix, and the weighted Unifrac distance matrix. The Adonis test and db-RDA analysis were performed on all 5 distance matrices to investigate the association of the beta diversity and the variables of risk factor and HPV status. A significant association was found between menstrual status and the bacterial community structure based on Bray-Curtis distance matrix, Janson-Shannon distance matrix, and weighted Unifrac distance matrix in Adonis analysis and db-RDA analysis; and in the analysis based on Unweighted Unifrac distance matrix, the factors of age and currently smoking were found to be significantly associated with the bacterial community structure. No significant association was found between the bacterial community structure and other factors (Supplementary Table S2). The variable of menstrual status was associated with all 3 distance matrices based on abundance data, indicating the importance of abundance characteristics of the bacterial community on the difference between premenopausal and postmenopausal women. Differences among taxa with higher abundance might play a significant role in the differences observed in the community structure among different menstrual statuses.

The cervical communities were classified into 5 CSTs in hierarchical clustering analyses based on the Jensen-Shannon distance matrix and Ward linkage. As shown in Fig. 2, the CST1 is dominated by Lactobacillus iners (81.0%); CST2 by Lactobacillus crispatus (84.2%); CST3 by Gardneri. amnionii (17.2%), Gardnerella ADEV_s (11.0%), and a variety of Pseudomonas spp that sum up to 36.3% of relative abundance; CST4 by Gardnerella ADEV_s (41.8%), Atopobium vaginae (22.0%), and Lactobacillus iners (10.6%); and CST5 by Salmonella enterica (10.2%), a variety of Lactobacillus spp, Pseudomonas spp, Prevotella spp, and Streptococcus spp, summing up to 19.5%, 9.5%, 8.7%, and 7.8% of relative abundance, respectively. Among these, the CST1 and CST2 were similar to the CST III (dominated by L. crispatus) and CST I (dominated by L. iners), respectively, previously described by Ravel and colleagues (13); while CST IV in a previous study was further clustered into 3 groups (CST 3, CST 4, and CST 5) in this study. The proportion of each CST in different groups of HPV status is shown in Supplementary Table S3. No significant differences were observed in the evaluation of the association between cervical CST and different HPV statuses or menstrual statuses.

Identification of cervical microbiota composition markers with respect to different HPV statuses

Linear discriminant analysis (LDA) effect size (LEfSe) analysis was performed to identify differences in microbiota composition that may be related to different statuses of HPV infection (Fig. 3). We identified 17 bacterial taxa that have different abundancies between the baseline HPV-positive and -negative women. Among them, genus Ureaplasma and the family, order, class, and the phylum to which it belongs, were more abundant in the HPV-positive
group; whereas Deinococcus and Thermus at the phylum level; Deinococci and Acidobacteria Gp4 at the class level; Listeriaceae, Ectothiorhodospiraceae, Coxiellaceae, and Fusobacteriaceae at the family level; and Faecalibacterium, Ezakiella, Brochothrix, Diplorickettsia, and Fusobacterium at the genus level were more abundant in the baseline
HPV-negative group. In the comparison of women with or without new acquisition of HPV within a year, only one taxon, the genus *Actinomyces* was found to manifest a different abundance between the 2 groups. In a comparison of women with or without any HPV type cleared in a year, Rhodobacterales at the order level and Rhodobacteraceae at the family level were associated with the clearing group. No taxon was found have a different abundance between women with a HPV type persisting for a year and women with no HPV type persisting for a year.

Figure 2.
2D PCoA plots of the cervical microbiota performed on the Janson–Shannon distance metric correlated with CSTs.
At the OTU level, 5 OTUs were identified to have different abundances between the baseline HPV-positive and -negative women. Of these, 2 OTUs, annotated to *Ureaplasma parvum* and *Fusobacterium nucleatum*, respectively, had a relatively higher abundance in the baseline HPV-positive group; whereas, the other 3 OTUs, annotated to *Faecalibacterium GL538271_s*, *Ezakiella massiliensis*, and *Pseudomonas aeruginosa*, respectively, had a relatively higher abundance in the baseline HPV-negative group. In the comparison of women with or without new-acquisition HPV within a year, 4 OTUs, annotated to *Alloprevotella tanenae*, *Prevotella nigrescens*, *Prevotella oulorum*, and *Dialister invisus*, respectively, possessed a relatively higher abundance in women with new-acquisition HPV within a year.

In a comparison of women with or without any HPV type cleared in a year, 11 OTUs were associated with the non-cleared group. The OTUs were annotated to EU728721_g, AY959069_s (belonging to the family Lachnospiraceae), *Gemella asaccharolytica*, *Prevotella amnii*, *Prevotella GQ037615_s*, *Sacccharmonas spp*, *Prevotella KQ959344_s*, *Prevotella DJQ815942_s*, *Sulirubrobacter EU1289476_s*, *Labilithrix AB286604_s*, and *Lachnolea glyciner*, respectively; and the remaining 1 (sharing 90.12% of similarity with a strain of HG-B01269 in the EzBioCloud 16S database), failed to be annotated to a known taxon. Nine OTUs were associated with the HPV-persisting group in the comparison of women with or without any HPV type persisting for a year. The OTUs were annotated to *Prevotella timonensis*, *Lactobacillus delbrueckii*, *Eubacterium_g23 HQ746544_s*, *Microstrep tus lupini*, *Deinococcus suerensis*, HM748650_g, FF936969_s (belonging to the family Haliangiaceae), *Dialister invisus*, *Anaerobacterium chartislovens*, and *Coprococcus comes*, respectively.

### Evaluation of the association of cervical microbiota composition and alpha diversity indices

To identify the cervical microbiota composition that correlated with each of the alpha diversity indices, we performed a pairwise Spearman correlation on all the identified genus and 5 alpha diversity indices, respectively. In the results, 79 taxa at the genus level were associated with Chao1 index, the rho ranging from 0.17 to 0.54. For the PD_whole_tree index, the correlations of 106 genera were shown to be statistically significant, the rho ranging between 0.27 and 0.62. A total of 110 genera were statistically correlated with the observed_otu index with the rho ranging from 0.20 to 0.56. For the Shannon index, the correlations of 52 genera were shown to be statistically significant, the rho ranging from −0.27 to 0.62. A total of 110 genera were statistically correlated with the observed_otu index with the rho ranging from −0.20 to 0.56. For the Shannon index, the correlations of 52 genera were shown to be statistically significant, the rho ranging from −0.27 to 0.62.
Bayesian network
A Bayesian network was constructed on the variables of 5 alpha diversity indices, HPV status, and menstrual status to assess potential associations (Fig. 4).

Figure 4A was built on overall datasets, and it was shown that the observed_otu index was the only alpha diversity index directly associated with baseline HPV positivity; and the new acquisition of HPV type in a year was also directly associated with baseline HPV positivity. No variable tested was associated with menstrual status.

Figure 4B was constructed from data on HPV-positive women to assess the potential associations of different variables with HPV clearance and persistence. Our results showed that none of the 5 alpha diversity indices was associated with baseline HPV positivity. Of these, the genus Ezakiella was the only microbiota composition that was shown to be statistically significant, the rho ranging from 0.71 to 0.67. The top 10 genera correlated with each of the alpha indices are shown in Table 2. Campylobacter was the genus most correlated with the Chao1 and PD_whole_tree indices, and it correlated third most with the observed_otu index. Genus Acinetobacter was the genus most correlated with the observed_otu index; and Lactobacillus and Prevotella were the 2 genera exhibiting the most negative and positive correlations, respectively, for both Shannon and Simpson indices.

Discussion
In this study, we investigated the association between cervical microbiota and variables that included HPV positivity and other risk factors (such as age, cigarette smoking, parity, menstrual status, and use of hormonal/oral contraceptives) in women with a negative cytologic screening diagnosis. One-year follow-up was conducted on the participants to further investigate the association of baseline cervical microbiota and the acquisition of new HPV-type infection, persistent infection of any HPV type during a year, and the clearance of any HPV type during a year.

The results indicated that the cervical microbial richness of HPV-positive women was lower than HPV-negative women to assess the potential associations of different variables with HPV clearance and persistence. Our results showed that none of the 5 alpha diversity indices was associated with baseline HPV positivity. Of these, the genus Ezakiella was the only microbiota composition that was shown to be statistically significant, the rho ranging from 0.71 to 0.67. The top 10 genera correlated with each of the alpha indices are shown in Table 2. Campylobacter was the genus most correlated with the Chao1 and PD_whole_tree indices, and it correlated third most with the observed_otu index. Genus Acinetobacter was the genus most correlated with the observed_otu index; and Lactobacillus and Prevotella were the 2 genera exhibiting the most negative and positive correlations, respectively, for both Shannon and Simpson indices.

Bayesian network
A Bayesian network was constructed on the variables of 5 alpha diversity indices, HPV status, and menstrual status to assess potential associations (Fig. 4).

Figure 4A was built on overall datasets, and it was shown that the observed_otu index was the only alpha diversity index directly associated with baseline HPV positivity; and the new acquisition of HPV type in a year was also directly associated with baseline HPV positivity. No variable tested was associated with menstrual status.

Figure 4B was constructed from data on HPV-positive women to assess the potential associations of different variables with HPV clearance and persistence. Our results showed that none of the 5 alpha diversity indices was associated with baseline HPV positivity. Of these, the genus Ezakiella was the only microbiota composition that was shown to be statistically significant, the rho ranging from 0.71 to 0.67. The top 10 genera correlated with each of the alpha indices are shown in Table 2. Campylobacter was the genus most correlated with the Chao1 and PD_whole_tree indices, and it correlated third most with the observed_otu index. Genus Acinetobacter was the genus most correlated with the observed_otu index; and Lactobacillus and Prevotella were the 2 genera exhibiting the most negative and positive correlations, respectively, for both Shannon and Simpson indices.

Discussion
In this study, we investigated the association between cervical microbiota and variables that included HPV positivity and other risk factors (such as age, cigarette smoking, parity, menstrual status, and use of hormonal/oral contraceptives) in women with a negative cytologic screening diagnosis. One-year follow-up was conducted on the participants to further investigate the association of baseline cervical microbiota and the acquisition of new HPV-type infection, persistent infection of any HPV type during a year, and the clearance of any HPV type during a year.

The results indicated that the cervical microbial richness of HPV-positive women was lower than HPV-negative
women, and the difference was more significant in the postmenopausal group than in the premenopausal group. Although no significant associations were found between HPV status (namely, baseline HPV positivity, the acquisition of new HPV type infection, persistent infection of any HPV type during a year, and the clearance of any HPV type during a year) and the beta diversity of the cervical microbiota, several taxa whose abundance differed significantly in subgroups with different HPV statuses were identified.

Figure 4.
Bayesian network showing the dependencies between/among variables selected on the basis of their association.
In the comparison of the alpha diversity, it is interesting to note that the richness of the cervical microbiota of HPV-positive women was lower than for HPV-negative women in our study. This result is controversial with respect to previous studies that showed that for the microbiota in the female reproductive tract, low microbial diversity was more closely associated with a healthier condition, and HPV-positive women have a higher species diversity (4, 10). A possible explanation for this controversy is that compared with the previous study, only cytologically normal women were included in the HPV-positive group in this study, so as to eliminate the potential impact of the microbiologic changes derived from HPV-related cervical lesions. It is known that the microbiota’s diversity increases with the severity of the cervical lesion and is significantly higher than in healthy women (7, 15). Therefore, comparisons between the HPV-positive and -negative groups, irrespective of histologic or cytologic status, might lead to an overestimation of the diversity of the HPV-positive group. Audirac–Chalifour and colleagues also showed no significant differences ($P = 1$, Student $t$ test) in the comparison of pd and Shannon index between HPV-positive and -negative groups when only NCL-negative cases were considered (7). Collectively, our results indicated that the alpha diversity indices of microbiota in the female reproductive tract were not a stable indicator of HPV positivity, especially when the infection was asymptomatic.

\textit{Lactobacillus} species are the most common inhabitants of the female reproductive tract, and reduced relative abundance of \textit{Lactobacillus} is often considered to lead to increased diversity of reproductive tract microbiota (16). In this study, by calculating the 2-by-2 Spearman rank correlation coefficient (Spearman rho) for all of the taxa and 5 alpha diversity indices uncovered, we found that \textit{Lactobacillus}—the most abundant genus in our overall microbiota data—is only significantly associated with evenness but not associated with richness diversity. It was the most highly correlated (negatively) genus in both Shannon and Simpson indices, whereas its correlations with pd\textunderscore whole\textunderscore tree, Chao1, and observed\textunderscore otu indices were all nonsignificant. This result supports the theory that the predominance of \textit{Lactobacillus} spp. prevents the colonization of other bacterial species and could lead to a decrease in evenness of the microbiota. Conversely, our results also suggest that the dominance of \textit{Lactobacillus} does not diminish the other bacterial species thoroughly, but that the abundance of other bacterial species might remain at a very low level and be detected with the increase in sequencing depth. The genera \textit{Prevotella}, \textit{Dialister}, \textit{Peptostreptococcus}, and \textit{Gardnerella} were found to be positively correlated with evenness diversity, indicating that as \textit{Lactobacillus} diminishes, the abundance of these other genera will significantly increase together with the Shannon and Simpson indices. For richness diversity, several other genera, such as \textit{Actinobacter}, \textit{Burkholderia}, \textit{Campylobacter}, \textit{Pseudomonas}, \textit{Corynebacterium}, \textit{Halorubrum}, and \textit{Halorotalia}—rather than \textit{Lactobacillus}—were found to be correlated significantly.

Regarding beta diversity, the db-RDA and adonis analysis found no significant association between different HPV statuses and distant matrices in this study, which is similar to the results of Di Paola and colleagues (10) and Di Pietro and colleagues (17). The unsupervised clustering based on distance matrix was another analyzing method commonly used in reproductive tract microbiota data to classify women into different CSTs. Several investigators using hierarchical clustering analyses based on the Jensen–Shannon distance matrix and Ward linkage reported classifying the participants into one of 5 distinct CSTs: CST-I (most often dominated by \textit{Lactobacillus crispatus}), CST-II (most often dominated by \textit{Lactobacillus gasseri}), CST-III (most often dominated by \textit{Lactobacillus iners}), CST-IV (characterized by a paucity of \textit{Lactobacillus} spp. and a wide array of facultative and strict anaerobes), and CST-V (most often dominated by \textit{Lactobacillus jenseni}; refs. 15, 18, 19). Among these, it was reported that CST-IV or a certain subgroup of CST-IV was more highly associated with a high Nugent score, increasing the severity of cervical lesions and positivity for HPV.

In this study, using the same clustering method, we also classified the participants into 5 groups with CST1 dominated by \textit{Lactobacillus iners} similar to CST-III in the aforementioned studies and CST 2 dominated by \textit{Lactobacillus crispatus} similar to CST-I. CSTs 3, 4, and 5 were all similar to CST-IV in the previous studies in not being dominated by any \textit{Lactobacillus} spp., and having their distinct indicator species, respectively. CST-II (dominated by \textit{Lactobacillus gasseri}) and CST-V (dominated by \textit{Lactobacillus jenseni}) described in the previous studies were not found in our study. A possible cause of the inconsistent clustering pattern in this study compared with previous studies is that with a very high sequencing depth, much more detailed information of the microbiota, which might be ignored with a lower sequencing depth, was demonstrated in this study. The minimal number of reads for each sample in this study was high (up to 61,096), and a total of 1,528 OTUs that were assigned to 386 genera were observed from the sequences. In the comparison of the proportion of each CST in different HPV statuses, we failed to observe any statistical association between CST and HPV status, indicating that the association between CST and HPV infection status might also be population-specific, but this requires further confirmation in a wider range of studies with larger sample sizes.

In this study, the taxonomic differences in cervical microbiota in groups with different HPV statuses were investigated with LEfSe analysis. We found several taxa to possess significantly different abundances with respect to the comparison of the baseline positive and negative groups. Among them, the \textit{Ureaplasma parvum} and their
related taxa manifested higher richness in the baseline HPV-positive group compared with the baseline HPV-negative group. This is consistent with previous studies in different populations of the world indicating that *Ureaplasma parvum* was associated with a significantly increased risk of overall HPV infection and increased risk of abnormal cervical cytology (20–22). It was reported that the infection of *Ureaplasma* could lead to scarring and damage of the epithelium (23), suggesting that infection of the reproductive tract with *Ureaplasma parvum* might also increase the risk of HPV infection by facilitating viral infections.

In contrast, the *Brochothrix*, *Diplorickettsia*, *Ezakiella*, *Faecalibacterium*, and *Fusobacterium* genera and their related taxa and *Pseudomonas aeruginosa* showed significantly higher abundance in the HPV-negative group compared with the HPV-positive group. It is interesting to note that although the abundance of *Fusobacterium* on a genus level was associated with HPV negativity, one of the 5 OTUs belonging to Fusobacterium—assigned as *Fusobacterium* JQ465429.5—was more abundant in the HPV-positive group. It was also reported in a previous study that *Fusobacterium nucleatum* (another species of *Fusobacterium*) was more abundant in the HPV-negative group (9). The controversial association of HPV positivity and different *Fusobacterium*-related taxa indicates that different species of *Fusobacterium* might play different roles in the process of HPV infection. Species of the *Faecalibacterium* genus were previously reported to demonstrate a higher abundance in women who cleared HPV infection after one year compared with women who developed persistent HPV infection (10). However, little is known about interactions between *Brochothrix*, *Diplorickettsia*, *Ezakiella*, and *Pseudomonas aeruginosa* and HPV.

The comparison between the baseline HPV-positive and HPV-negative groups was cross-sectional; therefore, we could not provide information about cause-and-effect relationships. Thus, a one-year follow-up was conducted in this study and the causal links between cervical microbiota and new infection, persistence, and clearance of HPV were investigated; and we found 5 taxa that might be linked to new HPV infection. Of these, the baseline abundance of *Actinomyces* was significantly associated with new HPV infection; and 3 taxa, *Alloprevotella tannerae*, *Prevotella nigrescens*, and *Prevotella oulorum* (which belong to *Prevotellaceae*) and *Dialister inwards* were positively associated with new HPV infection within a year’s time. The association of these taxa with HPV that we found in our study was similar to those of previous studies. *Actinomyces* was negatively associated with HPV positivity in the study by Di Paola and colleagues (10); whereas, the species of *Prevotellaceae* were reported to show greater abundance in the HPV-positive group relative to the HPV-negative group (24). *Dialister* was reported to be associated with HPV positivity in studies in Korea and Nigeria (4, 24).

In a comparison between HPV clearance and persistence, we found 22 taxa to be associated with the clearance or persistence of HPV over a year, and of these, 5 taxa belong to *Prevotella* and 1 taxon belongs to *Dialister*. These taxa were all negatively associated with the clearance or positively associated with persistence of HPV types. These outcomes are in accordance with the findings of Dareng and colleagues (24) and Lee and colleagues (4), indicating that for women with a higher abundance of *Prevotella* and *Dialister*, it is more difficult to clear HPV and they more easily develop HPV persistence, thereby exhibiting a higher probability of detection as being persistently HPV positive. Three of the remaining taxa found in our study belong to *Lachnospiraceae*, one of the bacteria characteristic of microbiota associated with bacterial vaginosis (BV; refs. 25, 26). Several studies have shown that BV was correlated with higher incidence, prevalence, and persistence of HPV infection and with development of CIN (27–29). *Lactobacillus delbrueckii* was found to be associated with the absence of HPV persistence in the HPV-positive group, which is consistent with a recent study by Zhang and colleagues (30), and supports the theory that certain species of *Lactobacillus* promote HPV clearance (18). However, in the 12 species of *Lactobacillus* found in this study, only 1 species was found to be significantly associated with HPV infection status, suggesting great heterogenicity of different subgroups of *Lactobacillus* in their association with HPV. It is known that the species of the genus *Lactobacillus* share less than 70% of nucleotide identity across their genomes (15), so further investigations using metagenomic techniques at the gene level would help us to understand how *Lactobacillus* plays protective or pathogenic roles with regard to HPV. The remaining 10 taxa associated with HPV clearance or persistence were novel and have not been reported elsewhere, but their association with HPV status need to be confirmed in further studies.

There are some limitations to our study. First, we were not able to obtain information regarding BV status of the participants; therefore, the relationship between cervical microbes, BV, and HPV could not be evaluated. Second, we used 16 sRNA gene sequencing to analyze the cervical microbiome, and this method could introduce bias during several processes, including the selection of the gene amplification area, the gene amplification procedure, the selection of the gene sequence database, and OTU clustering. The strengths of our work include its relatively large sample size, and the higher sequencing depth.

In conclusion, in this study, we investigated the associations of cervical microbiota with different HPV infection statuses in cytologically normal women. We observed several taxa that could differentiate baseline HPV infection status and predict the acquisition of new HPV infection and...
Persistence and/or clearance of HPV within a year. In the analysis of associations among alpha diversity indices, HPV status, and microbiota component, we found that the richness in HPV-positive women was higher than in HPV-negative women. The richness diversity was primarily correlated with the abundance of genera other than Lactobacillus, such as Acinetobacter, Burkholderia, Campylobacter, Pseudomonas, Corynebacterium, Halorubrum, and Haloribetals; although, Lactobacillus was shown to be the taxon most correlated with evenness diversity. No significant difference was observed among various HPV infection statuses with respect to evenness diversity. With our results—together with those of previous studies—we can conclude that certain cervical microbiota compositions associated with different HPV infection statuses may constitute a biomarker to help discriminate women who are at risk for developing persistent HPV infection. Furthermore, investigations of the mechanisms underlying these associations might provide useful insights into the development of new therapeutic strategies to boost the clearance of HPV infection by modifying reproductive tract microbiota.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

References


Evaluation of the Associations Between Cervical Microbiota and HPV Infection, Clearance, and Persistence in Cytologically Normal Women

Wu Ritu, Wu Enqi, Siriguleng Zheng, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1940-6207.CAPR-18-0233

Supplementary Material
Access the most recent supplemental material at:
http://cancerpreventionresearch.aacrjournals.org/content/suppl/2018/11/21/1940-6207.CAPR-18-0233.DC1

Cited articles
This article cites 30 articles, 3 of which you can access for free at:
http://cancerpreventionresearch.aacrjournals.org/content/12/1/43.full#ref-list-1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://cancerpreventionresearch.aacrjournals.org/content/12/1/43.
Click on “Request Permissions” which will take you to the Copyright Clearance Center's (CCC) Rightslink site.