

SCIENTIFIC REPORTS

OPEN

Association of cervical microbial community with persistence, clearance and negativity of Human Papillomavirus in Korean women: a longitudinal study

Selvaraj Arokiyaraj^{1,2}, Sang Soo Seo¹, Minji Kwon¹, Jae Kwan Lee³ & Mi Kyung Kim¹ 

The present study aimed to identify the cervical microbes that are associated with HPV negativity, HPV clearance and HPV persistence and to assess the microbes' longitudinal associations as related to HPV infection dynamics among Korean women. We enrolled 41 women with 107 samples, and classified them according to the HPV infection dynamics: HPV negativity (21 samples, 10 subjects), HPV clearance (42 samples, 15 subjects), and HPV persistence (44 samples, 16 subjects). Cervical swabs were collected at the baseline and six-month-interval follow-up visits. HPV positivity was determined by HPV DNA HC2 assay, and the microbiome was analyzed using 16SrRNA pyrosequencing, linear discriminant analysis effect size and multivariate logistic analysis. In the multivariate logistic analysis results, *Lactobacillus crispatus* (multivariate OR (mOR) = 8.25, 95% CI 2.13–32.0) was predominant in the HPV-negative group. We observed that *Eubacterium eligens* (mOR = 11.5, 95% CI 1.31–101.4), *Gardnerella vaginalis* (mOR = 17.0, 95% CI 2.18–131.8), and *Ureaplasma urealyticum* (mOR = 7.42, 95% CI 1.3–42.46) had the strongest associations with HPV clearance, and *Lactobacillus johnsonii* (mOR = 16.4, 95% CI 1.77–152.2) with HPV persistence. Overall, greater diversity was observed in HPV-persistence than in HPV-negative women. Our findings suggest that the presence and prevalence of a specific cervical microbiome are factors involved in HPV dynamics.

Cervical cancer is one of the leading causes of cancer death among women, and the most important causal agent in the development of cervical intraepithelial neoplasia (CIN) and cervical cancer is persistent infection with high-risk human papilloma virus (hrHPV)^{1,2,4,5}. The typical pattern of progression in HPV-mediated cervical cancer is as follows: (i) acquisition, (ii) persistence, (iii) progression to pre-cancer (CIN 1, 2 and 3), and invasive cancer. There is evidence that most cases of HPV infection are transient, which is to say, likely to regress naturally³. Several cofactors such as smoking, high parity, long-term use of oral contraceptives⁶, hormone treatment and co-infection with sexually transmitted infection agents are relevant to progression of cervical cancer among HPV-infected women⁷. Associations between cervical microbes and HPV infection and CIN have been investigated^{8,9}. Also, there have been reports on, for example, the frequency and determinants of acquisition and persistence of HPV infection among Danish soldiers¹⁰ and the viral and non-viral determinants of cervical HPV acquisition and clearance in Hawaiian women¹¹. Recently too, an American study reported temporal changes of cervical microbes and HPV detection among Baltimore women¹². Notwithstanding the many reports on cervical microbes and their HPV associations^{13,14}, more knowledge on the epidemiology and longitudinal dynamics of cervical microbes is required before the nature of HPV progression is clearly understood. The present study aimed to identify the cervical microbes that are associated with HPV negativity, HPV clearance and HPV persistence and to assess the microbes' longitudinal associations as related with HPV infection dynamics among Korean women.

¹Division of Cancer Epidemiology and Management, Center for Uterine Cancer, National Cancer Center, Ilsandong-gu, Goyang, Republic of Korea. ²Department of Food Science and Biotechnology, Sejong University, Gwangjin-gu, Seoul, Republic of Korea. ³Department of Obstetrics and Gynecology, Korea University College of Medicine, Seoul, Republic of Korea. Selvaraj Arokiyaraj and Sang Soo Seo contributed equally. Correspondence and requests for materials should be addressed to M.K.K. (email: alrud@ncc.re.kr)

Demographic characteristics	Negative (N = 10) 21 sample	Clearance (N = 15) 42 sample	Persistence (N = 16) 44 sample	p value
Age (years)	44.6 ± 9.0	44.4 ± 13.3	44.6 ± 13.0	0.982
HPV, viral load); Median (range)	—	34.6 (1.13–97.7)	37.4 (2.67–1466.8)	0.603
Body mass index (Kg/m ²)	22.1 ± 1.9	22.9 ± 3.4	21.8 ± 3.5	0.592
Parity(parous)	2.38 ± 0.5	2.27 ± 0.6	2.42 ± 1.4	0.311
Postmenopausal status	2 (20)	7 (50)	6 (37.5)	0.326
Oral contraceptive use (ever)	3 (30)	5 (33)	1(6)	0.157
HRT				
Never	9 (90)	13(87)	13(81)	1
Ever	1 (10)	2(13)	3(19)	
Smoking status				
Never	8 (80)	12(80)	15(94)	0.537
Ever	2 (20)	3(20)	1(6)	
Passive smoking				
No	6 (60)	8(57)	11(69)	0.792
yes	4 (40)	6(43)	5(31)	
Alcohol drinking status				
Yes	8 (80)	11(79)	12(75)	1.0
No	2 (20)	3(21)	4(25)	
Alcohol drinking (>1 week)	1 (10)	5 (36)	5 (31)	0.480

Table 1. Demographic characteristics of study participants at baseline among HPV groups (Negative, Clearance, and Persistence). Values are Mean ± SD or n (%). Only available variables were used in this study, due to missing response for several questions. Chi square test was used for testing categorical variables and Kruskal Wallis test for comparisons of continuous variables. When the number of expected frequencies was less than 5 and the number of cells is more than 25%, Fisher's exact test was performed. HRT: Hormone replacement therapy; HPV: human papillomavirus.

Results

General characteristics. We enrolled 41 women and classified them into three HPV groups according to the HPV infection dynamics: HPV negativity (21 samples, 10 subjects), HPV clearance (42 samples, 15 subjects), and HPV persistence (44 samples, 16 subjects). The epidemiological and clinical information on the subjects is presented in Table 1. There were no significant differences in the age at enrollment among the HPV groups. Also, the HPV viral load at enrollment did not differ between HPV clearance and HPV persistence.

Cervical microbiota among HPV group. We sequenced 107 cervical samples (41 subjects) from a total of 617, 044 high-quality reads with an average of 18.97 operational taxonomic units (OTUs) per sample after quality filtering. The sequence reads were assigned to 14.5 OTUs in HPV negativity, 20.2 OTUs in HPV clearance, and 24.5 OTUs in HPV persistence. Eleven phyla — Firmicutes, Actinobacteria, Bacteroidetes, Fusobacteria, Proteobacteria, Tenericutes, Cyanobacteria, Verrucomicrobia, Planctomycetes, Acidobacteria, and Candidatus Saccharibacteria — were found, six of which (Firmicutes, Actinobacteria, Bacteroidetes, Fusobacteria, Proteobacteria, and Tenericutes) were dominant (>1%) among the groups. *Lactobacillus crispatus* and *Lactobacillus iners* were the dominant species among the three HPV groups. A shift in the cervical microbiota was observed during the follow-up period (Supplementary Table S1). The HPV-negative women showed the highest abundance of *Lactobacillus crispatus* at the baseline (40.45%) and six-month-interval follow-up visits (35.6%). Compared with HPV-negative, HPV positive-women (HPV clearance + persistence group at baseline) had the highest abundance of *Atopobium vaginae* at the baseline and follow-up visits. Compared with HPV persistence, HPV clearance women showed increased percentage of *Eubacterium eligens*, *Ureaplasma urealyticum* at the baseline and follow-up visits. HPV persistence women showed high proportion of *Lactobacillus johnsonii* at the baseline and follow-up visits, compared with other groups (HPV negativity + clearance group).

Identification of cervical microbiome markers of HPV dynamics. The taxonomic groups that were relatively abundant in the HPV persistence, negativity and clearance groups were identified by linear discriminant analysis effect size (LEfSe with $\alpha = 0.05$), an LDA score of at least 2, and a relative abundance greater than 0.1, both at the baseline and for total visits. The changes at the baseline are depicted in a cladogram (Fig. 1A). In the HPV-persistence group, significant over-representation of *Haemophilus* (Order: Pasteurellales; Family: Pasteurellaceae) ($P = 0.0345$) (Fig. 1A,B,E) was observed. By contrast, the HPV-clearance group was found to have significantly higher levels of *Gardnerella vaginalis* (Order: Bifidobacteriales; Family: Bifidobacteriaceae) ($P = 0.0403$) (Fig. 1A,B,D), and the HPV-negative group was determined to have significant overrepresentation of *Lactobacillus crispatus* ($P = 0.0067$) (Fig. 1A–C).

The total-visit changes also are depicted in a cladogram (Fig. 2A). In the HPV-persistence group, the LDA score was highest for Mycoplasmataceae, which showed significant over-representation ($P = 0.0194$)

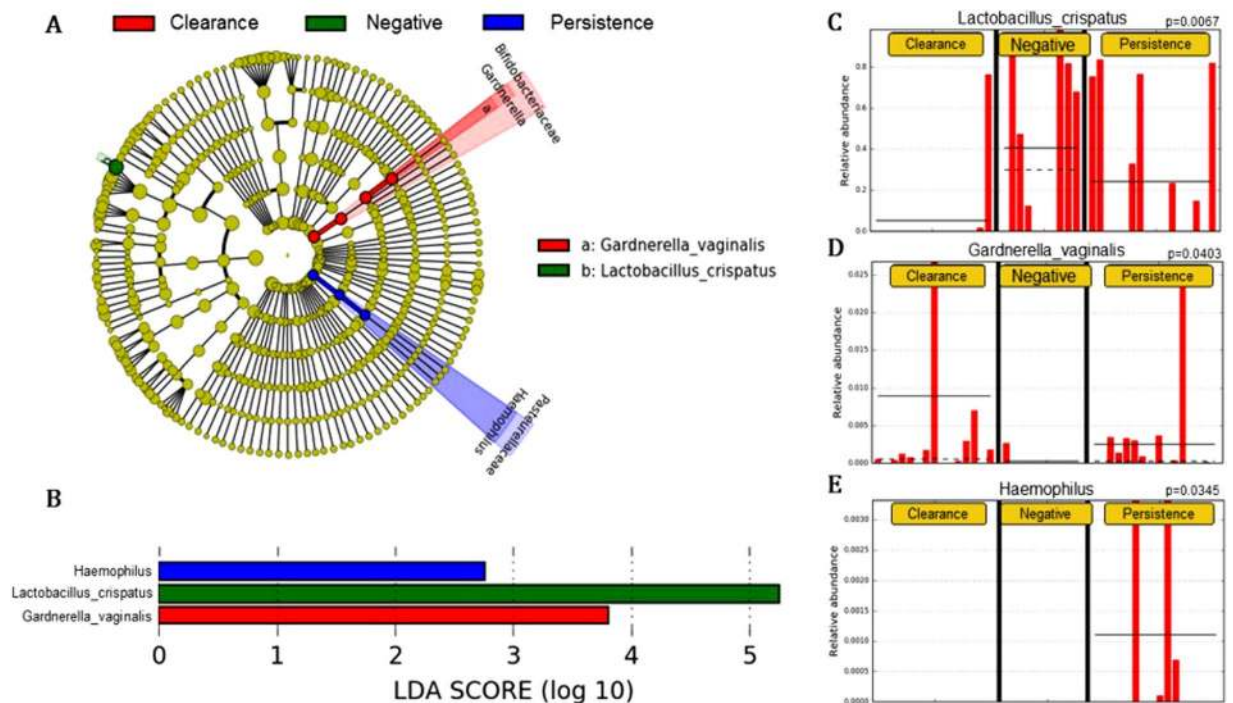


Figure 1. (A) Cladogram representing taxa with different abundances according to HPV groups at baseline visit. Red and green colors show taxa enriched in clearance and persistence women. Brightness is proportional to the abundance of taxa. (B) Histogram of Linear Discriminant Analysis (LDA) scores computed for features differentially abundant among HPV groups (clearance, negative and persistence). The LDA score on the log₁₀ scale is indicated at the bottom. The greater the LDA score is, the more significant the phylotype biomarker is in the comparison. Relative abundance counts of *Lactobacillus crispatus*, *Gardnerella vaginalis* and *Haemophilus* which were found to be significantly over-represented in the HPV negative (C) HPV clearance (D) and HPV persistence (E) groups.

(Fig. 2A,B,G). In the HPV-negative group, the highest LDA score was observed for *Lactobacillus crispatus*, followed by *Corynebacterium sundsvallense*, *Facklamia hominis*, *Fusobacterium naviforme*, *Antinobaculum schaalii*, and *Helcococcus ovis*, with significant over-representation of *Lactobacillus crispatus* ($P = 0.0016$) (Fig. 2A–C). In the HPV-clearance group, *Eubacterium eligens*, *Ureaplasma urealyticum*, and *Gardnerella vaginalis* were over-represented (Fig. 2A,B). The HPV-clearance group also was found to have significantly higher levels of *Gardnerella vaginalis* ($P = 0.0028$), *Eubacterium eligens* ($P = 0.0068$) and *Ureaplasma urealyticum* ($P = 0.0112$) (Fig. 2D–F).

We evaluated species richness (Chao 1 index), alpha-diversity (Shannon index) and beta-diversity of the cervical microbiome in baseline (Fig. 3A–C) and total visit samples (Fig. 3D–F). The results showed species richness (Chao1) was greater in the HPV-clearance and persistence subjects than in the HPV-negative subjects ($p < 0.01$) in total visits. Further, beta diversity differed between HPV negative, clearance and persistence group in total visit samples (bray $p < 0.005$). The baseline visit not showed any significance.

Association of microbes with HPV clearance and HPV persistence. There were significant median differences in the relative abundances of *Lactobacillus crispatus*, *Eubacterium eligens*, *Ureaplasma urealyticum*, *Gardnerella vaginalis* and *Lactobacillus johnsonii* among the HPV groups. A multivariate logistic analysis (Table 2) showed that *Lactobacillus crispatus* (multivariate OR (mOR) = 8.25, 95% CI 2.13–32.0) was the most highly associated with HPV negativity. The multivariate OR of *Eubacterium eligens* for HPV clearance was higher (mOR = 11.5, 95% CI 1.31–101.4) than for HPV negativity. *Ureaplasma urealyticum* (mOR = 7.42, 95% CI 1.3–42.5) and *Gardnerella vaginalis* (mOR = 17.0, 95% CI 2.18–131.8) were more highly associated with HPV clearance than with HPV negativity. *Lactobacillus johnsonii* (mOR = 16.4, 95% CI 1.77–152.2) was associated more with HPV persistence than with HPV clearance.

Discussion

In the present study, we assessed the longitudinal associations of cervical microbes in HPV-negativity, clearance and persistence women. The main findings showed that the cervical microbiome differs significantly by HPV group. We found that women with high proportions of *Lactobacillus johnsonii*, *Haemophilus* (genus) and *Mycoplasmataceae* (family) had the strongest association with HPV persistence, and that women with a cervical microbiome dominated by *Eubacterium eligens*, *Gardnerella vaginalis* and *Ureaplasma urealyticum* had the strongest associations with HPV clearance. Women with a high proportion of *Lactobacillus crispatus* were most

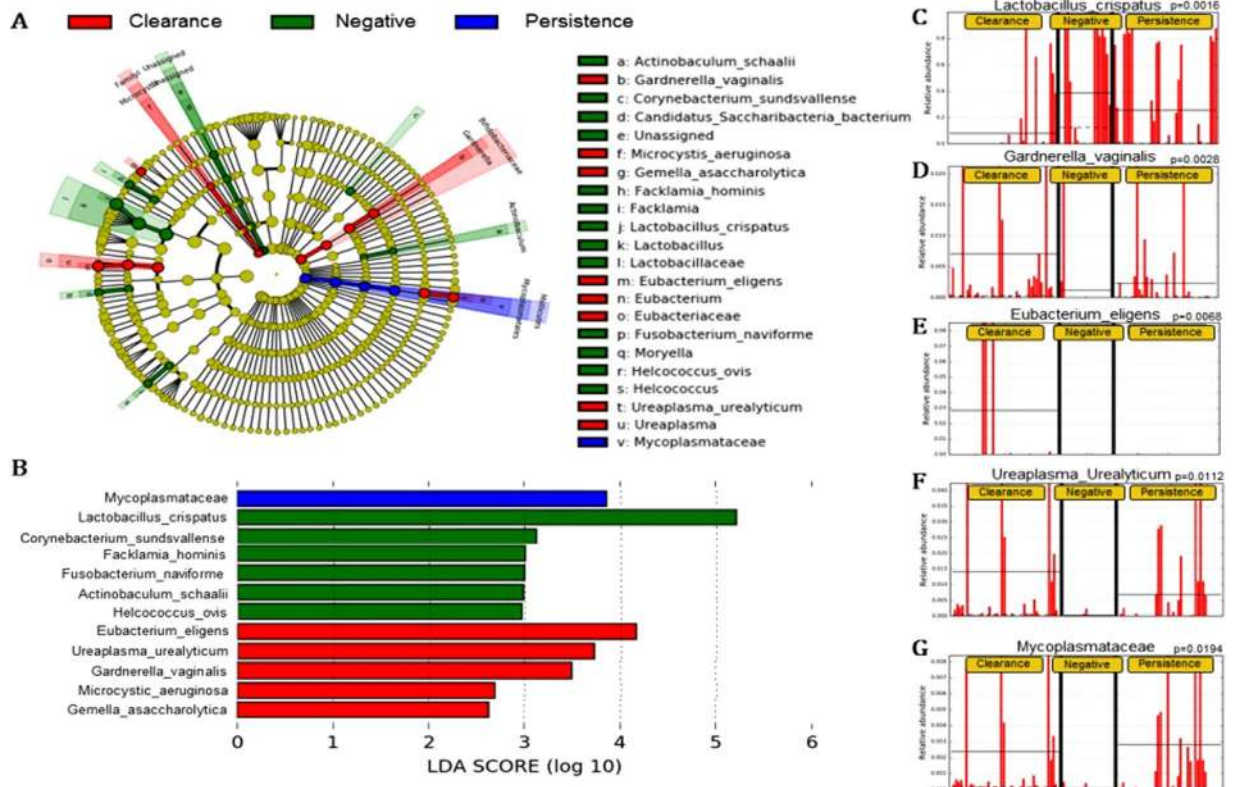


Figure 2. (A) Cladogram representing taxa with different abundances according to HPV groups in total visits. Red and green colors show taxa enriched in clearance and persistence women. Brightness is proportional to the abundance of taxa. (B) Histogram of Linear Discriminant Analysis (LDA) scores computed for features differentially abundant among HPV groups (clearance, negative and persistence). The LDA score on the log₁₀ scale is indicated at the bottom. The greater the LDA score is, the more significant the phylotype biomarker is in the comparison. The relative abundance counts of *Lactobacillus crispatus*, *Gardnerella vaginalis*, *Eubacterium eligens*, *Ureaplasma urealyticum* and Mycoplasmataceae were found to be over-represented in the HPV negative (C), HPV clearance (D–F) and HPV persistence (G) groups.

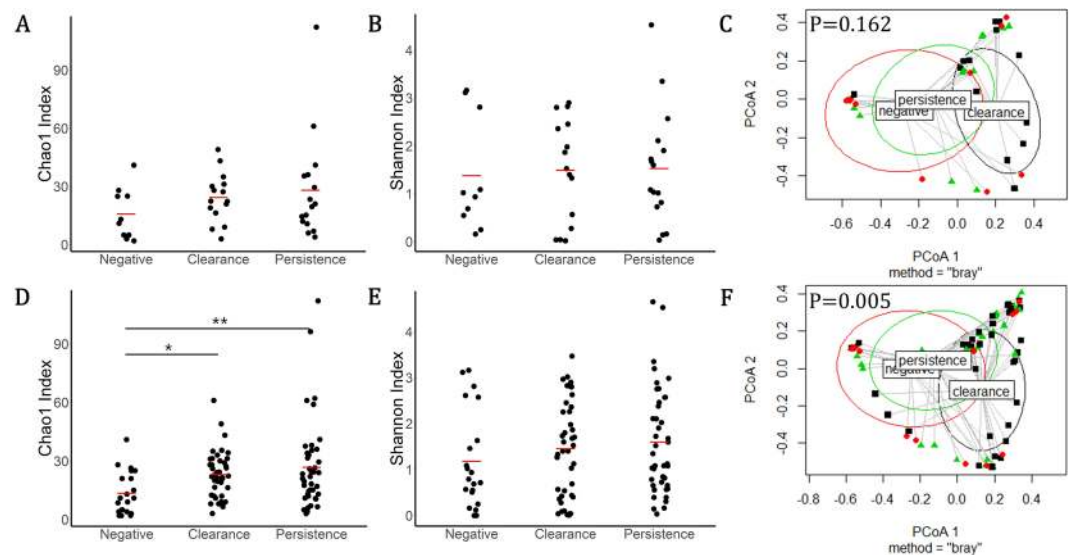


Figure 3. Chao1, alpha diversity (Shannon) and beta diversity index for three groups of women with baseline (A–C) and total-visit (C–E) samples. Values expressed in means. Statistical differences are represented by asterisks (<0.01) according to Kruskal-wallis test. *p < 0.05, **p < 0.01, ***p < 0.001.

Group	N (%) (no exist/exist) ^a	Median (Relative abundance)	p ^b	Multivariate OR (95% CI) ^c
<i>Lactobacillus crispatus</i>				
Clearance + Persistence (ref)	48 (55.8)/38 (44.2)	16.41		1 (ref)
Negative	5 (23.8)/16 (76.2)	38.39	0.009	8.25 (2.13–32.0)
<i>Eubacterium eligen</i>				
Negative + Persistence (ref)	61 (96.8)/2 (3.2)	0.009		1 (ref)
Clearance	34 (77.3)/10 (22.7)	12.32	0.001	11.5 (1.31–101.4)
<i>Ureaplasma urealyticum</i>				
Negative (ref)	18 (85.7)/3 (14.3)	0.015		1 (ref)
Clearance	21 (47.7)/23 (52.3)	1.253	0.003	7.42 (1.30–42.5)
<i>Gardnerella vaginalis</i>				
Negative (ref)	19 (90.5)/2 (9.5)	0	0.0002	1 (ref)
Clearance	18 (40.9)/26 (59.1)	0.03		17.0 (2.18–131.8)
<i>Lactobacillus johnsonii</i>				
Clearance (ref)	42 (95.5)/2 (4.5)	0	0.0101	1 (ref)
Persistence	33 (78.6)/9 (21.4)	0		16.4 (1.77–152.2)

Table 2. Multivariate odds ratios of species (relative abundance >0) for HPV dynamics according the relative abundance of four species. ^aThe number of women with/without each microbiome is presented. ^bp value was from Wilcoxon rank-sum test and p < 0.05 was regarded as significant. ^cMultivariate odds ratios and 95% CI of all variables were estimated by no exist (relative abundance = 0) as reference. All variables in this table are adjusted for age, menopausal status, oral contraceptive use and smoking habit as categorical types.

likely to have HPV-negative infection. Higher bacterial diversity was observed in HPV-persistence women than in women showing HPV negativity.

Based on these results of our two-year longitudinal study, we can suggest that bacterial dysbiosis is a factor associated with HPV dynamics (progression and regression). This study is, to the best of our knowledge, the first to examine the relationship between *Haemophilus* and *Lactobacillus johnsonii* in HPV-persistent women.

Our multivariate logistic analysis confirmed that *Lactobacillus johnsonii* was significantly associated (mORs = 16.4, 95% CI = 1.77–152.2) with HPV persistence among our Korean subjects. There is no previous study on *Lactobacillus johnsonii* in relation to cervical microbiomes, though it has been reported in saliva samples of HPV-positive and HPV-negative oropharyngeal cancer patients. A recent study by Guerrero-Preston *et al.* reported bacterial species *Lactobacillus gasseri/johnsonii* and *Lactobacillus vaginalis* in saliva of HPV-positive and HPV-negative oropharyngeal cancer patients¹⁵. However, the interaction between these microbes in saliva with HPV persistence and cancer risk was not discussed. The LefSe analysis in the present study also confirmed enrichment of the genus *Haemophilus* and family *Mycoplasmataceae* in HPV-persistence women. The results for *Mycoplasmataceae* are in agreement with the finding of Abedamowo *et al.*, that *Mycoplasma hominis* in cervical microbiota is significantly associated with hr-HPV infection¹⁶. It has also been reported that a few *Mycoplasma* species are efficient methylators and can promote cervical carcinogenesis through methylation of hr-HPV and cervical somatic cells¹⁶. One of the previous cervical microbiota studies (16-weeks samples) conducted by Brotman *et al.*¹² demonstrated that high proportions of *Atopobium* spp. (belonging to CST IV-B) were associated with the slowest HPV clearance rate and that the persistence effect may be due to the high proportion of *Atopobium vaginae* that can play a role in the disruption of the epithelial barrier. The plausible molecular mechanisms of HPV persistence infection have been proposed: the viral life cycle and immune system evasion (avoidance of triggering of an immune response by the host)¹⁷, the tethering mechanism on host mitotic chromosomes (to ensure that the viral genome is not lost during cell division)¹⁷, and E2 interaction with host receptors¹⁷. Then, 16S rRNA gene sequencing revealed an increased proportion of *L. johnsonii* in HPV persistence and depletion of *Lactobacillus crispatus* relative to HPV-negative women; additionally, an alpha-diversity analysis provided further confirmation that HPV persistence in the total-visit samples showed increased species richness. Therefore, it can be suggested that changes in cervical microbiota with depletion in *Lactobacillus crispatus* and increased microbial diversity promote HPV infection and might be involved in HPV persistence¹⁸. The alpha-diversity results (Chao 1 index) showed increased diversity in HPV-persistence women relative to HPV-negative women in the total-visit samples. Therefore, high bacterial diversity in HPV persistence can represent an environment that increases the risk of progressing HPV persistence.

In the HPV-clearance women, multivariate logistic analysis and LefSe analysis confirmed the significant enrichment of species *G. vaginalis*, *E. eligens* and *Ureaplasma urealyticum*. The clearance effect might be due to the innate immune response (Toll-like receptors-TLR) that is involved in a cascade of events promoting HPV clearance. TLR 3 and TLR9 play a role in regulating the pro-inflammatory cytokine and anti-viral environments of the lower female genital tract during viral and bacterial infection¹⁹; intense inflammatory response induced by *G. vaginalis* in viral clearance²⁰, as well as viral proteins²¹ and other intrinsic host factors (such as sexual behaviors)^{21,22} responsible for HPV clearance. Another plausible explanation is that this clearance effect is due to an intense inflammatory response or that higher concentrations of inflammation markers (IP-10 and MIG chemokines) induced by bacteria in the cervical region assist in viral clearances²³. One study conducted by Moscicki *et al.*²⁰ reported that the HPV viral clearance effect might be due to the serendipity effect of *Neisseria gonorrhoe*. One

previous study²⁴ reported the lowest clearance rate in women whose cervical microbiota were dominated by *Atopobium* and *Gardnerella*; therefore, further studies are needed to identify the mechanism or mechanisms whereby the composition of the cervical microbiota influence clearance.

Among the HPV-negative women, the *Lactobacillus crispatus* population remained stable at the baseline and follow-up visits, and *Lactobacillus crispatus* was significantly over-represented. Interestingly, in the LEfSe analysis results, we observed *Corynebacterium sandsvallense*, *Facklamia hominis*, *Actinobaculum schaalii* and *Helcococcus ovis* for the first time in the HPV-negative women. As reported earlier, in healthy women, hydrogen-peroxide producing *Lactobacillus* spp. are stable in maintaining the vaginal ecosystem²⁵. Klebanoff *et al.* reported that *Lactobacillus* sp. increase TNF- α and IL-1 α production, activate NF- κ B in THP-1 cells and increase TNF- α production by human monocytes. This suggests that a higher concentration of *Lactobacillus* in the vagina can influence the physiology and host defenses therein²⁵. In our previous study, we reported that *Lactobacillus crispatus* and *Lactobacillus inners* (Cluster III) were the dominant species in HPV-negative women and that high abundance of *L. crispatus* is associated with a low risk of CIN²⁶.

The strength of our present research is that the HPV samplings were longitudinal and that the differences between the cervical microbial compositions (species and genus level) among the HPV-negativity, clearance and persistence samples were identified by 16S rRNA pyrosequencing. One of the most notable findings is that the bacterial communities in the cervical region of the HPV-persistence and clearance women were distinctive from those in the HPV-negative women, suggesting the presence of certain selective pressures contributing to the shift in the cervical microbiota. However, we also must acknowledge certain study limitations: (1) small sample sizes among the study groups, which could have biased the results; (2) an advanced-age study population (mean age: 44). So, the results should be interpreted with caution, and especially, it should be recognized that they might not apply to the general population; (3) another minor limitation of this study is HPV grouping by the liquid hybridization assay Hybrid Capture 2 (HC2) assays. It has been noted that HC2 assay shows some false positive results of 5%. Despite these limitations, HC2 assay has several advantages over other commercially available PCR-based target amplification method^{27–29}. Moreover, HC2 assay is technically well designed and can be easily controlled and performed by the laboratory technician. Whereas PCR-based method contains several steps that must be carefully optimized which make it more difficult to standardize PCR in laboratory condition.

In conclusion, this study analyzed various cervical microbial taxa among different HPV groups, and found that increased bacterial diversity with reduced *L. crispatus* species could be related to HPV persistence and also that the presence and prevalence of a specific cervical microbiome could be relevant to HPV dynamics.

Methods

Subject recruitment. We collected cervical samples from Normal and ASCUS patients (age between 18 and 65 years) who had participated in the Korean human papillomavirus (HPV) cohort study from 2006 to 2013. Detailed information on the inclusion/exclusion criteria for the HPV cohort is provided in our previous paper³⁰. The selected participants were interviewed using a structured questionnaire on their socio-demographic characteristics (Table 1), after which they underwent physical and gynecological examinations. Cervical samples were collected using a Cervix brush (Rovers Medical Devices, Oss, The Netherlands), after which the brush was immediately soaked in a vial of PreservCyt solution (Cytoc Corporation, Marlborough, MA, USA) fixed within a Thin Prep processor. The cytological findings were grouped based on the Bethesda system³¹. The samples were examined for the presence of one or more hrHPV types using the Hybrid Capture 2 (HC2) assay (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68) according to the manufacturer's instructions. The values obtained were recorded in relative light units (RLUs). All relative light units (RLU) measured on a luminometer were divided by the RLU of the respective positive control to provide a ratio. The sample was classified as a positive when the RLU/PC ratio was 1 pg/mL or greater. After baseline sampling, follow-up visits were scheduled at six-month intervals. The collected samples were stored at -80°C for further analysis.

Grouping of HPV status for 2-year follow-up. To investigate the shifts of bacterial communities in the HPV states, we grouped the samples based on the following conditions: (1) HPV negative – persistently negative detection of hrHPV types throughout 24 months (all negative during observation period, 21 samples, 10 subjects); (2) HPV clearance – women who were initially hrHPV (baseline), and then regressed to negative during follow-up (42 samples, 15 subjects); (3) HPV persistence – persistently positive detection of hrHPV types at the baseline and during follow-up (all positive during observation period, 44 samples, 16 subjects) (Fig. 4). Of the 41 subjects, 35 participated at the 2nd visit, 20 at the 3rd visit, 10 at the 4th visit, and 1 at the final visit.

High-risk HPV DNA detection. HPV DNA was detected using the Digene HC2 high-risk DNA test (Qiagen, Gaithersburg, MD, USA) with signal amplification and chemiluminescence for 13 types of HR-HPV scored in RLU/PC. A positive result indicated a concentration of 1 pg/ml or higher than the RLU/cutoff ratio (RLU of the specimen/mean RLU of 2 positive controls).

DNA isolation and pyrosequencing. DNA from cervical samples were isolated using the Fast DNA SPIN kit (MP Biomedicals, Santa Ana, CA, USA). The 16S universal primers 27F (5' GAGTTTGATCMTGGCTCAG 3') and 518R (5' WTTACCGCGGCTGC-TGG 3') were used for amplification. A 20 ng aliquot of each sample was used for a 50 μ l PCR reaction containing 10 \times taq buffer, a dNTP mixture (Takara, shiga, Japan), 10 μ m of the bar-coded fusion primers, and 2 U of taq Polymerase (extaq, takara). The PCR conditions and pyrosequencing protocols are available elsewhere³². Beads recovered from emulsion PCR were deposited on a 454 Pico titer plate, and sequencing was executed using a Roche/454 GS-FLX plus. The sequencing run was performed by Macrogen Ltd. (Seoul, Korea).

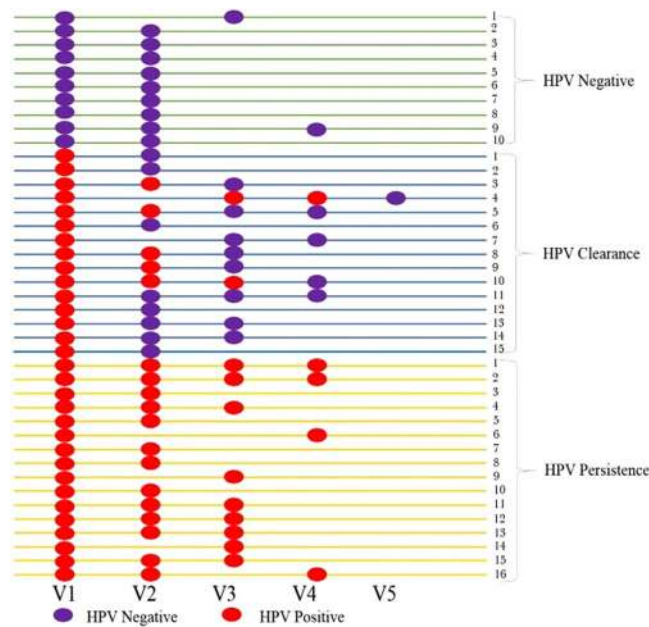


Figure 4. Grouping of HPV status for 2-year follow-up study. V1: baseline visit, V2: visit after 6 months, V3: visit after 12 months, V4: visit after 18 months, V5: visit after 24 months. Only available variables were used in this study, due to missing responses for several questions.

Next-Generation Sequencing using 454 GS-FLX plus. The raw sequences for the samples were arranged using a unique barcode, and low-quality reads (average quality score <25 or read length <300 bp) were removed³². The primer sequences were cut down by employing pairwise sequence alignment, and sequences were gathered to correct for sequencing errors. Taxonomic identification was performed using the EzTaxon-e public database³³ according to the highest pairwise similarity among the BLASTN search results. Possible chimera sequences were removed by the UCHIME algorithm³⁴, and the diversity indices were calculated in Mothur after normalization of the read number in each sample. The potential biomarkers linked to HPV negative, HPV clearance and HPV persistence were analyzed by LDA effect size (LEfSe). Finally, the effect relevance was predicted by LDA³⁵.

Statistical analysis. The Chi-square test was used for testing of the categorical variables, and the Kruskal Wallis test was employed for comparison of the categorical continuous variables. When the number of expected frequencies was less than 5 and the number of cells was more than 25%, Fisher's exact test was performed. To find a significant difference in alpha diversity, the Kruskal Wallis rank sum test was used to evaluate the differences in diversity among the three groups, followed by Dunn's test of multiple comparisons. Multivariate logistic analysis was performed after adjusting for age, menopausal status, oral contraceptive use, and smoking habit. Risk estimates are presented as OR with 95% CI. LEfSe analysis was conducted to find significant differences among the relative abundance taxa³⁵. Beta diversity was calculated with principal coordinates analysis (PCoA) according to the Bray-Curtis distances. A permutational multivariate analysis of variance (PERMANOVA) was implemented to determine significance in distance. Diversity and the PERMANOVA results were analyzed using the R packages "vegan"³⁶. The statistical analysis was performed with SAS 9.4, and R version 3.3.1 with ggplot2 packages was used for visualization³⁷.

Ethical Statement. The Ethics Committee of the National Cancer Center approved this study, and all of the experiments were performed in accordance with the approved guidelines and regulations (IRB No. NCC2016-0147). Written informed consent was obtained from all of the study participants in accordance with good clinical practice.

Data Availability

The datasets generated during the current study are available from corresponding author on reasonable request.

References

1. Bosch, F. X., Lorincz, A., Munoz, N., Meijer, C. J. & Shah, K. V. The causal relation between human papillomavirus and cervical cancer. *J Clin Path* **55**, 244–265 (2002).
2. Munoz, N., Castellsague, X., de Gonzalez, A. B. & Gissmann, L. Chapter 1: HPV in the etiology of human cancer. *Vaccine* **24**(Suppl 3), S3/1–10 (2006).
3. Ho, G. Y., Bierman, R., Beardsley, L., Chang, C. J. & Burk, R. D. Natural history of cervicovaginal papillomavirus infection in young women. *The N Engl J Med* **338**, 423–428, <https://doi.org/10.1056/nejm199802123380703> (1998).
4. Schettino, M. T. *et al.* Persistent papillomavirus type-31 and type-45 infections predict the progression to squamous intraepithelial lesion. *Taiwan J Obst Gyn* **53**, 494–497, <https://doi.org/10.1016/j.tjog.2014.06.001> (2014).

5. Walboomers, J. M. *et al.* Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *The J Pat* **189**, 12–19, 10.1002/(sici)1096-9896(199909)189:1<12::aid-path431>3.0.co;2-f (1999).
6. Castellsague, X. & Munoz, N. Chapter 3: Cofactors in human papillomavirus carcinogenesis—role of parity, oral contraceptives, and tobacco smoking. *J Nat Can Inst. Monographs*, 20–28 (2003).
7. Dahlstrom, L. A. *et al.* Prospective seroepidemiologic study of human papillomavirus and other risk factors in cervical cancer. *Can Epi Bio Prev* **20**, 2541–2550, <https://doi.org/10.1158/1055-9965.epi-11-0761> (2011).
8. Mitra, A. *et al.* The vaginal microbiota, human papillomavirus infection and cervical intraepithelial neoplasia: what do we know and where are we going next? *Microbiome* **4**, 58, <https://doi.org/10.1186/s40168-016-0203-0> (2016).
9. Song, D., Li, H., Li, H. & Dai, J. Effect of human papillomavirus infection on the immune system and its role in the course of cervical cancer. *Onc Let* **10**, 600–606, <https://doi.org/10.3892/ol.2015.3295> (2015).
10. Kjaer, S. K. *et al.* Acquisition and persistence of Human Papillomavirus Infection in Younger Men: A Prospective Follow-up Study among Danish Soldiers. *Can Epi Bio Prev* **14**, 1528–1533, <https://doi.org/10.1158/1055-9965.epi-04-0754> (2005).
11. Goodman, M. T. *et al.* Prevalence, acquisition, and clearance of cervical human papillomavirus infection among women with normal cytology: Hawaii Human Papillomavirus Cohort Study. *Can Res* **68**, 8813–8824, <https://doi.org/10.1158/0008-5472.can-08-1380> (2008).
12. Brotman, R. M. *et al.* Interplay between the temporal dynamics of the vaginal microbiota and human papillomavirus detection. *The J of Inf Dis* **210**, 1723–1733, <https://doi.org/10.1093/infdis/jiu330> (2014).
13. Srinivasan, S. *et al.* Temporal Variability of Human Vaginal Bacteria and Relationship with Bacterial Vaginosis. *PLoS One* **5**, e10197, <https://doi.org/10.1371/journal.pone.0010197> (2010).
14. Gajer, P. *et al.* Temporal dynamics of the human vaginal microbiota. *Sci Tran Med* **4**, 132ra152, <https://doi.org/10.1126/scitranslmed.3003605> (2012).
15. Preston, R. G. *et al.* High-resolution microbiome profiling uncovers *Fusobacterium nucleatum*, *Lactobacillus gasseri*/*johnsonii*, and *Lactobacillus vaginalis* associated to oral and oropharyngeal cancer in saliva from HPV positive and HPV negative patients treated with surgery and chemo-radiation. *Oncotarget* **8**, 110931–110948, <https://doi.org/10.18632/oncotarget.20677> (2017).
16. Adebamowo, S. N. *et al.* *Mycoplasma hominis* and *Mycoplasma genitalium* in the Vaginal Microbiota and Persistent High-Risk Human Papillomavirus Infection. *Front in Pub Health* **5**, 140, <https://doi.org/10.3389/fpubh.2017.00140> (2017).
17. Shanmugasundaram, S. S. & You, J. Targeting Persistent Human Papillomavirus Infection. *Viruses* **9** <https://doi.org/10.3390/v9080229> (2017).
18. Paola, M. D. *et al.* Characterization of cervico-vaginal microbiota in women developing persistent high-risk Human Papillomavirus infection. *Sci Rep* **7**, 10200, <https://doi.org/10.1038/s41598-017-09842-6> (2017).
19. Andersen, J. M., Khairy, A. A. & Ingalls, R. R. Innate Immunity at the Mucosal Surface: Role of Toll-Like Receptor 3 and Toll-Like Receptor 9 in Cervical Epithelial Cell Responses to Microbial Pathogens. *Bio Reprod* **74**, 824–831, <https://doi.org/10.1095/biolreprod.105.048629> (2006).
20. Moscicki, A. B. *et al.* Rate of and Risks for Regression of CIN-2 in adolescents and young women. *Obst & Gynecol* **116**, 1373–1380, <https://doi.org/10.1097/AOG.0b013e3181fe777f> (2010).
21. Bhat, P., Mattarollo, S. R., Gosmann, C., Frazer, I. H. & Leggatt, G. R. Regulation of immune responses to HPV infection and during HPV-directed immunotherapy. *Immunol Rev* **239**, 85–98, <https://doi.org/10.1111/j.1600-065X.2010.00966.x> (2011).
22. Schmeink, C. E. *et al.* Prospective follow-up of 2,065 young unscreened women to study human papillomavirus incidence and clearance. *Int J of Cancer* **133**, 172–181, <https://doi.org/10.1002/ijc.27986> (2013).
23. Shannon, B. *et al.* Association of HPV infection and clearance with cervicovaginal immunology and the vaginal microbiota. *Mucosal Immunol* **10**, 1310–1319, <https://doi.org/10.1038/mi.2016.129> (2017).
24. Ledger, J.W. W. S. S. Chapter 9. *Human Papillomavirus genital infections. Second Edition*. 93 (2016).
25. Klebanoff, S. J., Hillier, S. L., Eschenbach, D. A. & Waltersdorff, A. M. Control of the microbial flora of the vagina by H2O2-generating *Lactobacilli*. *J Infect Dis* **164**, 94–100 (1991).
26. Oh, H. Y. *et al.* The association of uterine cervical microbiota with an increased risk for cervical intraepithelial neoplasia in Korea. *Clinical Microbiol Infect* **21**, 674 e671–679, <https://doi.org/10.1016/j.cmi.2015.02.026> (2015).
27. Castle, P. E. *et al.* Human papillomavirus genotype specificity of hybrid capture 2. *J Clin Microbiol* **46**, 2595–2604, <https://doi.org/10.1128/JCM.00824-08> (2008).
28. Kurian, E. M. *et al.* Cervista HR and HPV 16/18 assays vs hybrid capture 2 assay: outcome comparison in women with negative cervical cytology. *Am J Clin Pathol* **136**, 808–816, <https://doi.org/10.1309/AJCPQD7B7DPVHLH> (2011).
29. Zhao, C. & Yang, H. Approved assays for detecting HPV DNA—design, indications, and validation. *CAP Today* **26**, 40–44 (2012).
30. Hwang, J. H., Lee, J. K., Kim, T. J. & Kim, M. K. The association between fruit and vegetable consumption and HPV viral load in high-risk HPV-positive women with cervical intraepithelial neoplasia. *Cancer Causes Cont* **21**, 51–59 <https://doi.org/10.1007/s10552-009-9433-9> (2010).
31. Solomon, D. *et al.* The 2001 Bethesda System: terminology for reporting results of cervical cytology. *JAMA* **287**, 2114–2119 (2002).
32. Jeon, Y. S., Chun, J. & Kim, B. S. Identification of household bacterial community and analysis of species shared with human microbiome. *Curr Microbiol* **67**, 557–563, <https://doi.org/10.1007/s00284-013-0401-y> (2013).
33. Kim, O. S. *et al.* Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *Int J Syst Evol Microbiol* **62**, 716–721, <https://doi.org/10.1099/ijs.0.038075-0> (2012).
34. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200, <https://doi.org/10.1093/bioinformatics/btr381> (2011).
35. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol* **12**, R60, <https://doi.org/10.1186/gb-2011-12-6-r60> (2011).
36. Oksanen, J. *et al.* Vegan: community ecology package, version 2.4–1, <https://cran.r-project.org/web/packages/vegan/index.html>
37. Wickham, H. *ggplot2: elegant graphics for data analysis*. (Springer, 2016).

Acknowledgements

This work was funded by the National Cancer Center of Korea (Grant No. 1610210).

Author Contributions

All of the authors contributed significantly to this work. M.K.K. and S.S.S. designed the research study; J.K.L. and S.S.S. enrolled the subjects; M.J.K. analyzed the results; S.A. and M.K.K., participated in drafting the paper. All of the authors reviewed and approved the final draft.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-33750-y>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018