Published in partnership with Nanyang Technological University

6

https://doi.org/10.1038/s41522-025-00747-1

Metagenomic insight into the vaginal microbiome in women infected with HPV 16 and 18

Check for updates

Da-Ryung Jung 0^1 , Yeseul Choi^{2,3}, Minsoo Jeong¹, Vineet Singh¹, Se Young Jeon^{4,5}, Incheol Seo^{5,6}, Nora Jee-Young Park^{5,7}, Yoon Hee Lee⁴, Ji Young Park⁷, Hyung Soo Han^{2,3,5,8}, Jae-Ho Shin^{1,9} \boxtimes & Gun Oh Chong^{4,5} \boxtimes

Human papillomavirus (HPV) 16 and 18 (HPV 16/18) account for over 70% of cervical cancer (CC) cases, yet their interaction with the vaginal microbiome (VM) remains unclear. This study explored the association between high-risk HPV types (HR-HPVs), VM composition and bacterial function using shotgun metagenomic sequencing. In early-stage cervical lesions, the HPV 16/18 group showed reduced *Lactobacillus*-dominant community state types compared to other HR-HPVs, while invasive CC exhibited increased pathogenic bacteria, including *Streptococcus agalactiae*, *Fannyhessea vaginae*, and *Sneathia vaginalis*. The VM associated with HPV 16/18 was enriched in immune response and inflammation pathways, whereas other HR-HPVs were linked to cellular metabolism and hormonal signaling. Notably, HPV 16/18 exhibited stronger bacterial-fungal correlations, indicating shifts in the microbial genomic signatures. Our study links VM differences with HPV 16/18 oncogenic potential across cervical lesion stages, urging further research for better diagnostics and treatments.

Human papillomavirus (HPV) is the most prevalent sexually transmitted infection worldwide and a primary cause of cervical cancer, as well as other cancers including anal, oropharyngeal, and penile cancers¹⁻³. Although the immune system clears most transient HPV infections, persistent infections by high-risk HPV (HR-HPV) can lead to the integration of viral genes into the host genome, thereby instigating disease^{4,5}. Particularly, HPV 16 and 18, which are responsible for over 70% of all cervical cancer cases, are the most oncogenic HPV types. These types are significantly more likely to progress to cervical intraepithelial neoplasia (CIN) and carcinomas compared to other HR-HPV types⁶. Recent studies have shown that women infected with HPV 16 or 18 are more likely to develop high-grade lesions than those infected with other HR-HPV types7. However, despite evidence of increased oncogenic potential and persistence of HPV 16 and 18, the underlying mechanisms remain poorly understood. Some researchers have noted that the cervicovaginal microenvironment, along with individual variations in immune function and genetic susceptibility, may play significant roles^{8,9}. Numerous studies have investigated the association between the vaginal microbiome (VM) and HPV infection, precancerous cervical lesions, or cervical cancer. The composition of vaginal bacterial species can be categorized into five community state types (CSTs): CST I (dominated by *Lactobacillus crispatus*), CST II (dominated by *Lactobacillus gasseri*), CST III (dominated by *Lactobacillus gasseri*), CST V (dominated by *Lactobacillus gasseri*), CST V (dominated by *Lactobacillus gasseri*), CST V (dominated by *Lactobacillus gasseri*). Furthermore, recent studies have employed the VAginaL community state type Nearest CentroId classifier (VALENCIA) tool to categorize CSTs further¹⁰.

A healthy VM, predominantly composed of lactobacilli that produce lactic acid, plays a crucial role in various defense mechanisms that protect women against diseases^{11,12}. Conversely, persistent HR-HPV infection is associated with a depletion of lactobacilli and an enrichment of diverse bacterial populations, including *Fannyhessea vaginae*, *Gardnerella vaginalis*, *Prevotella*, *Finegoldia*, *Dialister*, and Sneathia^{13–15}. These findings indicate

¹Department of Applied Biosciences, Kyungpook National University, Daegu, Republic of Korea. ²Department of Biomedical Science, Graduate School, Kyungpook National University, Daegu, Republic of Korea. ³BK21 Four Program, School of Medicine, Kyungpook National University, Daegu, Republic of Korea. ⁴Department of Obstetrics and Gynecology, Kyungpook National University Chilgok Hospital, Daegu, Republic of Korea. ⁵Clinical Omics Institute, Kyungpook National University, Daegu, Republic of Korea. ⁶Department of Immunology, School of Medicine, Kyungpook National University, Daegu, Republic of Korea. ⁷Department of Pathology, Kyungpook National University Chilgok Hospital, Daegu, Republic of Korea. ⁸Department of Physiology, School of Medicine, Kyungpook National University, Daegu, Republic of Korea. ⁹NGS Core Facility, Kyungpook National University, Daegu, Republic of Korea. ¹Demail: jhshin@knu.ac.kr; gochong@knu.ac.kr that persistent HR-HPV infection may induce aberrant changes in the cervicovaginal microenvironment. Furthermore, non-lactobacilli bacteria may exacerbate HR-HPV infection by producing enzymes and metabolites that influence various cellular pathways, demonstrating a bidirectional relationship. A recent study in a Greek cohort demonstrated that women infected with HPV 16 and 18 exhibited significantly lower levels of *Lacticaseibacillus* compared to those infected with other HR-HPV types, potentially explaining the higher risk associated with HPV 16 and 18 in the vaginal microenvironment¹⁶. Moreover, in a Puerto Rican cohort, increased fungal diversity was observed in HR-HPV-positive cervical samples, with *Malassezia* identified as a biomarker¹⁷. In addition, viral community analysis has validated that the diversity of human viruses is significantly diminished in the presence of HPV¹⁸.

Shotgun metagenomic sequencing overcomes the technical limitations of 16S rRNA amplicon sequencing by facilitating detailed compositional and functional analyses of the microbiota at the species level. This method is particularly crucial in studies of vaginal microbiota, where high-resolution species-level information is essential for CST classification. Additionally, it enables comprehensive investigation of the entire microbiome—including bacteria, fungi, and viruses—without the biases associated with amplification. This method also provides functional insights through genome-level assembly, which explains its increasing adoption in vaginal metagenomics studies.

To date, most studies have focused on the correlation between the VM and HPV infection status or cervical lesions, but few have investigated the distinctions between HPV 16 and 18 and other HR-HPV types. To address this gap, we employed shotgun metagenomic sequencing to comprehensively explore the associations between specific HR-HPV types, VM composition and function, and the severity of cervical lesions. Given the markedly higher oncogenic potential and persistence of HPV 16 and 18, we explored whether these types are linked to distinct microbial characteristics —such as reduced *Lactobacillus* dominance, enrichment of pathogenic taxa, and activation of immune-related pathways—that may underlie their enhanced pathogenicity compared to other HR-HPV types.

Results

Study cohort and HR-HPV status

In this study, 68 women with confirmed pathological diagnoses were classified into three groups: group 1 (normal and LSIL), group 2 (HSIL), and group 3 (ICC). These groups were further subdivided based on HR-HPV status: HPV 16/18 and other HR-HPVs. Demographic and clinical characteristics are comprehensively detailed in Table 1 and Supplementary Table 1. Across the entire cohort, only marital status showed a significant difference (p = 0.0383). However, in group 3, both age and smoking status differed significantly (p = 0.0093 and p = 0.0847, respectively).

Individuals infected with HPV 16/18 and other HR-HPVs were recruited in a relatively balanced manner across each cervical lesion group (Table 1 and Fig. 1a). A total of 13 HR-HPV types were identified, with HPV 16 (83.3%) being the most prevalent in the HPV 16/18 group. Among other HR-HPVs, HPV 52, HPV 58 (each 26.3%), and HPV 51 (15.8%) were the most common (Table 1 and Fig. 1a). Although most participants were infected with a single HR-HPV type, co-infections were observed in two individuals in group 2; one was infected with both HPV 31 and HPV 39, and the other with HPV 52 and HPV 58 (Fig. 1b).

Distribution of vaginal CSTs

The vaginal CST of each sample was characterized using VALENCIA. In our cohort, four CSTs (CSTs I, III, IV-B, and IV-C) were identified based on the composition and relative abundance of bacterial species (Table 2 and Supplementary Table 2). These CSTs were further subdivided into sub-CSTs, such as CST I-A (almost entirely *L. crispatus*), CST I-B (containing less *L. crispatus* but remaining predominantly this species), CST III-A (almost entirely composed of *L. iners*), CST III-B (less *L. iners* yet still primarily composed of this species), CST IV-B (high to moderate relative abundance of *G. vaginalis* and *F. vaginae*), CST IV-C0 (relatively even

community distribution including *Prevotella* spp.), CST IV-C1 (dominated by *Streptococcus* spp.), and CST IV-C4 (dominated by *Staphylococcus* spp.). There was no significant difference in the distribution of CSTs according to HR-HPV types. However, in group 2, women with CSTs I and III, which were dominated by *Lactobacillus* spp., exhibited a higher frequency of other HR-HPVs (Table 2). Furthermore, the proportion of CSTs I and III decreased while that of CST IV increased as the stage of cervical lesions progressed. Notably, in group 3, a significant number of women infected with other HR-HPV types presented a VM primarily consisting of *Streptococcus* spp. (CST IV-C1) and *Staphylococcus* spp. (CST IV-C4).

Comparing the average relative abundance of bacterial species used to assign CSTs across groups confirmed that *Lactobacillus* spp. comprised more than half of the total bacterial composition in each of the other HR-HPVs of groups 1 (66.46%) and 2 (56.69%), slightly higher than in HPV 16/18 of groups 1 (44.55%) and 2 (45.79%) (Fig. 2). However, in group 3, HPV 16/18 (32.33%) exhibited a higher proportion of *Lactobacillus* spp. compared to other HR-HPVs (11.97%). Additionally, across all cervical lesion groups, the combined proportions of *G. vaginalis* and *F. vaginae* were consistently higher in HPV 16/18 than in other HR-HPVs. Additionally, we ensured the reliability of our taxonomic classification approach by confirming a high level of concordance between our dataset and Human Vaginal Microbiome Genome Collection (VMGC) reference ¹⁹ (Supplementary fig. 1).

Differential composition of the vaginal microbiome

At the species level, Shannon diversity did not vary among HR-HPV types, either across the cohort or when categorized by cervical lesions (Supplementary Fig. 2). However, regardless of HR-HPV type, Shannon diversity significantly increased with both the severity of cervical lesions and age range.

NMDS analysis elucidated the associations between vaginal bacterial communities and specific bacterial species (L. crispatus, L. iners, G. vaginalis, F. vaginae, Streptococcus spp., and Staphylococcus spp.) utilized for CST assignment in our cohort (Fig. 3a-d). Each individual's bacterial composition was distinctly categorized by CST based on bacterial species, with five species showing significant associations: L. crispatus (p = 0.001), L. iners (p = 0.042), G. vaginalis (p = 0.002), F. vaginae (p = 0.002), and Staphylococcus spp. (p = 0.001) (Fig. 3a and Supplementary Table 3). L. crispatus exhibited a negative correlation with four bacterial species, excluding L. iners (Supplementary Fig. 3), and demonstrated a negative correlation with the severity of cervical lesions and age, but not with HR-HPV types (HPV 16/18 were assigned 2 and other HR-HPVs were assigned 1 considering clinical severity). F. vaginae showed a strong positive correlation with G. vaginalis and Streptococcus spp. Across all cervical lesion groups, L. crispatus, along with either G. vaginalis or F. vaginae, commonly influenced the distribution of bacterial communities (Fig. 3b-d and Supplementary Table 3). However, the delineation of groups based on HR-HPV types was not clearly discernible in the NMDS plots.

A comparison of species abundance between HPV 16/18 and other HR-HPVs revealed significant variations in 14 species across the entire cohort and in 3, 19, and 16 species within groups 1, 2, and 3, respectively (log2 fold change >1 or <-1, *p* < 0.05, DESeq2) (Fig. 3e and Supplementary Fig. 4). In the entire cohort, L. gasseri, Lactobacillus paragasseri, and Streptococcus agalactiae were predominantly associated with other HR-HPVs, while F. vaginae, Anaerococcus vaginalis, Anaerococcus obesiensis, and Escherichia coli were prevalent in HPV 16/18 (Supplementary fig. 4). Notably, Finegoldia magna was significantly enriched with other HR-HPVs in group 1. In contrast, Lactobacillus spp. and Prevotella spp., including L. gasseri, L. paragasseri, Lactobacillus sp. JM1, Prevotella buccalis, and Prevotella sp. oral taxon 299, were more abundant with the other HR-HPVs in group 2. F. vaginae demonstrated a significant increase in abundance in both groups 2 and 3 infected with HPV 16/18. All species that exhibited significant differences between HPV 16/18 and other HR-HPVs in group 3 were taxa predominantly enriched with HPV 16/18, such as L. iners, Limosilactobacillus vaginalis, Streptococcus dysgalactiae, Aerococcus christensenii, Peptoniphilus harei, P. buccalis, and Sneathia vaginalis.

Table 1 | Demographic and clinical characteristics of participants included in the study cohort

Variables	Group 1			Group 2			Group 3		
	Other HR- HPVs (n = 14)	HPV 16/ 18 (<i>n</i> = 9)	p value	Other HR- HPVs (n = 12)	HPV 16/ 18 (<i>n</i> = 11)	p value	Other HR- HPVs (<i>n</i> = 12)	HPV 16/ 18 (<i>n</i> = 10)	p value
Age (years)	44.2 ± 10.3	41.1 ± 12.7	0.5279	38.23 ± 5.4	37.5 ± 8.1	0.7825	56.2 ± 9.4	43.9 ± 6.7	0.0093
BMI (kg m ⁻²)	25.3 ± 3.2	24.5 ± 4.0	0.5819	23.1 ± 3.4	23.5 ± 4.3	0.8051	26.3 ± 3.5	25.2 ± 2.7	0.4262
Married women (n, %)	12 (85.7)	6 (66.7)	0.2905	11 (91.7)	8 (72.7)	0.2417	12 (100)	8 (80.0)	0.1124
Parity (n)	1.4 ± 0.9	1.2 ± 1.0	0.6168	1.7 ± 1.1	1.8 ± 1.3	0.7655	0	0	1.0000
Smoking (n, %)	0 (0)	1 (11.1)	0.3913	2 (16.7)	2 (18.2)	1.0000	1 (8.3)	4 (40.0)	0.0847
HPV type (n, %)			0.0017			0.0001			0.0049
16	0 (0)	7 (77.8)		0 (0)	11 (100)		0 (0)	7 (70.0)	
18	0 (0)	2 (22.2)		0 (0)	0 (0)		0 (0)	3 (30.0)	
31	0 (0)	0 (0)		1 (7.1)	0 (0)		1 (8.3)	0 (0)	
33	0 (0)	0 (0)		0 (0)	0 (0)		2 (16.7)	0 (0)	
36	1 (7.1)	0 (0)		0 (0)	0 (0)		0 (0)	0 (0)	
39	0 (0)	0 (0)		2 (14.3)	0 (0)		0 (0)	0 (0)	
45	0 (0)	0 (0)		0 (0)	0 (0)		2 (16.7)	0 (0)	
51	2 (14.3)	0 (0)		3 (21.4)	0 (0)		1 (8.3)	0 (0)	
52	3 (21.4)	0 (0)		2 (14.3)	0 (0)		4 (33.3)	0 (0)	
53	1 (7.1)	0 (0)		0 (0)	0 (0)		0 (0)	0 (0)	
56	3 (21.4)	0 (0)		0 (0)	0 (0)		0 (0)	0 (0)	
58	4 (28.6)	0 (0)		6 (42.9)	0 (0)		1 (8.3)	0 (0)	
62	0 (0)	0 (0)		0 (0)	0 (0)		1 (8.3)	0 (0)	
Multiple HPV infection (n, %)	0	0	1.0000	2 (16.7)	0 (0)		0 (0)	0 (0)	1.0000



Fig. 1 | Distribution of high-risk human papillomavirus (HR-HPV) types in the study cohort. a The number of participants with HPV 16, HPV 18, and other HR-HPV types in each cervical lesion group. b A heatmap illustrating the presence of HR-HPV types in each group.

Enrichment of microbial functional genes and pathways

From the shotgun metagenomic data, an average of 60% of the genes per sample were assigned to non-redundant genes. Functional annotations were assigned to 9224 KEGG orthologs (KOs), representing molecular functions, by aligning non-redundant genes with the KEGG database. Contrary to observations at the species level, Shannon diversity at the KO level was higher in HPV 16/18 compared to other HR-HPVs, although this difference was not statistically significant (p = 0.061, Wilcoxon test) (Supplementary Fig. 5). No differences in diversity were observed with respect to cervical lesions. Additionally, when categorized by the severity of cervical lesions, HPV 16/18 exhibited markedly higher diversity in group 3 compared to other HR-HPVs (p = 0.021, Wilcoxon test) (Supplementary Fig. 5). The overall KO composition did not differ significantly among HR-HPV types (p = 0.308, PERMANOVA); however, considerable variation in composition was observed depending on the cervical lesion (p = 0.001, PERMANOVA) (Supplementary fig. 6). Additionally, a minor yet significant difference in composition was noted in group 3 (p = 0.032, PERMANOVA), but not in groups 1 (p = 0.527, PERMANOVA) or 2 (p = 0.837, PERMANOVA) (Fig. 4a–c).

We then focused on KOs differentially enriched between HPV 16/18 and other HR-HPVs (Fig. 4d–f). Our results revealed distinct KO patterns across varying severities of cervical lesions: group 1 (14/110 KOs), group 2

Table 2 | Distribution of vaginal community state types in the study cohort

CST	Group 1 (n, %)			Group 2 (n, %)			Group 3 (n, %)		
	HPV 16/ 18 (n = 9)	Other HR- HPVs (<i>n</i> = 14)	p value	HPV 16/ 18 (n = 11)	Other HR- HPVs (<i>n</i> = 12)	p value	HPV 16/ 18 (n = 10)	Other HR- HPVs (n = 12)	p value
I + III	4 (44.4)	10 (71.4)	0.2294	6 (54.5)	7 (58.3)	0.8628	3 (30.0)	1 (8.3)	0.5895
I	2 (22.2)	4 (28.6)	0.7721	2 (18.2)	4 (33.3)	0.4425	1 (10.0)	1 (8.3)	0.9473
I-A	2 (22.2)	3 (21.4)	1	1 (9.1)	3 (25.0)	0.3490	1 (10.0)	1 (8.3)	0.9473
I-B		1 (7.1)	0.4760	1 (9.1)	1 (8.3)	1			
III	2 (22.2)	6 (42.9)	0.3403	4 (36.4)	3 (25.0)	0.5892	2 (20.0)		0.1282
III-A	2 (22.2)	5 (35.7)	0.5276	2 (18.2)		0.1474	1 (10.0)		0.3153
III-B		1 (7.1)	0.4760	2 (18.2)	3 (25.0)	0.7306	1 (10.0)		0.3153
IV	5 (55.6)	4 (28.6)	0.2913	5 (45.5)	5 (41.7)	0.8861	7 (70.0)	11 (91.7)	0.2178
IV-B	4 (44.4)	4 (28.6)	0.4686	2 (18.2)	3 (25.0)	0.7306	6 (60.0)	4 (33.3)	0.2365
IV-C0				1 (9.1)		0.3384	1 (10.0)		0.3153
IV-C1	1 (11.1)		0.2468	2 (18.2)	2 (16.7)	0.9626		1 (8.3)	0.4113
IV-C4								6 (50.0)	0.0118





display the proportion of dominant species in individual samples, categorized by their CST classification (**d**: group 1; **e**: group 2; **f**: group 3).

(34/48 KOs), and group 3 (19/1799 KOs). This indicated a stark contrast in KO abundance between other HR-HPVs and HPV 16/18. Next, we categorized the differentially expressed KOs into respective pathways, enabling a comprehensive functional analysis. Groups 1, 2, and 3 demonstrated significant differences, involving 8, 14, and 25 KEGG pathways, respectively, between HPV 16/18 and other HR-HPVs (log2 fold change >1 or <-1, p < 0.05) (Fig. 4g and Supplementary fig. 7). In group 1, HPV 16/18 was significantly associated with bacterial biosynthesis (streptomycin

biosynthesis), cell cycle regulation (p53 signaling pathway), and immune responses (*Yersinia* infection, leishmaniasis). Conversely, pathways associated with inflammation and immune cell migration, including arachidonic acid metabolism and leukocyte transendothelial migration, were predominant in other HR-HPVs. In group 2, the microbial community in HPV 16/18 exhibited an accumulation of pathways involved in immune response modulation (Janus kinase/signal transducer and activator of transcription (JAK-STAT) signaling pathway) and malignant transformation (endometrial



Fig. 3 | Comparison of the vaginal microbiome (VM) between human papilloma virus (HPV) 16/18 and other high-risk (HR)-HPVs. Non-metric multidimensional scaling (NMDS) of VM composition in (a) the entire group, b group 1, c group 2, and d group 3, based on the Bray–Curtis dissimilarities distance matrix. Black arrows highlight properties that significantly contribute to dispersion in each group, while gray arrows represent non-significant contributors. e Bar charts display

the differentially abundant bacterial species in each disease group comparing HPV 16/18 and other HR-HPVs. Log2 fold change values are used to denote significant enrichment, with positive values (red bars) indicating enrichment in HPV 16/18 and negative values (blue bars) indicating enrichment in other HR-HPVs (log2 fold change >1 or <-1, p < 0.05, DESeq2).

cancer, melanoma, and non-small cell lung cancer). The group pertaining to other HR-HPVs was characterized by enhanced pathways associated with cellular metabolism (flavone and flavonol biosynthesis) and hormonal signaling (oxytocin signaling pathway). In group 3, other HR-HPVs showed involvement in pathways associated with cell adhesion (extracellular matrix (ECM) receptor interaction and cell adhesion molecules), immune response (Toll-like receptor signaling pathway and natural killer cell-mediated cytotoxicity), and inflammation (tumor necrosis factor (TNF) signaling pathway).

We focused on the genes of microbial enzymes involved in the degradation of glycogen and mucin, which are potential sources of carbon and energy in the vaginal environment. We annotated enzymes involved in glycogen degradation, such as glycogen debranching enzyme and glycogen phosphorylase, and enzymes involved in mucin degradation, such as sialidase, fucosidase, and galactosidase, through CAZymes. Although not statistically significant, HPV 16/18 generally

showed higher abundance in most enzymes compared to other HR-HPVs (Fig. 4h).

Construction of MAG

The metagenome assembly revealed the taxonomic distribution from a genomic perspective, providing a comprehensive understanding of the genetic composition of microbial communities. A phylogenetic tree was constructed using 137 MAGs, annotated with information on completeness, contamination, genome size, and phylum (Fig. 5 and Supplementary Table 4). The genome sizes of the MAGs ranged from 0.6 to 6.8 Mb (average = 1.7 Mb). According to the minimum information about a metagenome-assembled genome (MIMAG) standard²⁰, 53.28% (73/137) of the MAGs were classified as high-quality genomes (completeness \geq 90%, contamination <5%), while the remaining 46.72% (64/137) were classified as medium-quality genomes (completeness \geq 50%, contamination <10%). Using a genomic similarity threshold (ANI \geq 95%) at the prokaryotic



Fig. 4 | **Bacterial functional profiling of the vaginal microbiome (VM).** The volcano plots displaying log2 fold change differences in KEGG orthologs (KOs) in (**a**) group 1, **b** group 2, and **c** group 3. Points in blue indicate KOs which were differentially abundant in other HR-HPV, and points in red in HPV 16/18, respectively. Non-metric multidimensional scaling (NMDS) of KOs composition in (**d**) group 1, **e** group 2, and **f** group 3, based on the Bray–Curtis dissimilarities distance matrix. **g** Bar charts illustrating differentially abundant KEGG pathways across each disease group (log2 fold change >1 or <-1, *p* < 0.05, DESeq2) between

HPV 16/18 and other HR-HPV types. Log2 fold change values are used to denote significant enrichment, with positive values (red bars) indicating enrichment in HPV 16/18 and negative values (blue bars) indicating enrichment in other HR-HPV types. **h** Differences in key functions associated with the metabolism of glycogen and mucin, two abundant host resources in the vagina. The boxplot displays the combined abundance of the specified enzymes with normalized square root transformation. Significant differences between groups are indicated (*p < 0.05 and *p < 0.1).

species level, we grouped the 137 MAGs into 120 species-level bins. These bins were assigned to eight bacterial phyla: *Actinomycetota* (consisting of 17 species), *Bacilliota* (80 species), *Bacteroidota* (16 species), *Patescibacteria* (1 species), *Desulfobacterota* (1 species), *Pseudomonadota* (1 species), *Fusobacteriota* (3 species), and *Campylobacterota* (1 species). Comparison with the VMGC reference showed that among the 137 MAGs in our dataset, 85.40% had ≥83% ANI and 76.64% had ≥95% ANI, demonstrating high concordance with the VMGC dataset. Furthermore, RPKM-based analysis of key MAGs classified at the species level further supported the high sensitivity of our classification method (Supplementary Fig. 8). Furthermore, to assess the reliability of the contig-based classification, we compared the key vaginal microbial taxa identified from contigs with those identified using direct short-read classification (Kraken2). The results showed a high degree of consistency in microbial composition between the two methods,

indicating that contig-based assembly did not introduce significant biases in taxonomic classification (Supplementary Fig. 9).

Integrated network of multi-kingdom virginal microbiome

To investigate the co-occurrence patterns of VM, we analyzed multikingdom networks that included bacteria, fungi, and viruses examining both intra-kingdom and inter-kingdom correlations (Fig. 6a–c). In the network of group 3, the highest number of nodes and edges was observed. Additionally, the HPV 16/18 networks in group 1 and 3 were more complex than those of other HR-HPVs based on the number of nodes and edges, but no differences were found in group 2. In groups 1 and 3, HPV 16/18 exhibited a greater number of inter-kingdom correlations than did other HR-HPVs (Fig. 6d). The only intra-kingdom negative correlations were observed in group 3. The co-occurrence networks showed that HPV 16/18 made the



Fig. 5 | Distribution of metagenomic assembled genomes (MAGs) in the vaginal microbiome (VM). Phylogenetic tree of MAGs, including their completeness (red), contamination (blue), and genome size (gray). Phylum information is displayed at the innermost level of the annotation layer.

network unstable and enhanced the inter-kingdom correlations between bacteria and fungi across all cervical lesion group (Fig. 6e).

Discussion

The cervical-vaginal microenvironment reflects host physiology and plays a crucial role in women's health. Indeed, several studies have reported a bidirectional association between HPV and the VM. HPV-positive women exhibited higher microbial diversity and a lower relative abundance of *Lactobacillus* spp. compared to HPV-negative women^{21,22}. Furthermore, women with a non-*Lactobacillus*-dominant VM are twice as likely to contract oncogenic HPV types as those with a *L. crispatus*-dominant VM²³. However, Lebeau et al. emphasized that not all HPV infections (including those involving high-risk genotypes) significantly alter the VM, suggesting that the impact on the microbial community may vary depending on the pathological outcomes of the infection²⁴. Thus, distinguishing between pathological outcomes is essential to elucidating the oncogenic potential and persistence of HPV 16/18 compared to other HR-HPVs.

Our study involved a cross-sectional analysis to profile the VM of women infected with HPV 16/18 and other HR-HPV types, categorized according to the severity of cervical lesions. This research employed metagenomic sequencing to explore microbial (bacterial, fungal, and viral) species, providing high-resolution and functional characteristics of the microbiome. Additionally, it expanded the existing database to include 120 bacterial species of cervical-vaginal microenvironment (as identified from 137 MAGs) that were identified through gene assembly. Our results suggest that VM composition differs in HPV 16/18 infections, particularly in relation to disease progression. An imbalanced VM may be associated with the oncogenic potential and persistence of HPV 16/18, indicating a possible interplay between VM and disease status.

In this study, the bacterial communities in samples were classified into four CSTs based on species composition and relative abundance to assess associations with HPV infection or cervical lesion severity. CSTs II and IV (predominantly L. gasseri and L. jensenii, respectively) were not observed in our cohort, which aligns with previous studies indicating a lower prevalence of these CSTs in Korean cohorts^{25,26}. Earlier studies suggest that L. gasseri could potentially aid in HPV clearance^{27,28}. However, the scarcity of L. gasseri-dominant communities among Korean women may suggest a reduced capability for HPV clearance. Although not significant, a higher prevalence of CSTs I and III, which are Lactobacillus spp.-dominant, was observed in other HR-HPVs compared to HPV 16/18 in group 1 (patients with normal or early-stage cervical lesions). This suggests that a healthy vaginal environment may assist in HPV 16/18 clearance when there is no or minimal pathological abnormality. Indeed, L. iners exhibits unique characteristics among vaginal lactobacilli, and its role continues to be debated due to its inconsistent associations with adverse reproductive health outcomes. Nevertheless, L. iners is frequently observed in healthy Asian women^{21,29}, and some studies have associated it with HPV clearance⁴. Additionally, a normal cervical microenvironment post-HPV infection



Fig. 6 | **Integrated multi-kingdom network of the vaginal microbiome (VM).** Networks depicting **a** group 1, **b** group 2, and **c** group 3 contained bacterial (aqua), fungal (orange), and viral (gray) taxa with positive (blue) and negative (red) correlations. Node size indicates the degree of connection. **d** The number of inter-

tends to shift from an *L. crispatus*-dominant community to *L. iners*dominant community³⁰, warranting further research to fully understand its clinical significance. In this study, CST differences were more pronounced across stages of cervical lesions than by HPV type, with groups 1 and 2 exhibiting significantly higher prevalence of CSTs I and III and lower prevalence of CST IV compared to group 3. This pattern indicates a decrease in *Lactobacillus* spp. and an increase in other aerobic bacteria with advancing cervical cancer stages, reaffirming the need to differentiate pathological outcomes when comparing VMs based on HPV type, as argued by Lebeau et al.²⁴.

Persistent HPV infection is a major risk factor for ICC, while transient infection often regresses without progressing to high-grade lesions. Identifying the infection type is crucial, but we lacked longitudinal data to directly distinguish between persistent and transient infections. To address this limitation, we compared HPV types in metagenomic data with qPCR results (Seegene Anyplex II HPV 28 Assay Kit), calculating relative abundance using the PaVE database³¹. Several samples positive for HPV by qPCR were not detected in the metagenomic data, likely due to the higher sensitivity and specificity of qPCR (Supplementary Table 5). qPCR, using specific primers, can effectively detect HPV even at low viral loads, whereas metagenomic sequencing depends on sequencing depth and the viral DNA abundance, which may reduce detection sensitivity. Previous studies have established that a high HPV viral load is associated with persistent HPV infection^{32,33}. Accordingly, we categorized our metagenomic data into high viral load samples (HPV detected, suggesting persistent infection) and low viral load samples (HPV not detected, indicating transient infection) (Supplementary fig. 10). In the HPV 16/18 group, high viral load samples exhibited a distinct VM profile, lacking the CST I (dominated by L. crispatus) type, and the proportion of high viral load samples increased with the severity of cervical lesions. In contrast, no clear pattern was observed in the other HR-HPVs group. This suggests that persistent HPV 16/18 infections may induce vaginal microbiome dysbiosis, potentially exacerbating cervical lesion progression, consistent with prior studies reporting that the absence of CST I reflects a reduction in protective lactobacilli and promotes carcinogenesis⁴. However, the lack of a pattern in the other HR-HPVs group suggests viral load effects on the VM vary by HPV type, warranting further type-specific kingdom and intra-kingdom correlations in each group. **e** The number of bacterialbacterial (BB), fungal-fungal (FF), and viral-viral (VV) correlations, which signify inter-kingdom correlations, as well as bacterial-fungal (BF), bacterial-viral (BV), and fungal-viral (FV) correlations, which signify intra-kingdom correlations.

research. While this approach does not directly confirm infection status, it provides a practical framework for indirectly inferring infection type using our current dataset.

In this study, strict inclusion criteria were applied to analyze the differences in VM between HPV 16/18 and other HR-HPVs, which inevitably resulted in a limited number of study participants. To address this limitation, we performed an additional validation analysis using the VM dataset from a Swedish cohort published by Norenhag et al.³⁴, obtained from the NCBI Sequence Read Archive (accession number: PRJEB72779). The dataset was analyzed using the same analytical methods applied in our study to ensure consistency and comparability. For this additional analysis, we selected participants from the HSIL group (CIN2 and CIN3) who met our study criteria and compared the vaginal microbiota composition between HPV 16 and other HR-HPV infection groups (Supplementary fig. 11). The results showed consistent patterns with our study findings, confirming that HPV 16 infection is more strongly associated with vaginal microbiota dysbiosis than other HR-HPV infections. While we acknowledge that the Swedish cohort may be influenced by geographic and demographic differences and that the inclusion criteria may not be fully identical to our study, the fact that similar trends were replicated in an independent cohort suggests a consistent association between HPV infection types and vaginal microbiota composition.

The bacterial species associated with different HR-HPV subtypes varied across pathological disease groups. In group 2, women infected with other HR-HPVs exhibited a higher prevalence of *L. gasseri*, *L. paragasseri*, and *Lactobacillus* sp. JM1 (similar to *L. paragasseri* based on GTDB-Tk) compared to those infected with HPV 16/18. Although a *L. gasseri*-dominant vaginal community (CST II) is rare in Korean cohorts, the fact that *L. gasseri* demonstrates the quickest remission rate among various CSTs. Therefore, our findings indicate that the vaginal microenvironment in women infected with other HR-HPVs differs from that in women with HPV 16/18, potentially reflecting differences in the clearance dynamics of these infections. This variation in bacterial flora raises significant questions regarding the role of the VM in HR-HPV infections. In group 3, bacterial species associated with HPV 16/18 were substantially more prevalent. *S. agalactiae*, commonly known as group B *Streptococcus*, has been extensively

studied in the context of maternal and fetal health during pregnancy. However, its increase in aerobic vaginitis could potentially alter the vaginal environment³⁵. *F. vaginae* and *S. vaginalis* are strongly associated with bacterial vaginosis, characterized by a decrease in normal lactobacillus bacteria and an increase in anaerobic bacteria³⁶. Notably, *Sneathia* spp. have been detected three times more frequently in women with HR-HPV infections, elucidating them as a microbiological marker of HR-HPV infections²¹. These findings illustrate the complex interactions between viruses and bacteria in the cervical microenvironment and underscore the necessity for comprehensive investigations into the interactions between HR-HPV types and the VM.

CST IV typically has fewer Lactobacillus spp. and more anaerobic bacteria such as *G. vaginalis* and *F. vaginae*, typically being considered an unhealthy state. However, a study of Gajer et al. revealed that CST IV is often found in healthy Hispanic and Black women, suggesting it does not necessarily reflect dysbiosis¹¹. They also found that VM is not static and can transition to different states over short periods. Additionally, since VM dominated by a single *Lactobacillus* spp. also exhibit significant intra-species genetic variations, being "healthy" cannot be defined solely by a low pH and *Lactobacillus*-dominated community³⁷. Therefore, assessing health based only on microbial community composition (such as CST) is insufficient, and it is important to understand the specific functions of vaginal microbes to address this issue.

We comprehensively evaluated the functional differences in the VM across according to HR-HPV types. Differences in pathway enrichment between groups may be associated with the underlying mechanisms of HPV infection, potentially influencing the clinical outcomes associated with HPV-related diseases. Notably, distinct functional compositions were prominent in group 3 of our cohort. Contrary to differences at the species level, a significantly greater abundance of functional genes was observed in microbiomes associated with other HR-HPVs. This suggests a complex interplay between the microbiome's functional state and the progression of HPV-related conditions, warranting further investigation into how these functional variances contribute to the pathogenesis and outcomes of HPV infections. Furthermore, our findings indicate that the microbial communities associated with HPV 16/18 are enriched in pathways that modulate immune responses and potentially promote malignant transformation, while the microbiomes associated with other HR-HPVs are characterized by a predominance of pathways that may influence broader aspects of cellular metabolism and hormonal signaling. This variation underscores the potential for these functional differences to impact the clinical outcomes of HPV-related diseases. The activation of the JAK-STAT signaling pathway is fundamental to cell proliferation, invasion, survival, inflammation, and immune regulation. This pathway is associated with carcinogenesis and metastasis, suggesting that HPV 16/18 may promote more aggressive disease phenotypes through this mechanism³⁸. The oncoproteins E6 and E7 expressed by HPV 16/18 can disrupt the normal regulatory mechanisms of the JAK-STAT signaling pathway, leading to persistent activation that promotes cellular malignancy³⁹. Additionally, pathways related to malignant transformation, such as those in endometrial cancer, melanoma, and non-small cell lung cancer, are up-regulated in HPV 16/18. In contrast, in other high-risk HPVs, the activation of cell adhesion pathways is increased, which can limit interactions between host cells and the virus, thereby inhibiting viral invasion and spread. The E6 and E7 oncoprotein of HPV plays a role in inhibiting cell adhesion pathways by degrading the p53 protein, disrupting cell cycle control and apoptosis, and weakening interactions with the ECM, thus reducing cell-to-cell adhesion^{40,41}. The ECM is a meshwork of extracellular proteins that maintain a barrier between cells and the external environment and provide structural support. Strengthening ECM-receptor interactions is important for inhibiting viral spread. It is important to note the role of these pathways in other high-risk HPVs in potentially inhibiting viral activity^{42,43}. However, this is limited to the function of microorganisms, and extensive additional research is needed to determine if these functions also impact the host.

Recent studies have revealed that enzymes derived from vaginal microbiota are involved in the metabolism of glycogen and mucin produced by the host. Glycogen, primarily produced by vaginal epithelial cells, serves as a crucial carbon source for vaginal microbes⁴⁴. The activity of glycogendegrading enzymes is essential for maintaining the balance of the VM and providing an energy source in the vaginal environment^{45,46}. Additionally, mucin is a significant glycoprotein in cervical mucus that protects against pathogens and helps maintain reproductive health. Mucin-degrading enzymes break down cervical mucus, weakening the protective barrier and potentially facilitating viral penetration and spread^{47,48}. Our study found that these enzymes are generally more observed in HPV 16/18 compared to other HR-HPVs, providing evidence for why HPV 16/18 exhibits more aggressive clinical phenotypes. France et al. emphasized through a metatranscriptomic study by CST that the degree and pattern of mucin degradation could vary depending on the community composition⁴⁹. This variation can have significant implications for fertility and susceptibility to sexually transmitted infections. Therefore, the increased activity of these enzymes in HPV 16/18 suggests that these viruses can dynamically alter the vaginal environment, increasing susceptibility to infection, replication, and persistence of the virus.

Our research builds on the premise that microbial genomes exhibit significant variation within the same species owing to adaptation to diverse environments. Previous studies utilizing extensive metagenomic data have suggested that evolutionary processes acting on the VM might adversely influence pregnancy outcomes⁵⁰. Notably, multiple genotypes of G. vaginalis have been identified at the species level, suggesting a potential link between microbial genomic variations and host phenotypic outcomes, which may impact pregnancy outcomes. In this study, we identified four distinct genotypes of G. vaginalis (referred to as Bifidobacterium vaginale in the GTDB), underscoring the genetic diversity within this species (Supplementary fig. 12). This diversity in genetic types may contribute to differential susceptibility to HR-HPV types. Furthermore, our assembly of 137 MAGs not only includes Lactobacillus but also encompasses a broad spectrum of vaginal bacteria such as Blautia, Finegoldia, Peptoniphilus, Dialister, and Streptococcus. Our analysis of MAG distribution across disease severity groups revealed that HPV 16/18 infections are associated with a significantly different bacterial population compared to other HR-HPV infections, with a particularly high prevalence of taxa in group 3. This finding underscores the intricate interactions between the virus and bacterial factors within the cervical microenvironment and highlights the need for comprehensive studies to explore the interactions between HR-HPV types and the VM. Such studies are essential to understanding how intraspecies genetic diversity at the microbial level may correlate with changes in the vaginal environment induced by different HR-HPV types.

The VM comprises a complex network of bacteria, fungi, viruses, archaea, and protozoa^{51,52}. While previous studies primarily focused on the relationship between vaginal bacteria and women's health, our research utilized metagenomic data to explore the interactions among bacteria, fungi, and viruses within the VM. We observed an increase in the number of nodes and edges in groups 1 and 3, indicating that HPV 16/18 infection may influence the dynamics of the entire VM, leading to complex interactions among various microbes. The presence of HPV 16/18 was associated with an overall increase in bacterial-fungal correlations, suggesting potential synergistic or interdependent relationships between these microbial components. A recent study on mixed vaginitis (bacterial vaginosis plus vulvovaginal candidiasis) found that key bacteria such as Gardnerella, Atopobium, and Lactobacillus play significant roles, and that the VM undergoes unique changes following treatment, distinct from those seen in single vaginitis⁵³. This study highlighted the importance of bacterial-fungal interactions in driving the dynamic changes within the VM. Similarly, our research indicates that the increased bacterial-fungal correlations in HPV 16/18 infections underscore the need for further studies to understand the mechanisms driving these microbial interactions and their potential links to cervical carcinogenesis.

The cross-sectional design of our study limited our ability to infer causality between risks of HPV 16/18, cervical lesions, and VM composition. The absence of HR-HPV negative controls precluded comparative analysis with uninfected populations, potentially distorting our understanding of the microbiome's impact on HR-HPV pathogenicity. Moreover, this study has a limitation in that it cannot directly track the persistence or clearance of HPV infections over time. The heterogeneous nature of our comparison groups might have led to an underestimation of the associations between aberrant VM and HPV 16/18 infection. Additionally, it remains unclear whether HPV disrupts the vaginal environment or if pre-existing dysbiosis facilitates HPV persistence. These limitations highlight the need for longitudinal studies including HR-HPV negative controls to better understand the complex interplay between HPV 16/18 infection, cervical carcinogenesis, and the VM.

As HPV vaccination programs continue to expand globally, it is important to consider the potential impact of vaccination on the VM. Although our study did not include information on participants' vaccination status, previous research suggests that HPV vaccination itself has minimal influence on the composition of VM. For instance, Cheng et al.²³ reported negligible microbiome changes following quadrivalent HPV vaccination. In contrast, HPV infection is known to induce microbial alterations, including a reduction in Lactobacillus and enrichment of anaerobic taxa such as *Gardnerella* and *Sneathia*. These patterns may also be observed in vaccinated individuals who experience breakthrough infections or acquire HPV types not covered by the vaccine. Therefore, our findings may be partially applicable to vaccinated populations, and future studies should incorporate vaccination status when investigating microbiome–HPV interactions.

Our results collectively suggest the dynamic and complex nature of the VM in the context of HR-HPV associated cervical tumors. Our findings show that microbial community patterns differ between HPV 16/18 and other HR-HPVs depending on the stage of cervical lesions. Notably, differences in VM composition in the early stages of cervical lesions may be associated with the oncogenic potential of HR-HPV types. As cervical cancer progresses, these microbial differences become more pronounced. These findings suggest a possible link between VM composition and HPV-related cervical tumors, emphasizing the need for further research to clarify the mechanisms underlying these associations. Understanding these dynamics holds significant potential for developing innovative diagnostic and treatment strategies for women's health issues.

Methods Study population

The ethical approval for this study was obtained from the Institutional Review Board of Kyungpook National University Chilgok Hospital (KNUCH 2023-08-054-001). This study included patients with histologically validated cervical lesions who presented to the Department of Obstetrics and Gynecology at Kyungpook National University Chilgok Hospital (Daegu, Republic of Korea). We recruited a total of 68 women and categorized them into three groups: group 1 consisted of 23 participants, including 11 with negative malignancy findings (normal) and 12 with lowgrade squamous intraepithelial lesions (LSIL); group 2 comprised 23 patients with high-grade squamous intraepithelial lesions (HSIL); and group 3 included 22 patients with invasive cervical cancer (ICC). Obstetric experts collected cervicovaginal specimens from the lower third of the vagina of all participants using sterile cotton swabs. The swab samples were immediately placed in DNase-, RNase-, and pyrogen-free conical tubes and stored at -80°C until further experiment. Basic clinical characteristics and detailed medical histories were documented during the participants' visits.

HPV genotyping

Cervicovaginal swab specimens were subjected to HPV genotyping using the Anyplex II HPV 28 Assay Kit (Seegene, Seoul, Republic of Korea) following the manufacturer's instructions. The assay detected a total of 28 HPV types, which included both HR-HPVs (16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 69, 73, and 82) and low-risk HPV types (6, 11, 40, 42, 43, 44, 54, 61, and 70). Based on the HPV genotyping results, the participants were subdivided into two groups: those infected with HPV 16 and 18 (HPV 16/18) and those with other HR-HPV types (other HR-HPVs). The final subdivided groups are shown in Table 1.

Microbial DNA extraction and shotgun metagenomic sequencing

Total microbial DNA was extracted from each cervicovaginal swab sample using the DNeasy PowerSoil Pro DNA Kit (QIAGEN, Hilden, Germany) in accordance with the manufacturer's instructions. The quality and quantity of the extracted DNA were assessed using the NanoDrop One Microvolume UV-Vis Spectrophotometer (Thermo Fisher, Waltham, MA, USA) and the Qubit Flex Fluorometer (Thermo Fisher, Waltham, MA, USA), respectively. High-quality DNA was used for library preparation with the DNBSEQ-G400RS High-Throughput Sequencing FCL PE100 Kit (MGI Tech, Shenzhen, China). Shotgun metagenome sequencing was performed on the DNBSEQ-G400RS platform (MGI Tech, Shenzhen, China) at the Kyungpook National University NGS Core Facility (Daegu, Republic of Korea).

Bioinformatic analysis

From the raw sequencing data, adapter sequences were trimmed and lowquality reads were removed using SOAPnuke v2.1.754. Subsequently, reads were aligned to the reference human genome GRCh38 using Bowtie2 v2.5.055. The aligned reads were then excluded from further analysis. The quality-controlled reads were employed for taxonomic classification using Kraken2 v2.1.2⁵⁶ with a custom database that included bacterial, fungal, and viral genomes from the National Center for Biotechnology Information (NCBI). Microbial abundance was estimated using Bracken v2.8⁵⁷ at each taxonomic rank based on the Kraken2 output. The abundance data derived from Kraken2 was used to assign vaginal CSTs employing a nearestcentroid-based algorithm using the VALENCIA tool¹⁰. To validate our taxonomic classification approach, we compared our dataset against the recently published Human Vaginal Microbiome Genome Collection (VMGC)¹⁹, which represents a comprehensive genome database of VM. To detect HPV genotype in metagenome data, we constructed a custom database by integrating the RefSeq Microbial Genomes (archaea, bacteria, fungi and virus) from NCBI with the Papilloma Virus Episteme (PaVE) database³¹. Using this custom database, we aligned the sequencing reads with Bowtie2 and calculated the reads per kilobase of exon per million reads mapped (RPKM) for HPV types using CoverM v0.6.1 (https://github.com/ wwood/CoverM).

To identify functional genes in each sample, we assembled the qualified reads. Microbial genome assembly was performed using MEGAHIT v1.1.3⁵⁸. Subsequent to assembly, protein-coding sequences (CDS) were predicted using Prodigal v2.6.3⁵⁹. Non-redundant gene categories were then clustered using CD-HIT v4.8.1⁶⁰ with an identity threshold exceeding 95%. The RPKM for each predicted gene was calculated using CoverM v0.6.1 (https://github.com/wwood/CoverM). The predicted CDS were functionally annotated using the Kyoto Encyclopedia of Genes and Genomes (KEGG) via GhostKOALA⁶¹. Carbohydrate active enzymes (CAZymes) were annotated using run_dbcan, a standalone tool of dbcan3 v4.1.4⁶², for the analysis of glycogen and mucin degradation enzymes.

The qualified reads from all samples were co-assembled using MEGAHIT, and contigs were clustered into metagenomic bins using the metaWRAP v1.3⁶³ binning module with the parameters -maxbin2, -concoct, and -metabat2. These bins were refined and reassembled to combine the outputs from three binning software applications (completeness >50%, contamination <10%). The metagenome-assembled genome (MAG) bins were taxonomically classified using GTDB-Tk v2.3.2, referencing the Genome Database Taxonomy (GTDB)⁶⁴ release 207 based on Average Nucleotide Identity (ANI). A phylogenetic tree of the MAGs was subsequently constructed using iTOL⁶⁵. The relative abundance of the MAGs was quantified using the CoverM pipeline, based on the coverage of mapped reads in "genome" mode with the "-m rpkm" option. Additionally, like short-read analysis, we compared our MAGs using the VMGC reference based on Average Nucleotide Identity (ANI)⁶⁶. MAGs with ANI \geq 83% were classified as belonging to the same genus, while those with ANI \geq 95% were classified as the same species.

Statistical analysis

All statistical analyses and visualizations were performed using R software v4.2.2 (http://www.r-project.org/). The non-metric multidimensional scaling (NMDS) analysis of VM based on the Bray–Curtis dissimilarity distance matrix was conducted using the vegan v2.6-4 and phyloseq v1.42.0 packages in R. Permutational multivariate analysis of variance (PERMANOVA) was performed to test the statistical significance of differences among groups. We identified differentially abundant taxa and microbial functions using the DESeq2 v1.38.3 package in R, based on fold change values. To analyze multi-kingdom microbial networks—encompassing bacteria, fungi, and viruses—we filtered microbial species by a prevalence of >50% and an abundance >0.01% in each group. Correlation network analysis was conducted using the igraph v1.5.1 R package in R. Significant correlations were determined using Spearman's correlation coefficient (q < 0.05, r > 0.85).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The shotgun metagenome sequencing raw data (short-read archives, SRA) are available in the GenBank repository with bioproject accession number PRJNA1054643.

Code availability

The underlying code for this study is not publicly available but may be made available to qualified researchers on reasonable request from the corresponding author.

Received: 31 May 2024; Accepted: 2 June 2025; Published online: 12 June 2025

References

- Forman, D. et al. Global burden of human papillomavirus and related diseases. Vaccine 30, F12–F23 (2012).
- Lehtinen, M., Baussano, I., Paavonen, J., Vänskä, S. & Dillner, J. Eradication of human papillomavirus and elimination of HPV-related diseases-scientific basis for global public health policies. *Expert Rev. Vaccines* 18, 153–160 (2019).
- Kidd, L. C. et al. Relationship between human papillomavirus and penile cancer—implications for prevention and treatment. *Transl. Androl. Urol.* 6, 791 (2017).
- Usyk, M. et al. Cervicovaginal microbiome and natural history of HPV in a longitudinal study. *PLoS Pathog.* 16, e1008376 (2020).
- Bosch, F. X. et al. Comprehensive control of human papillomavirus infections and related diseases. *Vaccine* **31**, H1–H31 (2013).
- de Sanjose, S. et al. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. *Lancet Oncol.* 11, 1048–1056 (2010).
- Aydoğmuş, H. & Aydoğmuş, S. Comparison of colposcopic biopsy results of patients who have cytomorphological normal but HPV 16-18 or other high-risk HPV subtypes positive. *Asian Pac. J. Cancer Prev.* 20, 417 (2019).
- Shannon, B. et al. Association of HPV infection and clearance with cervicovaginal immunology and the vaginal microbiota. *Mucosal Immunol.* **10**, 1310–1319 (2017).
- 9. Zou, J. et al. Variants in human papillomavirus receptor and associated genes are associated with type-specific HPV infection and lesion progression of the cervix. *Oncotarget* **7**, 40135 (2016).

- France, M. T. et al. VALENCIA: a nearest centroid classification method for vaginal microbial communities based on composition. *Microbiome* 8, 1–15 (2020).
- 11. Gajer, P. et al. Temporal dynamics of the human vaginal microbiota. *Sci. Transl. Med.* **4**, 132ra152–132ra152 (2012).
- Anahtar, M. N., Gootenberg, D. B., Mitchell, C. M. & Kwon, D. S. Cervicovaginal microbiota and reproductive health: the virtue of simplicity. *Cell Host Microbe* 23, 159–168 (2018).
- Bik, E. M. et al. A novel sequencing-based vaginal health assay combining self-sampling, HPV detection and genotyping, STI detection, and vaginal microbiome analysis. *PLoS ONE* 14, e0215945 (2019).
- 14. Caselli, E. et al. Vaginal microbiota and cytokine microenvironment in HPV clearance/persistence in women surgically treated for cervical intraepithelial neoplasia: an observational prospective study. *Front. cell. infect. microbiol.* **10**, 540900 (2020).
- Molina, M. A., Leenders, W. P., Huynen, M. A., Melchers, W. J. & Andralojc, K. M. Temporal composition of the cervicovaginal microbiome associates with hrHPV infection outcomes in a longitudinal study. *bioRxiv*, 2023.2006. 2002.543506 (2023).
- Sofou, E. et al. High risk HPV-positive women cervicovaginal microbial profiles in a Greek cohort: a retrospective analysis of the GRECOSELF study. *Front. microbiol.* 14, 1292230 (2023).
- Vargas-Robles, D. et al. The cervical microbiota of Hispanics living in Puerto Rico is nonoptimal regardless of HPV status. *mSystems* 8, e00357–00323 (2023).
- Sasivimolrattana, T. et al. Cervical microbiome in women infected with HPV16 and high-risk HPVs. *Int. J. Environ. Res. Public Health* 19, 14716 (2022).
- 19. Huang, L. et al. A multi-kingdom collection of 33,804 reference genomes for the human vaginal microbiome. *Nat. Microbiol.*, 1–16 (2024).
- Bowers, R. M. et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* 35, 725–731 (2017).
- Lee, J. E. et al. Association of the vaginal microbiota with human papillomavirus infection in a Korean twin cohort. *PLoS ONE* 8, e63514 (2013).
- 22. Mitra, A. et al. Cervical intraepithelial neoplasia disease progression is associated with increased vaginal microbiome diversity. *Sci. Rep.* **5**, 16865 (2015).
- Cheng, L. et al. Vaginal microbiota and human papillomavirus infection among young Swedish women. *NPJ Biofilms Microbiomes* 6, 39 (2020).
- 24. Lebeau, A. et al. HPV infection alters vaginal microbiome through down-regulating host mucosal innate peptides used by Lactobacilli as amino acid sources. *Nat. Commun.* **13**, 1076 (2022).
- Kim, M. J., Lee, S., Kwon, M. Y. & Kim, M. Clinical significance of composition and functional diversity of the vaginal microbiome in recurrent vaginitis. *Front. microbiol.* **13**, 851670 (2022).
- Lee, Y. H. et al. Vaginal microbiome-based bacterial signatures for predicting the severity of cervical intraepithelial neoplasia. *Diagnostics* **10**, 1013 (2020).
- Brotman, R. M. et al. Interplay between the temporal dynamics of the vaginal microbiota and human papillomavirus detection. *J. Infect. Dis.* 210, 1723–1733 (2014).
- Mitra, A. et al. The vaginal microbiota associates with the regression of untreated cervical intraepithelial neoplasia 2 lesions. *Nat. Commun.* 11, 1999 https://doi.org/10.1038/s41467-020-15856-y (2020).
- 29. Sun, N. et al. Comprehensive characterization of microbial community in the female genital tract of reproductive-aged women in china. *Front. cell. infect. microbiol.* **11**, 649067 (2021).
- Audirac-Chalifour, A. et al. Cervical microbiome and cytokine profile at various stages of cervical cancer: a pilot study. *PLoS ONE* 11, e0153274 (2016).

- Van Doorslaer, K. et al. The Papillomavirus Episteme: a major update to the papillomavirus sequence database. *Nucleic Acids Res.* 45, D499–D506 (2017).
- Trevisan, A. et al. Human papillomavirus type 16 viral load measurement as a predictor of infection clearance. *J. Gen. Virol.* 94, 1850–1857 (2013).
- Lu, X., Wang, T., Zhang, Y. & Liu, Y. Analysis of influencing factors of viral load in patients with high-risk human papillomavirus. *Virol. J.* 18, 1–8 (2021).
- Norenhag, J. et al. Compositional and functional differences of the vaginal microbiota of women with and without cervical dysplasia. *Sci. Rep.* 14, 11183 (2024).
- Choi, Y. et al. Updates on Group B Streptococcus infection in the field of obstetrics and gynecology. *Microorganisms* 10, 2398 (2022).
- Wei, Z.-T. et al. Depiction of vaginal microbiota in women with highrisk human papillomavirus infection. *Front. Public Health* 8, 587298 (2021).
- Ma, B. et al. A comprehensive non-redundant gene catalog reveals extensive within-community intraspecies diversity in the human vagina. *Nat. Commun.* **11**, 940 (2020).
- Stine, R. R. & Matunis, E. L. JAK-STAT signaling in stem cells. *Adv. Exp. Med. Biol.* 786, 247–267 (2013).
- Gutiérrez-Hoya, A. & Soto-Cruz, I. Role of the JAK/STAT pathway in cervical cancer: its relationship with HPV E6/E7 oncoproteins. *Cells* 9, 2297 (2020).
- Ben Khalifa, Y. et al. The human papillomavirus E6 oncogene represses a cell adhesion pathway and disrupts focal adhesion through degradation of TAp63β upon transformation. *PLoS Pathog.* 7, e1002256 (2011).
- Yuan, H. et al. Human papillomavirus type 16 oncoprotein E7 suppresses cadherin-mediated cell adhesion via ERK and AP-1 signaling. *Int. J. Oncol.* 35, 309–314 (2009).
- Tomlin, H. & Piccinini, A. M. A complex interplay between the extracellular matrix and the innate immune response to microbial pathogens. *Immunology* **155**, 186–201 (2018).
- Maginnis, M. S. Virus-receptor interactions: the key to cellular invasion. J. Mol. Biol. 430, 2590–2611 (2018).
- 44. Ayre, W. B. The glycogen-estrogen relationship in the vaginal tract. *J. Clin. Endocr.* **11**, 103–110 (1951).
- Jenkins, D. J. et al. Bacterial amylases enable glycogen degradation by the vaginal microbiome. *Nat. Microbiol.* 8, 1641–1652 (2023).
- Van Der Veer, C. et al. Comparative genomics of human Lactobacillus crispatus isolates reveals genes for glycosylation and glycogen degradation: implications for in vivo dominance of the vaginal microbiota. *Microbiome* 7, 1–14 (2019).
- Segui Perez, C. et al. Prevotella timonensis degrades the vaginal epithelial glycocalyx through high fucosidase and sialidase activities. *bioRxiv*, 2024.2001. 2009.574844 (2024).
- Vagios, S. & Mitchell, C. M. Mutual preservation: a review of interactions between cervicovaginal mucus and microbiota. *Front. cell. infect. microbiol.* **11**, 676114 (2021).
- France, M. T. et al. Insight into the ecology of vaginal bacteria through integrative analyses of metagenomic and metatranscriptomic data. *Genome Biol.* 23, 66 (2022).
- 50. Liao, J. et al. Microdiversity of the vaginal microbiome is associated with preterm birth. *Nat. Commun.* **14**, 4997 (2023).
- 51. Ursell, L. K., Metcalf, J. L., Parfrey, L. W. & Knight, R. Defining the human microbiome. *Nutr. Rev.* **70**, S38–S44 (2012).
- Dekaboruah, E., Suryavanshi, M. V., Chettri, D. & Verma, A. K. Human microbiome: an academic update on human body site specific surveillance and its possible role. *Arch. Microbiol.* **202**, 2147–2167 (2020).
- 53. Xiao, B., Qin, H., Mi, L. & Zhang, D. Correlation analysis of vaginal microbiome changes and bacterial vaginosis plus vulvovaginal

candidiasis mixed vaginitis prognosis. *Front. cell. infect. microbiol.* **12**, 860589 (2022).

- 54. Chen, Y. et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaScience* **7**, gix120 (2018).
- 55. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
- Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20, 1–13 (2019).
- Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* 3, e104 (2017).
- Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
- 59. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf.* **11**, 1–11 (2010).
- Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152 (2012).
- Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731 (2016).
- 62. Zheng, J. et al. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Res.* **51**, W115–W121 (2023).
- Uritskiy, G. V., DiRuggiero, J. & Taylor, J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6, 1–13 (2018).
- Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 38, 5315–5316 (2022).
- Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296 (2021).
- Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9, 5114 (2018).

Acknowledgements

This research was supported by a grant of the Korea Health Technology R&D Project though the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: RS-2023-KH135444), Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(RS-2023-00274032), and biological materials Specialized Graduate Program through the Korea Environmental Industry & Technology Institute (KEITI) funded by the Ministry of Environment (MOE). We would like to thank KNU NGS Core Facility for the sequencing using DNBSEQ-G400RS platform.

Author contributions

J.H.S. and G.O.C. designed and managed the study; Y.S.C., S.Y.J., N.J.Y.P., J.Y.P. and G.O.C. were responsible for clinical sampling and patient information collection. D.R.J. and M.J. performed bacterial DNA extraction and sequencing. D.R.J., Y.S.C., and M.J. performed the data analysis and statistical analysis, and I.S. checked the analyses. D.R.J. and M.J. drafted manuscript, and V.S., H.S.H., J.H.S. and G.O.C. revised the manuscript. All authors have read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

The study was approved by the Institutional Review Board (IRB) of Kyungpook National University Chilgok Hospital (IRB: KNUCH 2023-08-054-001). All participants signed a consent form.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41522-025-00747-1.

Correspondence and requests for materials should be addressed to Jae-Ho Shin or Gun Oh Chong.

Reprints and permissions information is available at

http://www.nature.com/reprints

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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit http://creativecommons.org/licenses/bync-nd/4.0/.

© The Author(s) 2025