



OPEN

## Leveraging existing 16S rRNA gene surveys to decipher microbial signatures and dysbiosis in cervical carcinogenesis

Xiaoxiao Li<sup>1,3</sup>, Fenfen Xiang<sup>1,3</sup>, Tong Liu<sup>2</sup>, Zixi Chen<sup>1</sup>, Mengzhe Zhang<sup>1</sup>, Jinpeng Li<sup>1</sup>, Xiangdong Kang<sup>1✉</sup> & Rong Wu<sup>1✉</sup>

The presence of dysbiotic cervicovaginal microbiota has been observed to be linked to the persistent development of cervical carcinogenesis mediated by the human papillomavirus (HPV). Nevertheless, the characteristics of the cervical microbiome in individuals diagnosed with cervical cancer (CC) are still not well understood. Comprehensive analysis was conducted by re-analyzing the cervical 16S rRNA sequencing datasets of a total of 507 samples from six previously published studies. We observed significant alpha and beta diversity differences in between CC, cervical intraepithelial neoplasia (CIN) and normal controls (NC), but not between HPV and NC in the combined dataset. Meta-analysis revealed that opportunistic pernicious microbes *Streptococcus*, *Fusobacterium*, *Pseudomonas* and *Anaerococcus* were enriched in CC, while *Lactobacillus* was depleted compared to NC. Members of *Gardnerella*, *Sneathia*, *Pseudomonas*, and *Fannyhessea* have significantly increased relative abundance compared to other bacteria in the CIN group. Five newly identified bacterial genera were found to differentiate CC from NC, with an area under the curve (AUC) of 0.8947. Moreover, co-occurrence network analysis showed that the most commonly encountered *Lactobacillus* was strongly negatively correlated with *Prevotella*. Overall, our study identified a set of potential biomarkers for CC from samples across different geographic regions. Our meta-analysis provided significant insights into the characteristics of dysbiotic cervicovaginal microbiota undergoing CC, which may lead to the development of noninvasive CC diagnostic tools and therapeutic interventions.

**Keywords** Microbiota, HPV, Cervical cancer, 16S rRNA sequencing, Biomarkers

CC is the fourth most common malignant tumor in women, but the incidence and mortality rank second place in countries with a low development index<sup>1</sup>. The key etiological factor in the development of CIN and CC is well recognised to be persistent infection with high-risk HPV. Nevertheless, it is widely recognised that HPV alone is insufficient to cause cervical malignant transformation. Numerous factors have been linked to the occurrence of CIN, and it has been suggested that the composition of cervical microbiota has a significant role in the progression of HPV infection, ultimately leading to the formation of CIN or CC<sup>2–4</sup>. The potential impact of this phenomenon on the development of CC lies in its ability to modify various aspects of the host's inflammatory, genomic, and metabolic processes<sup>5,6</sup>.

An optimal cervicovaginal status was commonly associated with low microbial diversity and prevalence of *Lactobacillus*. *Lactobacillus* spp., which produce lactic acid and bacteriocin, were the key microbes in healthy women, participating in creating the stability of the cervicovaginal microbial composition and maintaining a low pH environment<sup>7,8</sup>. With the development of next-generation sequencing technology, increasing evidence has shown that cervical microbiota dysbiosis, as an important environmental factor, may lead to bacterial vaginosis (BV)<sup>9</sup>, sexually transmitted infections (STIs)<sup>10</sup>, HPV infection and the subsequent development of cervical lesions<sup>11–16</sup>. There were five community state types (CSTs). CST I, II, III, and V were dominated by *Lactobacillus crispatus* (*L. crispatus*), *Lactobacillus gasseri* (*L. gasseri*), *Lactobacillus iners* (*L. iners*) and *Lactobacillus jensenii* (*L. jensenii*) respectively. CST IV was mainly composed by a high abundance of anaerobic bacteria<sup>17</sup>. Prior

<sup>1</sup>Department of Laboratory Medicine, Putuo Hospital, Shanghai University of Traditional Chinese Medicine, 164 Lanxi Road, Shanghai 200062, China. <sup>2</sup>Department of Molecular Science, Uppsala Biocenter, Swedish University of Agricultural Science, Uppsala, Sweden. <sup>3</sup>These authors contributed equally: Xiaoxiao Li and Fenfen Xiang. ✉email: xd\_kang@163.com; rong701@126.com

research has indicated that *L. crispatus* is more prevalent among women who do not have HPV infection or cancer lesions, but *L. iners* and non-Lactobacillus species are more frequently observed in women with HPV infection and patients who have cancer lesions<sup>18–20</sup>. Numerous research have been conducted to investigate the cervical microbiota and its association with cervical lesions in various populations. However, there have been conflicting findings regarding microbial distinctions between individuals with and without cervical lesions. For example, some taxa, including *Aerococcus*, *Coriobacteria* and *Fannyhessea* were reported to be enriched in CIN<sup>21</sup>. Whereas, Audirac-Chalifour et al.<sup>22</sup> found that *Sneathia* and *Fusobacterium* were the predominant taxa in CIN and CC, respectively. The key issue related to the differences in cervical microbiota between cervical lesions and health controls is the lack of apparent reproducibility in different studies when identifying the microbiome characteristics of cervical lesions. Besides, the development of non-invasive and sensitive early diagnosis tests for CC and CIN based on cervical microbiota is meaningful.

Variation in study objectives and analytical processing can influence findings. Herein, we systematically reviewed the literature on cervical carcinogenesis microbiota, and used a consistent analytical approach on pooled 16S rRNA gene raw sequences from 6 studies to identify CC-associated microbiota. We sought to determine the differences in alpha-diversity, beta-diversity, microbial compositions and taxonomic alterations across the stages of CC development. Moreover, the bacterial biomarkers for classifications of different disease groups and bacterial taxa correlation were also examined.

## Materials and methods

### Dataset acquisition and study inclusion criteria

A systematic searches for public available data were performed on the NCBI Sequence Read Archive (SRA, <http://www.ncbi.nlm.nih.gov/sra>), European Nucleotide Archive (ENA, <http://www.ebi.ac.uk/ena>) and Genome Sequence Archive (GSA, <https://ngdc.cnbc.ac.cn/gsa/>) databases using the search term “cervical cancer”, “HPV”, “cervical intraepithelial neoplasia” and “16S rRNA”, and limiting search results by “Bioproject”. Bioproject accession numbers containing high throughput sequencing reads and associated metadata were collected (Table 1). Then, we retrieved articles from NCBI PubMed or Google Scholar using submitted Bioproject information. The studies incorporated in this meta-analysis were mandated to meet the following criteria: (a) they had to involve samples obtained from the human cervix, (b) they needed to have have sequenced by NGS for 16S rRNA gene, (c) they had non-use of douches in previous days of sampling, (d) not having records of antibiotic or antifungal or antiviral usage within the previous three months of sampling, (e) to have associated metadata, sequencing data and barcodes from the public database or provided by the authors until November 2022 upon request by emails. In order to facilitate comparisons of data across studies, a preformatted metadata file including sample ID, study ID, sequencing type, country and diagnosis was collected from public database or request from authors. Finally, we obtained 16S rRNA sequence data and sufficient metadata (containing information differentiating samples by diagnosis) from 6 studies for subsequent analyses<sup>22–27</sup>. An additional 23 studies related to cervical microbiota of cervical lesions were excluded due to incomplete information on sequences, metadata or barcodes or the same sequences<sup>15,16,21,28–47</sup>. The downloaded datasets were grouped according to the following stages: NC, HPV, CIN and CC.

### Sequence processing

The raw data from all datasets were downloaded in the sequence read archive (SRA) and converted into fastq format using SRA Toolkit. Next, all downloaded 16S rRNA gene sequences were processed using the open-source DADA2 for quality control and denoising using parametric error model in R (version 1.16)<sup>48</sup>. Filtering, learning errors, dereplication, amplicon sequence variant (ASV) inference and chimera removal were conducted for each study. Subsequently, taxonomy was assigned using Silva 138 rRNA database using the Naïve Bayesian Classifier algorithm default in DADA2<sup>49</sup>. The merging of paired-end reads was conducted when the overlapping region between the reads exceeded a length of 20 base pairs following the truncation process. Otherwise, only forward reads were used for the following analysis. The within-sample (alpha) diversity metrics including evenness and Shannon diversity were calculated based on the ASV table in each sample in R using vegan package. The between-sample (beta) diversity was assessed based on Bray–Curtis distance. Principal coordinate analysis (PCoA) was used to visualize the distance using PAST (version 3.0) software, and the differences between groups were determined using permutational multivariate analysis of variance (PERMANOVA) with 999 permutations.

### Statistical analysis

Significantly altered taxa among different stage groups were determined by the linear discriminant analysis (LDA) effect size (LEfSe) method with a cutoff LDA score > 3 and *p*-value < 0.05. STAMP software (version 2.1.3) with extend error bar method was used to explore differentially abundant bacteria at genus level between two groups (i.e. NC vs CC), and the significance criteria were *p*-value < 0.05 using two-sided Fishers exact test. Afterward, logistic regression models were built using the selected bacterial genera biomarker with a backward stepwise selection algorithm using leaps package. The receiver operating characteristic (ROC) analysis was used to illustrate the performances of classification models. The statistical analyses were performed using GraphPad Prism software (version 9.3). Pairwise and multiple groups comparisons were performed using the two-sided Wilcoxon rank-sum test and the Kruskal–Wallis test, respectively. Spearman correlation test was used for correlation analyses for the representative (top 15) genera and displayed using R software (version 4.1.2) with corplot, ggplot2 and ggpubr packages. Bubble plots were made using the ggplot2 and reshape 2 packages, and a heatmap plot was conducted with the pheatmap package in R.

Data set	Bioproject	Country	*Source	NC	HPV	LSIL	HSIL	CC	Instrument	Region
Cheng	PRJEB34755	Sweden	Author	113	144	0	0	0	MiSeq	V3-V4
Xie	PRJNA595048	China	SRA	25	0	1	18	18	NovaSeq	V4
Audirac-Chalifour	PRJNA308947	Mexico	SRA	7	10	0	4	8	454	V3-V4
Ilhan	PRJNA518153	United States	SRA	18	11	12	27	10	MiSeq	V4
Kang	PRJNA692362	Korea	Author	7	0	0	8	8	PGM	V3
Wei	PRJNA551647	China	SRA	30	30	0	0	0	MiSeq	V3-V4
Mei	PRJCA007388	China	NA	42	28	0	0	0	MiSeq	V3-V4
Chen	PRJNA415526	China	NA	68	78	51	23	9	MiSeq	V3-V4
Onywera	PRJNA473351	South Africa	NA	56	29	0	0	0	MiSeq	V3-V4
Borgogna	PRJNA391039	United States	NA	13	26	0	0	0	454	V1-V3
McKee	PRJNA622998	United States	NA	89	110	45	64	0	MiSeq	V4
Lin	PRJCA007758	China	NA	60	0	60	0	0	MiSeq	V3-V4
Wang	PRJNA815961	China	NA	14	0	9	29	10	MiSeq	V4
Tango	PRJEB5760	Korea	NA	50	0	0	25	17	454	-
Nieves-Ramírez	PRJNA766648	Mexico	NA	107	0	90	31	0	HiSeq	V3
Mitra	PRJEB7756	England	NA	20	0	52	92	5	MiSeq	V1-V2
Bokulich	PRJNA518153	United States	NA	18	9	10	27	8	MiSeq	V4
Chao	PRJCA005712	China	NA	103	86	0	83	0	HiSeq	V4
Piyathilake	PRJNA392354	United States	NA	0	0	340	90	0	MiSeq	V4
Fang	PRJNA846153	China	NA	20	20	0	0	0	NovaSeq	V3-V4
Zhou	PRJNA548879	China	NA	20	42	0	0	0	HiSeq	V3-V4
Godoy-Vitorino	PRJNA429969	United States	NA	10	52	0	0	0	MiSeq	V4
Paola	PRJEB18720	Italy	NA	17	55	0	0	0	454	V3-V5
Lee	PRJEB3342	Korea	NA	45	23	0	0	0	454	V2-V3
Fan	PRJNA725946	China	NA	54	0	0	0	65	HiSeq	V3-V4
Wang2021	PRJNA687331	China	NA	40	0	0	0	26	MiSeq	V3-V4
Lam	PRJNA431248	United States	NA	0	0	0	0	58	MiSeq	V4
Zhang2018	PRJNA319915	China	NA	64	0	62	40	0	HiSeq	V3-V4
Zhang2022	PRJNA787754	China	NA	113	159	47	37	0	NovaSeq	-

**Table 1.** Summary of 16S-tag studies on cervicovaginal microbiota-related information. NC normal control, HPV HPV infection, CIN cervical intraepithelial neoplasia; Low-grade squamous intraepithelial lesion (women diagnosed with HPV infection and cervical intraepithelial neoplasia 1) [CIN1]; HSIL, high-grade squamous intraepithelial lesion (included CIN2 and CIN3). CC cervical cancer. \*NA indicates studies were not included in the analysis due to the datasets without sufficient metadata, without barcode sequences to splits the datasets, or low sequencing quality.

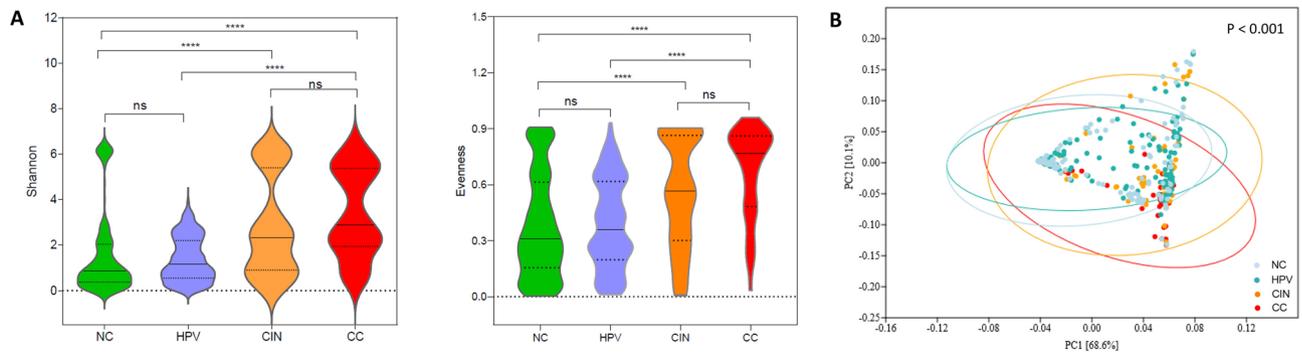
## Results

### Alpha and beta diversity differences

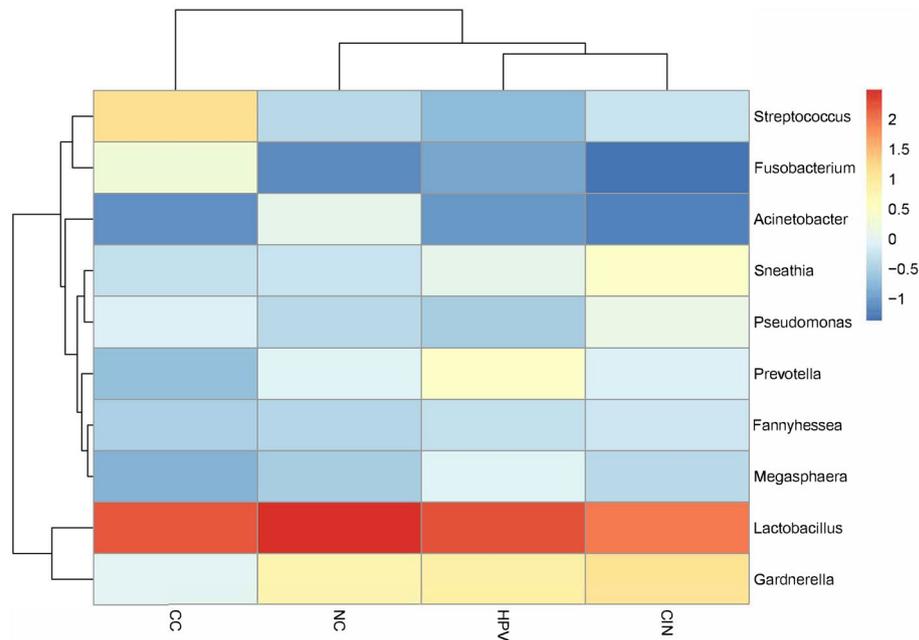
Quality-filtered 16S rRNA gene sequences from 507 cervical epithelial scrapings samples were available for meta-analysis. Sequences were from six independent studies and five countries (Table 1). To compare the alpha diversity between different disease stages, we observed that the NC group had significantly lower Shannon diversity and evenness when compared CIN and CC groups ( $p < 0.0001$ ) (Fig. 1A). There were no significant differences between NC and HPV groups, CIN and CC groups ( $p > 0.05$ ). Moreover, the microbial diversities exhibited an increasing trend with disease progression (Fig. 1A). We next analyzed whether there were differences in the structures of microbial communities associated with different disease stages. Beta diversity was visualized by principal coordinate analysis (PCoA) based on the Bray–Curtis distance. Overall cervical microbial community structures of four disease stages were significantly different (PERMANOVA,  $F = 11.54$ ,  $p < 0.001$ ) (Fig. 1B).

### Characteristics of the core cervicovaginal microbiota

The present study investigated the comprehensive microbial compositions during the course of cervical cancer. The cervical microbiota was shown to be predominantly composed of five bacterial phyla at the phylum level, namely *Firmicutes*, *Actinobacteria*, *Proteobacteria*, *Bacteroidetes*, and *Fusobacteria*. There were 45 genera identified as core microbiota when combining all datasets. Among them, *Lactobacillus*, *Gardnerella*, *Streptococcus*, *Sneathia*, *Prevotella*, *Pseudomonas*, *Fannyhessea*, *Megasphaera*, *Fusobacterium* and *Acinetobacter* were the most prevalent and abundant genera across the datasets. Hierarchical clustering analysis showed that the HPV group and CIN group shared a similar distribution of the top 10 genera (Fig. 2). For *Lactobacillus*, the most common encountered species was *L. iners* (Fig. 3). Moreover, *Lactobacillus* was significantly lower in CIN and CC groups when compared to NC and HPV groups ( $p < 0.001$ ). While, *Streptococcus* and *Fusobacterium* were significantly higher in CIN and CC groups when compared to NC and HPV groups ( $p < 0.05$ ) (Fig. 4).



**Figure 1.** Bacterial diversity for patients in each stage group. (A) Alpha-diversity was estimated by Shannon and evenness indexes in each group. The solid black line indicated the corresponding median value in each group. Pairwise comparisons were performed using Wilcoxon rank-sum test. Asterisks mean differences between the two groups are statistically significant ( $p < 0.0001$ ). ns, no significant difference. (B) Principal coordinate analysis (PCoA) for all included samples based on Bray–Curtis distance.  $p$ -value was estimated by permutational multivariate analysis of variance (PERMANOVA).



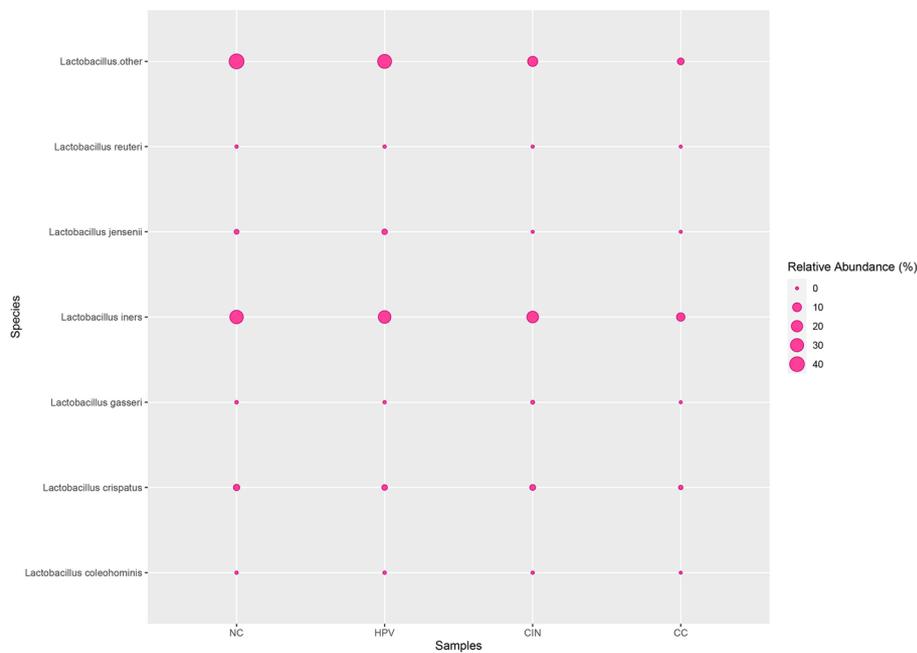
**Figure 2.** Heatmap plot showing the top 10 most abundant bacterial communities in each group at the genus level. The abundance was log-transformed to reduce the skewness of the data.

### Differential taxa across different disease stages

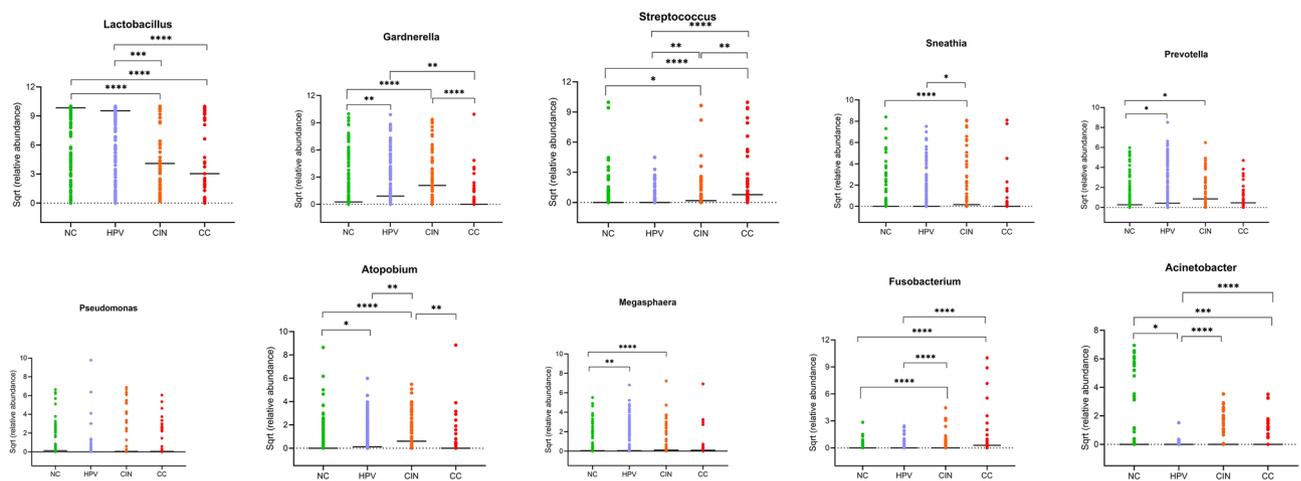
LEfSe was performed to identify the specific taxa with significantly higher abundance among the four groups. As shown in Fig. 5, a total of 59 clades were screened out with a LDA threshold score of 3.0. Class *Bacilli* (including *Lactobacillus* genus) was enriched in the NC group. Genus *Acinetobacter* and *Bacteroides* were also enriched in the NC than in other groups. The CIN group was more highly colonized by class *Actinobacteria* and *Gammaproteobacteria*. As for the CC group community, class *Clostridia* and *Epsilonproteobacteria* had significantly higher relative abundances.

### Bacterial biomarkers for distinguishing CC from NC

Among the ten bacterial genera that significantly differed between the two stages, four genera including *Streptococcus*, *Fusobacterium*, *Pseudomonas* and *Anaerococcus*, were found to be abundant in the CC stage compared to the NC stage. On the other hand, six were depleted including *Lactobacillus*, *Acinetobacter* and *Bacteroides* (Fig. 6A). We assessed the significantly altered bacterial genera for their potential as diagnostic biomarkers for discriminating CC from NC. Four CC-enriched and one CC-depleted were identified as potential biomarkers using the backward stepwise selection algorithm. A logistic regression model was built based on the five



**Figure 3.** Bubble plot showing relative abundance of *Lactobacillus* species across stage groups.

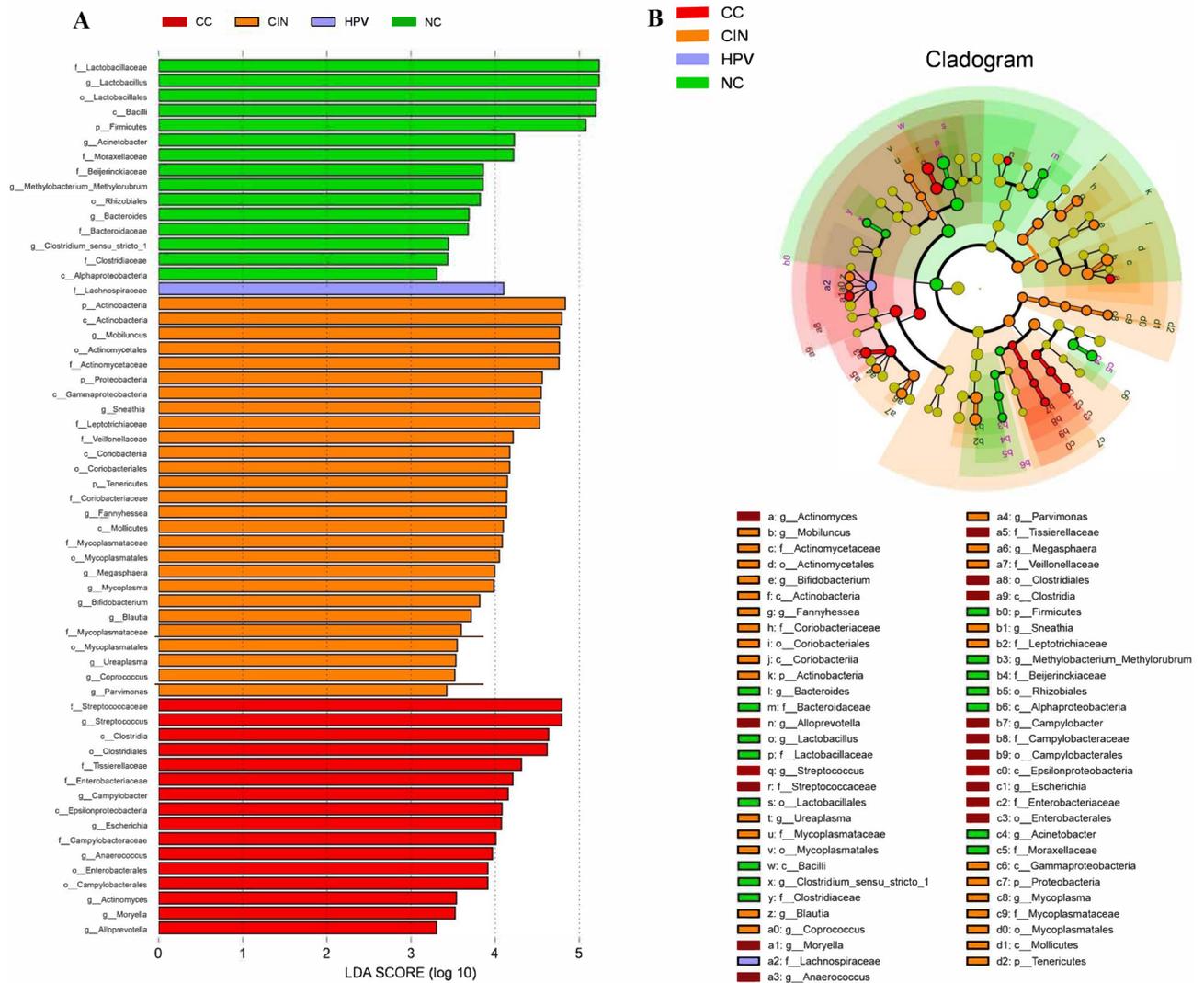


**Figure 4.** Comparison of top 10 abundant bacterial genera across disease progression stages. Multiple groups comparisons were performed using Kruskal–Wallis test. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ; \*\*\*\* $p < 0.0001$ .

biomarkers. To evaluate the performance of the model, ROC analysis was conducted, yielding an area under the curve (AUC) of 0.8947 (95%CI: 0.8343, 0.9551) (Fig. 6D).

### Potential biomarkers to classify other disease stages

Microbial biomarkers were identified with the purpose of developing invasive diagnostic procedures for distinguishing between CIN and NC, HPV infection and NC, CIN and HPV infection, CC and HPV infection, and CC and CIN. The microbial taxa that exhibited significant differences between CIN and NC samples consisted of 14 genera that were enriched in CIN samples and 4 genera that were deficient in CIN samples (Fig. 6B). The logistic regression model, which utilised three genera, effectively differentiated between CIN and NC. It achieved an AUC of 0.7673 (95% Confidence Interval [CI]: 0.7007, 0.8339) (Fig. 6E). Nevertheless, the model's ability to differentiate between HPV and NC was deemed inadequate, as indicated by the AUC of 0.6568 (95% confidence interval: 0.6035, 0.7101) (Fig. 6C, F). Similarly, the performance of the models for distinguishing the CIN from HPV was good (AUC: 0.7669, 95%CI: 0.6919, 0.8418) (Figure S1A). The taxa were significantly changed between CC and HPV including 9 CC-enriched and 5 CC-depleted genera (Figure S1B). Five genera were capable of differentiating samples between CC and HPV (AUC: 0.8836, 95%CI: 0.8217, 0.9455) (Figure S1E). For CC vs CIN, 2 CC-enriched genera and 1 CC-depleted genus were significantly altered (Figure S1C), these



**Figure 5.** Bacterial taxa differences among the four groups using Wekemo Bioincloud (<https://bioincloud.tech/>). (A) Linear discriminant analysis (LDA) effect size (LEfSe) analysis on selected core taxa among the four groups. Only lineages with LDA values > 3 are displayed. (B) Cladogram showing the phylogenetic distribution of lineages associated with the four groups.

genera were capable of discriminating samples between CC and CIN by achieving an AUC of 0.7353 (95%CI: 0.6422, 0.8284) (Figure S1F).

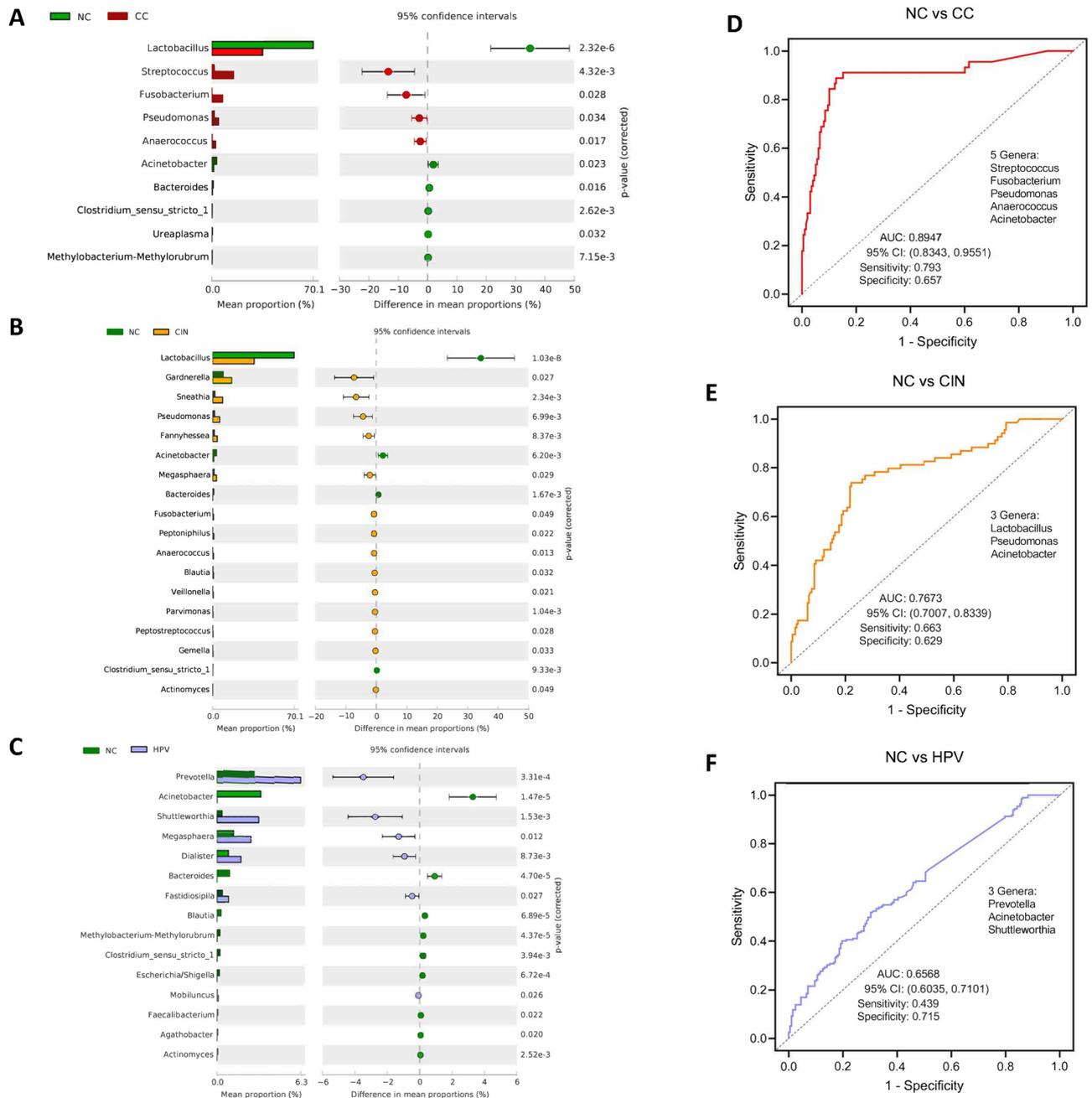
### Correlation network analysis

To further our comprehension of the potential interaction among core taxa, we performed co-occurrence network analysis in the datasets. Significant correlations were found in 35 genera pairs ( $r > 0.6$  or  $r < -0.6$ ,  $p < 0.05$ ) (Table S1). There were three genera pair showing negative correlations. *Lactobacillus* exhibited a negative correlation compared with all other dominant genera (Fig. 7A). The most strongest negative correlation was found between *Lactobacillus* and *Prevotella*. In addition, *Lactobacillus* also had negative correlations with *Fannyhessea* and *Dialister*. Positive co-occurrence correlations occurred in the genera of *Prevotella*, *Sneathia*, *Megasphaera*, *Fannyhessea*, *Acinetobacter*, *Dialister*, *Anaerococcus* and *Bifidobacterium* (Fig. 7B). These microorganisms may play a crucial role in the network.

### Discussion

Cervical cancer is a multifactorial disease involving the interactions among host, microbial and environmental factors. Despite HPV being demonstrated as the main well-established risk factor, cervicovaginal microbiome dysbiosis has emerged as a key risk factor in inflammation<sup>50</sup>, HPV acquisition, persistence and cervical carcinogenesis<sup>19</sup>. We integrated six datasets to present a cervical-related microbial landscape along the different stages, and explored the potential bacterial taxa as biomarkers to monitor cervical carcinogenesis.

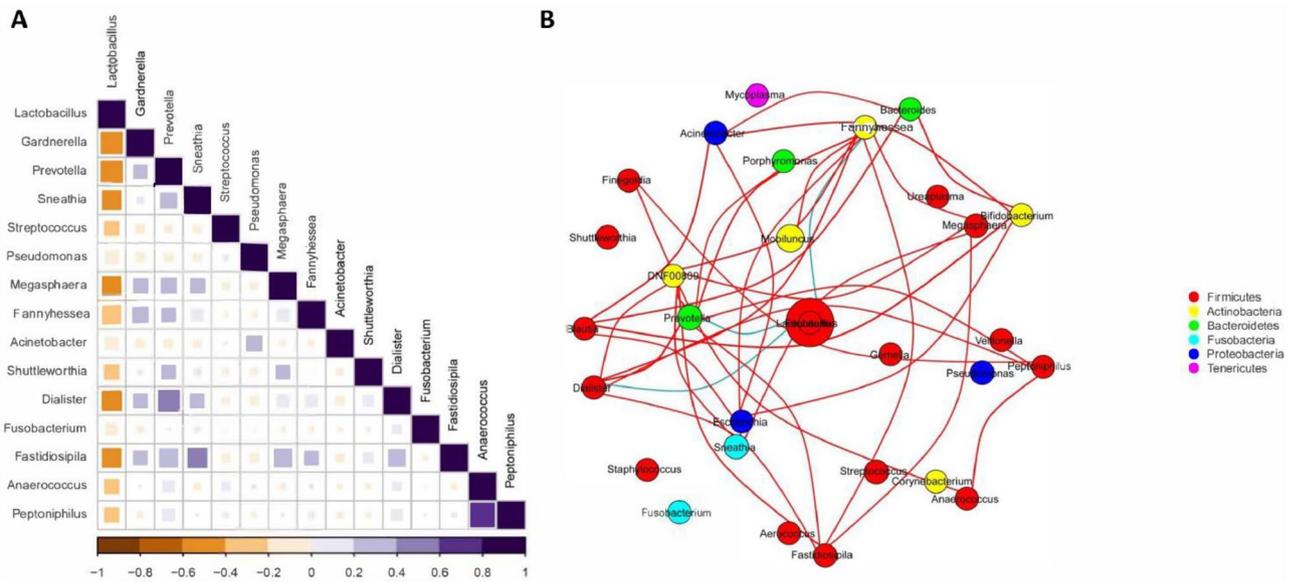
As an important component of the cervicovaginal self-purification function and biological barrier, the resident microbiota and opportunistic pathogens of the cervix maintain the microecological balance. The dominant



**Figure 6.** Differentially taxa between CC, CIN, HPV and NC and the diagnostic genera markers. (A–C) The significantly altered genera as revealed by the extended error bar method using the Wilcoxon rank-sum test. (D–F) Receiver operating characteristic (ROC) analysis for the identified genera markers with logistic regression model discriminating CC, CIN, and HPV from NC.

bacteria *Lactobacillus* can decompose cervical epithelial glycogen to produce lactic acid, maintain a weakly acidic environment, which can inhibit the proliferation of pathogens<sup>51</sup>. In addition, they can enhance anti-infection ability by producing various metabolites or stimulating immune cells to produce various cytokines<sup>52</sup>. Following imbalance of this defense system, it may induce histological alterations of the vaginal mucosa and the cervical epithelium, thereby exerting a selective pressure on the microbiota<sup>53</sup>. Some cervicovaginal taxa, such as *Gardnerella*, *Fusobacteria*, *Dialister* and *Prevotella*, as well as a decrease in the proportion of *Lactobacillus* spp. have been linked to dysbiosis that would generate an unstable microenvironment, which in turn may affect key risk factor in cervical cancer<sup>54,55</sup>. Acid-producing *Lactobacillus* dysbiosis are responsible for increasing the levels of mucin-degrading enzymes, which may affect the mucous layer stability of the cervicovaginal epithelium<sup>56</sup>.

We observed no significant difference between HPV and NC group in bacterial richness and diversity, which is consistent with several previous studies<sup>15,57,58</sup>. Although a few studies results showed that HPV infection can increase bacterial diversity<sup>13,43,59</sup>. The prevailing bacterial taxa that exhibited enrichment in women with HPV



**Figure 7.** (A) Correlation analysis of the top 15 dominant genera with spearman's rank correlation test method using R software (version 4.1.2). (B) Correlation network of the core genera in the combined dataset using Wekemo Bioincloud (<https://bioincloud.tech/>). The blue line indicates a negative correlation ( $r < -0.6$ ), and the red line indicates a positive correlation ( $r > 0.6$ ). Each circle represents a core bacterial taxa, and the colour of the circle represents the phylum to which it belongs.

infection were *Prevotella*, *Megasphaera*, *Shuttleworthia*, and *Dialister*. This finding aligns with prior investigations conducted in this field<sup>60,61</sup>. *Prevotella* has been considered to play an important role in HPV infection and persistence<sup>43</sup>. This bacterium was indicated as a major modulator of host inflammatory responses in the female genital tract by increasing the number of cytokines in cervicovaginal fluid<sup>62</sup>. Moreover, *Prevotella* can secrete proteases to degrade host antibodies, and transfer ammonia to *Gardnerella*, resulting reduction in host mucosal immunity<sup>63</sup>. Lebeau et al.<sup>64</sup> found that HPV infection alters vaginal microbiome through down-regulating host mucosal innate peptides used by *Lactobacilli* as amino acid sources. However, several studies have shown that HPV does not necessarily induce significant changes in the cervicovaginal microbial communities<sup>15,65</sup>. Borgogna et al.<sup>31</sup> found that the vaginal metabolome of HPV-positive women differed from normal individuals in terms of several metabolites, including biogenic amines, glutathione and lipid-related metabolites.

Higher richness and diversity were observed in CIN and CC groups when compared to NC group. *Gardnerella*, one of the most frequently reported in CIN studies, was enriched in CIN group compared with NC, as well as other undesirable genera, like *Sneathia*, *Pseudomonas*, *Fannyhessea* and *Megasphaera*. A longitudinal cohort study revealed a positive association between *Gardnerella* and CIN progression caused by elevated microbial diversity<sup>66</sup>. Previous studies have posited that there may be an increased risk of CIN associated with the enrichment of *Gardnerella vaginalis* (*G. vaginalis*) and *Fannyhessea vaginae*<sup>67</sup>. *G. vaginalis* is classified as a facultative anaerobe, capable of adhering to the vaginal epithelium. This adherence serves as a framework for the production of biofilms and promotes the proliferation of microorganisms<sup>68,69</sup>. It was associated with higher levels of inflammatory cytokines<sup>70</sup>. It has been reported that *Fannyhessea vaginae* can activate the proinflammatory transcription factor NF- $\kappa$ B in the cervicovaginal epithelial cells, triggering abundant inflammation and innate immune responses<sup>71,72</sup>.

In relation to CC, it has been observed that some bacterial taxa, namely *Fusobacterium*, *Pseudomonas*, and *Anaerococcus*, exhibit not only higher abundance in women with precancerous lesions but also in women with cervical cancer. Other identified taxon in this meta-analysis to be associated with cervical cancer included *Streptococcus*. Previously, species of *Streptococcus* have been reported that likely involve the activation of multiple inflammatory cytokines and may affect human vaginal and cervical epithelial cells. Soares et al.<sup>73</sup> revealed that *Streptococcus* possesses metallopeptidases which could help them invade tissue or cause bacterial transmission. *Fusobacterium* species have attracted widespread attention due to their pro-inflammatory properties<sup>74,75</sup>. *Fusobacterium nucleatum* has been shown to potentiate intestinal tumorigenesis and modulate the tumor-immune microenvironment, indicating potential as a diagnostic biomarker for colorectal cancer<sup>76,77</sup>. Besides, *Fusobacterium nucleatum* also has been studied as a possible diagnostic biomarker of CC as it is positively correlated with tumor differentiation<sup>78</sup>. *Pseudomonas* has long been considered to be an opportunistic pathogen in vaginal inflammation, the human urogenital system. They can disrupt the mucosal defense against extracellular pathogens by secreting protease IV and inactivating interleukin 22<sup>79</sup>. It has been indicated that *Pseudomonas aeruginosa* has a potential role in the development of cervical cancer by promoting the expression of integrins in cervical cancer cells<sup>80</sup>.

In this meta-analysis, we observed distinct distributions of the CC group compared with other sample groups. The most relevant genera in each disease stage were revealed by our study, which allowed us to discover robust diagnostic biomarkers. Finally, five common genera including *Streptococcus*, *Fusobacterium*, *Pseudomonas*,

*Anaerococcus* and *Acinetobacter* were identified as the most important features for distinguishing CC from the normal population (AUC > 0.8), indicating their possible role in cervical carcinogenesis as well as their clinical applications. Additionally, we also identified the depletion of potentially beneficial microorganisms, such as lactic acid-producing *Lactobacillus*. It has been reported that *L. iners* can create a protective micro-ecological environment by regulating the core fucosylation of the vaginal epithelial cell against CC<sup>44</sup>. Moreover, we identified additional biomarkers to discriminate CC from CIN (AUC > 0.7) and HPV (AUC > 0.8).

Profiling cervicovaginal microbial communities using 16S rRNA genes is a straightforward and cost-effective method, and is generally cheaper than shotgun metagenomic sequencing. Nevertheless, shotgun metagenomics and circularizing probes-based RNA (ciRNA) sequencing targeted approaches with species-level resolution can provide in-depth insights into cervicovaginal microbiome. A growing number studies have employed shotgun sequencing of the vaginal metagenome<sup>41,81–84</sup>, metatranscriptome<sup>85,86</sup> and ciRNA<sup>87,88</sup>. Liu et al.<sup>84</sup> have observed a total of 111 species in vaginal microbiome of healthy Chinese women, including all dominant vaginal *Lactobacillus* species, such as *L. iners*, *L. crispatus*, *L. gasseri*, and *L. jensenii*. It has been reported that *L. crispatus*, *L. iners* and *G. vaginalis* were the top three species in both HPV16-positive and control groups<sup>83</sup>. A previous study by Macklaim et al.<sup>85</sup> focused on the transcriptional activity of *L. iners* in four reproductive age women and found variation in the species, transcriptional activity. Communities dominated by *L. crispatus* were found to exhibit higher expression of phosphate and phosphonate transporters<sup>86</sup>.

The meta-analysis conducted for CIN and CC in our study was constrained by limitations in the study number and sample size. Additionally, the study lacks comprehensive demographic and clinical data. Therefore, we appeal researchers to accurately and comprehensively disseminate their sequencing data and accompanying metadata. Furthermore, the division of CIN samples into low-grade squamous intraepithelial lesions (LSIL) and high-grade squamous intraepithelial lesions (HSIL) was not feasible due to the limited sample size. Moreover, the fact that LSIL are considered non-progressive lesions, which could affect the analyses in the study.

## Conclusions

In summary, the present study conducted a comprehensive analysis of several cervical 16S rRNA gene sequencing datasets in a standardised manner, leading to the identification of distinct microbial characteristics in the cervical region throughout different stages of disease progression. The findings of our investigation suggest a potential correlation between the presence of certain microorganisms and the development of cervical cancer, which provides great clinical significance and feasibility for the development of noninvasive screening or diagnosis methods for cervical cancer.

## Data availability

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Received: 3 November 2023; Accepted: 17 May 2024

Published online: 21 May 2024

## References

- Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**(6), 394–424 (2018).
- Misra, J. S. et al. Role of different etiological factors in progression of cervical intraepithelial neoplasia. *Diagn. Cytopathol.* **34**(10), 682–685 (2006).
- Wilkinson, E. J. et al. Evolution of terminology for human-papillomavirus-infection-related vulvar squamous intraepithelial lesions. *J. Low. Genit. Tract Dis.* **19**(1), 81–87 (2015).
- Forman, D. et al. Global burden of human papillomavirus and related diseases. *Vaccine* **30**, F12–F23 (2012).
- White, B. A. et al. The vaginal microbiome in health and disease. *Trends Endocrinol Metab.* **22**(10), 389–393 (2011).
- Torcia, M. G. Interplay among vaginal microbiome, immune response and sexually transmitted viral infections. *Int. J. Mol. Sci.* **20**(2), 266 (2019).
- Wilson, W. A. et al. Regulation of glycogen metabolism in yeast and bacteria. *FEMS Microbiol. Rev.* **34**(6), 952–985 (2010).
- Zacharof, M.-P. & Lovitt, R. Bacteriocins produced by lactic acid bacteria a review article. *APCBEE Proceed.* **2**, 50–56 (2012).
- Haahr, T. et al. Reproductive outcome of patients undergoing in vitro fertilisation treatment and diagnosed with bacterial vaginosis or abnormal vaginal microbiota: A systematic PRISMA review and meta-analysis. *BJOG Int. J. Obstet. Gynaecol.* **126**(2), 200–207 (2019).
- Lewis, F. M., Bernstein, K. T. & Aral, S. O. Vaginal microbiome and its relationship to behavior, sexual health, and sexually transmitted diseases. *Obstet. Gynecol.* **129**(4), 643 (2017).
- Lin, W. et al. Changes of the vaginal microbiota in HPV infection and cervical intraepithelial neoplasia: A cross-sectional analysis. *Sci. Rep.* **12**(1), 2812 (2022).
- 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**(1), 1–15 (2016).
- Brotman, R. M. et al. Interplay between the temporal dynamics of the vaginal microbiota and human papillomavirus detection. *J. Infect. Dis.* **210**(11), 1723–1733 (2014).
- Norenhag, J. et al. The vaginal microbiota, human papillomavirus and cervical dysplasia: A systematic review and network meta-analysis. *BJOG Int. J. Obstet. Gynaecol.* **127**(2), 171–180 (2020).
- Godoy-Vitorino, F. et al. Cervicovaginal fungi and bacteria associated with cervical intraepithelial neoplasia and high-risk human papillomavirus infections in a hispanic population. *Front. Microbiol.* **9**, 2533 (2018).
- Lee, J. E. et al. Association of the vaginal microbiota with human papillomavirus infection in a Korean twin cohort. *PLoS One* **8**(5), e63514 (2013).
- Ravel, J. et al. Vaginal microbiome of reproductive-age women. *Proc. Natl. Acad. Sci.* **108**(Suppl 1), 4680–4687 (2011).
- Huang, X. et al. Cervicovaginal microbiota composition correlates with the acquisition of high-risk human papillomavirus types. *Int. J. Cancer* **143**(3), 621–634 (2018).

19. Brusselaers, N. *et al.* Vaginal dysbiosis and the risk of human papillomavirus and cervical cancer: Systematic review and meta-analysis. *Am. J. Obstet. Gynecol.* **221**(1), 9–18 e8 (2019).
20. Mitra, A. *et al.* The vaginal microbiota associates with the regression of untreated cervical intraepithelial neoplasia 2 lesions. *Nat. Commun.* **11**(1), 1999 (2020).
21. Zhang, Y. *et al.* Vaginal microbiota changes caused by HPV infection in Chinese women. *Front. Cell. Infect. Microbiol.* <https://doi.org/10.3389/fcimb.2022.814668> (2022).
22. Audirac-Chalifour, A. *et al.* Cervical microbiome and cytokine profile at various stages of cervical cancer: A pilot study. *PLoS One* **11**(4), e0153274 (2016).
23. Cheng, L. *et al.* Vaginal microbiota and human papillomavirus infection among young Swedish women. *NPJ Biofilms Microb.* **6**(1), 39 (2020).
24. Xie, Y. *et al.* Revealing the disturbed vaginal microbiota caused by cervical cancer using high-throughput sequencing technology. *Front. Cell. Infect. Microbiol.* **10**, 538336 (2020).
25. Ilhan, Z. E. *et al.* Deciphering the complex interplay between microbiota, HPV, inflammation and cancer through cervicovaginal metabolic profiling. *EBioMedicine* **44**, 675–690 (2019).
26. Kang, G.-U. *et al.* Potential association between vaginal microbiota and cervical carcinogenesis in Korean women: A cohort study. *Microorganisms* **9**(2), 294 (2021).
27. Wei, Z.-T. *et al.* Depiction of vaginal microbiota in women with high-risk human papillomavirus infection. *Front. Public Health* **8**, 587298 (2021).
28. Mei, L. *et al.* Dysbiosis of vaginal microbiota associated with persistent high-risk human papilloma virus infection. *J. Transl. Med.* **20**(1), 1–8 (2022).
29. Chen, Y. *et al.* Human papillomavirus infection and cervical intraepithelial neoplasia progression are associated with increased vaginal microbiome diversity in a Chinese cohort. *BMC Infect. Dis.* **20**, 1–12 (2020).
30. Onywera, H. *et al.* Predictive functional analysis reveals inferred features unique to cervicovaginal microbiota of African women with bacterial vaginosis and high-risk human papillomavirus infection. *PLoS One* **16**(6), e0253218 (2021).
31. Borgogna, J. *et al.* The vaginal metabolome and microbiota of cervical HPV-positive and HPV-negative women: A cross-sectional analysis. *BJOG Int. J. Obstet. Gynaecol.* **127**(2), 182–192 (2020).
32. McKee, K. S. *et al.* The vaginal microbiota, high-risk human papillomavirus infection, and cervical cytology: Results from a population-based study. *Gynecol. Pelvic Med.* **3**, 18 (2020).
33. Lin, S. *et al.* Dysbiosis of cervical and vaginal microbiota associated with cervical intraepithelial neoplasia. *Front. Cell. Infect. Microbiol.* **12**, 20 (2022).
34. Wang, H. *et al.* Observation of the cervical microbiome in the progression of cervical intraepithelial neoplasia. *BMC Cancer* **22**(1), 362 (2022).
35. Tango, C. N. *et al.* Taxonomic and functional differences in cervical microbiome associated with cervical cancer development. *Sci. Rep.* **10**(1), 9720 (2020).
36. Nieves-Ramirez, M. *et al.* Cervical squamous intraepithelial lesions are associated with differences in the vaginal microbiota of Mexican women. *Microbiol. Spectr.* **9**(2), e00143–e221 (2021).
37. Mitra, A. *et al.* Cervical intraepithelial neoplasia disease progression is associated with increased vaginal microbiome diversity. *Sci. Rep.* **5**(1), 16865 (2015).
38. Bokulich, N. A. *et al.* Multi-omics data integration reveals metabolome as the top predictor of the cervicovaginal microenvironment. *PLoS Comput. Biol.* **18**(2), e1009876 (2022).
39. Chao, X. *et al.* Research of the potential vaginal microbiome biomarkers for high-grade squamous intraepithelial lesion. *Front. Med.* **8**, 565001 (2021).
40. Piyathilake, C. J. *et al.* Cervical microbiota associated with higher grade cervical intraepithelial neoplasia in women infected with high-risk human papillomavirus microbiome and CIN Risk. *Cancer Prev. Res.* **9**(5), 357–366 (2016).
41. Fang, B. *et al.* Exploring the association between cervical microbiota and HR-HPV infection based on 16S rRNA gene and metagenomic sequencing. *Front. Cell. Infect. Microbiol.* <https://doi.org/10.3389/fcimb.2022.922554> (2022).
42. Zhou, Y. *et al.* Patients with LR-HPV infection have a distinct vaginal microbiota in comparison with healthy controls. *Front. Cell. Infect. Microbiol.* **9**, 294 (2019).
43. Di Paola, M. *et al.* Characterization of cervico-vaginal microbiota in women developing persistent high-risk human papillomavirus infection. *Sci. Rep.* **7**(1), 1–12 (2017).
44. Fan, Q. *et al.* Lactobacillus spp. Create a protective micro-ecological environment through regulating the core fucosylation of vaginal epithelial cells against cervical cancer. *Cell Death Dis.* **12**(12), 1094 (2021).
45. Wang, Z. *et al.* The diversity of vaginal microbiota predicts neoadjuvant chemotherapy responsiveness in locally advanced cervical cancer. *Microb. Ecol.* **84**, 1–12 (2021).
46. Lam, K. C. *et al.* Transkingdom network reveals bacterial players associated with cervical cancer gene expression program. *PeerJ* **6**, e5590 (2018).
47. Zhang, C. *et al.* The direct and indirect association of cervical microbiota with the risk of cervical intraepithelial neoplasia. *Cancer Med.* **7**(5), 2172–2179 (2018).
48. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**(7), 581–583 (2016).
49. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**(D1), D590–D596 (2012).
50. Ness, R. B. *et al.* A cluster analysis of bacterial vaginosis-associated microflora and pelvic inflammatory disease. *Am. J. Epidemiol.* **162**(6), 585–590 (2005).
51. McMillan, A. *et al.* Disruption of urogenital biofilms by lactobacilli. *Colloids Surf. B Biointerfaces* **86**(1), 58–64 (2011).
52. Wang, C. *et al.* The Effect of Lactobacillus isolates on growth performance, immune response, intestinal bacterial community composition of growing Rex Rabbits. *J. Anim. Physiol. Anim. Nutr.* **101**(5), e1–e13 (2017).
53. Kim, T. K. *et al.* Heterogeneity of vaginal microbial communities within individuals. *J. Clin. Microbiol.* **47**(4), 1181–1189 (2009).
54. Shigehara, K. *et al.* Prevalence of genital Mycoplasma, Ureaplasma, Gardnerella, and human papillomavirus in Japanese men with urethritis, and risk factors for detection of urethral human papillomavirus infection. *J. Infect. Chemother.* **17**(4), 487–492 (2011).
55. Kovachev, S. M. Cervical cancer and vaginal microbiota changes. *Arch. Microbiol.* **202**(2), 323–327 (2020).
56. Lamont, R. F. *et al.* The vaginal microbiome: New information about genital tract flora using molecular based techniques. *BJOG Int. J. Obstet. Gynaecol.* **118**(5), 533–549 (2011).
57. Tuominen, H. *et al.* HPV infection and bacterial microbiota in the placenta, uterine cervix and oral mucosa. *Sci. Rep.* **8**(1), 1–11 (2018).
58. Chao, X.-P. *et al.* Correlation between the diversity of vaginal microbiota and the risk of high-risk human papillomavirus infection. *Int. J. Gynecol. Cancer* **29**(1), 28 (2019).
59. Shannon, B. *et al.* Association of HPV infection and clearance with cervicovaginal immunology and the vaginal microbiota. *Mucosal Immunol.* **10**(5), 1310–1319 (2017).
60. Łaniewski, P. *et al.* Linking cervicovaginal immune signatures, HPV and microbiota composition in cervical carcinogenesis in non-Hispanic and Hispanic women. *Sci. Rep.* **8**(1), 7593 (2018).
61. Santella, B. *et al.* Microbiota and HPV: The role of viral infection on vaginal microbiota. *J. Med. Virol.* **94**(9), 4478–4484 (2022).

62. Amano, A. *et al.* Variations of *Porphyromonas gingivalis* fimbriae in relation to microbial pathogenesis. *J. Periodontol Res.* **39**(2), 136–142 (2004).
63. Pybus, V. & Onderdonk, A. B. Evidence for a commensal, symbiotic relationship between *Gardnerella vaginalis* and *Prevotella bivia* involving ammonia: Potential significance for bacterial vaginosis. *J. Infect. Dis.* **175**(2), 406–413 (1997).
64. 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**(1), 1076 (2022).
65. Bienkowska-Haba, M. *et al.* A new cell culture model to genetically dissect the complete human papillomavirus life cycle. *PLoS Pathog.* **14**(3), e1006846 (2018).
66. Usyk, M. *et al.* Cervicovaginal microbiome and natural history of HPV in a longitudinal study. *PLoS Pathogens* **16**(3), e1008376 (2020).
67. Oh, H. *et al.* The association of uterine cervical microbiota with an increased risk for cervical intraepithelial neoplasia in Korea. *Clin. Microbiol. Infect.* **21**(7), 674.e1–74.e9 (2015).
68. Fethers, K. *et al.* Bacterial vaginosis (BV) candidate bacteria: Associations with BV and behavioural practices in sexually-experienced and inexperienced women. *PLoS One* **7**(2), e30633 (2012).
69. Harwich, M. D. *et al.* Drawing the line between commensal and pathogenic *Gardnerella vaginalis* through genome analysis and virulence studies. *BMC Genom.* **11**, 1–12 (2010).
70. Moscicki, A. B., Shi, B., Huang, H., Barnard, E. & Li, H. Cervical-vaginal microbiome and associated cytokine profiles in a prospective study of HPV 16 acquisition, persistence, and clearance. *Front. Cell. Infect. Microbiol.* **10**, 569022 (2020).
71. Libby, E. K. *et al.* Atopobium vaginae triggers an innate immune response in an in vitro model of bacterial vaginosis. *Microb. Infect.* **10**(4), 439–446 (2008).
72. Doerflinger, S. Y., Throop, A. L. & Herbst-Kralovetz, M. M. Bacteria in the vaginal microbiome alter the innate immune response and barrier properties of the human vaginal epithelia in a species-specific manner. *J. Infect. Dis.* **209**(12), 1989–1999 (2014).
73. Soares, G. C. M. T. *et al.* Metalloproteinases produced by group B Streptococcus: Influence of proteolytic inhibitors on growth and on interaction with human cell lineages. *Int. J. Mol. Med.* **22**(1), 119–125 (2008).
74. Yang, Y. *et al.* *Fusobacterium nucleatum* increases proliferation of colorectal cancer cells and tumor development in mice by activating toll-like receptor 4 signaling to nuclear factor- $\kappa$ B, and up-regulating expression of microRNA-21. *Gastroenterology* **152**(4), 851–66.e24 (2017).
75. Tang, B. *et al.* *Fusobacterium nucleatum*-induced impairment of autophagic flux enhances the expression of proinflammatory cytokines via ROS in Caco-2 cells. *PLoS One* **11**(11), e0165701 (2016).
76. Kostic, A. D. *et al.* *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microb.* **14**(2), 207–215 (2013).
77. Yu, J. *et al.* Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* **66**(1), 70–78 (2017).
78. Huang, S.-T. *et al.* Intratumoral levels and prognostic significance of *Fusobacterium nucleatum* in cervical carcinoma. *Aging (Albany NY)* **12**(22), 23337 (2020).
79. Bradshaw, J. L. *et al.* *Pseudomonas aeruginosa* protease IV exacerbates pneumococcal pneumonia and systemic disease. *Mosphere* **3**(3), e00212–e218 (2018).
80. Werner, J. *et al.* Expression of integrins and Toll-like receptors in cervical cancer: Effect of infectious agents. *Innate Immun.* **18**(1), 55–69 (2012).
81. Feehily, C. *et al.* Shotgun sequencing of the vaginal microbiome reveals both a species and functional potential signature of preterm birth. *NPJ Biofilms Microb.* **6**(1), 50 (2020).
82. Goltsman, D. S. A. *et al.* Metagenomic analysis with strain-level resolution reveals fine-scale variation in the human pregnancy microbiome. *Genome Res.* **28**(10), 1467–1480 (2018).
83. Yang, Q. *et al.* The alterations of vaginal microbiome in HPV16 infection as identified by shotgun metagenomic sequencing. *Front. Cell. Infect. Microbiol.* **10**, 286 (2020).
84. Liu, F. *et al.* Comparative metagenomic analysis of the vaginal microbiome in healthy women. *Synth. Syst. Biotechnol.* **6**(2), 77–84 (2021).
85. Macklaim, J. M. *et al.* Comparative meta-RNA-seq of the vaginal microbiota and differential expression by Lactobacillus iners in health and dysbiosis. *Microbiome* **1**(1), 1–11 (2013).
86. France, M. T. *et al.* Insight into the ecology of vaginal bacteria through integrative analyses of metagenomic and metatranscriptomic data. *Genome Biol.* **23**(1), 1–26 (2022).
87. Andralojc, K. M. *et al.* Novel high-resolution targeted sequencing of the cervicovaginal microbiome. *BMC Biol.* **19**(1), 1–18 (2021).
88. Molina, M. A. *et al.* Longitudinal analysis on the ecological dynamics of the cervicovaginal microbiome in hrHPV infection. *Comput. Struct. Biotechnol. J.* **21**, 4424–4431 (2023).

## Acknowledgements

The authors thank those authors who provided details of the data in their published articles. Special thanks to Yinfan Liu, from East China University of Science and Technology, for his valuable feedback and advice on our manuscript.

## Author contributions

XL designed this study, analyzed the data and drafted this manuscript; FX and JL collected and organized the data; ZC and MZ analyzed the data; TL, XK and RW contributed a lot to the design and review the manuscript. All authors have read and approved the final version of the manuscript.

## Funding

This work was funded by Science and Technology Innovation Project of Putuo District Health System (No. ptkwws202305, No. ptkwws202307), the One Hundred Talents Project of Putuo Hospital, Shanghai University of Traditional Chinese Medicine (No. 2022-RCQH-03).

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-62531-z>.

**Correspondence** and requests for materials should be addressed to X.K. or R.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](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 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 licence, and indicate if changes were made. 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/by/4.0/>.

© The Author(s) 2024