

Ecological and inferred functional landscape of the cervicovaginal microbiome in relation to HPV infection and disease in Kenyan women

Mariano A. Molina

`mariano.molina.beitia@ki.se`

Karolinska Institutet

Kosgei Noah

Gynocare Women's and Fistula Hospital

Ivy Akinyi

Jaramogi Oginga Odinga University of Science and Technology

Lisse Tavernier

AML Sonic Healthcare

Sadaf Sakina Hassan

Karolinska University Hospital, Karolinska Institutet

Milan Stosic

Akershus University Hospital

Davy Vanden Broeck

AML Sonic Healthcare

Hillary Mabeya

Gynocare Women's and Fistula Hospital

Laila Sara Arroyo Mühr

Karolinska University Hospital, Karolinska Institutet

Article

Keywords: Cervicovaginal microbiome, HPV, Microbial ecology, Inferred functional profiling

Posted Date: October 7th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11267936/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Persistent high-risk human papillomavirus (HPV) infection is the central cause of cervical cancer, but the association of the microbiome with disease severity from normal cytology to cervical lesions remains incompletely understood, particularly in African populations. Here, we characterized the cervicovaginal microbiome of 91 Kenyan women with negative cytology, low-grade, or high-grade squamous intraepithelial lesions using 16S rRNA gene sequencing, HPV genotyping, and *Trichomonas vaginalis* detection. Cervicovaginal communities resolved into six distinct ecological states or cervicotypes that differed in dominant taxa and composition. Increasing lesion severity was associated with depletion of *Lactobacillus* communities and shifts in specific bacterial taxa. Microbial co-occurrence analysis further identified *Lactobacillus*-, *Gardnerella*-, and *Delftia*-associated ecological modules, with the *Lactobacillus*-associated module inversely related to cytological severity and HPV infection. Inferred functional profiling identified 35 pathways significantly associated with increasing lesion severity, including changes in carbohydrate, amino acid, fatty acid, aromatic compound, nucleotide, and cofactor metabolism. These inferred functional signatures correlated with the *Delftia*-associated module. Exploratory penalized regression analyses showed that ecological features provided modest discriminatory information in addition to clinical variables. Altogether, these findings identify coordinated ecological and inferred functional changes in the cervicovaginal microbiome associated with cervical disease severity and support further investigation of microbial ecological organization in HPV-associated cervical disease.

Introduction

Persistent infection with human papillomavirus (HPV) is the necessary cause of virtually all cervical cancer cases, but HPV infection alone is not sufficient to determine disease outcome (1, 2). Most infections are cleared or become undetectable, while a smaller proportion persists and contributes to the development of low-grade (LSIL) and high-grade squamous intraepithelial lesions (HSIL) and, eventually, invasive cancer (3, 4). This divergence in clinical outcome points to an important role for the local microenvironment in shaping the natural history of HPV infection (5). Host immunity, epithelial integrity, coinfections, hormonal and behavioral factors, and the composition of the microbiome may all influence whether an HPV infection resolves or persists (5, 6). Understanding these interacting factors is particularly important in sub-Saharan Africa, where the burden of cervical cancer remains high and where the biological and epidemiological context of HPV infection differs substantially from that of many populations in which microbiome studies have traditionally been performed (7–9).

The cervicovaginal microbiome is increasingly recognized as an important component of this microenvironment. In many women, the vaginal microbiota is dominated by one *Lactobacillus* species, particularly *Lactobacillus crispatus*, *Lactobacillus iners*, *Lactobacillus gasseri*, and *Lactobacillus jensenii* (10). These lactobacilli contribute to a low pH environment, typically around 3.5–4.5, through lactic acid production and have been linked to epithelial homeostasis and protection against several reproductive tract infections (11, 12). Other women harbor more polymicrobial communities enriched for facultative

and obligate anaerobes, including *Gardnerella*, *Prevotella*, *Sneathia*, *Fannyhessea*, and *Megasphaera* (10, 13). This variation has commonly been summarized using community state types or related ecological classifications (14). However, these microbial communities are not simply divided into *Lactobacillus*-dominated and non-*Lactobacillus*-dominated states. High-resolution studies have revealed different dominant species, coexisting bacterial populations, and community structures that may have distinct biological implications (10, 15, 16).

Multiple studies have linked cervicovaginal microbial composition to HPV infection and cervical disease. Persistent HPV infection and cervical lesions have frequently been associated with lower abundance of protective lactobacilli, greater bacterial diversity, and enrichment of anaerobic taxa (17, 18). Several studies have also suggested that specific community configurations are associated with HPV persistence, lesion progression, or lesion regression (10, 19, 20). These observations have led to the notion that the cervicovaginal microbiome may participate in a local microenvironment that favors viral persistence and epithelial transformation (21, 22). Possible mechanisms include altered mucosal barrier function, changes in local inflammation, microbial metabolite production, and interactions with host immune responses (23). At the same time, associations reported for individual taxa are not always consistent between studies. The same bacterial genus may contain species with different ecological roles, and similar taxonomic profiles may occur in communities with very different internal organization (22). This makes it difficult to understand disease-associated microbiome changes from individual taxa alone. Additionally, HPV genotype may further shape these relationships, as individual high-risk HPV types differ in persistence and long-term risk of cervical precancer (24, 25). Coinfections may also modify the local microenvironment; in particular, *Trichomonas vaginalis* has been associated with cervical carcinogenesis (26). Examining these factors alongside microbial ecology may therefore help distinguish disease-associated community patterns from those related to viral or other genital infections.

Vaginal microbiome composition varies with geography, ancestry, sexual behavior, reproductive factors, sexually transmitted infections, antibiotic exposure, and other environmental and host characteristics (27, 28). African populations show a higher prevalence of polymicrobial or non-*Lactobacillus*-dominated communities than European and North American populations (8, 27). This variation limits the extent to which microbial signatures identified in one population can be assumed to apply elsewhere. It also raises a broader biological question: whether the relevant signal lies in the presence of particular organisms, in the way bacteria assemble into ecological communities, or in the inferred functional capacity produced by those communities (20, 22). Most studies of HPV-associated cervicovaginal microbiota have focused primarily on taxonomic composition or broad community state classifications. How these bacteria organize into co-occurring ecological modules, how those modules relate to inferred functional potential, and whether these relationships are evident in underrepresented African populations remain poorly defined (15). Inferred functional characterization is particularly relevant because taxonomically distinct communities may converge on similar metabolic activities, while communities that appear compositionally similar may differ in their inferred functional output (29).

To address these gaps, we characterized the cervicovaginal microbiome of Kenyan women with negative cytology, LSIL, and HSIL. We combined 16S rRNA gene profiling with HPV genotyping and *T. vaginalis* detection to examine microbiome composition in relation to cervical lesion severity and HPV-related characteristics. We first defined cervicovaginal ecological states and identified bacterial taxa associated with clinical features. We then moved beyond individual taxa by reconstructing microbial co-occurrence networks and deriving ecological modules that capture coordinated community structure. Lastly, we inferred microbial metabolic pathways and examined how inferred diagnosis-associated functional changes relate to microbial ecological organization. We also tested whether clinical, ecological, taxonomic, and inferred functional features contain discriminatory information for cervical disease states.

Results

Study cohort and cervicovaginal microbiome composition

The study included 91 women with evaluable cytological diagnoses and cervicovaginal microbiome profiles, comprising 52 samples classified as negative for intraepithelial lesion or malignancy (NILM), 18 as LSIL, and 21 as HSIL. Following taxonomic filtering, 1,618 bacterial ASVs were retained for downstream analyses. The cervicovaginal microbiome varied substantially between participants and was commonly dominated by *Lactobacillus* (Fig. 1A). The remaining samples were enriched for anaerobic genera, including *Gardnerella*, *Prevotella*, *Sneathia*, *Fannyhessea*, and *Megasphaera*. These compositional patterns were observed among cytological diagnoses, HPV risk groups (**Supplementary Table S1**), and *T. vaginalis* infection (Fig. 1A).

Principal coordinates analysis based on Bray–Curtis dissimilarity showed that clinical diagnosis explained a small but statistically significant proportion of the total microbial community variation (PERMANOVA, $R^2 = 0.035$, $p = 0.0225$; Fig. 1B). Shannon diversity and observed richness were comparable between NILM, LSIL, and HSIL (Fig. 1C, **Supplementary Figure S1A**). Shannon diversity tended to be higher among women positive for *T. vaginalis*, although this difference did not reach statistical significance ($p = 0.056$; **Supplementary Figure S1B**). Although the overall distribution of HPV risk groups among cytological diagnoses was not significantly different (Fisher's exact test, $p = 0.256$), HPV risk groups showed a small significant positive correlation with increasing diagnosis severity ($\rho = 0.21$, $p = 0.049$; Fig. 1D).

Unsupervised clustering resolves six ecologically distinct cervicotypes

Unsupervised hierarchical clustering of species-level relative abundance identified six cervicotypes (CT), designated CT1, CT2, CT3, CT4A, CT4B, and CT4C (Fig. 2A). CT1 was predominantly characterized by *Lactobacillus crispatus*/others, whereas CT2 contained communities enriched for *Lactobacillus iners*. CT3

represented more heterogeneous *Gardnerella*-abundant communities (Fig. 2A). In contrast, CT4A-CT4C were comparatively depleted of *Lactobacillus* and represented three distinct polymicrobial community profiles. CT4A was characterized by enrichment of *Hoylesella*, *Megasphaera*, and *Gardnerella* species, CT4B by *Sneathia*, *Prevotella*, and *Gardnerella* species, whereas CT4C was distinguished by dominance of *Delftia*, *Bifidobacterium*, *Streptococcus*, and other anaerobic taxa (Fig. 2A).

tSNE projections showed a visualization concordant with the assigned cervicotypes, with samples assigned to the same cervicotype generally occupying similar regions of the ordination space (Fig. 2B). This organization was closely related to the identity of the most abundant taxon, while the second-most abundant taxon further distinguished samples with otherwise similar dominant community profiles (Fig. 2C–2D). Samples with stronger dominance by a single taxon also occupied distinct regions of the ordination compared with more evenly distributed communities (Fig. 2E).

Cervicotype distributions varied among cytological diagnoses. CT1 and CT3 contained higher proportions of NILM samples, whereas CT4B and CT4C contained larger proportions of LSIL and HSIL samples (**Supplementary Figure S2A**). HPV risk group composition also varied across cervicotypes, with higher-risk HPV categories occurring more frequently in CT4A and CT4C (**Supplementary Figure S2B**). Likewise, CT4 cervicotypes accounted for a larger proportion of SIL samples (**Supplementary Figure S3A**), whereas their distribution was comparable between single and multiple HPV infections (**Supplementary Figure S3B**). As expected, Shannon diversity significantly differed among cervicotypes ($p < 0.0001$), with significantly lower values in *Lactobacillus*-dominated communities ($p < 0.05$) and significantly higher values in polymicrobial anaerobic communities ($p < 0.05$; **Supplementary Figure S1C**, **Supplementary Table S2**).

Bacterial taxa are associated with cervical lesion severity and *T. vaginalis* infection

Next, MaAsLin2 was used to identify bacterial taxa associated with cytological diagnosis severity (NILM–LSIL–HSIL), HPV infection, and *T. vaginalis* infection. At an exploratory FDR threshold of $q < 0.10$, increasing cytological diagnosis severity was associated with positive coefficients for *Delftia*, whereas *Lactobacillus* and *G. vaginalis* showed inverse associations (Fig. 3A–3B, **Supplementary Figure S4A**). Likewise, HPV infection showed a weak positive association with *Anaerococcus*, whereas *Lactobacillus* taxa generally showed negative coefficients (**Supplementary Figure S4A**). HPV16/18 infection was also associated with *Corynebacterium*, *Anaerococcus*, and *Staphylococcus* species (Fig. 3A, **Supplementary Figure S4A**). Notably, *T. vaginalis* infection showed positive associations with *Mycoplasma* ($q < 0.05$), *Prevotella bivia*, and *Fannyhessea* ($q < 0.10$; Fig. 3A and **Supplementary Figure S4B**). After adjusting for HPV risk group and *T. vaginalis*, we found broadly consistent effect directions for *Delftia*, although associations were attenuated after FDR correction (**Supplementary Table S3**).

Microbial co-occurrence analysis identifies ecological modules

To investigate microbial composition beyond individual taxa, SparCC associations were calculated among prevalent bacterial taxa and clustered using the Markov Cluster Algorithm. This analysis identified three ecological modules characterized by *Gardnerella*-, *Delftia*-, and *Lactobacillus*-associated taxa (Fig. 4A). Taxa within each module showed predominantly positive associations, while the corresponding association matrix revealed distinct patterns of co-occurrence among the three ecological communities (**Supplementary Figure S5**). Of these, the *Lactobacillus*-associated module showed a significant negative correlation with increasing diagnosis severity (NILM–LSIL–HSIL; $\rho = -0.28$, $q < 0.05$) and with HPV infection ($\rho = -0.26$, $q < 0.05$; Fig. 4B), whereas the *Gardnerella*- and *Delftia*-associated modules were not significantly associated with these clinical characteristics. Likewise, the *Lactobacillus*-associated module showed an inverse correlation with HPV risk groups ($\rho = -0.27$, $p = 0.01$). However, this association was no longer significant after adjustment for cytological diagnosis ($\beta = -0.12$, $p = 0.107$; **Supplementary Figure S6**)

Inferred metabolic pathways associated with cytological diagnosis

The functional potential of the cervicovaginal microbiome was then examined using PICRUSt2-inferred MetaCyc pathway abundances. MaAsLin2 identified 35 inferred pathways significantly associated with cytological diagnosis after FDR correction ($q < 0.05$; Fig. 5A). These inferred pathways included functions related to carbohydrate and organic acid degradation, amino acid metabolism, fatty acid metabolism, nucleotide biosynthesis, and cofactor biosynthesis. Hierarchical clustering of standardized pathway abundances revealed heterogeneous inferred functional profiles within each cytological diagnosis, with visually distinct patterns also corresponding to cervicotypes and microbial module scores (Fig. 5A). These inferred pathways included UMP biosynthesis II and III, adenosylcobalamin biosynthesis II, fatty acid salvage and β -oxidation, and several inferred degradation pathways involving galactarate, galactate, glucarate, erythronate, histidine, tryptophan, methylglyoxal, catechol, salicylate, and related aromatic compounds (Fig. 5A). Additional inferred pathways included ectoine, glucosylglycerol, and 2-methylbutanoate.

Interestingly, of the 35 inferred cytological diagnosis-associated pathways, 17 were also associated with cervicotype, 13 with the *Delftia*-associated module, and two with the *Gardnerella*-associated module, whereas none overlapped with inferred pathways associated with the *Lactobacillus*-associated module (**Supplementary Figure S7**). To summarize the magnitude and direction of these inferred functional changes, cytological diagnosis-associated pathways were ranked according to their MaAsLin2 effect sizes (Fig. 5B). Among the inferred pathways with the largest positive effect sizes, inferred abundances increased with ordinal cytological diagnosis severity (NILM–LSIL–HSIL), including pathways involved in fatty acid salvage and β -oxidation, methylglyoxal degradation, D-glucarate and D-galactarate degradation, histidine degradation, and adenosylcobalamin biosynthesis. In contrast, UMP biosynthesis II and III showed decreasing inferred abundances with increasing cytological diagnosis severity (Fig. 5B).

To further characterize the ecological distribution of these inferred cytological diagnosis-associated functional alterations, we calculated a composite pathway score summarizing the inferred cytological

diagnosis-associated pathway signature and compared it among cervicotypes (Fig. 5C). Notably, the composite score differed significantly among cervicotypes (Kruskal–Wallis, $p < 0.001$), with CT4C exhibiting significantly higher scores than CT2, CT3, CT4A, and CT4B ($q < 0.01$). In contrast, CT3, CT4A, and CT4B showed comparatively similar scores. CT1 also showed significantly higher scores than CT2 ($q < 0.01$).

Microbial ecological modules exhibit defined inferred functional signatures

Next, we further examined the ecological basis of these inferred functional alterations by assessing relationships between the *Gardnerella*- and *Delftia*-associated microbial modules and the inferred cytological diagnosis-associated pathways. The *Delftia*-associated module showed broad positive correlations with inferred diagnosis-associated pathways, including inferred pathways involved in carbohydrate and organic-acid degradation, fatty acid metabolism, amino acid metabolism, and aromatic compound degradation ($q < 0.05$; Fig. 6A). In contrast, UMP biosynthesis II and III were negatively correlated with the *Delftia*-associated module. The *Gardnerella*-associated module showed substantially fewer significant associations, with the strongest signal corresponding to a negative correlation with L-tryptophan degradation X ($q < 0.05$; Fig. 6A). These patterns are summarized in the conceptual framework shown in Fig. 6B.

Ecological features provide discriminatory information across cervical disease severity

Finally, we explored whether clinical, taxonomic, ecological, and inferred functional features could distinguish cervical cytological states using penalized regression with repeated nested five-fold cross-validation. In the binary NILM-versus-SIL analysis, discrimination was modest across all feature sets (**Supplementary Figure S8A**). Ecological features, including cervicotype and the *Lactobacillus*-, *Gardnerella*-, and *Delftia*-associated module scores, performed better than clinical features alone, and the combined clinical and ecological model showed the most consistent improvement across resampling iterations. Across repeated cross-validation, this model achieved significantly higher AUCs than the clinical-only and ecological-only models (paired Wilcoxon $p = 2.3 \times 10^{-5}$ and $p = 0.0014$, respectively), although differences based on averaged out-of-fold predictions were not significant by DeLong testing after FDR correction. Given that several microbiome associations followed an ordered NILM–LSIL–HSIL gradient, we additionally modeled cytological diagnosis as an ordinal outcome (**Supplementary Figure S8B**). The combined clinical and ecological model again showed the strongest overall discrimination, with a mean macro-AUC of 0.60 across repeated cross-validation. This was significantly higher than the clinical-only model and the clinical, ecological, and functional model after FDR correction (both $q = 0.012$), while the comparison with the ecological-only model did not remain significant after correction ($q = 0.055$).

Discussion

The cervicovaginal microbiome is increasingly viewed as part of the local microenvironment that may shape HPV persistence and cervical disease. In this Kenyan cohort, our findings point to a relationship between cervical lesion severity and several levels of microbial organization, from community composition to ecological structure and inferred metabolic function. The clearest pattern was not a single disease-associated taxon, but a broader shift in microbial ecology. *Lactobacillus*-associated community structure decreased with increasing cytological severity, while distinct non-*Lactobacillus* ecological states showed different taxonomic and inferred functional profiles (30, 31).

The identification of distinct cervicotypes is consistent with the established view that vaginal microbial communities occupy a limited number of recurrent configurations, while also highlighting considerable heterogeneity within these broad states (32–34). In our cohort, *Lactobacillus*-dominated communities included profiles characterized by *L. iners* and other lactobacilli, whereas *Lactobacillus*-depleted communities separated into several polymicrobial configurations enriched for anaerobic taxa. One community was distinguished by *Delftia*, which has been described as a potential contaminant and an emerging opportunistic pathogen in low-biomass studies (35). In our cohort, however, it was concentrated within a distinct polymicrobial cervicotype and co-occurrence module. *Delftia* has also been reported in vaginal microbiome studies and cultured from vaginal specimens, supporting the biological plausibility of this ecological pattern (30, 36, 37). Further longitudinal validation will be needed to determine whether *Delftia* could contribute to risk stratification alongside established HPV- and cytology-based markers. Large vaginal microbiome studies have likewise identified reproducible co-occurrence structures within these communities (15). In our cohort, the greater representation of CT4 states among women with SIL and the higher prevalence of higher-risk HPV categories in some of these communities support an association between specific ecological configurations and cervical disease. These patterns, nevertheless, should not be interpreted as healthy or unhealthy microbiome states, particularly in African populations where polymicrobial vaginal communities are more frequent (22, 27).

To our knowledge, this is the first study to characterize co-occurring microbial ecological modules in relation to HPV infection and cervical lesion severity. Increasing cytological severity was inversely associated with *Lactobacillus* and positively associated with *Delftia*, while *T. vaginalis* infection was associated with *Mycoplasma* (12). These associations further illustrate how genital coinfections may contribute to the ecological heterogeneity observed across cervical disease states. The negative association between the *Lactobacillus*-associated module and both diagnosis severity and HPV infection further suggests that *Lactobacillus* communities may be more informative than the abundance of a single *Lactobacillus* species. The genotype-resolved HPV data also allowed us to examine microbial ecology across HPV risk groups, although these cross-sectional associations should be interpreted cautiously given the modest sample size within individual risk strata. This is consistent with longitudinal evidence showing that *Lactobacillus*-dominated vaginal communities are associated with a greater probability of disease regression, while depletion of lactobacilli and enrichment of anaerobic taxa such as *Gardnerella* and *Prevotella* are associated with persistent disease (18, 19, 38, 39). Our findings differ

from a simple model in which increasing cervical disease is associated with enrichment of classical anaerobic taxa. The inverse association of *G. vaginalis* with lesion severity, together with the lack of a significant association between the broader *Gardnerella*-associated module and cytological diagnosis, emphasizes that the ecological significance of individual bacterial species and strains depends on the community in which they occur (40–42). This may help explain why associations for taxa such as *Gardnerella* and *L. iners* have varied between studies (41, 43).

The inferred functional analyses extend this ecological interpretation. Thirty-five inferred pathways were associated with increasing cytological severity, with most showing higher inferred abundance from NILM to HSIL. These included inferred pathways involved in carbohydrate and organic acid degradation, amino acid metabolism, fatty acid metabolism, aromatic compound degradation, and cofactor biosynthesis, whereas UMP biosynthesis II and III decreased with increasing severity. Similar functional changes have been reported in metagenomic studies of HPV-associated cervical disease, including alterations in carbohydrate, amino acid, nucleotide, and cofactor metabolism (20, 23, 44, 45). Of note, nearly half of the inferred diagnosis-associated pathways were also associated with cervicotype, and more than one third overlapped with pathways associated with the *Delftia*-associated module, whereas no comparable overlap was observed with the *Lactobacillus*-associated module. This pattern suggests that the inferred functional changes associated with cytological severity are linked preferentially to particular non-*Lactobacillus* ecological configurations and are not explained solely by depletion of *Lactobacillus*-associated communities. The composite diagnosis-associated pathway score also differed markedly among cervicotypes. CT4C showed significantly higher scores than CT2, CT3, CT4A, and CT4B, consistent with its distinct polymicrobial composition and association with the *Delftia*-associated module. CT1 also showed a significantly higher score than CT2, despite both representing predominantly *Lactobacillus*-dominated communities. The contrast between *L. crispatus*-dominated CT1 and *L. iners*-dominated CT2 suggests that *Lactobacillus*-dominated communities can differ substantially in their inferred metabolic profiles (10, 46, 47).

These findings should nevertheless be interpreted in the context of how functional profiles were derived. PICRUSt2 infers microbial gene content from 16S rRNA ASV composition and reference genomes, so the inferred pathways are not independent measurements of microbial function (48). Their association with cervicotypes and ecological modules may therefore partly reflect the same taxonomic information used to define those community structures. This limitation is particularly relevant for the vaginal microbiome, where direct comparisons with shotgun metagenome sequencing have shown only modest overall agreement and substantial variation in inference accuracy according to community composition (49). In particular, PICRUSt2 performs considerably less well in *L. iners*-dominated communities than in *L. crispatus*-dominated communities (49). The difference observed between CT1 and CT2 may therefore reflect both genuine biological variation and community-dependent differences in metagenome inference accuracy. Direct metagenomic, metatranscriptomic, or metabolomic studies will be needed to confirm the underlying microbial functions (23, 41).

The strong relationship between the *Delftia*-associated module and inferred diagnosis-associated pathways also raises potential links between microbial ecology and metabolic changes implicated in cervical disease. The module showed broad positive correlations with inferred metabolic pathways, consistent with a community profile with greater potential for nutrient utilization and metabolic turnover. These pathway classes closely resemble metabolic alterations reported in cervicovaginal metabolomic studies of HPV persistence and cervical lesions, where changes in amino acid, lipid, carbohydrate, nucleotide, redox, and tricarboxylic acid cycle metabolism have been repeatedly observed (23, 50, 51). Such changes could influence the cervicovaginal microenvironment through altered production of organic acids, lipid intermediates, redox-active metabolites, and amino acid-derived compounds, with potential consequences for epithelial barrier integrity, inflammatory signaling, and local immune regulation. Tryptophan metabolism may be especially relevant in this context, as *Lactobacillus*-associated indole derivatives can support epithelial and immune homeostasis through aryl hydrocarbon receptor signaling, whereas disruption of these pathways has been associated with HPV persistence and cervical disease (52, 53). The inverse association between the *Gardnerella*-associated module and predicted L-tryptophan degradation may therefore reflect a distinct metabolic configuration within this ecological state (51). Still, these associations should not be attributed to *Delftia*, *Gardnerella*, or any single bacterial taxon, as the modules represent consortia of co-occurring bacteria with collective metabolic potential.

The exploratory discriminatory analyses also highlighted the limits of the microbiome signal in this cohort. Clinical variables alone showed little discriminatory ability, whereas ecological features produced the strongest ordinal model when combined with clinical information, supporting the possibility that higher-order microbial community organization contains disease-relevant discriminatory information (10, 18). Adding taxonomic or inferred functional features did not consistently improve performance, suggesting that increasingly detailed microbiome features alone may not be sufficient to capture the biological heterogeneity underlying cervical lesion severity. The modest discrimination observed here may therefore reflect not only sample size, heterogeneity within LSIL and HSIL, and the absence of several host and behavioral covariates, but also the contribution of other components of the cervicovaginal microenvironment. Integrating microbial ecology with host transcriptomic, metabolomic, epigenomic, and immune profiles may provide a more complete representation of the biological states associated with HPV persistence and cervical disease severity (16, 23, 42).

Several limitations should be considered when interpreting our findings. The cross-sectional design prevents determination of whether microbiome changes precede cervical lesion development, arise as a consequence of HPV-associated epithelial changes, or reflect bidirectional interactions. Longitudinal studies have linked vaginal microbiome composition to subsequent lesion outcomes, but causal direction remains unresolved (17–19, 39). The cohort was also relatively small, particularly within the LSIL and HSIL groups, and detailed participant-level information on age, HIV status, pregnancy, menopausal status, recent antibiotic exposure, contraceptive use, and other potential confounders was unavailable. This is particularly relevant because many of these factors are known to influence vaginal microbial composition. In addition, 16S rRNA gene sequencing provides limited species and strain

resolution. Lastly, the ecological modules and exploratory feature modeling were derived within this cohort and require external validation in independent African and other populations.

Despite these limitations, our study provides an integrated assessment of the cervicovaginal microbiome in relation to HPV-associated cervical disease in a population that remains underrepresented in microbiome research. Cervical lesion severity was associated with coordinated differences in taxonomic composition, ecological community structure, and inferred functional potential. The most consistent signal was not a single disease-associated organism, but a broader shift in ecological structure, including loss of a *Lactobacillus*-associated community pattern and the emergence of distinct inferred functional signatures linked to other microbial modules. Future longitudinal studies integrating host immunity, direct metagenomic profiling, and metabolite measurements will be important for determining whether these ecological states actively shape HPV persistence and lesion progression or reflect changes in the cervicovaginal microenvironment. These findings support further investigation of the cervicovaginal microbiome as an interacting microbial ecosystem whose organization and inferred functional potential are associated with cervical disease severity.

Methods

Ethical statement

The study was conducted in accordance with the principles of the Declaration of Helsinki and was approved by Jaramogi Oginga Odinga University Of Science and Technology Institutional Ethics Review Committee (JOOUST IERC), Kenya [approval/reference number: ERC/09/2025(2)]. All participants received information about the study and provided written informed consent before enrolment and sample collection. All additional approvals required for the transfer and analysis of samples in Belgium and Sweden were obtained from the relevant institutional and national authorities.

Study participants

Cervical samples were collected from women attending Gynocare Women's and Fistula Hospital, Kenya, in January 2026 as part of a community outreach initiative aimed at increasing awareness of HPV infection and cervical cancer. Each participant provided two cervical samples: one sample was collected for HPV testing and subsequent molecular analyses, including cervical microbiome profiling, and a separate sample was collected for cytological evaluation. Samples intended for HPV testing and molecular analyses were obtained using a Sumbow Orange Sterile Stick (S.M.I. Co. Ltd/KEMSA), placed in eNAT transport and preservation medium, stored at -20°C , and shipped to the AML laboratory in Belgium for molecular testing.

Cytological evaluation was performed locally using direct cervical smears, and diagnoses were classified as negative for intraepithelial lesion or malignancy (NILM), low-grade squamous intraepithelial lesion (LSIL), high-grade squamous intraepithelial lesion (HSIL), or insufficient for evaluation. Of the 100 participants enrolled, 98 had evaluable cytology. Microbiome sequencing data was available for 93

participants, comprising 52 NILM, 18 LSIL, 21 HSIL, and two samples with insufficient cytology. The two samples with insufficient cytology were excluded from analyses involving cytological diagnosis, resulting in a primary analytical cohort of 91 women. Additional participant-level clinical and demographic variables, including age, HIV status, pregnancy status, menopausal status, recent antibiotic use, and contraceptive use, were not available for the present analysis.

DNA extraction, HPV genotyping, and *Trichomonas vaginalis* detection

Samples were processed, and DNA extraction was performed as previously described (54-56). Extracted DNA was subjected to HPV genotyping using the RIATOL quantitative real-time PCR assay, a type-specific multiplex assay that detects individual HPV genotypes through amplification of genotype-specific viral targets (HPV16 E7, HPV18 E7, HPV31 E6, HPV33 E6, HPV35 E6, HPV39 E7, HPV45 E7, HPV51 E6, HPV52 E7, HPV56 E7, HPV58 E6, HPV59 E7, HPV66 E6, HPV68 E7) on the LightCycler 480 (Roche Molecular Systems) (54-56). Human beta-globin amplification was included as an internal quality control to confirm adequate cellular DNA content. Samples were reported as HPV-negative when all 14 hrHPV targets were undetected and beta-globin levels exceeded 0.12 ng/ μ L. The clinically optimized positivity threshold corresponded to a log₁₀ viral concentration above 6.493 copies/mL (54). HPV genotypes were recorded individually, allowing samples to be classified as HPV-negative or as having single or multiple HPV infections. Detection of *T. vaginalis* DNA was performed using TaqMan-based real-time quantitative PCR (57). *T. vaginalis* results were recorded as positive or negative.

HPV risk group classification

HPV genotypes were classified into hierarchical risk groups according to their long-term risk of cervical intraepithelial neoplasia grade 3 or worse (CIN3+), based on Thomsen *et al.* (25). Briefly, HPV16 was classified as the highest-risk genotype; HPV18, HPV31, and HPV33 as high risk; HPV45, HPV52, and HPV58 as medium risk; and HPV35, HPV39, HPV51, HPV56, HPV59, HPV66, and HPV68 as lower-risk genotypes. Samples containing multiple HPV genotypes were assigned to the highest-risk category represented. HPV-negative samples were classified separately as "Negative" (**Supplementary Table S1**).

16S rRNA gene library preparation and sequencing

Aliquots of extracted DNA were shipped from the AML laboratory in Belgium to the International HPV Reference Center in Stockholm, Sweden, for cervicovaginal microbiome analysis. Microbiome profiling was performed using 16S rRNA gene amplicon sequencing. Library preparation and sequencing were conducted at the National Genomics Infrastructure, SciLifeLab, Stockholm, Sweden. The V3–V4 hypervariable regions of the bacterial 16S rRNA gene were amplified using a two-step PCR protocol. In the first PCR, locus-specific primers 341F and 805R containing 5' overhang adapter sequences were used. Amplification was performed using KAPA HiFi HotStart ReadyMix (Roche) under the following conditions: initial denaturation at 98°C for 2 min; 28 cycles of denaturation at 98°C for 20 s, annealing at 54°C for 20 s, and extension at 72°C for 15 s; followed by a final extension at 72°C for 2 min.

PCR products were purified using magnetic bead-based purification. A second PCR was performed to add dual indices and Illumina sequencing adapters, followed by an additional purification step. Library quality was evaluated based on fragment-size distribution and DNA concentration, and libraries with concentrations above 1 ng/μL were considered to have passed quality control. No-template controls and positive controls were included to assess potential contamination and assay performance. Libraries were sequenced on an Illumina NextSeq 2000 platform (Illumina, USA) using paired-end 151-bp reads and dual 10-bp index reads on a P1 flow cell, according to the manufacturer's instructions.

Bioinformatic analysis

Raw sequencing data were processed using the nf-core/ampliseq pipeline, version 2.17.0. Quality filtering, denoising, chimera removal, and amplicon sequence variant inference were performed using DADA2. Taxonomic classification was conducted using the SILVA reference database, version 138.2. Non-bacterial sequences were removed before downstream analysis. Samples with fewer than 10,000 reads after quality filtering were excluded. The resulting ASV abundance table, representative sequences, and taxonomic assignments were used for downstream microbiome analyses.

Microbiome data analysis

All downstream analyses were performed in R v4.3.3 using the packages phyloseq, vegan, ggplot2, pheatmap, tidyverse, MaAsLin2, SpiecEasi, glmnet, ordinalNet, pROC, and ComplexUpset. Alpha diversity was evaluated using the observed richness and Shannon indices. Taxon relative abundance was calculated within each sample by dividing the read count assigned to each taxon by the total number of bacterial reads retained for that sample. Microbial community structure was explored using principal coordinate analysis (PCoA) based on Bray–Curtis dissimilarities. Taxonomic composition was summarized at the genus and species levels and visualized using relative abundance heatmaps. Cervicotypes (CTs) were identified using unsupervised hierarchical clustering of species-level relative abundance profiles, using Pearson correlation as the distance metric and average linkage. Species-level assignments were retained when available, whereas remaining taxa were represented at the genus level. The resulting clusters were assigned cervicotype labels based on their dominant taxa, overall community composition, and correspondence with previously described cervicovaginal community configurations (10, 32). Community organization was additionally visualized using t-distributed stochastic neighbor embedding (t-SNE). Microbial co-occurrence networks were inferred using SparCC and clustered into microbial ecological modules using the Markov Cluster Algorithm (MCL) as described previously (15). Module eigentaxa were calculated to summarize the dominant ecological patterns within each module and were used for downstream association analyses.

Differential abundance and inferred functional analyses

Differential abundance analyses were performed using MaAsLin2 (58), a multivariable framework for microbiome association testing with FDR control among high-dimensional microbial features. Independent MaAsLin2 models were fitted for cytological diagnosis (categorical), HPV infection,

HPV16/18 infection, *T. vaginalis* infection, and ecological module scores. Taxa detected in at least 10% of samples and exhibiting a mean relative abundance of at least 0.1% were retained for analysis. To assess taxa associated with disease severity, cytological diagnosis was additionally modeled as an ordinal variable (NILM–LSIL–HSIL), allowing identification of taxa whose abundance changed with increasing cervical lesion severity. Predicted microbial functional potential was inferred from ASV profiles using PICRUSt2 (48). We mapped inferred gene families to MetaCyc pathways and analyzed pathway abundances using MaAsLin2 to identify associations with cervical lesion severity, cervicotypes, HPV-related variables, *T. vaginalis* infection, and microbial ecological modules. To summarize coordinated inferred functional alterations associated with cytological diagnosis severity, a composite diagnosis-associated pathway score was calculated for each sample as the mean standardized abundance (Z-score) of pathways positively associated with cytological diagnosis severity minus the mean standardized abundance of pathways negatively associated with cytological diagnosis severity. Overlap between inferred cytological diagnosis-associated pathways and microbial ecological features was summarized using an UpSet plot.

Exploratory discriminatory analyses

Elastic-net penalized regression was used to evaluate the discriminatory capacity of clinical, ecological, taxonomic, and inferred functional feature sets. Binary logistic regression was first used to distinguish NILM from SIL (LSIL and HSIL combined), with performance evaluated using nested five-fold cross-validation repeated 50 times and summarized by the area under the receiver operating characteristic curve (AUC). A complementary proportional-odds ordinal elastic-net model was used to model increasing cytological severity (NILM–LSIL–HSIL). Ordinal model performance was evaluated using nested five-fold cross-validation repeated 10 times and summarized using the ranked probability score (RPS), macro-averaged AUC, and AUCs at the cumulative thresholds NILM versus LSIL + HSIL and NILM + LSIL versus HSIL. Hyperparameters were selected by inner five-fold cross-validation, and taxonomic and inferred functional feature filtering was performed within the outer training folds to minimize information leakage. Performance estimates incorporating cervicotypes should be interpreted as exploratory, as these features were defined using the full cohort.

Statistical analysis

All statistical analyses were conducted in R v4.3.3. Differences in alpha diversity were evaluated using Kruskal–Wallis or Wilcoxon rank-sum tests as appropriate. Differences in microbial community composition were assessed using permutational multivariate analysis of variance (PERMANOVA). Associations between continuous variables were evaluated using Spearman correlation analyses. Differential abundance analyses were performed using MaAsLin2. For analyses involving multiple taxa, pathways, or correlations, *p*-values were adjusted using the Benjamini–Hochberg FDR correction. Unless otherwise specified, FDR-adjusted *q*-values <0.05 and *p*-values <0.05 were considered statistically significant. For exploratory high-dimensional microbiome analyses, including differential abundance and ecological module analyses, associations with FDR-adjusted *q*-values <0.10 were considered. Pairwise

comparisons of ROC curves were performed with DeLong's test. All statistical tests were two-sided unless otherwise specified.

Declarations

Data availability

The 16S rRNA gene sequencing data generated in this study have been deposited in the NCBI Sequence Read Archive under BioProject accession PRJNA1539008 and will be publicly available upon publication.

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.

Acknowledgements

The authors acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure (NGI), and UPPMAX for assistance with massively parallel sequencing and computational infrastructure. Open access funding provided by Karolinska Institute. This work was supported by Karolinska Institutet through the LMIC in Africa 2024 research grant (2024-03558), awarded to L.S.A.M. for the project “Microbiome influences on HPV and cervical cancer in Africa: a pilot study on disease development.” The funder had no role in study design, data collection, analysis, interpretation, or preparation of the manuscript.

Competing interests

The authors declare no competing interests.

Author contributions

M.A.M.: conceptualization, methodology, formal analysis, microbiome data analysis, visualization, writing—original draft, and writing—review and editing. K.N., I.A., and H.M.: investigation, clinical data collection, study coordination, and writing—review and editing. L.T. and D.V.B.: investigation, laboratory analyses, and writing—review and editing. S.S.H. and M.S.: writing—review and editing. L.S.A.M.: conceptualization, supervision, funding acquisition, interpretation, and writing—review and editing. All authors approved the final manuscript.

References

1. Molina MA, Steenbergen RDM, Pumpe A, Kenyon AN, Melchers WJG. HPV integration and cervical cancer: a failed evolutionary viral trait. *Trends in Molecular Medicine*. 2024;30(9):890–902.

2. Wei F, Georges D, Man I, Baussano I, Clifford GM. Causal attribution of human papillomavirus genotypes to invasive cervical cancer worldwide: a systematic analysis of the global literature. *The Lancet*. 2024;404(10451):435–44.
3. 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*. 2020;10.
4. Lycke KD, Steben M, Garland SM, Woo YL, Cruickshank ME, Perkins RB, et al. An updated understanding of the natural history of cervical human papillomavirus infection—clinical implications. *American Journal of Obstetrics and Gynecology*. 2025;232(5):453–60.
5. Spurgeon Megan E. The cervicovaginal microbiome: an emerging determinant of HPV infection and disease. *Journal of Virology*. 2026;100(8):e01706-25.
6. Łaniewski P, Cui H, Roe DJ, Chase DM, Herbst-Kralovetz MM. Vaginal microbiota, genital inflammation, and neoplasia impact immune checkpoint protein profiles in the cervicovaginal microenvironment. *npj Precision Oncology*. 2020;4(1):22.
7. Osafo KS, Zhang Y, Obiri-Yeboah D, Sun B, Yang H, Sun P. Advancing the elimination of cervical cancer across Africa. *Nature Communications*. 2026;17(1):3907.
8. Onywera H, Mbulawa ZZA, Brink A, Williamson A-L, Mwapagha LM. Unravelling the Biological Interplay Between Genital HPV Infection and Cervicovaginal Microbiota in Sub-Saharan Africa: Implications for Cervical (Pre)cancer Prevention. *Venereology [Internet]*. 2024; 3(4):[211 – 31 pp.].
9. Ayeni KI, Berry D, Ezekiel CN, Warth B. Enhancing microbiome research in sub-Saharan Africa. *Trends in Microbiology*. 2024;32(2):111–5.
10. Molina MA, Andralojc KM, Huynen MA, Leenders WPJ, Melchers WJG. In-depth insights into cervicovaginal microbial communities and hrHPV infections using high-resolution microbiome profiling. *npj Biofilms and Microbiomes*. 2022;8(1):75.
11. Mejía-Caballero A, Marco ML. Lactobacilli biology, applications and host interactions. *Nature Reviews Microbiology*. 2025.
12. Chen M, Qi C, Qing W, Zhou Z, Zhang Y, Chen R, et al. Vaginal microbiome and sexually-transmitted pathogens in Chinese reproductive-age women: a multicentre cross-sectional and longitudinal cohort study. *Nature Communications*. 2025;16(1):10002.
13. Mancilla V, Jimenez NR, Bishop NS, Flores M, Herbst-Kralovetz MM. The Vaginal Microbiota, Human Papillomavirus Infection, and Cervical Carcinogenesis: A Systematic Review in the Latina Population. *Journal of Epidemiology and Global Health*. 2024;14(2):480–97.
14. France MT, Ma B, Gajer P, Brown S, Humphrys MS, Holm JB, et al. VALENCIA: a nearest centroid classification method for vaginal microbial communities based on composition. *Microbiome*. 2020;8(1):166.
15. Lebeer S, Ahannach S, Gehrman T, Wittouck S, Eilers T, Oerlemans E, et al. A citizen-science-enabled catalogue of the vaginal microbiome and associated factors. *Nature Microbiology*. 2023.

16. Usyk M, Schlecht NF, Pickering S, Williams L, Sollecito CC, Gradissimo A, et al. molBV reveals immune landscape of bacterial vaginosis and predicts human papillomavirus infection natural history. *Nature Communications*. 2022;13(1):233.
17. Usyk M, Zolnik CP, Castle PE, Porras C, Herrero R, Gradissimo A, et al. Cervicovaginal microbiome and natural history of HPV in a longitudinal study. *PLoS Patho*. 2020;16(3):e1008376.
18. Molina MA, Leenders WPJ, Huynen MA, Melchers WJG, Andralojc KM. Temporal composition of the cervicovaginal microbiome associates with hrHPV infection outcomes in a longitudinal study. *BMC Infectious Diseases*. 2024;24(1):552.
19. Mitra A, MacIntyre DA, Ntritsos G, Smith A, Tsilidis KK, Marchesi JR, et al. The vaginal microbiota associates with the regression of untreated cervical intraepithelial neoplasia 2 lesions. *Nature Communications*. 2020;11(1):1999.
20. Stosic M, Molina MA, Mukhedkar D, Hassan SS, Sahlin A, Ballester CR, et al. Inferred functional microbiome features are associated with clinical trajectories of low-grade cervical lesions. *BMC Microbiology*. 2026;26(1):802.
21. Lebeau A, Bruyere D, Roncarati P, Peixoto P, Hervouet E, Cobraiville G, et al. HPV infection alters vaginal microbiome through down-regulating host mucosal innate peptides used by Lactobacilli as amino acid sources. *Nat Commun*. 2022;13(1):1076.
22. Molina MA. From dysbiosis to mechanisms: Why cervicovaginal microbiome-HPV studies must catch up with biology. *PLOS Pathogens*. 2026;21(12):e1013830.
23. Molina MA, Dai W. Microbiome-metabolome crosstalk in HPV pathogenesis: from ecosystem dynamics to translational biomarkers. *Computational and Structural Biotechnology Journal*. 2026;0(ja).
24. Hidjo M, Mukhedkar D, Masimirembwa C, Lei J, Arroyo Mühr LS. Cervical cancer microbiome analysis: comparing HPV 16 and 18 with other HPV types. *Scientific Reports*. 2024;14(1):22014.
25. Thomsen LT, Frederiksen K, Munk C, Junge J, Iftner T, Kjaer SK. Long-term risk of cervical intraepithelial neoplasia grade 3 or worse according to high-risk human papillomavirus genotype and semi-quantitative viral load among 33,288 women with normal cervical cytology. *International Journal of Cancer*. 2015;137(1):193–203.
26. Hamar B, Teutsch B, Hoffmann E, Hegyi P, Váradi A, Nyirády P, et al. *Trichomonas vaginalis* infection is associated with increased risk of cervical carcinogenesis: A systematic review and meta-analysis of 470 000 patients. *International Journal of Gynecology & Obstetrics*. 2023;163(1):31–43.
27. Condori-Catachura S, Ahannach S, Ticlla M, Kenfack J, Livo E, Anukam KC, et al. Diversity in women and their vaginal microbiota. *Trends in Microbiology*. 2025;33(11):1163–72.
28. Jie Z, Liang W, Ding Q, Liu X, Zhang Y, Chen N, et al. Genomic landscape of the human vaginal microbiome is linked to host genetics and population of origin. *Nature Genetics*. 2026;58(7):1595–609.
29. Silverstein MR, Bhatnagar JM, Segrè D. Metabolic complexity drives divergence in microbial communities. *Nature Ecology & Evolution*. 2024;8(8):1493–504.

30. Liu Z, Wu Y, Wang C, Hou N, Liu X. Changes in the vaginal microbiome in female patients with condyloma acuminatum and its impact on persistent HPV infection. *Scientific Reports*. 2025;15(1):34996.
31. Tango CN, Seo S-S, Kwon M, Lee D-O, Chang HK, Kim MK. Taxonomic and Functional Differences in Cervical Microbiome Associated with Cervical Cancer Development. *Scientific Reports*. 2020;10(1):9720.
32. Munoz A, Hayward MR, Bloom SM, Rocafort M, Ngcapu S, Mafunda NA, et al. Modeling the temporal dynamics of cervicovaginal microbiota identifies targets that may promote reproductive health. *Microbiome*. 2021;9(1):163.
33. Mazibuko-Motau N, Sobia P, Xu J, Elsherbini JA, San JE, Lewis L, et al. Vaginal microbial shifts are unaffected by oral pre-exposure prophylaxis in South African women. *Scientific Reports*. 2022;12(1):16187.
34. Zhu M, Frank MW, Radka CD, Jeanfavre S, Xu J, Tse MW, et al. Vaginal *Lactobacillus* fatty acid response mechanisms reveal a metabolite-targeted strategy for bacterial vaginosis treatment. *Cell*. 2024;187(19):5413-30.e29.
35. Ryan MP, Tony Pembroke J. *Delftia* spp as Opportunistic Pathogens: a narrative review. *New Microbes and New Infections*. 2026;70:101705.
36. Wu M, Gao J, Wu Y, Li Y, Chen Y, Zhao F, et al. Characterization of vaginal microbiota in Chinese women with cervical squamous intra-epithelial neoplasia. *International Journal of Gynecological Cancer*. 2020;30(10):1500–4.
37. Zhang L, Zhang X, Bai H, Li T, Zhang Z, Zong X, et al. Characterization and Genome Analysis of the *Delftia lacustris* Strain LzhVag01 Isolated from Vaginal Discharge. *Current Microbiology*. 2024;81(8):232.
38. Mitra A, MacIntyre DA, Lee YS, Smith A, Marchesi JR, Lehne B, et al. Cervical intraepithelial neoplasia disease progression is associated with increased vaginal microbiome diversity. *Scientific Reports*. 2015;5(1):16865.
39. Banila C, Ladoukakis E, Scibior-Bentkowska D, Santiago LR, Reuter C, Kleeman M, et al. A longitudinal pilot study in pre-menopausal women links cervicovaginal microbiome to CIN3 progression and recovery. *Communications Biology*. 2025;8(1):883.
40. France MT, Fu L, Rutt L, Yang H, Humphrys MS, Narina S, et al. Insight into the ecology of vaginal bacteria through integrative analyses of metagenomic and metatranscriptomic data. *Genome Biology*. 2022;23(1):66.
41. Holm JB, France MT, Gajer P, Ma B, Brotman RM, Shardell M, et al. Integrating compositional and functional content to describe vaginal microbiomes in health and disease. *Microbiome*. 2023;11(1):259.
42. Dos Santos SJ, Copeland C, Macklaim JM, Reid G, Gloor GB. Vaginal metatranscriptome meta-analysis reveals functional BV subgroups and novel colonisation strategies. *Microbiome*. 2024;12(1):271.

43. Petrova MI, Reid G, Vaneechoutte M, Lebeer S. *Lactobacillus iners*: friend or foe? *Trends Microbiol.* 2017;25(3):182–91.
44. İlhan ZE, Łaniewski P, Thomas N, Roe DJ, Chase DM, Herbst-Kralovetz MM. Deciphering the complex interplay between microbiota, HPV, inflammation and cancer through cervicovaginal metabolic profiling. *eBioMedicine.* 2019;44:675–90.
45. Jimenez NR, Mancilla V, Łaniewski P, Herbst-Kralovetz MM. Immunometabolic Contributions of Atopobiaceae Family Members in Human Papillomavirus Infection, Cervical Dysplasia, and Cancer. *The Journal of Infectious Diseases.* 2025;232(4):767–78.
46. Stosic M, Molina MA, Mukhedkar D, Hassan SS, Sahlin A, Ballester CR, et al. Inferred functional microbiome features are associated with clinical trajectories of low-grade cervical lesions. 2026.
47. Molina MA, Melchers WJG, Andralojc KM, Leenders WPJ, Huynen MA. Longitudinal analysis on the ecological dynamics of the cervicovaginal microbiome in hrHPV infection. *Computational and Structural Biotechnology Journal.* 2023;21:4424–31.
48. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, et al. PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology.* 2020;38(6):685–8.
49. Carter Kayla A, Fodor Anthony A, Balkus Jennifer E, Zhang A, Serrano Myrna G, Buck Gregory A, et al. Vaginal Microbiome Metagenome Inference Accuracy: Differential Measurement Error according to Community Composition. *mSystems.* 2023;8(2):e01003-22.
50. Yu T, Gao S, Jin F, Yan B, Wang W, Wang Z. Characteristics of the vaginal microbiota and vaginal metabolites in women with cervical dysplasia. *Frontiers in Cellular and Infection Microbiology.* 2024;Volume 14–2024.
51. Kawasaki R, Kukimoto I, Tsukamoto T, Nishio E, Iwata A, Fujii T. Cervical mucus can be used for metabolite screening in cervical cancer. *Cancer Science.* 2024;115(11):3672–81.
52. Berard AR, Brubaker DK, Birse K, Lamont A, Mackelprang RD, Noël-Romas L, et al. Vaginal epithelial dysfunction is mediated by the microbiome, metabolome, and mTOR signaling. *Cell Reports.* 2023;42(5).
53. Zhang Z, Zheng K, Zhang Z, Cao L, Lin L, Sun W, et al. *Lactobacillus gasseri* LGV03-derived indole-3-lactic acid ameliorates immune response by activating aryl hydrocarbon receptor. *Microbial Cell Factories.* 2025;24(1):34.
54. Benoy I, Xu L, Vanden Broeck D, Poljak M, Oštrbenk Valenčak A, Arbyn M, et al. Using the VALGENT-3 framework to assess the clinical and analytical performance of the RIATOL qPCR HPV genotyping assay. *Journal of Clinical Virology.* 2019;120:57–62.
55. Depuydt CE, Benoy IH, Beert JFA, Criel AM, Bogers JJ, Arbyn M. Clinical Validation of a Type-Specific Real-Time Quantitative Human Papillomavirus PCR against the Performance of Hybrid Capture 2 for the Purpose of Cervical Cancer Screening. *Journal of Clinical Microbiology.* 2012;50(12):4073–7.
56. Dhillon SK, Chung PYJ, Padalko E, Praet M, Pereira AR, Redzic N, et al. Intra- and interlaboratory reproducibility of the RIATOL qPCR HPV genotyping assay. *Journal of Medical Virology.* 2023;95(9):e29093.

57. Donders GGG, Depuydt CE, Bogers J-P, Vereecken AJ. Association of *Trichomonas vaginalis* and Cytological Abnormalities of the Cervix in Low Risk Women. PLOS ONE. 2014;8(12):e86266.
58. Mallick H, Rahnavard A, McIver LJ, Ma S, Zhang Y, Nguyen LH, et al. Multivariable association discovery in population-scale meta-omics studies. PLOS Computational Biology. 2021;17(11):e1009442.

Figures

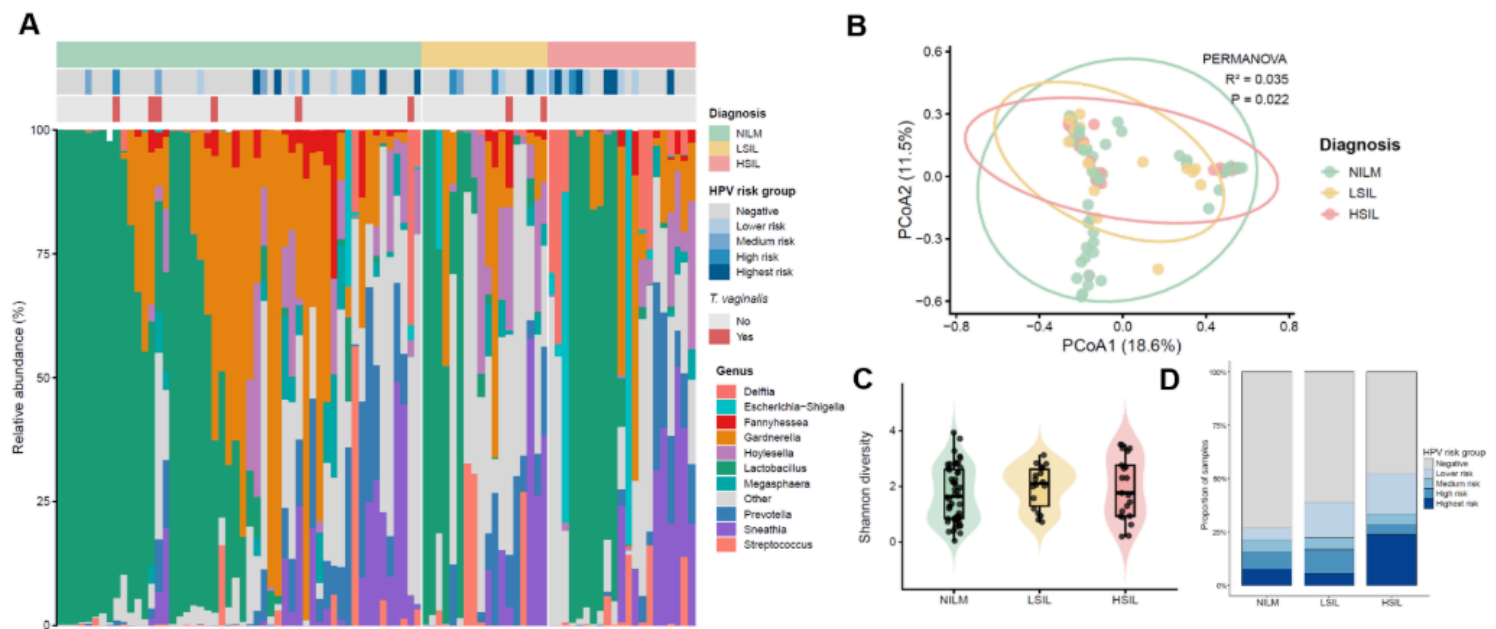


Figure 1

Overview of the cervicovaginal microbiome among clinical characteristics. (A) Relative abundance of the dominant bacterial genera across the cohort, ordered by cytological diagnosis. Annotation bars indicate cytological diagnosis, HPV risk groups (see Supplementary Table S1), and *T. vaginalis* infection status. Genera contributing <1% of the total relative abundance are grouped as Other. (B) Principal coordinates analysis (PCoA) based on Bray–Curtis dissimilarity showing differences in overall microbial community composition according to cytological diagnosis. Ellipses represent the 95% confidence interval around each group centroid. (C) Shannon diversity among cytological diagnoses. Points represent individual samples; boxplots indicate the median and interquartile range. (D). Distribution of HPV risk groups across cytological diagnoses, displayed as the proportion of samples within each diagnosis category. NILM, negative for intraepithelial lesion or malignancy; LSIL, low-grade squamous intraepithelial lesion; HSIL, high-grade squamous intraepithelial lesion.

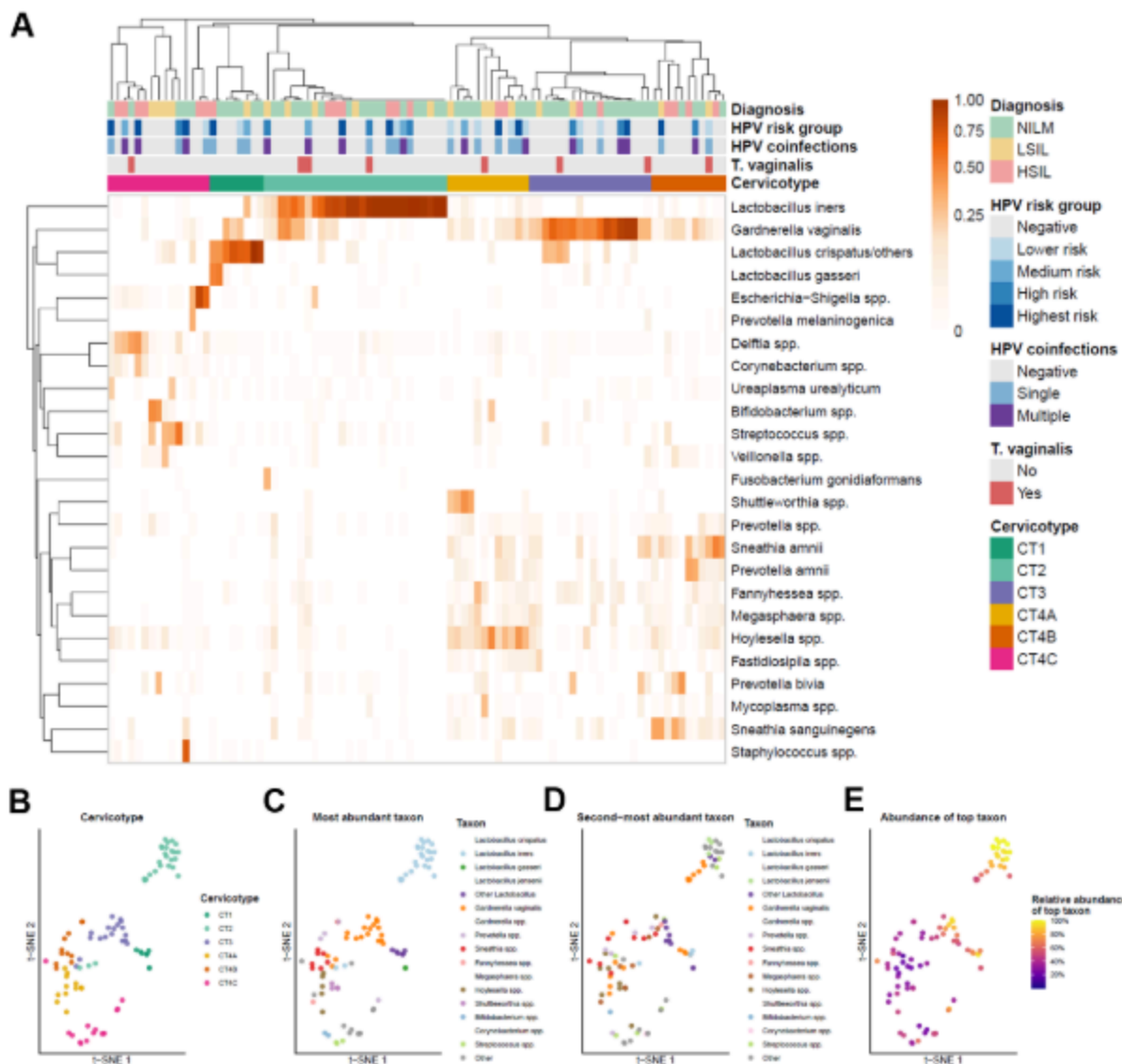


Figure 2

Unsupervised clustering identifies distinct cervicovaginal microbial ecological states. (A) Heatmap with relative abundances of the 25 most abundant bacterial species across the cohort, with hierarchical clustering applied to both samples (columns) and taxa (rows). Samples are annotated according to cytological diagnosis (NILM, LSIL, HSIL), HPV risk groups (see Supplementary Table S1), HPV coinfections (negative, single, or multiple HPV infections), *T. vaginalis* infection status, and cervicotypes (CT). Heatmap colors indicate species relative abundance, with darker shades representing higher abundance. t-distributed stochastic neighbor embedding (t-SNE) projections of the same samples colored by cervicotype (B), the most abundant taxon (C), the second most abundant taxon (D), and the relative abundance of the dominant taxon (E). NILM, negative for intraepithelial lesion or malignancy; LSIL, low-grade squamous intraepithelial lesion; HSIL, high-grade squamous intraepithelial lesion.

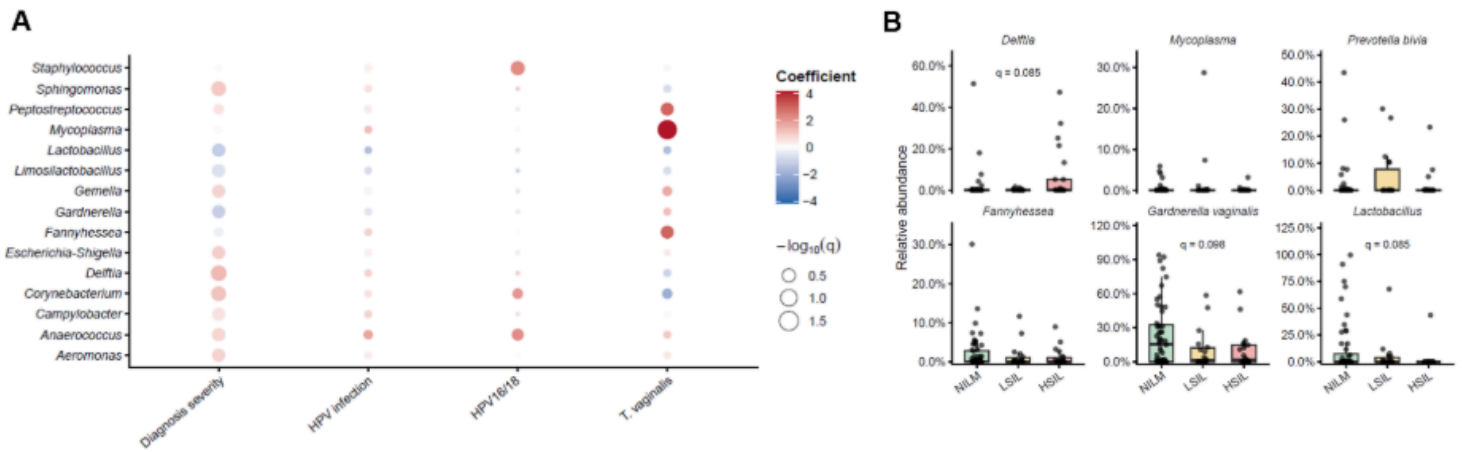


Figure 3

Differential bacterial abundance analysis among clinical characteristics. (A) Bubble plot summarizing associations between bacterial genera and clinical variables identified with MaAsLin2. Associations are shown for cytological diagnosis severity (NILM–LSIL–HSIL), HPV infection, HPV16/18 infection, and *T. vaginalis* infection. Bubble color indicates the direction and magnitude of the regression coefficient (red, positive association; blue, negative association), while bubble size corresponds to statistical significance. (B) Relative abundances of selected bacterial taxa associated with cytological diagnosis severity. Points represent individual samples, boxplots indicate the median and interquartile range, and whiskers extend to 1.5× the interquartile range. NILM, negative for intraepithelial lesion or malignancy; LSIL, low-grade squamous intraepithelial lesion; HSIL, high-grade squamous intraepithelial lesion.

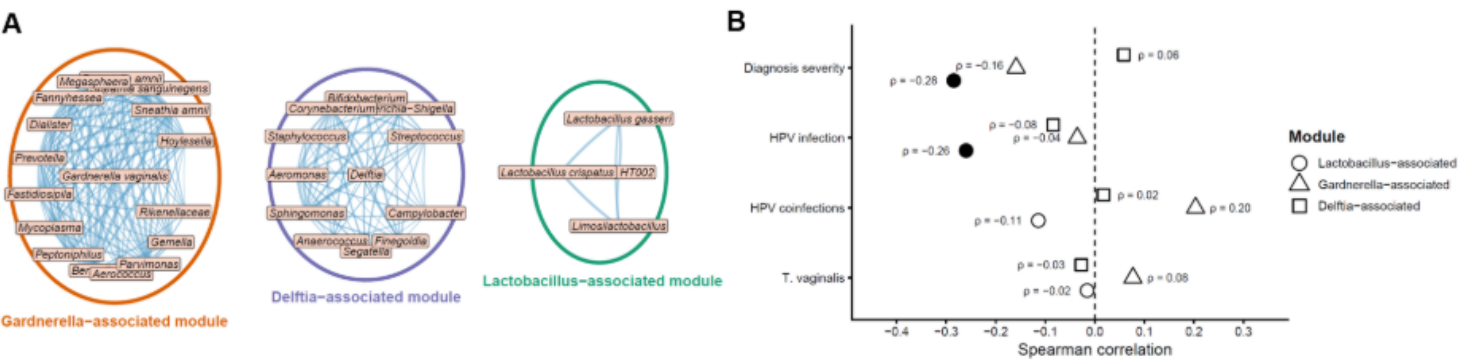


Figure 4

Microbial co-occurrence modules show distinct associations with cervical disease and HPV-related characteristics. (A) SparCC co-occurrence networks showing three microbial modules identified by Markov Cluster Algorithm (MCL) clustering: the *Gardnerella*-associated, *Delftia*-associated, and *Lactobacillus*-associated modules. Nodes represent bacterial taxa, and edges indicate positive SparCC associations above the predefined correlation threshold. (B) Associations between standardized module eigentaxon scores and clinical characteristics, including cytological diagnosis severity (NILM–LSIL–HSIL), HPV infection, HPV coinfections, and *T. vaginalis* infection, assessed using Spearman correlation.

Black and white symbols indicate $q < 0.05$ and $q \geq 0.10$, respectively. Values shown next to each point are the corresponding Spearman correlation coefficients (ρ).

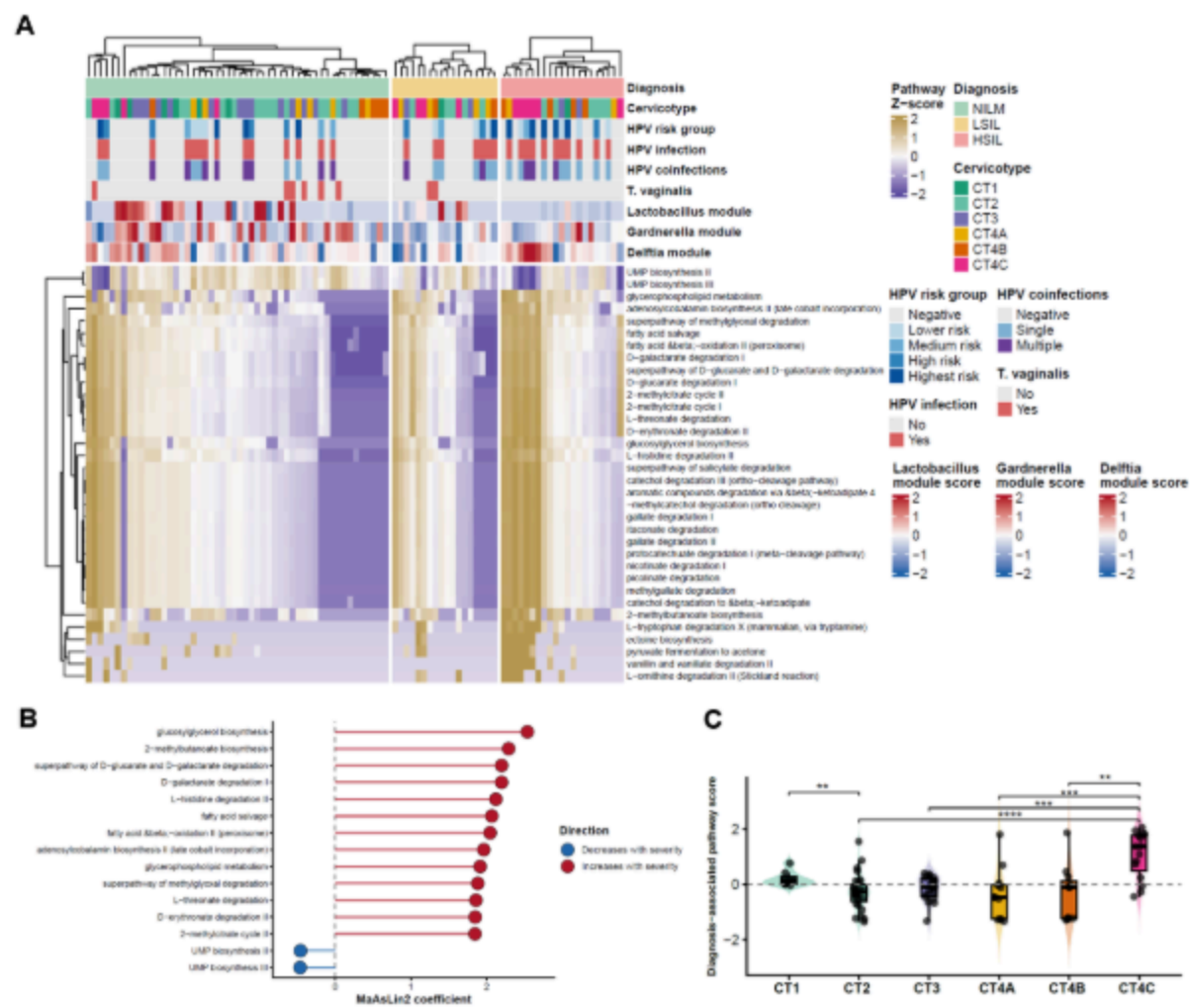


Figure 5

Inferred cytological diagnosis-associated metabolic pathways reflect cervicovaginal microbial ecological states. (A) Heatmap of PICRUSt2-inferred MetaCyc pathways significantly associated with cytological diagnosis in MaAsLin2 analyses after FDR correction ($q < 0.05$). Predicted pathway abundances were log10-transformed and standardized within each pathway as Z-scores. Brown indicates relatively higher predicted pathway abundance, and purple indicates relatively lower predicted pathway abundance. Samples are grouped by cytological diagnosis (NILM, LSIL, and HSIL) and hierarchically clustered within each group using average linkage; pathways are clustered in the same way. Annotations indicate cytological diagnosis, cervicotypes, HPV risk group, HPV infection, HPV coinfections, *T. vaginalis* infection, and standardized eigentaxon scores for the *Lactobacillus*, *Gardnerella*, and *Delftia* microbial modules. (B) Top inferred MetaCyc pathways associated with cytological diagnosis severity (NILM–LSIL–HSIL), ranked by MaAsLin2 effect size. Positive coefficients (red) indicate pathways with increasing inferred abundance across the ordinal cytological diagnosis severity, whereas negative coefficients (blue) indicate pathways with decreasing inferred abundance. (C) Composite

diagnosis-associated functional pathway score across cervicotypes (CT), calculated as the mean standardized abundance of pathways positively associated with cytological diagnosis severity minus the mean standardized abundance of pathways negatively associated with cytological diagnosis severity. Differences between cervicotypes were assessed using Kruskal–Wallis test followed by FDR correction. ** $q < 0.01$, *** $q < 0.001$, **** $q < 0.0001$. NILM, negative for intraepithelial lesion or malignancy; LSIL, low-grade squamous intraepithelial lesion; HSIL, high-grade squamous intraepithelial lesion.

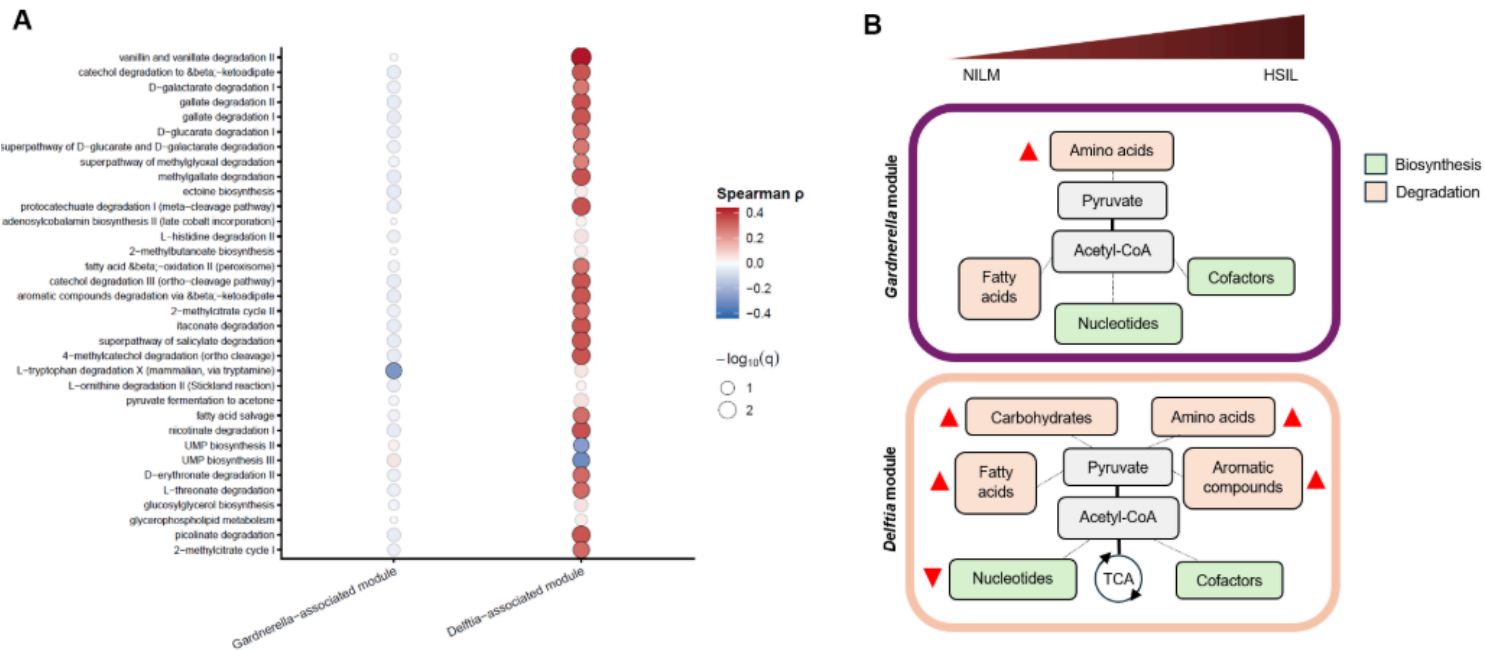


Figure 6

Correlation of inferred cytological diagnosis-associated metabolic pathways with microbial ecological modules. (A) Bubble plot showing Spearman correlations between the *Gardnerella* and *Delftia* microbial module scores and PICRUSt2-inferred MetaCyc pathways. Bubble color indicates the direction and magnitude of the Spearman correlation coefficient (ρ), with red representing positive correlations and blue representing negative correlations. Bubble size corresponds to statistical significance. (B) Conceptual representation summarizing the major inferred metabolic pathway classes associated with each microbial module based on the correlations shown in panel A. Colored boxes denote inferred pathway classes (biosynthesis or degradation). Red triangles indicate the direction in which the corresponding inferred diagnosis-associated pathways change with increasing cytological severity (NILM–LSIL–HSIL).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterials.pdf](#)