Applied Microbiology Open Access

OPEN ACCESS Freely available online

Research Article

A Metagenomics-Based Study on Functional Associations between Vaginal Microbiota and Phenotypes of HPV-Infected Patients

Qi Chen[1,3], Jhih-Hua Jhong[2], Shangfu Li[1,3], Chien-Hsun Huang[4*], Tzong-Yi Lee[1,3*]

[1]Department of Computational Biology, The Chinese University of Hong Kong, Shenzhen, China; [2]Department of Computer Science and Information Engineering, National Central University, Taoyuan, Taiwan; [3]Department of Life and Health Sciences, The Chinese University of Hong Kong, Shenzhen, Guangdong, China; [4]Department of Obstetrics and Gynecology, Taoyuan General Hospital, Ministry of Health&Welfare, Taoyuan, Taiwan

ABSTRACT

Aim: In this study, we focus on the complex and mutable ecological of vaginal microbiome to expound the association between phenotype.

Methods and results: In a study of 39 Taiwan women, 4 microbial community types were identified by initial unsupervised clustering analysis base on the microbial abundance using 16S rRNA sequencing data. We found that species diversity was different among these community types, which was mainly reflected in evenness ($p<0.001$). Compared to other community types, low evenness group shown high relative abundance of *Lactobacillus* , higher Human Papilloma Virus (HPV) infection and more serious cervical intraepithelial neoplasia and squamous intraepithelial lesion. In other groups dominated by *Gardnerella* and *Streptococcus* , we found *Gardnerella* gradually reduced with severity from ASC-US to CIN2 and HSIL. Functional analysis from inferred metagenomes showed that the low evenness group was showing some pathways including DNA repair, recombination protein, DNA replication protein and DNA activity such as pyrimidine and amino sugar, nucleotide. Amino acid metabolism and biosynthesis were enriched in high evenness group.

Conclusion: According to our finding, we emphasized that the species evenness might play an important role in vaginal microbiota, potentially leading to shifts in functional pathways underlying the influence of vaginal diseases.

Significance and impact of the study: This study reveals that the ecological balance, especially species evenness, can shape the vaginal microbiota.

Keywords: Vaginal microbiota; Community evenness; Complex and mutable ecological

INTRODUCTION

Bacteria account for 50% of the cells of the human body, and together with archaea and lower eukaryotes are collectively termed 'human microbiota [1,2]. The link between health, disease and the human microbiota is a fast-moving and contentious area of research. An appreciation of the variation in microbiota composition amongst individuals is expanding our understanding of the pathophysiology underlying a variety of diseases affecting many body systems, including many physiological phenotypes of vaginal health. However, imbalance of vaginal microbiome can cause serious diseases, such as bacterial vaginosis [3-5], Human Papilloma Virus (HPV)-induced cervical carcinogenesis [6,7], local and systemic immune responses[8,9]. Therefore, the composition of vaginal microbiota is not fixed under different phenotypes of the human body. Besides what we have known some main factors that modify vaginal microbiota such as menopause [10,11], sexual behaviour [12,13], pregnancy [14-16], age [17,18], menopause [19] and even smoking [20,21] are diverse in different study. In recent years, more and more studies have clarified that the content of *Lactobacillus* could be variable in different countries, regions, races and even in the same region and race [22-25]. The vagina and its microbiota form a balanced ecosystem, and microbes normally play an important role in preventing colonization by pathogenic

Correspondence to: Tzong-Yi Lee, Department of Computational Biology, The Chinese University of Hong Kong, Shenzhen, China, E-mail: leetzongyi@cuhk.edu.cn

Chien-Hsun Huang, Department of Obstetrics and Gynecology, Taoyuan General Hospital, Ministry of Health&Welfare, Taoyuan, Taiwan, E-mail: lithsunh@gmail.com

Received: March 01, 2021; **Accepted:** March 15, 2021; **Published:** March 22, 2021

Citation: Chen Q, Jhong JH, Li S, Huang CH, Lee TY (2021) A Metagenomics-Based Study on Functional Associations between Vaginal Microbiota and Phenotypes of HPV-Infected Patients. Appli Microbiol Open Access.7: 195.

organisms. *Lactobacillus* has been reported to defend against pathogens and sexually transmitted infections [12,26] through maintenance of a hostile pH [27,28], production of species-specific metabolites [27,29] bacteriocins [30] and disruptionof biofilms [31]. Microbial cell–cell interactions in the vaginal flora are clarified to play an integral role in the development of biofilms and, ultimately, they can also generate an array of serious gynecological and obstetric complications [32]. The description of a polymicrobial biofilm on the epithelial surface from Bacterial Vaginosis (BV) vaginal biopsy specimens puts *Gardnerella*, the major component of these multispecies communities, at the center of BV pathogenesis [33-35]. The vaginal microbiota is different from the intestinal microbiome because the vaginal microbiota colonizes in an open lumen. Similarly, they do not belong to the human body, but only exist as an independent microbial community that mutually benefits with the human body. However, previous research focused on factors associated with vaginal health [36]. Therefore, it is meaningful to analyze the microbial community structure from an ecological perspective. Early research has shown that community biodiversity is mainly determined by species richness and evenness. Its balance affects the structure and function of the entire ecological community [37,38]. It is known that complex and mutable ecological of vaginal microbiome is associated with vaginal diseases and potentially with systemic disease such as cervical cancer. Previous research has mainly focused on certain types of microbes, and rarely involves the ecological balance with vaginal microbes as the core. In this study, based on the perspective of the diversity of the vaginal microbiota, we will analyze the impact of changes in the species richness and evenness of the vaginal microbiota on the vagina microbial community, thereby causing effects on the human vaginal microenvironment.

MATERIALS AND METHODS

Sample collection and DNA extraction

Samples are from department of OBS/GYN in TaoYuan Hospital. Subjects had abnormal smear results and no history of autoimmune diseases. History of childbearing and sex life and medication are also collected. This study was approved by the Medical Ethics and Institutional Review Board of Taoyuan General Hospital, Ministry of Health and Welfare (Date: 2019. 09. 20, #IRB: TYGH106076) and was conducted in accordance with the ethical standards of the declaration of Helsinki. Vaginal swab samples were collected using ∑ -Transwab® Sigma Swab and transport to the lab in Health GeneTech Taiwan on 4°C within four hours. Samples were centrifuged at 14,000 rpm for 1 min upon arrival. After removing the supernatant, the samples were stored at -20°C till DNA and RNA extraction. Physical bead beating method was applied to the samples to achieve better bacterial lysis before the DNA was purified by Zymo BIOMICS™ DNA Miniprep Kit. RNA was isolated by Trizol method and purified by Direct-zol RNA Miniprep Kit. 3 ug of total RNA was subjected to rRNA depletion with the MICROBExpress kit before library preparation.

Library construction and sequencing

16S metagenomics library was constructed using two rounds of PCR reactions. The first PCR amplified the V3-V4 region of 16S rRNA gene as well as adding adaptor sequences. Ilumina index sequences were added into the PCR product in the second PCR products using AMPure beads at 1:1 ratio. The library DNA was QC for its fragment size by electrophoresis and for its concentration using real-time quantitative PCR before sequenced using Miseq V3 600 cycles kit.

Bacterial 16S rRNA analyze

We got 40 subjects contained 80 paired end sequence. Formatted reads (FASTQ file formats) were subjected to the FastQC (Babraham Bioinformatics, Cambridge, CBE) for quality filtering (at quality score=20(Q20)) using the split_libraries_fastq.py script of QIIME (version 1.9.1) [39]. We compared the difference among Q19, Q20 and Q30 which Q20 was the best choice (Supplement Figure 1). Operational Taxonomic Units (OTUs) were assigned to the reads using QIIME'spick_closed_reference_otus.py script against the Greengenes Database. Sequences were clustered into OTUs using a 97% similarity threshold. The OUT, less than 2 sequences, was discarded and remaining were summarized from phylum to genus level. The reconstruction of 40 subjects was also performed using Uparse [40] and taxa were assigned to consensus sequences using rdp against the Greengenes Database. In the end we got the composition of microorganism in each subject from phylum to genus. The microorganism construction was used to investigate possible divergence in taxonomic assignment between the two methods. The microbial community diversity, alpha diversity and beta diversity, was performed using QIIME. Unsupervised hierarchical clustering based on relative abundance at genus level was applied to define clusters according to abundance and taxa diversity of each sample. The results were classed into 3 groups and we its association with the regard to known risk factor and smear result. The alpha diversity of each defined group was estimated using the Shannon index, Phylogenetic Diversity (PD) whole tree index, chao1 index and observed otus index. clustering analysis was carried out using a website tool named 'Morpheus' Bacterial metagenome content was predicted from 16S rRNA gene-based microbial compositions, and functional inferences were made from the Kyoto Encyclopedia of Gene and Genomes (KEGG) catalog, using the PICRUSt algorithm [41]. A total of 6,909 inferred genes were categorized into 328 KEGG functional pathways. The KEGG pathways of each subjects were subjected to the Linear Discriminative Analysis (LDA) Effect Size (LEfSe) [42] of linux environment to get the significant difference of KEGG functional pathways in 3 groups obtained from clustering result. And the differentially abundant KEGG pathways were used to be correlation with risk factor.

Statistical analysis

Patients' sociodemographic and clinical data have been summarized with frequency and average (average ± SD). The alpha diversity of the different group was compared using the Mann-Whitney test in GraphPad Prim 8. The different distribution of all regarded risk factors associated with 3 groups was investigated by GraphPad Prim 8 (for discontinuous data) and Origin 2018 (for continuous data), and the statistical analyses was carried out using R 3.6.1 (Kruskal-Wallis rank sum test and Fisher's Exact Test were used). LEfSe based on linux, using Kruskal-Wallis and estimating the effect size of the comparisons, was used for evidencing significantly different KEGG pathways obtained from PICRUSt. A 95% confidence interval has been used in this analysis and the LDA score threshold was set to one.

RESULTS

Sociodemographic and clinical baseline characteristics

We enrolled 39 women into the study who were from Taiwan. The average age of participants was 55 years old and the average BMI

index is 23.56. Approximately 69% of the women were insomnia, and 38% for vaginal inflammation, 44% for pelvis inflammation 18% for bladder urethra inflammation (Table 1).

Table 1: Sociodemographic characteristics of the studied participants.

Characteristics index	Index	Sample
Age (average ± SD)	55 ± 13	37
BMI	23.52 ± 3.73	36
Insomnia	69%(27/39)	38
Vaginal inflammation	38%(15/39)	38
Pelvis inflammation	44%(7/39)	38
Bladder urethra inflammation	18%(7/39)	38
Diabetes	18%(7/39)	38
HPV infected	54%(21/39)	39
First sex (average ± SD)	23 ± 6	38
Sexual partner (average ± SD)	2 ± 1	38
Pregnancy (average ± SD)	3 ± 2	38
Fertility (average ± SD)	2 ± 1	38
Menstrual regularity	46% (18/39)	38
Menopause	62% (24/39)	38
Stay up late	23% (9/39)	37
Sports	72% (28/39)	38

Microbiome community diversity

To further understand the microbial community composition of each sample, the software QIIME1 was applied. A total of 6053730 reads with an average of 151343 was processed and 3950423 high quality reads with an average of 98761 were involved into study after quality control using QIIME1. As a result, 2541 Operational Taxonomic Units (OTUs) were constructed, with an average of 63 OTUs per sample. We further analyzed the distribution of microbial composition in phylum and genus (Figures 1A and 1B). 17 phyla were enriched where 5 phyla were detected as highly abundant (within relative abundant >1%): *Firmicutes* (55.35%) *Proteobacteria* (11.45%), *Actinobacteria* (18.96%), *Bacteroidetes* (10.56%), *Fusobacteria* (1.35%) (Figure 1B). We also identified 233 genera in which Lactobacillus, *Streptococcus, Gardnerella, Prevotella* and *Bifidobacterium* are most abundant (top 5 highly abundant). In this study, we performed 2 strategies to construct microbial community, showing a high degree of similarity. Fitting using initial unsupervised clustering analysis resulted in four community types among all 39 individuals at genus level, named Group1,2,3,4 (Figure 1C). Relative abundance of different community types was analyzed except Group 4 (only 1 sample: N23). Group 1 was dominated by *Lactobacillus*; the total of *Gardnerella*, *Prevotella* and *Lactobacillus* were more than 40% in Group 2; *Streptococcus* are highly enriched in Group 3. Obviously, Group 2 and 3 showed higher bacterial diversity and lower *Lactobacillus* abundance than Group 1. On the contrary, *Gardnerella* was only highly enriched in Group 2 (Figure 1D). To further analyze the diversity of microbial species among the three groups, we selected Shannon index, Phylogenetic Diversity (PD) whole tree index, Chao1 index, Observed otus index for evaluation. The Shannon index showed a significantly lower in Group 1 compared to the remaining community types (Group 2 and Group 3) (Figure 2A), but no significant difference was detected among the three community types at the level of Phylogenetic Diversity (PD) whole tree, Chao1 and Observed otus index (Figures 2B-2D). Suggesting the community evenness was mainly contribution to the community types difference rather than the community richness. For exploring the impact of the difference in the dominant microorganism among the three groups, we used smear results as a feature to describe the distribution in three groups. Some clinical indicators were involved in these subjects that we could make a different distribution in three community types. We summarized the smear test results in Figure 3A, including adenocarcinoma, Atypical Glandular Cell (AGC), Atypical Squamous Cells of Undetermined Significance (ASC-US), cervical intraepithelial neoplasia 1/2/3(CIN1, CIN2, CIN3) and High grade Squamous Intraepithelial Lesion (HSIL). *Gardnerella* was dominant in Group 2, In particular, we found it gradually reduced with severity from ASC-US to CIN2 and HSIL (Figure 3B). *Streptococcus*, main component of Group 3, showed no distinct trends in ASC-US and HSIL, but *Veillonella* and *Prevotella* were enriched in ASC-US (Figure 3C). In this study, we combined CIN2, CIN3, and HSIL, named 'high risk group' because of more serious cervical intraepithelial neoplasia and squamous intraepithelial lesion. The 'high risk group' was contained in Group 1 with 60% ratio. In comparison, 47% in Group 2 and 33% in Group 3 (Figure 3A).

Analysis of different community types correlation

With regards to some known factors (age, height, weight, first sex age, sexual partner, pregnancy, fertility, abortion, insomnia, stay up, virginal inflammation, pelvis inflammation, balder urethra inflammation, diabetes, HVP infection, menstrual regularity, menstruation, sports), we examined the correlation between these factors in different community types to explore the feature of each group (Table 2). Sexual partner, sports and menstruation, compared to each other in three community types, showed significant differences (Figures 4A-4C). More sexual partners were detected in Group 1 with high relative abundance of *Lactobacillus*, slightly loweraverage age and higher HPV infection (Figures 4A and 4D, Figure 1D). More subjects in Group 1 were menstrual, but major part of subjects were menopause in group 2 and group 3 (Figure 4B). In addition, we found the number of sports was increasing from group 1 to group 3 (Figure 4C). Moreover, all subjects in group 3 participated in sports with lower ration of HPV infection (Figure 4D). The remaining factors were insignificantly different in three community types. Next, for illustrating the important role of microorganisms, we carried out analyses to explore microbiota function based on inferred metagenomes through the Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt). The PICRUSt analysis comparing the microbiota between three community types showed different KEGG Orthologs (KO) and its related KEGG pathway. In the end, 6,909 inferred genes were categorized into 328 KEGG functional pathways. In order to characterize the degree of enrichment of pathways between three community types, another tool, named Linear Discriminant Analysis (LDA) effect size (LEfSe), was used to excavate the significant different KEGG pathways among three community types. We chose $p<0.01$ and LDA score >3 as the threshold to pick out the most significant KEGG pathway, which 15 KEGG pathways included in Group 1, 7 KEGG pathways included in Group 2 and 10 KEG pathways included in Group 3 (Figure 5A) (Table S5). A broad overview of these significant KEGG pathways was indicated that all these 3 community types were contained more metabolism pathways and less biosynthesis

pathways. More interesting was some pathways matched DNA or chromosome was only found in Group 1. Besides, more about the metabolism pathway which included amino acid metabolism was mainly enrichment in Group 3, including Histidine metabolism and Glycine, Serine and Threonine metabolism. Moreover, Phenylalanine, Tyrosine, Tryptophan biosynthesis and Valine, Leucine, Isoleucine biosynthesis exclusively appeared in Group 3. Compared to others groups, the number of KEGG pathway in Group 2 was a bit less, but we found two functions about protein, protein folding and associated processing and bacterial motility protein (Figure 5B).

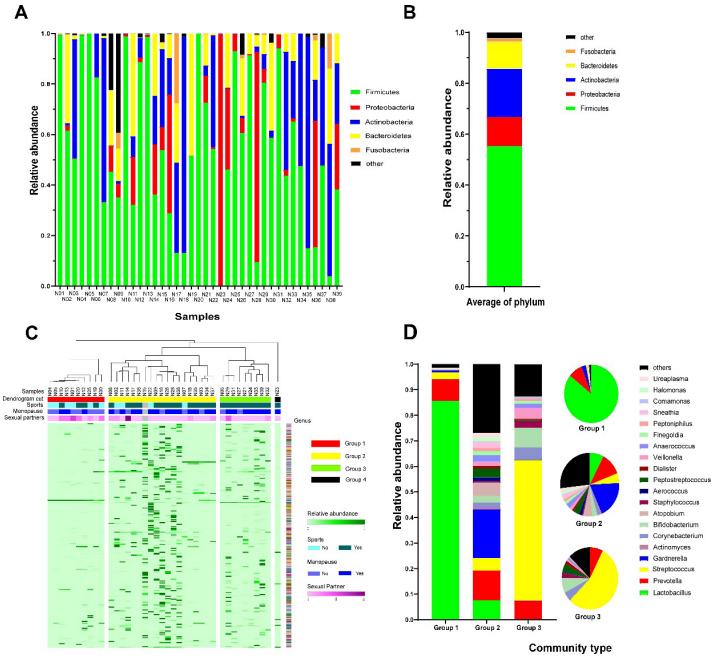

Figure 1: Microbial community structure and clustering result in vaginal metagenomes (A) Composition of 39 subjects (within relative abundant >1%) at phylum level; (B) The average of phylum from 39 subjects; (C) Bacterial genera distributions detected in vagina using QIIME1. Clustering of subjects was performed using unsupervised hierarchical linkage with average Euclidean distances of the proportional bacterial abundance in each subject. Clustering result in four groups and some 'risk factor', such as sports, HPV infection, menopause, and sexual partners, were annotated within 4 groups; (D) Composition of high abundance genus (containing at least 10% of each group subjects within relative abundant >1%) in each of the three community types. The group 4 is not contained because of few number at genus level. Subjects distribution of the community types are fellows: group 1, n=10; group 2, n=19; group 3, n=9.

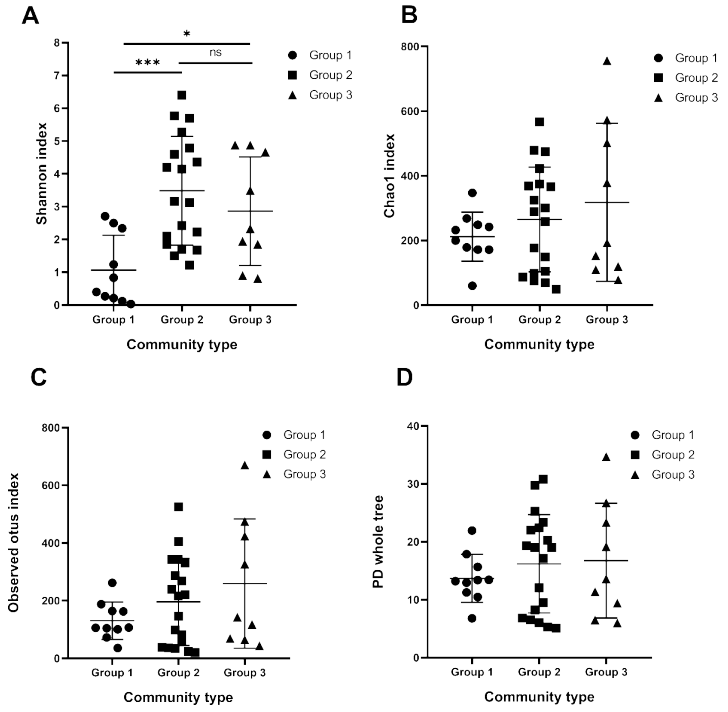

Figure 2: Diversity of the vaginal microbiota in each of community types.The alpha diversity indexes, including Shannon index (A) Chao1 index; (B) Observed tus index; (C) Phylogenetic Diversity (PD) whole tree index; (D) were performed to evaluate the species diversity in each of community types. Only Shannon index showed significant difference among three different groups ($p<0.05$; one-side Mann-Whitney U test). Significant changes: *, elevation with $p<0.05$; ***, elevation with $p<0.001$.

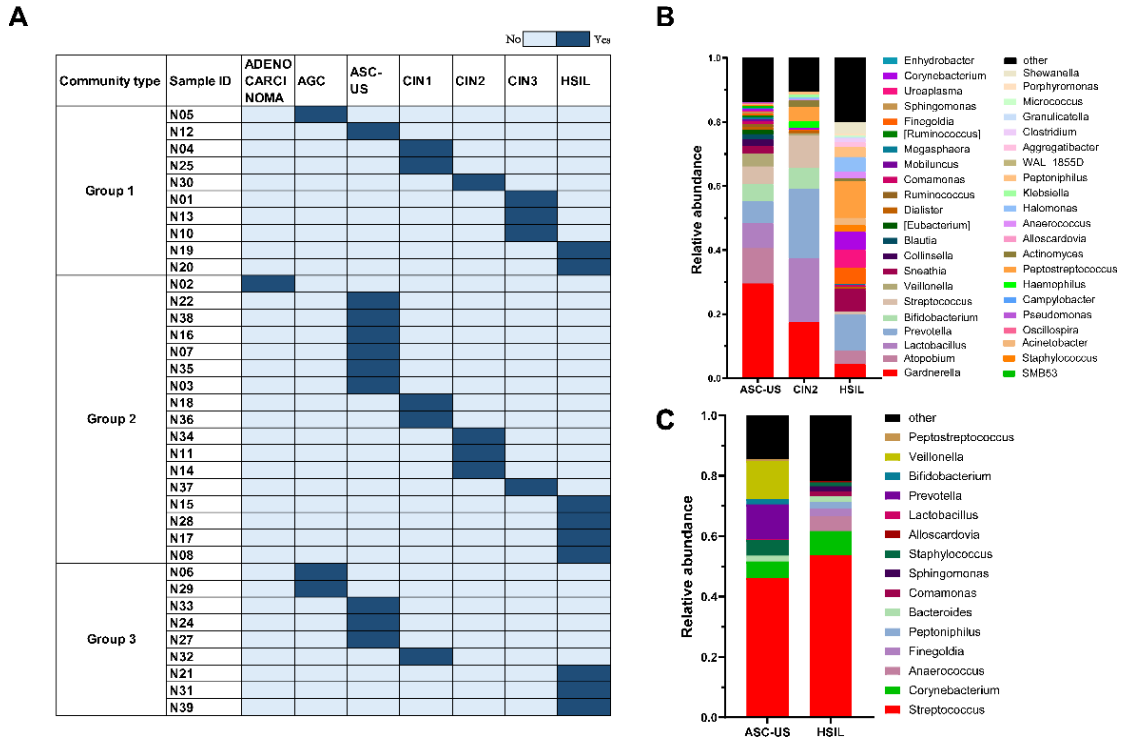

Figure 3: The distribution of smear results and composition genus in Group 2 and Group 3 A) The distribution of smear results in each of community types; (B) Composition of high abundance genus (containing at least 10% of each group subjects within relative abundant >1%) in women who are detected as ASC-US, CIN2 and HSIL from group 2; (C) Composition of high abundance genus (containing at least 10% of each group subjects within relative abundant >1%) in women who are detected as ASC-US and HSIL from group 3.

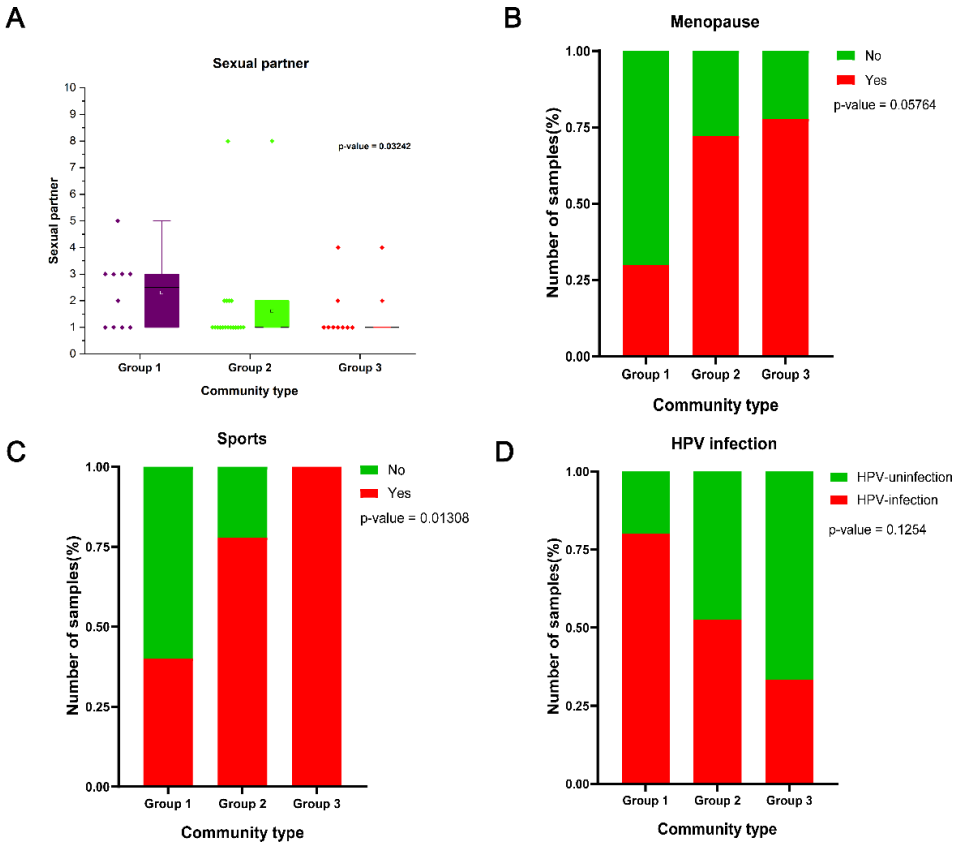

Figure 4: The significantly different of regard to known some factors in each other community types (A) The number of sexual partner showed different distribution in three groups (p=0.03242, Kruskal-Wallis rank sum test); (B) Menopause, important factor for women, was much less in Group 1 (p=0.05784, Fisher's Exact Test for Count Data); (C) Sports was more likely enriched in Group 3 (p=0.01308, Fisher's Exact Test for Count Data); (D) The number of HPV infection was gradually decrease from Group 1 to Group 3 (p=0.1254, Fisher's Exact Test for Count Data). The boxes represent 25th–75th percentiles, black lines indicate the median and whiskers extend to the maximum and minimum values within 1.5 × the interquartile range and dots indicate outliers.

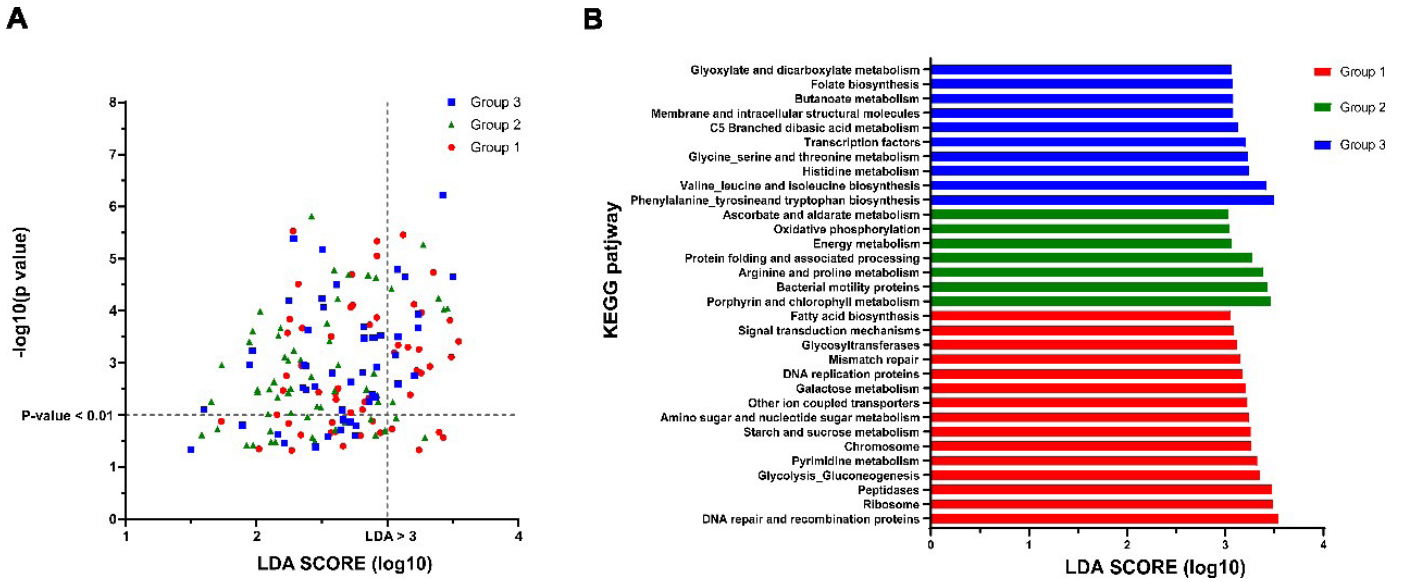

Figure 5: Bacterial taxa associated with clustering result were related to several gene functional pathways. Bacterial gene function was predicted from 16S rRNA gene-based microbial compositions using the PICRUSt algorithm to make inferences from KEGG annotated databases (A) Genus and KEGG pathway counts were submitted to LEfSe to evaluate the significantly different KEGG pathways; (B) 32 KEGG pathways are selected which 15 in Group 1, 7 in Group 2 and 10 in Group 3 using p<0.01 and LDA score >3 as threshold.

Table 2: Clustering results characteristics according to the findings of clinical data.

Characteristics	Group 1	Group 2	Group 3	p-value
Height	159.60 ± 6.70	154.84 ± 4.83	157.22 ± 5.59	0.8246a
Weight	60.7 ± 9.70	56.06 ± 1025	58.78 ± 7.32	0.3304 a
First sex	21.70 ± 5.06	23.28 ± 7.35	23.33 ± 4.03	0.2134 a
Sexual partner	2.3 ± 1.34	1.61 ± 1.65	1.44 ± 1.01	0.0324 a
Pregnancy	2.5 ± 1.08	3.22 ± 1.44	3.11 ± 0.610	0.7771 a
Fertility	2.00 ± 1.05	2.44 ± 0.78	2.89 ± 0.60	0.1032 a
Abortion	0.5 ± 0.71	0.78 ± 1.11	0.22 ± 0.44	0.9719 a
Insomnia (never/slight/seriou)	4:2:4	5:6:7	2:2:5	0.8813 b
Vaginal inflammation (never/slight/serious)	5:2:3	12:2:4	5:2:2	
0.8728 b	55 ± 13	55 ± 13	55 ± 13	55 ± 13
Pelvis inflammation (never/slight/serious)	9:0:1	14:3:1	7:1:1	
0.8505b	55 ± 13	55 ± 13	55 ± 13	55 ± 13
Bladder urethra inflammation (never/slight/serious)	8:0:2	15:1:2	7:1:1	
0.9134 b	55 ± 13	55 ± 13	55 ± 13	55 ± 13
Diabetes (never/slight/serious)	9:1:0	14:4:0	7:2:0	
0.7512 b	55 ± 13	55 ± 13	55 ± 13	55 ± 13
Menstrual regularity (regular/irregular)	6:4	10:8	5:4	1 b
Menstruation (yes/no)	3:7	13:5	7:2	0.0576 b
Sports (yes/no)	4:6	14:4	9:0	0.0131 b
Age	48 ± 12.16	55.88 ± 11.20	60.56 ± 13.18	0.255 a
stay up (yes/no)	3:7	5:12	2:7	1 b
HPV-infection (infection/uninfection)	8:2	10:9	3:6	0.1254 b
Note: a: Kruskal-wallis rank sum test; b: Fisher's exact test for count data	55 ± 13	55 ± 13	55 ± 13	55 ± 13

DISCUSSION

In this vaginal microbiome analysis, we constructed the vaginal microbial community using 39 subjects from Taiwan. Relatively consistent microbial composition was observed in QIIME and Uparse, suggesting that some pre-processions (dereplication, chimera filtering) was almost no effect on detected microorganisms at phylum level. In this study, we used result of QIIME to further analysis. 233 genera were detected in which *Lactobacillus*, *Streptococcus*, *Gardnerella*, *Prevotella* and *Bifidobacterium* were most abundant (top 5 highly abundant). In order to investigate the differences between these subjects, we performed clustering analysis at genus level, hoping to classify samples with similar microbiota. Three groups with different microbial relative abundance were obtained. We observed the vaginal microbiome among three groups were difference. To our surprised was only Shannon index presented significantly different among three groups. In contrast, the Phylogenetic Diversity (PD) whole tree, chao1 and observed outs index were not different, indicating the community evenness was main factor to influence three groups. Biodiversity was a complex term that includes taxonomic, functional, spatial and temporal aspects of organismic diversity, with species richness (the number of species) and evenness (the relative abundance of species)

considered among the most important measures [43,44]. In earlier scientific research, the majority of studies of biodiversity-functioning and biodiversity–stability theory had predominantly examined richness [37,38,45-47]. Wittebolle [48,49] has confirmed that community evenness was a key factor in preserving the functional stability of an ecosystem using experimental manipulations of both richness and initial evenness in microcosms with denitrifying bacterial communities. Therefore, when communities were highly uneven, or there was extreme dominance by one or a few species, their functioning was more likely less resistant to environmental stress. In this study, we got verified that the community evenness was different among three groups but no differed in species richness. Interestingly, we observed highest ratio of HPV infection in Group 1 and lowest in Group 3, the trend was gradually decreasing from Group 1 to Group 3. Besides, the *Lactobacillus* was main composition in Group 1 with 'over threshold' more than 85%. In previous research by Ravel [50], for example, vaginal bacterial communities dominated by *Lactobacillus* were found in 80.2% and 89.7% of Asian and white women, respectively, but just 59.6% and 61.9% of black and Hispanic women, respectively. On the other hand, occurrence of communities with low proportions or no detectable *Lactobacillus* species community type were elevated in Hispanic (38.1%) and black (40.4%) women compared to Asian (19.8%) and white (10.3%) women. These findings were in accordance with results obtained by Xia Zhou [22,25,50]. And more interesting, although a low pH environment promoted by lactic acid may be considered generally protective because of inhibition of growth potentially pathogenic species, HPV infection and development of CIN may be additionally influenced by the chemical structure of the lactic acid molecule [51]. Some research has revealed that women exhibit a higher ratio of L to D-lactate, which can lead to increased expression of Extracellular Matrix Metalloproteinase Inducer (EMMPRIN) and Activation of Matrix Metallo Proteinase (MMP-8). This expression could feasibly lead to altered cervical integrity and facilitate entry of HPV to the basal keratinocytes, where the virus thrives [51]. In addition, a high concentration of D-lactate produced has been recently shown to increase cervicovaginal mucus viscosity and enhance its viral particle trapping potential [52]. Therefore, we suspected that the distribution of vaginal microbiota in women in Taiwan was distinct, showing 'over threshold' high *Lactobacillus* was associated with HPV infection because of weaker resistance. To further investigate the different distribution of Shannon index, we got smear results involve in this study. CIN2, CIN3 and HSIL were classed into 'high risk group', and we observed the number of ratio in each group was gradually decreasing from Group 1 to Group 3. Combined with conclusion from previous analysis, we made a conjecture that if communities were highly uneven showing out of balance, their functioning is less resistant to environmental stress, such as high-grade disease (CIN 2 or 3, or High-grade Squamous Intraepithelial Lesion (HSIL). In addition, we analyzed the microbiota composition of Group 2 and Group 3, showing high Shannon index, associated with smear result. In Group 3, dominant by *Streptococcus*, no significant difference of *Streptococcus* was observed in ASC-US and HISL. In Group2, dominant by *Gardnerrella*, three states (ASC-US, CIN2, HSIL) were involved to show *Gardnerrella* gradually decreased with severity. Besides we also observed *Peptostreptococcus* was higher level in CIN2, HSIL and lower level in ASC-US, which was consistent with previous research [53]. Gardnerrella was first described in 1953 Leopold [54], and subsequently identified as the causative agent of a cluster of vaginal symptoms currently known as vaginosis. Evy Gillet [55] had investigated the association between Bacterial Vaginosis (BV) and cervical intraepithelial neoplasia, confirming a positive association between BV and cervical pre-cancerous lesions. However, half of the included studies showed no significant association between BV and cervical pre-cancerous lesions. And as far as we know, Swidsinski [56] demonstrated that a biofilm was present on 90% of the epithelial surfaces of BV vaginal biopsy specimens, with Gardnerrella the major component of these multispecies communities. Therefor we suspected that Gardnerrella may influenced the cervical pre-cancerous lesions by forming biofilms. With regard to known risk factors (age, height, weight, first sex age, pregnancy, fertility, vaginitis, pelvic inflammation, cystourethritis, diabetes, menstrual regularity, insomnia and stay up late) to vaginal microbiome, we detected no significant ($p<0.05$) difference among the three community types. More sexual partners, fewer numbers of menopause and sports were detected in Group 1 with lower average of age, higher HPV infection and higher relative abundance *Lactobacillus*. We had known that following menopause, reduced oestrogen and resulting vaginal atrophy were thought to lead to Lactobacillus depletion and increased diversity [20], which was consistent with our results. The Group 1, associated with more serious cervical pre-cancerous lesions, was showing some pathways including DNA repair and recombination protein, DNA replication protein, and more about substrate of DNA activity such as pyrimidine and amino sugar, nucleotide. Simple explanations for our finding was that due to the drastic changes in the external environment, the genetic and mutation of the bacteria itself were affected, which reflected in the enhanced activity of DNA and chromosomes. A recent study also showed that women with high-risk HPV strains had lower concentrations of amino acids and peptides compared with women who had only low-risk HPV [57]. As to what effect this change will have on the vaginal microenvironment, it is not yet clearly known. This requires further research on metabolomics of the vagina. In summary, in this study of female vaginal microbiome from Taiwan, we observed that community evenness was a main factor to influence function of vaginal microbiome. We provided an ecological perspective on the association between complex vaginal microbial composition and some phenotypes of the human body, including HPV infection, smear result and regard to know risk factors. We suspected that 'over threshold' *Lactobacillus* accompanied by lower community evenness and more enriched in DNA and chromosome activate, resulting more serious cervical pre-cancerous lesions (CIN2, SHIL) and higher HPV infection. We additionally described the *Gardnerella* was gradually decreasing serious pre-cancerous lesions (from ASC-US, CIN2 to HSIL). About this correlation trend, a simple hypothesis was proposed: *Gardnerella* may further mediate the occurrence and development of pre-cancerous lesions by forming biofilms, which may involve the participation of other anaerobic microorganisms. But unfortunately, the complex relationship between vaginal microbes and hosts was only observed in this study, as we did not directly obtain data on their interactions. The establishment of a clearer relationship needs to be further revealed by vaginal metabolomics. In our view, most of the early study so far has placed much emphasis on the cataloguing species names and the new species. In fact, we think that interventions that could help to treat conditions such as vaginal disease, cervical cancer will be discovered only if we move beyond species catalogues and mutable ecological and evolutionary relationships that microbes have with each other and with their hosts [58].

CONCLUSION

In this study, we found species evenness were significant different among three community groups, while the species richness showed no significant, which emphasized that the species evenness might play an important role in vaginal microbiota, potentially leading to shifts in functional pathways underlying the influence of vaginal diseases.

DECLARATIONS

Ethics approval and consent to participate

This study was approved by the Medical Ethics and Institutional Review Board of Taoyuan General Hospital, Ministry of Health and Welfare (Date: 2019. 09. 20, #IRB: TYGH106076) and was conducted in accordance with the ethical standards of the Declaration of Helsinki.

Consent for publication

Not applicable

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article (and its additional files). The vaginal microbiome data supporting the conclusions of this article are available in the NCBI Sequence

Read Archive (BioProject ID: PRJNA648790)

Competing interests

The authors declare that they have no competing interests.

Acknowledgement

This work was supported by the Warshel Institute for Computational Biology, School of Life and Health Sciences, The Chinese University of Hong Kong, Shenzhen, China and Taoyuan General Hospital (PTH10773).

Authors' contributions

QC carried out the data collection and curation, participated in the bioinformatics analyses, and drafted the manuscript. JHJ and SL participated in the design of the study and performed the draft revision. CHH and TYL conceived of the study, participated in its design and coordination, and helped to revise the manuscript. All authors read and approved the final manuscript.

REFERENCES

1. Marchesi JR, Ravel J. The vocabulary of microbiome research: A proposal. Microbiome . 2015; 3:31.
2. Sender R, Fuchs S, Milo R. Are we really vastly outnumbered? Revisiting the ratio of bacterial to host cells in humans. Cell. 2016 ;164(3):337-340.
3. Goje O, Shay E, Padmanabhan R, Eng C. Characterizing the vaginal microbiome in women with recurrent bacterial vaginosis and treatment implications. American Journal of Obstetrics & Gynecology. 2019 ;221(6):671.
4. Muzny CA, Taylor CM, Swords WE, Tamhane A, Chattopadhyay D, Cerca N, et al. An updated conceptual model on the pathogenesis of bacterial vaginosis. J Infect Dis. 2019;220(9):1399-1405.
5. Lev-Sagie A, Goldman-Wohl D, Cohen Y, Dori-Bachash M, Leshem A, Mor U, et al. Vaginal microbiome transplantation in women with intractable bacterial vaginosis. Nat Med. 2019;25(10):1500-1504.
6. Kwasniewski W, Wolun-Cholewa M, Kotarski J, Warchol W, Kuzma D, Kwasniewska A, et al. Microbiota dysbiosis is associated with HPV-induced cervical carcinogenesis. Oncol lett. 2018 ;16(6):7035-7047.
7. Wu P, Xiong H, Yang M, Li L, Wu P, Lazare C, et al. Co-infections of HPV16/18 with other high-risk HPV types and the risk of cervical carcinogenesis: A large population-based study. Gynecol Oncol. 2019 ;155(3):436-443.
8. Stoner KA, Beamer MA, Avolia HA, Austin MN, Meyn LA, Hillier SL, et al. Fewer cervical immune cells in women with abnormal vaginal microbiota compared to women with normal vaginal microbiota. Am J Obstet Gynecol. 2016; 215(6):S814.
9. Chan D, Lee YS, Ahmed S, Teoh TG, Bennett PR, MacIntyre DA, et al. Local and systemic immune responses in preterm and term pregnancies and their association with the vaginal microbiota. Int J Gynaecol Obstet. 2019; 126: E128-E128.
10. Muhleisen AL, Herbst-Kralovetz MM. Menopause and the vaginal microbiome. Maturitas. 2016 ;91:42-50.
11. Alvisi S, Gava G, Orsili I, Giacomelli G, Baldassarre M, Seracchioli R,et al. Vaginal health in menopausal women. Medicina. 2019;55(10):615.
12. Wessels JM, Lajoie J, Vitali D, Omollo K, Kimani J, Oyugi J, et al. Association of high-risk sexual behaviour with diversity of the vaginal microbiota and abundance of Lactobacillus. PLoS One. 2017 2;12(11):e0187612.
13. Lewis FM, Bernstein KT, Aral SO. Vaginal microbiome and its relationship to behavior, sexual health, and sexually transmitted diseases. Obstet Gynecol. 2017; 129(4):643.
14. Subramaniam A, Kumar R, Cliver SP, Zhi D, Szychowski JM, Abramovici A, et al. Vaginal microbiota in pregnancy: evaluation based on vaginal flora, birth outcome, and race. Am J Perinatol. 2016 ;33(4):401-408.
15. Freitas AC, Chaban B, Bocking A, Rocco M, Yang S, Hill JE, et al. The vaginal microbiome of pregnant women is less rich and diverse, with lower prevalence of Mollicutes, compared to non-pregnant women. Sci Rep. 2017;7(1):1-6.
16. McKee, K. The vaginal microbiota in pregnancy and spontaneous preterm birth. Am J Obstetric and Gynecol. 2019; 221(6):694-694.
17. Dei M, Di Maggio F, Di Paolo G, Bruni V. Vulvovaginitis in childhood. Best Pract Res Clin Obstet Gynaecol. 2010;24(2):129-137.
18. Ranđelović G, Mladenović V, Ristić L, Otašević S, Branković S, Mladenović-Antić S, et al. Microbiological aspects of vulvovaginitis in prepubertal girls. Eur J Pediatr. 2012;171(8):1203-1208.
19. Brotman RM, Shardell MD, Gajer P, Fadrosh D, Chang K, Silver M, et al. Association between the vaginal microbiota, menopause status and signs of vulvovaginal atrophy. Menopause. 2018 Nov;25: 1321-1330.
20. Brotman RM, He X, Gajer P, Fadrosh D, Sharma E, Mongodin EF, et al. Association between cigarette smoking and the vaginal microbiota: A pilot study. BMC Infect Dis. 2014;14(1):471.
21. Nelson TM, Borgogna JC, Michalek RD, Roberts DW, Rath JM, Glover ED, et al. Cigarette smoking is associated with an altered vaginal tract metabolomic profile. Sci Rep. 2018; 8(1):852.
22. Zhou X, Bent SJ, Schneider MG, Davis CC, Islam MR, Forney LJ, et al. Characterization of vaginal microbial communities in adult healthy women using cultivation-independent methods. Microbiology. 2004;150(8):2565-2573.
23. Hyman RW, Fukushima M, Diamond L, Kumm J, Giudice LC, Davis RW, et al. Microbes on the human vaginal epithelium. Proc Natl Acad Sci USA. 2005;102(22):7952-7957.

24. Anukam KC, Osazuwa EO, Ahonkhai I, Reid G. *Lactobacillus* vaginal microbiota of women attending a reproductive health care service in Benin city, Nigeria. Sex Transm Dis. 2006; 33(1):59-62.

25. Zhou X, Brown CJ, Abdo Z, Davis CC, Hansmann MA, Joyce P, et al. Differences in the composition of vaginal microbial communities found in healthy Caucasian and black women. ISME J. 2007;1(2):121-133.

26. Martin DH. The microbiota of the vagina and its influence on women's health and disease. Am J Med Sci. 2012;343(1):2-9.

27. O'Hanlon DE, Moench TR, Cone RA. Vaginal pH and microbicidal lactic acid when lactobacilli dominate the microbiota. PloS one. 2013;8(11):e80074.

28. Godha K, Tucker KM, Biehl C, Archer DF, Mirkin S. Human vaginal pH and microbiota: An update. Gynecol Endocrinol. 2018; 34(6):451-455.

29. Ocaña VS, de Ruiz Holgado AA, Nader-Macías ME. Characterization of a bacteriocin-like substance produced by a vaginal *Lactobacillus* salivariusstrain. Appl Environ Microbiol. 1999; 65(12):5631-5635.

30. Boris S, Barbés C. Role played by lactobacilli in controlling the population of vaginal pathogens. Microbes Infect. 2000 ;2(5):543-546.

31. McMillan A, Dell M, Zellar MP, Cribby S, Martz S, Hong E, et al. Disruption of urogenital biofilms by lactobacilli. Colloid Surf B Biointerfaces. 2011; 86(1):58-64.

32. Menard JP, Mazouni C, Salem-Cherif I, Fenollar F, Raoult D, Boubli L, et al. High vaginal concentrations of atopobium vaginae and *gardnerella* vaginalis in women undergoing preterm labor. Obstet Gynecol. 2010;115(1):134-140.

33. Hardy L, Jespers V, Dahchour N, Mwambarangwe L, Musengamana V, Vaneechoutte M, et al. Unravelling the bacterial vaginosis-associated biofilm: a multiplex *gardnerella* vaginalis and atopobium vaginae fluorescence in situ hybridization assay using peptide nucleic acid probes. PLoS One. 2015; 10(8):e0136658.

34. Hardy L, Jespers V, Van den Bulck M, Buyze J, Mwambarangwe L, Musengamana V, et al. The presence of the putative *gardnerella* vaginalis sialidase: A gene in vaginal specimens is associated with bacterial vaginosis biofilm. PloS one. 2017 ;12(2):e0172522.

35. Castro J, Machado D, Cerca N. Unveiling the role of *gardnerella* vaginalis in polymicrobial bacterial vaginosis biofilms: The impact of other vaginal pathogens living as neighbors. ISME J. 2019; 13(5):1306-13017.

36. Greenbaum S, Greenbaum G, Moran-Gilad J, Weintraub AY. Ecological dynamics of the vaginal microbiome in relation to health and disease. Am J Obstet Gynecol. 2019 ;220(4):324-335.

37. Naeem S, Li S. Biodiversity enhances ecosystem reliability. Nature. 1997; 390(6659):507-509.

38. Sankaran M, Mc NaughtonSJ. Determinants of biodiversity regulate compositional stability of communities. Nature. 1999; 401(6754):691-693.

39. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010; 7(5):335-336.

40. Edgar RC. UPARSE: Highly accurate OTU sequences from microbial amplicon reads. Nat Methods. 2013;10(10):996-998.

41. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol. 2013; 31(9):814-821.

42. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. Genome Biol. 2011; 12(6):1-8.

43. Purvis A, Hector A. Getting the measure of biodiversity. Nature. 2000; 405:212-219.

44. Wilsey BJ, Potvin C. Biodiversity and ecosystem functioning: Importance of species evenness in an old field. Ecology. 2000; 81(4):887-892.

45. Loreau M, Hector A. Partitioning selection and complementarity in biodiversity experiments. Nature. 2001 ;412:72-76.

46. Cardinale BJ, Palmer MA, Collins SL. Species diversity enhances ecosystem functioning through interspecific facilitation. Nature. 2002; 415:426-429.

47. Bell T, Newman JA, Silverman BW, Turner SL, Lilley AK. The contribution of species richness and composition to bacterial services. Nature. 2005; 436:1157-1160.

48. Wittebolle L, Marzorati M, Clement L, Balloi A, Daffonchio D, Heylen K, et al. Initial community evenness favours functionality under selective stress. Nature. 2009 ;458:623-626.

49. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, et al. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci. 2011; 108:4680-4687.

50. Zhou X, Hansmann MA, Davis CC, Suzuki H, Brown CJ, Schütte U, et al. The vaginal bacterial communities of Japanese women resemble those of women in other racial groups. FEMS Immunol Medi Microbiol. 2010; 58(2):169-181.

51. Witkin SS, Mendes-Soares H, Linhares IM, Jayaram A, Ledger WJ. Influence of vaginal bacteria and D-and L-lactic acid isomers on vaginal extracellular matrix metalloproteinase inducer: Implications for protection against upper genital tract infections. MBio. 2013; 4(4): e00460-13.

52. Nunn KL, Wang YY, Harit D, Humphrys MS, Ma B, Cone R, et al. Enhanced trapping of HIV-1 by human cervicovaginal mucus is associated with *Lactobacillus* crispatus-dominant microbiota. MBio. 2015; 6(5): e01084-15.

53. 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. Sci Rep. 2015;5(1):1-1.

54. Leopold S. Heretofore undescribed organism isolated from the genitourinary system. US Armed Forces Med J. 1953; 4(2):263-236.

55. Gillet E, Meys JF, Verstraelen H, Verhelst R, De Sutter P, Temmerman M, et al. Association between bacterial vaginosis and cervical intraepithelial neoplasia: systematic review and meta-analysis. PloS one. 2012 ;7(10):e45201.

56. Swidsinski A, Mendling W, Loening-Baucke V, Ladhoff A, Swidsinski S, Hale LP, et al. Adherent biofilms in bacterial vaginosis. Obstet Gynecol. 2005;106:1013-1023.

57. Borgogna JC, Shardell MD, Santori EK, Nelson TM, Rath JM, Glover ED, et al. The vaginal metabolome and microbiota of cervical HPV‐positive and HPV-negative women: a cross-sectional analysis. BJOG. 2020; 127(2):182-192.

58. Proctor L. Priorities for the next 10 years of human microbiome research. Nature. 2019; 569: 623-625.