Altered vaginal eukaryotic virome is associated with different cervical disease status

Yanpeng Li,1, Le Cao,1, Xiao Han,1, Yingying Ma, Yanmei Liu, Shujun Gao,*

Shanghai Public Health Clinical Center, Fudan University, Shanghai, 201508, China
Center of Diagnosis and Treatment for Cervical Diseases, Obstetrics and Gynecology Hospital of Fudan University, Shanghai, 200080, China
Shanghai Key Laboratory of Female Reproductive Endocrine-Related Disease, Fudan University, Shanghai, 200080, China

ARTICLE INFO

Keywords:
Vaginal virome
Cervical cancer
Anellovirus
Papillomavirus

ABSTRACT

Viruses are important components of the human body. Growing evidence suggests that they are engaged in the physiology and disease status of the host. Even though the vaginal microbiome is involved in human papillomavirus (HPV) infection and cervical cancer (CC) progression, little is known about the role of the vaginal virome. In this pilot exploratory study, using unbiased viral metagenomics, we aim to investigate the vaginal eukaryotic virome in women with different levels of cervical lesions, and examine their associations with different cervical disease status. An altered eukaryotic virome was observed in women with different levels of lesions and Lactobacillus depletion. Anelloviruses and papillomaviruses are the most commonly detected eukaryotic viruses of the vaginal virome. Higher abundance and richness of anelloviruses and papillomaviruses were associated with low-grade squamous intraepithelial lesion (LSIL) and CC. Besides, higher anellovirus abundance was also associated with lactobacillus-depleted microbiome profiles and bacterial community state (CST) type IV. Furthermore, increased correlations between Anelloviridae and Papillomaviridae occurred in the women with increased cervical disease severity level from LSIL to CC. These data suggest underlying interactions between different microbes as well as the host physiology. Higher abundance and diversity of both anelloviruses and papillomaviruses shared by LSIL and CC suggest that anellovirus may be used as a potential adjunct biomarker to predict the risk of HPV persistent infection and/or CC. Future studies need to focus on the clinical relevance of anellovirus abundance with cervical disease status, and the evaluation of their potential as a new adjunct biomarker for the prediction and prognoses of CC.

1. Introduction

More than 600,000 new cases and 342,000 deaths worldwide were related to cervical cancer (CC) in 2020, making it the fourth most prevalent cancer and the fourth leading cause of cancer death in women (Sung et al., 2021). Persistent infection with human papillomavirus (HPV), especially high-oncogenic-risk HPV (hrHPV), is a major risk factor for cervical lesion and CC development (Curty et al., 2019; Castanheira et al., 2021). However, hrHPV infection alone is a necessary but not sufficient factor for CC development since most hrHPV infections are transient and regress spontaneously (Castellsagué, 2008). Many additional factors are also involved in HPV persistence and cervical disease progression or regression, such as host-related factors including immunity status, smoking, parity and sexual behavior, mechanical factors like vaginal douching, and other biological factors like sexually transmitted infections (STIs) (Castellsagué, 2008; Audirac-Chalifour et al., 2016).

The microbiome, in most cases the bacterial microbiome, has long been determined to play an essential role in maintaining homeostasis and physiological functions (Gilbert et al., 2018). The vaginal bacterial microbiome (VMB) has been shown to be associated with cervical disease progression (Mitra et al., 2016; Kyrgiou et al., 2017; Champer et al., 2018; Laniewski et al., 2020). The bacterial microbiome in the healthy vaginal tract is mainly dominated by Lactobacillus and is less diverse than other anatomical sites such as the gut. According to relative abundance of the bacterial profiles, five distinct bacterial community state types (CSTs I–V) are established (Ravel et al., 2011). The

* Corresponding authors.
E-mail addresses: 031108282@fudan.edu.cn (S. Gao), chiyu_zhang1999@163.com (C. Zhang).
1 Yanpeng Li, Le Cao, and Xiao Han contributed equally.
Local microenvironment factors that influence the local immune response and the vaginal microenvironment, thereby involving in HPV infection control and cervical oncogenesis (Chase et al., 2015; Castanheira et al., 2021). For example, the decreasing abundance of certain Lactobacillus species and increasing microbiome diversity (especially the enrichment of different anaerobes) are associated with increasing lesion severity and CC progression (Mitra et al., 2015; Audirac-Chalifour et al., 2016; Ustyk et al., 2020; Gardella et al., 2022; Zhang et al., 2022). However, the change of vaginal bacterial microbiome could also be a consequence of altered local immune response and vaginal environment.

The collection and complex network of commensal and pathogenic eukaryotic viruses and bacteriophages termed the virome (Virgin, 2014; Carroll et al., 2018; Ren et al., 2021). Many studies that focused on limited disease types and body compartments suggested an important role of the virome in human physiology and disease (Virgin, 2014; Norman et al., 2015; Li et al., 2019; Liang and Bushman, 2021; Cao et al., 2022). However, whether it is the cause of disease status and/or adverse conditions is not clear. As it was discussed in a recent review (Madere and Monaco, 2022), the virome could also be involved in different gynecological health and disease. For example, higher vaginal eukaryotic viral diversity was found to be associated with preterm birth, exposure to antibiotics and reproduction failure (Wylie et al., 2018; Eskew et al., 2020). Furthermore, the bacterial compositions may play an important role in the bacterial vaginosis (BV) status. A recent study investigated the association of cervicovaginal DNA virome with cervical lesion severity and CC development. In this exploratory study, we enrolled women with different cervical disease status, and investigated their vaginal eukaryotic virome profiles using an unbiased viral metagenomic sequencing method. We aim to determine the total viral components that present in women’s vaginal tract and examine possible associations between the eukaryotic virome trait and certain lesion stages.

2. Materials and Methods

2.1. Study design and cohort

To investigate the vaginal virome and its association with different cervical disease status, women who visited the cervical clinic during summer of 2021 at the Obstetrics and Gynecology Hospital of Fudan University were enrolled in this cohort. The exclusion criteria included ongoing pregnancy, immunosuppression by drugs, the use of antibiotics during the last one month, and previous history of cervical treatment or surgery. Women were classified into different cervical disease groups according to hrHPV test, cytology, colposcopy inspections and pathological diagnosis. In total, 46 women with low-grade squamous intraepithelial lesion (LSIL), 36 women with high-grade squamous intraepithelial lesion (HSIL) and 26 women with cervical cancer (CC) were included. Women without any signs of cervical tissue lesions (normal tissue, NT) were selected based on the 2019 American Society of Colposcopy and Cervical Pathology (ASCCP) guideline (Schiffman et al., 2020): HPV negative and normal cytology result; or HPV positive and normal colposcopy inspections. Fifty-three women with NT were included. Demographics, such as age, douching habits and contraception methods, as well as basic clinical features were recorded upon their visits (Table 1). The sample collection was approved by the Ethics Committees of Obstetrics and Gynecology Hospital of Fudan University. Written informed consents were obtained before sample collection.

Table 1

<table>
<thead>
<tr>
<th>Patient characteristic of 161 patients included in study cohort.</th>
<th>NT (n = 53)</th>
<th>LSIL (n = 46)</th>
<th>HSIL (n = 36)</th>
<th>CC (n = 26)</th>
<th>Total (n = 161)</th>
<th>P-valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years</td>
<td>39.4</td>
<td>41.4</td>
<td>38.4</td>
<td>48.3</td>
<td>41.2</td>
<td>0.005</td>
</tr>
<tr>
<td>Mean (SD, range)</td>
<td>(11.1,23–74)</td>
<td>(9.8, 26–69)</td>
<td>(10.8, 23–69)</td>
<td>(13.5, 20–75)</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Ethnicity (Han), n (%)</td>
<td>53 (100)</td>
<td>46 (100)</td>
<td>36 (100)</td>
<td>26 (100)</td>
<td>161 (100)</td>
<td>0.555</td>
</tr>
<tr>
<td>Vaginal douching, n (%)</td>
<td>8 (15.1)</td>
<td>3 (6.5)</td>
<td>5 (13.9)</td>
<td>3 (11.5)</td>
<td>19 (11.8)</td>
<td>–</td>
</tr>
<tr>
<td>Yes</td>
<td>45 (84.9)</td>
<td>42 (93.5)</td>
<td>31 (86.1)</td>
<td>20 (76.9)</td>
<td>139 (86.3)</td>
<td>–</td>
</tr>
<tr>
<td>No</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>3 (11.5)</td>
<td>3 (1.9)</td>
<td>–</td>
</tr>
<tr>
<td>Contraception, n (%)</td>
<td>25 (47.2)</td>
<td>15 (32.6)</td>
<td>11 (30.5)</td>
<td>12 (46.1)</td>
<td>63 (39.2)</td>
<td>0.239</td>
</tr>
<tr>
<td>Condom</td>
<td>24 (45.3)</td>
<td>21 (45.6)</td>
<td>17 (47.2)</td>
<td>7 (26.9)</td>
<td>69 (42.8)</td>
<td>–</td>
</tr>
<tr>
<td>Copper IUD + Ligation</td>
<td>4 (7.5)</td>
<td>9 (19.6)</td>
<td>7 (19.4)</td>
<td>5 (19.2)</td>
<td>25 (15.5)</td>
<td>–</td>
</tr>
<tr>
<td>Other oral hormonal contraception</td>
<td>0 (0)</td>
<td>1 (2.2)</td>
<td>1 (2.8)</td>
<td>2 (7.7)</td>
<td>4 (2.5)</td>
<td>–</td>
</tr>
<tr>
<td>Times of abortion</td>
<td>1.2 (1.3, 0–5)</td>
<td>1.4 (1.10–4)</td>
<td>1.2 (2.0, 0–11)</td>
<td>1.4 (1.2, 0–4)</td>
<td>1.3 (1.4, 0–11)</td>
<td>0.270</td>
</tr>
<tr>
<td>Mean (SD, range)</td>
<td>2.6</td>
<td>2.0</td>
<td>2.3</td>
<td>2.3</td>
<td>2.4</td>
<td>0.161</td>
</tr>
<tr>
<td>Number of sexual partners</td>
<td>(4.7, 1–24)</td>
<td>(1.5, 1–4)</td>
<td>(1.5, 1–8)</td>
<td>(6.3, 1–24)</td>
<td>(3.8, 1–24)</td>
<td>–</td>
</tr>
<tr>
<td>Complications, n (%)</td>
<td>4 (7.5)</td>
<td>4 (8.7)</td>
<td>4 (11.1)</td>
<td>4 (15.4)</td>
<td>16 (9.9)</td>
<td>0.684</td>
</tr>
<tr>
<td>Multifocal lesions, n (%)</td>
<td>0 (0)</td>
<td>8 (17.4)</td>
<td>2 (5.5)</td>
<td>0 (0)</td>
<td>10 (6.2)</td>
<td>–</td>
</tr>
<tr>
<td>LV and STIs</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>–</td>
</tr>
<tr>
<td>Lactobacillus profiles</td>
<td>33 (62.3)</td>
<td>30 (65.2)</td>
<td>24 (66.7)</td>
<td>7 (26.9)</td>
<td>94 (58.4)</td>
<td>0.004</td>
</tr>
<tr>
<td>Lac-dominant, n (%)</td>
<td>19 (35.8)</td>
<td>15 (32.6)</td>
<td>12 (33.3)</td>
<td>19 (73.1)</td>
<td>65 (40.4)</td>
<td>–</td>
</tr>
<tr>
<td>CSTs</td>
<td>18 (34)</td>
<td>12 (26.1)</td>
<td>12 (33.3)</td>
<td>19 (73.1)</td>
<td>61 (37.9)</td>
<td>–</td>
</tr>
<tr>
<td>CST 1, n (%)</td>
<td>15 (28.3)</td>
<td>9 (19.5)</td>
<td>5 (13.9)</td>
<td>0</td>
<td>29 (18)</td>
<td>–</td>
</tr>
<tr>
<td>CST III, n (%)</td>
<td>18 (34)</td>
<td>22 (47.8)</td>
<td>19 (52.8)</td>
<td>7 (26.9)</td>
<td>66 (41)</td>
<td>–</td>
</tr>
<tr>
<td>CST IV, n (%)</td>
<td>18 (34)</td>
<td>12 (26.1)</td>
<td>12 (33.3)</td>
<td>19 (73.1)</td>
<td>61 (37.9)</td>
<td>–</td>
</tr>
<tr>
<td>CST V, n (%)</td>
<td>1 (1.9)</td>
<td>2 (4.3)</td>
<td>0</td>
<td>0</td>
<td>3 (1.9)</td>
<td>–</td>
</tr>
</tbody>
</table>

* Fisher’s exact test and Kruskal-Wallis (with Dunn’s correction) test were used as appropriate for the comparison of the variables between different groups (NT, LSIL, HSIL and CC). NT, normal tissue; LSIL, low-grade squamous intraepithelial lesion; HSIL, high-grade squamous intraepithelial lesion; CC, cervical cancer. BV, bacterial vaginosis; STIs, sexually transmitted infections; NA, not available.
2.2. Sample collection

Vaginal swabs were collected upon clinic visit, and were immersed in virus transport medium (Yocon, Beijing, China) immediately. Samples were first stored at 4 °C and then transferred to –80 °C within 24 h. A total of 161 swabs were collected.

2.3. Vaginal virome-sample processing, library construction and sequencing

The enrichment of viral particles was performed as previously described (Li et al., 2019; Li et al., 2021) with a few modifications. Briefly, vaginal swab was resuspended in 400 μL PBS, and the suspension was homogenized with ceramic beads beating twice at 30 Hz for 30 s each on a Tissuelyser-24 (Shanghai Jing Xin), with a 2-min interval between bead-beating cycles. A negative control sample (the same PBS used for swab suspension) was also performed in parallel to monitor potential background contaminations. The suspension was centrifuged at 12,000 × g for 10 min. Supernatant was passed through a 0.45 μm sterile filter to reduce background noise (Costar Spin-X centrifuge tube filters, Corning, USA). Filters were incubated with a cocktail of nuclease for 2 h at 37 °C. The reaction was terminated with 30 mmol/L EDTA at 65 °C for 10 min. Both DNA and RNA were extracted using QIAamp MinElute virus kit (Qiagen, Germany), and then amplified using random RT-PCR as described (Li et al., 2020a; Li et al., 2020b). Briefly, reverse transcription was performed with the following primer, 5′GGCGGACTAATGCGTAGTCAACGTNNNNNNNNGGC-3′, and the second strand synthesis was performed using Klenow Fragment DNA polymerase (New England Biolabs, Massachusetts, USA). Both cDNA and RNA were then amplified using AmpliTaq Gold DNA polymerase. DNA libraries were prepared using the transposon-based Nextera XT Sample Preparation Kit and were sequenced on the Illumina Novaseq platform (Illumina, USA) with 2 × 150-bp paired reads.

2.4. Virome bioinformatic analyses

Sequencing data were analyzed as previously described (Zhao et al., 2017; Li et al., 2022). The metagenomic sequencing data were first filtered by Trimmomatic v.0.38 (Bolger et al., 2014) by removing adapters and low-quality sequences. Human sequences (using HG38 database) were subtracted from the data using Bowtie2 v.2.3.4.3 (Langmead and Salzberg, 2012). Remaining high-quality reads were subjected to de-novo assembly using Megahit v.1.1.3 (Li et al., 2015). Assembled contigs, as well as singlets, were mapped against the viral nucleic acid and protein database using BLAST 2.11.0+ (E < 10^{-10}) and BLASTx (E < 10^{-5}) (DIAMOND v.0.9.24) (Buchfink et al., 2015), respectively. All the viral hit candidates were then searched against the NCBI nonredundant nt and nr database to further identify sequences that have higher similarity to non-viral sequences. To reduce potential cross-library contamination (Geoghegan et al., 2021), viruses with a read count less than 0.1% of the highest count for that virus among the other libraries were removed from subsequent analyses. Viral abundance of each sample was calculated by reads per million (RPM) of total clean reads.

2.5. Vaginal bacterial microbiome-sample processing, library construction, sequencing and analyses

Total DNA was extracted from the swabs using TIANGEN bacterial DNA Kit (Tiangen Biotech Co., Ltd, China). The library was generated as previously described (Xu et al., 2021). Briefly, the V4 hypervariable region (515–806 nt) of 16S rRNA gene was amplified by two rounds of PCR. The amplification libraries were purified and sequenced using the Illumina Miseq platform by Novogene Co., Ltd. The NGS data were first filtered by Trimmomatic v.0.39 to remove adaptors and low-quality sequences (SLIDINGWINDOW:5:20 MINLEN:50). Reads were merged using Vsearch v.2.18 and demultiplexed by Fastx-toolkit. DADA2 was used to quality filtering, dereplication, denoising with default settings. All the sequences were truncated to 250 bp using QIIME 2 platform (https://qiime2.org). The taxonomic classification of amplicon sequence variants (ASVs) representative sequences was performed using the Naïve Bayesian Classifier algorithm based on the Silva database at genus level (Quast et al., 2013). The bacterial microbiome composition at the genus level was analyzed using the Statistical Analysis of Metagenomic Profiles (STAMP) package (v.2.1.3) (Parks and Beiko, 2010). The vaginal Lactobacillus community state types (CSTs) I-V were determined as previously described (Ravel et al., 2011). Based on the hierarchical clustering analysis (HCA) of bacterial microbiome, samples with a Lactobacillus relative abundance of > 65.7% were classified as Lactobacillus-dominated (lac-dominant), and samples with a Lactobacillus relative abundance of < 65.7% were classified as Lactobacillus-depleted (lac-depleted).

2.6. Phylogenetic analyses

For the phylogenetic analyses of anellovirus and papillomavirus, all the viral sequences annotated as Anelloviridae and Papillomaviridae were first extracted and assembled into contigs. ORFs of the contigs longer than 1500 bp were extracted using NCBI’s ORF Finder tool under the “any sense codon” option. The ORF1 region of Anelloviridae and the L1 region of Papillomaviridae were further identified and curated by aligning against an in-house database containing the ORF1 region of anellovirus reference sequences and the L1 region of papillomavirus reference sequences, respectively. Viral nucleic acid sequences were aligned using MAFFT (Katoh et al., 2019). Phylogenetic trees were inferred using the maximum likelihood method with IQ-Tree (Minh et al., 2020). ModelFinder was used to determine the best substitution models. Phylogenetic trees based on nucleotide sequences were generated using the bootstrap method (1000 times). All the sequences used in the phylogenetic analyses were deposited in GenBank under the accession numbers OP721120–OP721159 (annelovirus) and OP721160–OP721243 (papillomavirus).

2.7. Detection of anellovirus and papillomavirus with qPCR methods

Anellovirus was detected using a SYBR qPCR method. Primers targeting the conserved region of anellovirus were used (Thijssen et al., 2020; Li et al., 2022). Anellovirus copy number in each sample was calculated based on a standard curve constructed by plasmid standards. qPCR program was as follows: 95 °C for 1 min and 40 cycles at 95 °C for 15 s, and 63 °C for 1 min. HPV was detected by BioPerfectus HPV Genotyping Real Time PCR kit (Bioperfectus Tech-nologies, Taizhou, China), which covers 13 high risk HPV types (HPV-16, -18, -31, -33, -35, -39, -45, -51, -52, -56, -58, and -68) and 8 low risk HPV types (HPV-26, -53, -56, -73, -82, -6, -11, and -81).

2.8. Alpha- and beta-diversities of the vaginal viral community

The alpha diversity was calculated by the Shannon diversity index (H = −Σ(pi log(pi)) and chao richness score, which measures species diversity and relative abundance, and computes the number of species in a community, respectively. The beta diversity of viral community was indicated by PCoA analyses based on Bray-Curtis distances. All indexes were calculated at species level or genotypes (HPV).

2.9. Co-occurrence network analyses

Co-occurrence networks of the viral genus/family were analyzed in different groups. Rho > 0.6 and P < 0.05 (Spearman correlation coefficient) were used as cutoff for including edges into the graph. Co-occurrence network of correlated viral taxa was visualized by Gephi v.0.9.2.
2.10. Statistics

Continuous variables between groups were compared by the nonparametric Mann-Whitney U test or Kruskal-Wallis Test with Dunn's correction. Frequencies were compared by Fisher's exact test. Statistical significance for principal coordinate analyses (PCoA) was determined by PERMANOVA testing. The correlations between the qPCR and metagenomic sequencing (RPM), age and anellovirus or HPV abundance were calculated by Spearman's correlation. Viral abundance was shown as the reads per million (RPM). LEfSe analysis was performed to identify discriminative vaginal viral features (Segata et al., 2011). The linear discriminant analysis (LDA) was used to analyze specific taxa among different groups based on the relative abundance of each taxon. A logarithmic LDA score of $>4$ was used to determine discriminative features. All statistical analyses were performed using R studio v.3.8, Graphpad Prism 8 and SPSS 26.0.

3. Results

3.1. Demographics and clinical characteristics

Demographic, clinical characteristics and vaginal Lactobacillus profiles of the cohort are listed in Table 1. The vaginal swabs from 53 women with NT, 46 with LSIL, 36 with HSIL and 26 with CC were subjected to viral metagenomic sequencing and 16s sequencing. The mean age of all women is 41.2 years old, and women with CC were significantly older than those in other groups ($P = 0.005$). There was no significant difference in douching habits, contraception methods, abortion history, sexual partners and clinical complications among different groups. Multifocal lesions were only detected in 6.2% of all individuals (8 and 2 women from LSIL and HSIL groups, respectively). In order to analyze the association between virome alterations and the Lactobacillus profiles, samples were classified as either lac-dominant or lac-depleted, and different CSTs (Table 1) as described in the method.

3.2. Overview of the vaginal virome

In total, 1526 million clean reads were generated. After quality control and annotation, 367 million reads were recognized as viral sequences. Overall, seven families of human eukaryotic viruses were detected in the vaginal swabs (Fig. 1 and Supplementary Fig. S1), including Papillomaviridae, Anelloviridae, Herpesviridae, Polyomaviridae, Adenoviridae, Retroviridae and Hepadnaviridae. At least one eukaryotic virus was detected in 90.7% of the samples. Papillomaviruses (HPV) and anelloviruses were the most commonly detected viruses, which were present in 78.3% and 69.6% of women, respectively. They were also the most abundant viruses, accounting for more than 90% of the total eukaryotic viral reads (Fig. 1B). Other five viral families were detected at low prevalence (0.6%–7.4%). Bacteriophages were present in all samples, with Siphoviridae, Myoviridae and other Caudovirales being the most abundant phages (Fig. 1C).

3.3. Vaginal virome with different cervical lesion severity and bacterial microbiome profiles

We first investigated the relationship between vaginal eukaryotic virome and different cervical lesion severity (Fig. 2A–D). Viral abundance was significantly higher in the LSIL group ($P < 0.001$) than the NT and HSIL groups. Women with CC showed a similar trend, but no significant difference was reached ($P = 0.054$ and 0.002) (Fig. 2A). Shannon diversity of the vaginal virome was highest in the CC group, and viral richness was significantly higher in both the LSIL and CC groups than the NT and HSIL groups ($P < 0.01$) (Fig. 2B and C). Principal coordinate analyses (PCoA) showed that the virome community of the CC group is significantly separated from the NT, LSIL and HSIL groups that shared similar virome community (PERMANOVA, adjusted $P = 0.015$, 0.024, 0.034 respectively, Fig. 2D). LEfSe analyses revealed that specific HPV types were associated with the LSIL and HSIL groups, and two anellovirus species were associated with the CC group (Fig. 2E). Since the Lactobacillus profile is associated with cervical lesion severity and CC, samples were classified as either lac-dominant or lac-depleted, and different CSTs (Table 1) as described in the method, and the relationship between the vaginal eukaryotic virome and different Lactobacillus profiles was investigated (Fig. 2F–M). The virome community was significantly separated between the lac-dominant and lac-depleted groups (PERMANOVA, $P = 0.009$) (Fig. 2D), and slightly higher viral abundance and diversity indexes were observed in the lac-depleted group than the lac-dominant group ($P > 0.05$) (Fig. 2F–H). The virome community was also significantly separated among CSTs I, III and IV (PERMANOVA, adjusted $P = 0.003$, 0.032 and 0.035 respectively, Fig. 2M). We found significantly higher viral abundance in the CSTs III ($P = 0.017$) and IV ($P = 0.004$) groups, and only a slightly higher richness was observed in CST IV compared to CST I ($P = 0.054$) (Fig. 2J–L).

3.4. Anellovirus

Since anellovirus was one of the most abundant and prevalent viruses in women's vagina, we further analyzed vaginal anellome in different groups (Fig. 3). The CC and LSIL groups had substantially higher anellovirus abundance, richness and Shannon diversity than the other two groups, with the highest level in the CC group (Fig. 3A–C). In addition, substantially higher anellovirus abundance and richness were detected in the lac-depleted group ($P < 0.05$ than the lac-dominant group (Fig. 3D and E). Anellovirus Shannon diversity was also higher in lac-depleted group, but no significance was reached ($P = 0.08$) (Fig. 3F). Higher anellovirus abundance and richness were detected in the CST III ($P > 0.05$) and IV ($P < 0.05$) groups than the CST I group (Fig. 3G and H). No significant difference of anellovirus Shannon diversity was found between CSTs (Fig. 3I). We further performed a qPCR assay to validate the anellovirus abundance of NGS, and a strong positive correlation was found between the number of reads and the qPCR viral loads (Fig. 3J, Spearman's correlation $r = 0.755$, $P < 0.0001$). All three genera of anelloviruses, as well as some unclassified anelloviruses were detected in women's vagina (Fig. 4, and Supplementary Fig. S2), and gammatorquevirus and unclassified anelloviruses mainly contributed to higher anellovirus abundance in the CC group. Slightly higher age of the CC group did not contribute to the higher anellovirus abundance since anellovirus abundance was not significantly associated with age (Supplementary Fig. S3).

3.5. Papillomavirus

HPV was the second most abundant eukaryotic virus in women's vagina. The unbiased NGS method detected 65 HPV genotypes in 77.6% women, covering all 21 genotypes that were detected by qPCR method plus 44 additional genotypes (Supplementary Table S1). The agreement for overall HPV positivity between the two methods was 80.1% (73.3% by qPCR kit and 77.6% by NGS). The most prevalent hrHPVs by NGS were HPV 58 (37.3%), 51 (28.6%), 56 (22.4%), 52 (22.4%) and 16 (21.1%) (Supplementary Fig. S4). The hrHPVs positivity rates were significantly higher in LSIL and CC groups and were not associated with Lactobacillus profiles (Fig. 5A–C). Most of the HPV were alphapapillomavirus, followed by gammapapillomavirus (Fig. 6), and one potential novel gammapapillomavirus genotype (with 86% identity to HPV226) was identified from a woman with no tissue lesions (Supplementary Fig. S5).

Women with LSIL had significantly higher HPV abundance than those with NT and HSIL, and both the LSIL and CC groups had significantly higher HPV richness and Shannon diversity than the other two groups, with LSIL showing the highest levels (Fig. 5D–F). No significant difference was found in HPV abundance, richness and Shannon diversity between the Lac-dominant and Lac-depleted groups (Fig. 5G–I). Slightly
higher HPV abundance and richness were found in CST III and IV groups as compared to those in CST I, but no significance was reached (Fig. 4J and K). No difference of HPV Shannon diversity was found between CSTs (Fig. 4L). Similar to the observation in anellovirus abundance, age did not contribute to the higher HPV abundance in the LSIL and CC groups (Supplementary Fig. S6).

3.6. The interactions of virome in different cervical disease severity

To reveal potential correlations among different viruses, the co-occurrence networks of virome from NT, LSIL, HSIL, and CC groups were compared (Fig. 7). The vast majority of the viral correlations occurred within the same viral family (mainly Anelloviridae and...
A

B

C

D

E

F

G

H

I

J

K

L

M

(caption on next page)
Papillomaviridae). The correlations between Anelloviridae and Papillomaviridae were rarely observed in the NT group (Fig. 7A), and correlations between the two viral families increased along with the aggravation of cervical disease severity from LSIL to CC (Fig. 7B-D and Supplementary Fig. S7). In particular, a substantially increased network complexity was observed in the HSIL and CC groups, as more co-occurrence connections were detected between different viral taxa (Supplementary Figs. S7–C, P < 0.001, Fisher’s exact test).

4. Discussion

We revealed that the eukaryotic virome was altered in the vagina of women with different levels of cervical lesions or cancer. The altered virome was mainly characterized by increased abundance and diversity of both Anelloviridae and Papillomaviridae in women with LSIL and CC. In particular, higher anellovirus abundance was significantly associated with CC, as well as non-protective vaginal bacteria profiles (e.g., lact-depleted and CST IV). Furthermore, increased correlations between Anelloviridae and Papillomaviridae occurred in the women with increased cervical disease severity from LSIL to CC. These results suggest that the vaginal virome may be involved in the development of various cervical diseases, especially CC.

Even though we investigated the vaginal virome using a method targeting both DNA and RNA viruses (Siqueira et al., 2018; da Costa et al., 2021; Li et al., 2021), the overall vaginal viral types are mainly DNA viruses, which are similar to previous studies (Madere and Monaco, 2022), with Papillomaviridae and Anelloviridae being the most common eukaryotic viruses and Caudovirales being the dominant bacteriophages. Higher eukaryotic viral abundance and richness were associated different cervical lesion severity or CC status, and this was consistent with the few studies investigating the vaginal virome in the context of women genital health. For example, higher eukaryotic viral diversity was significantly associated with preterm birth and reproductive outcomes in asymptomatic women (Wylie et al., 2018; Eskevi et al., 2020). In women with BV, more eukaryotic viruses were detected (Zhang et al., 2021), and bacteriophages profiles were associated with bacterial community and BV status (Jakobsen et al., 2020). A recent study of the virome in HIV-1/HPV co-infected women reported that a higher abundance and diversity of HPV was associated with premalignant cervical lesions (LSIL and HSIL) (Siqueira et al., 2019). These data indicate that vaginal virome may play an unnoticed role in the local health and disease status.

Different from anellovirus, we found that HPV abundance and diversity were highest in LSIL, followed by CC. HPV is often detected extra-chromosomally in benign and low-grade lesions, and the integrated HPV (or persistent HPV infection) highly correlates with the severity of cervical disease (Klaes et al., 1999; Hudelet al., 2004; IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, 2008). This means a more active replication of HPV at LSIL stage rather than HSIL. Besides, because the samples and methods in this study mainly capture cell-free viral particles rather than the cells, the integrated HPV's have relatively low possibility to be detected. These may explain why the higher HPV abundance was observed in the LSIL group than in the HSIL and CC groups. Similar to a recent study (Kaelin et al., 2022), we found no significant difference in HPV abundance and diversity by the Lactobacillus profile. This indicates that other virome alterations (such as anellovirus) other than HPV may better explain the changes in the vaginal Lactobacillus profile.

Anelloviruses are ubiquitous in human population and different anatomical sites (Tozetto-Mendoza et al., 2020; Taylo et al., 2022). Even though no convincing evidence of anellovirus involvement in particular disease types, they are believed to interact with the host immune system (Kaczorowska and van der Hoek, 2020). We found that vaginal anelloviruses were highly associated with CC and the Lactobacillus profiles, which were consistent with two recent studies (Kaelin et al., 2022; Tozetto-Mendoza et al., 2022). In these studies, they found that higher anellovirus abundance was associated with the genital inflammation and lower levels of Lactobacillus. Vaginal anellovirus abundance also negatively correlated with CD4+ T-cell count (Siqueira et al., 2019), highlighting possible interactions of them with our local immune response. The co-infection of multiple anelloviruses, as well as with other viruses are common in different body compartments, such as blood, respiratory tract, semen and vagina (Siqueira et al., 2019; Kaczorowska and van der Hoek, 2020; Liu et al., 2021; Li et al., 2022). In particular, the co-infection of anellovirus and HPV was found to be associated with increased laryngeal tumor progression (Szlak et al., 2005). Increased co-occurrence of anelloviruses with HPVs and other viruses was found to be associated with the lesion severity and CC; these data suggest that different anellovirus compositions and their cross-interactions with HPVs may involve in the HPV persistence, microenvironment change and progression to CC. Considering the widely reported association between anellovirus and the immune status, our findings suggest that anellovirus may be a potential indicator of the local vaginal immune and/or disease status. Whether the change in anellovirus composition is the result or cause of the altered vaginal microenvironment, such as HPV persistence and disease progression, needs to be further investigated.

This study further highlights the power of metagenomic sequencing for the diagnosis of viruses in clinic samples, as well as the mining of specific viral types or groups that may be associated with different disease conditions. Even though further works in a larger cohort with longitudinal samples are needed, our exploratory results indicate that, similar to the bacterial microbiome, specific viral signatures, especially the commensal anelloviruses that are often considered harmless and are neglected in clinical settings, could be linked with women's vaginal health. Anelloviruses are sensitive to host immune status, and their abundance are related to many disease conditions (Kaczorowska and van der Hoek, 2020; Liu et al., 2021; Hoeck et al., 2022; Kaelin et al., 2022; Li et al., 2022; Redondo et al., 2022). Even though, it’s not clear whether the high anellovirus abundance, diversity, as well as its correlation with HPVs in CC was the cause or result of different disease conditions, these data indicate a potential use of these commensal viruses to predict disease progression. For example, it was recently shown that a higher anellovirus abundance was associated with higher risk of the subsequent acquiring of bloodborne pathogen in drug users (Kandahil et al., 2021). Thus, with more evidence acquired in future studies, the early surveillance of the anellovirus traits may be used as an alternative adjunct biomarker to evaluate and identify women that are at higher risk of the development of cervical diseases.

Currently, limited knowledge is available about the role of the virome in women's genital health. The results of our study together with previous data suggest an important role of specific vaginal virome profiles in different cervical disease status. However, there are several unanswered questions could be considered in future studies. First, inflammation and local immunity are associated with HPV persistence and CC development
fig. 3. the association of anelloviridae with disease status and lactobacillus profiles. relative abundance and richness/shannon of anelloviruses are compared and shown for different disease groups (a, b, c), lac-dominant and lac-depleted groups (d, e, f) and cst (g, h, i). spearman’s correlations between ngs abundance (rpm) and qpcr viral load (copy numbers) of anelloviruses (j). statistical significance was assessed by mann-whitney test between two groups, and kruskal-wallis test with dunn’s correction among multiple groups. the number of individuals in each group is same to that in figs. 1 and 2. nt, normal tissue; lsil, low-grade squamous intraepithelial lesion; hsil, high-grade squamous intraepithelial lesion; cc, cervical cancer; cst, community state types.

fig. 4. maximum-likelihood phylogenetic tree (grt + f + r7 model) of anellovirus orf1 sequences. sequences identified in this study are highlighted with different colors.
Fig. 5. The association of Papillomaviridae with different disease status. The frequencies of high-risk HPVs (hrHPVs) are shown in disease groups (A), lac-dominant and lac-depleted groups (B) and CSTs (C). Relative abundance and richness/Shannon of papillomaviruses are compared and shown for different disease groups (D, E, F), lac-dominant and lac-depleted groups (G, H, I) and CSTs (J, K, L). Statistical significance was assessed by Mann-Whitney test between two groups; the differences between groups in panel A–C were compared using Fisher's exact test (adjusted by the Benjamini and Hochberg method) and Kruskal-Wallis Test with Dunn's correction among multiple groups in panel D–L. NT, normal tissue; LSIL, low-grade squamous intraepithelial lesion; HSIL, high-grade squamous intraepithelial lesion; CC, cervical cancer; CSTs, community state types.
Fig. 6. Maximum-likelihood phylogenetic tree (GTR + F + R10 model) of papillomavirus full length L1 sequences. Sequences identified in this study are highlighted with different colors.
the women's genital health (Madere and Monaco, 2022). This study preferentially focused on the eukaryotic viruses, and the role of prokaryotic virome in vaginal diseases deserves to be further investigated using an improved bioinformatic pipeline dedicated to phages.

5. Conclusions

The vaginal eukaryotic virome compositions are associated with different levels of lesion severity. Both the abundance and diversity of anelloviruses and papillomaviruses are associated with LSIL and CC. Besides, an altered viral correlation network, especially that between anelloviruses and papillomaviruses, was related to women with high lesion severity. These data indicate that the vaginal eukaryotic viral communities may play a critical role in maintaining a healthy microenvironment and influencing CC progression, and anellovirus may be used as a potential adjunct biomarker to predict the risk of developing to persistent HPV infection and/or CC.

Data availability

The short-read sequencing data are available at the NCBI Sequence Read Archive (SRA) under BioProject number PRJNA865010.

Ethics statement

Sample collection was approved by the Ethics Committees of Obstetrics and Gynecology Hospital of Fudan University. Written informed consents were obtained before sample collection.
Author contributions

Yanpeng Li: conceptualization, formal analysis, data curation, writing-original draft. Le Cao: formal analysis, methodology. Xiao Han: investigation, resources. Yingying Ma: methodology. Yanmei Liu: investigation, resources. Shujun Gao: conceptualization, resources, data curation, writing-review & editing, supervision. Chiyu Zhang: conceptualization, data curation, writing-review & editing, supervision.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

This study was supported by the funding from Shanghai Public Health Clinical Center (to CZ). We thank all the patients who participated in the study and the nurses who contributed to the sample collections.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.vjirs.2022.12.004.

References


