Transkingdom Analysis of the Female Reproductive Tract Reveals Bacteriophages Form Communities

Ferralita S. Madere[1], Michael Sohn[2], Angelina Winbush[3], Breóna Barr[4], Alex Grier[5], James Java[5], Tracy Meiring[6], Anna-Lise Williamson[6,7], Linda-Gail Bekker[8], David H. Adler[9], Cynthia L. Monaco[1,10*]

[1]*Department of Microbiology and Immunology, University of Rochester Medical Center, Rochester, NY, USA*

[2]*Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, NY, USA*

[3]*University of Rochester School of Medicine & Dentistry, Rochester, NY, USA*

[4] *Department of Rural Family Medicine, West Virginia University, Morgantown, WV*

[5] *UR Genomics Research Center, University of Rochester Medical Center, Rochester, NY, USA*

[6]*Institute of Infectious Diseases & Molecular Medicine and Division of Medical Virology, Faculty of Health Sciences, University of Cape Town, Anzio Road, Observatory, Cape Town, South Africa*

[7]*National health Laboratory Service, Groote Schuur Hospital, Cape Town, South African*

[8]*Desmond Tutu HIV Centre, Institute of Infectious Diseases & Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Anzio Road, Observatory, Cape Town, South Africa*

[9]*Department of Emergency Medicine, University of Rochester Medical Center, Rochester, NY, USA*

[10]*Department of Internal Medicine, Division of Infectious Diseases, University of Rochester Medical Center, Rochester, NY, USA*

[*]*Correspondence: Cynthia_Monaco@URMC.Rochester.edu*

Abstract

The female reproductive tract (FRT) microbiome plays an important role in maintaining vaginal health. Viruses play a key role in regulating other microbial ecosystems, but little is known about how the FRT viruses (virome), particularly bacteriophages, impacts FRT health and dysbiosis. We hypothesize that bacterial vaginosis is associated with alterations in the FRT virome, and these changes correlate with bacteriome shifts. We conducted a retrospective, longitudinal analysis of vaginal swabs collected from 54 bacterial vaginosis (BV)-positive and 46 BV-negative South African women. Bacteriome analysis revealed samples clustered into five distinct bacterial community groups (CG). Bacterial alpha diversity was significantly associated with BV. Virome analysis on a subset of baseline samples showed FRT bacteriophages clustering into novel viral state types (VSTs), a viral community clustering system based on virome composition and abundance. Distinct BV bacteriophage signatures included increased alpha diversity along with *Bacillus*, *Burkholderia* and *Escherichia* bacteriophages. Discriminate bacteriophage-bacteria transkingdom associations were also identified between *Bacillus* and *Burkholderia* viruses and BV-associated bacteria, providing key insight for future studies elucidating transkingdom interactions driving BV-associated microbiome perturbations. In this cohort, bacteriophage-bacterial associations suggest complex interactions which may play a role in the establishment and maintenance of BV.

45 **Keywords:** virome, microbiome, bacterial vaginosis, bacteriophage, transkingdom associations,

46 female genital tract

Introduction

The female reproductive tract (FRT) houses a compositionally dynamic environment where the host participates in an intricate interplay with a microbiome composed of bacteria and archaea (bacteriome), fungi (fungome), viruses (virome), and occasional prokaryotic parasites [1, 2]. The FRT microbiome plays an important protective role in maintaining vaginal health and preventing urogenital diseases such as bacterial vaginosis (BV), yeast infections, pre-term birth and sexually transmitted infections including HIV [3-5]. Prior studies of the FRT microbiome have primarily focused on determining bacterial composition and function. At least five different bacterial community groupings have been described within the FRT, distinguishable by the dominance of *Lactobacillus* species or presence of more diverse anaerobes [6, 7]. Prevalence of these communities varies by ethnic group, with majority of Caucasian women hosting *Lactobacillus*-dominant FRT microbiomes, whereas African women tend to be asymptomatically colonized by higher diversity FRT microbiota [7, 8]. *Lactobacillus*-dominant FRT bacteriomes, especially *L. crispatus*, protect against vaginal diseases by competitive exclusion against pathogenic bacteria for space and nutrients or by promoting an acidic vaginal environment via production of lactic acid and maintaining a low inflammatory state [9-11]. BV, the most common cause of vaginal discharge in reproductive-age women, is a symptomatic clinical condition characterized by a shift in the FRT microbiota away from a low inflammatory, *Lactobacillus*-dominant microbiome to more diverse community including facultative anaerobes. BV is associated with an increased risk of sexually transmitted infection (STI) acquisition and pre-term birth [12, 13]. FRT bacteriome shifts can occur rapidly and may be related to shifts in bacteriophage populations [14, 15]. Specific BV-

associated bacteria include *Gardnerella vaginalis*, *Prevotella*, *Fusobacterium*, *Atopobium vaginae*, *Megasphaera*, and *Sneathia* among others [16].

Viruses rival bacterial numbers in the microbiome and are more diverse [1]. However, studies of the virome have been limited in part due to lack of a common viral genetic element analogous to the bacterial 16S rRNA gene, as well as the high genetic diversity between viral species [1, 17]. The FRT virome is home to eukaryotic viruses and bacteriophages [1]. Compared to the FRT bacteriome, little is known about the viral communities of the FRT and how their interactions with bacteria contribute to disease states such as bacterial vaginosis (BV). The few prior studies that have examined the FRT virome have mainly concentrated on the DNA eukaryotic virome, finding *Papillomaviridae, Polyomaviridae, Herpesviridae, Poxviridae, Adenoviridae*, and *Anelloviridae* present [18, 19]. However, bacteriophages are the largest and most abundant viral group and can modulate bacterial composition and abundance, suggesting that they may play an important role in regulating bacterial composition of the FRT microbiome [17, 20]. Bacteriophages may be lytic, hijacking bacterial host replication machinery in order to replicate and then lysing the bacterial host to release virions, or lysogenic, whereby bacteriophage DNA is integrated into the bacterial host genome as a prophage and replicates with bacterial genome replication [21]. This lysogenic lifestyle can result in generalized transduction of bacterial genes between bacterial hosts that can confer increased fitness through methods such as toxin production, carbohydrate metabolism, or antibiotic resistance [22]. Upon environmental stress, lysogenic bacteria can become lytic, and therefore may serve to regulate bacterial populations in unfavorable host conditions.

Data on bacteriophage populations in the FRT are limited. One study revealed numerous *Caudovirales* order bacteriophage in the FRT; however, no relationship was found between FRT *Caudovirales* bacteriophages and bacterial populations [4]. Additionally, other investigations studying BV-positive and BV-negative samples in a cohort of Danish women found no significant difference in viral nor bacterial alpha diversity between BV-positive and BV-negative women [23]. While brief characterization of the vaginal microbiome in pregnant women found an absence of *Microviridae* and *Herelleviridae* bacteriophage families in a pregnant woman without vaginitis [24]. Neither of these studies, however, employed significant sequencing depths necessary to perform a through characterization of the FRT phageome and identify distinct perturbations between health and disease. *Lactobacillus* bacteriophages have been isolated from human FRT *Lactobacillus* species and have been observed to have broad host ranges, capable of infecting multiple *Lactobacillus* species [24]. Upon induction, lysogenic phages within lactobacilli can lyse their host and enable pathogenic bacteria to flourish, suggesting their possible role in regulating bacteriome composition and promoting the growth of BV-associated bacteria [25-28]. Further in support of this hypothesis, high gene expression of the CRISPR anti-bacteriophage defense system occurs in BV, suggesting that an altered phage load could contribute to the hallmark dysbiosis observed in BV [29-32].

Herein we investigated FRT virome composition and transkingdom bacterial-bacteriophage associations within the FRT. We assessed the vaginal samples of 100 young, sexually active BV-positive and -negative South African women to identify discriminate viral and bacterial signatures in health and FRT disease. We show for the first time that FRT DNA bacteriophage populations cluster into community groupings that correspond to bacterial community groupings, and that

specific bacteriophages correlate with bacteria associated with and protective against BV. These findings improve our understanding of the transkingdom associations in the FRT microbiome and the impact that these could have on the induction and pathogenesis of disease.

Methods

Study Cohort

De-identified vaginal swabs from the University of Cape Town HPV-HIV study [33] were retrospectively used for bacteriome and virome analysis. This cohort was comprised of 50 HIV-positive and 50 HIV-negative young, sexually active women between the ages of 16 and 21 recruited from the youth community center and clinic in two urban disadvantaged communities (Masiphumelele and Mthatha townships) in Cape Town, South Africa between October 2012 and October 2014. Informed consent was obtained from all participants above 18 years of age and parental consent was obtained for participants of age 17 or younger. This study was approved by the Institutional Review Board at University of Rochester Medical Center and the Human Research Ethics committee at the University of Cape Town. Vaginal swabs were self-collected approximately every 6 months by subjects using Dacron swabs high within the vagina, placed in Digene transport media, and frozen at -80°C until use. Pap smears were taken at baseline and at least one other visit, and tests for HPV, *Trichomonas*, and BV performed [33]. HIV status was confirmed upon enrollment, and CD4+ T cell count for those who were HIV-positive was determined. Exclusion criteria included prior HPV vaccination or cervical surgery. Sexual history, recent contraceptive methods and current HIV treatment interventions were also queried at study entry.

Bacterial 16S rRNA gene amplicon sequencing

Total nucleic acid was extracted from 253 resuspended vaginal swab samples using the MagNA Pure Compact Nucleic Acid Isolation Kit (Roche) [33]. 16S rRNA gene amplicon sequencing was performed with primers specific to the V3-V4 region [34] followed by amplicon pooling, bead-based normalization and sequencing on the Illumina MiSeq platform at 312 bp paired-end reads (University of Rochester Genomics Research Center, UR GRC). Water processed similarly to samples and pre-defined bacterial mixtures (Zymo) were run as negative and positive controls, respectively. Eleven samples failed 16S rRNA gene amplification.

Bacterial 16S rRNA gene amplicon analysis

Raw data from the Illumina MiSeq was first converted into FASTQ format 2×312 paired-end sequence files using the bcl2fastq program (v1.8.4) provided by Illumina. Format conversion was performed without de-multiplexing, and the EAMMS algorithm was disabled. All other settings were default. Reads were multiplexed using a configuration described previously [34]. The extract_barcodes.py script from QIIME (v1.9.1) [35] was used to split read and barcode sequences into separate files suitable for import into QIIME 2 (v2018.11) [36], which was used to perform all subsequent read processing and characterization of sample composition. Reads were demultiplexed requiring exact barcode matches, and 16S primers were removed allowing 20% mismatches and requiring at least 18 bases. Cleaning, joining, and denoising were performed using DADA2 [37]: reads were truncated (forward reads to 260 bps and reverse reads to 240 bps), error profiles were learned with a sample of one million reads per run, and a maximum expected error of two was allowed. Taxonomic classification was performed with custom naïve Bayesian classifiers trained on target-region specific subsets of the August, 2013 release of GreenGenes

[38]. Sequence variants that could not be classified to at least the phylum level were discarded. Sequencing variants observed fewer than ten times total, or in only one sample, were discarded. Vaginal samples with fewer than 10,000 reads and/or features present in less than 20 samples were discarded. Four samples and all negative controls did not achieve sufficient sequence variants for downstream analysis leaving 238 samples that were included in the final analysis. Phylogenetic trees were constructed using MAFFT [39] for sequence alignment and FastTree [40] for tree construction. For the purposes of diversity analyses, samples were rarefied to a depth of 10,000 reads. Faith's PD and the Shannon index were used to measure alpha diversity, and weighted Unifrac [41] was used to measure beta diversity.

Lactobacilli DNA Extraction and qPCR Analysis

Known concentrations of *Lactobacillus iners* (ATCC 55195), *Lactobacillus crispatus* (ATCC 33820) *Lactobacillus gasser*i (ATCC 33323), and *Lactobacillus jensenii* (ATCC25258) genomic DNA extracted using isopropanol precipitation, was used to measure the detection limit of the species-specific 16S rRNA gene qPCR assays and to generate standard curves for quantifying assay results. DNA was previously extracted from vaginal swabs as described in [42] and along with lactobacilli genomic DNA was quantified using the Qubit 4 flurometer (Invitrogen Inc., Carlsbad, CA) with the Qubit dsDNA high-sensitivity assay (Invitrogen Inc., Carlsbad, CA). 10-fold serial dilutions (10^0 to 10^8 copies) were used to generate standard curves of extracted genomic DNA. The range of slopes for the qPCR assays was from -3.7 to -4.7, and r^2 values were all > 0.99.

qPCR primers targeting the 16S rRNA genes of *L. iners*, *L. crispatus*, *L. gasser*i , and *L. jensenii* were previously developed in [43]. SYBR green-based qPCR assays were performed on a CFX96

Touch Real-Time PCR Detection System (Bio-Rad, Hercules, CA). The reaction mixture (20 µL) contained 10 µL iQ SYBR Green Supermix (Bio-Rad), 2 µL each of forward and reverse primers to a final primer concentration of 100 nM each except for the assay for *L. iners* which had used 4 µL each to a concentration of 200 nM., and 10 µL template DNA (10 ng). Temperature cycling for all assays was polymerase activation at 95°C for 3 min, followed by 40 cycles of amplification with denaturation at 95°C for 15 seconds, followed by annealing/ extension at 55 to 58°C for 1 min (*L. gasser*i , *L. iners* and *L. jensenii* at 55°C and *L. crispatus* at 58°C). Fluorescence was measured at the final step of each cycle. Following amplification, melting curve analysis was performed by heating at 0.5°C increments from 55 to 95°C, temperature was held for 5 s at each step, and fluorescence was acquired at each temperature increment. For each qPCR assay, vaginal swabs and extracted lactobacilli genomic DNA standards were run in triplicate, and the average values were used to calculate 16S rRNA gene copy number per 10 ng total vaginal swab DNA. Negative (no DNA water) controls were run with every assay to check for contamination.

Virus-Like Particle Preparation, Library Construction and Sequencing

Virus-like particle (VLP) preparation was adapted from methods previously described in [44]. Briefly, vaginal specimens were resuspended in 200 µL of Digene transport media and an equal volume of SM buffer was added (50mM Tris-HCl, 8mM magnesium sulfate, 100mM sodium chloride, and 0.01% gelatin, pH 7.5, Fisher) and mixed by vortexing for 5 minutes. Specimens were centrifuged at 2000 x g and filtered using a 0.45 µM filter to remove intact cells and bacteria. Samples underwent lysozyme treatment (1ug/mL at 37C° for 30 minutes) (Sigma-Aldrich) to degrade remaining host cell and bacterial membranes. DNase digestion (Turbo DNase Buffer, Turbo DNase, Baseline Zero) (Ambion) was performed to remove contaminating bacterial and

host DNA followed by heat inactivation of the DNase at 75C° for 15 minutes. Enriched VLPs were lysed with 10% SDS and 20mg/mL Proteinase K (Ambion) at 56C° for 20 minutes, followed by treatment with CTAB (10% Cetyltrimethylammonium bromide, 0.5 M NaCl, nuclease-free water, filtered through 0.22 μM filter) at 65C° for 20 minutes. Phenol: Chloroform: Isoamyl Alcohol (Invitrogen, pH 8.0) nucleic acid extraction was performed, the resulting aqueous fraction then washed with an equal volume of chloroform and concentrated through isopropanol precipitation. Extracted nucleic acid was aliquoted and stored at -80C° until use. NEBNext Ultra II FS DNA Library Prep Kit (New England Biolabs) was used for library construction with NEBNext Multiplex Oligos for Illumina Dual Index Primers (New England Biolabs). Following equimolar pooling, DNA libraries were sequenced on the Illumina NovaSeq platform (UR-GRC) generating an average of over 29 million 150bp paired-end reads per sample.

Virome Analysis Pipeline

The VirusSeeker pipeline [45] was deployed on a Linux cluster on raw sequence data. Briefly, raw sequences from samples went through sample pre-processing steps that included adapter removal, stitching of reads, quality filtering, and CD-HIT was used to minimize sequence redundancy and define unique sequences (98% identity over 98% of the sequence length). Sequencing reads underwent human genome filtering, then unmapped reads were sequentially queried against a customized viral database comprised of all viral sequences in NCBI using BLASTn (e-value cutoff 1E−10), followed by BLASTx (e-value cutoff: 1E−3). False positive viral sequences were identified by successively querying the candidate viral reads against the NCBI NT database using MegaBLAST (e-value cutoff: 1E−10), BLASTn (e-value cutoff: 1E−10), and the NCBI NR

database using BLASTx (e-value cutoff: 1E−3). All sequences that aligned to viruses were further classified into viral genera and species based on the NCBI taxonomic identity of the top hit.

Statistical Analysis

Descriptive statistics were used to summarize the characteristics of the study population. Mean was used for continuous variables, and frequency or proportion was used for categorical variables. Alpha diversity was measured by Shannon's diversity index and analyzed using a linear mixed effects model for bacteria data and a linear regression model for bacteriophage data [46]. Beta diversity was measured by the weighted UniFrac distance for bacteria data and by the Bray-Curtis distance for bacteriophage data [44, 47]. PERMANOVA was used to quantify dissimilarity in beta diversity [48]. Clustering was performed using partitioning around medoids, with the number of clusters estimated by maximum average silhouette width [49]. For differential abundance analysis, the relative abundance was first arcsine-transformed, and then univariate analysis was performed using mixed effects models for bacteria data and linear regression models for bacteriophage data. Univariate analysis of correlations between bacteria and bacteriophage was performed using Kendall's rank correlation coefficient [50]. All analyses were performed in R and Prism version 8.3.0 for Windows (GraphPad Software, La Jolla California USA, www.graphpad.com) [51]. *P* values in the univariate analyses were adjusted for multiplicity using the Benjamini-Hochberg procedure [52].

Results

Cohort Characteristics

Fifty HIV-positive and 50 HIV-negative young, sexually active South African women ages 16-21 were recruited in Cape Town, South Africa between October 2012 and October 2014 as part of the University of Cape Town HPV-HIV study to investigate high-risk HPV persistence in HIV-positive women as previously reported [33, 42]. Vaginal swabs were self-collected at six-month intervals for up to six consecutive visits along with medical and sexual history, laboratory data and demographic information. Pap smears and STI testing were performed at the baseline visit and each year. HIV-positive subjects had an average CD4+ T cell count of 477.5 cells/μL, and 22 (44%) were not on ART at the baseline study visit. There was an increased risk of HPV for HIV-positive participants at baseline (OR 5.299, 95% CI 2.048 to 13.73). This cohort included 54 BV-positive and 46 BV-negative subjects at baseline (Table 1). The median age of both BV-positive and BV-negative women was 19. BV-positive subjects were more likely to have HPV (p=0.0940) and a higher prevalence of high-risk HPV subtypes (p=0.0005) compared to BV-negative subjects. BV-positive subjects also exhibited higher rates of abnormal pap smears (p=0.1181). These data demonstrate that our cohort behaves similarly to other published cohorts and was suitable for further study [53].

Table 1: Cohort Characteristics

Subject Characteristics	**BV Positive (n=54)**	**BV Negative (n=46)**	***p*-value**
Age (years), mean (Interquartile range; IQR)	19.2 (16-21)	18.8 (16-21)	0.2352
Laboratory Results			
HIV Positive Samples, n (%)	29 (53.70)	21 (45.65)	0.5475
HPV Positive Samples, n (%)	39 (60.94)	25 (39.06)	0.0940

High Risk HPV Types Present in Positive Samples, n (%)	29 (76.32)	9 (23.68)	0.0005
Visits with Abnormal Pap Smear, n (%)	13	5	0.1181
Smoking History			
Smoker, n (%)	5 (5)	4 (4)	>0.9999
Non-Smoker, n (%)	49 (49)	42 (42)	
Sexual History			
History of STI, n (%)	27 (57.45)	20 (42.5)	0.5514
Lifetime Sexual Partners			
1, n (%)	11 (11)	4 (4)	
2-5, n (%)	39 (39)	39 (39)	0.2482
>5, n (%)	4 (4)	3 (3)	
Sexual Partners in the Last 6 Months			
1, n (%)	51 (51)	44 (44)	>0.9999
2-5, n (%)	3 (3)	2 (2)	
Form of Contraception			
None, n (%)	1 (1)	1 (1)	
Condom, n (%)	50 (50)	40 (50)	0.3712
Injection, n (%)	30 (30)	31 (31)	
Pill, n (%)	2 (2)	3 (2)	

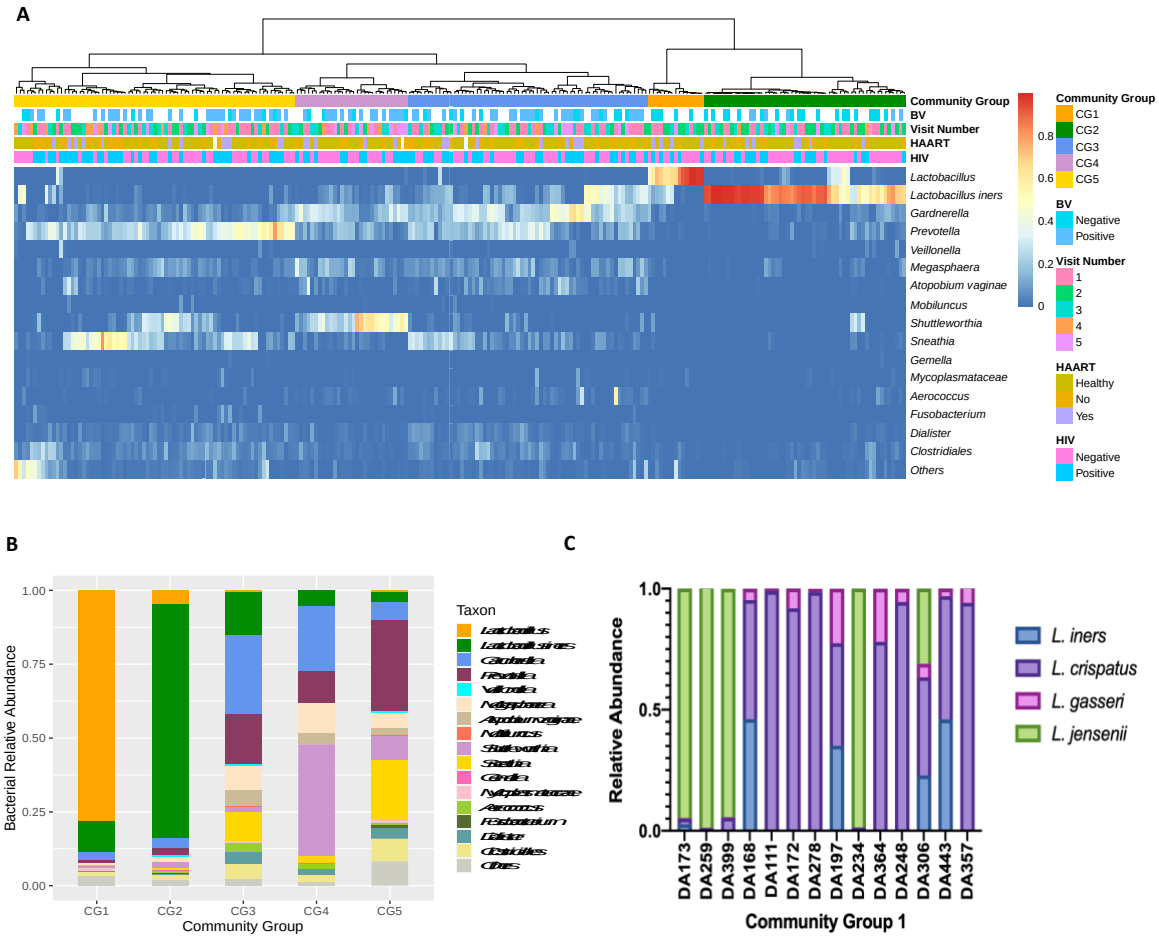
HIV viral load and HAART regimen not available for HIV-negative subjects, 22 (44%) HIV-positive participants on HAART. Hight risk HPV include 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73 and 82. For continuous variables, Mann-Whitney and Kruskal Wallis tests were used; for comparing categorical variables, Chi-square and Fisher's exact tests were used.

The FRT bacteriome clusters into community groups

Prior studies examining Western cohorts have identified five bacterial community state types (CSTs), of which type I-III and V were *Lactobacillus*-dominant while CST IV was comprised of polymicrobial communities [7]. However, studies examining the FRT bacteriome in African cohorts have revealed a different pattern in hierarchical clustering analysis, with more community groupings of high diversity bacteriomes, consistent with the increased prevalence of high diversity FRT bacteriomes in this population [4, 54]. To further asses the bacterial communities within the FRT bacteriome seen in African women, 253 vaginal swabs were processed and underwent 16S rRNA gene amplicon sequencing of the V3-V4 region [34]. Eleven samples did not amplify, and four failed to achieve sufficient reads for downstream analysis, leaving 238 samples for bacteriome analysis and sequence identification using QIIME2 (Fig.1A).

Hierarchical clustering analysis of all visits by community abundance and composition identified five unique bacterial community clusters named herein bacterial community groups (CG), which were distinguished by *Lactobacillus*-dominance (CG1 and 2) or higher diversity bacteriomes (CG3-5; Fig. 1A). The identified CG were compositionally different from conventional CSTs [7]. Similar to other African cohorts [7], the majority (n=173, 72.7%) of subjects had high diversity FRT bacteriomes with a low prevalence of *Lactobacillus*-dominant bacteriomes. Samples that were dominated by a single species, defined as >50% community composition, made up 42.0% of all samples and mostly clustered with CG1 and 2, while the remaining samples showed no individual dominant species and mainly clustered in CG3-5. CG1 (n=15, 6.3%), a low diversity FRT bacteriome, was comprised almost exclusively of *Lactobacillus* species (78%) that were unable to be further delineated by 16S rRNA gene amplicon sequencing (Fig. 1B). To further define the primary bacterial constituents of CG1, qPCR of 16S rRNA gene sequences from

dominant FRT *Lactobacillus* species *L. iners*, *L. crispatus*, *L. gasseri* and *L. jensenii* [8, 54] was performed, revealing that approximately 70% of samples in GC1 were *L. crispatus*-dominant, similar to what has been previously described as CST I [55] (Fig. 1C). One vaginal swab in CG1 contained sufficient volume to only perform qPCR for *L. iners* and *L. crispatus*, and *L. crispatus* was most abundant (not shown). *L. jensenii* tended to predominate when present (Fig. 1C). CG2 (n=54, 22.7%) was *L. iners*-dominant, with a few samples showing notable amounts of *Gardnerella* and *Prevotella* as well (Fig. 1B). The second to largest and most diverse group, CG3 (n=64, 26.9%), consisted mainly of *Gardnerella*, *Prevotella* and *L. iners* (Fig. 1B). The smallest high diversity CG was CG4 (n=30, 12.6%), in which *Shuttleworthia* and *Gardnerella* were predominant. CG5 contained the largest number of samples (n=75, 31.5%) and was dominated by *Sneathia* and *Prevotella* (Fig. 1B). CG3, CG4 and CG5 exhibited significantly higher alpha diversity than CG1 and 2 (Fig. 2A; p = 0.0001, 0.0065, and <0.0001 respectively). Beta-diversity significantly differed between these five bacterial CG (p=0.0001; Fig. 2B).



309

310 **Figure 1: Bacteriome Profiling by Community Group of Self-Collected Vaginal Swabs from**

311 **South African Women.** (A) Relative abundance of 16 most frequent bacterial taxa (y-axis) by

312 sample (x-axis), grouped by community group (CG), BV Status, visit number, HAART status and

313 HIV status (color key shown). Percent abundance is indicated by gradient key. Using Ward's

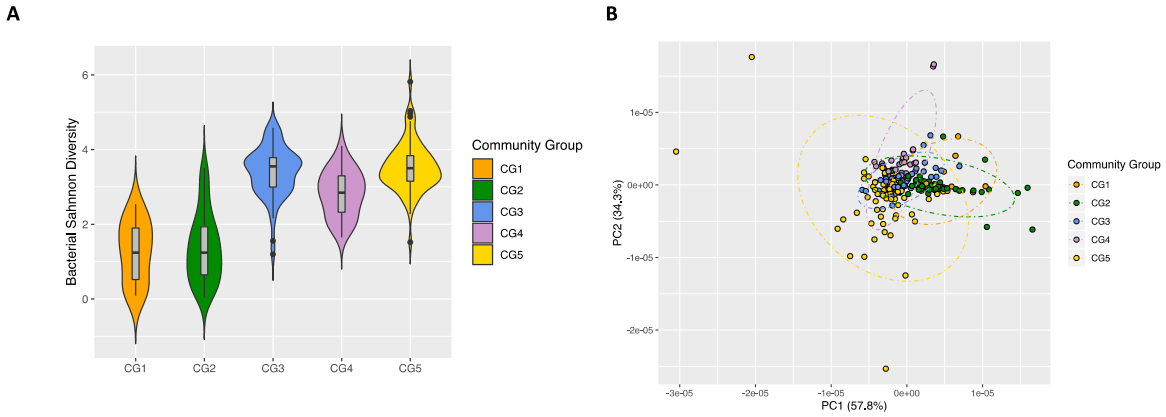
314 linkage hierarchical clustering, samples clustered into five distinct bacterial community profiles

315 called community groups (CG). (B) Average bacterial community group structure for each of the

316 five community groups (x-axis) based on relative abundance (y-axis). (C) Bar plot showing the

317 relative abundance of L. iners, L. crispatus, L. gasseri and L. jensenii bacterial species (relative

318 abundance of 16S rRNA copies per 10 ng total DNA) as determined by qPCR of vaginal swabs
319 that clustered into CG1.



320

321 **Figure 2: FRT Bacteriome Clusters into Distinct Community Groups That Differ by Alpha**
322 **and Beta Diversity.** (A) Bacterial Shannon diversity (y-axis) by CG (x-axis) as determined by
323 linear mixed effects model. Center bar represents median; grey box bounded by upper/lower
324 interquartile ranges (IQR); Whiskers represent range; Dots represent outliers; Color-filled areas
325 are representative of density/distribution of diversity values. (B) Principle Coordinate Analysis
326 (PCoA) plot of the weighted UniFrac distances colored by community group.

327

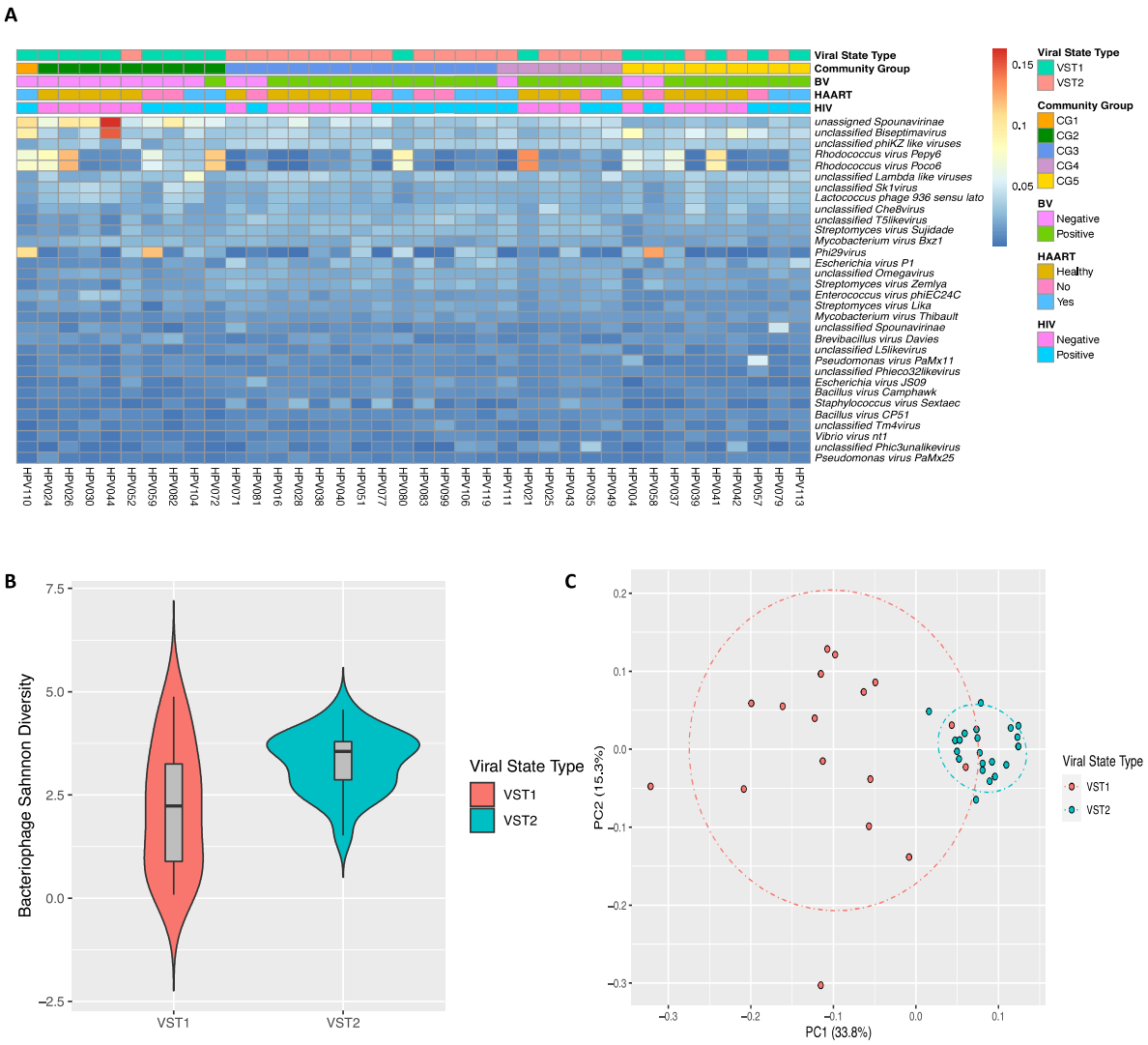
328 ## Bacteriophages comprise the majority of the FRT DNA virome

329 While the FRT bacteriome has been well-studied, the FRT virome, especially bacteriophage
330 populations, is relatively unknown. We therefore characterized the FRT DNA virome using a
331 subset of baseline samples by enriching for virus-like particles (VLPs) from resuspended vaginal
332 swabs and extracting viral nucleic acid [44]. Libraries were constructed and sequenced using the
333 Illumina NovaSeq platform for 38 baseline samples, 14 of which were BV-negative and 24 were
334 BV-positive. Resulting viral sequences underwent quality control, removal of bacterial and human

reads, and then were assigned to known viral taxa using VirusSeeker, a BLAST-based NGS virome analysis pipeline [45].

On average there were 29 million reads per sample with 86.8% of them being high quality. The DNA eukaryotic virome was comprised almost entirely of *Papillomaviridae.* However, FRT bacteriophages were abundant. Samples contained sequences identified as belonging to *Myoviridae, Siphoviridae, Podoviridae, Inoviridae, Ackermannviridae, Microviridae, Lipothrixviridae, Plasmaviridae* and *Tectiviridae* bacteriophage families. Sequences assigned to members of the *Caudovirales* order, lytic tailed dsDNA bacteriophages, including *Myoviridae, Siphoviridae* and *Podoviridae*, were the most abundant in all samples regardless of BV, HAART, or HIV status.

Bacteriophage communities within the FRT clustered into two distinct, novel bacteriophage community groups based on composition and abundance that we have termed viral state types (VSTs) (Fig. 3A). VST1 represented 44.7% (n=17) of all samples while VST2 contained the remaining 55.3% (n=21). Bacteriophage Shannon diversity differed between VSTs (Fig. 3B), with VST2 having the highest diversity bacteriophage populations. The VSTs also grouped distinctly by beta diversity analysis (Fig. 3C, PERMANOVA, p =0.0001). Neither VST exhibited a dominant bacteriophage member. VST1 contained several samples with high relative abundance of *Rhodococcus viruses*, *Spounavirinae*, *phi29 virus* and *Biseptimavirus*-assigned bacteriophage reads. VST2 was comprised of a more even distribution. This is the first study to identify bacteriophage community groups in the FRT.



357

358 **Figure 3: FRT DNA Bacteriophages Cluster into Two Unique Community Groups.** Self-
359 collected vaginal swabs were processed for DNA virome analysis by enriching for viral nucleic
360 acid, libraries built, and sequenced. (A) Relative abundance of the 32 most frequent bacteriophage
361 species (y-axis) by sample (x-axis). Ward's linkage hierarchical clustering analysis was used to
362 cluster samples into distinct bacteriophage community profiles called viral state types (VSTs).
363 VST, BV Status, HAART status and HIV status (color key) are shown. Percent abundance is
364 indicated by gradient key. (B) Bacteriophage Shannon diversity (y-axis) by VST (x-axis) as
365 determined by linear regression model. Center bar represents median; grey box is bounded by

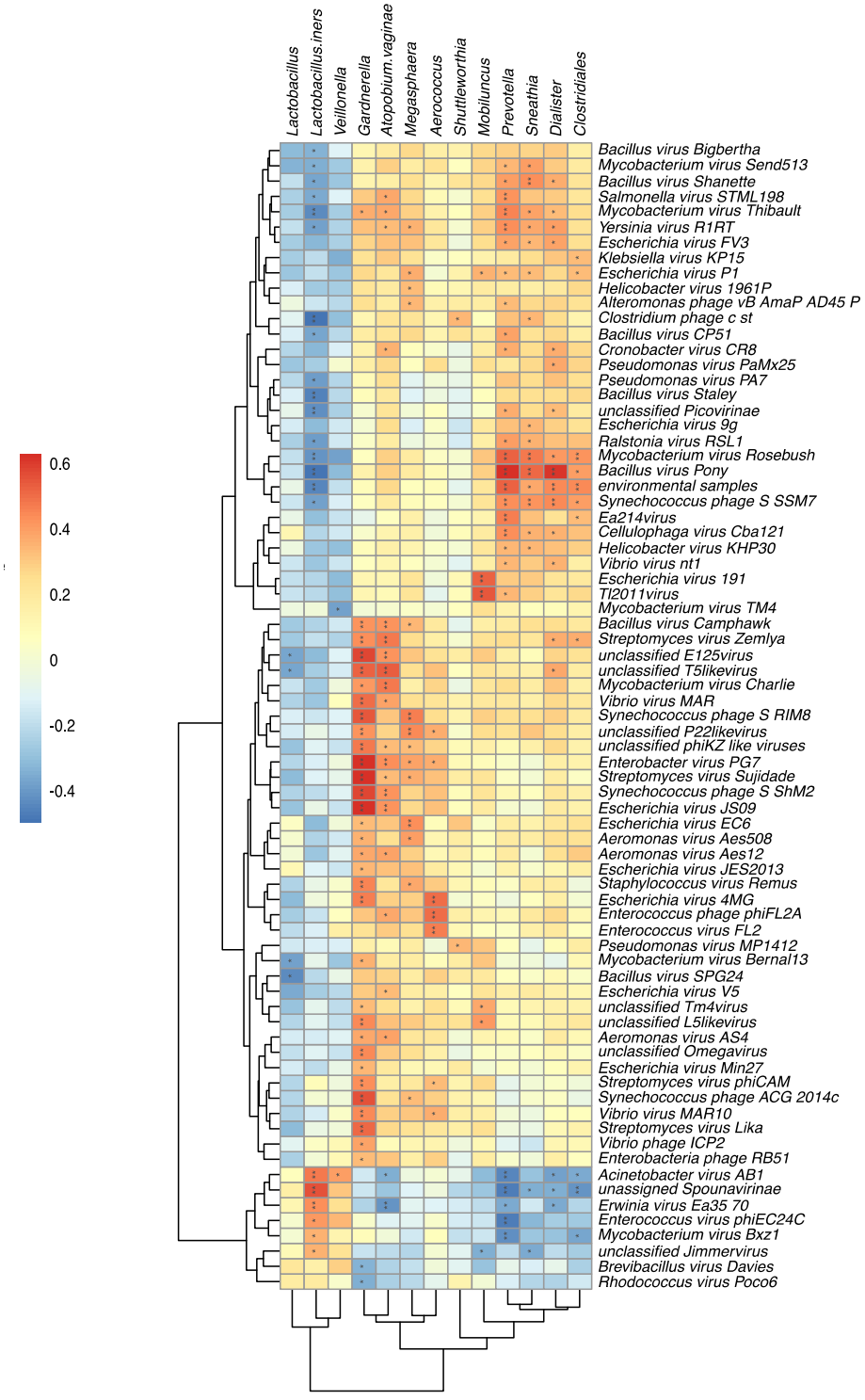
upper/lower interquartile ranges (IQR); whiskers represent range; dots represent outliers; color-filled areas are representative of density/distribution of diversity values. (C) Principle Coordinate Analysis (PCoA) plots of beta diversity distances, as determined by Permutational multivariate analysis of variance, colored by VST.

Transkingdom Associations within the FRT Microbiome of South African Women

Bacteriophage can directly impact bacterial composition and abundance through infection of their host. Therefore, we investigated the transkingdom associations between bacteriophage and bacteria in the FRT. The VSTs significantly correlated with bacterial CG (p=0.00015; Fig. 3A), with VST1 associated with CG1 and 2, the *Lactobacillus*-dominant groups, and VST2 associated with CG3 and 4, both higher diversity CG. CG5 contained samples belonging to both VSTs (Fig. 3A). These data indicate a strong association between bacteriophage communities and the host bacterial populations.

To further investigate specific bacteriophage-bacterial interactions, correlations between FRT bacterial composition and bacteriophage composition were identified by Kendell's rank correlation coefficient. Reads assigned to bacteriophages *Bacillus virus Camphawk* and *Bacillus virus Pony*, which infect members of the *Bacillus* genus, positively correlated with the BV-associated bacteria *Gardnerella, A. vaginae*, *Prevotella*, *Sneathia* and *Dialister* (Fig. 4). Reads assigned to unclassified bacteriophages of the *E125* genus also positively associated with *Gardnerella* and *A. vaginae* (Fig. 4). A number of *Bacillus*-infecting phage including, *Bacillus virus Pony and Bacillus virus Staley*, were inversely associated with bacteria protective from BV,

389 particularly *L. iners* (Fig. 4). Interestingly, in this cohort, bacteriophage associations revealed
390 *Veillonella* more closely grouped with *Lactobacillus* rather than BV-associated bacteria despite
391 the known role of *Veillonella* in lactose fermentation and higher diversity FRT microbiomes [7].
392 These data suggest that bacteriophages directly or indirectly interact with FRT bacterial
393 populations in disease states, and identify putative FRT bacteriophage-host networks that may play
394 a role in development and maintenance of BV.



395

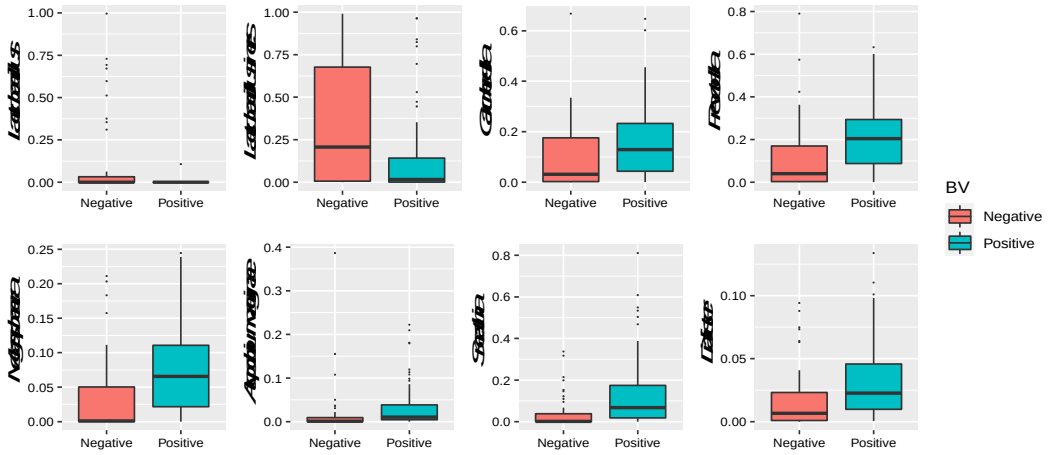
Figure 4: Transkingdom Associations within the FRT of South African Women. Heatmap of estimated Kendall's correlation coefficients between FRT bacterial taxa (x-axis) and sequences assigned to bacteriophage (y-axis). Multiple comparisons correction by the Benjamini-Hochberg procedure. * indicates $p<0.05$, ** $p< 0.01$. Magnitude and sign of the Kendall's rank correlation coefficient is indicated by gradient key. Red indicates positive correlations; blue indicates negative correlations.

Effects of Bacterial Vaginosis on the FRT Virome

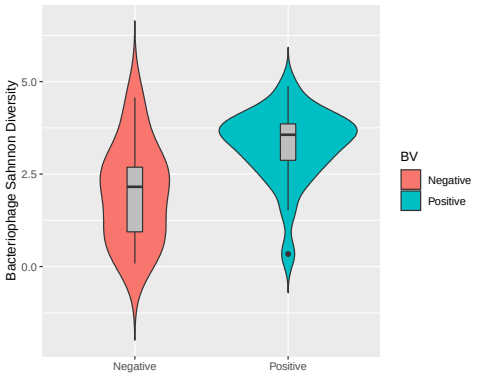
We next examined the impact of different disease states on this cohort. BV is a clinically significant condition with high morbidity characterized by high bacterial diversity [12, 13]. Our data showed significant associations between high diversity bacterial CG and VSTs, and further suggested specific bacteriophage interactions with BV-associated bacteria. Similar to published cohorts [2], BV in our cohort was positively associated with increased bacterial alpha diversity compared to healthy subjects (p=0.0001). The high diversity CG3-5 (p=0.0001) also correlated positively with BV. Bacterial taxa including *Gardnerella, Prevotella, Sneathia and Megasphaera* (p=0.000439; Fig. 5A) were linked to clinical BV status, corroborating distinct bacterial signatures associated with BV. Of particular clinical relevance, we additionally sought to identify specific bacterial taxa that were predictive of initiation and recovery from BV. Relative abundance of bacterial genera less frequently observed in clinical BV, Aerococcus (p=0.00216) and Gemella (p=0.00746), was significantly increased among participants who recovered from or transitioned to BV, suggesting members of this genera as possible regulators of FRT bacteriome structure in health and BV.

Since clinical BV was associated with significant changes in FRT bacterial composition and correlated with bacterial CG, and bacteriophage VSTs also correlated with bacterial CG, we sought correlations between bacteriophage populations and clinical BV. Regression analysis accounting for HIV status and VST revealed bacteriophage Shannon diversity significantly differed by clinical BV diagnosis (p=0.0193) (Fig. 5B). We then identified specific bacteriophage taxa differentially abundant in the FRT of BV-positive and -negative women. *Bacillus*-infecting bacteriophages are known to belong to the *Herelleviridae* and *Podoviridae* families [56, 57]. Members of these families, including *Bacillus virus Camphawk* and *Bacillus virus Pony* (Fig. 5C), which were significantly associated with BV diagnosis (p= 0.019058 and p=0.014546, respectively). BV diagnosis also strongly correlated with the bacteriophages *Escherichia virus FV3* and *unclassified E125 virus*, which are known to infect the BV-associated bacteria *Escherichia coli* and *Burkholderia*, respectively [58, 59]. Together, these data uncover a link between highly diverse FRT bacteriophage populations, a distinct subset of bacterial hosts, and BV.

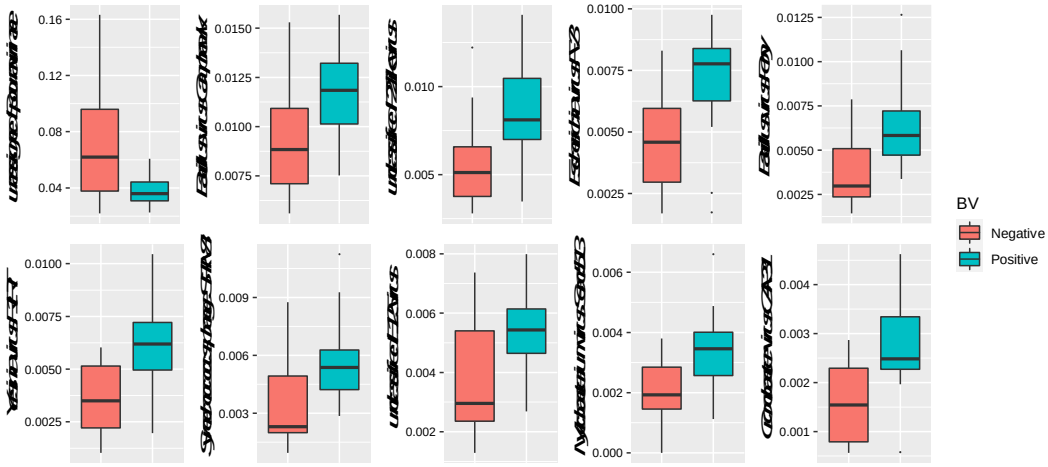
A



B



C



433

Figure 5: Discriminant FRT Bacterial and Bacteriophage Species Associated with Bacterial Vaginosis. (A) Discriminant bacterial taxa by BV status was determined by univariate analysis using mixed effects models. Relative abundance is represented on the y-axis, BV-negative subjects are shown in orange and BV-positive in green (x-axis). (B) Bacteriophage Shannon diversity (y-axis) by BV status (x-axis) as determined by a linear regression model. (C) Discriminant bacteriophage species by clinical BV diagnosis was determined by univariate analysis using a linear regression model. Relative abundance is represented on the y-axis, BV-negative subjects are shown in green and BV-positive in orange (x-axis).

The Effect of HIV and HPV on the FRT Microbiome

We also examined the effect of HIV on the FRT microbiome in this cohort. We found no significant difference in bacterial richness, alpha or beta diversity between HIV-positive and HIV-negative subjects. Further, there were no significant alterations in bacteriome or bacteriophage diversity by HIV status. Thus, FRT bacterial and bacteriophage communities were not detectably altered by HIV infection in this cohort, suggesting that localized infections may play a more important role in the FRT composition than systemic infections.

We also examined the relationship between HPV, the main eukaryotic virus found, and bacterial populations. Upon examination of 63 HPV-positive and 37 HPV-negative baseline samples, there were no significant associations between HPV infection and bacteriophage or bacterial diversity. Analysis of HPV subtypes revealed increased bacterial alpha diversity in HPV6-positive subjects (p=0.0181), suggesting certain HPV subtypes may directly or indirectly benefit from the presence

of higher diversity bacterial populations, although this analysis may have been underpowered for less prevalent subtypes.

Discussion

The FRT is a dynamic ecosystem in which bacteriophage and bacterial communities establish complex connections that influence the host environment and the physical manifestation of gynecological diseases. Bacterial population shifts are well-established contributors to FRT disease states including BV [7]. However, whether shifts in the FRT viral community occur in disease states concurrently with bacterial community perturbation was unknown. This study is the first to offer a comprehensive characterization of both the FRT bacteriome and DNA virome, with a particular emphasis on bacteriophage composition and transkingdom interplay, utilizing a South African cohort of BV-affected women.

Globally, the most common clinical presentation of vaginal bacterial dysbiosis is BV, a condition which poses a significant threat to female reproductive health and STI acquisition [2, 12, 13]. Here, similar to other studies, we find that women suffering from BV have distinct and compositionally diverse FRT bacteriomes defined by a loss of *Lactobacillus* dominance and gain of facultative anaerobes [7]. However, for the first time, our data revealed novel bacteriophage communities we have termed VSTs, which correlated to bacterial communities and clinical diagnosis of BV. Although bacteriophages have been studied at other mucosal sites in humans including the gut [60], this is the first description of bacteriophage community groupings and may be unique to the FRT environment due to the distinctive bacterial communities present. Few prior studies have examined the bacteriophage populations in the FRT. Nonetheless, in contrast to our findings, one

group examining FRT bacteriophage populations found no distinct bacteriophage community structures within their cohort, nor differences in bacteriophage composition among the bacterial communities [4]. Similar to our study, they also utilized a South African cohort; however, they focused analysis on the *Caudovirales* family rather than all bacteriophage populations, which likely impacts differences observed. Additionally, our findings contrast with those preciously described by Jakobsen et. al [23], who found no significant difference in viral nor bacterial alpha diversity between BV-positive and BV-negative samples in a cohort of Danish women undergoing IVF treatment for non-female factor infertility. However, their analysis was limited by the cross-sectional nature of the study and meager modest sequencing depth. Our viral sequencing depth of 29 million reads per sample was greater than 640-fold higher than that employed in Jakobsen et al. (average of 44,686 reads per sample) and is likely a major contributor to our unique findings. Additionally, Zhang et al. briefly characterized the vaginal microbiome in pregnant women finding absence of Microviridae and Herelleviridae in a pregnant woman without vaginitis [24]. This study was limited in statistical analysis due to the pooling of subjects into size sequencing libraries and similar to Jakobsen et al., Zhang et al. used a sequencing depth an order of magnitude lower that what was employed in our current investigation. Validation of our novel FRT VSTs using other cohorts is currently underway.

As the etiology of rapid microbiome shifts between health and BV disease states remains unclear, we sought to identify specific bacterial taxa associated with recovery and transition to BV that could be involved in these dynamic changes. We observed a significant increase in relative abundance of the gram-positive facultatively anaerobic cocci *Aerococcus* and *Gemella* during transition to and from clinical BV. However, previous literature has recognized these bacterial

genera as being less frequently detected during clinical BV [61]. This could indicate these taxa as important facilitators of disease-associated microbiome composition and as clinical biomarkers of BV. Similar to what is seen in the gut, these bacteria may be assuming the role of "primary species" commonly seen after an environmental disturbance, whereby fast-growing facultative anaerobes transiently bloom and are essential for the presence of other taxa along with ecological diversity and structure [62]. Interestingly, *Enterococcus virus FL2* A and *Enterococcus phage phiFL2A* both were found to be significantly associated with the transitory bacteria *Aerococcus*. Because of the sequence similarity between *Aerococcus* and *Enterococcus* bacteria, along with the identification of these bacteriophage prior to the introduction of more sensitive next generation sequencing technologies, it is probable that *Enterococcus virus FL2* A and *Enterococcus phage phiFL2A* have a misidentified host [63]. These bacteriophages may actually target *Aerococcus* species, facilitating their significant role in recovery and transition to BV. A larger longitudinal patient population would be needed to further distinguish between specific taxa associated solely with BV incidence vs BV recovery and the roles they play in these FRT microbiome transitions.

Our analysis also is the first to identify discriminant bacteriophage taxa by BV status and assess transkingdom associations in the FRT. Correlation analysis between bacterial taxa and assigned bacteriophage species identified bacteriophages that positively correlated with BV-associated bacteria and inversely correlated with *Lactobacillus*, suggesting that transkingdom interactions between bacteriophages and bacterial species could be the driver of BV-associated bacterial community alterations. *Bacillus virus Camphawk* and *Bacillus virus Pony*, previously demonstrated to be lytic to *Bacillus* members [64, 65], were associated with both BV and BV-associated bacteria. In the gut, *Bacillus* strains have been shown to antagonize enteropathogenic

bacteria, while concurrently promoting the growth of *Lactobacillus* [66]. If *Bacillus* species act in a similar manner in the FRT, the lytic nature of the *Bacillus* bacteriophages *Bacillus virus Camphawk* and *Bacillus virus Pony* could at least partially explain the shift in vaginal microbiota away from *Lactobacillus* species and toward more diverse bacterial species, including the facultative anaerobes seen in BV. *E. coli* and *Burkholderia* bacteriophages *Escherichia virus FV3* and *unclassified E125* virus, respectively, were also associated with BV. While less predominant than other bacterial species, *E. coli* and *Burkholderia* are also implicated in BV-associated bacterial communities [1, 2, 58]. *Escherichia virus FV3* and *unclassified E125* virus may act to regulate bacterial abundance and community composition in BV via a predator-prey relationship to allow for growth of primary BV-associated bacterial members such as *Gardnerella* and *Prevotella* [22, 67]. BV risk factors such as new or multiple sexual partners provide a plausible mechanism for introduction of novel bacteriophage that could target and deplete commensal FRT bacteria [68, 69].

Interestingly, we did not find any significant association between sequences assigned to bacteriophage know to infect the more common BV-associated bacteria, including *Gardnerella* or *Prevotella*, by either BV-discriminant taxa or transkingdom analysis. The absence of bacteriophage that infect hallmark BV bacteria such as *Gardnerella vaginalis* in our analysis may be attributable to the high proportion of *Gardnerella* species that contain CRISPR/Cas-9 bacteriophage defense loci, making them more resistant to bacteriophage infection and establishing a uneven bacteriophage burden between bacteria in BV [70]. These data suggest a very active, dynamic environment of bacteriophage warfare against numerous bacterial hosts to regulate the bacterial populations. Further *in vitro* studies will be necessary to determine host range

and bacteriophage lifestyle. A thorough examination of the FRT microbiome in the context of BV contributes to a deeper understanding of BV pathogenesis.

While we focused on BV as a common FRT-localized disease, we also were interested in investigating the effect of other disease states on the FRT microbiome. In addition to BV, we studied the impact of HIV infection on the FRT but observed no difference in FRT bacterial or bacteriophage diversity based on HIV status. One possible explanation is that HIV, as a systemic disease, does not have a major impact on the localized FRT mucosal environment. Prior literature shows that acute HIV infection leads to distinct changes in local inflammatory marker profiles, including elevation of pro-inflammatory cytokines IL-1α, IL-1β, and IL-6 [4, 54] however, it remains unclear if this initial inflammatory response persists long-term. Alternatively, we may not have detected differences due to the well-controlled nature of HIV infection in this cohort. We previously showed that immunocompromise was a major factor affecting enteric microbiome diversity in HIV infection [44], a finding replicated at other mucosal sites [71], suggesting increased likelihood of observing microbiome alterations with immunodeficiency. Since the mean CD4+ T cell count for this cohort was 478 cells/μL, and no subjects were known to be immunosuppressed, this could explain the dearth of associations of the FRT microbiome with HIV seen in our study.

Limitations to this study include the cross-sectional nature of the virome analysis, preventing speculation on the longitudinal impact of bacteriophage changes, and low RNA integrity of the samples, blocking assessment of the FRT RNA virome. The initial limited patient consent also

precluded significant *in vitro* validation of bacteriophage-bacterial pairs. Finally, there may be inaccuracies or biases in sequence assignments.

Conclusions

In this retrospective longitudinal study, we performed a novel in-depth investigation of the FRT virome and bacteriome using a cohort of young, sexually active, South African women. We discovered significant alterations in FRT bacterial and bacteriophage diversity and community structure associated with BV. Transkingdom analysis revealed associations of specific bacteriophages with bacteria protective of and associated with BV. This study is the first to describe VST structure within the FRT and its associations with bacterial diversity and composition. The nature of the FRT in health and disease is both complex and dynamic and our findings provide insight into putative interactions between bacteriophage and bacteria that may contribute to development and maintenance of FRT dysbiosis. Further studies are needed to investigate direct mechanisms employed by bacteriophages to promote dysbiosis.

List of abbreviations

BV: Bacterial Vaginosis

CG: Bacterial Community Group

HIV: Human Immunodeficiency Virus

FRT: Female Reproductive Tract

NGS: Next Generation Sequencing

STI: Sexually Transmitted Infection

VST: Viral State Type

Conflicts of Interest

The authors declare that they have no conflicts of interests.

Ethics approval and consent to participate

This study was reviewed by the Research Subjects Review Board of the University of Rochester and granted human exemption status.

Informed Consent

Informed consent was obtained from all subjects involved in the study

Authors' contributions

Conceptualization and Methodology: CLM and FSM; Formal Analysis – MS and AG; Investigation: FSM, BB, AW; Resources and Sample Acquisition: TM, ALW, LGB, DHA; Data curation: FSM, BB, AW; Writing – original draft: FSM; Writing – reviewing and editing: all authors; Visualization: FSM, MS, AG; Supervision: CLM; Funding acquisition: CLM.

Funding

This research was supported in part by a grant from the University of Rochester Center for AIDS Research (CFAR), an NIH-funded program (P30AI078498). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National

Institutes of Health. FM was a recipient of a National Institutes of Health HIV T32 Training Grant AI1049815.

Acknowledgments

We thank the study subjects for their participation as well as study team nurses and personnel. We also would like to thank James Java, Ph.D. for assistance with the VirusSeeker virome pipeline and Cassandra Newkirk for experimental support. We thank the Center for Integrated Research Computing (CIRC) at the University of Rochester for providing computational resources and technical support. Microbiome and virome sequencing in this study was completed by the University of Rochester Genomics Research Center (GRC).

REFERENCES

[1] H. W. Virgin, "The virome in mammalian physiology and disease," *Cell*, vol. 157, no. 1, pp. 142-50, Mar 27 2014, doi: 10.1016/j.cell.2014.02.032.

[2] C. Ceccarani *et al.*, "Diversity of vaginal microbiome and metabolome during genital infections," (in eng), *Sci Rep*, vol. 9, no. 1, p. 14095, Oct 2019, doi: 10.1038/s41598-019-50410-x.

[3] M. I. Petrova, M. van den Broek, J. Balzarini, J. Vanderleyden, and S. Lebeer, "Vaginal microbiota and its role in HIV transmission and infection," *FEMS microbiology reviews*, vol. 37, no. 5, pp. 762-92, Sep 2013, doi: 10.1111/1574-6976.12029.

[4] C. Gosmann *et al.*, "Lactobacillus-Deficient Cervicovaginal Bacterial Communities Are Associated with Increased HIV Acquisition in Young South African Women," *Immunity*, vol. 46, no. 1, pp. 29-37, Jan 17 2017, doi: 10.1016/j.immuni.2016.12.013.

[5] S. Greenbaum, G. Greenbaum, J. Moran-Gilad, and A. Y. Weintraub, "Ecological dynamics of the vaginal microbiome in relation to health and disease," (in eng), *Am J Obstet Gynecol*, vol. 220, no. 4, pp. 324-335, 04 2019, doi: 10.1016/j.ajog.2018.11.1089.

[6] J. P. Brooks *et al.*, "Changes in vaginal community state types reflect major shifts in the microbiome," (in eng), *Microb Ecol Health Dis*, vol. 28, no. 1, p. 1303265, 2017, doi: 10.1080/16512235.2017.1303265.

[7] S. B. Smith and J. Ravel, "The vaginal microbiota, host defence and reproductive physiology," *J Physiol*, vol. 595, no. 2, pp. 451-463, Jan 15 2017, doi: 10.1113/JP271694.

[8] J. Ravel *et al.*, "Vaginal microbiome of reproductive-age women," (in eng), *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108 Suppl 1, pp. 4680-7, Mar 15 2011, doi: 1002611107 [pii]
10.1073/pnas.1002611107.

[9] C. Nardis, L. Mosca, and P. Mastromarino, "Vaginal microbiota and viral sexually transmitted diseases," (in eng), *Ann Ig*, vol. 25, no. 5, pp. 443-56, 2013 Sep-Oct 2013, doi: doi:10.7416/ai.2013.1946.

[10] J. J. Schellenberg and F. A. Plummer, "The Microbiological Context of HIV Resistance: Vaginal Microbiota and Mucosal Inflammation at the Viral Point of Entry," (in eng), *Int J Inflam*, vol. 2012, p. 131243, 2012, doi: 10.1155/2012/131243.

[11] D. W. White *et al.*, "Latent herpesvirus infection arms NK cells," (in eng), *Blood*, Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't vol. 115, no. 22, pp. 4377-83, Jun 3 2010, doi: 10.1182/blood-2009-09-245464.

[12] M. R. Joesoef and G. Schmid, "Bacterial vaginosis," (in eng), *Clin Evid*, no. 13, pp. 1968-78, Jun 2005.

[13] J. E. Bilardi *et al.*, "The burden of bacterial vaginosis: women's experience of the physical, emotional, sexual and social impact of living with recurrent bacterial vaginosis," (in eng), *PloS one*, vol. 8, no. 9, p. e74378, 2013, doi: 10.1371/journal.pone.0074378.

[14] S. Srinivasan *et al.*, "Temporal variability of human vaginal bacteria and relationship with bacterial vaginosis," *PloS one*, vol. 5, no. 4, p. e10197, 2010, doi: 10.1371/journal.pone.0010197.

[15] R. M. Brotman, J. Ravel, R. A. Cone, and J. M. Zenilman, "Rapid fluctuation of the vaginal microbiota measured by Gram stain analysis," (in eng), *Sex Transm Infect*, vol. 86, no. 4, pp. 297-302, Aug 2010, doi: 10.1136/sti.2009.040592.

[16] B. M. Ngugi *et al.*, "Effects of bacterial vaginosis-associated bacteria and sexual intercourse on vaginal colonization with the probiotic Lactobacillus crispatus CTV-05," (in eng), *Sex Transm Dis*, vol. 38, no. 11, pp. 1020-7, Nov 2011, doi: 10.1097/OLQ.0b013e3182267ac4.

[17] K. Cadwell, "The virome in host health and disease," *Immunity*, vol. 42, no. 5, pp. 805-13, May 19 2015, doi: 10.1016/j.immuni.2015.05.003.

[18] K. M. Wylie, T. N. Wylie, A. G. Cahill, G. A. Macones, M. G. Tuuli, and M. J. Stout, "The vaginal eukaryotic DNA virome and preterm birth," (in eng), *Am J Obstet Gynecol,* vol. 219, no. 2, pp. 189.e1-189.e12, 08 2018, doi: 10.1016/j.ajog.2018.04.048.

[19] K. M. Wylie, K. A. Mihindukulasuriya, Y. Zhou, E. Sodergren, G. A. Storch, and G. M. Weinstock, "Metagenomic analysis of double-stranded DNA viruses in healthy adults," *BMC biology,* vol. 12, p. 71, Sep 10 2014, doi: 10.1186/s12915-014-0071-7.

[20] K. M. Parmar, S. L. Gaikwad, P. K. Dhakephalkar, R. Kothari, and R. P. Singh, "Intriguing Interaction of Bacteriophage-Host Association: An Understanding in the Era of Omics," *Frontiers in microbiology,* vol. 8, p. 559, 2017, doi: 10.3389/fmicb.2017.00559.

[21] D. Lawrence, M. T. Baldridge, and S. A. Handley, "Phages and Human Health: More Than Idle Hitchhikers," (in eng), *Viruses,* vol. 11, no. 7, Jun 2019, doi: 10.3390/v11070587.

[22] A. Fillol-Salom *et al.*, "Bacteriophages benefit from generalized transduction," (in eng), *PLoS pathogens,* vol. 15, no. 7, p. e1007888, 07 2019, doi: 10.1371/journal.ppat.1007888.

[23] R. R. Jakobsen *et al.*, "Characterization of the Vaginal DNA Virome in Health and Dysbiosis," (in eng), *Viruses,* vol. 12, no. 10, 10 2020, doi: 10.3390/v12101143.

[24] H. T. Zhang, H. Wang, H. S. Wu, J. Zeng, and Y. Yang, "Comparison of viromes in vaginal secretion from pregnant women with and without vaginitis," (in eng), *Virol J,* vol. 18, no. 1, p. 11, 01 2021, doi: 10.1186/s12985-020-01482-z.

[25] L. H. Damelin, M. Paximadis, D. Mavri-Damelin, M. Birkhead, D. A. Lewis, and C. T. Tiemessen, "Identification of predominant culturable vaginal Lactobacillus species and

associated bacteriophages from women with and without vaginal discharge syndrome in South Africa," *Journal of medical microbiology,* vol. 60, no. Pt 2, pp. 180-3, Feb 2011, doi: 10.1099/jmm.0.024463-0.

[26] A. O. Kilic, S. I. Pavlova, S. Alpay, S. S. Kilic, and L. Tao, "Comparative study of vaginal Lactobacillus phages isolated from women in the United States and Turkey: prevalence, morphology, host range, and DNA homology," *Clinical and diagnostic laboratory immunology,* vol. 8, no. 1, pp. 31-9, Jan 2001, doi: 10.1128/CDLI.8.1.31-39.2001.

[27] S. I. Pavlova, A. O. Kilic, S. M. Mou, and L. Tao, "Phage infection in vaginal lactobacilli: an in vitro study," *Infectious diseases in obstetrics and gynecology,* vol. 5, no. 1, pp. 36-44, 1997, doi: 10.1155/S1064744997000094.

[28] A. Wahida, K. Ritter, and H. P. Horz, "The Janus-Face of Bacteriophages across Human Body Habitats," (in eng), *PLoS pathogens,* vol. 12, no. 6, p. e1005634, 06 2016, doi: 10.1371/journal.ppat.1005634.

[29] J. M. Macklaim, A. D. Fernandes, J. M. Di Bella, J. A. Hammond, G. Reid, and G. B. Gloor, "Comparative meta-RNA-seq of the vaginal microbiota and differential expression by Lactobacillus iners in health and dysbiosis," *Microbiome,* vol. 1, no. 1, p. 12, 2013, doi: 10.1186/2049-2618-1-12.

[30] M. Rho, Y. W. Wu, H. Tang, T. G. Doak, and Y. Ye, "Diverse CRISPRs evolving in human microbiomes," *PLoS genetics,* vol. 8, no. 6, p. e1002441, 2012, doi: 10.1371/journal.pgen.1002441.

[31] Q. Zhang, M. Rho, H. Tang, T. G. Doak, and Y. Ye, "CRISPR-Cas systems target a diverse collection of invasive mobile genetic elements in human microbiomes," *Genome biology,* vol. 14, no. 4, p. R40, 2013, doi: 10.1186/gb-2013-14-4-r40.

[32] H. Zhao *et al.*, "Mice deficient in Epg5 exhibit selective neuronal vulnerability to degeneration," (in eng), *The Journal of cell biology,* vol. 200, no. 6, pp. 731-41, Mar 18 2013, doi: 10.1083/jcb.201211014
jcb.201211014 [pii].

[33] L. Dylla, B. Abar, A. L. Williamson, T. L. Meiring, L. G. Bekker, and D. H. Adler, "Human papillomavirus clustering patterns among HIV-infected and HIV-uninfected adolescent females in South Africa," (in eng), *J AIDS HIV Res,* vol. 9, no. 10, pp. 202-206, Oct 2017, doi: 10.5897/JAHR2017.0445.

[34] D. W. Fadrosh *et al.*, "An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform," (in eng), *Microbiome,* vol. 2, no. 1, p. 6, Feb 2014, doi: 10.1186/2049-2618-2-6.

[35] J. G. Caporaso *et al.*, "QIIME allows analysis of high-throughput community sequencing data," (in eng), *Nat Methods,* vol. 7, no. 5, pp. 335-6, May 2010, doi: 10.1038/nmeth.f.303.

[36] E. Bolyen *et al.*, "Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2," (in eng), *Nat Biotechnol,* vol. 37, no. 8, pp. 852-857, 08 2019, doi: 10.1038/s41587-019-0209-9.

[37] B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. Johnson, and S. P. Holmes, "DADA2: High-resolution sample inference from Illumina amplicon data," (in eng), *Nature methods,* vol. 13, no. 7, pp. 581-3, 07 2016, doi: 10.1038/nmeth.3869.

[38] T. Z. DeSantis *et al.*, "Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB," (in eng), *Appl Environ Microbiol,* vol. 72, no. 7, pp. 5069-72, Jul 2006, doi: 10.1128/AEM.03006-05.

[39] K. Katoh and D. M. Standley, "MAFFT multiple sequence alignment software version 7: improvements in performance and usability," (in eng), *Mol Biol Evol,* vol. 30, no. 4, pp. 772-80, Apr 2013, doi: 10.1093/molbev/mst010.

[40] M. N. Price, P. S. Dehal, and A. P. Arkin, "FastTree 2--approximately maximum-likelihood trees for large alignments," (in eng), *PloS one,* vol. 5, no. 3, p. e9490, Mar 2010, doi: 10.1371/journal.pone.0009490.

[41] C. Lozupone, M. E. Lladser, D. Knights, J. Stombaugh, and R. Knight, "UniFrac: an effective distance metric for microbial community comparison," (in eng), *The ISME journal,* vol. 5, no. 2, pp. 169-72, Feb 2011, doi: 10.1038/ismej.2010.133.

[42] D. Adler *et al.*, "High risk human papillomavirus persistence among HIV-infected young women in South Africa," *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases,* vol. 33, pp. 219-21, Apr 2015, doi: 10.1016/j.ijid.2015.02.009.

[43] M. Zozaya-Hinchliffe, R. Lillis, D. H. Martin, and M. J. Ferris, "Quantitative PCR assessments of bacterial species in women with and without bacterial vaginosis," (in eng), *J Clin Microbiol,* vol. 48, no. 5, pp. 1812-9, May 2010, doi: 10.1128/JCM.00851-09.

[44] C. L. Monaco *et al.*, "Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome," *Cell host & microbe,* vol. 19, no. 3, pp. 311-22, Mar 09 2016, doi: 10.1016/j.chom.2016.02.011.

[45] G. Zhao *et al.*, "VirusSeeker, a computational pipeline for virus discovery and virome composition analysis," *Virology*, vol. 503, pp. 21-30, Jan 18 2017, doi: 10.1016/j.virol.2017.01.005.

[46] J. C. Pinheiro and D. M. Bates, *Mixed-Effects Models in S and S-PLUS*. New York: Springer-Verlag, 2000.

[47] K. Aagaard, J. Ma, K. M. Antony, R. Ganu, J. Petrosino, and J. Versalovic, "The placenta harbors a unique microbiome," (in eng), *Science translational medicine*, vol. 6, no. 237, p. 237ra65, May 2014, doi: 10.1126/scitranslmed.3008599.

[48] M. J. Anderson, "A new method for non-parametric multivariate analysis of variance," *Austral Ecology*, vol. 26, pp. 32-46, 2001.

[49] K. L and R. PJ, *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, 1990.

[50] M. Kendall, "A New Measure of Rank Correlation," *Biometrika*, vol. 30, no. 1-2, pp. 81-93, 1938.

[51] R. C. Team. "A language and environment for statistical computing." R Foundation for Statistical Computing. (accessed.

[52] Y. Hochberg and Y. Benjamini, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society, Series B*, vol. 57, no. 1, pp. 289-300, 1995.

[53] M. G. Torcia, "Interplay among Vaginal Microbiome, Immune Response and Sexually Transmitted Viral Infections," *International journal of molecular sciences,* vol. 20, no. 2, Jan 11 2019, doi: 10.3390/ijms20020266.

[54] M. N. Anahtar *et al.*, "Cervicovaginal bacteria are a major modulator of host inflammatory responses in the female genital tract," *Immunity,* vol. 42, no. 5, pp. 965-76, May 19 2015, doi: 10.1016/j.immuni.2015.04.019.

[55] M. T. France *et al.*, "VALENCIA: a nearest centroid classification method for vaginal microbial communities based on composition," (in eng), *Microbiome,* vol. 8, no. 1, p. 166, 11 2020, doi: 10.1186/s40168-020-00934-6.

[56] M. P. Ritz, A. L. Perl, J. M. Colquhoun, K. R. Chamakura, and G. F. Kuty Everett, "Complete Genome of Bacillus subtilis Myophage CampHawk," (in eng), *Genome announcements,* vol. 1, no. 6, Dec 2013, doi: 10.1128/genomeA.00984-13.

[57] J. H. Grose, G. L. Jensen, S. H. Burnett, and D. P. Breakwell, "Correction: genomic comparison of 93 Bacillus phages reveals 12 clusters, 14 singletons and remarkable diversity," (in eng), *BMC genomics,* vol. 15, p. 1184, Dec 2014, doi: 10.1186/1471-2164-15-1184.

[58] Q. Xia *et al.*, "Identification of vaginal bacteria diversity and it's association with clinically diagnosed bacterial vaginosis by denaturing gradient gel electrophoresis and correspondence analysis," (in eng), *Infect Genet Evol,* vol. 44, pp. 479-486, 10 2016, doi: 10.1016/j.meegid.2016.08.001.

[59] M. I. Petrova, E. Lievens, S. Malik, N. Imholz, and S. Lebeer, "Lactobacillus species as biomarkers and agents that can promote various aspects of vaginal health," (in eng), *Front Physiol,* vol. 6, p. 81, 2015, doi: 10.3389/fphys.2015.00081.

[60] A. Edlund, T. M. Santiago-Rodriguez, T. K. Boehm, and D. T. Pride, "Bacteriophage and their potential roles in the human oral cavity," *J Oral Microbiol*, vol. 7, p. 27423, 2015, doi: 10.3402/jom.v7.27423.

[61] J. S. Coleman and C. A. Gaydos, "Molecular Diagnosis of Bacterial Vaginosis: an Update," (in eng), *J Clin Microbiol*, vol. 56, no. 9, 09 2018, doi: 10.1128/JCM.00342-18.

[62] S. M. Gibbons, "Keystone taxa indispensable for microbiome recovery," (in eng), *Nat Microbiol*, vol. 5, no. 9, pp. 1067-1068, 09 2020, doi: 10.1038/s41564-020-0783-0.

[63] A. Yasmin *et al.*, "Comparative genomics and transduction potential of Enterococcus faecalis temperate bacteriophages," (in eng), *J Bacteriol*, vol. 192, no. 4, pp. 1122-30, Feb 2010, doi: 10.1128/JB.01293-09.

[64] I. M. Willms, M. Hoppert, and R. Hertel, "Characterization of Bacillus Subtilis Viruses vB_BsuM-Goe2 and vB_BsuM-Goe3," (in eng), *Viruses*, vol. 9, no. 6, 06 2017, doi: 10.3390/v9060146.

[65] B. E. Khatemi, C. C. Chung On, K. R. Chamakura, and G. F. Kuty Everett, "Complete Genome of Bacillus megaterium Podophage Pony," (in eng), *Genome announcements*, vol. 1, no. 6, Dec 2013, doi: 10.1128/genomeA.00860-13.

[66] S. Mingmongkolchai and W. Panbangred, "Bacillus probiotics: an alternative to antibiotics for livestock production," (in eng), *J Appl Microbiol*, vol. 124, no. 6, pp. 1334-1346, Jun 2018, doi: 10.1111/jam.13690.

[67] M. A. Brockhurst, A. D. Morgan, A. Fenton, and A. Buckling, "Experimental coevolution with bacteria and phage. The Pseudomonas fluorescens--Phi2 model system," (in eng), *Infect Genet Evol*, vol. 7, no. 4, pp. 547-52, Jul 2007, doi: 10.1016/j.meegid.2007.01.005.

[68] Y. Turovskiy, K. Sutyak Noll, and M. L. Chikindas, "The aetiology of bacterial vaginosis," (in eng), *J Appl Microbiol,* vol. 110, no. 5, pp. 1105-28, May 2011, doi: 10.1111/j.1365-2672.2011.04977.x.

[69] A. A. Vostrov, O. A. Vostrukhina, A. N. Svarchevsky, and V. N. Rybchin, "Proteins responsible for lysogenic conversion caused by coliphages N15 and phi80 are highly homologous," (in eng), *J Bacteriol,* vol. 178, no. 5, pp. 1484-6, Mar 1996, doi: 10.1128/jb.178.5.1484-1486.1996.

[70] M. Pleckaityte, M. Zilnyte, and A. Zvirbliene, "Insights into the CRISPR/Cas system of Gardnerella vaginalis," (in eng), *BMC Microbiol,* vol. 12, p. 301, Dec 2012, doi: 10.1186/1471-2180-12-301.

[71] R. M. Presti *et al.*, "Alterations in the oral microbiome in HIV-infected participants after antiretroviral therapy administration are influenced by immune status," (in eng), *AIDS,* vol. 32, no. 10, pp. 1279-1287, 06 2018, doi: 10.1097/QAD.0000000000001811.