METAGENOMIC CHARACTERIZATION OF THE VAGINAL MICROBIOME

by

Keshini Devakandan

B.Sc., Western University, 2013

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Reproductive and Developmental Sciences)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

December 2016

Abstract

Background: The vaginal microbiome is a dynamic environment colonized by a wide array of microorganisms. Although bacterial vaginosis (BV) is characterized by a disruption in the normal bacterial microbiome of the vagina, the factors contributing to recurrent BV remain unknown. In addition, very little is known about the role of viruses in the vaginal microbiome and associated dysbioses. Objectives: 1) characterize the vaginal bacteriome of women with recurrent BV using *cpn*60 sequencing, compare bacterial profiles to healthy-asymptomatic cohort, and correlate profiles to descriptive characteristics; and 2) characterize the vaginal virome of healthy-asymptomatic, HIV-positive women and women with recurrent BV, and correlate profiles to descriptive characteristics.

Methods: Twenty-six women were recruited into the recurrent BV bacteriome study. Vaginal swabs were obtained for *cpn*60 sequencing and Gram stain Nugent scoring. Additionally, samples from 54 women were analyzed in the virome study: 21 healthy-asymptomatic, 25 HIV-positive and eight recurrent BV. The vaginal swabs were processed to enrich for viruses and then subjected to metagenomics shotgun sequencing. Demographic, behavioural and clinical information was collected for all participants, in both bacteriome and virome studies.

Results: Bacteriome analyses detected 122 *cpn*60 operational taxonomic units (OTUs). Bacterial profiles clustered into six community state types (CSTs). Trends suggested a relationship between BV-associated CSTs and number of sexual partners (past year), oral sex, use of (hormonal) contraception, abnormal discharge (past 48 hours), lifetime history of trichomoniasis, and number of BV episodes (past two months and year). Virome analyses detected a total of 477 species. Viral profiles clustered into seven groups. Viral patterns were identified within bacteriome CSTs, Nugent scores, viral loads, between *Lactobacillus*-dominant,

Lactobacillus iners-dominant, and heterogeneous profiles, and were associated with a number of descriptive characteristics.

Conclusions: The vaginal microbiome is highly diverse and potentially associated with many clinical factors. Our ability to use the microbiome data to subdivide women into clusters, and detect trends between clusters and characteristics will expand our knowledge on the vaginal microbiome as a whole.

Preface

This research project received funding and support from the Canadian Institutes of Health Research and Genome British Columbia. Ethical approval for this project was obtained by the University of British Columbia – Children's & Women's Health Centre of BC Research Ethics Board (UBC C&W REB) with Certificate Number H11-01912. Preliminary results from Sections 3.1 and 4.1 were presented as posters at the 12th Annual Genomics Forum - Genomics: Driving BC's Bioeconomy and the 41st Annual Infectious Diseases Society for Obstetrics and Gynecology Meeting with the following abstract publications:

Devakandan, K, Wagner, EC, Chaban, B, Albert, AYK, van Schalkwyk, J, Hill, JE, Money, D, Vogue Team. Metagenomic Characterization of the Vaginal Microbiome of Women with Recurrent Bacterial Vaginosis. Genome BC 2014. May 2014.

Devakandan, K, Wagner, EC, Chaban, B, Albert, AYK, van Schalkwyk, J, Links, MG, Hill, JE, Money, D, Vogue Study Team. Metagenomic Characterization of the Vaginal Microbiome of Women with Recurrent Vulvovaginitis. IDSOG 2014. August 2014.

This research project could not have been made possible without the contribution and support of a number of select individuals. The collection of vaginal swabs for the recurrent bacterial vaginosis bacteriome analyses were conducted by Dr. Deborah Money, Dr. Julie van Schalkwyk, Dr. Wally Unger and Dr. Janet Lyons from the Reproductive Infectious Diseases clinic at BC Women's Hospital (Vancouver, BC), Dr. Julie van Schalkwyk's Obstetrics and Gynecology (OBGYN) consulting office (Vancouver, BC), and another speciality OBGYN practice in the Vancouver community. In addition to myself, the collection of demographic, behavioural and clinical information was carried out by Melissa Watt, Sandy Lee, Shivinder

Dhillon, Lilija Berngards and Kristi Engle Folchert from the Women's Health Research Institute. DNA extraction, PCR amplification and sequencing outlined in Section 2.1 were performed by Dr. Bonnie Chaban, Niradha Withana Gamage, and Aline Costa de Freitas from the University of Saskatchewan. Clustering analyses of the bacteriome profiles as outlined in Section 2.1 were conducted by Dr. Bonnie Chaban and Aline Costa de Freitas from the University of Saskatchewan.

Vaginal swabs from the healthy-asymptomatic cohort were self-collected by study participants, and collection of vaginal swabs from the HIV-positive population was conducted by Bal Dhesi and Laura Vicol from the Oak Tree Clinic at BC Women's Hospital (Vancouver, BC). Collection of demographic, clinical and behavioural information from healthy-asymptomatic and HIV-positive women were collected by Shivinder Dhillon, Cara Bergen, Kristi Engle Folchert, Lilija Berngards, Gisela Gosse and Daljeet Mahal. As indicated in Section 2.2, the design and implementation of the virome optimization protocol on mock samples, and the centrifugation and filtration spiking experiments were carried out by myself, along with the help of Dr. Waren Baticados and Dr. Patrick Tang from the BC Centre for Disease Control (BCCDC) (Vancouver, BC). Subsequent processing of the vaginal viral swabs was aided by Dr. Miguel Uyaguari Diaz from the BCCDC. Library size selection was conducted by Coastal Genomics (Burnaby, BC Canada), and sequencing was conducted by the Canada's Michael Smith Genome Sciences Centre (Vancouver, BC Canada). Metagenomic analyses as mentioned in Section 2.2 were aided by Dean Attali, Dr. Miguel Uyaguari Diaz and Dr. William Hsiao.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BV: Bacterial vaginosis

OTU: operational taxonomic unit

CST: community state type

UBC C&W REB: University of British Columbia – Children's & Women's Health Centre of BC Research Ethics Board

OBGYN: Obstetrics and Gynecology

BCCDC: BC Centre for Disease Control

VOGUE: Vaginal Microbiome Project/Vaginal Microbiome Group Initiative

NIH: National Institutes of Health

CIHR: Canadian Institutes of Health Research

HIV: human immunodeficiency virus

HPV: human papillomavirus

STI: sexually transmitted infection

H_2O_2: hydrogen peroxide

DNA: deoxyribonucleic acid

pH: power of hydrogen

RNA: ribonucleic acid

16S rRNA: 16S ribosomal RNA

PPROM: preterm prelabor rupture of membranes

KOH: potassium hydroxide

IUD: intrauterine device

HR: hazard ratio

BVAB: BV-associated bacteria

PCR: polymerase chain reaction

RDP: Ribosome Database Project

*cpn*60: chaperonin-60

CMV: *Cytomegalovirus*

HSV: herpes simplex virus

HSIL: high grade squamous intraepithelial lesion

LSIL: low grade squamous intraepithelial lesion

SEN-V: SEN-virus

JC virus: *John Cunningham virus*

WHRI: Women's Health Research Institute

BMI: body mass index

THC: tetrahydrocannabinol

REDCap: Research Electronic Data Capture

CFRI: Child and Family Research Institute

PBS: phosphate buffered saline

Ct: cycle threshold

ddH_2O: double-distilled water

qPCR: quantitative or real-time PCR

qRT-PCR: quantitative reverse transcription PCR

cDNA: complementary DNA

NCBI: National Center for Biotechnology Information

SD: standard deviation

UTI: urinary tract infection

HCV: Hepatitis C virus

HBV: Hepatitis B virus

sAb: surface antibody

sAg: surface antigen

cAb: core antibody

HHV: human herpesvirus

WSW: women who have sex with women

STD: sexually transmitted disease

IDU: injection drug users

BC: British Columbia

GBS: Group B Streptococcus

EBV: Epstein-Barr virus

rRNA: ribosomal RNA

Acknowledgements

First of all I want to thank my supervisor Dr. Deborah Money for taking me on as a student, for her mentorship, and cultivating my passion for women's health and microbiome research. I want to thank my committee members: Dr. Lori Brotto, Dr. Melanie Murray, and Dr. Patrick Tang for your guidance and constructive feedback throughout my Master's. I also want to extend my thanks to all those who made this project possible. Dr. Janet Hill, Dr. Bonnie Chaban, Niradha Withana Gamage, and Aline Costa De Freitas for your respective roles in the bacteriome analyses, and for being so helpful and answering all my questions.

I would like to thank Dr. Waren Baticados and Dr. Miguel Uyaguari Diaz for their assistance in the optimization of the virome protocol, and its implementation on our vaginal viral samples. Thank you for taking the time to train me, and always being available to discuss experiments and provide advice. I would also like the thank Dr. Miguel Uyaguari Diaz again, in addition to Dr. William Hsiao with their help with the bioinformatics. And a tremendous thank you to my partner, Dean Attali, for not only keeping me sane and motivated, but providing an immense amount of help with bioinformatic protocols, and automating manual processes.

I want to also thank the following individuals, who helped along the way, being there to discuss analyses, answer questions, provide advice, as well as providing moral support. Thank you to Dr. Arianne Albert, Kristi Engle Folchert, Ozge Goktepe, Dr. Janet Hill, Nancy Lipsky, Dr. Aroha Miller, Zahra Pakzad, Emily Wagner, Fontayne Wong, and my family.

This project would not have been possible without each and every one of the individuals mentioned here, thank you.

1 Introduction

1.1 Human Microbiome

The human microbiome is an ecological community of microorganisms, with which we share a symbiotic relationship; whether it is mutualistic, commensal, or parasitic [1,2]. Our microbiome consists of trillions of microbes, which have previously been suggested to outnumber our host cells 10:1 [3]. However, more recent work places this ratio closer to 1:1, but stresses that this should not affect the importance of host-microbiota interactions [4]. There is a cooperative interplay between the human host, and the microbes that reside in and on our body such that microbes play a role in host functions such as defense, metabolism and reproduction [5]. Due to the close symbiotic relationship between the host and the microbiome, it is not surprising that microbial dysbiosis may result in adverse consequences in the human host [1]. Gastrointestinal dysbiosis has been associated with obesity, irritable bowel disease, asthma, diabetes, colorectal cancer, as well as other conditions [1,2,6,7]. Similarly, vaginal dysbiosis has been associated with bacterial vaginosis (BV), preterm birth, and other conditions [2,7].

A multi-interdisciplinary project has been launched worldwide aiming to characterize the microbiome, and the factors which influence the distribution and evolution of its microorganisms, in order to understand the range of human genetic and physiological diversity [8]. The National Institutes of Health (NIH) has funded the Human Microbiome Project in the United States, and the Canadian Institutes of Health Research (CIHR) has funded a Canadian Microbiome Initiative. The International Human Microbiome Consortium coordinates these two initiatives, as well as others around the world [3]. To date, the human microbiome; predominantly the bacteriome; in the oral and nasal cavities, the gastrointestinal and genito-urinary tracts, and the skin have been studied [7,9]

1.2 Vaginal Microbiome

The vaginal microbiome is a dynamic environment colonized by a wide array of microorganisms [10,11]. It is influential in reproductive success, defense against pathogens, and quality of life. Activities such as the use of antimicrobials or vaginal hygiene products, or sexual behaviours may disrupt the vaginal microbiome and potentially lead to adverse health consequences [11]. These health consequences include but are not limited to, preterm birth, the development of pelvic inflammatory disease, and an increased risk of acquisition of the human immunodeficiency virus (HIV), human papillomavirus (HPV) and other sexually transmitted infections (STIs) [12–20]. Not only is the vaginal microbiome vital in determining the reproductive health of women, but has been suggested to establish and shape the microbiome of her children too, suggesting long-lasting effects on the immune system and health of her children [2,6,21].

1.3 Bacteriome

Healthy Vaginal Microbiota

Initially, culture-dependent studies have been used to establish what constitutes a "normal" vaginal flora. These methods detected a predominance of *Lactobacillus* species, which are thought to promote a healthy vaginal environment though their production of hydrogen peroxide (H_2O_2), lactic acid and bacteriocins [22–24]. These metabolites are thought to act cooperatively in order to prevent the overgrowth of other, less favourable bacterial genera [23].

Hydrogen Peroxide

H_2O_2 has been suggested be to an important defense mechanism in the promotion of a healthy vaginal microbiota [9,23,25]. It is thought to exert its protective bactericidal effects through oxidative stress, which cause breaks in the DNA of the cell. These bactericidal effects are also thought to be enhanced in the acidic environment created by *Lactobacillus'* metabolite, lactic

acid [23,25]. Hawes et al. showed that the frequency of symptomatic vaginosis was much lower in women who possessed a majority of H_2O_2-producing lactobacilli (3%), compared to women who possessed a majority of lactobacilli not producing H_2O_2 (25%), or other bacterial species (46%) [26]. Their group also noted that the most effective H_2O_2-producing strains were *Lactobacillus (L.) crispatus* and *L. jensenii* [26]. This is in line with Antonio et al. who found that 75% of women without BV had *L. crispatus* or *L. jensenii*; while only 12% of women with BV possessed these species [27]. Similarly, Arouthcheva et al. noted that 96% of *Lactobacillus* species found in healthy vaginal ecosystems produced H_2O_2 whereas only 6% of *Lactobacillus* species recovered from women with BV produced H_2O_2 [22]. This suggests that lactic acid is not the only compound needed to maintain a healthy vaginal environment.

However, a recent study showed that production of H_2O_2 was constitutive upon exposure of cells to oxygen [25,28,29]. Previous *in vitro* studies, indicating the protective effects of H_2O_2 did not account for the antioxidant, anaerobic conditions of the vaginal environment [28,29]. O'Hanlon et al. concluded that H_2O_2 could not be produced at inhibitory levels within the hypoxic environment of the vagina, the vaginal fluid has sufficient anti-oxidant capabilities to block the bactericidal effects of H_2O_2, and H_2O_2 was more toxic to vaginal lactobacilli than it was to the BV-associated bacteria they tested [28,29]. With the conflicting evidence, H_2O_2's role in the vaginal microbiota is controversial.

Lactic Acid

Unlike the uncertain role of H_2O_2, lactic acid is present at active levels in the vaginal environment and exert its protective effects [28,29]. Lactic acid is thought to promote a healthy vaginal environment by reducing the pH of the vaginal environment [20,30,31]. Vaginal epithelial cells produce a sugar known as glycogen, which lactobacilli ferment to produce lactic acids [11,25]. Lactic acid production maintains the vaginal pH at less than 4.5, creating an inhospitable

environment for most pathogenic bacteria [22]. Lactic acid is suggested to disrupt the cell membrane of gram-negative bacteria, preventing their overgrowth in the vaginal environment [28]. Lactic acid may be an important metabolite, providing women with a *Lactobacillus*-dominated vaginal microbiota, with significantly more lactic acid mediated protection against infections [29].

Bacteriocins

Additional *Lactobacillus* metabolites, which are suggested to promote a healthy vaginal environment are bacteriocins. These proteins are thought to have a diverse range of activity, from inhibiting the growth of closely related *Lactobacillus* species, to inhibiting the growth of a wide-range of other gram-positive and gram-negative bacteria such as *Gardnerella (G.) vaginalis* [22]. Several studies identified a link between hydrogen ion production and bacteriocins activity. They postulated that at a low pH, when hydrogen ion concentrations are high, bacteriocin are active, and this activity decreases as the pH increases [22]. Thus, there may be a link between lactic acid and bacteriocins, since lactic acid maintains a low vaginal pH.

Healthy-Asymptomatic Women

Historically, culture-dependent studies set out to determine what constituted a “normal” vaginal flora. These early methods detected a predominance of *Lactobacillus* species, which was initially believed to be a single organism *L. acidophilus*, but later *L. acidophilus* was deemed the characteristic member of a complex of related species. This complex included *L. acidophilus, L. amylolyticus, L. amylovorus, L. crispatus, L. gallinarium, L. gasseri, L. iners, L. jensenii,* and *L. johnsoni* [9,32–34]. As previously mentioned, *Lactobacillus* is thought to promote a healthy vaginal microbiota by creating an inhospitable environment for other bacterial genera, thus preventing their overgrowth in the vaginal environment.

However, despite the identification of important vaginal flora, culture-dependent methods have major limitations. A large number of organisms cannot be cultured in standard media, due to their fastidious or anaerobic nature [9,35]. Consequently, this selects for the identification of bacterial organisms with certain characteristics. As a result, culture-dependent methods do not provide a good representation of the vaginal microbiota, and overestimate the diversity of *Lactobacillus* species in the normal vaginal environment [9,36]. Due to limitations of culture-dependent techniques, culture-independent methods are needed to identify those bacteria deemed "unculturable" [37], and provide insight into what constitutes a healthy vaginal microbiota.

Culture-independent microbiome studies using the 16S ribosomal RNA (16S rRNA) gene target, also found *Lactobacillus* to be the dominant genus found in healthy-asymptomatic women, and associated with a healthy vaginal microbiome [13,38–40]. Several species of *Lactobacillus* have been found, where *L. iners* and *L. crispatus* are the most common in women, followed by *L. jensenii* and *L. gasseri* [9,13,38,39,41,42]. Past 16S rRNA gene target studies, have shown that not all healthy-asymptomatic women necessarily have a vaginal microbiome dominated by *Lactobacillus* species [13,38–40,43,44]. Burton et al. found that of their 19 participants, six women had a dominance of *G. vaginalis* [43]. Of those six participants, three women had other microorganisms detected that were not commonly found in the vaginal environment, *Arthrobacter sp.*, *Caulobacter sp.*, and *Butyrivibrio fibrisolvens* [43]. In addition, three women in this study were also found to have a dominance of *Lactobacillus* and *G. vaginalis* [43]. Hyman et al. also saw that not all participants had vaginal microbiomes dominated by *Lactobacillus* species [44]. Seven of 20 participants had very little or no *Lactobacillus* species present on their vaginal epithelium, and nine participants had a mixture of *Lactobacillus* species and other microbes (*Bifidobacterium*, *Gardnerella*, *Atopobium*, *Corynebacterium, Janthinobacterium*) [44]. When looking at relative microorganism abundance in each participant, Hyman et al. noted that *Lactobacillus* was dominant in half of participants and the other half had other bacteria such as

Bifidobacterium, *Streptococcus*, *Gardnerella*, *Pseudomonas*, *Prevotella*, *Eubacterium* [44]. Additional studies conducted by Zhou et al. showed that a proportion of healthy women lacked *Lactobacillus* dominance in their vaginal microbiome, and instead other lactic acid-producing bacteria like *Atopobium vaginae, Megasphaera,* and *Leptotrichia* species were detected and found to be normal constituents of the vaginal environment [13,38]. Ravel et al. also noted that although not all community clusters were dominated by *Lactobacillus* species, all community clusters contained bacteria genera known to produce lactic acid [39]. Additionally, Ling et al. observed that in addition to *Lactobacillus* species, the lactic acid bacteria *Alloiococcus* was also identified in the vaginal microbiome of healthy women [16]. This highlights an important notion that perhaps microorganism function rather than type, contributes to a healthy vaginal microbiota [13,16,38,39,41].

More recently, research was conducted by our team to examine the stability of the vaginal microbiota in healthy women throughout their menstrual cycle. Findings show that 23% of women experienced transitioning of their vaginal microbiota between different *Lactobacillus* species, and 8% transitioned from a heterogeneous cluster to a *Lactobacillus*-dominated cluster. The majority of women (69%) had little variation and slight fluctuations through their menstrual cycle. Overall, this indicated that there was little association between a woman's menstrual cycle and vaginal microbiota. Results also detected a dominance of *Lactobacillus* species: *L. crispatus, L. iners* and *L. jensenii*. Of interest, two clusters were dominated by *Bifidobacterium (B.) breve,* or a heterogeneous mixture of non-lactobacilli. The role of *B. breve* in the vaginal environment has yet to be elucidated, but in view of the fact that *B. breve* is a lactic acid producing bacteria, it is suggested to promote a healthy vaginal environment [45].

Research has also been conducted by our team to characterize the vaginal microbiome of healthy-asymptomatic women. Our research identified additional community state types (CST), in conjunction with the CSTs identified by Ravel et al. and Gajer et al. [39,40]. A growing consensus has been established with the collective research on the vaginal microbiota, that the vaginal environment exists in a limited number of community configurations; CSTs [39,40,46,47]. CSTs are defined by the dominant bacterial species present within the vaginal microbiota, and can further describe health and disease states of this environment [39,40,46,47]. The vaginal communities of our population of healthy-asymptomatic women could be clustered into six CSTs. Three CSTs were dominated by *Lactobacillus* species; CST I: *L. crispatus*, CST III: *L. iners*, CST V: *L. jensenii*; two were dominated by two *G. vaginalis* strains; CST IVC: G. vaginalis subgroup A, CST IVD: G. vaginalis subgroup C; and CST IVA was our mixed dominant cluster. CST IVA was comprised of a mixture of multiple bacteria; *Stapylococcus, Streptococcus, Prevotella, Alloscardovia, Gardnerella* and *Lactobacillus*; where no single dominant species was seen. Additionally this cluster could also be dominated by *G. vaginalis* subgroup B, *B. breve, B. dentium,* or *Atopobium vaginae*. Approximately 75% of women in our population fell into the three CSTs dominated by *Lactobacillus* species, where 50% had vaginal microbiotas dominated by *L. crispatus*. We also found that approximately 15% of women fell into the two clusters dominated by *G. vaginalis* species [47].

Bacterial Vaginosis

A shift in proportionality of the vaginal microbiota, from a *Lactobacillus* dominated environment to one with higher levels of anaerobic and gram-negative bacteria (i.e. *G. vaginalis*), is known as bacterial vaginosis (BV) [28,29,48]. Initially this genital condition was referred to as nonspecific vaginitis, but upon the detection and characterization of *G. vaginalis*, this genital condition was renamed *G. vaginalis* vaginitis. However it is now understood to be a

dysbiosis of the genital tract with high levels of multiple endogenous organisms and with higher diversity and a shift from *Lactobacillus* dominant – hence the term “bacterial vaginosis” [49].

BV is the most common cause of vaginal complaints for reproductive women worldwide [17,20,50]. In the United States alone, there are approximately 22 million annual cases of BV [50]. Almost one-third of women will experience BV in their lifetime, but this rate can be as high as 1 in 2 women depending on race, geography, and history of a female sexual partner [51–54]. BV can manifest as abnormal discharge, and malodor, however approximately 50% of women are asymptomatic [15,55]. BV is associated with an increased risk of acquisition and transmission of STIs, HIV and upper genital tract infections in non-pregnant women. In pregnant women, BV is associated with risk of post-abortal sepsis, early miscarriage, recurrent abortion, late miscarriage, preterm prelabor rupture of membranes (PPROM), spontaneous preterm labor, preterm birth, histological chorioamnionitis, and postpartum endometritis [9,55,56].

Existing diagnostic methods include Amsel’s criteria and the Nugent score. Amsel’s criteria are four markers used for clinical diagnosis; the presence of at least three indicates BV [57]. Amsel’s criteria includes a pH greater than 4.5, a homogenous milky discharge, a ‘fishy’ odor upon adding 10% KOH solution to the discharge (the “ whiff test”), and the presence of clue cells on microscopy [57]. Clue cells are epithelial cells of the vagina, whose cell border is stippled by bacteria, thus obscuring the border [57]. All four markers can be obtained through a vaginal exam, to identify discharge, for collection of swabs to assess pH, for the whiff test, and for microscopy to detect clue cells. The Nugent score is a standardized system used in microscopy to classify bacterial morphotypes present on Gram stained vaginal smear slides [58]. Samples are given a score between 0 and 10; higher scores indicate the presence of BV (Figure 1.1). A score of 0-3 is inconsistent with BV, a score of 4-6 indicates intermediate BV, and a score of 7-10 is consistent with BV [58]. The Nugent score is dependent on the presence of *Lactobacillus*, *G.*

vaginalis, and *Mobiluncus* species. An abundance of *Lactobacillus* species will decrease the score, whereas an abundance of *G. vaginalis* and/or *Mobiluncus* species will increase the score [58]. The Nugent Criteria is deemed the gold standard with regards to BV diagnosis.

Score[b]	Lactobacillus morphotypes	*Gardnerella* and *Bacteroides* spp. morphotypes	Curved gram-variable rods
0	4+	0	0
1	3+	1+	1+ or 2+
2	2+	2+	3+ or 4+
3	1+	3+	
4	0	4+	

[a] Morphotypes are scored as the average number seen per oil immersion field. Note that less weight is given to curved gram-variable rods. Total score = lactobacilli + *G. vaginalis* and *Bacteroides* spp. + curved rods.

[b] 0, No morphotypes present; 1, <1 morphotype present; 2, 1 to 4 morphotypes present; 3, 5 to 30 morphotypes present; 4, 30 or more morphotypes present.

Figure 1.1 Nugent scoring system (0-10) for Gram stained vaginal smears [58]

Currently, treatment for BV is aimed at eliminating *G. vaginalis* with either metronidazole or clindamycin [20,55]. Treatments target *G. vaginalis* because previous studies utilizing microscopy and microbial culture have linked BV to the presence of *G. vaginalis.* However, *G. vaginalis* has also been detected in the vaginal microbiome of healthy and asymptomatic women, and is no longer a definitive diagnostic marker [16,19,39,40,59–63]. Current treatments have an initial modest success rate of 80%, but with a recurrence rate up to 30% [20,55,60]. This recurrence rate increases the longer women are off of treatment [20,64,65]. Although *G. vaginalis* is not a definite diagnostic marker, it has still been found in the vaginal microbiota of women with BV, and may play an influential role in BV pathogenesis. Virulence factors associated with *G. vaginalis* have been suggested to be biofilm formation and sialidase production [63].

Biofilm Formation by *G. Vaginalis*

In addition to current treatment methods, the high frequency of recurrence may be a result of the formation of resistant biofilms providing effective survival niches [66,67]. In comparing biopsies from women with and without BV, a characteristic dense biofilm covering at least 50% of vaginal epithelial surface was seen in 90% of samples in women with BV, versus 10% of samples in women without BV [66]. Swidsinski et al. found that BV was associated with the development of an adherent polymicrobial biofilm abundant with *G. vaginalis* on the vaginal epithelium [66]. *G. vaginalis*; suggested to be a predominant and obligatory component of the biofilm; was stacked in confluent or patchy layers tightly attached to the vaginal epithelium surface [66]. *G. vaginalis*' role in biofilm formation is also noted by Patterson et al., where they suggest that *G. vaginalis* possess adaptive systems which allow for increased tolerance to H_2O_2 and lactic acid allowing it to form biofilms in this environment [68]. More recent research has demonstrated similar findings, where *G. vaginalis* is suggested to be an early colonizer; able to adhere strongly to the vaginal epithelium, paving the way for additional species with lower pathogenic potential to adhere, grow and become established in the vaginal environment [67,69,70].

Additionally Swidsinski et al. noted that *Atopbium (A.) vaginae* was found in 80% of *G. vaginalis* positive biofilms, and contributed to 40% of the whole biofilm [66]. Similarly, Hardy et al. found *A. vaginae* to be an important constituent of the vaginal epithelium biofilm, where 41% of biofilms containing *G. vaginalis* biofilms were also positive for *A. vaginae* [70]. No samples contained *A. vaginae* alone, and when *A. vaginae* was found with *G. vaginalis,* both bacterial species were present at higher concentrations [70]. This suggests that *G. vaginalis* is the initial colonizer creating a favourable environment for *A. vaginae* [70]. Alternatively, Patterson et al. highlighted the importance of *Peptoniphilus sp*. in the formation of vaginal biofilms [67]. They found that not only *G. vaginalis* but *Peptoniphilus sp.* adhered strongly to vaginal epithelial cells [67]. Its association to BV has been relatively recent, as Mazzarro et al. found that *Peptoniphilus*

sp. could be isolated from 36% of their persistent BV cases [71]. Together, this research suggests the association of *G. vaginalis* and *Peptoniphilus sp.* could lead to BV pathogenesis [67,71].

Other bacterial genera have been noted to be present in vaginal biofilms: *Bacteriodes, Corynebacterium, Lactobacillus, Veillonella, Ruminococcus* and *Streptococcus* but were not highly organized like *G. vaginalis and A. vaginae* [66]. This finding is in line with other more recent studies; outlined below; suggesting increased quantity and diversity of multiple bacterial groups in women with BV [9,16,19,41,42]. This may speak to the frequent recurrence of BV which is polymicrobial, rather than the single species it was previously believed to be, and treated as such.

The Role of Sialidase in *G. Vaginalis* Colonization

Sialidase, production by *G. vaginalis*, is thought to be detrimental to the mucous layer of the vagina, as it has mucin degradation abilities, and is also suggested to be associated with biofilm production. Our team has successfully classified *G. vaginalis* strains into four subgroups, Group A, B, C and D. When analyzing sialidase gene expression and activity of these subgroups, our analyses did not show sialidase enzymatic activity in each of the subgroups, but positive gene detection was seen in all isolates of Group B and C, and 50% of isolates in Group D [63]. Our team went on to further investigate the distribution of *G. vaginalis* subgroups in women with and without BV, and found Group B was significantly more abundant in women with BV than other subgroups, or the “normal” vaginal microbiome [63]. More recently, using 112 *G. vaginalis* isolates obtained from Canada, Belgium and Kenya, our group confirmed the classification of these subgroups through chaperonin-60 gene sequencing, and also further explored the sialidase gene expression and activity in these subgroups with more recent methods. We found that sialidase gene expression was detected in all isolates of Group B, C and D and one isolate of Group A. Sialidase gene activity was also positive in all Group B

isolates, in addition to three of the 35 isolates in Group C. The presence of sialidase activity in a specimen is currently used in a screening test for BV, thus making sialidase enzyme production important for clinical diagnosis, when microscopic capabilities are not available. Differential sialidase production by *G. vaginalis* subgroups may lead to different clinical statuses in patients, and each subgroup may have distinct roles in BV pathogenesis [72].

Risk Factors for BV

Risk factors for BV are multiple, further making its etiology difficult to define. Studies have suggested multiple risk factors for BV including demographic characteristics, sexual behaviours and health practices, and baseline vaginal microbiome composition. Moreover, certain ethnicities have been suggested to have higher rates of BV than others. It is well recognized that African women have the highest prevalence of BV worldwide [17]. This may be a result of sociodemographic and lifestyle risk factors for BV, rather than ethnicity [73,74]. Studies are conflicting; some having shown that when risk factors are controlled for, Black race does not increase risk of BV [26,75], while others do demonstrate an association between race and BV after controlling for associated risk factors [51,52,76–79]. Cherpes et al. found that women who were of a black ethnicity may have an episode of BV 1.6 times more frequently than women who were not of a black ethnicity [80] Alternatively, some studies suggest a biologic or genetic link between race and BV [73,74]. Ravel et al. found that only 59.6% of Black women had vaginal communities dominated by *Lactobacillus* species compared to 80.2% of Asians and 89.7% of Caucasians [39]. Furthermore, Black women were more likely than Asian or Caucasian women to have vaginal communities cluster with the heterogeneous CST consisting of diverse bacterial species. [39]. Some of these diverse bacterial species have been linked to BV, i.e. *G. vaginalis*, *Mobiluncus*, *Atopobium vaginae* [39]. Additionally, Srinivasan et al. found that in Black women without BV, there was greater abundance of several BV-associated bacterial species, which may contribute to increased risk for BV [41]. The role that race has in BV is controversial.

Certain sexual behaviours and health practices have also been suggested to have a role in BV, and include: recent sexual activity [80,81], higher number of heterosexual sexual partners [26,52,56,78,80–83], vaginal intercourse after anal sex [80], intercourse with an uncircumcised male partner [80], unprotected vaginal intercourse [81–84], lack of hormonal contraceptive use [52,79,84–87], inconsistent condom use [83,84], intrauterine device (IUD) use [86], female sexual partners [53,64,81–83,85,88–92], and vaginal douches [52,93,94].

A study conducted by Cherpes et al. found that the hazard ratio (HR) for the presence of BV in women who engaged in sexual activity in the past four months was 2.4 [80]. Hawes et al. also conducted a study and found that having a new sexual partner was associated with 2.5 times the risk of developing BV [26]. The risk conferred by sexual activity with a male partner is suggested to be a result of the alkaline property of semen, which causes a loss of vaginal acidity for hours after intercourse. It is possible that this loss of acidity may allow for the overgrowth of anaerobic bacteria and/or affect the maintenance of H_2O_2 and lactic acid producing lactobacilli, resulting in an imbalance of the vaginal environment and subsequently the acquisition of BV [80,82]. Additionally, males may be carriers of BV-associated organisms; as studies have isolated BV-associated organisms from the male genital tract [83,95–97]. Alternatively, Li et al. noted that woman who had a greater frequency of sexual intercourse each month were less likely to have BV [86]. They attributed this to the antibacterial effect of prostasomes (prostate-derived organelles in semen) or prostasome-derived proteins [86]. However, frequency of sexual intercourse was not adjusted by number of sexual partners, as this information was not collected [86]. This association also been noted by Mitchell et al. where after adjusting for number of partners, reporting greater than 5 episodes of vaginal intercourse in the past 90 days was associated with an increase in lactobacilli quantities [18].

Additionally, Cherpes et al. found the HR for vaginal intercourse after anal sex (within the past four months) to be 2.1 [80]. This sexual behaviour may directly disrupt the vaginal microbiota, or increase the risk of acquiring certain microorganisms that lead to acquisition of BV [26,80]. Furthermore, Cherpes et al. noted a HR of 1.9, for intercourse with an uncircumcised male partner (within the past four months) [80]. Uncircumcised males have higher proportions of gram-negative rods in the subpreputial space compared to circumcised males [80]. Therefore, males may transmit infections, or introduce certain microorganisms into the vaginal environment, which cause a state of dysbiosis resulting in a BV infection [80,83].

Bradshaw et al. observed that women using estrogen-containing contraceptives had a 50% decline in their risk for recurrent BV [84]. Estrogen may increase the glycogen content of epithelial cells, which is an essential substrate for the generation of lactic acid by *Lactobacillus* species, and in turn this may promote a healthy vaginal environment [84]. Several other studies have also noted the protective effect of hormonal contraceptives [79,85,87]. Bradshaw et al. found that inconsistent condom use almost doubled the risk of recurrent BV [84]. As mentioned above, male sexual partners may transmit BV-associated organisms and/or cause an imbalance in the vaginal environment; condoms may provide protection against these factors [82,97].

In women who have sex with women there is a twofold increase in risk of BV [82,83,85]. Having a female sexual partner with a history of BV has also been noted to increase the risk of BV acquisition [53,88,90]. In addition, having a symptomatic female sexual partner was associated with a 3-4 fold increase in BV incidence [90–92]. Women may have a longer duration of infectiousness of BV or increased sensitivity to BV-associated bacteria than men, thus female sexual partners may confer a risk of BV through the exchange of vaginal secretions during sexual contact [83].

Reports have also suggested that intravaginal applicants and practices (such as feminine hygiene products, vaginal gels, vaginal lubricant, vaginal suppositories, intravaginal rings, and vaginal douching) may alter the microbiome composition, and can have differential impacts [50,53,86,98–101]. Vaginal douches were found to be associated with BV [86], where those who used vaginal douches were found to have a 21% increase in the risk of developing BV [93]. Similar findings were found by Klebanoff et al., where a significant causal association between vaginal douching and BV was found, and where douching was not in response to BV symptoms [94]. Some douche products contain surfactant detergents, substances that may inhibit lactobacilli or disturb cell membranes within the vaginal environment, causing irritation to the mucosal surfaces which can subsequently lead to an increase in susceptibility to genital tract infections [93,102]. Douching may also alter the vaginal environment to one that is more hospitable to BV-associated organisms [26]. Additional studies also found a positive association [26,103,104], but conversely others [89] found that douching was not associated with BV; different definitions of douching could explain the differences seen between these studies. Research has also noted differential findings on the cessation of these vaginal practices and the risk of BV, where cessation may not decrease the risk of BV [93,100,101,105], or may promote a healthy vaginal microbiota [103,106].

In addition to these sexual behaviours and health practices, cigarette smoking [80,82,90,107], HSV-2 infection [80,100] and IUD use [86] have also been shown to be positively associated with BV. Cherpes et al. found that the HR for cigarette smoking (in the past four months) and HSV-2 infections were 1.5 and 1.7, respectively [80]. Cigarette smoking and HSV-2 infections may result in an immune response in the genital mucosa, and as a result cause an alteration in the vaginal microbiota and thus promote the overgrowth of BV-associated anaerobes [80,100]. Smart et al. also found that smoking was their strongest non-sexual factor for BV; the odds of developing BV as a smoker was almost double than those who were non-smokers [82]. Similar findings were also

found by Bradshaw et al., and Brotman et al. [90,107]. Several chemicals found in cigarette smoke have also been found in the cervical mucus of smokers [82,107]. Either directly causing an imbalance in the vaginal flora or depleting Langerhans immune cells in cervical epithelium leading to local immunosuppression, these chemicals may cause a disruption in the vaginal flora leading to BV [82,107].

IUD use may also have a role in BV, where the presence of the IUD tail in the endocervix or vaginal environment may favour the growth of anaerobic bacteria and *G. vaginalis* [108]. Li et al. found that those who used IUDs were at a higher risk of having BV, which verifies findings from older studies [86,108–111]

The composition of a woman's vaginal microbiome can act as a precursor to the development of BV. The lack of H_2O_2-producing vaginal lactobacilli [26,27,80], and intermediate vaginal flora [80,111] has been associated positively with BV. The role of hydrogen peroxide to inhibit the overgrowth of anaerobic organisms is controversial[22,26–29,80]. However, Cherpes et al. found that the HR for intermediate vaginal flora and lack of H_2O_2-producing vaginal lactobacilli was 2.7 and 2.0, respectively [80]. Intermediate vaginal flora may be positively associated with BV, as it is thought to be unstable. Cherpes et al. noted that two-thirds of women with intermediate vaginal flora will either shift to a 'normal' flora or a BV-associated flora [80]. Previous studies support these findings [26,27]. Several studies have also suggested an association between BV-associated bacteria (BVAB) species and BV, where detection of BVAB species in the vaginal microbiota was associated with a greater BV acquisition risk [88,112]. At various times and in different women, these risk factors and precursors may play a role in BV pathogenesis [50].

Women with Bacterial Vaginosis

Culture-dependent methods identified the presence of *G. vaginalis* species in the vaginal microbiota of women with BV [31]. However, as a result of its fastidious nature, cultures are not a reliable means to detect *G. vaginalis* [31]. As mentioned previously, culture-dependent methods cannot identify all organisms present in the vaginal microbiota as many such as *Bacteroides spp, Peptostreptococcus spp, Mobiluncus spp,* and *Mycoplasma hominis* are not easily cultured due to their fastidious nature [31].

More recently, culture-independent studies have set out to characterize the vaginal microbiome of women with recurrent BV. Metagenomic data suggests that BV does not possess a single universal profile, but a collection of different microbial profiles with great diversity [9,16,41]. Many bacterial microorganisms have been found to be associated with BV including; *G. vaginalis, A. vaginae, BVAB -1, -2, -3, L. iners* and species in the following genera: *Prevotella, Eggerthella, Dialister, Megasphaera, Sneathia, Leptotrichia, Parvimonas (*formerly *Peptostreptococcus), Veillonella, Bacteroides, Mobiluncus, Porphyromonas, Mycoplasma, Ureaplasma, Streptococcus, Staphylococcus, Gemella, Escherichia/Shigella, Aerococcus, Papillibacter* [9,16,19,41,42,62,113,114]. Women with recurrent BV may not possess a single dominant species unlike that seen in most healthy asymptomatic women [9,16]. Molecular-based techniques have shown that the number of phylotypes associated with BV is far greater than that seen in intermediate and normal vaginal microbiomes [9,16]. Vitali et al. found that the BV-related bacteria, *A. vaginae*, *Prevotella* sp. and *Mycoplasma hominis*, were significantly more abundant in women who had BV compared to healthy women [42]. Although these bacterial species were also present in healthy women, suggesting that abundance levels rather than mere presence may be considered the hallmark of recurrent BV [42].

Additionally, research suggests that *G. vaginalis* acts in combination with other microorganisms to cause BV, while others suggest perhaps *G. vaginalis* and *A. vaginae* in combination are the root of BV [115–120], or that *G. vaginalis* and *Prevotella sp.* synergistically act together to aggravate the BV process [16,19]. Alternatively, additional research has shown the presence of *G. vaginalis* in the vaginal microbiomes of healthy asymptomatic women [16,19,39,40,59–63], suggesting that the presence of certain microorganisms does not always indicate an abnormal vaginal microbiome either [9,31].

Another important species is *L. iners. L. iners* and *L. crispatus* are both common *Lactobacillus* species in women [9,41,42]. Although *L. crispatu*s is rarely dominant in women with BV, *L. iners* has been detected at high levels in women with and without BV. The reasons behind this have yet to be identified, but it is postulated that it may be due to *L. iners* ability to adapt to conditions associated with BV [9]. It is suggested that *L. iners*, in comparison to other *Lactobacillus* species, has a more pronounced ability to resist unknown factors allowing its survival during the onset of BV [9,121]. An alternative presumption is the relative lack of antagonism of *L. iners* to BV-associated anaerobes, such that *L. iners* dominance predisposes women to acquiring BV. Moreover, *L. iners* is a poor producer of H_2O_2; thus its dominance in the vaginal microbiota could permit the growth of disruptive bacterial species at higher proportions, which would result in BV symptoms, but a normal Nugent score [9].

As previously mentioned, H_2O_2 promotes a healthy vagina microbiota through preventing the overgrowth of other bacterial genera. Antonio et al. found that H_2O_2 was produced by 95% of *L. crispatus* and 94% of *L. jensenii* isolates, compared with only 9% of *L. iners* isolates [27]. Women lacking H_2O_2 producing lactobacilli were more likely to develop BV, as they could not inhibit the growth of BV-related genera [26]. Overall, *L. iners* is suggested to be a transitional bacterial species; pushing a healthy microbiota into an unhealthy state [9,16,19,41]. *L. iners* is now

known to be an important player in the vaginal microbiome. This was not always evident as *L. iners* grows only on blood agar and this was not used in traditional culture methods [9].

Current metagenomic data reveals the complexity behind the vaginal microbiome and how little is known, especially with regards to BV. The heterogeneity in the composition of bacterial taxa in women with BV aids in its difficulty to diagnose and treat. Incorporation of culture-independent methods is important in the construction of vaginal microbiome profiles.

Universal Gene Targets

The 16S rRNA gene encodes the small subunit 16S ribosomal RNA, which is an important structural component of ribosomes; machines which have a role in protein synthesis [122,123]. The 16S rRNA gene target has been traditionally used in phylogenetic studies of microbial communities and sequence-based taxonomy [124,125]. One key advantage, is its ability to overcome the 'great plate count anomaly', which is the fact that a large number of organisms cannot be cultured on standard media, and thus are not represented in culture-based communities [126]. A key example would be its role in identifying the causative agent of Whipple's disease; the uncultivable bacteria *Tropheryma whipplei* [127]. Thus, 16S rRNA is able to identify uncultivable and cultivable organisms [128]. Also, 16S rRNA is beneficial in identifying slow-growing bacteria, which would take a significant amount of time to be grown in culture before phenotypic tests could be applied [127].

In addition, other advantages conferred by 16S rRNA include its universality [123,127,129,130], the presence of multiple gene copies thus making it an abundant and easily detectable target [124], its highly conserved regions allowing construction of broad-range universal polymerase chain reaction (PCR) primers, and presence of highly variable regions for the identification of individual species [123,127,128]. Moreover, the Ribosome Database Project (RDP) is an extensive

reference database consisting of many divergent taxa from a wide range of environments [123,130]. Thus far, the 16S rRNA target has reclassified and renamed numerous bacterial genera and species, classified uncultivable bacteria, determined phylogenetic relationships, and allowed for the discovery and classification of novel bacterial species [127].

Despite the above advantages, several limitations exist. The 16S rRNA gene target consists of nine hypervariable regions (V1 to V9), and there is no consensus on which region is ideal to sequence [131,132]. As a result, different studies sequence different regions and thus results are not necessarily directly comparable [131]. Moreover, some regions are better at resolving certain taxa than others [125]. Huse et al. noted that *Lactobacillus* species could be differentiated using hypervariable regions V1 to V3 but not hypervariable regions V3 to V5 [125]. Hummelen et al. also noted that primers targeting hypervariable region V6 were biased against *Sneathia*, *Leptotrichia*, *Ureaplasma* and *Mycoplasma*; bacterial species which have been associated with BV [9,16,19,41,42,62,113,114,133]. In addition, the 16S rRNA target does not provide sufficient information to discriminate between closely related species [123,124,134]. Nucleotide differences between closely related species are not disturbed throughout the gene but clustered in hypervariable regions, thus only sequencing the correct region of the 16S rRNA gene will allow closely related species to be distinguished from one another [132]. We also see that although having multiple gene copies within the genome can be advantageous, it can also complicate its use in quantification studies [123,124]. Furthermore, the gene structure itself proves to be disadvantageous as the alternating regions of variable and conserved regions can result in the formation of stable secondary structures, resulting in PCR artifacts [124,135].

The chaperonin-60 (*cpn*60) universal target gene is an alternative gene target which can be used in phylogenetic studies and sequence-based taxonomy. The *cpn*60 universal target gene has important advantages over the 16S rRNA target, and may overcome limitations which

arise when using the 16S rRNA target. Chaperonin-60 is a 60-kilodalton heat shock protein. It is an important player in cellular protein folding, and universally found in eukaryotes and most prokaryotes [136]. The *cpn*60 universal target gene is a 552-558 base pair region present on the *cpn*60 gene; unlike 16S rRNA there are not several regions which can be sequenced [122]. Also unlike 16S rRNA, the *cpn*60 universal target gene is highly discriminatory being able to provide sufficient species-level identification, and uniform interspecies variability [136]. Research conducted by our team, showed that *cpn*60 was able to yield higher numbers of operational taxonomic units (OTUs) at different cut-off points, compared to the 16S rRNA target, when specifically looking at *Prevotella* identified sequences [137]. Also mentioned previously, our team has successfully classified *G. vaginalis* strains into four subgroups, and this was accomplished using the *cpn*60 universal target gene [63,72]. We have also provided evidence that supports the eventual reclassification of subgroups as different species of *Gardnerella*. The *cpn*60 universal target sequence offers a robust tool for identification of subgroups within *G. vaginalis* that may not be possible using other targets [63,72]. The *cpn*60 coding region consists of uniform sequence variation and thus has a low likelihood of forming chimeric PCR products [124,135]. Additionally, *cpn*60 genes are present as single copies within the genome. This may be more difficult to detect, but this target is less complicated in quantification studies, and the presence of a single gene copy eliminates sequencing artifacts associated with multiple gene copies that are not identical (as may be seen with 16S sequencing) [124]. Using the *cpn*60 universal target gene in culture-independent sequencing methods may identify new bacterial taxa, over the traditionally used 16S rRNA gene target, in women with recurrent BV, and may lead to the development of targeted treatments.

1.4 Virome

Overview

The human microbiome consists of other microorganisms besides bacteria, however most studies explore the bacterial organisms present in the human microbiome. The human virome is also an important, complex aspect of the human microbiome which profoundly influences host physiology [138]. Viruses are parasitic biologic entities, which require host cells for replication [139]. They are considered ubiquitous as they infect nearly every cell type, and are found in every ecosystem [139]. Viruses are made up of either single or double stranded RNA and/or DNA, and a protein capsid. Some viruses may also have a lipid envelope modified from their host cells [139]. Viruses encode proteins that regulate the cell cycle and host gene expression, suppress or disrupt host immune response, and encode microRNAs that regulate cellular processes [140]. From their functions, one can see that viral communities are continually interacting with the host [140]. These dependent microorganisms interact with genetic material of most cells on the planet, including the trillions of bacteria within the human microbiome [139].

The mammalian virome includes viruses that infect eukaryotic cells and archaea, bacteriophages that infect bacteria, and virus-derived genetic elements in host chromosomes that can change host-gene expression, express proteins or even generate infectious viruses [141,142]. Analyses of human viral communities are complex, as viruses are small in size, rapidly evolve, are difficult to culture, and lack a universal phylogenetic marker (i.e. 16S rRNA, *cpn*60) [143]. The virome contains a range of diverse elements, and is only now being characterized on a sequence level due to advancements in sequence and bioinformatic technologies [138,139].

Background

With regards to the female genital tract, not a lot is known about the dynamics of the vaginal virome [144]; in fact only one exploratory study exists which aims to define the microbial

communities which exist in the human microbiome. This first large-scale molecular analysis of the viral flora was completed by the Human Microbiome Project. Wylie et al. sampled 102 healthy-asymptomatic individuals, from five major body sites; nose, skin, mouth, stool and vagina [140]. Vaginal samples showed a dominance of papillomavirus, where 37.5% of subjects sampled carried one or more *Alphapapillomavirus* [140]. Subjects carried as many as 4 distinct *Alphapapillomavirus*, including oncogenic types HPV-16, and HPV-18 [140]. Additionally Wylie et al. suggested that some components of the virome were stable over time, including the vaginal virome [140]. Thirty to fifty percent of subjects who had *Alphapapillomavirus*, *Betapapillomavirus*, and *Gammapapillomavirus* detected in their samples collected at one visit, had the viruses detected again in their follow-up visit [140]. Furthermore, this group suggested that viruses and bacteria may have a dynamic relationship within the microbiome [140]. As mentioned above, *Alphapapillomavirus* were detected in 37.5% of vaginal samples. *Alphapapillomavirus* were more common in an individual with more bacterial diversity; whose vaginal microbiome showed less than 85% *Lactobacillus*-dominance, and an increase in anaerobic bacteria [140]. It is unknown whether this viral-bacterial relationship is causative or if both are present due to simultaneous exposure [140]. Overall, Wylie et al. found that not only was the virome highly diverse like the bacterial component of the microbiome, but there was also high interpersonal diversity [140].

Although there is limited information on the vaginal virome, there are several persistent pathogenic viruses known to be present within the female genital tract of some women which include; *Cytomegalovirus* (CMV), the human immunodeficiency virus (HIV), herpes simplex virus (HSV), and the human papillomavirus (HPV).

Cytomegalovirus

Cytomegalovirus (CMV) is a linear double-stranded DNA virus in the *Herpesviridae* family. CMV is highly complex, and different strains have been clustered into four distinct groups

on the basis of genomic variation in the amino-terminal region of the envelope glycoprotein [145]. CMV is prevalent not only in the immunocompromised population, but it is also present in its active and/or latent form in those who are immunocompetent and asymptomatic [146,147]. However, even though we see a high prevalence of CMV in the population, CMV of the genital tract is an uncommon phenomenon [146]. A study conducted in a STI clinic found only 4-12% of patients were CMV positive using cervical swabs; accounting for age, geographical and socioeconomic distribution [146,148]. While high-risk CMV patients are screened and identified, asymptomatic low-risk individuals often fall under the radar [146].

CMV is an opportunistic pathogen, causing serious consequences in immunocompromised patients, as well as a common cause of congenital infection [145,147]. CMV is present in 40-100% of the population[147,149–151], and can found in the genital tracts of a majority of women regardless of whether BV is present (91%) or not (83%) [145]. Ross et al. showed that women with BV had increased rates of CMV seroprevalence and seroconversion [145]. Additionally, women with BV are 4 times more likely to shed CMV virus than those without BV, irrespective of whether or not they were symptomatic, suggesting that the presence of BV in women may facilitate local CMV replication in the genital tract. [145].

In addition to the relationship between BV, and CMV; a relationship between CMV and HPV may also exist. The prevalence of cervical CMV in healthy women is 1-2%, but the prevalence of CMV seen in low grade squamous intraepithelial lesion (LSIL) and high grade squamous intraepithelial lesion (HSIL) is significantly higher at 30% and 53%, respectively [146,152]. Although the exact mechanism remains to be determined, this finding suggests that CMV is transmitted along with HPV or sequentially allowing it to influence HPV-induced dysplastic changes in the cervix [146]. One possible mechanism is that CMV viral replication involves the expression of multiple genes; immediate-early, early, and late genes [147]. It has been previously

suggested that the immediate-early gene products can activate other viral and cellular genes [146], which may be the case here. There are conflicting studies however, where CMV was found to not play a role in HPV lesions [146]. The viral interactions within the genital tract remain to be determined.

Human Immunodeficiency Virus

The human immunodeficiency virus (HIV) results in the progressive destruction of $CD4^+$ T lymphocytes, cells that are vital to keep the immune system functioning normally [153,154]. There are over 34 million individuals infected with HIV globally, and up to 7,000 new infections each day [50]. Individuals who are infected with HIV are at a higher risk for opportunistic infection due to immune suppression. Women with HIV can be infected by HSV, HPV, and/or CMV among viral and bacterial pathogens [154–156]. Ameur et al. aimed to profile the vaginal microbiome of 20 HIV positive women to determine the spectrum of coinfection [156]. Forty-six different HPV types were present in this cohort, with 5 to 21 HPV types present per woman. The most abundant type present in this population was HPV-58, followed by a number of other high-risk HPV types. Two novel HPV types were found, which indicates the presence of more undiscovered HPV types. HPV prevalence is seen to be higher in HIV-positive individuals, and it is suggested that low-risk HPV types in immunocompromised people may result in tumour formation due to immune suppression [156].

In addition to HPV, three other viruses were found to be present in HIV-infected individuals. *Torque teno* and SEN-virus (SEN-V) were found in multiple samples, and the *John Cunningham* (JC) virus was found in one woman. *Torque teno* is a virus that has not been associated with any clinical syndromes while SEN-V is associated with post-transfusion hepatitis, and JC viruses may become active in immunosuppressed individuals, resulting in brain and kidney infections [156].

There is limited information on the viral dynamics within the genital tract. Money et al. suggested that local hormonal variations, seen through the menstrual cycle, may affect HIV viral load in the genital tract which may not be seen in the plasma [144]. The genital HIV viral load was found to be statistically lower at the periovulatory phase, despite not being noted in the plasma [144]. It is suggested that this decline may be a result of a decrease in local HIV replication/production, or perhaps of enhanced cervical mucus production resulting in a washout effect of the virus [144]. In addition, detectable HIV virus levels were seen in the genital tract, even when it was not in the plasma [144], which speaks to how local factors may play a role in the dynamics of microbial populations.

HSV and BV are known to increase the risk of HIV transmission and acquisition [50,154]. HSV episodes have been shown to increase not only plasma HIV shedding but mucosal shedding too [154]. Furthermore, HSV-2 has been found to be a significant risk factor for the acquisition of HIV [100]. HSV-2 is also positively associated with a 2-fold increase in the risk of HIV transmission [157]. Not only is HSV a suggested risk factor for HIV, but women who are HIV positive also have a higher prevalence of HSV [155,156].

As mentioned before, BV increases the risk of HIV acquisition. Women with BV are suggested to have elevated levels of pro-inflammatory cytokines, which allowed for a favourable state within macrophages for HIV replication, thus placing women with BV at a significantly higher risk of infection [50].

Herpes Simplex Virus

Herpes simplex virus (HSV) -1 and 2 are the second most prevalent sexually transmitted viral infection worldwide, and the most common cause of genital ulcers [100,155,157,158]. In Canada, the annual incidence of HSV-1 and 2 is unknown [159], however the HSV seroprevalence of

reproductive-age women is 17%, which increases with age [155]. Also, it is estimated that 23% and over 50% of women in the United States and Sub-Saharan African, respectively, are infected with HSV-2 [100]. With regards to HSV-1 genital infections, its incidence and prevalence are increasing globally [157,159]. In a study conducted by Fife et al., the HSV-2 incidence was 7.3 per 100 persons-years; it is suggested that every year 7.3 persons will become infected with HSV-2 [157]. Several factors were found to be positively associated with HSV-2 infection, where age, years of sexual activity, number of lifetime sexual partners and history of STDs were the strongest predictors [157]. Number of lifetime partners has been previously noted to be a risk factor, where individuals with more than 10 lifetime partners had a 50% rate of genital herpes [155]. Females are found to be at a greater risk for genital herpes infections [100,159]. The annual transmission rate in HSV-discordant couples when the male is the source of infection is 11-17%, compared to the 3-4% transmission rate in couples when females are the genital herpes source [159].

Eighty percent of women who are symptomatic will present with typical genital symptoms and signs, and the remaining 20% will present with atypical symptoms such as nonspecific ulceration, erosions, fissures, excoriations, erythematous patches or other non-specific lesions, genital pain or urethritis, aseptic meningitis and cervicitis [159,160]. HSV-2 is associated with genital infections, and HSV-1 is associated with orolabial disease and up to 50% of new genital ulcers [158]. Viral shedding can be seen in individuals who have diagnosed with HSV-2, and even in those who are seropositive and have mild to no symptoms [158]. This proves problematic, for individuals who are sexually active as they may not 1) be aware of viral shedding in the absence of symptoms, and 2) be aware of their HSV status; up to 75% of individuals who are HSV-2 seropositive are not diagnosed [158].

Women infected with HSV-2 have a higher prevalence of BV compared to those who are seronegative. Masese et al. found that women with HSV-2 had a 30% increase in the odds of having a BV episode; the mechanism behind this association remains to be determined [100]. It is suggested that the vaginal microbiota is altered due to immune activation by HSV-2. BV is also believed to increase HSV-2 transmission, as BV increases genital shedding of HSV-2 [100]. BV has also been suggested to increase the risk of acquisition of HSV -1 and 2 [161]. There also exists a relationship between HSV and HIV which is mentioned above, studies have suggested an association between the two where some indicate HSV-2 to be a significant risk factor for the acquisition of HIV [100], and others suggest women who are HIV positive have a higher prevalence of HSV [155].

Human Papillomavirus

There are over 176 different types of the human papillomavirus (HPV) present in the human population today [162]. Over 40 of these infect the female genital tract [163–165], and at least 15 of these are deemed high-risk for cervical cancer [164]. HPV types can be designated as low and high risk, for wart-causing and cancer-causing, respectively [162]. Collectively, types 16, 18, 45, 31, 33, 52, 58 and 35 are designated as high-risk HPV types and account for 9 out of 10 cervical cancer cases [165]. HPV 16 and 18, have shown a strong association for cervical cancer [162,165], and account for 70% of all cervical cancers [17].

These high-risk HPV types play a key role in the development of cervical cancer [17,162,166]. Persistence of HPV infections leads to the development of high grade lesions, which progresses to cervical cancer [17,162]. However, not all infections will persist and progress to cervical cancer [17]. In fact, most women will have mild or no cytological abnormalities, which will go unnoticed or regress to normalcy [17,162]. It is unknown why, some high-risk HPV infections lead to cervical cancer in some women, and not in others [17]. There are several high-risk HPV types, which are

commonly associated with normal cytology in women, including HPV types 16, 18, 31, 52, and 58 [165].

Ma et al. sought to define HPV types present in the human microbiome, and discovered the prevalence of HPV in the vaginal microbiome to be 41.5% [162]. Forty-three HPV types were identified, including types that had not been previously characterized. Most HPV types identified in the vagina belonged to the genus *Alphapapillomavirus*. Also, of those types that have been previously characterized to be high-risk, ten were found in nine women in their cohort; who did not have clinical disease. Ma et al. also found that more than 50% of their healthy subjects were coinfected with 2-3 HPV types [162]. Coinfection with more than one type has been previously documented; the risk of HPV acquisition of a certain type does not diminish if one is already infected with a phylogenetically related type [163].

The relationship between BV and HPV is complicated, and controversial. There are conflicting studies, evaluating whether BV results in an increased risk of HPV acquisition or vice versa [17], and whether there is a biological link between the two conditions or are they present frequently in similar environments [167]. Mao et al. identified a temporal relationship between BV and HPV, where through their time lag analysis it suggested a coinfection of HPV and BV, or HPV occurring first [167]. Gillet et al. was also able to show a positive relationship between BV and HPV, where the odds of a woman with BV being infected with HPV was 1.43 times greater than in a woman without BV [17]. However, the order of infection is uncertain, it is possible women were infected with HPV before a BV diagnosis [17]. In addition to the relationship between HPV and BV, the rate of cervical infection has been shown to be higher in women with HIV [165].

Both the vaginal bacteriome and virome are complex aspects of the vaginal microbiome, and like the bacteriome, we see that the virome is dynamic, diverse and understudied.

Additionally, there seems to be a complex interplay within the vaginal virome of which we do not fully understand.

1.5 Vaginal Microbiome Project (VOGUE)

The Vaginal Microbiome Project is a multi-site interdisciplinary team looking to characterize the vaginal microbiome. Knowledge about the vaginal microbiome is limited, and as it is a vital determinant of a woman's reproductive health, it is of the utmost importance to widen our knowledge base, define the microbial populations of the vaginal microbiome and understand its dynamic nature. In order to do this, the Vaginal Microbiome Project objectives are as follows:

1. Expansion of our existing team to enhance capabilities and broaden our research focus
2. Conducting our comprehensive research program, the goals of which are to:
    a. establish comprehensive clinical cohorts, complete with a sample repository cross linked to a clinical database
    b. refine methods and approaches for characterization of the entire vaginal microbiome, including prokaryotes, eukaryotes and viruses
    c. apply knowledge of vaginal microbiome structure and function for diagnostics, prediction of outcomes, and evaluation of intervention strategies
2. Exploring social and cultural determinants of behaviours that influence vaginal health
3. Implementing a real-time knowledge translation program

This project encompasses several different sub-studies, which aim to characterize the vaginal microbiome of distinct clinical populations. The five studies under the Vaginal Microbiome Project umbrella are: Vogue 1A (healthy-asymptomatic women), Vogue 1B (HIV-positive women), Vogue 1B2 (women with recurrent bacterial vaginosis), Vogue 1C (women

with low-risk pregnancies) and Vogue 1D (women with high-risk pregnancies). Ultimately these five studies will allow us to broadly characterize the vaginal microbiome of healthy non-pregnant and pregnant women, identify microbiome profiles associated with abnormal diseases or conditions, and determine microbiome profile changes associated with treatments or interventions.

Of these five sub-studies, my study focuses on Vogue 1B2; characterization of the vaginal bacteriome of women with recurrent BV; as well as utilizing the available vaginal viral samples and metadata from Vogue 1A, 1B and 1B2 to characterize the virome of these three populations of women.

1.6 Bacteriome

Rationale

The etiology of recurrent BV remains a topic of debate, as this is a poorly defined genital condition with a multifactorial nature. Not only is our knowledge in this field limited, but current diagnostic tools and treatments have a limited ability to aid women with recurrent BV. We will explore the vaginal bacteriome in women with recurrent BV, in hopes that not only will our project aid in the generation of pointed hypotheses in future studies by our group, but help us in better being able to define recurrent BV, and understand of the distinction between vaginal health and disease.

Hypothesis

We hypothesize that utilizing *cpn*60 sequencing methods, we will obtain predictive vaginal microbiome profiles from women with recurrent bacterial vaginosis as the foundation for diagnostic tools.

In order to test our hypothesis, bacterial profiles within our population will be analyzed through our clustering pipeline. We anticipate certain distinct profiles to be more prevalent in this population compared to the healthy-asymptomatic population. Gajer et al. has noted that some taxa in CST IVB, *Atopobium, Gardnerella, Mobiluncus, Prevotella,* and *Sneathia*, have previously been associated with BV [40,62,113,114]. Another study conducted by Mehta et al. found that the proportion of women clinically diagnosed with BV was lower in CSTs containing a higher relative abundance of *Lactobacillus* species [168]. Even a previous study conducted by our own group has indicated associations between BV-positive Nugent scores with CST IVA, IVC and IVD, whereas BV-negative Nugent scores were associated with *Lactobacillus*-dominated CST I, III, and V [47]. We hypothesize fewer women to possess profiles dominated by *Lactobacillus* species, and more women to possess distinct profiles dominated by *G. vaginalis* species and heterogeneous profiles.

Additionally, to examine the relationship between our distinct profiles and metadata, we will look for correlations between known risk factors and specific bacterial species. Literature has noted an association between *L. crispatus* and the absence of the four Amsel's criteria, where there is a strong association between *L. crispatus* and low pH, negative whiff test, absence of clue cells and normal vaginal discharge [41]. Additionally, an association between *G. vaginalis* and *A. vaginae,* and three of the four clinical markers was detected, where these species were associated with each criterion except abnormal vaginal discharge and amine odor, respectively [41]. *Leptotrichia amnionii* and *Eggerthella sp.*, species previously associated with BV [62,113,114], were the only two bacteria species to have an associated with all four Amsel's criteria [41]. Literature has also noted an association between recent sexual activity [80,81], higher number of heterosexual sexual partners [26,52,56,80–83], vaginal intercourse after anal sex [80], intercourse with an uncircumcised male partner [80], unprotected vaginal intercourse [81–84], lack of hormonal contraceptive use [52,79,84–87], IUD use [86], female sexual partners [53,64,81–83,85,88–92], and vaginal

douches [52,93,94], cigarette smoking [80,82,90,107], and HSV-2 infection [80,100] with BV. We hypothesize that each distinct profile will be associated with its own set of behavioural, clinical and demographic variables.

Objectives

Aim #1:

To characterize the vaginal bacteriome of women with recurrent BV, using culture-independent *cpn*60 gene sequencing.

Aim #2:

To compare vaginal bacteriome profiles between our cohort of women with recurrent BV, and healthy-asymptomatic women.

1.7 Virome

Rationale

Our knowledge on the vaginal microbiome in its entirety is also sparse. The majority of vaginal microbiome studies aim to characterize the bacterial species which inhabit the vaginal environment. There are no studies whose primary aim is to define the viral populations of the vaginal environment. Thus, our knowledge on the vaginal microbiome is limited to the bacterial species which occupy this environment.

We will explore the vaginal virome in each of the CSTs as defined by our group [47]. In order to analyze the virome in each cohort, the characterization protocol must be first optimized. This will give us a refined methodology to look at the virome in this study, as well as in future virome studies. We hope this project will also aid in the generation of pointed hypotheses in future studies by our group and in conjunction with the vaginal bacteriome, gain insight on the vaginal microbiome as a whole.

Hypothesis

We hypothesize that each CST will exhibit distinct and separate viral patterns, similar to the diversity defined in the CSTs of the bacteriome. Additionally, we hypothesize, that the diversity in viral patterns can be correlated with demographic, behavioral and clinical data.

In order to test our hypothesis, we will apply our refined virome methodology to explore the viral patterns of these CSTs. Most of the vaginal microbiome work today, consists of defining bacterial populations. However, Brotman et al. identified a differential distribution of HPV-positive samples across CSTs [107]. CST I and CST II; dominated by *L. crispatus* and *L. gasseri*, respectively, had lowered proportions of HPV-positive samples, compared to CST III and IV; dominated by *L. iners* and low-*Lactobacillus* groups, respectively. In addition, women who clustered into CST III and CST IVB were seen to have a higher proportion of high risk HPV-positive samples [107]. Literature has illustrated the diversity in microorganism populations and we anticipate the same diversity in our viral populations.

Additionally, as seen with the vaginal bacteriome, we mentioned that certain bacterial patterns could be correlated with our metadata, and we anticipate the vaginal virome will also exhibit correlations between viral diversity and metadata. In order to test this, we will look at variables which have been associated with bacterial CSTs and diversity, and determine if associations can be made to viral communities.

Objectives

Aim #1:

To refine methods and approaches for characterization of the vaginal virome.

Aim #2:

To use our optimized methods to define the vaginal virome in three different cohort of women; healthy-asymptomatic women, HIV-positive women, and women with recurrent BV.

2 Materials and Methods

2.1 Bacteriome

Ethics Statement

This research project entitled, Vogue (Vaginal Microbiome Group Initiative) Study 1B2, received ethics approval by the University of British Columbia – Children's & Women's Health Centre of BC Research Ethics Board (UBC C&W REB) with Certificate Number H11-01912 on June 1, 2012, and has been annually renewed.

Participant Recruitment

Women who were attending for gynecologic care for vaginitis/vaginosis symptoms potentially consistent with BV, were approached for enrolment into this study, named Vogue 1B2. Recruitment sites included the Reproductive Infectious Diseases clinic at BC Women's Hospital, Vancouver, BC, Dr. Julie van Schalkwyk's OBGYN consulting office, in Vancouver, BC, and several other speciality OBGYN practices in the Vancouver community. Study advertising material was also posted on social media, Craigslist, Walk-in Clinics, and in the community. Women recruited through advertisement were seen by a clinician in our Women's Health Research Institute (WHRI) Research Clinic at BC Women's Hospital.

We aimed to enroll 50 women, who fit the inclusion criteria into the study. The inclusion criteria were: had an adequate comprehension of the English language to sign written informed consent, non-pregnant, between the ages 18-49, not menopausal, HIV negative, and had a clinically documented or self-reported history of recurrent BV which was defined as at least 4 episodes in 12 months with BV or BV-like symptoms; including excessive, troublesome discharge and malodor. The exclusion criteria were: inability to provide informed consent, currently pregnant, younger than 18 or older than 49, menopausal, HIV positive, and no history of recurrent BV.

Our target sample size was based on a power calculation that took into consideration the ability to compare the profiles of women with recurrent BV to healthy asymptomatic women. As part of the CIHR funded team grant to study the vaginal microbiome, multiple cohorts of women were being recruited. The baseline cohort, Vogue 1A, recruited non-pregnant women with no gynecologic concerns or symptoms. Vogue 1A had an enrolment target of 300 women, thus for an 80% power to detect differences between these populations, enrolment of 50 women was needed; power calculations described in the Statistical Analyses section below. However, recruitment proved to be very challenging and despite many enhancements of the recruitment efforts, a lower than anticipated target was reached. Due to budgetary and time constraints, samples from 26 women were available for analysis at time of completion of this project arm.

Data Collection

Prior to data and sample collection, informed consent and optional consent for sample storage for 25 years was obtained from study participants (Appendix A and B) Behavioural, demographic and clinical information was collected via interview. The demographic data included: age, height, weight, body mass index (BMI), ethnicity, marital status, education level, and general residential location. General medical history was obtained and a genital infection history was documented for bacterial vaginosis, yeast infections, urinary tract infections, trichomoniasis, genital warts (condylomatas), genital herpes, chlamydia, gonorrhea, and syphilis, as well as the treatment method for the most recent infection. Antimicrobial use was recorded for the 3-month period prior to the study visit, and prescription and non-prescription drugs for the 2-month period prior to the study visit. Information about reproductive health included the first day of the subject's last menstrual period and regularity, tampon and menstrual cup usage, pregnancy history, recent vulvo-vaginal symptoms, feminine hygiene product usage, and contraception usage. Sexual activity information was documented, and including gender of sexual partners, recent vaginal intercourse occurrence, number of recent sexual partners, pain

experienced during vaginal intercourse, frequency of receiving oral sex, frequency of anal sex, and sex toy usage. History of substance use included both current and previous use of heroin, cocaine, crack, crystal meth, tetrahydrocannabinol(THC)/marijuana, opiates/opiods, benzodiazepines, methadone, alcohol, tobacco, and any other substances reported by the study subject. Additionally, the attending physician documented clinical findings, based on examination on the day of the study visit. All information was de-identified by assigning each participant a code number. The data collection forms can be seen in Appendix C and D.

Database Design

An electronic database was created using a secure web application; Research Electronic Data Capture (REDCap). This database is securely stored on a protected server at the Child and Family Research Institute (CFRI). All data collected via data collection forms through patient interviews, and study lab results were entered into REDCap for more organized, efficient and accessible data storage

Sample Collection

Vaginal samples were collected during the clinically indicated pelvic exam by the attending clinician. At the time of the speculum exam, swab samples were taken from the posterior fornix and lateral vaginal wall. Three vaginal swabs were collected, one to generate vaginal profiles using the *cpn*60 universal target gene, one for Gram stain (Nugent's) scoring and a back-up sample. All samples were identified only by the patient's study number. The Gram stain swab used was the Copan Sterile Transport Swab Suitable for Aerobes and Anaerobes (Copan Diagnostics Inc., Murrieta, CA) for the purposes of storage and transportation to the laboratory. The Gram stain swab was transported to the Children's and Women's hospital pathology laboratory, Vancouver, BC, for Nugent scoring using validated methodology. The swabs for genomic analysis were PurFlock Ultra Sterile Flocked Swab Applicators with Tips (Fisher Scientific Company, Ottawa, Ontario). They were stored in empty

Copan 1mL Universal Transport Medium tubes (Copan Diagnostics Inc., Murrieta, CA), and stored in a -80˚ Celsius freezer within four hours of swab collection at the WHRI laboratory. The swabs for genomic analysis were batched and shipped to Dr. Janet Hill's laboratory at the University of Saskatchewan, Saskatoon, for analysis of vaginal microbiome profiles through *cpn*60 profiling. The second swab was stored in the WHRI laboratory as a back-up sample.

Metagenomic Analysis

Generation of bacterial profiles using *cpn*60 methodology comprised of *cpn*60 amplicon generation, pyrosequencing and bioinformatic analysis. To generate *cpn*60 amplicons, DNA was extracted from vaginal samples using MagMAX™ DNA extraction (Life Technologies Inc., Burlington, ON).

For DNA extraction, the dry vaginal swabs were homogenized with 300 µL of sterile phosphate buffered saline (PBS) buffer (pH 7.4) by vortexing the solution and sample for 30 seconds. Afterwards, 200 µL of the sample solution was removed from the swab container and placed into a 1.5 mL tube. In a separate tube, 235 µL of MagMAX lysis/binding buffer solution was added to a prepared tube of zirconian beads in a guanidinium thiocyanate-based solution. 175 µL of the sample solution was then added to the prepared tube of zirconian beads. This tube was vortexed for 15 minutes, and then centrifuged for three minutes at 16,000 x g. This allowed for mechanical disruption of cells, releasing the nucleic acid content into the guanidinium thiocyanate-based solution, which provided nucleic acid protection by inactivating nucleases. Secondly, an 8x12 plate required for MagMAX DNA extraction was prepared. Ten microliter of lysis binding enhancer was added to row A of the plate. This was followed by the addition of 10 µL of magnetic beads to row A. Rows B and C of the plate had 150 µL of Wash Solution 1 Concentrate added to each well, , while rows D and E of the plate had 150 µL of Wash Solution 2 Concentrate added to each well. Row F of the plate had 50 µL of Elution buffer

added. The supernatant liquid from the centrifuged sample solution was then be added to row A of the plate, followed by the addition of 65 µL of 100% isopropanol. Next, this prepared plate was then run on the Magmax™ instrument for 20 minutes. During the run the magnetic beads bound the nucleic acids, the Wash Solution 1 Concentrate removed proteins, and other contaminants, while the Wash Solution 2 Concentrate removed residual binding solution. The nucleic acid was then eluted using the elution buffer. After the run, the sample yield from row F of the processed plate was extracted.

Following DNA extraction, forward and reverse *cpn*60 PCR primers were incorporated. These bacterial primers recognize the *cpn*60 universal target, which is a 552-558 base pair long region of *cpn*60 to be amplified. The amplified *cpn*60 universal target gene was further amplified and sequenced via 454 titanium pyrosequencing. After the sequence had been determined, the final step of *cpn*60 profiling was bioinformatic analysis. Through computational algorithms using the software GS Assembler (454 Life Sciences (Roche), Branford, CT), sequences reads were pooled together, and all like sequences were grouped together to determine the number of unique DNA sequences; the number of isotigs. Sequences were then cleaned, by removing any *cpn*60 PCR primers attached to the isotig sequence, using the Seqclean program. The cpnDB, a curated sequence database [124], was used where, based on >55% cut-off identity, isotigs were related to OTUs; their closest species neighbour. Any isotigs that matched the same species in the database, as another isotig, were removed to create a unique list of OTUs. Methods were in accordance with published protocols [47,136]; *cpn*60 amplicon generation conditions and primers can be seen in the Appendix E.

*cpn*60 based microbial profiling is unable to detect *Mycoplasma* and *Ureaplasma* species, as these lack the *cpn*60 gene target [45,47]. Thus, specific PCR assays were conducted in order to identify the presence of these species in our samples. Specifically, Mollicutes

(*Mycoplasma* or *Ureaplasma*) were detected by targeting the 16S rRNA gene using a conventional, semi-nested PCR [169], and *Ureaplasma* spp. were identified via specific PCR for the multiple-banded antigen gene [170].

Statistical Analyses

Power calculations were conducted, using the pwr package [171] in R, to determine the sample size required for an 80% power to detect differences between our recurrent BV and healthy-asymptomatic cohort (Vogue 1A). Power calculations used a two-tail test with an alpha error level of 0.05. This power calculation was formed on differences of the vaginal microbiota based on Gram stain analyses [172]. Our target sample size was 50 women; however we were only able to successfully enroll 26 women. We now have a 55% power to detect differences between our two populations. Power calculations were also conducted to determine what power existed to detect differences between our different clusters.

Our 26 women, clustered into six groups with respects to CSTs. For an 80% power to detect differences between these CSTs, enrolment of 51 women was needed in each CST. However, as indicated, recruitment proved to be challenging, and 26 women were successfully enrolled; where there was on average 4 participants in each CST. With our numbers, there was a 46.7% power to detect differences between these CSTs.

Our analysis plan included analyzing the bacterial profiles of our cohort of women with recurrent BV, relating profiles to clinical, behavioural and demographic data, and comparing bacterial profiles with earlier studies conducted in healthy-asymptomatic women.
To determine the bacterial diversity in our study population, we employed the Shannon Diversity Index (H):

$$H' = -\sum_{i=1}^{R} p_i \ln p_i$$

Where,

R = the total number of species in the community

p_i = the proportion of species *i*.

The Shannon Diversity Index speaks to the diversity of the dataset with regards to richness and evenness. Richness is the number of different species present in the population; a population increases in richness the more different species exist. Evenness is the number of individuals present in the given species type; a population increases in evenness the more equally abundant each species type is. The index value increases with richness and evenness; thus higher values are more diverse in comparison to lower values. The Shannon Diversity Index was calculated via the vegan package in R [173]. Furthermore, we analyzed Chao1 (species richness) and Pielou's evenness independently for our cohort through the BiodiversityR and vegan package in R, respectively [173,174]. Good's coverage estimator (C) was used to determine how well our cohort covered the total bacterial species present in the vaginal environment, and was calculated via the entropart package in R [175]. Good's coverage estimator was calculated by:

$C = (1 - (n_1/N))$

Where,

n_1 = the number of operational taxonomic units (OTUs) represented by one sequence

N = the total number of individuals in the sample

Furthermore, rarefaction curves were constructed to determine if sampling depth was sufficient. Sampling depth is deemed sufficient when rarefaction curves plateau reaching a slope of zero. The slope of each curve increases as more common species are found, and as the number of bacterial species discovered decreases the curve plateaus, indicating enough samples were collected to accurately characterize the community. Slopes of each curve were also calculated at five different points on the curve, total reads less 1, 5, 10, and 100 read(s), to determine whether the slope of the curve was decreasing and approaching zero. Rarefaction curves and slopes were completed via the vegan package in R [173].

In order to look at the descriptive characteristics we obtained, a Fisher's exact test and One-Way ANOVA were conducted to determine if significant associations existed between CSTs, and our demographic, behavioural and clinical variables. A two-independent sample t-test and correlation analyses were conducted to determine if significant associations existed between Shannon's Diversity Index and descriptive characteristics of our cohort. Due to our small sample size, p-values were not the ideal indicator of significant associations, thus effect sizes were also examined. Cramer's V, F-Statistic, Cohen's D and Pearson's R were calculated, and represented effect sizes for Fisher's exact test, one-way ANOVA, two-independent sample t-test, and correlation analyses, respectively. The stats, vcd and effsize packages were used to calculate effect sizes in R [176–178]. This allowed us to compare associations between CSTs and Shannon's Diversity Index with our variables, relatively and independent of sample size. To note, one woman fell into CST II, and thus was treated as an outlier and removed from our CSTs analyses, but this participant remained in subsequent analyses.

Variables which were deemed to have a large association to CSTs via effect size were further analyzed using a contingency table, to determine whether these associations were associated with healthy-CSTs or BV-associated CSTs. Variables which were deemed to have a

large association to Shannon's Diversity Index via effect size were further analyzed by determining the mean diversity of each group, and whether these associations were positive or negative.

Descriptive characteristics were grouped where appropriate. The contingency table was created using the stats package in R. Initially, univariate analyses were conducted, and as no variables were found to be statistically significant, multivariate analyses were not conducted.

Furthermore, the ALDEx2 package [179] was used to determine associations between bacterial species abundance, and our demographic, clinical and behavioural variables within our cohort. As these variables were not normally distributed, the nonparametric Kruskal-Wallis test was used. P-values were adjusted to control for false discovery rate through the Benjamini Hochberg procedure. Median relative abundances of bacterial species (centre-log-ratio transformed) were indicated for each group in our demographic, clinical and behavioural variables.

Cluster analyses were also conducted to identify how individual bacterial profiles grouped together, which allowed for the construction of CSTs. In order to conduct hierarchical clustering analyses, the Jensen-Shannon divergence index was first applied to determine the ecological distance between the bacterial species from different subject samples. This determined the compositional (dis)similarity between individual samples. Once the ecological distance was calculated, the stats and labdsv packages in R aided in hierarchical clustering analyses, and determined how our cohort clustered and which bacterial species dominated each cluster [176,180]

The Vogue 1A study, previously conducted by our group, recruited 312 women who were HIV negative, not pregnant, nor had recurrent BV. Demographic, clinical and behavioural variables were compared between this group and our group of women with recurrent BV, to determine whether these two populations were significantly different from each other. Bacterial diversity indexes between these two cohorts were also compared, in addition to cluster analyses. It was determined if any overlap existed between bacterial populations present in these cohorts, as well as the differences in bacterial diversity.

2.2 Virome Methods

Prior research was conducted by our team to analyze viruses within the vaginal environment of healthy-asymptomatic women, and HIV and BV-positive women. However, very few viral reads were detected in our populations; where on average 2.1 (0-18) viral reads were detected per sample, including a large abundance of bacterial species. The challenges from this previous study illustrated that in the future we needed to focus on a single variable; the virome; and process samples prior to nucleic acid extraction such to remove bacteria and human cells, therefore reducing nucleic acid competition. Proceeding with the vaginal virome study, our first step was to optimize the recovery of viruses.

Virome Protocol Optimization Methods

The sample processing protocol was first optimized to improve its sensitivity for the detection of viruses in the vaginal virome samples. Simulated samples consisting of A549 tissue culture cells, *Escherichia (E.) coli*, and human adenovirus and enterovirus were used to optimize the ratio of viral:bacterial:human nucleic acids, namely to increase the proportion of viral nucleic acids. The qPCR cycle threshold (Ct) values were used to quantitate the various components in the simulated samples prior to and after processing. The optimization experimental flow is depicted in Figure 2.1.

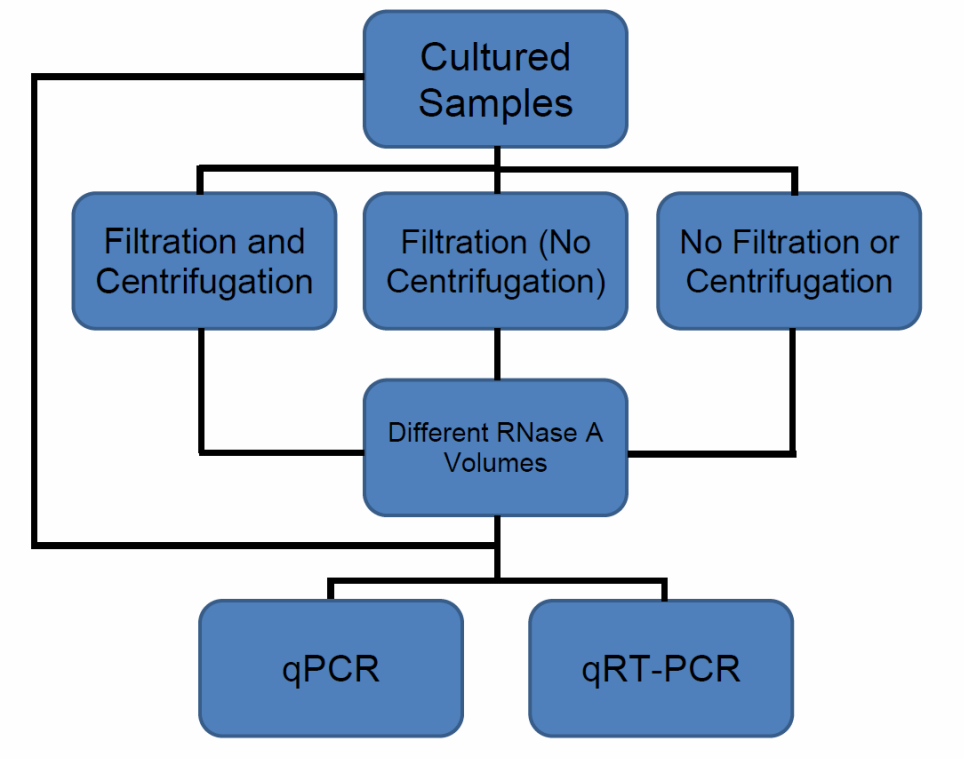

Figure 2.1 Optimization protocol workflow

Sample Processing

As seen in Figure 2.1, three methods were investigated for their efficiency at increasing the ratio of viral nucleic acids; centrifugation and filtration, filtration, and no centrifugation or filtration. Cultures were passed through each arm, where they were subjected to varying concentrations of ribonuclease A (RNase A) (2 ng/µl) (Life Technologies): 0 ng/µl, 0.01 ng/µl, and 0.02 ng/µl. RNase A removes the free RNAs present in our sample, but we were concerned that this enzyme could also degrade the viral RNAs we wanted to analyze. By testing different concentrations, we determined the optimal amount of RNase A to use in our experiment. Outlined below is the full sample processing protocol, as used in our 'Centrifugation and Filtration' arm.

Culture samples were vortexed vigorously for 30 seconds. Samples were centrifuged for five minutes at 15,000 x g. The supernatant was then filtered through a 0.45 µm filter (Fisher) to remove eukaryotic and bacterium-sized particles [181–183]. After filtration, samples were treated with DNases, and RNases to remove free DNA and RNA. One microliter of Turbo DNase I (2 U/µl) (Life Technologies) was added to samples along with 2 µl of 10X TURBO DNase Buffer (Life Technologies) and varying volumes of RNase A (2 ng/µl) (Life Technologies) [181–184]. These enzymes were incubated with samples for 30 minutes at 37˚C. Afterwards, 2 µl of DNase Inactivation Reagent (Life Technologies) was added to inactivate the DNase. Samples were then vortexed vigorously for 30 seconds, and centrifuged for 2 minutes at 15,000 x g to remove the inactivating agent, which was contained in the pellet. To inactivate RNase A, 1 µl of SUPERase• In™ RNase Inhibitor (Life Technologies) was added to the sample and incubated at 37˚C for 60 minutes [184,185]

The NucliSens® easyMag® (BioMerieux) extracted the viral, bacterial and human cell nucleic acids from the sample, as per the manufacturer's protocol. Cultured samples were placed in 2ml of lysis buffer and incubated at room temperature for at least 10 minutes. Afterwards, 100 µl of easyMag® magnetic silica beads and 100 µl of double-distilled water (ddH_2O) were added to the samples, and then samples were loaded onto the easyMag® extraction machine.

Once in the machine, the target nucleic acids were captured by the easyMag® magnetic silica beads. The silica beads were then attracted towards the NucliSens® easyMag™ magnetic device, which along with several washing steps, allowed for the purification of the nucleic acids. Heating steps released the nucleic acid from the silica beads, and finally the silica beads were separated from the eluate by the magnetic device (BioMerieux). Samples were eluted into 25 µl.

Polymerase Chain Reaction

Once the nucleic acids were extracted from the sample, the next step was amplification. Samples were used in two separate PCR reaction techniques, in order to amplify viral, bacterial and human targets; quantitative or real-time PCR (qPCR), and quantitative reverse transcription PCR (qRT-PCR). qPCR was able to quantify DNA targets, while qRT-PCR quantified RNA targets.

qPCR measures the amount of a DNA target present in a sample. Two qPCR assays were used, a Taqman® based assay and a SYBR® Green based assay (Thermo Fisher Scientific). In this Taqman® based qPCR assay a Taqman® probe was utilized. Attached to the 5' and 3' end of the Taqman probe were the fluorescent reporter dye and a quencher, respectively. When the probe was intact, the quencher prevents emission of the reporter dye. However, during the extension phase of PCR, the *Taq* DNA polymerase cleaves the Taqman® probe as it extends the primer. When the probe was no longer intact, and the quencher and fluorescent dye were separated, the report dye emits fluorescence, which was then detected to quantitatively analyze the PCR products. It was through the accumulated fluorescence signals we were able to quantify the PCR products [186]. This qPCR assay was used in our optimization experiments, to amplify *E.coli* nucleic acids targeting the *uidA* gene target, as well as amplify human nucleic acids targeting the *RNaseP* gene.

The second, SYBR® Green based, qPCR assay was used to amplify human adenovirus nucleic acids targeting the hexon gene, in our optimization experiments. This assay makes use of the SYBR® Green dye, which binds to double-stranded DNA. Once bound to double-stranded DNA, the dye emits fluorescence. After each round of PCR, more PCR products were created thus more double-stranded DNA was available for the SYBR® Green dye to bind on to.

Thus, as PCR processes round after round, the fluorescence intensity increased, and these accumulated fluorescence signals allow us to quantify the PCR products [186].

The reaction conditions for both assays can be seen in Appendix E, along with the primer and probe sets.

qRT-PCR quantifies RNA targets in a sample. Reverse transcriptase converts the RNA template into a complementary DNA (cDNA), which was then used as a template in a Taqman® based qPCR assay. Reverse transcription and qPCR reagents were added together in the same reaction tube for a one-step procedure. qRT-PCR was used to quantify human enterovirus in our optimization protocol by targeting the 5' *UTR* gene. The reaction conditions can be seen in Appendix E, along with the primer and probe sets.

Centrifugation and Filtration Spiking Experiments

The centrifugation and filtration pipeline, along with the use of 10 µl of RNase A (2 ng/µl) (Life Technologies) was determined to allow optimal viral recovery rates.

To further validate this protocol, 500 µl of vaginal samples were spiked with 100 µl of *E. coli*, and human adenovirus and enterovirus. Samples were processed as outlined above. Pellets from the initial centrifugation step were saved, and along with our processed spiked vaginal samples, nucleic acids were extracted and amplified; protocol outlined above. To support this protocol's ability to maintain the viral communities in our sample, yet reduce the bacterial and human cells present; bacterial and human cell DNA should be present in the pellet and viral nucleic acids should be evenly distributed between the pellet and sample.

Virome Methods for Clinical Samples

Experiment flow is depicted in Figure 2.2

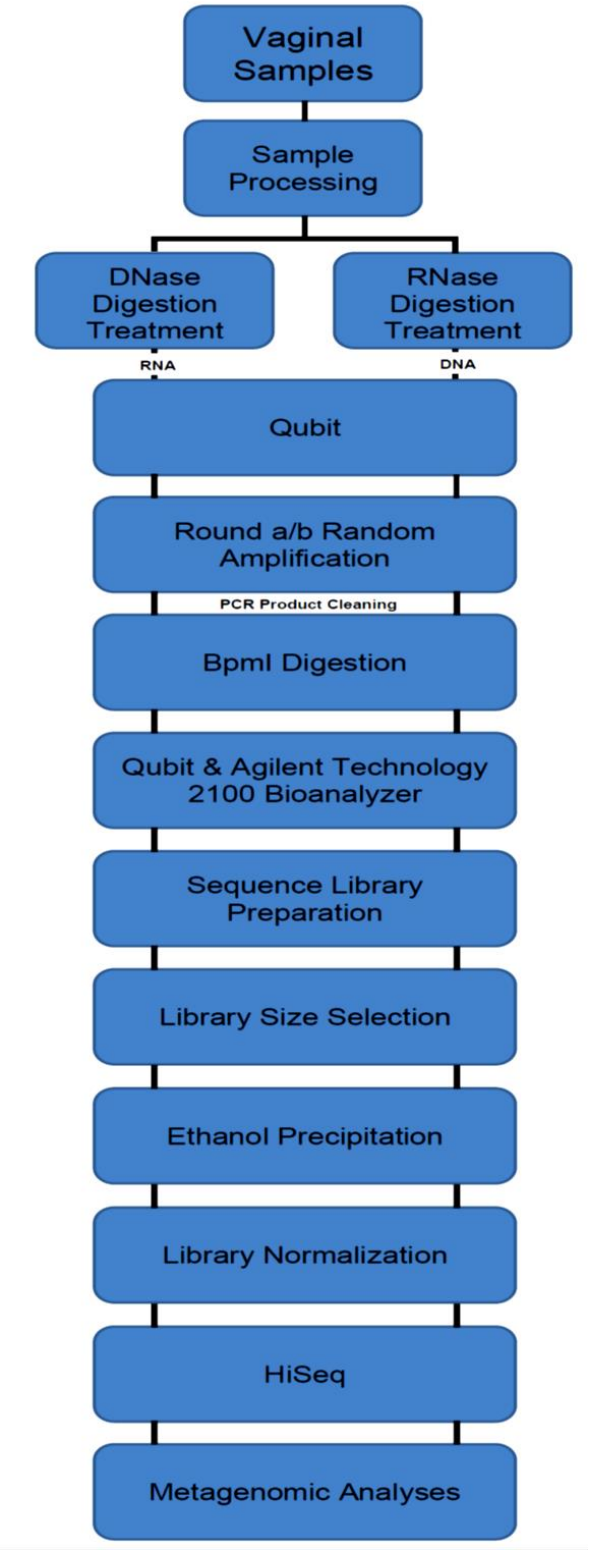

Figure 2.2 Experimental workflow for vaginal viral samples

Participant Characteristics

Vogue 1A. Metagenomic characterization of the vaginal microbiome of healthy, non-pregnant women.

This arm of the study sought to characterize the vaginal microbiome of 300 healthy-asymptomatic women. Women were deemed healthy and asymptomatic through clinical assessment by an OBGYN specialist or by a physician at a family practice or student health clinic. Those women fitting the inclusion criteria were invited to participate in this study. The inclusion criteria were sufficient comprehension of English in order to complete informed consent, non-pregnant, between the ages of 18-49, a regular menstrual cycle (~28 days). Women were not enrolled into this study if they met the exclusion criteria; lack of ability to provide written informed consent, or have used either systemic or topical antimicrobial therapy within the prior month. Our team enrolled 312 women into this arm of the study, and 62 viral vaginal samples are available for analysis. After removing women who had missing data, or were found to infringe on the antimicrobial use component of the inclusion criteria. 21 of these samples were chosen at random via the sample() function in R for virome analyses, along with their accompanying metadata.

Vogue 1B. Metagenomic characterization of the vaginal microbiome of HIV-positive women.

This study aimed to characterize the vaginal microbiome of 50 HIV-positive. Women were approached at the Oak Tree Clinic at BC Women's Hospital, and those who fit the study inclusion criteria were offered enrollment into the study. The inclusion criteria for this study were between 18 and 49 years of age, HIV positive, not menopausal and not currently pregnant. The exclusion criteria were inability to provide informed consent, currently pregnant, younger than 18 or older than 49, menopausal, and HIV negative. Our team enrolled 54 women in this arm of the study, and there were 54 viral vaginal swabs available for analysis. 25 of these samples were used for my viral analyses, along with their accompanying metadata. Samples were chosen to

reflect an even split between those who suppressed and unsuppressed; those with missing data were omitted.

Vogue 1B2. Metagenomic characterization of the vaginal microbiome of women with recurrent BV.

The goal of this study was to characterize the vaginal microbiome of 25 women with recurrent BV. Women were approached at the Infectious Disease Clinic at BC Women's Hospital, and various OBGYN clinics in the community. Additionally, study advertising was used to target women who were interested in partaking in this study. Women who fit the inclusion criteria of the study were enrolled. The inclusion criteria were: non pregnant, between the ages 18-49, not menopausal, HIV negative, and had recurrent BV which was defined as at least 4 episodes in 12 months with BV or BV-like symptoms. The exclusion criteria were: inability to provide informed consent, currently pregnant, younger than 18 or older than 49, menopausal, no history of recurrent BV, HIV positive. Our team enrolled 26 women in this arm of the study, and there were 8 viral vaginal swabs available for this cohort of women; all of which were analyzed.

Data Collection

Data collection for Vogue 1A and 1B2, can be seen above in the "Data Collection" bacteriome section. Data collection for 1B was also the same as Vogue 1B2. Additional information was asked, which included information pertaining to mode of HIV acquisition, duration of HIV infection, primary HIV positive test, HAART regimen, CD4 count and viral load at study visit, lowest CD4 count, highest viral load, the HIV clade, hepatitis B and C immune status, and antiretroviral medication history.

Sample Collection

As per Bacteriome Sample Collection with an additional sample collected for viral analysis. A Copan 3mL Universal Transport Medium Kits for the Collection and Preservation of Virus, *Chlamydia spp*., *Mycoplasma spp*., and *Ureaplasma spp* was used to collect the viral sample and it was also de-identified and stored in a -80˚ Celsius freezer within 4 hours of swab collection at the WHRI clinic

Sample Processing

Viral vaginal samples were vortexed vigorously for 30 seconds. 1 mL was extracted for further analysis and the rest stored at -80˚C. Samples were centrifuged for five minutes at 15,000 x g. The supernatant was then filtered through a 0.45 µm filter (Fisher) to remove eukaryotic and bacterium-sized particles [181–183]. After filtration, samples were treated with DNases, and RNases to remove free DNA and RNA. One microliter of Turbo DNase I (2 U/µl) (Life Technologies) were added to samples along with 2 µl of 10X TURBO DNase Buffer (Life Technologies) and 10 µl of RNase A (2 ng/µl) (Life Technologies) [181–184]. These enzymes were incubated with samples for 30 minutes at 37˚C. Afterwards, 2 µl of DNase Inactivation Reagent (Life Technologies) was added to inactivate the DNase. Samples were then vortexed vigorously for 30 seconds, and centrifuged for 2 minutes at 15,000 x g to remove the inactivating agent, which was contained in the pellet. To inactivate RNase A, 1 µl of SUPERase• In™ RNase Inhibitor (Life Technologies) was added to the sample and incubated at 37˚C for 60 minutes [184,185].

The NucliSens® easyMag® (BioMerieux) extracted the viral nucleic acids from the sample, as per the manufacturer's protocol and mentioned above.

DNase and RNase Digestion Treatment

Samples were split evenly into two microcentrifuge tubes, where one tube was treated with DNase to remove free DNA and analyze RNA viruses, and the other was treated with RNase to remove free RNA and analyze DNA viruses. For DNA digestion, 0.56 µl of Turbo DNase I (2 U/µl) (Life Technologies) was added to samples along with 1.28 µl of 10X TURBO DNase Buffer (Life Technologies). For RNA digestion, 1.25 µl of RNase A (2 ng/µl) (Life Technologies) was added to samples [181–184]. These enzymes were incubated with samples for 30 minutes at 37˚C. Afterwards, for DNA digestion, 1.28 µl of DNase Inactivation Reagent (Life Technologies) was added to inactivate the DNase. Samples were incubated at room temperature for 5 minutes, then vortexed vigorously for 30 seconds, and centrifuged for 2 minutes at 15,000 x g to remove the inactivating agent, which was contained in the pellet. The supernatant was transferred to a fresh tube. For RNA digestion, to inactivate RNase A, 1.25 µl of SUPERase• In™ RNase Inhibitor (0.2 U/µl) (Life Technologies) was added to the sample and incubated at 37˚C for 60 minutes [184,185].

Qubit

Each sample was analyzed quantitatively through the Quant-iTTM high-sensitivity DNA assay kit using a Qubit® fluorometer (Invitrogen). Fluorescent dyes will bind the single-stranded RNA molecules and intercalate between double-stranded DNA molecules. When bound to their targets, these dyes will emit a detectable fluorescence which is proportional to the amount of nucleic acids in our sample. This step allowed us to quantitatively analyze the nucleic acids in our sample post-Round a/b Random Amplification, prior to Library Preparation, post-Size Selection and prior to sequencing via HiSeq. A step-by-step protocol can be seen in Appendix F.

Round a/b Random Amplification Protocol

Through Qubit analysis, we determined that the amount of nucleic acids present in our samples was not sufficient enough to proceed to NEBNext® Library Preparation (New England Biolabs Inc.) (Qubit reading: Out of Range “Too Low” <0.50 ng/mL). Round a/b Random Amplification was utilized to increase the number of copies in our samples; validated by Qubit analyses after cleanup of PCR products as mentioned below. A step-by-step protocol can be seen in Appendix G [187].

PCR Product Cleaning

After Round a/b random amplification, PCR products were cleaned using Agencourt AMPure XP beads (Beckman Coulter) to purify PCR amplicons. This cleaning step also followed post-BpmI digestion. A step-by-step protocol can be seen in Appendix H.

BpmI Digestion

The BpmI enzyme (New England Biolabs Inc.) was used to digest primer sequences used in Round a/b Random Amplification, removing ~20bp from each end of the sequence and preventing the loss of sequencing information by sequencing known primer sequences. As mentioned above, products were cleaned via AMPure XP beads, and Qubit quantitatively analyzed the nucleic acids in our sample post-BpmI digestion. A step-by-step protocol can be seen in the Appendix I.

Agilent Technology 2100 Bioanalyzer

In order to assess the quality of our samples, the Agilent Technology 2100 Bioanalyzer was used to quantitatively and qualitatively analyze fragments pre- and post-BpmI digestion, post-library preparation and post-Coastal Genomics size selection. If BpmI digestion was successful, post-digestion samples should have been ~40bp shorter than pre-digestion samples. Samples were analyzed post-library preparation and post-Coastal Genomics size

selection to determine the presence of adapter dimers, as well determine if size selection was successful. A narrow curve should appear if size selection was successful. A step-by-step procedure is shown in Appendix K [188].

Sequence Library Preparation

The NEBNext Ultra DNA library preparation (Illumina) pipeline was as follows: end repair, adapter ligation, PCR amplification, PCR clean-up, and validation of the library. Briefly, our fragmented DNA samples were incorporated with End Prep Enzyme Mix and End Repair Reaction Buffer (10X) (NEB) resulting in repaired DNA with 5' phosphorylated dA-tailed ends. Adapters were incorporated and bound to each end of our of DNA sample, this was followed by incorporation of the USER™ enzyme and size selection using AMPure XP beads, selecting for fragments between 400-600bp. This allowed for precise selection of the DNA library such that sequences without adapter sequences, and impurities were removed. Next samples were incorporated with different combinations of Index Primer/i7 Primer (i701-i712) and Universal PCR Primer/i5 Primer (i501-i508), such that each sample would have a unique barcode. Samples were once again cleaned using AMPure XP beads. Qubit quantitatively analyzed the nucleic acids in our samples post-library preparation, and Bioanalyzer analyses validated our libraries. A step-by-step procedure is shown in Appendix J [188].

Library Size Selection

Through BioAnalyzer analyses, an abundance of adapter dimers were detected within samples at the 140 bp mark, and size selection aimed to remedy this problem. To test the size selection protocol, two samples were run on a 2% low-melting point agarose gel for 1.5 hours at 120V, and DNA was excised at 300-500bp. DNA was extracted from the agarose gel using the QIAquick Gel Extraction Kit (Qiagen). Qubit quantitatively analyze the nucleic acids in our sample, and the BioAnalyzer analyses detected an abundance of adapter dimers still present within our samples. These two samples were run through this protocol again, along with three

additional samples. These five samples were run on a 2% low-melting point agarose gel for 3 hours at 80V. The two samples from the previous run were excised again at 300-500bp, and the remaining three samples were excised at 400-600bp. Qubit analyses showed a drop in DNA concentration with double size selection, and BioAnalyzer analyses detected an abundance of adapter dimers still present within our samples excised at 400-600bp. Several different study arms were designed, using varying adapter ratios, AMPure XP bead ratios, and number of clean ups, with and without gel size selection, to examine the resulting effect on adapter dimer abundance and DNA concentration. Step-by-step procedures are shown in Appendix L.

Due to difficulties in size selection using the above methods, the samples were sent to Coastal Genomics (Burnaby, BC Canada) who through their Ranger Technology can automate this process of agarose gel loading, electrophoretic analysis, and recovery of targeted DNA fragment. Prior to sending samples to Coastal Genomics, samples were set to a standardized volume of 25 ul, and when necessary were diluted such that there was no more than 2 ug of DNA. Coastal Genomics selected DNA fragments within the size range of 300-500 bp.

Ethanol Precipitation

Samples from Coastal Genomics were eluted in a saline buffer at 250 ul, and required an ethanol precipitation to extract DNA. A step-by-step procedure is shown in Appendix M. Qubit quantitatively analyzed the nucleic acids in our sample, and the BioAnalyzer quantitatively and qualitatively analyzed the distribution of our DNA fragments.

Library Normalization

Each sample was diluted in 10 mM Tris pH 8.5, such that an equimolar solution (~ 6nM) of the DNA/cDNA library could be pooled. DNA and RNA samples were combined into two separate pools, and sent to Canada's Michael Smith Genome Sciences Centre (Vancouver, BC Canada) for sequencing.

HiSeq

Each pool was loaded onto a separate 125 base indexed PET HiSeq 2500 lane (Illumina). Samples were sequenced in accordance to the standardized manufacturer's protocol, which generated sequences at a maximum length of 125bp. The HiSeq machine also performed quality filtration, by filtering reads that contained greater than 5 nucleotides with quality scores below 15.

Metagenomic Analyses

Sequences were referenced automatically through Taxonomer [189], and manually through the National Center for Biotechnology Information (NCBI) database, in order to be related to their closest-species neighbour and viral family. Briefly, adapter and primer sequences were trimmed, and sequences were filtered for quality removing any sequences which were too short, had too many N's, and too many repeats, among other abnormal aspects. Sequences found within our negative controls were omitted. Forward and reverse reads were then merged, and matched to their closest-species neighbour, and viral family. Subsequently, classifications from our three cohorts were pooled together, a biologically relevant frequency threshold was placed to omit taxons which low representation. Taxons represented by ≤ 2 reads were removed.

Statistical Analyses

Power calculations were conducted, using the pwr package [171] in R, to determine what power existed to detect differences between our viral groups. Based on a two-tail test with an alpha error level of 0.05, we had 64% power to detect differences between our viral groups.

Our statistical analyses were similar to the analyses completed for the vaginal bacteriome. To determine the viral diversity in our three cohorts, we employed the Shannon Diversity Index, as well as analyzing Chao1 (species richness) and Pielou's evenness. Good's coverage estimator, rarefaction curves and slope values were used to determine how well our

cohorts covered the total viral species present in the vaginal environment. See Bacteriome Methods Section for details.

Cluster analyses were conducted to identify if individual viral profiles clustered together, in a similar manner to the bacteriome, and could allow for the construction of viral groups. These viral groups could then be related to demographic, behavioural and clinical variables, using Fisher's exact test and One-Way ANOVA. Furthermore, the ALDEx2 package [179] was used to determine associations between viral species abundance, and our demographic, clinical and behavioural variables within our cohort. See Bacteriome Methods Section for details.

Demographic, clinical and behavioural variables were also compared between of three cohorts to determine if these populations were significantly different from each other. Viral diversity indexes and viral profiles of the three cohorts were also compared, to determine if any overlap existed between viral populations present in these cohorts, as well as the differences in viral diversity.

3 Results

3.1 Bacteriome Results

Demographic Characteristics

Sixty-one women were approached who were attending for gynecologic care for BV-like symptoms between June 2012 and May 2015. Twenty women were excluded due to lack of symptoms on the day of sampling or with Nugent's scores inconsistent for BV, ten women were lost to follow-up and five women withdrew. The remaining 26 women with BV-like symptoms had a mean age of 32 years (range 20 – 47 years) (Table 3.1). Self-reported ethnicity was, 16 Caucasian, 7 Asian, 1 Indigenous origins, 1 Arab origins and 1 of mixed ethnicities. The highest education level attained of our study population was, one did not complete high school, two completed their high school diploma, seven received some post-secondary education, 14 completed post-secondary education, and two completed their graduate degree. The average BMI of this group was 22.7 kg/m^2 (range 15.5 - 31.2 kg/m^2) (Table 3.1).

Table 3.1 Demographic characteristics. Continuous variables are reported as mean ± SD, range), and categorical variables are reported as n

Demographics	
Age (Mean ± SD, Range) (years)	32 ± 7.9 (20 – 47)
BMI (Mean ± SD, Range) (kg/m^2)	22.7 ± 4.0 (15.5 – 31.2)
Ethnicity	Caucasian: 16 Asian: 7 Aboriginal origins: 1 Other: 2
Highest Education Level Attained	Graduate degree: 2 Post-secondary education: 14 Some post-secondary education: 7 High school diploma: 2 Did not complete high school: 1

Clinical Characteristics

Antimicrobial use was reported by the majority of women (18/26) in the past 3 months, and six women did not take any antimicrobials in the past 3 months; two women had missing antimicrobial use data. Of the 18 women using antimicrobials in the past 3 months, most were taking antimicrobials for BV treatment (13/18) (Table 3.2).

Since their last menstrual period, women in our cohort reported using various types of contraception. Seven women reported using no form of contraception, eight women used hormonal contraception, five women used barrier methods, three women had surgical sterilization, and two women were not sexually active; one woman had missing data regarding contraceptive use. In the 48 hours prior to their study visit, three women used condoms whereas 23 women did not (Table 3.2). Approximately half the women in our cohort were nulliparous (14/26), and ten women were multiparous. The remaining two women in our cohort had single pregnancies (Table 3.2).

Table 3.2 Clinical characteristics. Categorical variables are reported as n. Antimicrobial data was missing for two women, and data regarding contraceptive use was missing for one woman

Clinical Characteristics	
Pregnancy History	Nulliparous: 14 Multiparous: 10
Form of Contraception	None: 10 Not sexually active: 2 Hormonal: 8 Barrier: 5
Use of Condoms (in past 48 hours)	Yes: 3 No: 23
Symptoms	Abnormal discharge: 15 Abnormal odor: 6 Irritation or discomfort: 17 Painful intercourse: 14 Other: 8
Antimicrobial Use (in the past three months)	Yes: 18 No: 6

Almost all of the women in our cohort had been previously diagnosed with a genital infection (25/26). Reported lifetime histories of documented genital conditions included, 18 reports of a history of BV, 23 reports of a history of yeast, 16 reports of urinary tract infections (UTIs), one *Trichomonas vaginalis* infection, four reports of genital warts, four cases of genital herpes, and five cases of *Chlamydia trachomatis* infection. None of the women in our cohort reported a history of gonorrhea, or syphilis infections. Women had experienced an average of 4.7 BV infections in their lifetime (range 0 - 20), 2.3 BV infections in the last year (range 0 -12), and 0.6 BV infections in the past two months (range 0 - 2). Women also self-reported 11.5 'yeast' infections in their lifetime (range 0 - 40), 5.5 'yeast' infections in the last year (range 0 - 40), and 0.7 'yeast' infections in the past two months (range 0 - 3)**.** In addition, women reported 3.1 UTI infections in their lifetime (range 0 - 20), with 0.9 UTI infections in the past year (range 0 - 6) and 0.2 UTI infections in the past two months (range 0 - 3) (Table 3.3).

Table 3.3 Genital infection history. Continuous variables are reported as mean ± SD, range), and categorical variables are reported as n

Genital Infection History	
Bacterial Vaginosis (Mean ± SD, Range)	
Lifetime	4.7 ± 6.1 (0 – 20)
In the past year	2.3 ± 3.1 (0 – 12)
In the past two months	0.6 ± 0.8 (0 – 2)
Yeast	23/26
Urinary Tract Infections	16/26
Trichomoniasis	1/26
Genital Warts	4/26
Genital Herpes	4/26
Chlamydia	5/26
Gonorrhea	0/26
Syphilis	0/26

All women experienced BV-associated symptoms 48 hours prior to the study visit; 15 women experienced abnormal discharge, six abnormal odor, 17 irritation or discomfort, and eight women experienced other abnormal vaginal symptoms. Over 50% of women experience painful intercourse (14/26), and on average women experienced painful intercourse 86% of the

time (range 20 - 100%) (Table 3.2). Of the 26 women, seven had Nugent scores consistent with BV, 13 with intermediate BV, and six inconsistent with BV (Table 3.4).

Table 3.4 Nugent scores. Categorical variables are reported as n

Nugent Scores	
Inconsistent for BV (0 – 3)	6/26
Intermediate BV (4 – 6)	13/26
Consistent for BV (7 – 10)	7/26

Behavioural Characteristics

Tampon use was reported in 20 women, and six women reported never using tampons during their menstrual periods. Tampon use in the past month was reported in 15 women, whereas 11 women did not report tampon use in the past month. With regards to feminine hygiene, five women used feminine wipes or genital deodorant products, and four of these women used these products in the past 48 hours. One woman used douche products, and had douched in the past 48 hours (Table 3.5).

In the past year, women reported 1.2 sexual partners (range 0 - 4), and in the past two months women reported one sexual partner (range 0 -3). With regards to sexual activity, six women reported having had vaginal intercourse in the past 48 hours. Of our 26 women, the majority reported receiving oral sex (21/26), and few reported never receiving oral sex (4/26). Two women reported receiving oral sex in the past 48 hours. Seven women reported engaging in anal sex, and 18 reported never having engaged in anal sex. None of the women in our population had anal sex in the past 48 hours. Fifteen women reported sex toy use, and ten have never used sex toys. None of the women in our population used sex toys in the past 48 hours. Sexual activity data was missing for one woman (Table 3.5).

Reported illicit substance/alcohol use included four women never using illicit substances or consuming alcohol, and 21 women currently using illicit substances or consuming alcohol. Of the 26 women, most had never smoked (22/26), and three women were current smokers. Data on substance use was missing for one woman (Table 3.5).

Table 3.5 Behavioural characteristics. Continuous variables are reported as mean ± SD, range), and categorical variables are reported as n. Sexual activity and substance use data were missing for one woman.

Behavioural Characteristics	
Tampon Use (ever)	Yes: 20 No: 6
Tampon Use (in the past month)	Yes: 15 No: 11
Feminine Hygiene Products Use (ever)	Yes: 5 No: 21
Use of Feminine Hygiene Products (past 48 hours)	Yes: 4 No: 22
Substance Use	Yes: 21 No: 4
Current Smoker	Yes: 3 No: 22
Number of Sexual Partners (past year) (Mean ± SD, Range)	1.2 ± 0.9 (0 – 4)
Vaginal Intercourse (past 48 hours)	Yes: 6 No: 19
Oral Sex	Yes: 21 No: 4
Anal Intercourse	Yes: 7 No: 18
Use of Sex Toys	Yes: 15 No: 10

Bacteriome Data

A total of 742 isotigs were assembled from 127,768 reads, and based on a >55% cut-off identity to the *cpn*60 gene target; a total of 122 unique *cpn*60 OTUs were identified. Four OTUs were detected in all samples, corresponding to *L. crispatus, L. iners, L. jensenii,* and *G. vaginalis* Group *A*. Four OTUs were detected in almost all samples, with the exception of one or

two individuals, these corresponded to *G. vaginalis* Group C*, Clostridia sp*. (BVAB2)*, L. gasseri* and *Prevotella timonensis* (Figure 3.1).

Among women with intermediate (4-6) and consistent for BV (7-10) Nugent scores, 13 profiles were dominated by *G. vaginalis, Lactobacillus* species or both genera. Additionally, seven profiles were dominated by *Escherichia coli.*, *Megasphaera sp.*, *Actinobacteria sp.*, *Prevotella amnii*, *Prevotella timonensis*, *Atopobium vaginae*, *Clostridia sp.* BVAB2 and BVAB3, other *Streptococcus* and other *Prevotella* species, as shown in Figure 4. Furthermore, six women with a Nugent score inconsistent with BV, had profiles dominated by *L. gasseri*, *Klebsiella pneumoniae*, *L. crispatus*, *L. iners* or *G. vaginalis* Group A (Figure 3.1).

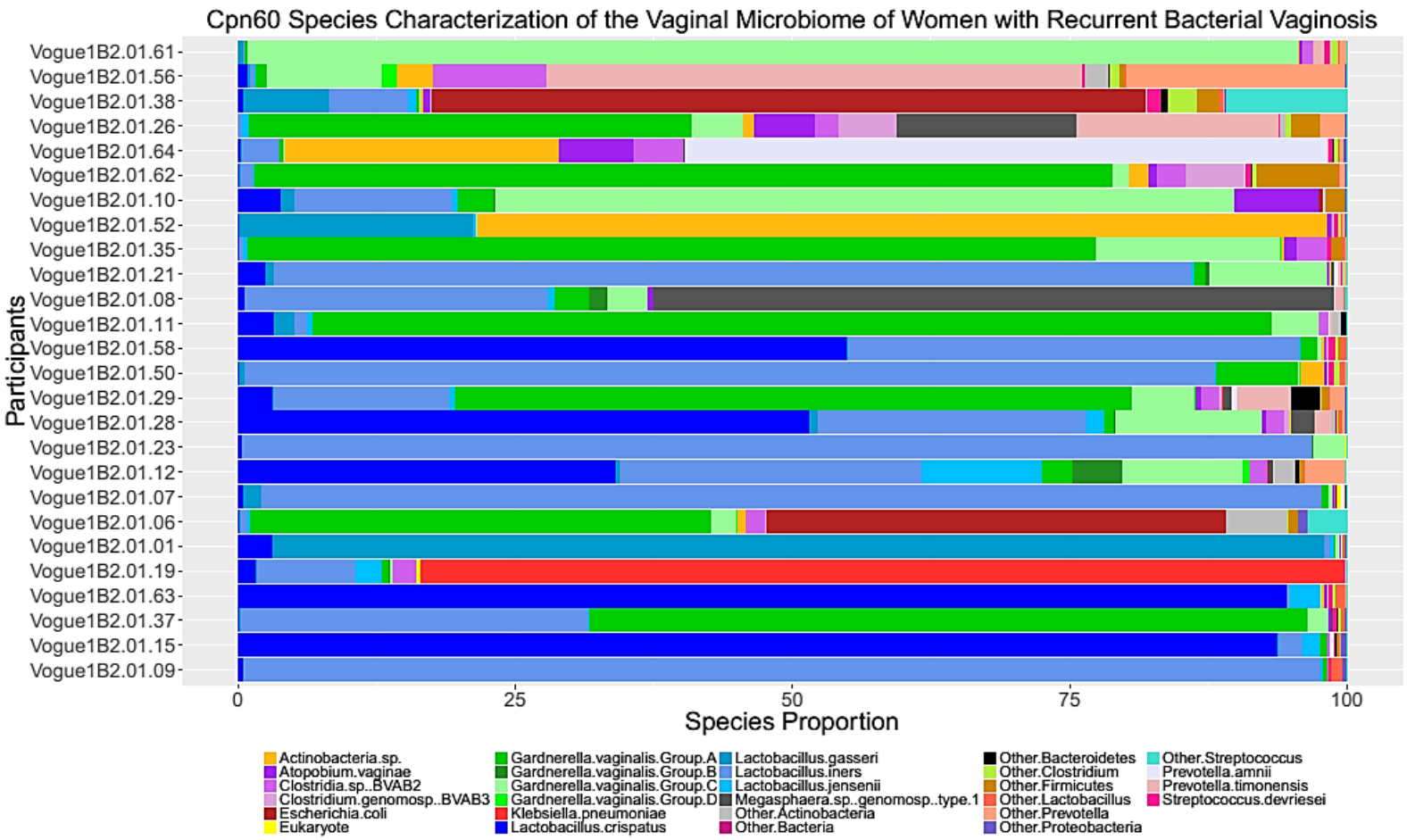

Figure 3.1 Vaginal bacteriome profiles ordered by Nugent score. Nugent scores inconsistent for BV present at the bottom, and Nugent scores consistent for BV present at top; Nugent scores in descending order

Hierarchical clustering showed that our participants clustered into six groups. Three groups were dominated by *Lactobacillus* species. Group one was dominated by *L. crispatus* (CSTI: 5/26), Group two was dominated by *L. gasseri* (CST II: 1/26), and Group three was dominated by *L. iners* (CST III:6/26). The fourth group was a mixed dominance group, where profiles were dominated by *G. vaginalis* Group B*, Prevotella amnii, Shigella boydii, Bifidobacterium breve, Bifidobacterium dentium*, and *L. delbrueckii* (CST IVA: 6/26). The fifth and sixth clusters were dominated by *G. vaginalis* species. Group five was dominated by *G. vaginalis* Group A, and *Megasphaera micronuciformis* (CST IVC: 6/26). And Group six was

dominated by *G. vaginalis* Group C (CST IVD: 2/26). Additional novel CSTs were identified in addition to CSTs previous studies have shown [39,40] (Table 3.6, Figure 3.2).

Table 3.6 Community state types. Categorical variables are reported as n

Community State Types		
CST	**Dominant OTU**	**N**
I	*L. crispatus* dominated	**5**
II	*L. gasseri* dominated	**1**
III	*L. iners* dominated	**6**
IVA	Mixed Dominant Phenotype: *Gardnerella* subgroup B, *Bifidobacterium* sp., *Prevotella amnii, Shigella boydii, and L.delbrueckii*	**6**
IVC	*Gardnerella* A dominated	**6**
IVD	*Gardnerella* C dominated	**2**

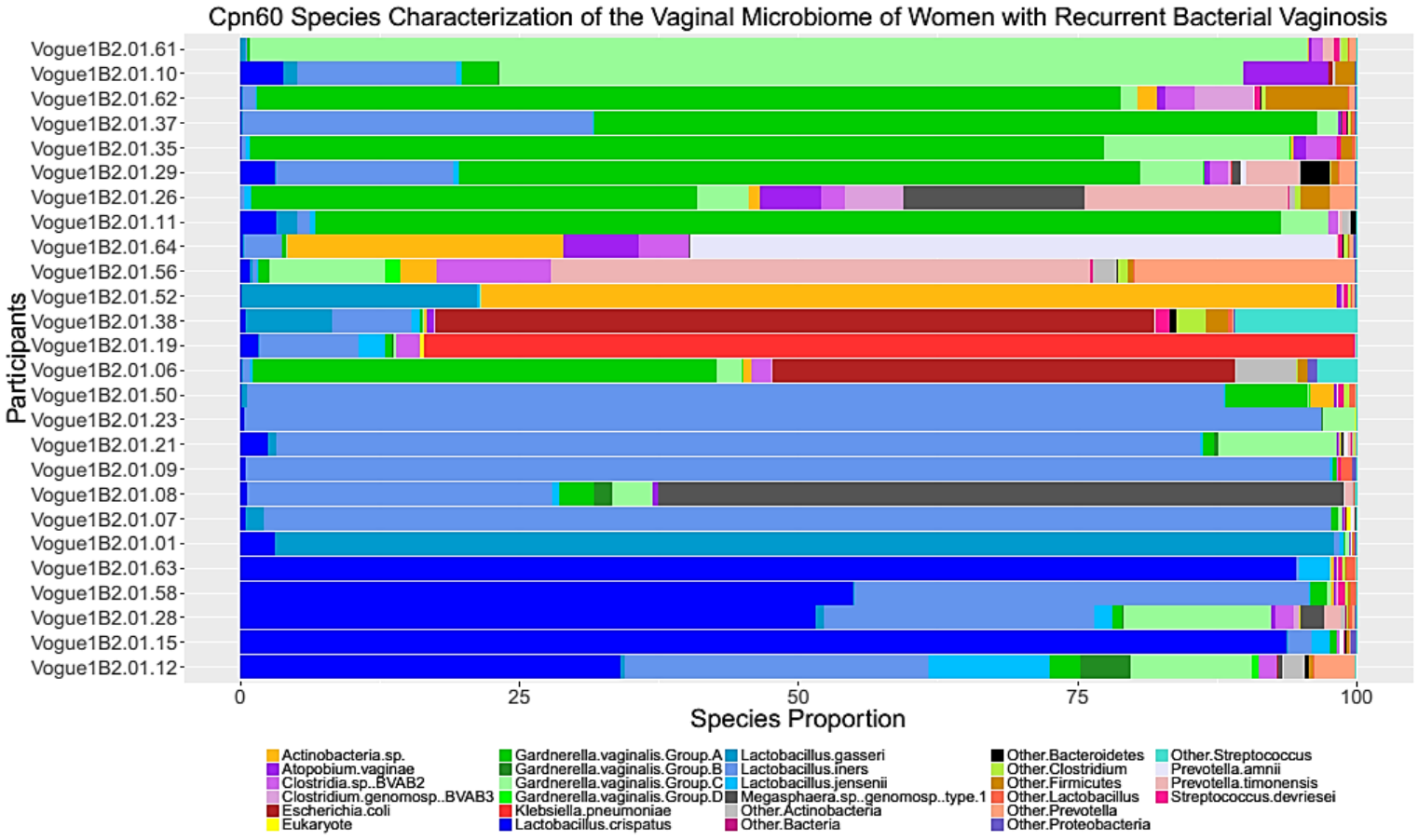

Figure 3.2 Vaginal bacteriome profiles ordered by CSTs. CST I present at bottom and CSTIVD present at top

Mollicutes (*Mycoplasma* and *Ureaplasma*)

Results of the Mollicutes and *Ureaplasma* specific PCR assays showed six samples were negative for Mollicutes, and 18 were positive for Mollicutes. Of these six negative samples, five were negative for *Ureaplasma* species, and one was positive for *Ureaplasma* (*U.*) *urealyticum*. *Ureaplasma* are mollicutes, yet one participant was negative for Mollicutes and positive for *Ureaplasma*. This indicates there was a failure of the Mollicutes specific PCR either technical or due to sequence variation, or Mollicutes were present in this participant below the detection threshold. Overall, Mollicutes PCR is not as robust as the *Ureaplasma* assay. Of the 18 positive Mollicutes samples, 14 were positive for *U. parvum*, and four were positive for *U.urealyticum*. We could not conclude whether these positive Mollicutes samples also contained *Mycoplasma* species. On average, there were 4.12×10^8 16S rRNA gene copies per swab in our negative Mollicutes samples (range 2.15×10^7 - 9.30×10^8), and 4.05×10^9 16S rRNA gene copies per swab in our positive Mollicutes samples (range 2.77×10^7 - 3.75×10^{10}) (Table 3.7). Detection of Mollicutes and *Ureaplasma* species differed across CSTs (Figure 3.3, 3.4).

Table 3.7 Mollicutes and *Ureaplasma* specific PCR assays results. * Participant 07 had a negative Mollicutes result, but was positive for *U. urealyticum*

Participants	Read Counts	16S rRNA gene copies per swab	Mollicutes	*Ureaplasma*
Vogue1B2.01.01	2175	4.67E+07	Positive	*Urealyticum*
Vogue1B2.01.06	1914	8.66E+07	Positive	*Urealyticum*
Vogue1B2.01.07	2226	1.20E+08	Negative	*Urealyticum**
Vogue1B2.01.08	18132	1.16E+09	Positive	*Urealyticum*
Vogue1B2.01.09	5710	2.77E+07	Positive	*Parvum*
Vogue1B2.01.10	1879	1.76E+09	Positive	*Parvum*
Vogue1B2.01.11	2454	6.23E+09	Positive	*Parvum*
Vogue1B2.01.12	2195	3.04E+08	Positive	*Parvum*
Vogue1B2.01.15	1602	2.15E+07	Negative	*Negative*
Vogue1B2.01.19	1096	6.29E+08	Negative	*Negative*
Vogue1B2.01.21	8065	1.80E+08	Positive	*Parvum*
Vogue1B2.01.23	6913	2.20E+09	Positive	*Parvum*
Vogue1B2.01.26	3868	3.75E+10	Positive	*Parvum*
Vogue1B2.01.28	1272	NA	NA	NA
Vogue1B2.01.29	2208	NA	NA	NA
Vogue1B2.01.35	3043	1.13E+09	Positive	*Parvum*
Vogue1B2.01.37	6091	2.05E+09	Positive	*Parvum*
Vogue1B2.01.38	4336	6.63E+08	Negative	*Negative*
Vogue1B2.01.50	12681	2.62E+09	Positive	*Parvum*
Vogue1B2.01.52	12966	4.94E+09	Positive	*Parvum*
Vogue1B2.01.56	5091	8.38E+09	Positive	*Urealyticum*
Vogue1B2.01.58	7460	9.17E+08	Positive	*Parvum*
Vogue1B2.01.61	4155	9.30E+08	Negative	*Negative*
Vogue1B2.01.62	7801	3.22E+09	Positive	*Parvum*
Vogue1B2.01.63	5605	1.08E+08	Negative	*Negative*
Vogue1B2.01.64	6956	1.81E+08	Positive	*Parvum*

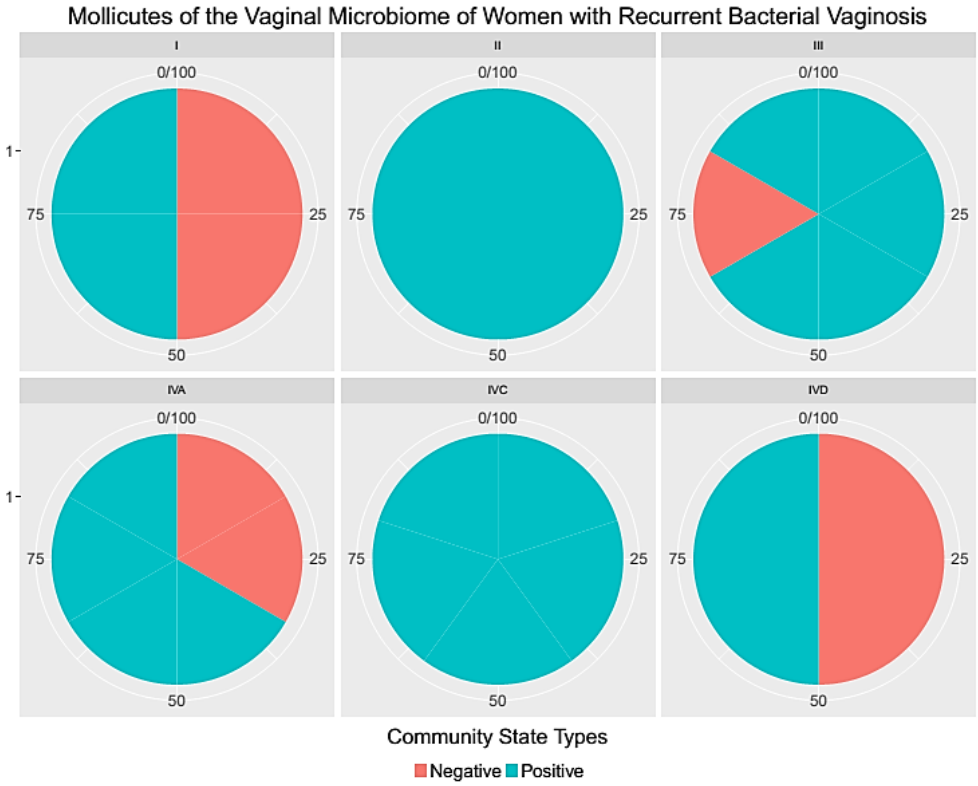

Figure 3.3 Mollicutes specific PCR results for each CST

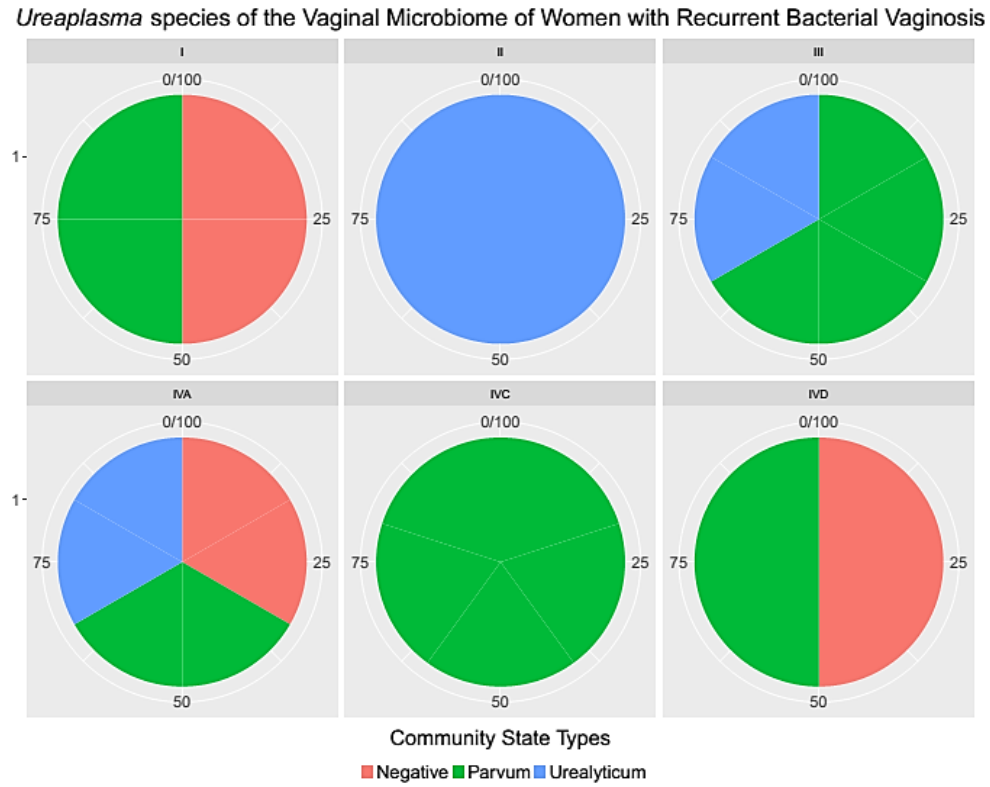

Figure 3.4 *Ureaplasma* species specific PCR results for each CST

Diversity Statistics

To determine how well our sample population covered the total bacterial species present in the vaginal environment, Good's coverage was employed (Table 3.8). In addition, rarefaction curves illustrated whether sampling depth was sufficient. Slope values for each curve were calculated to determine whether sampling depth was sufficient, or whether a large fraction of diversity remained to be discovered (Figure 3.5, Table 3.9). Furthermore, to determine bacterial diversity within our population of women, Shannon's Diversity Index, Species Richness, Pielou's evenness and Chao1, were calculated for each participant (Table 3.8).

Table 3.8 Diversity statistics. Diversity statistics for each participant.

Diversity Statistics				
Participants	**Good's Coverage**	**Shannon's Diversity**	**Species Richness**	**Pielou's Evenness**
Vogue1B2.01.01	0.9975	0.28	8.09	0.11
Vogue1B2.01.06	0.9989	1.36	12.68	0.50
Vogue1B2.01.07	0.9978	0.27	9.83	0.11
Vogue1B2.01.08	0.9999	1.11	11.26	0.36
Vogue1B2.01.09	0.9996	0.19	7.96	0.07
Vogue1B2.01.10	0.9983	1.19	10.92	0.45
Vogue1B2.01.11	0.9996	0.65	10.00	0.27
Vogue1B2.01.12	0.9995	1.85	15.48	0.64
Vogue1B2.01.15	0.9984	0.36	10.09	0.14
Vogue1B2.01.19	0.9964	0.70	10.00	0.30
Vogue1B2.01.21	0.9997	0.72	12.43	0.24
Vogue1B2.01.23	0.9996	0.18	4.28	0.07
Vogue1B2.01.26	0.9992	1.86	15.36	0.62
Vogue1B2.01.28	0.9969	1.45	16.90	0.49
Vogue1B2.01.29	0.9985	1.43	15.52	0.49
Vogue1B2.01.35	0.9992	0.82	11.74	0.30
Vogue1B2.01.37	0.9994	0.84	9.55	0.30
Vogue1B2.01.38	0.9989	1.35	15.51	0.45
Vogue1B2.01.50	0.9997	0.56	10.59	0.19
Vogue1B2.01.52	0.9998	0.67	9.67	0.22
Vogue1B2.01.56	1.0000	1.64	15.96	0.54
Vogue1B2.01.58	0.9997	0.93	11.47	0.32
Vogue1B2.01.61	0.9995	0.33	10.54	0.12
Vogue1B2.01.62	0.9999	0.99	14.15	0.34
Vogue1B2.01.63	0.9996	0.30	9.20	0.11
Vogue1B2.01.64	0.9995	1.26	12.78	0.41

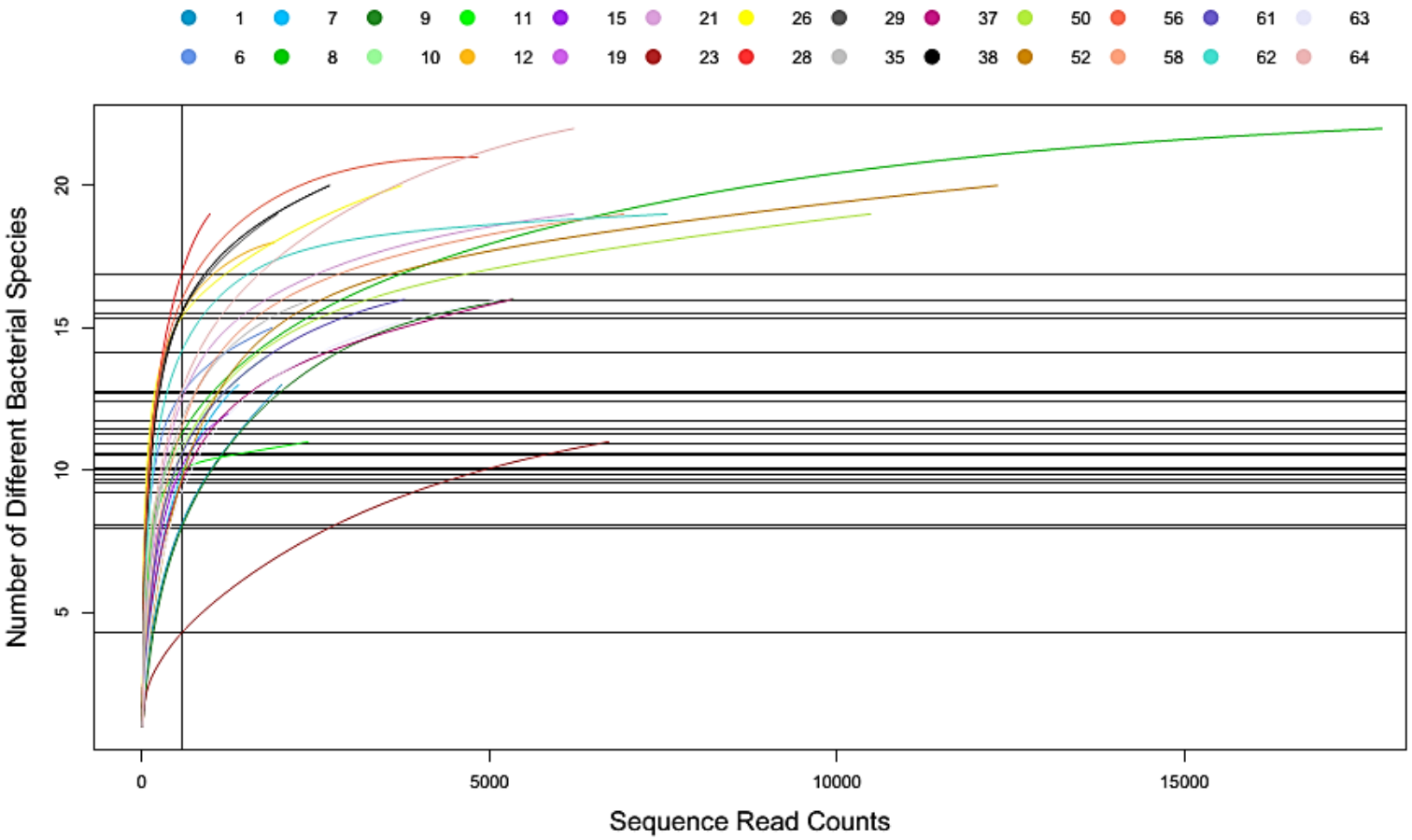

Figure 3.5 Rarefaction curves. Each rarefaction curve represents a participant

Table 3.9 Rarefaction curve slopes. Slope values for each rarefaction curve. Total Reads indicate the end point of the rarefaction curve. Each column represents the slope of the rarefaction curve, at Total Reads – indicated number of reads

Rarefaction Curve Slopes				
Participants	**Total Reads$_{-100}$**	**Total Reads$_{-10}$**	**Total Reads$_{-5}$**	**Total Reads$_{-1}$**
Vogue1B2.01.01	2.550E-03	2.501E-03	2.499E-03	2.497E-03
Vogue1B2.01.06	1.128E-03	1.076E-03	1.073E-03	1.071E-03
Vogue1B2.01.07	2.493E-03	2.199E-03	2.183E-03	2.171E-03
Vogue1B2.01.08	1.134E-04	1.122E-04	1.122E-04	1.121E-04
Vogue1B2.01.09	3.889E-04	3.757E-04	3.750E-04	3.744E-04
Vogue1B2.01.10	1.801E-03	1.689E-03	1.683E-03	1.678E-03
Vogue1B2.01.11	4.184E-04	4.184E-04	4.184E-04	4.184E-04
Vogue1B2.01.12	6.927E-04	5.427E-04	5.344E-04	5.277E-04
Vogue1B2.01.15	1.777E-03	1.626E-03	1.619E-03	1.614E-03
Vogue1B2.01.19	4.849E-03	3.673E-03	3.609E-03	3.559E-03
Vogue1B2.01.21	3.277E-04	3.228E-04	3.225E-04	3.223E-04
Vogue1B2.01.23	4.561E-04	4.479E-04	4.474E-04	4.471E-04
Vogue1B2.01.26	8.324E-04	8.066E-04	8.052E-04	8.040E-04
Vogue1B2.01.28	3.957E-03	3.161E-03	3.118E-03	3.084E-03
Vogue1B2.01.29	1.648E-03	1.552E-03	1.547E-03	1.543E-03
Vogue1B2.01.35	8.882E-04	8.240E-04	8.207E-04	8.180E-04
Vogue1B2.01.37	5.706E-04	5.643E-04	5.639E-04	5.636E-04
Vogue1B2.01.38	1.143E-03	1.116E-03	1.115E-03	1.114E-03
Vogue1B2.01.50	2.880E-04	2.864E-04	2.863E-04	2.862E-04
Vogue1B2.01.52	2.451E-04	2.439E-04	2.438E-04	2.438E-04
Vogue1B2.01.56	1.760E-05	1.630E-06	7.730E-07	8.580E-08
Vogue1B2.01.58	2.968E-04	2.894E-04	2.889E-04	2.886E-04
Vogue1B2.01.61	5.714E-04	5.335E-04	5.314E-04	5.297E-04
Vogue1B2.01.62	1.325E-04	1.324E-04	1.324E-04	1.324E-04
Vogue1B2.01.63	4.102E-04	3.899E-04	3.888E-04	3.879E-04
Vogue1B2.01.64	5.037E-04	4.849E-04	4.839E-04	4.830E-04

Univariate Analyses

No statistically significant associations were seen between our clinical, demographic and behavioural information, and specific CSTs. However, effect size illustrated trends between BV-associated CSTs (III, IVA, IVC, and IVD), and number of sexual partners in the past year, oral sex, use of (hormonal) contraception, abnormal discharge in the past 48 hours, lifetime history of trichomoniasis, and an increasing number of BV episodes experienced in the past two months and year (Table 3.10).

Table 3.10 Univariate Analyses of Community State Types. Effect size of Fisher's exact test: Cramer's V (small: 0.1, medium: 0.3, large: 0.5). Effect size of one-way ANOVA: F-Statistic (small: 0.1, medium: 0.25, large: 0.4)

Univariate Analyses of Community State Types		
Fisher's Test		
Variable	**Cramer's V**	**Association**
Number of Sexual Partners (in past year)	0.74	CSTI
Lifetime History of Trichomoniasis	0.69	BV-associated CSTs
Number of BV Episodes (past 2 months)	0.68	BV-associated CSTs
No Method of Contraception	0.61	CSTI
Oral Sex Frequency	0.60	BV-associated CSTs
Number of BV Episodes (past year)	0.55	BV-associated CSTs
Hormonal Contraception	0.54	BV-associated CSTs
Abnormal Discharge (in past 48 hours)	0.54	CSTI
One-Way ANOVA		
Variable	**F-Statistic**	**Association**
Shannon's Diversity	1.76	CSTI

A statistically significant association between Shannon's Diversity Index, and lifetime history of chlamydia was detected; where mean diversity was higher in those who had a lifetime history of chlamydia compared to those who did not (Table 3.11). No other statistically significant associations were found, but based on effect size, trends suggested associations between Shannon's Diversity Index and use of illicit substances and alcohol, use of antimicrobials in the past three month, lifetime history of yeast, use of prescription and non-prescription drugs in the past two months, and number of BV episodes in the past year and lifetime (Table 3.11). Trends were seen between increased bacterial diversity and women who used illicit substances and consumed alcohol, used antimicrobials in the past three month, had a lifetime history of yeast, and number of BV episodes in the past year and lifetime; whereas there was a trend between decreased bacterial diversity, and use of prescription and non-prescription drugs in the past two months (Table 3.11).

Table 3.11 Univariate Analyses of Shannon's Diversity. P-values adjusted via Benjamini Hochberg; * p < 0.01. Effect size of two-independent sample t-test: Cohen's D (small: 0.2, medium: 0.5, large: 0.8). Effect size of one-way ANOVA: F-Statistic (small: 0.1, medium: 0.25, large: 0.4)

Univariate Analyses of Shannon's Diversity		
Two-Independent Sample T-Test		
Variable	**Cohen's D**	**Association**
Lifetime History of Chlamydia*	2.30	Positive
Substance Use	1.04	Positive
Use of Antimicrobials (in the past three months)	1.03	Positive
Lifetime History of Yeast	0.96	Positive
Use of (Non) Prescription Drugs (in the past two months)	0.88	Negative
One-Way ANOVA		
Variable	**F-Statistic**	**Association**
Community State Types	1.75	Negative
Number of BV episodes (lifetime)	1.18	Positive
Number of BV episodes (in the past year)	0.69	Positive

In addition to viewing the demographic, clinical and behavioural information in relation to CSTs and Shannon's Diversity Index, bacterial species abundance was viewed in relation to these variables via the ALDEx2 package. Associations were suggested between specific bacterial species and Nugent score, CSTs, study arm (healthy asymptomatic cohort and recurrent BV cohort), and number of BV episodes in the past two months and year (Table 3.12 - 3.15). Median relative abundances of bacterial species (centre-log-ratio transformed) were indicated for each variable (Table 3.12 – 3.15).

Table 3.12 ALDEx2 Nugent score. Kruskal Wallis analyses adjusted via Benjamini Hochberg principle. Median centre-log-ratio transformed relative abundance of bacterial species for each Nugent category; intermediate BV and consistent for BV Nugent scores combined into single category

ALDEx2: Nugent Score			
Bacterial Species	**Inconsistent (0-3)**	**Intermediate/Consistent (4-10)**	**P-Value**
Clostridia sp. BVAB2	2.11	7.46	<0.001
Lactobacillus crispatus	13.19	6.82	<0.001
Atopobium vaginae	1.72	6.56	<0.001
Gardnerella vaginalis Group C	2.65	6.63	<0.001
Lactobacillus gasseri	7.93	5.07	<0.001
Lactobacillus jensenii	7.78	4.86	<0.001
Gardnerella vaginalis Group A	7.02	9.78	<0.001
Prevotella timonensis	3.08	5.87	<0.001
Gardnerella vaginalis Group B	5.06	9.19	<0.001
Lactobacillus reuteri	3.20	1.15	<0.01
Megasphaera sp. type 1	2.40	4.39	<0.01
Clostridium saccharoperbutylacetonicum	2.38	0.54	<0.01
Peptoniphilus harei	0.85	2.25	<0.01
Prevotella buccalis	1.10	2.37	<0.01
Gardnerella vaginalis Group D	0.77	2.31	<0.01
Prevotella bergensis	0.95	1.95	<0.05
Lactobacillus iners	8.49	6.95	<0.05
Clostridium genomosp. BVAB3	1.23	2.65	<0.05
Prevotella bivia	0.82	2.12	<0.05

Table 3.13 ALDEx2 study arm. Kruskal Wallis analyses adjusted via Benjamini Hochberg principle. Median centre-log-ratio transformed relative abundance of bacterial species for Vogue 1A and Vogue 1B2

ALDEx2: Study Arm			
Bacterial Species	**1A**	**1B2**	**P-Value**
Gardnerella vaginalis Group C	3.08	9.17	<0.001
Gardnerella vaginalis Group B	5.99	1.04	<0.001
Streptococcus agalactiae	3.68	0.60	<0.001
Streptococcus constellatus	3.19	0.43	<0.01
Clostridia sp. BVAB2	2.62	6.21	<0.01
Propionibacterium granulosum	2.99	0.38	<0.01
Lactobacillus crispatus	12.24	7.36	<0.05
Actinobacteria sp.	0.39	3.32	<0.05
Streptococcus devriesei	0.41	3.04	<0.05
Lactobacillus plantarum	0.37	2.87	<0.05

Table 3.14 ALDEx2 Community State Types. Kruskal Wallis analyses adjusted via Benjamini Hochberg principle. Median centre-log-ratio transformed relative abundance of bacterial species for each CST category; CST III, IVA, IVC and IVD combined into single category (BV-associated CSTs/BV-CSTs)

ALDEx2: Community State Types			
Bacterial Species	**CSTI**	**BV-CSTs**	**P-Value**
Lactobacillus crispatus	14.48	7.05	<0.001
Lactobacillus iners	7.77	9.21	<0.001
Lactobacillus jensenii	8.43	4.68	<0.001
Gardnerella vaginalis Group A	6.82	8.75	<0.001
Atopobium vaginae	1.49	4.08	<0.001
Clostridia sp. BVAB2	2.07	4.51	<0.001
Gardnerella vaginalis Group C	2.39	4.90	<0.001
Lactobacillus gasseri	8.12	5.79	<0.001
Gardnerella vaginalis Group B	4.67	7.56	<0.001
Clostridium saccharoperbutylacetonicum	3.01	0.81	<0.001
Lactobacillus salivarius	1.78	1.00	<0.001
Lactobacillus reuteri	3.70	1.61	<0.001
Gardnerella vaginalis Group D	0.66	1.67	<0.001
Megasphaera sp. genomosp. type *1*	2.29	3.34	<0.001
Eubacterium cellulosolvens	0.89	1.33	<0.001
Prevotella timonensis	2.82	4.71	<0.01
Staphylococcus hominis	2.36	2.23	<0.05

Table 3.15 ALDEx2 genital infection history. Kruskal Wallis analyses adjusted via Benjamini Hochberg principle. Median centre-log-ratio transformed relative abundance of bacterial species for each category; zero episodes of BV in given timeframe, and more than one episodes of BV in give timeframe

ALDEx2: Genital Infection History				
Variable	**Bacterial Species**	**0**	**1+**	**P-Value**
Number of BV episodes (past two months)	*Gardnerella vaginalis* Group C	3.18	9.50	<0.01
	Gardnerella vaginalis Group A	7.29	12.15	<0.05
Number of BV episodes (past year)	*Gardnerella vaginalis* Group C	2.99	7.66	<0.001
	Clostridia sp. BVAB2	2.42	6.54	<0.01
	Gardnerella vaginalis Group A	7.17	10.34	<0.05

Healthy-Asymptomatic Comparison

In comparing our population of women with recurrent BV, to our healthy-asymptomatic cohort, we found that there were no statistically significant differences across demographics, nor most clinical and behavioural characteristics (Table 3.16, 3.17, 3.19, 3.20, 3.21). However ,significantly more women with recurrent BV used antimicrobials in the past three months,

experienced vaginal symptoms in the past 48 hours and 2 weeks, and reported substance use (within three months or three months prior to study visit) (Table 3.17, 3.18, 3.19). In our recurrent BV population, significantly fewer women had Nugent scores inconsistent with BV, and more Nugent scores consistent for BV, and intermediate BV (Table 3.22). We also saw a statistically significant difference in the CSTs present in each population; more profiles were dominated by *L. crispatus* and fell into the CSTI cluster in our healthy asymptomatic cohort, whereas in our recurrent BV cohort more women fell into CST clusters (III, IVA, IVC, IVD) (Table 3.23). Diversity statistics were not significantly different between the healthy-asymptomatic and recurrent BV cohorts (Table 3.24).

Table 3.16 Comparison of demographic characteristics: healthy-asymptomatic (1A) and recurrent BV (1B2) cohorts. Continuous variables are reported as mean ± SD, range, and effect size of two-independent sample t-test: Cohen's D (small: 0.2, medium: 0.5, large: 0.8). Categorical variables are reported as %, and effect size of Fisher's exact test: Cramer's V (small: 0.1, medium: 0.3, large: 0.5)

Demographic Characteristics				
	1A	**1B2**	**P-Value**	**Effect Size**
Age	30 ± 7.6 (18 – 49)	32 ± 7.9 (20 – 47)	1.00	0.26
BMI	23.9 ± 5.2 (15.1 – 50.0)	22.7 ± 4.0 (15.5 – 31.2)	1.00	0.22
Ethnicity	Caucasian: 64.5% Asian: 19.0% South Asian: 3.9% Black: 2.9% Aboriginal origins: 1.9% Hispanic: 1.6% Other: 6.1%	Caucasian: 61.5% Asian: 26.9% Aboriginal origins: 3.9% Other: 7.7%	1.00	0.02

Table 3.17 Comparison of clinical characteristics: healthy-asymptomatic (1A) and recurrent BV (1B2) cohorts. Categorical variables are reported as %. Effect size of Fisher's exact test: Cramer's V (small: 0.1, medium: 0.3, large: 0.5)

Clinical Characteristics				
	1A	**1B2**	**P-Value**	**Effect Size**
Pregnancy History	Nulliparous: 69.7% Multiparous: 15.8%	Nulliparous: 53.8% Multiparous: 38.5%	0.47	0.16
Form of Contraception	None: 16.5% Hormonal: 41.9% Barrier: 35.8%	None: 38.5% Hormonal: 30.8% Barrier: 19.2%	0.16	0.17
			1.00	0.05
			1.00	0.05
Use of Condoms (past 48 hours)	4.5%	11.5%	1.00	0.09
Antimicrobial Use (in the past three months)	13.9%	69.2%	<0.001*	0.41
(Non) Prescription Drug Use (past two months)	75.5%	69.2%	1.00	0.02

Table 3.18 Comparison of vaginal symptoms: healthy-asymptomatic (1A) and recurrent BV (1B2) cohorts. Categorical variables are reported as %. Effect size of Fisher's exact test: Cramer's V (small: 0.1, medium: 0.3, large: 0.5)

Vaginal Symptoms				
	1A	**1B2**	**P-Value**	**Effect Size**
Symptoms (past 48 hours)	Abnormal discharge: 3.9% Abnormal odor: 0.6% Irritation or discomfort: 5.5% Other: 2.9%	Abnormal discharge: 57.7% Abnormal odor: 23.1% Irritation or discomfort: 65.4% Other: 30.8%	<0.001*	0.51
			<0.001*	0.40
			<0.001*	0.47
			<0.01*	0.29
Symptoms (past 2 weeks)	Abnormal discharge: 7.1% Abnormal odor: 1.6% Irritation or discomfort: 9.4% Other: 3.2%	Abnormal discharge: 69.2% Abnormal odor: 30.8% Irritation or discomfort: 69.2% Other: 26.9%	<0.001*	0.55
			<0.001*	0.41
			<0.001*	0.53
			<0.01*	0.34
Painful Intercourse	29.4%	53.8%	0.47	0.15

Table 3.19 Comparison of behavioural characteristics: healthy-asymptomatic (1A) and recurrent BV (1B2) cohorts. Categorical variables are reported as %. Effect size of Fisher's exact test: Cramer's V (small: 0.1, medium: 0.3, large: 0.5)

Behavioural Characteristics				
	1A	**1B2**	**P-Value**	**Effect Size**
Tampon Use	Never: 26.8% Sometimes: 20.6% Every period, part of the time: 25.5% Every period, exclusively: 25.5%	Never: 23.1% Sometimes but not for every period: 15.4% Every period, part of the time: 34.6% Every period, exclusively: 26.9%	1.00	0.03
Tampon Use (past month)	40.6%	57.7%	1.00	0.10
Use of Feminine Products	13.9%	19.2%	1.00	0.05
Use of Feminine Products (past 48 hours)	3.5%	15.4%	0.75	0.15
Illicit Substance Use	Never: 45.2% Past: 24.5% Current: 17.7%	Never: 26.9% Past: 46.2% Current: 23.1%	<0.05*	0.20
Alcohol Consumption	None: 7.7% Occasional drink: 48.1% 2-3 drinks/week: 27.7% Daily: 2.9%	None: 19.2% Occasional drink: 46.2% 2-3 drinks/week: 30.8%	<0.05*	0.20
Smoker	Never: 61.6% Past: 13.5% Current:12.3%	Never: 50.0% Past: 34.6% Current: 11.5%	1.00	0.02

Table 3.20 Comparison of sexual activity characteristics: healthy-asymptomatic (1A) and recurrent BV (1B2) cohorts. Continuous variables are reported as mean ± SD, range. Categorical variables are reported as %, and effect size of Fisher's exact test: Cramer's V (small: 0.1, medium: 0.3, large: 0.5). Number of sexual partners (past year) analyzed as categorical variable.

Sexual Activity Characteristics				
	1A	**1B2**	**P-Value**	**Effect Size**
Sexual Partners	Male: 89.7% Female: 2.9% Both: 4.2% Virgin: 3.2%	Male: 92.3% Female: 3.8%	1.00	0.08
Number of Sexual Partners (past year)	1.5 ± 1.9 (0 – 25)	1.2 ± 0.9 (0 – 4)	1.00	0.04
Vaginal intercourse (past 48 hours)	Yes:14.5% No: 85.5%	Yes: 23.1% No: 73.1%	1.00	0.07
Frequency of Oral Sex	Never: 26.5% Ever: 73.5%	Never: 15.4% Ever: 80.8%	1.00	0.06
Frequency of Anal Sex	Never: 83.9% Weekly: 1.3% Monthly: 2.9% Other: 11.9%	Never: 69.2% Weekly: 3.8% Monthly: 3.8% Other: 19.2%	1.00	0.09
Frequency of Sex Toy use	Never: 63.5% Daily: 0.6% Weekly: 13.9% Twice a month: 1.0% Monthly: 11.6% Other: 9.4%	Never: 38.5% Weekly: 7.7% Monthly: 15.4% Other: 34.6%	0.57	0.23

Table 3.21 Comparison of genital infection history: healthy-asymptomatic (1A) and recurrent BV (1B2) cohorts. Continuous variables are reported as mean ± SD, range, and effect size of two-independent sample t-test: Cohen's D (small: 0.2, medium: 0.5, large: 0.8). Effect size of Fisher's exact test: Cramer's V (small: 0.1, medium: 0.3, large: 0.5). BV episodes (in the past two months, past year, and lifetime) analyzed as categorical variable.

Genital Infection History				
	1A	**1B2**	**P-Value**	**Effect Size**
Bacterial Vaginosis	16.8%	69.2%	*<0.001	0.39
Lifetime	2.4 ± 3.9 (0 – 20)	4.7 ± 6.1 (0 – 20)	*<0.001	0.45
In the past year	0.4 ± 0.9 (0 – 6)	2.3 ± 3.1 (0 – 12)	*<0.001	0.58
In the past two months	0.08 ± 0.3 (0 – 1)	0.6 ± 0.8 (0 – 2)	*<0.001	0.50
Yeast	67.1%	88.5%	1.00	0.07
Urinary Tract Infections	51.6%	61.5%	1.00	0.01
Trichomoniasis	1.0%	3.8%	1.00	0.07
Genital Warts	6.8%	15.4%	1.00	0.09
Genital Herpes	4.8%	15.4%	1.00	0.11
Chlamydia	7.7%	19.2%	1.00	0.10
Gonorrhea	1.6%	0%	1.00	NA
Syphilis	0.3%	0%	1.00	NA

Table 3.22 Comparison of Nugent scores: healthy-asymptomatic (1A) and recurrent BV (1B2) cohorts. Categorical variables are reported as %. Effect size of two-independent sample t-test: Cohen's D (small: 0.2, medium: 0.5, large: 0.8).

Nugent Scores				
	1A	**1B2**	**P-Value**	**Effect Size**
Inconsistent for BV (0 – 3)	77.7%	23.1%	*<0.001	1.10
Intermediate BV (4 – 6)	8.1%	50.0%		
Consistent for BV (7 – 10)	10.3%	26.9%		

Table 3.23 Comparison of community state types: healthy-asymptomatic (1A) and recurrent BV (1B2) cohorts. Categorical variables are reported as %. Effect size of Fisher's exact test: Cramer's V (small: 0.1, medium: 0.3, large: 0.5)

Community State Types					
CST	**Dominant OUT**	**1A**	**1B2**	**P-Value**	**Effect Size**
Type I	*L. crispatus* dominated	50.3%	19.2%	*<0.05	0.30
Type II	*L. gasseri* dominated	0%	3.8%		
Type III	*L. iners* dominated	16.1%	23.1%		
Type 1VA	Mixed Dominant Phenotype: Mixed Dominant Phenotype: *Gardnerella* subgroup B, *Bifidobacterium* sp., *Prevotella amnii, Shigella boydii, and L.delbrueckii*	11.6%	23.1%		
Type IVC	*Gardnerella* A dominated	7.1%	23.1%		
Type IVD	*Gardnerella* C dominated	7.7%	7.7%		
Type V	*L. jensenii* dominated	7.1%	0%		

Table 3.24 Comparison of diversity statistics: healthy-asymptomatic (1A) and recurrent BV (1B2) cohorts. Continuous variables are reported as mean ± SD, range. Effect size of two-independent sample t-test: Cohen's D (small: 0.2, medium: 0.5, large: 0.8)

Diversity Statistics				
	1A	**1B2**	**P-Value**	**Effect Size**
Shannon's Diversity	0.86 ± 0.59 (0.04 – 2.35)	0.90 ± 0.52 (0.18 – 1.86)	1.00	0.06
Pielou's Evenness	0.30 ± 0.20 (0.02 – 0.78)	0.31 ± 0.17 (0.07 – 0.64)	1.00	0.09
Species Richness	18.63 ± 4.59 (5 – 30)	17.77 ± 3.54 (11 – 23)	1.00	0.19
Good's Coverage	1.00 ± 0.0 (0.97 – 1.00)	1.00 ± 0.0 (0.996 – 1.00)	1.00	0.03

3.2 Virome Results

Virome Protocol Optimization Results

The qPCR cycle threshold (Ct) values were determined to quantitate the various components in the simulated samples; A549 tissue culture cells, *Escherichia coli*, and human adenovirus and enterovirus; prior to processing (Table 3.25). Ct values were also determined for various dilutions of A549 tissue culture cells, *Escherichia coli*, and human adenovirus and

enterovirus. To avoid concentrated samples, dilutions were chosen to maintain a Ct value within an ideal detectable range of 25-30.

Table 3.25 qPCR Ct values for diluted and undiluted samples. Ct values for undiluted and diluted human adenovirus and enterovirus, *Escherichia coli*, and human cell samples. *Bolded Ct values were chosen for mock samples

qPCR Ct Values				
	Undiluted	**5x Dilution**	**10x Dilution**	**100x Dilution**
Adenovirus	10.7625		13.36	**17***
Enterovirus		14.53	**15.22***	
Escherichia coli	**24.77***		26.73	28.48
Human (DNA)	26.535		30.3	**33.985***
Human (RNA)	26.89		**30.49***	33.815

These Ct values were compared to Ct values of the various components in the simulated samples after processing (Table 3.26, 3.27). Contamination was present in our negative controls, and additional runs were conducted where no A549 tissue culture cells, *Escherichia coli*, and human adenovirus and enterovirus were present in the negative controls (Table 3.28). Centrifugation and filtration was the optimal pipeline, as indicated by results.

Table 3.26 Pipeline optimization run #1. qPCR Ct values for human adenovirus and enterovirus, and *Escherichia coli* components of mock samples, and for positive and negative controls. Mock samples ran through three pipelines, centrifugation and filtration, filtration, or neither centrifugation or filtration, and varying amounts of RNase A (ng) used. Undet: undetected; target species not present

Pipeline Optimization Run #1: Samples										
	Initial Ct	**Centrifugation and Filtration**			**Filtration**			**Neither Centrifugation or Filtration**		
Rnase A (ng)		**0**	**10**	**20**	**0**	**10**	**20**	**0**	**10**	**20**
Adenovirus	15	15.61	15.62	15.66	14.86	15.06	14.92	14.96	15.01	14.89
Enterovirus	21	22.79	22.63	22.6	22.52	22.96	22.77	22.14	22.32	22.31
Escherichia coli	26	39.84*	39.82*	undet	38.73	38.55	39.16	26.54	26.35	26.17
Pipeline Optimization Run #1: Positive and Negative Controls										
	Positive Control	**Centrifugation and Filtration**			**Filtration**			**Neither Centrifugation or Filtration**		
Rnase A (ng)		**0**	**10**	**20**	**0**	**10**	**20**	**0**	**10**	**20**
Adenovirus	14.4	35.66	31.78	33.33	37.64	34.79	33.79	38.12	36.78	35.92
Enterovirus	23.54	undet	undet	undet	undet	undet	undet	undet	undet	undet
Escherichia coli	28.72	undet	undet	undet	undet	undet	undet	undet	undet	undet

Table 3.27 Pipeline optimization run #2. qPCR Ct values for human adenovirus and enterovirus, *Escherichia coli* and human cell components of mock samples, and for positive and negative controls. Mocks samples ran through three pipelines, centrifugation and filtration, filtration, or neither centrifugation or filtration, and varying amounts of RNase A (ng) used. Undet: undetected; indicated species not present

Pipeline Optimization Run #2: Samples										
	Initial Ct	**Centrifugation and Filtration**			**Filtration**			**Neither Centrifugation or Filtration**		
Rnase A (ng)		**0**	**10**	**20**	**0**	**10**	**20**	**0**	**10**	**20**
Adenovirus	17	24.93	24.68	24.71	23.77	23.62	23.54	23.34	23.31	23.31
Enterovirus	15.22	24.12	24.16	24.34	24.01	24.11	24.06	23.77	23.91	23.91
Escherichia coli	24.77	undet	undet	undet	undet	undet	undet	27.35	27.28	26.91
Human (DNA)	33.985	undet	undet	undet	undet	undet	undet	undet	undet	undet
Human (RNA)	30.49	undet	undet	undet	undet	undet	undet	undet	undet	undet
Pipeline Optimization Run #2: Positive and Negative Controls										
	Positive Control	**Centrifugation and Filtration**			**Filtration**			**Neither Centrifugation or Filtration**		
Rnase A (ng)		**0**	**10**	**20**	**0**	**10**	**20**	**0**	**10**	**20**
Adenovirus	10.84	24.55	24.44	24.41	23.8	23.87	23.82	23.48	23.34	23.29
Enterovirus	12.16	24.13	24.12	24.06	24.13	24.05	24.13	24.08	23.94	23.97
Escherichia coli	26.61	undet	undet	undet	undet	undet	undet	27.45	27.27	27.19
Human (DNA)	NA	undet	undet	undet	undet	undet	undet	undet	undet	undet
Human (RNA)		undet	undet	undet	undet	undet	undet	undet	undet	undet

Table 3.28 Pipeline optimization run #3. qPCR Ct values for human adenovirus and enterovirus, *Escherichia coli* and human cell components of mock samples and positive control. Mocks samples ran through three pipelines, centrifugation and filtration, filtration, or neither centrifugation or filtration, and varying amounts of RNase A (ng) used. Undet: undetected; indicated species not present. Ct values could not be determined for negative controls as the indicated species were not detected; exception human adenovirus in the Filtration Pipeline using 20ng of RNAse A had a Ct of 39.24

Pipeline Optimization Run #3: Samples and Positive Control											
	Initial Ct	**Centrifugation and Filtration**			**Filtration**			**Neither Centrifugation or Filtration**			**Positive Control**
Rnase A (ng)		**0**	**10**	**20**	**0**	**10**	**20**	**0**	**10**	**20**	
Adenovirus	17	24.89	27.88	28.74	23.51	23.46	23.41	23.06	23.06	31.48	10.34
Enterovirus	15.22	24.26	24.31	24.21	24.05	24.13	24.16	23.86	23.99	23.99	14.43
Escherichia coli	24.77	undet	undet	undet	undet	39.27*	undet	26.99	27.32	26.91	26.58
Human (DNA)	33.985	undet	undet	undet	undet	undet	undet	undet	undet	undet	na
Human (RNA)	30.49										

Results support the centrifugation and filtration protocol's ability to maintain the viral communities in our sample, yet reduce the bacterial and human cells present. Bacterial and human cell nucleic acid presence was greater in the pellet versus the post-processed sample, and viral nucleic acids were evenly distributed between the pellet and post-processed sample (Table 3.29, 3.30).

Table 3.29 Spiking experiment #1. qPCR Ct values for human adenovirus and enterovirus, *Escherichia coli* and human cells, which were spiked in healthy-asymptomatic (Vogue 1A) vaginal samples. Pellets: recovered from centrifugation. Raw+emLB: raw healthy-asymptomatic spiked sample treated with lysis buffer and filtered. Positive control: for filtration protocol. Undet: undetected; indicated species not present

Spiking Experiment #1													
	Initial Ct	**Vogue 1A Sample**			**Pellets from 1A Sample**			**Raw+emLB**		**Positive Control**			**Negative Control**
		1	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**1**	**2**	**3**	**1&2&3**
Adenovirus	17	23.86	23.96	23.96	24.09	24.72	24.51	38.15	undet	23.63	23.58	23.56	undet
Enterovirus	15.22	26.91	26.95	27	26.93	27.45	28.19	undet	undet	28.44	28.57	28.48	undet
Escherichia coli	24.77	undet	undet	undet	undet	undet	undet	undet	undet	25.78	25.81	NA	undet
Human (DNA)	33.985	undet	undet	undet	30.02	30.98	30.41	29.56	29.49	undet			undet
Human (RNA)	30.49												

Table 3.30 Spiking experiment #2. qPCR Ct values for human adenovirus and enterovirus, *Escherichia coli* and human cells, which were spiked in healthy-asymptomatic (Vogue 1A) vaginal samples. Pellets: recovered from centrifugation. Positive control: for filtration protocol. Undet: undetected; indicated species not present

Spiking Experiment #2											
	Initial Ct	**Vogue 1A Samples**			**Pellets from Vogue 1A Samples**			**Positive Control**	**qPCR Positive Control**		**Negative Control**
		1	**2**	**3**	**1**	**2**	**3**	**1**	**1**	**2**	**1&2**
Adenovirus	17	25.1	24.06	24.3	24.06	24.38	23.94	23.62	9.55	9.53	undet
Enterovirus	15.22	26.42	26.27	26.28	25.8	26.28	25.94	26.39	14.25	14.15	undet
Escherichia coli	26.18	undet	39.93	undet	26.95	26.85	27	39.19	25.17	25.3	undet
Human (DNA)	33.985	38.58	39.32	undet	27.92	27.98	28.09	undet	NA	NA	undet
Human (RNA)	30.49										

Virome Study Population

Healthy-Asymptomatic Cohort

Demographic Characteristics

Our team enrolled 310 healthy-asymptomatic women between January 2011 and April 2014, and 21 viral vaginal swabs were selected from this cohort of women. These 21 healthy-asymptomatic women had a mean age of 28 years (range 20 – 42 years). Self-reported ethnicity was 11 Caucasian, nine Asian, and one South Asian. The highest education level attained of our study population was, one did not complete high school, one received their high school diploma, six received some post-secondary education, seven completed post-secondary education, and six completed their graduate degree. The average BMI of this group was 24.8 kg/m^2 (range 16.7 - 39.8 kg/m^2) (Table 3.31).

Table 3.31 Demographics characteristics for healthy-asymptomatic cohort. **Continuous variables are reported as mean ± SD, range, and categorical variables are reported as n**

Demographics	
Age (Mean ± SD, Range)	28.4 ± 5.9 (20 – 42)
BMI (Mean ± SD, Range)	24.8 ± 6.4 (16.7 – 39.8)
Ethnicity	Caucasian: 11 Asian: 9 South Asian: 1
Highest Education Level Attained	Graduate degree: 6 Post-secondary education: 7 Some post-secondary education: 6 High school diploma: 1 Did not complete high school: 1

Clinical Characteristics

Antimicrobial use was reported in none of women in the past 3 months (Table 3.32). Most women reported using prescription and non-prescription drugs in the past two months (17/21), and four women did not use any prescription and non-prescription drugs in the past two months.

Since their last menstrual period, women in our cohort used different types of contraception. Four women reported using no form of contraception, six women used hormonal contraception, eight women used barriers, one woman used an IUD, and two women were not sexually active (Table 3.32). The majority of women in our cohort were nulliparous (15/21), four women were multiparous and two women had single pregnancies (Table 3.32).

Table 3.32 Clinical characteristics for healthy-asymptomatic cohort. Categorical variables are reported as n

Clinical Characteristics	
Pregnancy History	Nulliparous: 15 Multiparous: 4
Form of Contraception	None: 4 Hormonal: 6 Barrier: 8
Use of Condoms (in past 48 hours)	Yes: 1 No: 20
Symptoms (in past 48 hours)	Abnormal discharge: 0 Abnormal odor: 0 Irritation or discomfort: 2 Painful intercourse: 8 Other: 0
Antimicrobial Use (in the past three months)	Yes: 0 No: 21

Mostly all women in our cohort have been previously diagnosed with a genital infection (20/21). Reported lifetime histories of genital conditions included, six reports of a history of BV, 15 reports of a history of yeast, nine reports of UTIs, one *Trichomonas vaginalis* infection, one report of genital warts, two cases of genital herpes, and one case of *Chlamydia trachomatis* infection. There were no women with a history of gonorrhea or syphilis. Women experienced an average of 1.3 BV infections in their lifetime (range 0 - 20), 0.3 BV infections in the last year (range 0 - 6), and 0 BV infections in the past 2 months (Table 3.33).

Table 3.33 Genital infection history for healthy-asymptomatic cohort. Continuous variables are reported as mean ± SD, range, and categorical variables are reported as n

Genital Infection History	
Bacterial Vaginosis (Mean ± SD, Range)	
Lifetime	1.3 ± 4.4 (0 – 20)
In the past year	0.3 ± 1.3 (0 – 6)
In the past two months	0
Yeast	15/21
Urinary Tract Infections	9/21
Trichomoniasis	1/21
Genital Warts	1/21
Genital Herpes	2/21
Chlamydia	1/21
Gonorrhea	0/21
Syphilis	0/21

Three women experienced abnormal vaginal symptoms within two weeks of the study visit; these three women experienced irritation or discomfort. Two of these three women continued to experience irritation or discomfort within 48 hours of the study visit. Several women experience painful intercourse (8/21), and on average women experienced painful intercourse 47.5% of the time (range 10 - 80%) (Table 3.32). Nugent scores between 0 - 3 are inconsistent for BV, scores between 4 - 6 reflect intermediate BV, and scores between 7 - 10 are consistent for BV. 13 women had Nugent scores inconsistent with BV, three with intermediate BV, and four consistent with BV (Table 3.34). One woman had a missing Nugent score result.

Table 3.34 Nugent scores for healthy-asymptomatic cohort. Categorical variables are reported as n

Nugent Scores	
Inconsistent for BV (0 – 3)	13/21
Intermediate BV (4 – 6)	3/21
Consistent for BV (7 – 10)	4/21

Behavioural Characteristics

Tampon use was reported in 14 women, and seven women reported never using tampons during their menstrual periods. With regards to feminine hygiene, five women used feminine wipes or genital deodorant products, and two women used these products in the past 48 hours (Table 3.35).

In the past year, women reported 1.5 sexual partner (range 0 - 6), and in the past two months women reported 1.0 sexual partners (range 0 -4). With regards to sexual activity, five women reported vaginal intercourse in the past 48 hours. Most women reported receiving oral sex (17/21), and four reported never receiving oral sex. One woman reported receiving oral sex in the past 48 hours. All 21 women reported never having anal sex. Seven women reported sex toy use, and 14 have never used sex toys. One woman in our population used sex toys in the past 48 hours (Table 3.35).

Reported illicit substance/alcohol use includes, 15 women never using illicit substances or consuming alcohol, and six women currently using illicit substances or consuming alcohol. Of our 21 women, 18 have never smoked, and three are current smokers (Table 3.35).

Table 3.35 Behavioural characteristics for healthy-asymptomatic cohort. Continuous variables are reported as mean ± SD, range, and categorical variables are reported as n

Behavioural Characteristics	
Tampon Use (ever)	Yes: 14 No: 7
Tampon Use (in the past month)	Yes: 10 No: 11
Feminine Hygiene Products Use (ever)	Yes: 5 No: 16
Use of Feminine Hygiene Products (past 48 hours)	Yes: 2 No: 19
Substance Use	Yes: 6 No: 15
Current Smoker	Yes: 3 No: 18
Number of Sexual Partners (past year) (Mean ± SD, Range)	1.5 ± 1.6 (0 – 6)
Vaginal Intercourse (past 48 hours)	Yes: 5 No: 16
Oral Sex	Yes: 17 No: 4
Anal Intercourse	Yes: 0 No: 21
Use of Sex Toys	Yes: 7 No: 14

Bacteriome Data

Based on >55% cut-off identity to the *cpn*60 gene target; a total of 168 unique *cpn*60 OTUs were identified from 160, 576 reads. Only one OTUs was detected in all samples, and this corresponded to *G. vaginalis* Group A. *L. crispatus* and *L. iners* were detected in all samples, with the exception of one or two individual(s).

Among women with inconsistent for BV scores, nine profiles were dominated by *Lactobacillus* species, two profiles were dominated by *G. vaginalis* species, one profile was dominated by *Bifidobacterium breve* and one profile had a mixed dominance profile of *Lactobacillus* species, Other *Proteobacteria* and *Actinobacteria*. Furthermore as seen in Figure 3.6, seven women had intermediate and consistent for BV scores; two profiles were dominated

by *Lactobacillus* species, one by *Lactobacillus* and *G. vaginalis* species, one by *G. vaginalis* species, one by *G. vaginalis* Group B, *Megasphaera* and *Prevotella* species, one by *Bifidobacterium breve*, and one profile had a mixed dominance of *Atopobium vaginae*, *Clostridium* BVAB3, *G. vaginalis* species, *Megasphaera* and *Prevotella* species. One woman had missing Nugent score data; her profile was dominated by *G. vaginalis* species, and *Prevotella timonensis* (Figure 3.6). CSTs are as follows; CST I: 5/21, CST III: 4/21, CST IVA:3/21, CST IVC: 3/21, CST IVD: 3/21, CST V: 3/21. Additional CSTs were identified in addition to CSTs previous studies have shown [39,40] (Table 3.36).

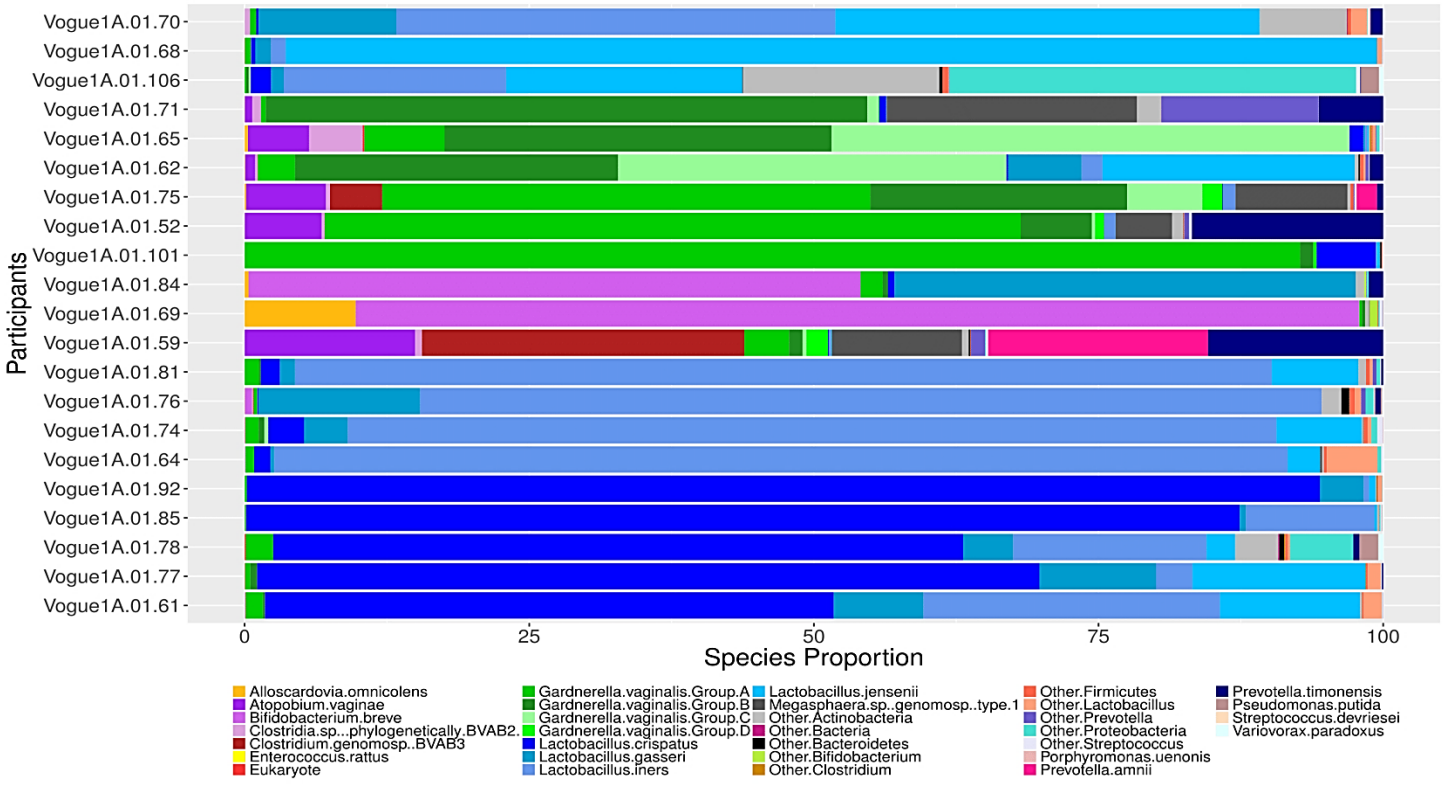

Figure 3.6 Healthy-asymptomatic vaginal bacteriome profiles ordered CST

Table 3.36 Community state types for healthy-asymptomatic cohort. Categorical variables are reported as n

Community State Types		
CST	**Dominant OUT**	**N**
I	*L.crispatus* dominated	**5**
III	*L.iners* dominated	**4**
IVA	Mixed Dominant Phenotype: *Gardnerella* subgroup B, *Bifidobacterium* sp., *Prevotella amnii, Shigella boydii, and L.delbrueckii*	**3**
IVC	*Gardnerella* A dominated	**3**
IVD	*Gardnerella* C dominated	**3**
V	*L.jensenii* dominated	**3**

HIV-Positive Cohort

Demographic Characteristics

Our team enrolled 54 women who were attending for Oak Tree clinic between April 2011 and October 2011, and 25 viral vaginal swabs were selected from this cohort of women. These 25 HIV-positive women had a mean age of 37 years (range 23 – 48 years). Self-reported ethnicity was ten Caucasian, six First Nations, four Black, four mixed ethnicities, and one South Asian. The highest education level attained of our study population was, 12 did not complete high school, four received their high school diploma, six received some post-secondary education, two completed post-secondary education, and one completed their graduate degree. The average BMI of this group was 26.0 kg/m^2 (range 17.8 - 42.9 kg/m^2) (Table 3.37).

Table 3.37 Demographics characteristics for HIV-positive cohort. Continuous variables are reported as mean ± SD, range, and categorical variables are reported as n

Demographics	
Age (Mean ± SD, Range)	36.9 ± 7.1 (23 – 48)
BMI (Mean ± SD, Range)	26 ± 6.8 (17.8 – 42.9)
Ethnicity	Caucasian: 10 Aboriginal origins: 6 Black: 4 South Asian: 1 Other: 4
Highest Education Level Attained	Graduate degree: 1 Post-secondary education: 2 Some post-secondary education: 6 High school diploma: 4 Did not complete high school: 12

Clinical Characteristics

Antimicrobial use was reported in almost half of women (10/25) in the past 3 months, and 15 women did not take any antimicrobials in the past 3 months (Table 3.38). Most women reported using prescription and non-prescription drugs in the past two months (20/25), and five women did not use any prescription and non-prescription drugs in the past two months.

Since their last menstrual period, women in our cohort used different and multiple types of contraception. Two women reported using no form of contraception, six women used hormonal contraception, 12 women used barriers, two woman used surgical sterilization, and eight women were not sexually active (Table 3.38). Most women in our cohort were multiparous (19/25). Two women had single pregnancies, and four women were nulliparous (Table 3.38).

Table 3.38 Clinical characteristics for HIV-positive cohort. Categorical variables are reported as n

Clinical Characteristics	
Pregnancy History	Nulliparous: 4 Multiparous: 19
Form of Contraception	None: 4 Hormonal: 6 Barrier: 12
Use of Condoms (in past 48 hours)	Yes: 2 No: 23
Symptoms (in past 48 hours)	Abnormal discharge: 4 Abnormal odor: 3 Irritation or discomfort: 1 Painful intercourse: 3 Other: 0
Antimicrobial Use (in the past three months)	Yes: 10 No: 15

Mostly all women in our cohort have been previously diagnosed with a genital infection (24/25). Reported lifetime histories of genital conditions included, 14 reports of a history of BV, 20 reports of a history of yeast, 14 reports of UTIs, five *Trichomonas vaginalis* infections, eight reports of genital warts, three cases of genital herpes, five cases of *Chlamydia trachomatis* infection, five reports of gonorrhea, and one report of syphilis. Women experienced an average of 1.4 BV infections in their lifetime (range 0 - 6), 0.4 BV infections in the last year (range 0 - 3), and 0.1 BV infections in the past 2 months (range 0 - 2). Women also reported 7.1 yeast infections in their lifetime (range 0 - 50), 0.5 yeast infections in the last year (range 0 - 6), and 0.1 yeast infections in the past 2 months (range 0 - 1)**.** In addition, women reported 5.3 UTI infections in their lifetime (range 0 - 100), with 0.8 UTI infections in the past year (range 0 - 15) and 0.1 UTI infections in the past two months (range 0 - 2) (Table 3.39).

Table 3.39 Genital infection history for HIV-positive cohort. Continuous variables are reported as mean ± SD, range, and categorical variables are reported as n

Genital Infection History	
Bacterial Vaginosis (Mean ± SD, Range)	
Lifetime	1.4 ± 1.6 (0 – 6)
In the past year	0.4 ± 0.8 (0 – 3)
In the past two months	0.1 ± 0.4 (0 – 2)
Yeast	20/25
Urinary Tract Infections	14/25
Trichomoniasis	5/25
Genital Warts	8/25
Genital Herpes	3/25
Chlamydia	5/25
Gonorrhea	5/25
Syphilis	1/25

Six women experienced abnormal vaginal symptoms within two weeks of the study visit; six women experienced abnormal discharge, four abnormal odor, one irritation or discomfort, and one other vaginal symptoms. Four women experienced abnormal vaginal symptoms within 48 hours of the study visit; four women experienced abnormal discharge, three abnormal odor, and one irritation or discomfort. Few women experience painful intercourse (3/25), and on average women experienced painful intercourse 23.3% of the time (range 10 - 50%) (Table 3.38). Nugent scores between 0 - 3 are inconsistent for BV, scores between 4 - 6 reflect intermediate BV, and scores between 7 - 10 are consistent for BV. 13 women had Nugent scores inconsistent with BV, four with intermediate BV, and eight consistent with BV (Table 3.40).

Table 3.40 Nugent scores for HIV-positive cohort. Categorical variables are reported as n

Nugent Scores	
Inconsistent for BV (0 – 3)	13/25
Intermediate BV (4 – 6)	4/25
Consistent for BV (7 – 10)	8/25

Behavioural Characteristics

Tampon use was reported in 16 women, and nine women reported never using tampons during their menstrual periods. With regards to feminine hygiene, six women used feminine wipes or genital deodorant products, and one woman used these products in the past 48 hours (Table 3.41).

In the past year, women reported 1 sexual partner (range 0 - 3), and in the past two months women reported 0.72 sexual partners (range 0 -1). With regards to sexual activity, four women reported vaginal intercourse in the past 48 hours. Almost 50% of women reported receiving oral sex (13/25), and 12 women reported never receiving oral sex. Two women reported receiving oral sex in the past 48 hours. Two women reported engaging in anal sex, and 23 reported never having anal sex. One woman in our population had anal sex in the past 48 hours. Four women reported sex toy use, and 21 have never used sex toys. One woman in our population used sex toys in the past 48 hours (Table 3.41).

Reported illicit substance/alcohol use includes, 19 women never using illicit substances or consuming alcohol, and six women currently using illicit substances or consuming alcohol. Of our 25 women, half have never smoked (13/25), and half are current smokers (12/25) (Table 3.41).

Table 3.41 Behavioural characteristics for HIV-positive cohort. Continuous variables are reported as mean ± SD, range, and categorical variables are reported as n

Behavioural Characteristics	
Tampon Use (ever)	Yes: 16 No: 9
Tampon Use (in the past month)	Yes: 11 No: 14
Feminine Hygiene Products Use (ever)	Yes: 6 No: 19
Use of Feminine Hygiene Products (past 48 hours)	Yes: 1 No: 24
Substance Use	Yes: 6 No: 19
Current Smoker	Yes: 12 No: 13
Number of Sexual Partners (past year) (Mean ± SD, Range)	1.0 ± 0.7 (0 – 3)
Vaginal Intercourse (past 48 hours)	Yes: 4 No: 21
Oral Sex	Yes: 13 No: 12
Anal Intercourse	Yes: 2 No: 23
Use of Sex Toys	Yes: 4 No: 21

HPV Data

On average, women had 2.45 HPV types (range 0-13) (HPV data was missing for five women). HPV-6, HPV-11, HPV-31, HPV-34, HPV-39, HPV-40, HPV-44, HPV-58, HPV-59, HPV-69, HPV-82 were not present in any of our participants. Seven women had high risk HPV types, one woman had an HPV type classified as probably carcinogenic, nine women had HPV types which are believed to be possibly carcinogenic and 13 women had low risk HPV types (Table 3.42) [165,190,191].

Table 3.42 HPV type characteristics. Continuous variables are reported as mean ± SD, range), and categorical variables are reported as n. HPV types as defined by [165,190,191]. HPV data was missing for five women.

HPV Characteristics	
Number of HPV Types (Mean ± SD, Range)	2.45 ± 3.1 (0 – 13)
Carcinogenic HPV Types	
HPV-16	2
HPV-18	1
HPV-33	2
HPV-35	1
HPV-45	1
HPV-51	2
HPV-52	1
HPV-56	3
Probably Carcinogenic HPV Types	
HPV-68	1
Possibly Carcinogenic HPV Types	
HPV-26	1
HPV-53	1
HPV-66	2
HPV-67	2
HPV-70	4
HPV-73	1
Low Risk HPV Types	
HPV-42	2
HPV-54	2
HPV-61	6
HPV-62	5
HPV-71	1
HPV-72	1
HPV-81	3
HPV-83	1
HPV-84	2
HPV-89	1

HIV Data

One woman fell into HIV clade A, 19 in HIV clade B and three women in HIV clade C. HIV clade information was missing for two women. There were several likely modes of HIV acquisition. Twelve women reported sexual contact, five drug use, one reported blood products or percutaneous, one reported perinatal transmission, and mode of transmission was unknown in six women. Most women were not antiretroviral naive (24/25). On average, women were on antiretroviral therapy for a total of 322 weeks (range 0 - 1155) (Table 3.43).

In our population, the average duration of HIV infection was 150 months (range 21 - 292 months). Median CD4 nadir was 170 mm^3 (range 20-440 mm^3), and median CD4 at study visit was 420 mm^3 (range 90 - 870 mm^3). The median highest plasma HIV viral load (VL) for this population was 75,300 copies/mL (range 759 – 699,939 copies/mL), and the median VL at study visit was 39 copies/mL (range 39 – 81,747 copies/mL). Sixteen women had suppressed viral loads, and nine women had unsuppressed viral loads (Table 3.43).

Hepatitis C virus (HCV) test results were as follows: 16 women had negative antibody results, and eight women had positive antibody results. HCV antibody data was missing for one woman. Five women had negative PCR results, and four had positive PCR results. HCV PCR data was missing for 16 women. Hepatitis B virus (HBV) surface antibody (sAb) results were negative for ten women, and positive for 12 women. Three women had missing HBV sAb data. Twenty-one women had negative HBV surface antigen (sAg) results. HBV sAg data was missing for four women. HBV core antibody (cAb) results were negative for ten women, and positive for four women. Eleven women had missing HBV cAb data (Table 3.43).

Table 3.43 HIV characteristics. Continuous variables are reported as (median) mean ± SD, range, and categorical variables are reported as n. HIV clade information and HCV antibody data were missing for one woman. HCV PCR data was missing for 16 women, HBV sAB data was missing for three women, HBV sAg data was missing for four women, and HBV cAb data was missing for 11 women.

HIV Characteristics	
HIV Clade	Clade A: 1 Clade B: 19 Clade C: 3
Mode of HIV Acquisition	Sexual Contact: 12 Drug Use: 5 Blood Products or Percutaneous: 1 Perinatal Transmission: 1 Unknown: 6
Antiretroviral Naïve	Yes: 1 No: 24
Number of Days on Antiretroviral Therapy (weeks)	(220.9) 322.4 ± 271.3 (0 – 1155)
Duration of HIV Infection (months)	(136) 150 ± 75.3 (21 – 292)
CD4 Nadir (mm^3)	(170) 172.3 ± 105.9 (20 – 440)
CD4 (at study visit) (mm^3)	(420) 450.8 ± 199.3 (90 – 870)
Highest Viral Load (copies/mL)	(75300) 142256.8 ± 189484.2 (759 – 699939)
Viral Load (at study visit) (copies/mL)	(39) 4413 ± 16660.4 (39 – 81747)
HCV Antibody	Negative: 16 Positive: 8
HCV PCR	Negative: 5 Positive: 4
HBV sAB	Negative: 10 Positive: 12
HBV sAG	Negative: 21
HBV cAB	Negative: 10 Positive: 4

Bacteriome Data

A total of 479 isotigs were assembled from 724,859 reads, and based on >55% cut-off identity to the *cpn*60 gene target; a total of 54 unique *cpn*60 OTUs were identified. No OTUs were detected in all samples, but one major genera were detected in all samples and corresponded to *Lactobacillus*, and three major genera were detected in almost all samples with

the exception of one or two individual(s); these corresponded to *G. vaginalis*, *Prevotella* and Other Bacteria.

Among women with intermediate and consistent for BV scores, fives profiles were dominated by *G. vaginalis*, one profile was dominated by *G. vaginalis* and *Lactobacillus*, one profile was dominated by *G. vaginalis* and *Megasphaera*, one was dominated by *G. vaginalis* and *Prevotella*, one was dominated by *G. vaginalis*, *Prevotella* and Other Bacteria, and three profiles were dominated by the 12 major genera present in this cohort. Furthermore as seen in Figure 3.7, eight profiles were dominated by *Lactobacillus*, two profiles were dominated by *Shigella*, one by *G. vaginalis*, one by *G. vaginalis* and *Lactobacillus*, and one by *Prevotella* (Figure 3.7). The CSTs are as follows; CST I: 5/25, CST III: 3/25, CST IVA: 4/25, CST IVC: 6/25, CST IVD: 6/25, CST V: 1/25. Additional CSTs were identified in addition to CSTs previous studies have shown [39,40] (Table 3.44).

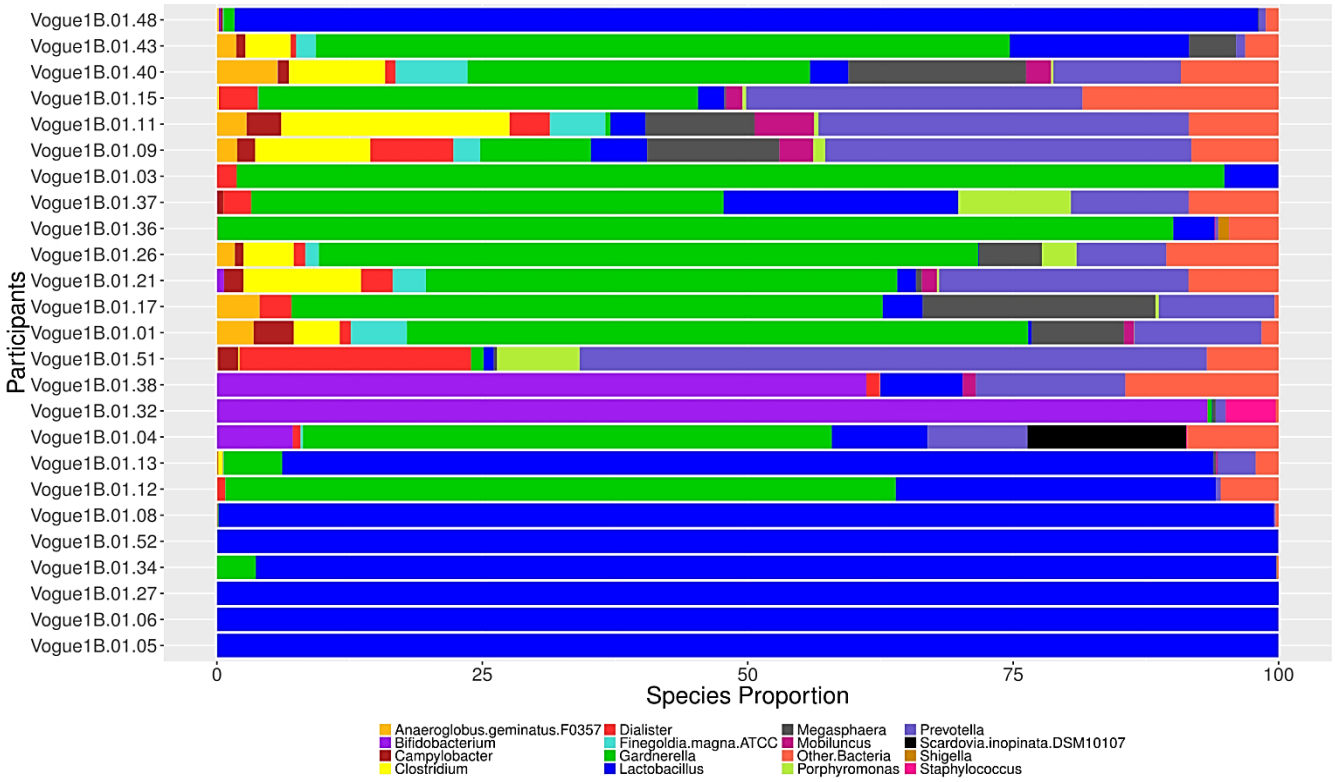

Figure 3.7 HIV-positive vaginal bacteriome profiles ordered by CST

Table 3.44 Community state types for HIV-positive cohort. Categorical variables are reported as n

Community State Types		
CST	**Dominant OTU**	**N**
I	*L.crispatus* dominated	**5**
III	*L.iners* dominated	**3**
IVA	Mixed Dominant Phenotype: *Gardnerella* subgroup B, *Bifidobacterium* sp., *Prevotella amnii, Shigella boydii, and L.delbrueckii*	**4**
IVC	*Gardnerella* A dominated	**6**
IVD	*Gardnerella* C dominated	**6**
V	*L.jensenii* dominated	**1**

Recurrent Bacterial Vaginosis Cohort

Demographic Characteristics

Our team enrolled 26 women who were attending for gynecologic care for BV-like symptoms between June 2012 and May 2015, and there were eight viral vaginal swabs available for this cohort of women. These eight women with BV-like symptoms had a mean age of 29 years (range 21 – 42 years). Self-reported ethnicity was six Caucasian and two Asian. The highest education level attained of our study population was, one did not complete high school, three received some post-secondary education, three completed post-secondary education, and one completed their graduate degree. The average BMI of this group was 20.3 kg/m^2 (range 15.5 - 27.2 kg/m^2) (Table 3.45).

Table 3.45 Demographics characteristics for recurrent BV cohort. Continuous variables are reported as mean ± SD, range, and categorical variables are reported as n

Demographics	
Age (Mean ± SD, Range)	29.2 ± 7.1 (21 – 42)
BMI (Mean ± SD, Range)	20.3 ± 3.3 (15.5 – 27.2)
Ethnicity	Caucasian: 6 Asian: 2
Highest Education Level Attained	Graduate degree: 1 Post-secondary education: 3 Some post-secondary education: 3 Did not complete high school: 1

Clinical Characteristics

Antimicrobial use was reported in half of women (4/8) in the past three months, and two women did not take any antimicrobials in the past three months; two women have missing antimicrobial use data. Of the four women using antimicrobials in the past three months, most were taking antimicrobials for BV treatment (3/4) (Table 3.46).

Since their last menstrual period, women in our cohort used different types of contraception. One woman reported using no form of contraception, one woman used hormonal contraception, two women used barriers, one woman used surgical sterilization, and two women were not sexually active; one woman has missing contraception data (Table 3.46). Most of the women in our cohort were nulliparous (5/8), and three women were multiparous (Table 3.46).

Table 3.46 Clinical characteristics for recurrent BV cohort. Categorical variables are reported as n. Contraception data was missing for one woman.

Clinical Characteristics	
Pregnancy History	Nulliparous: 5 Multiparous: 3
Form of Contraception	None: 3 Hormonal: 2 Barrier: 2
Use of Condoms (in past 48 hours)	Yes: 2 No: 6
Symptoms (in past 48 hours)	Abnormal discharge: 5 Abnormal odor: 2 Irritation or discomfort: 5 Painful intercourse: 4 Other: 0
Antimicrobial Use (in the past three months)	Yes: 4 No: 4

Mostly all women in our cohort have been previously diagnosed with a genital infection (7/8). Reported lifetime histories of genital conditions included, six reports of a history of BV, six reports of a history of yeast, four reports of UTIs, one report of genital warts, and one case of *Chlamydia trachomatis* infection. None of the women in our cohort reported a history of *Trichomonas vaginalis*, genital herpes, gonorrhea, or syphilis infections. Women experienced an average of 7.4 BV infections in their lifetime (range 0 - 20), 3.1 BV infections in the last year (range 0 - 10), and 0.7 BV infections in the past 2 months (range 0 - 2). Women also reported 8.1 yeast infections in their lifetime (range 0 - 20), 3.3 yeast infections in the last year (range 0 - 10), and 0.6 yeast infections in the past 2 months (range 0 - 2)**.** In addition, women reported 2.8 UTI infections in their lifetime (range 0 - 12), with 1.3 UTI infections in the past year (range 0 - 6) and 0.5 UTI infections in the past two months (range 0 - 3) (Table 3.47).

Table 3.47 Genital infection history for recurrent BV cohort. Continuous variables are reported as mean ± SD, range, and categorical variables are reported as n

Genital Infection History	
Bacterial Vaginosis (Mean ± SD, Range)	
Lifetime	6.5 ± 6.9 (0 – 20)
In the past year	2.8 ± 3.4 (0 – 10)
In the past two months	0.6 ± 0.7 (0 – 2)
Yeast	7/8
Urinary Tract Infections	4/8
Trichomoniasis	0/8
Genital Warts	1/8
Genital Herpes	0/8
Chlamydia	1/8
Gonorrhea	0/8
Syphilis	0/8

All women experienced BV symptoms; six women experienced abnormal discharge, two abnormal odor, and five irritation or discomfort. Fifty percent of women experience painful intercourse (4/8), and on average women experienced painful intercourse 91% of the time (range 80 - 100%); one woman had missing information on painful intercourse (Table 3.46). Nugent scores between 0 - 3 are inconsistent for BV, scores between 4 - 6 reflect intermediate BV, and scores between 7 - 10 are consistent for BV. Of out eight women, Nugent scores were one consistent with BV, five with intermediate BV, and two inconsistent with BV (Table 3.48).

Table 3.48 Nugent scores for recurrent BV cohort. Categorical variables are reported as n

Nugent Scores	
Inconsistent for BV (0 – 3)	2/8
Intermediate BV (4 – 6)	5/8
Consistent for BV (7 – 10)	1/8

Behavioural Characteristics

Tampon use was reported in seven women, and one woman reported never using tampons during their menstrual periods. With regards to feminine hygiene, one woman used

feminine wipes or genital deodorant products, and used these products in the past 48 hours (Table 3.49).

In the past year, women reported one sexual partner (range 1 - 2), and in the past two months women reported 0.86 sexual partners (range 0 -1). With regards to sexual activity, three women reported vaginal intercourse in the past 48 hours. Of our eight women, the majority reported receiving oral sex (6/8), and one reported never receiving oral sex (1/8). Two women reported receiving oral sex in the past 48 hours. One woman reported engaging in anal sex, and six reported never having anal sex. None of the women in our population had anal sex in the past 48 hours. Five women reported sex toy use, and two have never used sex toys. None of the women in our population used sex toys in the past 48 hours. Sexual activity data is missing for one woman (Table 3.49).

Reported illicit substance/alcohol use includes, one woman never using illicit substances or consuming alcohol, and six women currently using illicit substances or consuming alcohol. Of our 26 women, the most have never smoked (5/8), and one woman is a current smoker. Data on substance use is missing for one woman (Table 3.49).

Table 3.49 Behavioural characteristics for recurrent BV cohort. Continuous variables are reported as mean ± SD, range, and categorical variables are reported as n. Sexual activity and substance use data were missing for one woman.

Behavioural Characteristics	
Tampon Use (ever)	Yes: 7 No: 1
Tampon Use (in the past month)	Yes: 6 No: 2
Feminine Hygiene Products Use (ever)	Yes: 1 No: 7
Use of Feminine Hygiene Products (past 48 hours)	Yes: 1 No: 7
Substance Use	Yes: 0 No: 8
Current Smoker	Yes: 1 No: 7
Number of Sexual Partners (past year) (Mean ± SD, Range)	1.1 ± 0.4 (1 – 2)
Vaginal Intercourse (past 48 hours)	Yes: 3 No: 4
Oral Sex	Yes: 6 No: 1
Anal Intercourse	Yes: 1 No: 6
Use of Sex Toys	Yes: 5 No: 2

Bacteriome Data

A total of 426 isotigs were assembled from 35,563 reads, and based on >55% cut-off identity to the *cpn*60 gene target; a total of 61 unique *cpn*60 OTUs were identified. Seven OTUs were detected in all samples, and these correspond to *L. iners*, *G. vaginalis* Group A, *L. crispatus*, *L. jensenii*, *Prevotella timonensis*, *Clostridia sp.* probably BVAB2, *L. gasseri*. *G. vaginalis* Group C was detected in all samples, with the exception of one individual.

Among women with intermediate and consistent for BV scores, four profiles were dominated by *G. vaginalis, Lactobacillus* species or both genera. Additionally, two profiles were dominated by *Escherichia coli.*, or *Megasphaera sp.*, as can be seen in Figure 3.8. Furthermore

as seen in Figure 3.8, two women with a Nugent score inconsistent with BV, had profiles dominated by *L. crispatus*, or *L. iners*. The CSTs are as follows; CST I: 2/8, CST III: 3/8, CST IVA: 1/8, CST IVC: 1/8, CST IVD: 1/8. Additional CSTs were identified in addition to CSTs previous studies have shown [39,40] (Table 3.50).

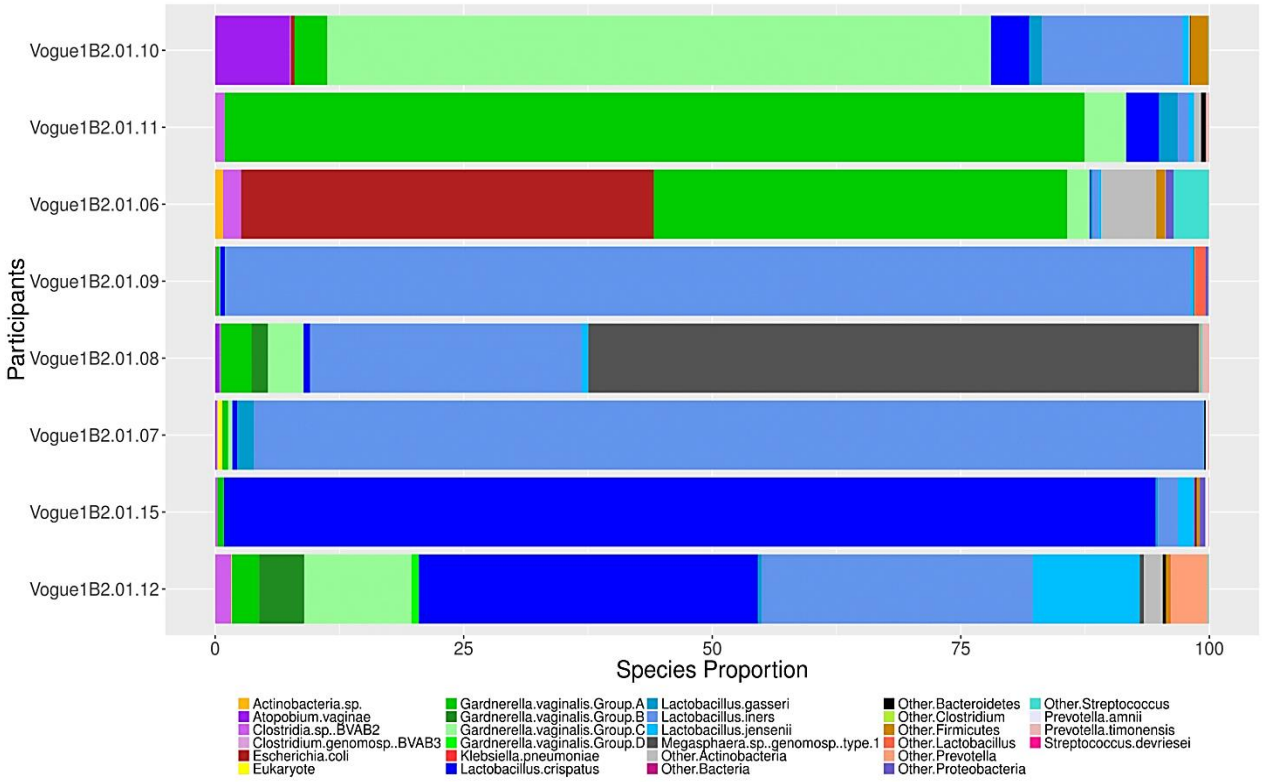

Figure 3.8 Recurrent BV vaginal bacteriome profiles ordered by CST

Table 3.50 Community state types for recurrent BV cohort. Categorical variables are reported as n

Community State Types		
CST	**Dominant OTU**	**N**
I	*L.crispatus* dominated	**2**
III	*L.iners* dominated	**3**
IVA	Mixed Dominant Phenotype: *Gardnerella* subgroup B, *Bifidobacterium* sp., *Prevotella amnii, Shigella boydii, and L.delbrueckii*	**1**
IVC	*Gardnerella* A dominated	**1**
IVD	*Gardnerella* C dominated	**1**
V	*L.jensenii* dominated	**0**

Virome Results

Hierarchical Clustering

Cluster analyses were conducted to identify how individual viral profiles grouped together, which allowed for the construction of viral groups that were related to demographic, behavioural and clinical variables. Viral profiles clustered into seven groups (Groups I to VII) (Figure 3.9). Group I was dominated by several *Alphapapillomavirus* species (5, 6, 9 and 14), as well as unclassified *Papillomaviridae* species. Group II was dominated by *Alphapapillomavirus 3* and *Coccolithovirus*. Group III was a mixed dominance cluster, containing *Acanthocystis turfacea chlorella virus canal-1*, *Alphapapillomavirus 8*, *Cronobacter phage CR5*, *Paramecium bursaria chlorella virus NY2A* and unclassified *Siphoviridae* species. Group IV was dominated by *Alphapapillomavirus 3*, and unclassified *Papillomaviridae* species. Group V was dominated by phage species; *Lactobacillus phage LL-H*, *Lactobacillus phage phi jlb1* and *Lactobacillus phage phiadh*. Group VI was largely dominated by human herpesvirus 4, as well as the *Lymphocryptovirus* genera and unclassified *Siphoviridae* species. Lastly, Group VII was dominated by *Gammaretrovirus* and *Podoviridae*, where the exact species could not be classified (Table 3.51).

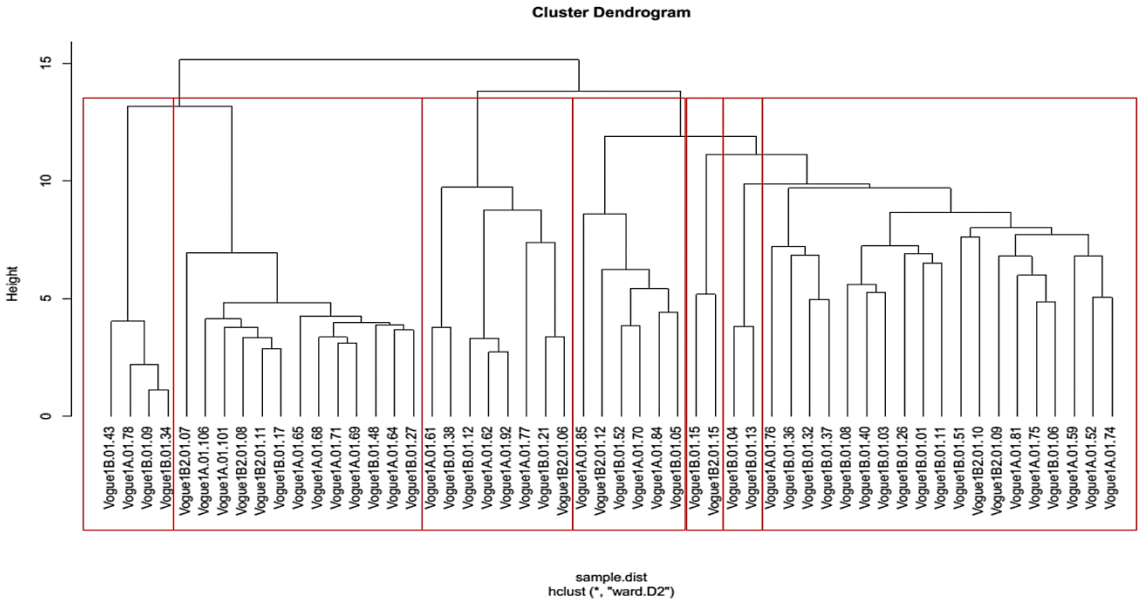

Figure 3.9 Cluster dendrogram tree cut into seven distinct viral groups

Table 3.51 Viral and phage species which were dominant in each one of our seven viral groups. Values are average abundance of each species within each viral group, and reported as percentage (%)

	Viral Groups						
Viral Species	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**
Acanthocystis turfacea chlorella virus canal-1	.	.	38.56	.	.	.	.
Alphapapillomavirus 14	14.71	.	.	.	.	.	.
Alphapapillomavirus 3	.	19.22	.	81.81	.	.	.
Alphapapillomavirus 5	15.78	.	.	.	.	.	.
Alphapapillomavirus 6	11.89	.	.	.	.	.	.
Alphapapillomavirus 8	.	.	19.14	.	.	.	.
Alphapapillomavirus 9	22.45	.	.	.	.	.	.
Cronobacter phage CR5	.	.	37.16	.	.	.	.
Human herpesvirus 4	.	.	.	.	.	70.45	.
Lactobacillus phage LL-H	.	.	.	.	25.39	.	.
Lactobacillus phage phi jlb1	.	.	.	.	11.38	.	.
Lactobacillus phage phiadh	.	.	.	.	31.89	.	.
Coccolithovirus	.	15.88	.	.	.	.	.
Gammaretrovirus	.	.	.	.	.	.	35.91
Lymphocryptovirus	.	.	.	.	.	11.25	.
Unclassified *Papillomaviridae* species	26.47	.	.	10.48	.	.	.
Unclassified *Podoviridae* species	.	.	.	.	.	.	67.72
Unclassified *Siphoviridae* species	.	.	11.22	.	.	11.58	.
Paramecium bursaria chlorella virus NY2A	.	.	12.9	.	.	.	.

Papillomaviridae was abundant in our cohorts, and 82% of participants fell into the four of the seven viral groups dominated by *Papillomaviridae* species. *Papillomaviridae* was removed to determine how viral profiles would then cluster. Viral profiles clustered into seven groups, all of which were mixed dominant clusters largely made up of different phage species (Table 3.52).

Table 3.52 Viral and phage species which were dominant in each one of our seven viral groups, in the absence of *Papillomaviridae*. Values are average abundance of each species within each viral group, and reported as percentage (%)

	Viral Groups						
Viral Species	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**
Acanthocystis turfacea chlorella virus canal-1	.	.	.	.	45.26	.	.
Bacillus phage PBC1	.	.	.	.	.	.	61.41
Corynebacterium phage P1201	.	.	14.63	.	.	.	.
Cronobacter phage CR5	.	.	19.26	.	.	.	.
Human herpesvirus 4	.	10.7	.	.	.	.	27.52
Human herpesvirus 5	.	.	.	.	.	.	15.42
Lactobacillus phage LL-H	.	.	.	.	.	50.9	.
Lactobacillus phage phi jlb1	.	.	.	.	.	14.81	.
Lactobacillus phage phiadh	12.15	.	.	.	.	44.38	.
Lactobacillus phage phig1e	.	10.01	.	.	.	.	.
Lactobacillus prophage Lj771	16.76	.	.	.	.	.	.
Merkel cell polyomavirus	.	.	.	.	15.84	.	.
Molluscum contagiosum virus	10.37	.	.	.	.	.	.
Other *Alphabaculovirus*	.	.	.	.	16.43	.	.
Other *Chlorovirus*	.	.	.	.	32.09	.	.
Other *Coccolithovirus*	14.79	23.77	.	.	.	.	.
Other *Enterovirus*	.	.	.	10.18	.	.	.
Other *Gammaretrovirus*	.	.	.	36.96	.	.	.
Other *Podoviridae*	.	.	.	63.05	.	.	.
Other *Siphoviridae*	15.17	.	29.55	.	.	.	.
Pseudomonas phage PPpW-4	.	.	.	.	10.66	.	.
Staphylococcus phage phiMR11	.	26.74	.	.	.	.	.
Streptococcus phage PH15	.	.	10.46	.	.	.	.
Streptococcus phage phiNJ2	.	.	15.82	.	.	.	.
Wuhan ant virus	.	.	.	.	.	.	22.84

Virome Characterization

In all three cohorts, from our 54 samples 97,493,869 reads were sampled and 29.23% were classified. 50,050,970 human reads were identified, 7.15% of human reads were classified and 92.85% could not be classified. 24,810,812 bacterial reads were identified, and 80.20% were classified. 5,464,755 viral reads were identified, and 90.75% were classified. 122,229 phage reads were identified, and 33.08% were classified. 69,643 fungal reads were identified, and 20.34% were classified. Slightly over one percent (1.06%) of total reads identified were labelled 'ambiguous', when reads cannot be distinguished between two or more taxonomic categories, and 16.34% of total reads identified could not be classified to any taxonomic category and were labelled 'unknown'.

With regards to viral and phage reads, within participants 51,731 reads were identified on average, of which on average 46,292 were classified. A total of 477 species were detected. Six species were detected in all samples, and corresponded to *Alphapapillomavirus 3*, *6* and *10*, human herpesvirus 4, *Coccolithovirus*, and unclassified *Papillomaviridae* species. DNA viruses were the most prevalent making up 96% of the viruses we detected, of which 74% were papillomaviruses (Figure 3.10, 3.11). When specifically looking at the abundance of human herpesviruses (HHV), HPV and HIV in our populations, *Alphapapillomavirus* were generally more abundant than HHV and HIV (Figure 3.12). Overall, separate viral patterns could not be seen within each bacteriome CST, as well as when *Lactobacillus* dominated, *L. iners* dominated and heterogeneous profiles were compared (Figure 3.13, 3.14). Viral profiles were viewed in relation to Nugent scores, and the majority of women within viral group V had Nugent scores consistent for BV (Figure 3.15, 3.16). This was not seen in other viral groups. Viral groups were as follows; Group I: 8, Group II: 13, Group III: 19, Group IV: 4, Group V: 6, Group VI: 2, and Group VII: 2 (Table 3.53).

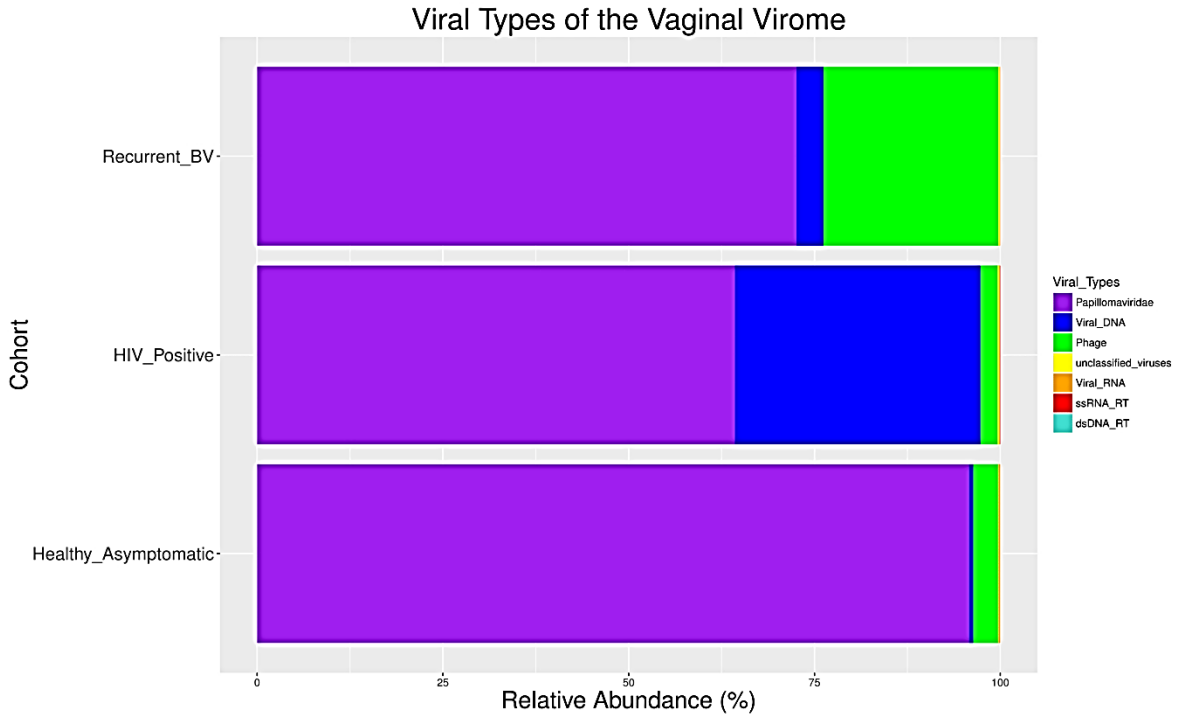

Figure 3.10 Viral types detected in the healthy-asymptomatic, HIV-positive and recurrent BV cohorts

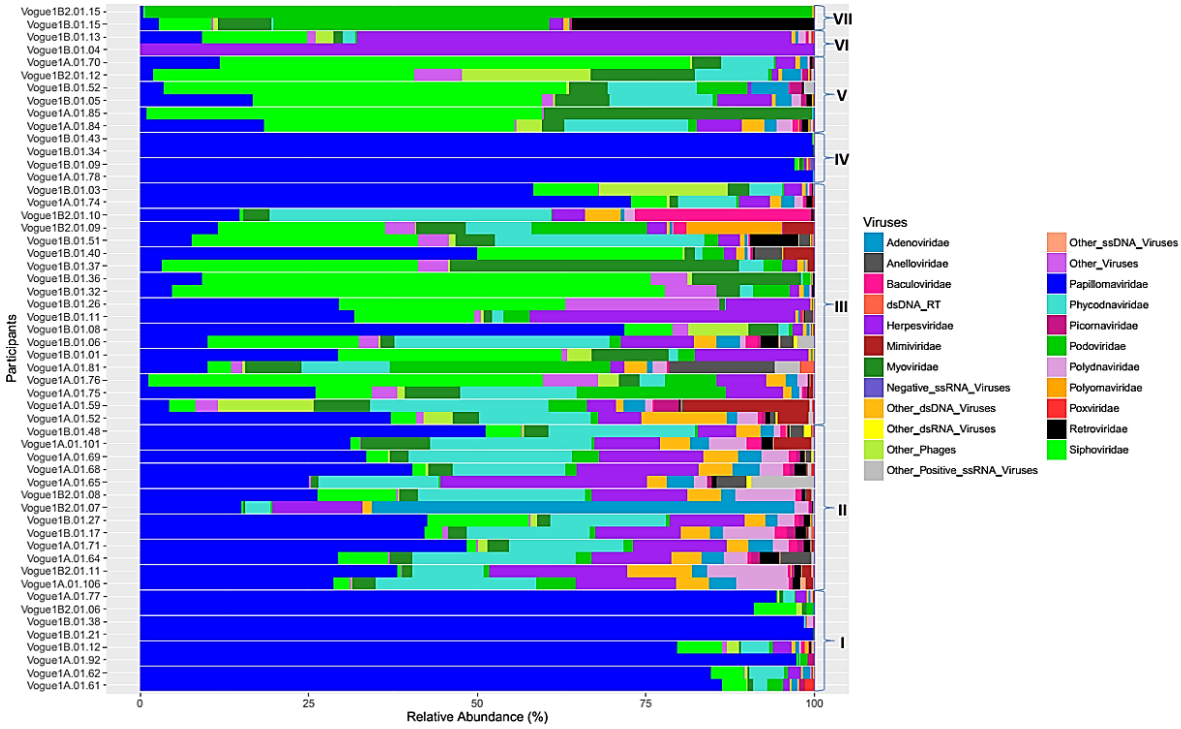

Figure 3.11 Viral profiles for healthy-asymptomatic (1A), HIV-positive (1B) and recurrent BV (1B2) cohorts. Profiles ordered based on viral groups

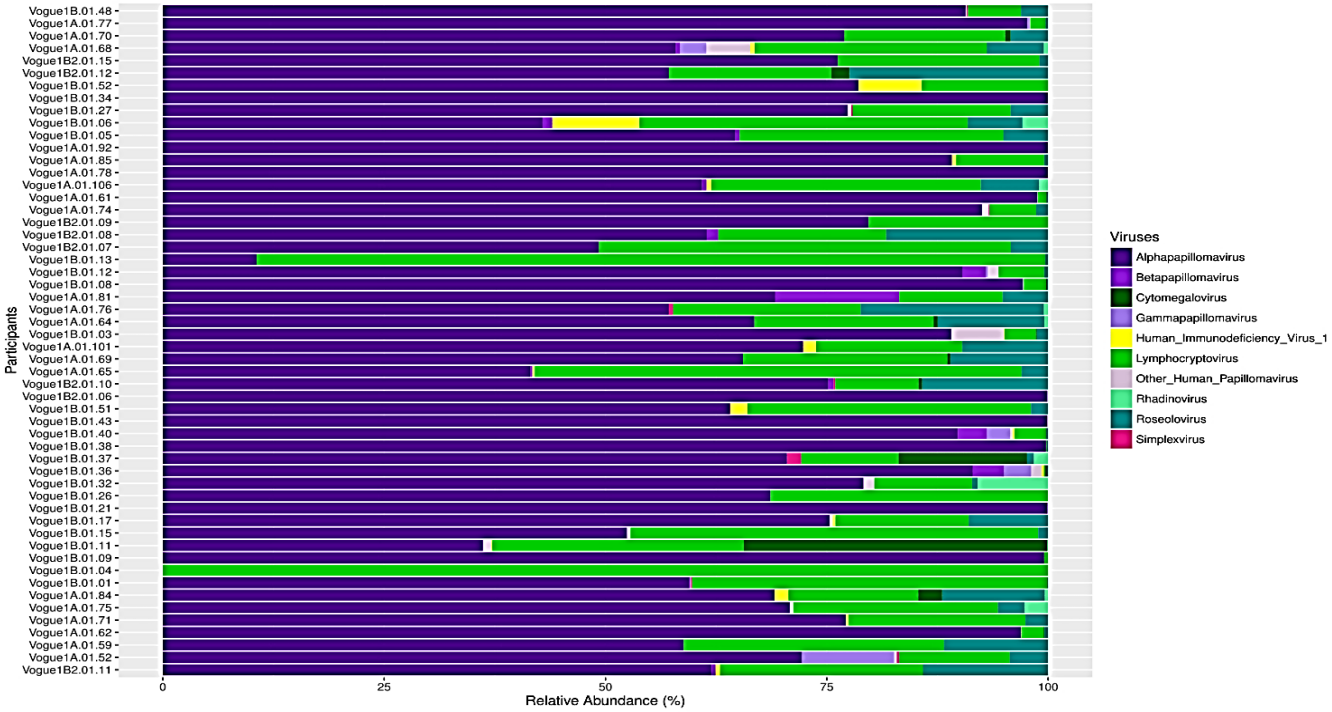

Figure 3.12 HPV, HHV and HIV for healthy-asymptomatic (1A), HIV-positive (1B) and recurrent BV (1B2) cohorts

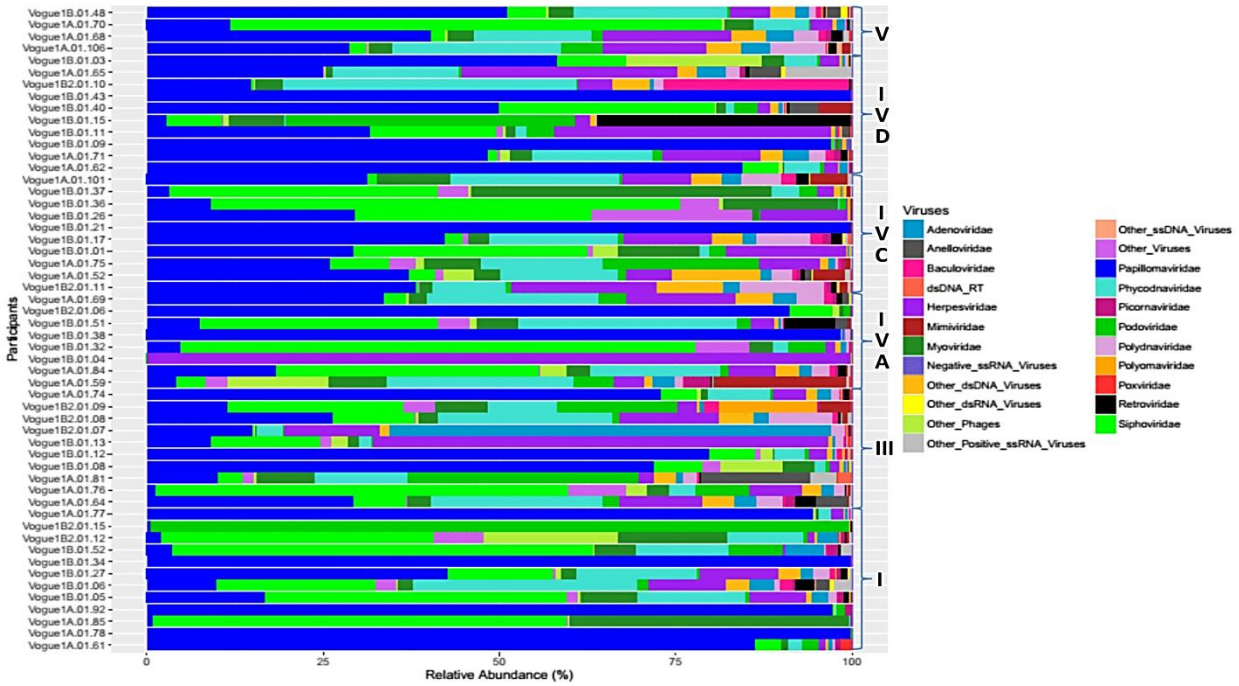

Figure 3.13 Vaginal virome profiles for healthy-asymptomatic (1A), HIV-positive (1B) and recurrent BV (1B2) cohorts ordered by bacteriome CSTs

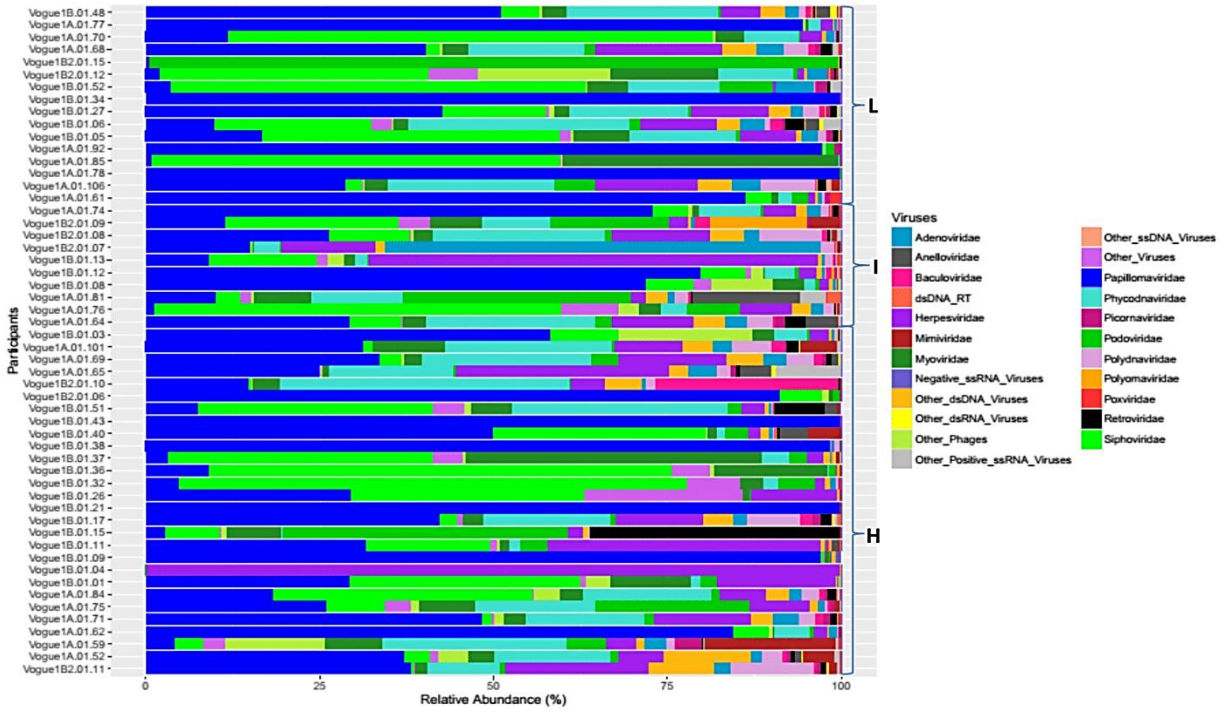

Figure 3.14 Vaginal virome profiles for healthy-asymptomatic(1A), HIV-positive(1B) and recurrent BV(1B2) cohorts grouped. *Lactobacillus* dominated profiles (L), *L. iners* dominance (I) and heterogeneous profiles (H)

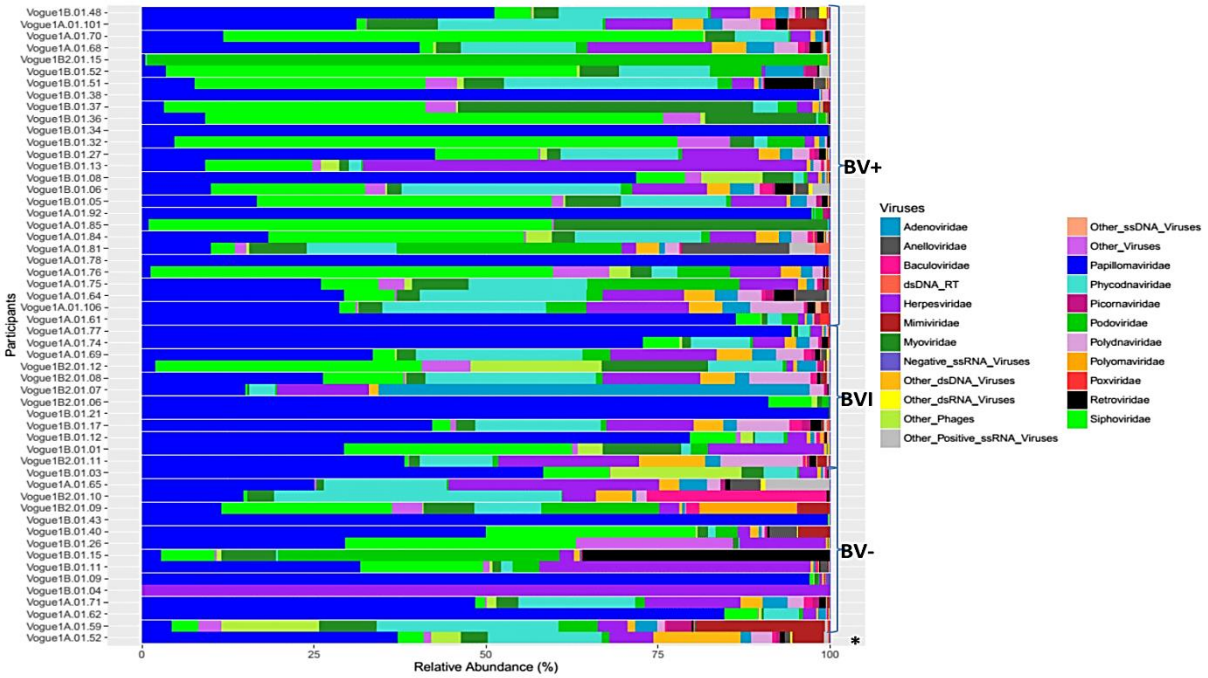

Figure 3.15 Vaginal virome profiles for healthy-asymptomatic (1A), HIV-positive (1B) and recurrent BV (1B2) cohorts ordered based on Nugent scores. Nugent scores consistent for BV (BV+), Nugent scores for intermediate BV (BVI), and Nugent scores inconsistent for BV

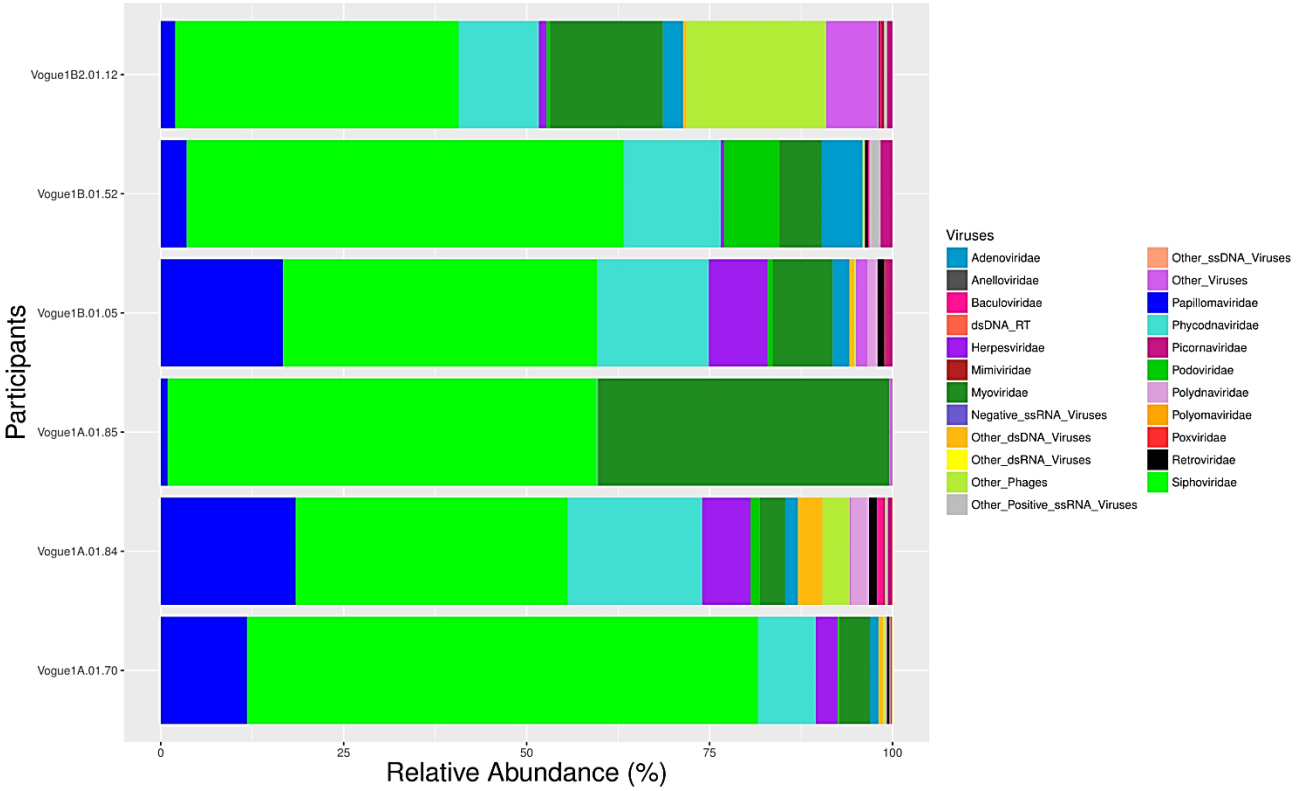

Figure 3.16 Healthy-asymptomatic(1A), HIV-positive(1B) and recurrent BV(1B2) individuals with viral profiles which fall into viral group V; dominated by *Lactobacillus* phage species. Nugent scores consistent for BV (BV+) and Nugent scores for intermediate BV (BVI)

Table 3.53 Viral groups for healthy-asymptomatic, HIV-positive and recurrent BV cohorts. Categorical variables are reported as n

Viral Groups		
Viral Group	**Viral Species Present**	**N**
I	*Alphapapillomavirus 5, 6, 9* and *14*, and unclassified *Papillomaviridae* species	**8**
II	*Alphapapillomavirus 3* and *Coccolithovirus*	**13**
III	*Acanthocystis turfacea chlorella virus canal-1*, *Alphapapillomavirus 8*, *Cronobacter phage CR5*, *Paramecium bursaria chlorella virus NY2A* and unclassified *Siphoviridae* species	**19**
IV	*Alphapapillomavirus 3*, and unclassified *Papillomaviridae* species	**4**
V	*Lactobacillus phage LL-H, Lactobacillus phage phi jlb1* and *Lactobacillus phage phiadh*	**6**
VI	Human herpesvirus 4, *Lymphocryptovirus* genera and unclassified *Siphoviridae* species	**2**
VII	*Gammaretrovirus* and *Podoviridae*	**2**

Healthy-Asymptomatic Cohort

In our healthy-asymptomatic cohort, 38,219 viral and phage reads were sampled, and 35,333 reads were classified on average in each participant. A total of 251 species were identified. Thirteen species were found in all participants, and corresponded to *Alphapapillomavirus 3, 6, 10* and *14*, *Coccolithovirus*, *Glypta fumiferanae ichnovirus*, human herpesvirus 4 an 6A, human mastadenovirus D, *Lymphocryptovirus*, unclassified *Papillomaviridae* species, *Pandoravirus salinus*, and *Staphylococcus virus G1*. The majority of profiles were dominated by the DNA viruses; five profiles were dominated by *Papillomaviridae*, and one profile was dominated by *Siphoviridae* and *Myoviridae*. There were 11 mixed dominant profiles, which included different abundances of *Papillomaviridae*, *Siphoviridae*, *Phycodnaviridae*, *Herpesviridae*, *Myoviridae*, *Podoviridae*, *Polydnaviridae Adenoviridae* and other dsDNA viruses. There were four profiles similar to the mixed dominant profiles, where three profiles also included *Mimiviridae*, and one profile had *Anelloviridae* (Figure 3.17). In this cohort, the viral groups were as follows: Group I: 4, Group II: 7, Group III: 6, Group IV: 1 and Group V: 3 (Table 3.54).

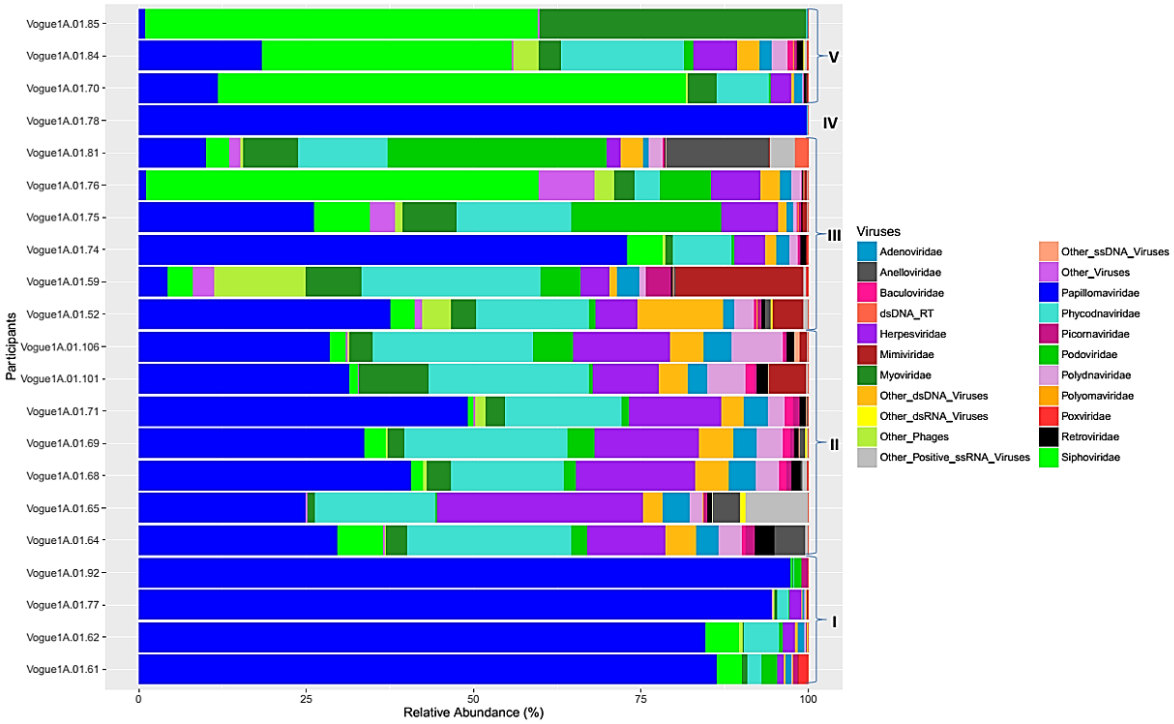

Figure 3.17 Viral profiles for healthy-asymptomatic cohort. Profiles ordered based on viral groups.

Table 3.54 Viral groups for healthy-asymptomatic cohort. Categorical variables are reported as n

Viral Groups		
Viral Group	**Viral Species Present**	**N**
I	*Alphapapillomavirus 5, 6, 9* and *14*, and unclassified *Papillomaviridae* species	**4**
II	*Alphapapillomavirus 3* and *Coccolithovirus*	**7**
III	*Acanthocystis turfacea chlorella virus canal-1*, *Alphapapillomavirus 8*, *Cronobacter phage CR5*, *Paramecium bursaria Chlorella virus NY2A* and unclassified *Siphoviridae* species	**6**
IV	*Alphapapillomavirus 3*, and unclassified *Papillomaviridae* species	**1**
V	*Lactobacillus phage LL-H, Lactobacillus phage phi jlb1* and *Lactobacillus phage phiadh*	**3**
VI	Human herpesvirus 4, *Lymphocryptovirus* genera and unclassified *Siphoviridae* species	**0**
VII	*Gammaretrovirus* and *Podoviridae*	**0**

The majority of women with *L. crispatus* dominated profiles (CST I) had viral profiles dominated by *Papillomaviridae* species (Figure 3.18). This pattern was specific to *L. crispatus* dominated profiles, and was not seen in other *Lactobacillus*-dominated profiles (Figure 3.19). No other patterns were seen within bacteriome CSTs, as well as between *L. iners* dominated, and heterogeneous profiles. Viral profiles were viewed in relation to Nugent scores, and women with Nugent scores consistent for BV had profiles largely dominated by phages species, compared to women with Nugent scores inconsistent for BV or intermediate BV (Figure 3.20). The abundance of HPV, HHV and HIV were looked at within our healthy-asymptomatic cohort. In the majority of women, HPV genera *Alphapapillomavirus* was more abundant that the HHV and HIV, except in one participant where *Lymphocryptovirus* (predominantly human herpesvirus 4) was more abundant (Figure 3.21). *Simplexvirus* genera (HSV -1, -2) were not seen in this population, and HIV was.

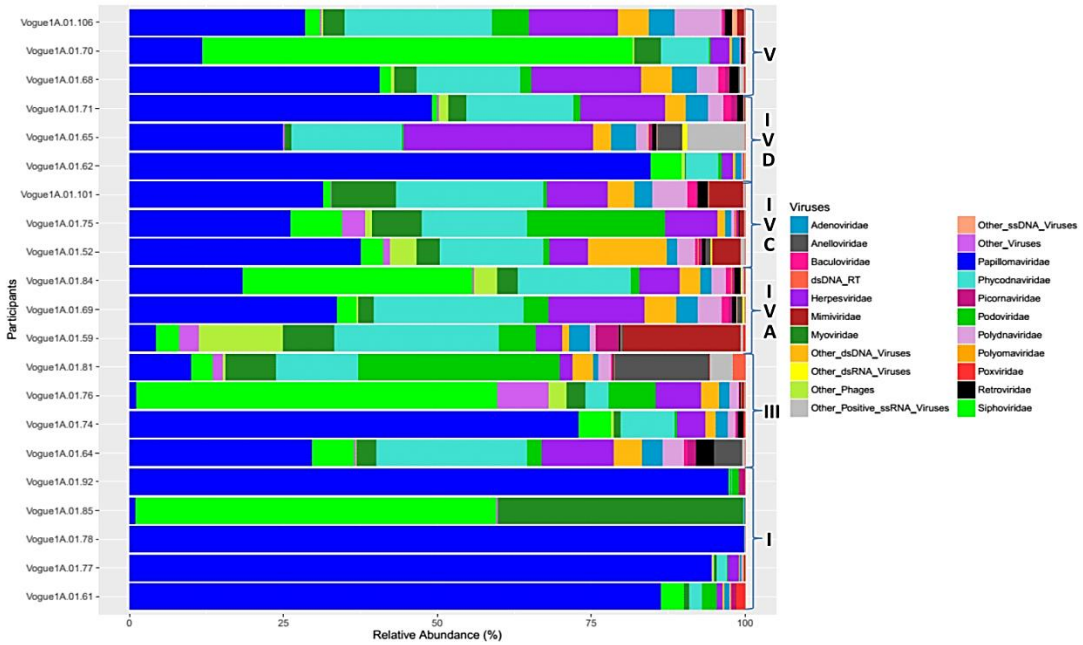

Figure 3.18 Vaginal virome profiles for healthy-asymptomatic ordered by bacteriome CSTs

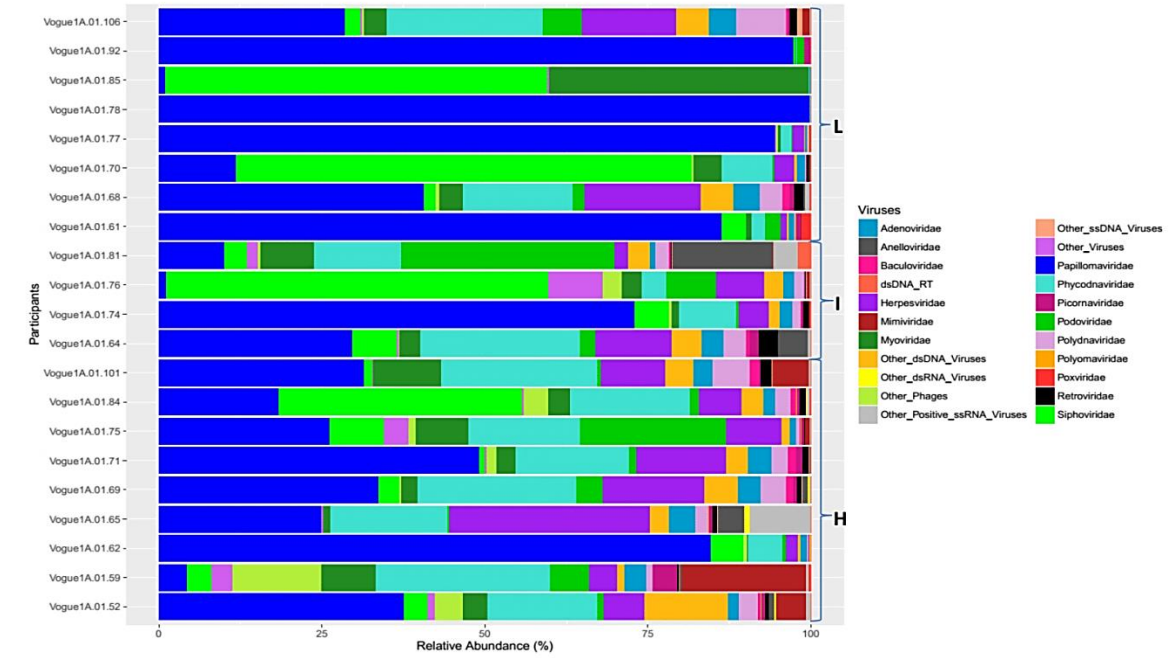

Figure 3.19 Vaginal virome profiles for healthy-asymptomatic cohort grouped. *Lactobacillus* dominated profiles (L), *L. iners* dominance (I) and heterogeneous profiles (H)

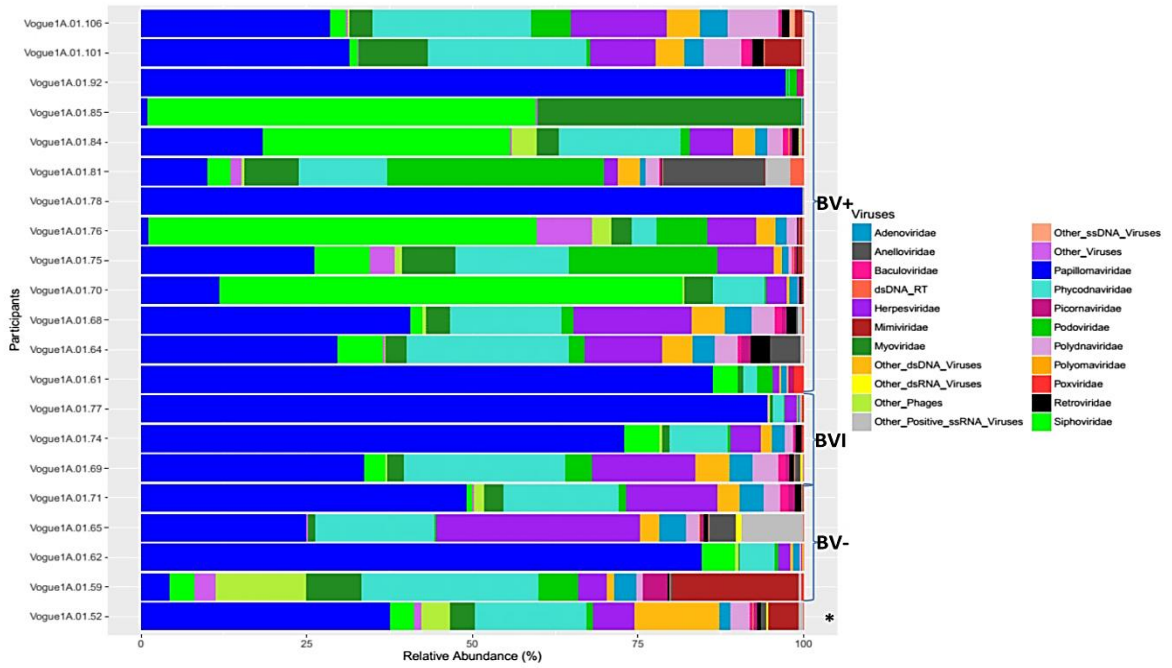

Figure 3.20 Vaginal virome profiles for healthy-asymptomatic cohort ordered based on Nugent scores. Nugent scores consistent for BV (BV+), Nugent scores for intermediate BV (BVI), Nugent scores inconsistent for BV (BV-). * missing Nugent score

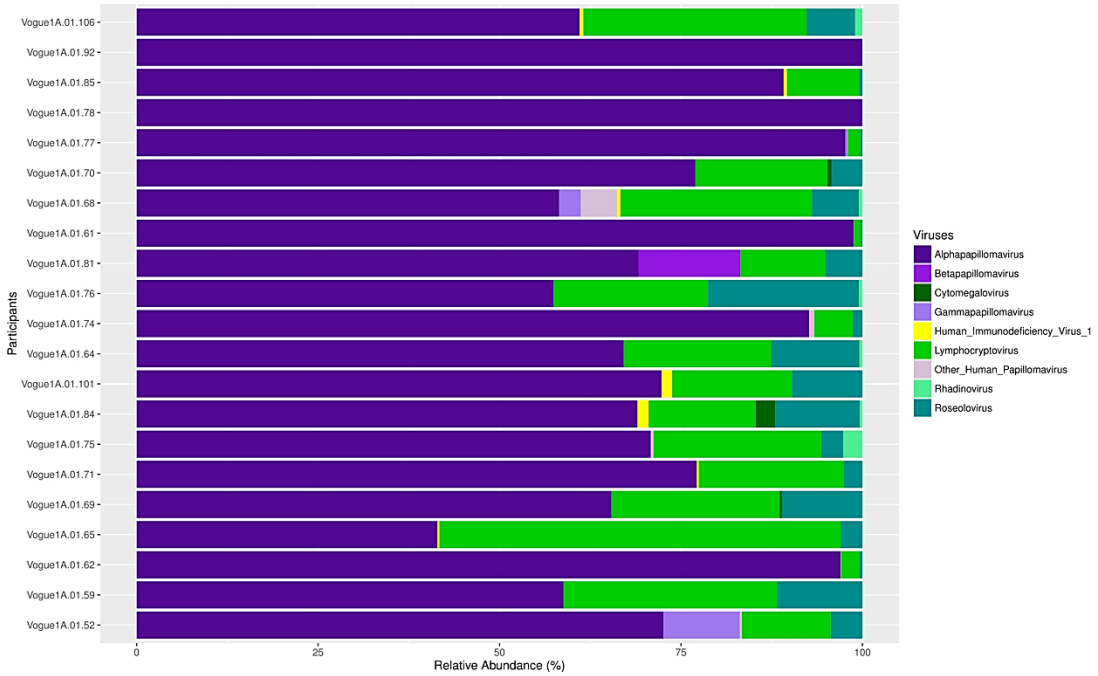

Figure 3.21 HPV, HHV and HIV for healthy-asymptomatic cohort

HIV-Positive Cohort

In our HIV-positive population, 76,566 viral and phage reads were sampled, and 67,494 were classified on average in each participant. A total of 339 species were detected. Seven species were detected in all participants, and corresponded to *Alphapapillomavirus 3, 6, 9* and *10*, *Coccolithovirus*, human herpesvirus 4, and unclassified *Papillomaviridae* species. Profiles were dominated by DNA viruses, and/or phage species. Five profiles were dominated by *Papillomaviridae*, and one profile was dominated by *Herpesviridae*. There were 18 mixed dominant profiles, which included different abundances of *Papillomaviridae*, *Siphoviridae*, *Phycodnaviridae*, *Herpesviridae*, *Myoviridae* and *Podoviridae*. Additionally, in several profiles *Polydnaviridae*, *Retroviridae*, other viruses and phages species were present (Figure 3.22). Three women fell into Group I, II and IV. Eleven women fell into Group III, two women fell into Group V and VI, and one woman fell into Group VII (Table 3.55).

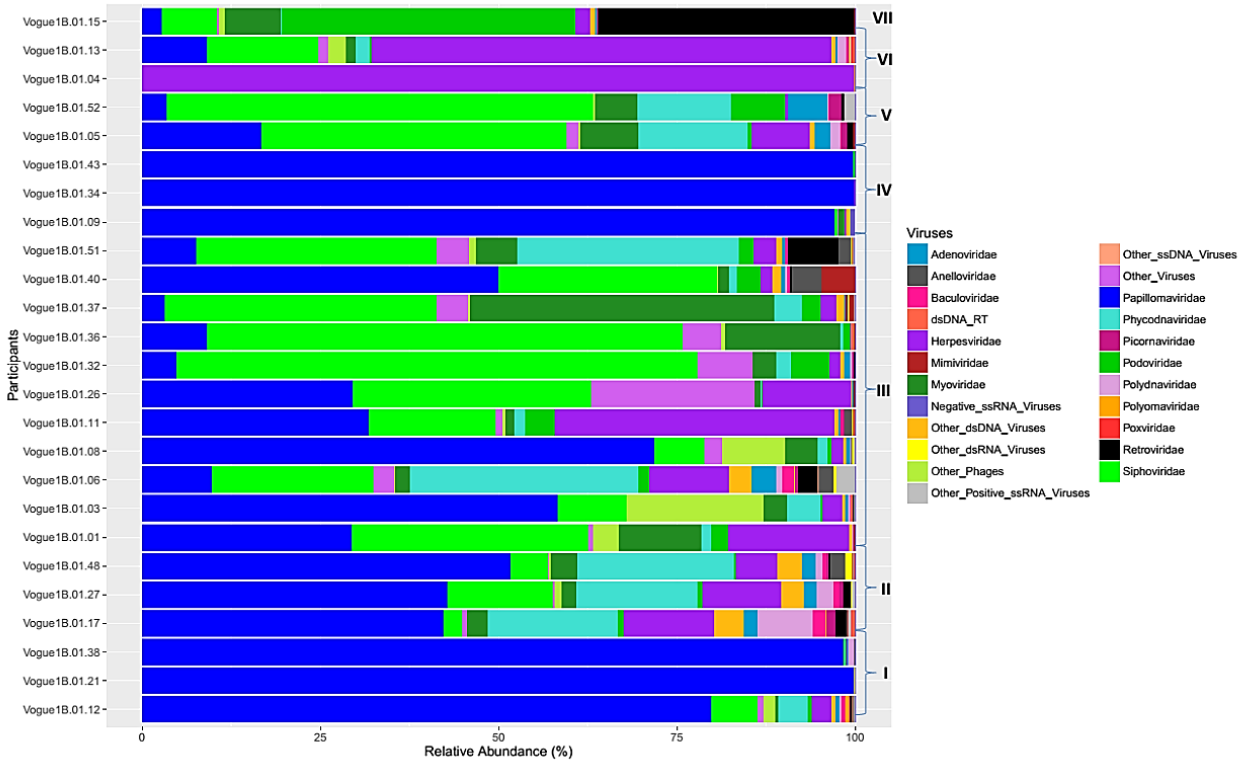

Figure 3.22 Viral profiles for HIV-positive cohort. Profiles ordered based on viral groups.

Table 3.55 Viral groups for HIV-positive cohort. Categorical variables are reported as n

Viral Groups		
Viral Group	**Viral Species Present**	**N**
I	*Alphapapillomavirus 5, 6, 9* and *14*, and unclassified *Papillomaviridae* species	**3**
II	*Alphapapillomavirus 3* and *Coccolithovirus*	**3**
III	*Acanthocystis turfacea chlorella virus canal-1*, *Alphapapillomavirus 8*, *Cronobacter phage CR5*, *Paramecium bursaria chlorella virus NY2A* and unclassified *Siphoviridae* species	**11**
IV	*Alphapapillomavirus 3*, and unclassified *Papillomaviridae* species	**3**
V	*Lactobacillus phage LL-H, Lactobacillus phage phi jlb1* and *Lactobacillus phage phiadh*	**2**
VI	Human herpesvirus 4, *Lymphocryptovirus* genera and unclassified *Siphoviridae* species	**2**
VII	*Gammaretrovirus* and *Podoviridae*	**1**

In this cohort, distinct viral patterns were not seen within each bacteriome CST, Nugent score categories, nor when *Lactobacillus* dominated, *L. iners* dominated and heterogeneous profiles were compared (3.23, 3.24). No viral patterns were seen between the Nugent score categories (3.25).

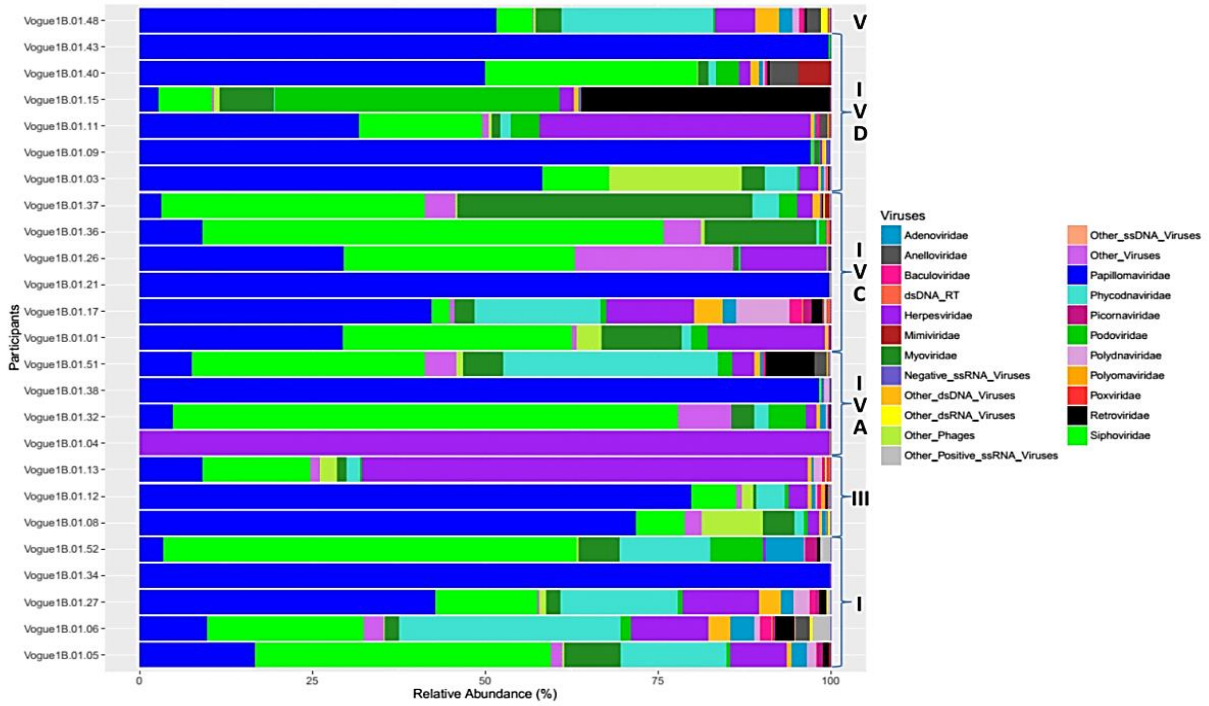

Figure 3.23 Vaginal virome profiles for HIV-positive cohort ordered by bacteriome CSTs

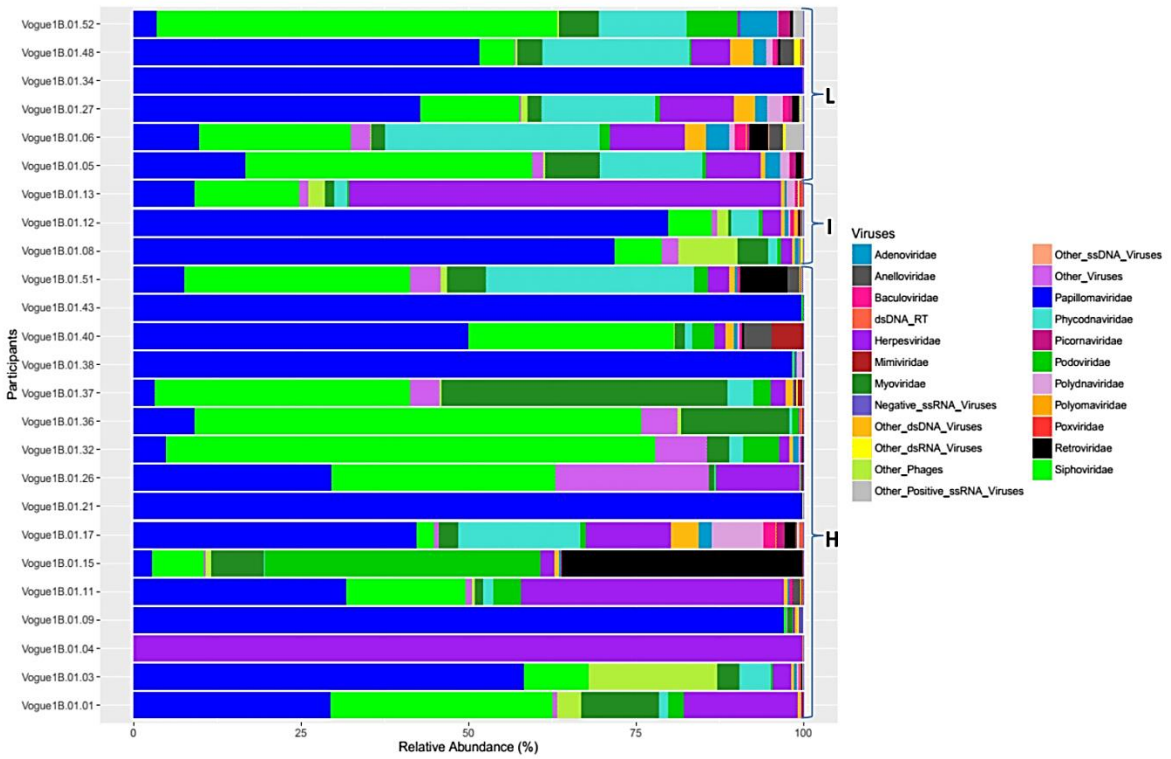

Figure 3.24 Vaginal virome profiles for HIV-positive cohort grouped. *Lactobacillus* dominated profiles (L), *L. iners* dominance (I) and heterogeneous profiles (H)

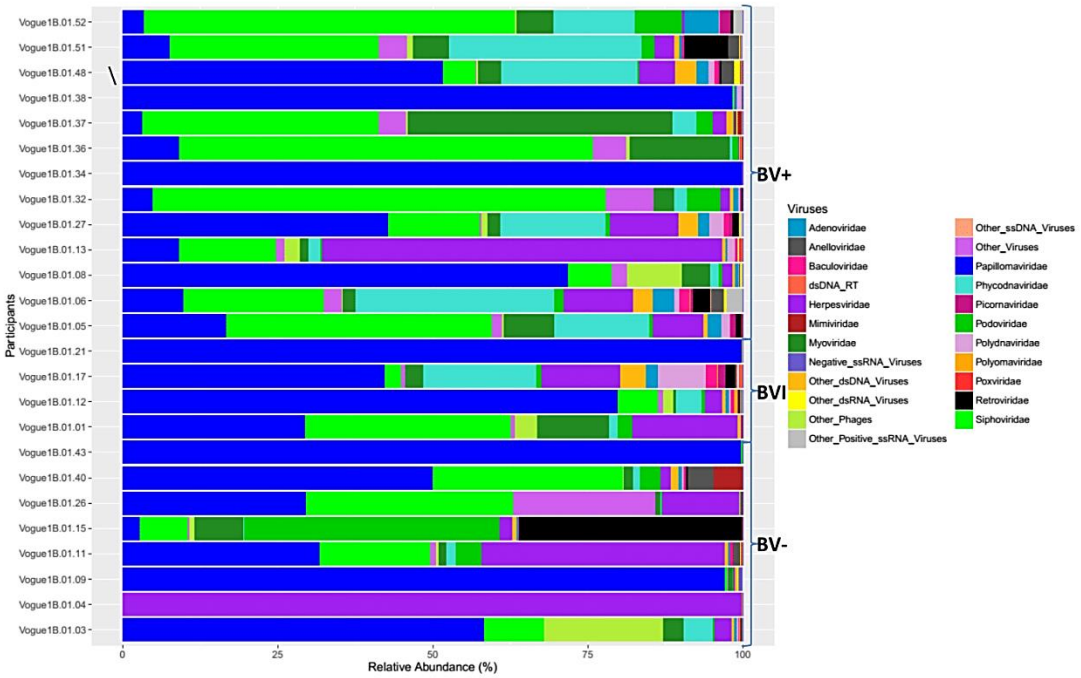

Figure 3.25 Vaginal virome profiles for HIV-positive cohort ordered based on Nugent scores. Nugent scores consistent for BV (BV+), Nugent scores for intermediate BV (BVI), Nugent scores inconsistent for BV (BV-)

Viral profiles were also viewed based on plasma HIV viral load. The majority of women with suppressed viral load (< 40 copies/mL) fell into Groups I - VII; 69% (11/16) of women had profiles which fell into Groups I - IV, and 31% (5/16) of women had profiles which fell into Groups V - VII (Figure 3.26). Women with unsuppressed viral loads (> 400 copies/mL) had and women with low level viral loads (40 - 400 copies/mL) had profiles which fell into Groups I - IV; predominately Group III (Figure 3.27, 3.28). *Papillomaviridae* was more abundant in low level and unsuppressed viral load groups.

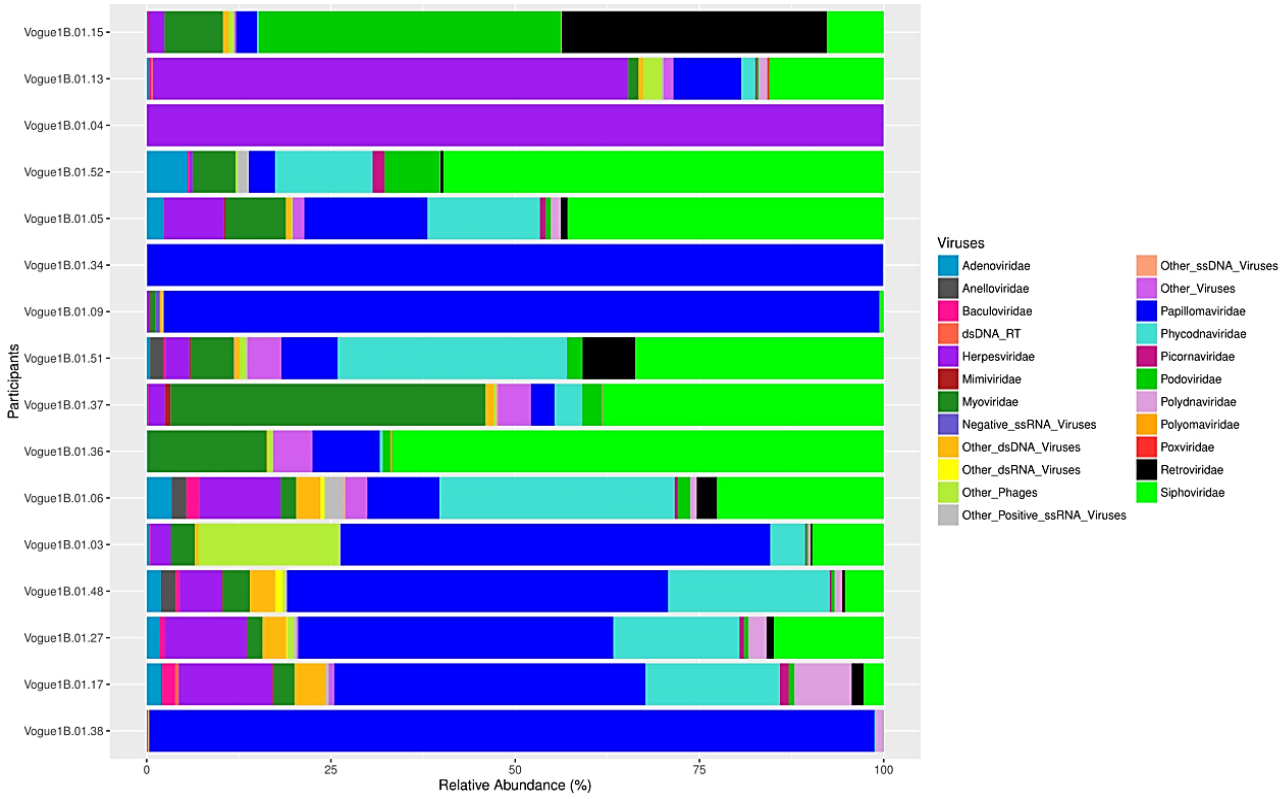

Figure 3.26 Vaginal virome profiles of suppressed women (viral load <40 copies/mL) in HIV-positive cohort. Profiles ordered based on viral groups

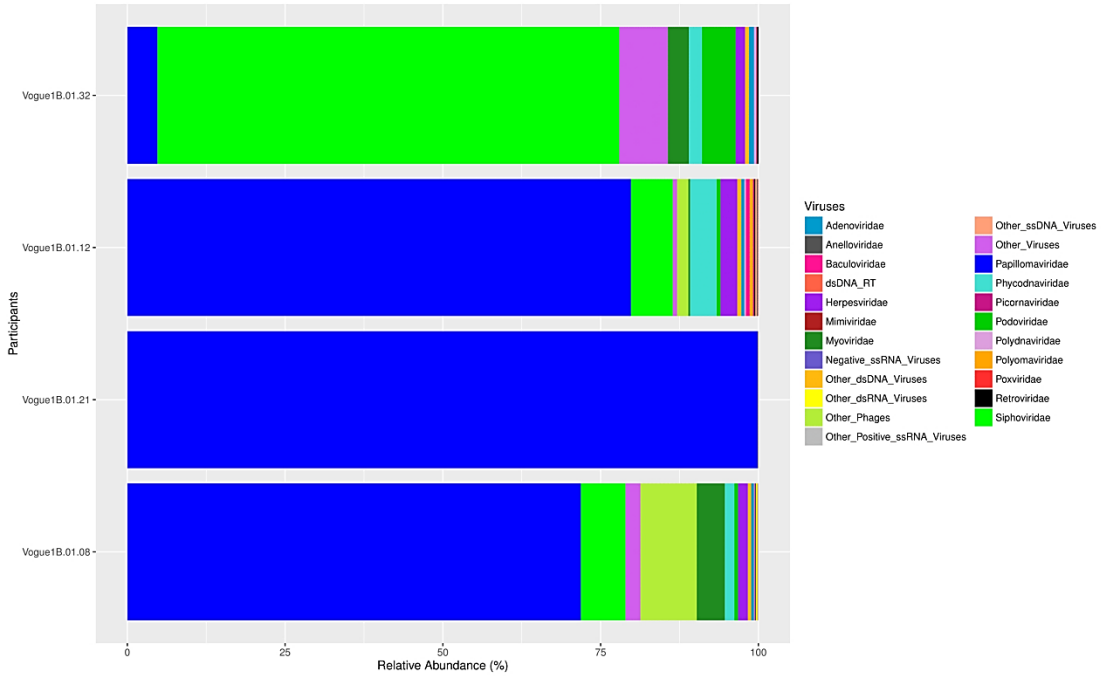

Figure 3.27 Vaginal virome profiles of women with low level viral loads (40 - 400 copies/mL) in HIV-positive cohort. Profiles ordered based on viral load in descending order from top to bottom

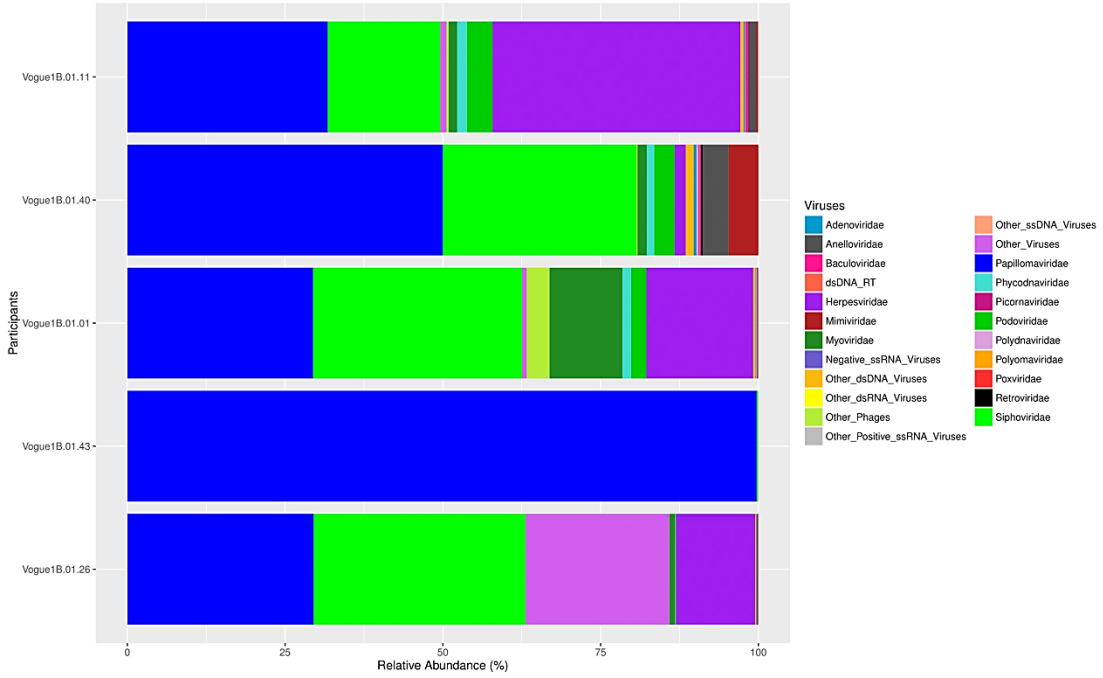

Figure 3.28 Vaginal virome profiles of unsuppressed women (viral load >400 copies/mL) in HIV-positive cohort. Profiles ordered based on viral load in descending order from top to bottom

A trend between greater *Papillomaviridae* abundance and increasing bacterial diversity in the low level viral load group was detected (Figure 3.29). This trend was not seen in suppressed and unsuppressed viral load groups (Figure 3.30, 3.31).

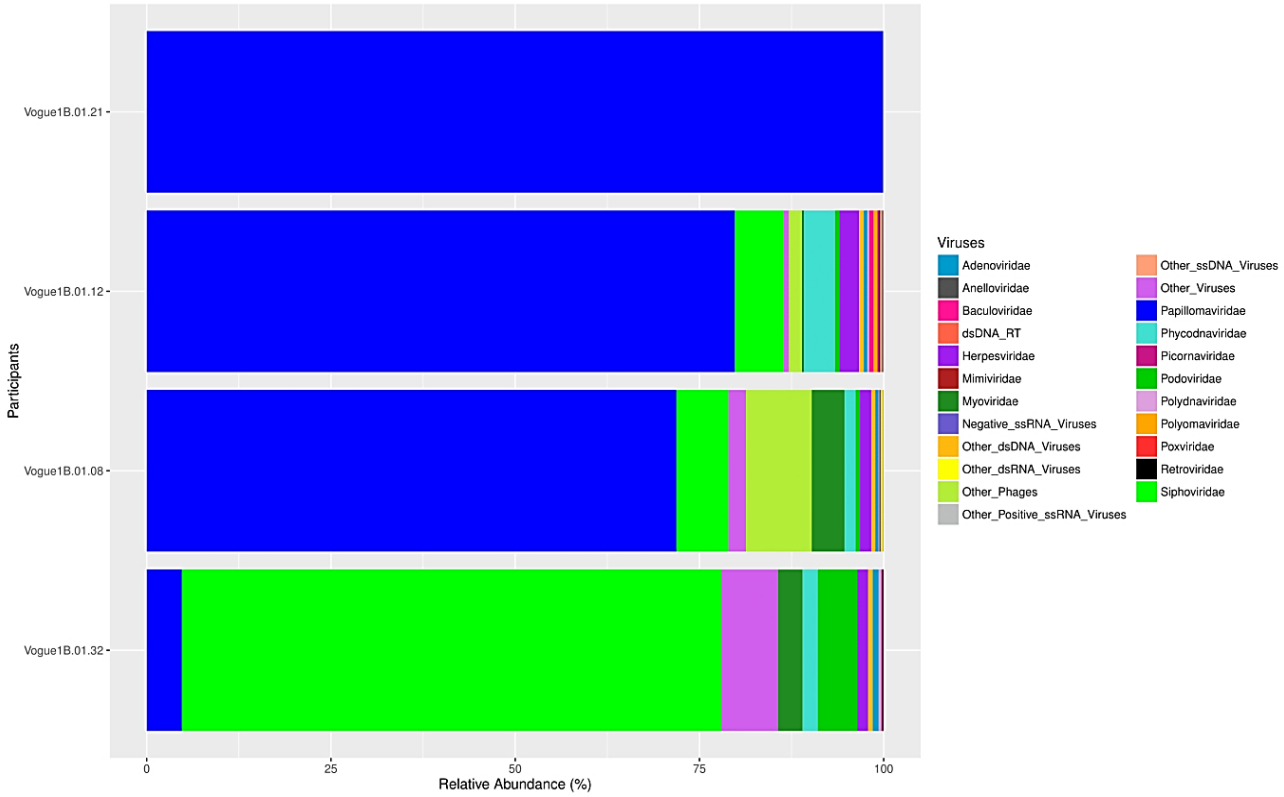

Figure 3.29 Vaginal virome profiles of women with low level viral loads (40 - 400 copies/mL) in HIV-positive cohort. Profiles ordered based on Shannon's Diversity Index (bacterial diversity) in descending order from top to bottom

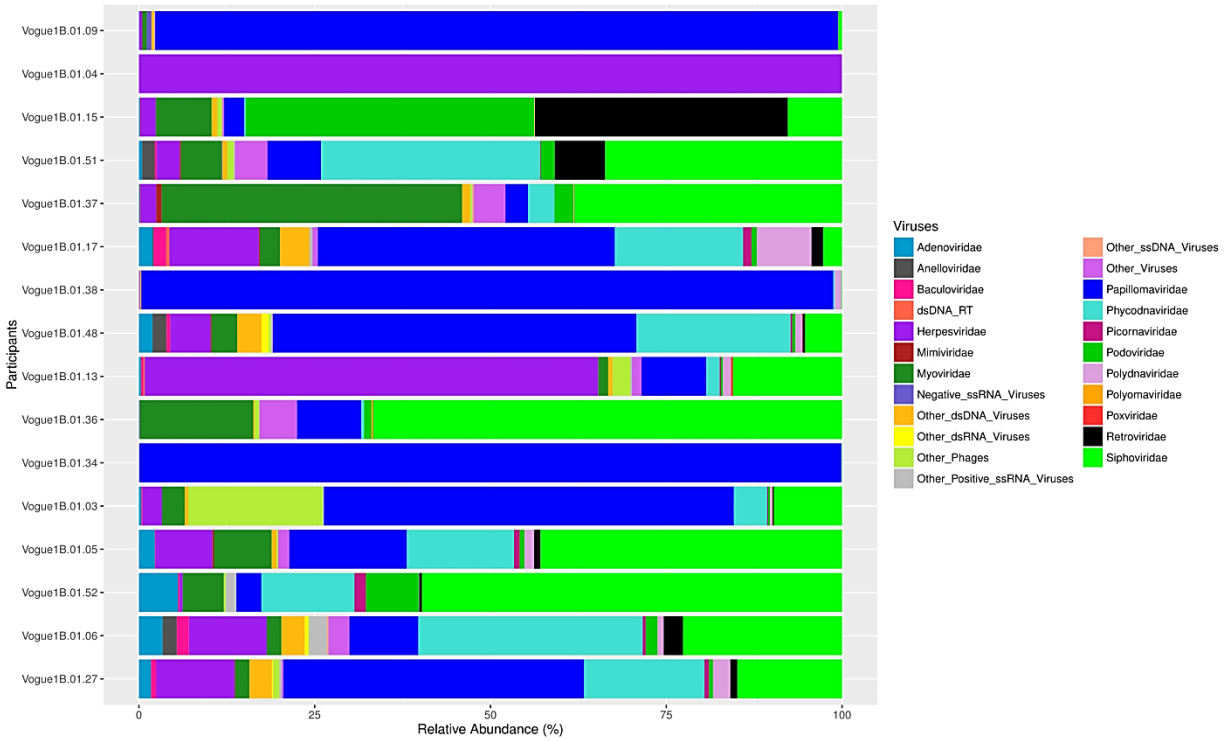

Figure 3.30 Vaginal virome profiles of suppressed women (viral load <40 copies/mL) in HIV-positive cohort. Profiles ordered based on Shannon's Diversity Index (bacterial diversity) in descending order from top to bottom

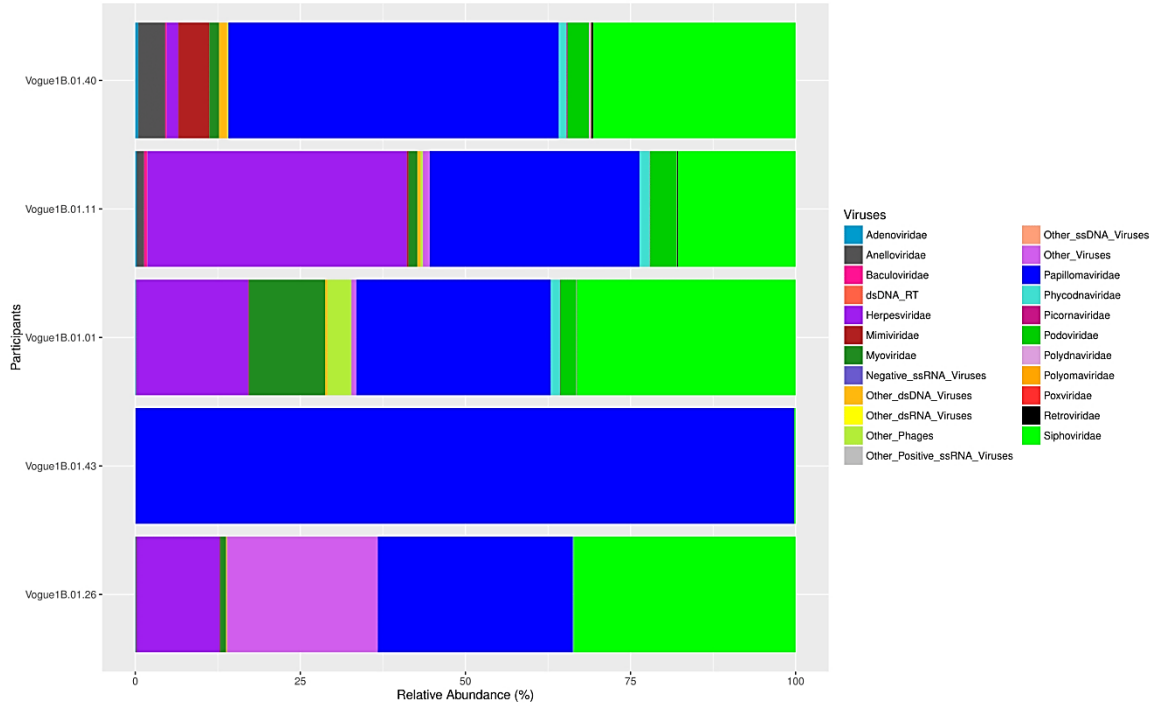

Figure 3.31 Vaginal virome profiles of unsuppressed women (viral load >400 copies/mL) in HIV-positive cohort. Profiles ordered based on Shannon's Diversity Index (bacterial diversity) in descending order from top to bottom

Profiles were viewed in the absence of *Papillomaviridae*, and the majority of profiles from all three viral load groups were mainly dominated by *Herpesviridae*, *Phycodnaviridae*, and phage species; *Siphoviridae*, *Myoviridae* and *Podoviridae* and other phage species. Women with low level viral loads, showed a pattern of *Siphoviridae* dominance with increasing viral loads (Figure 3.32). This pattern was not seen in the unsuppressed viral group (Figure 3.33, 3.34).

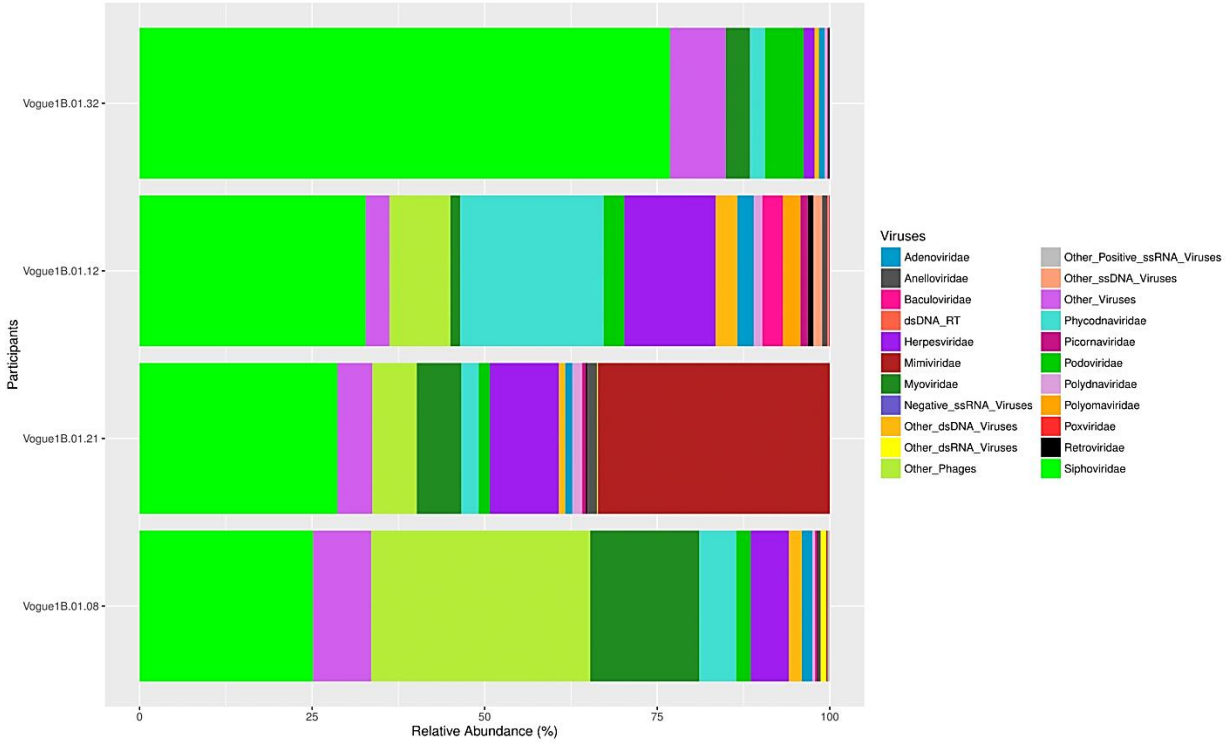

Figure 3.32 Vaginal virome profiles of women with low level viral loads (40 - 400 copies/mL) in HIV-positive cohort, in absence of *Papillomaviridae*. Profiles ordered based on viral load in descending order from top to bottom

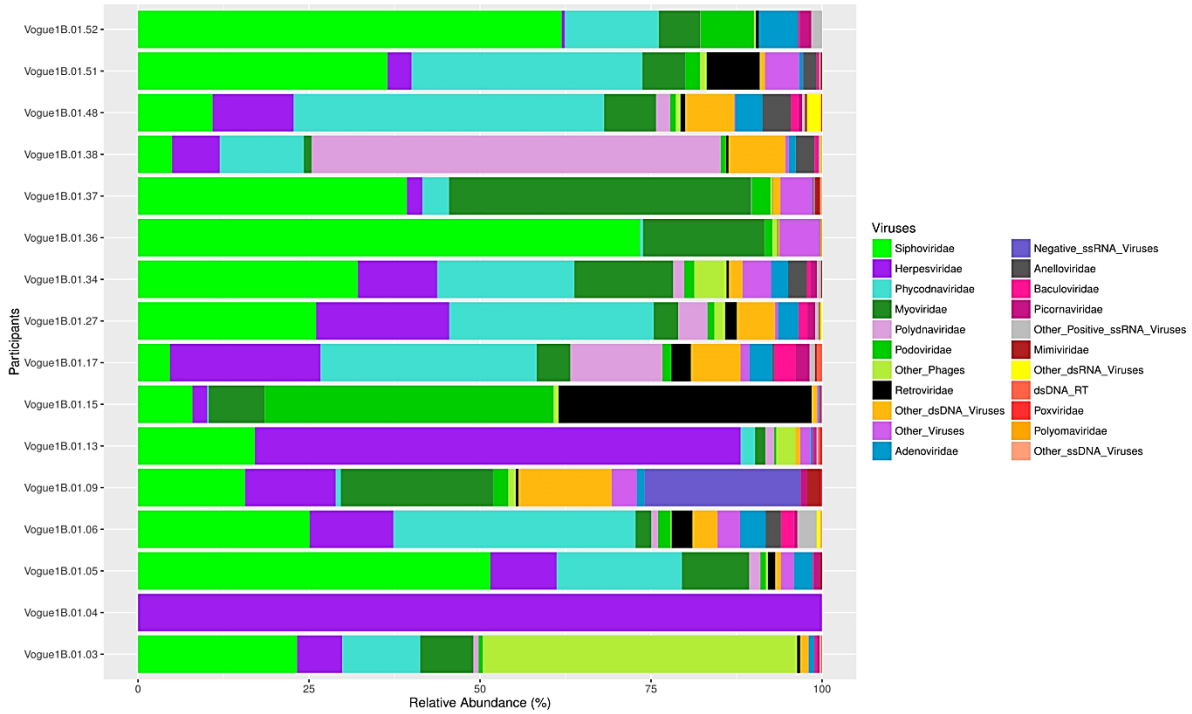

Figure 3.33 Vaginal virome profiles of suppressed women (viral loads <40 copies/mL) in HIV-positive cohort, in absence of *Papillomaviridae*. Profiles ordered based on viral load in descending order from top to bottom

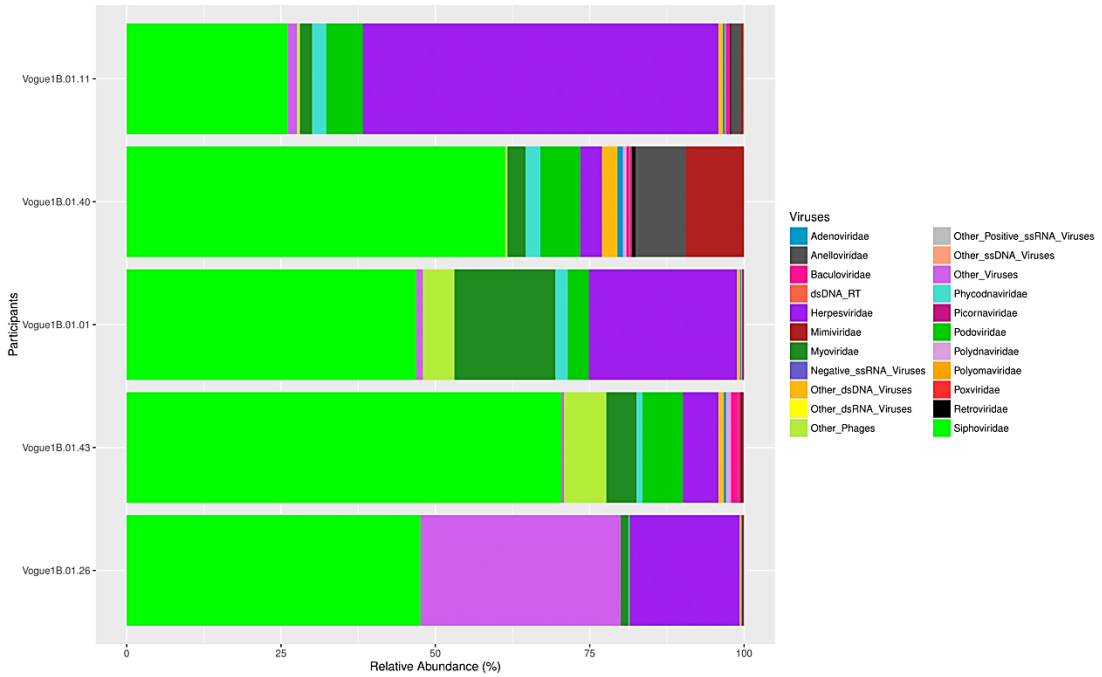

Figure 3.34 Vaginal virome profiles of unsuppressed women (viral loads >400 copies/mL) in HIV-positive cohort, in absence of *Papillomaviridae*. Profiles ordered based on viral load in descending order from top to bottom

A number of HIV-positive participants were also enrolled in an alternative study directed by our team, looking at HPV in this population. HPV type data is available for these individuals (Table 3.56). Several women positive for HPV types in our alternative study, showed an abundance of reads associated with those same HPV types (Table 3.56).

Table 3.56 HPV information for HIV-positive population. Viral reads detected in this study in a subset of HIV-positive women, with corresponding HPV type data collected from the HPVinHIV study. HPV types in rows, and participants in columns. Values in red indicate HPV types which women were found to be positive for in HPVinHIV study. * Participant 12 and 36 also positive for HPV54, and Participant 51 positive for HPV26. Numerical variables are reported as n

	HPV Type Classification														
	1	**3**	**4**	**8**	**9**	**11**	**12***	**15**	**17**	**21**	**26**	**34**	**36***	**38**	**51***
HPV-16	0	0	**0**	0	0	0	0	0	0	0	0	0	0	**0**	0
HPV-18	0	0	0	0	0	0	0	0	0	0	0	0	**199**	0	0
HPV-33	0	0	0	0	0	0	**247**	0	0	1	0	0	**1249**	1	0
HPV-35	0	0	0	0	0	0	0	0	**2**	0	0	0	0	0	0
HPV-42	0	0	0	0	1	**441**	0	0	0	699	0	0	0	**7**	0
HPV-45	0	0	0	0	0	0	1	1	4	2	5	2	**38**	1	0
HPV-51	**0**	0	0	0	0	0	0	0	0	47	0	1	**2**	0	0
HPV-52	5	5	4	4	6	5	5	2	4	**81353**	14	16	2	4	4
HPV-53	1	0	**0**	0	1	1	0	1	1	4	2	1	0	0	0
HPV-56	13	6	9	9	6	5	8	12	16	**299198**	23	17	**12**	**15**	4
HPV-61	0	2	0	0	1	**300**	1	**3**	3	0	0	**43**	**38**	**146**	**0**
HPV-62	7	12	**7**	**114**	6	26	5	38	**25**	63	18	346351	**1645**	49	**10**
HPV-66	0	0	3	0	0	0	0	0	0	772	0	0	**0**	0	**0**
HPV-67	0	0	0	0	0	**75**	**845**	0	0	0	0	0	0	0	0
HPV-68	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0
HPV-68b	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0
HPV-70	305	**87**	1	**22**	1	1	1	2	1	**13838**	1	4	1	**161**	0
HPV-71	1	2	2	0	0	3	0	3	7	4	2	5	1	**67293**	3
HPV-72	3	**7**	1	4	7	167	5	7	20	47	23	26	5	144	2
HPV-72b	3	**14**	4	9	10	131	6	4	24	16	25	71	9	167	6
HPV-73	146	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0
HPV-81	0	0	0	1	1	41	1	0	0	417	**751**	22	**117**	**2803**	0
HPV-83	1	0	0	3	**15638**	5	3	0	1	2	2	13	7	9	2
HPV-84	0	**5**	0	0	1	4	0	0	0	0	0	6	**0**	1	0
HPV-89	21	46	27	109	137	23	17	31	68	23	55	127	59	47	**27**

HPV genera *Alphapapillomavirus* was more abundant that HHV and HIV, except in a few cases where *Lymphocryptovirus* (predominantly human herpesvirus 4) was more abundant (Figure 3.35). *Simplexvirus* genera (HSV-1, -2) and HIV were seen in this population.

Figure 3.35 HPV, HHV and HIV for HIV-positive cohort

Recurrent BV Cohort

In our recurrent BV population, 9,592 viral and phage reads were sampled, and 8,799 classified on average in each participant. A total of 173 species were identified. Thirteen species were detected in all participants, and correspond to *Alphapapillomavirus 3, 6, 9, 10* and *14*, *Coccolithovirus*, *Glypta fumiferanae ichnovirus*, human herpesvirus 4, human mastadenovirus D, *Lymphocryptovirus*, *Pandoravirus salinus*, and unclassified *Papillomaviridae* and *Siphoviridae* species. Profiles were dominated by DNA viruses, and/or phage species. One profile was dominated by *Papillomaviridae*, and one profile was dominated by *Podoviridae*. One profile was dominated mainly by *Adenoviridae*, also containing *Papillomaviridae*, *Phycodnaviridae* and *Herpesviridae*. Four profiles were mixed dominance consisting of

Papillomaviridae, *Podoviridae*, *Phycodnaviridae*, *Siphoviridae*, *Herpesviridae*, *Myoviridae*, *Baculoviridae* and other dsDNA viruses. The remaining profile was dominated by a mixture of *Phycodnaviridae*, *Siphoviridae*, *Myoviridae*, and other unclassified phage and viral species (Figure 3.36). Viral groups in this cohort were as follows: one woman in Group I, V and VII, three women in Group II, and two women in Group III (Table 3.57).

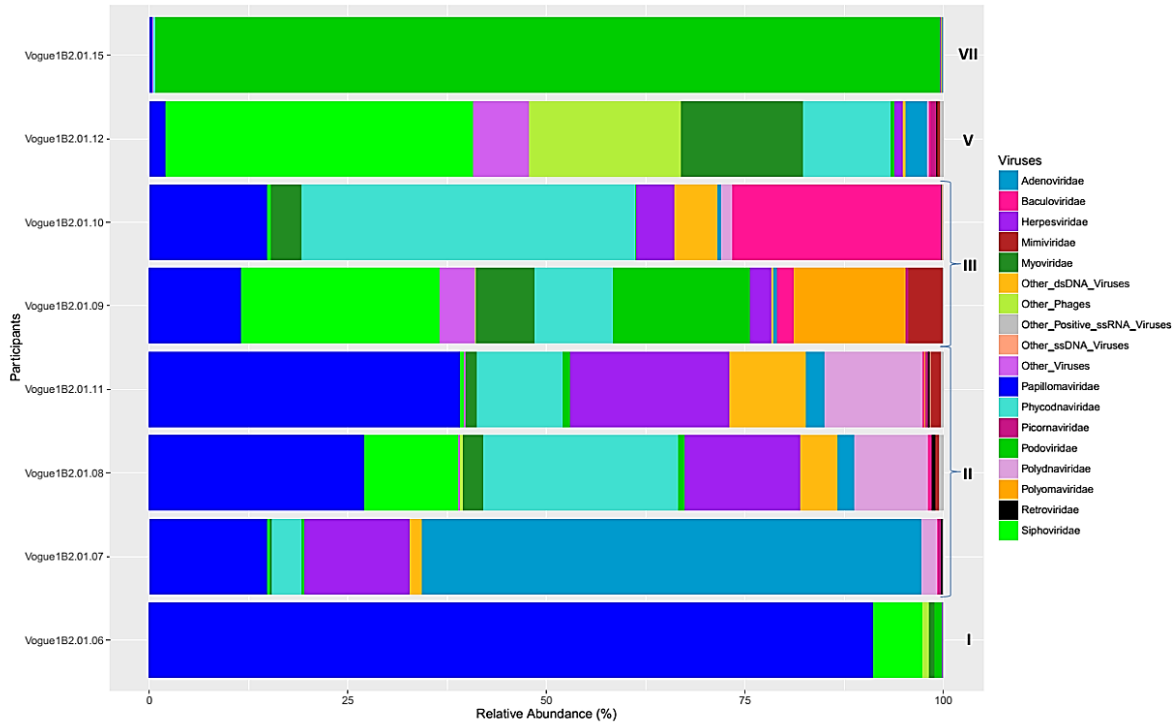

Figure 3.36 Viral profiles for recurrent BV cohort. Profiles ordered based on viral groups

Table 3.57 Viral groups for recurrent BV cohort. Categorical variables are reported as n

Viral Groups		
Viral Group	**Viral Species Present**	**N**
I	*Alphapapillomavirus 5, 6, 9* and *14*, and unclassified *Papillomaviridae* species	**1**
II	*Alphapapillomavirus 3* and *Coccolithovirus*	**3**
III	*Acanthocystis turfacea chlorella virus canal-1*, *Alphapapillomavirus 8*, *Cronobacter phage CR5*, *Paramecium bursaria chlorella virus NY2A* and unclassified *Siphoviridae* species	**2**
IV	*Alphapapillomavirus 3*, and unclassified *Papillomaviridae* species	**0**
V	*Lactobacillus phage LL-H*, *Lactobacillus phage phi jlb1* and *Lactobacillus phage phiadh*	**1**
VI	Human herpesvirus 4, *Lymphocryptovirus* genera and unclassified *Siphoviridae* species	**0**
VII	*Gammaretrovirus* and *Podoviridae*	**1**

Lactobacillus-dominated profiles (CSTI) were dominated by phage species, and did not have an abundance of *Papillomaviridae*. In comparison, *L. iners* dominated profiles (CST III) had a greater abundance of *Papillomaviridae*, and heterogenous profiles (CST IVA, IVC, IVD) were the most abundant in *Papillomaviridae* (Figure 3.37, 3.38). One woman had a Nugent score consistent for BV, and her viral profile was dominated by *Podoviridae*, and women with Nugent scores inconsistent for BV and for intermediate BV had viral profiles dominated by *Papillomaviridae*, as well as phage species (Figure 3.39).

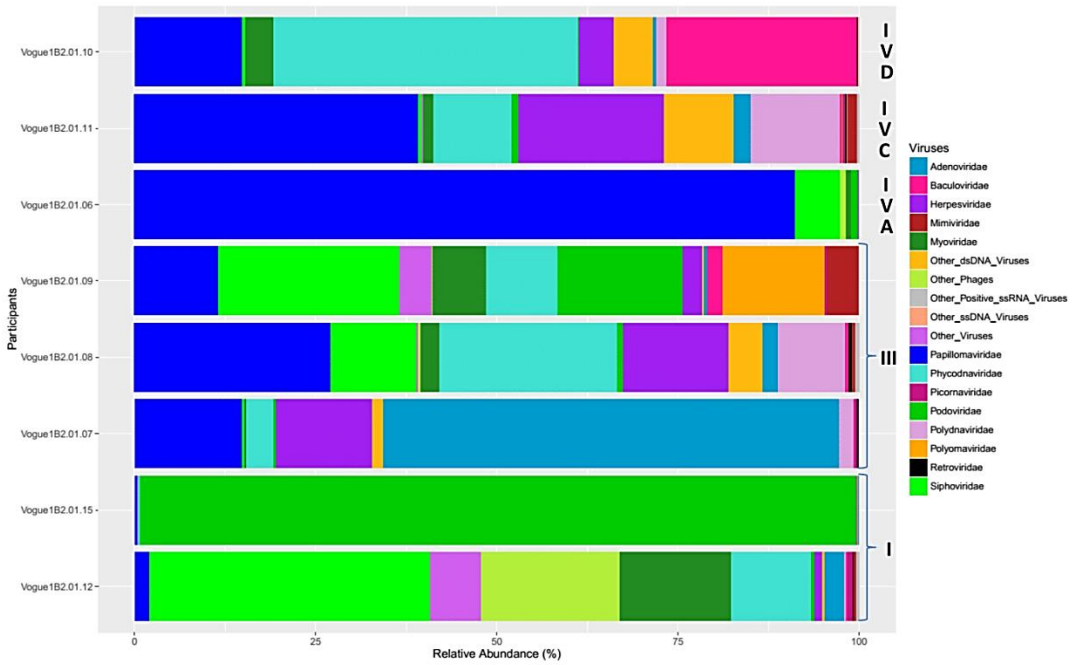

Figure 3.37 Vaginal virome profiles for recurrent BV cohort ordered by bacteriome CSTs

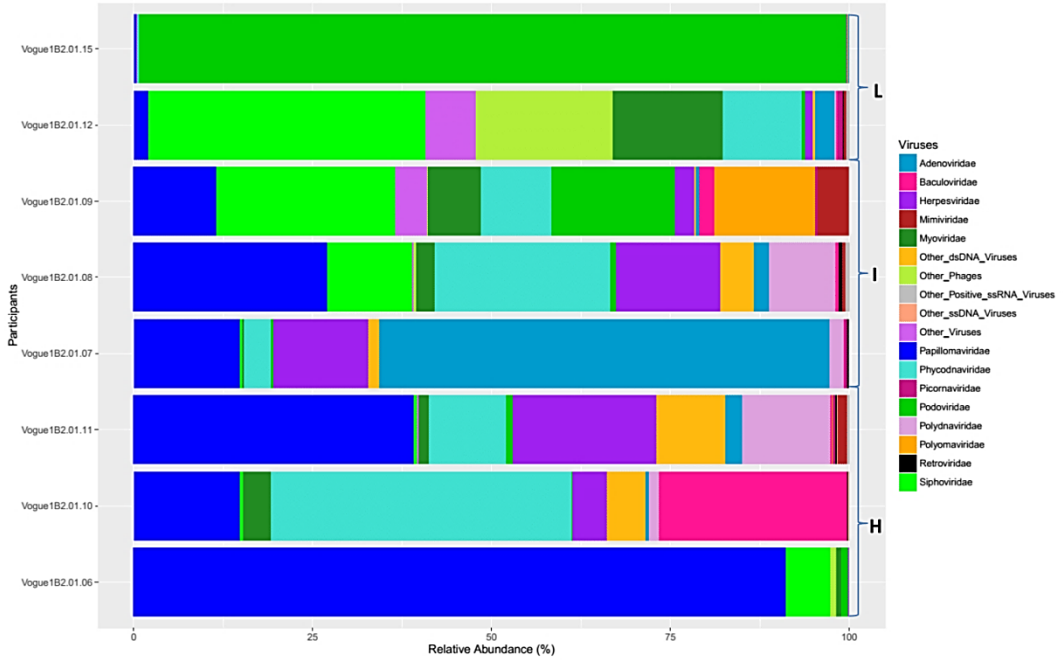

Figure 3.38 Vaginal virome profiles for recurrent BV cohort grouped. *Lactobacillus* dominated profiles (L), *L. iners* dominance (I) and heterogeneous profiles (H)

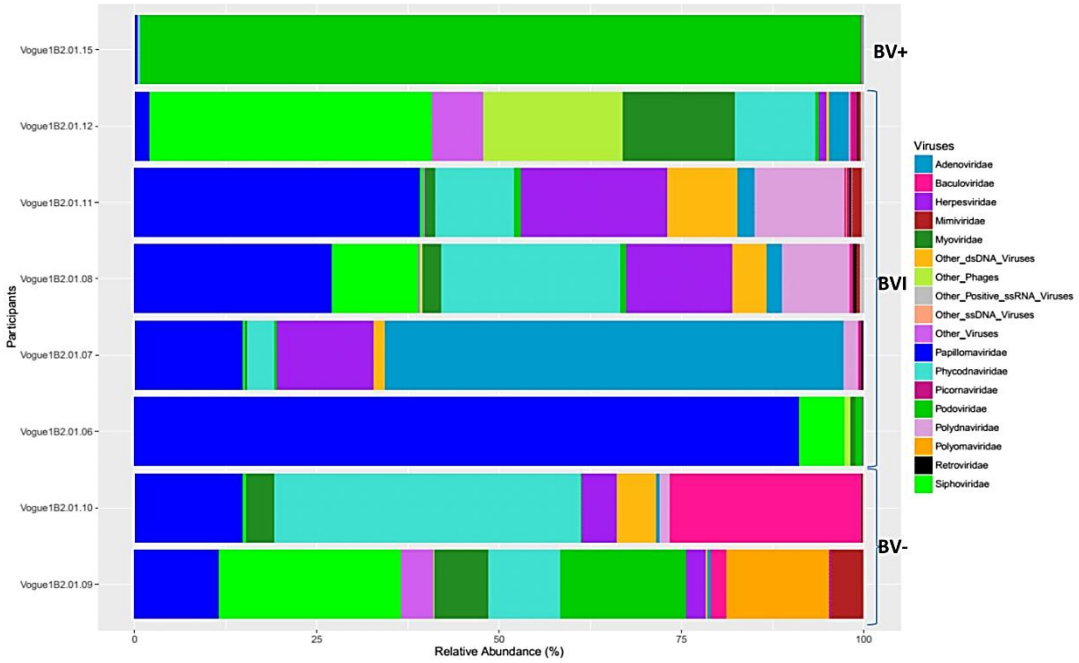

Figure 3.39 Vaginal virome profiles for recurrent BV cohort ordered based on Nugent scores. Nugent scores consistent for BV (BV+), Nugent scores for intermediate BV (BVI), Nugent scores inconsistent for BV (BV-)

The abundance of HPV, HHV and HIV were also looked at within our recurrent BV cohort. *Alphapapillomavirus* was more abundant in all participants, followed by *Lymphocryptovirus* (mainly human herpesvirus 4) and *Roseolovirus* (predominantly human herpesvirus 6). HIV, *Simplexvirus* (HSV-1, -2), and *Rhadinovirus* (human herpesvirus 8) were not seen in this population (Figure 3.40).

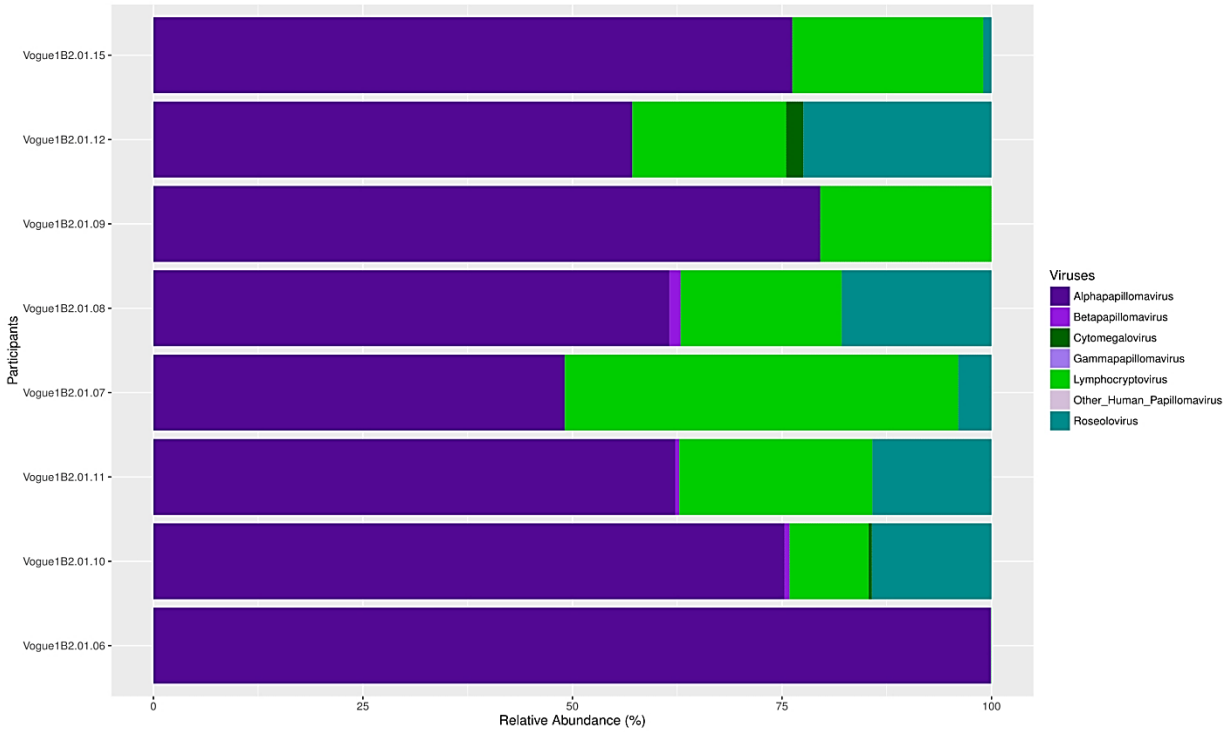

Figure 3.40 HPV, HHV and HIV for recurrent BV cohort

Diversity Statistics

To determine how well our sample population covered the total viral species present in the vaginal environment, Good's coverage was employed (Table 3.58-3.60). In addition, rarefaction curves illustrated whether sampling depth was sufficient; each curve represented a participant in our population. Slope values for each curve were also calculated to determine whether sampling depth was sufficient (Figure 3.41-3.43, Table 3.61-3.63). Furthermore, to determine bacterial diversity within our population of women, Shannon's Diversity Index, Species Richness (Chao1), and Pielou's evenness were calculated for each participant (Table 3.58-3.60).

Table 3.58 Diversity statistics for healthy-asymptomatic cohort

Diversity Statistics				
Participants	**Shannon's Diversity**	**Pielou's Evenness**	**Chao1**	**Good's Coverage**
Vogue1A.01.52	3.35	0.75	88	0.89
Vogue1A.01.59	3.09	0.77	57	1.00
Vogue1A.01.61	1.38	0.30	95	0.96
Vogue1A.01.62	1.50	0.35	72	0.99
Vogue1A.01.64	3.35	0.75	86	0.95
Vogue1A.01.65	2.84	0.69	64	0.99
Vogue1A.01.68	3.26	0.77	68	1.00
Vogue1A.01.69	3.02	0.73	63	0.99
Vogue1A.01.70	1.87	0.46	62	0.95
Vogue1A.01.71	2.70	0.64	70	0.97
Vogue1A.01.74	2.10	0.49	71	0.99
Vogue1A.01.75	3.12	0.73	73	0.95
Vogue1A.01.76	3.06	0.63	131	0.97
Vogue1A.01.77	1.45	0.34	71	0.99
Vogue1A.01.78	0.11	0.03	52	0.98
Vogue1A.01.81	3.10	0.68	96	0.96
Vogue1A.01.84	2.79	0.63	83	0.98
Vogue1A.01.85	1.04	0.26	54	0.97
Vogue1A.01.92	0.81	0.18	85	1.00
Vogue1A.01.101	3.07	0.74	66	0.94
Vogue1A.01.106	3.23	0.77	68	0.96

Table 3.59 Diversity statistics for HIV-positive cohort

Diversity Statistics				
Participants	**Shannon's Diversity**	**Pielou's Evenness**	**Chao1**	**Good's Coverage**
Vogue1B.01.01	2.84	0.63	92	0.99
Vogue1B.01.03	2.01	0.48	69	0.99
Vogue1B.01.04	0.43	0.12	40	1.00
Vogue1B.01.05	3.15	0.76	64	0.99
Vogue1B.01.06	3.13	0.67	109	0.98
Vogue1B.01.08	1.94	0.45	75	1.00
Vogue1B.01.09	0.58	0.15	49	1.00
Vogue1B.01.11	2.79	0.62	93	0.99
Vogue1B.01.12	1.91	0.42	99	0.99
Vogue1B.01.13	1.96	0.47	69	1.00
Vogue1B.01.15	1.70	0.39	77	0.98
Vogue1B.01.17	3.05	0.71	75	0.98
Vogue1B.01.21	1.36	0.31	85	1.00
Vogue1B.01.26	2.26	0.51	89	1.00
Vogue1B.01.27	2.91	0.67	80	1.00
Vogue1B.01.32	1.85	0.47	51	0.99
Vogue1B.01.34	0.45	0.10	101	0.99
Vogue1B.01.36	2.42	0.49	136	1.00
Vogue1B.01.37	2.13	0.48	87	0.95
Vogue1B.01.38	0.90	0.20	88	1.00
Vogue1B.01.40	2.84	0.64	86	1.00
Vogue1B.01.43	1.03	0.23	86	0.98
Vogue1B.01.48	2.47	0.55	93	1.00
Vogue1B.01.51	2.69	0.66	58	1.00
Vogue1B.01.52	1.80	0.48	43	1.00

Table 3.60 Diversity statistics for recurrent BV cohort

Diversity Statistics				
Participants	**Shannon's Diversity**	**Pielou's Evenness**	**Chao1**	**Good's Coverage**
Vogue1B2.01.06	1.78	0.36	131	1.00
Vogue1B2.01.07	1.75	0.46	45	1.00
Vogue1B2.01.08	2.98	0.74	58	1.00
Vogue1B2.01.09	2.95	0.74	55	0.98
Vogue1B2.01.10	2.10	0.48	80	1.00
Vogue1B2.01.11	3.06	0.75	59	0.97
Vogue1B2.01.12	2.66	0.64	66	0.97
Vogue1B2.01.15	0.28	0.08	38	0.99

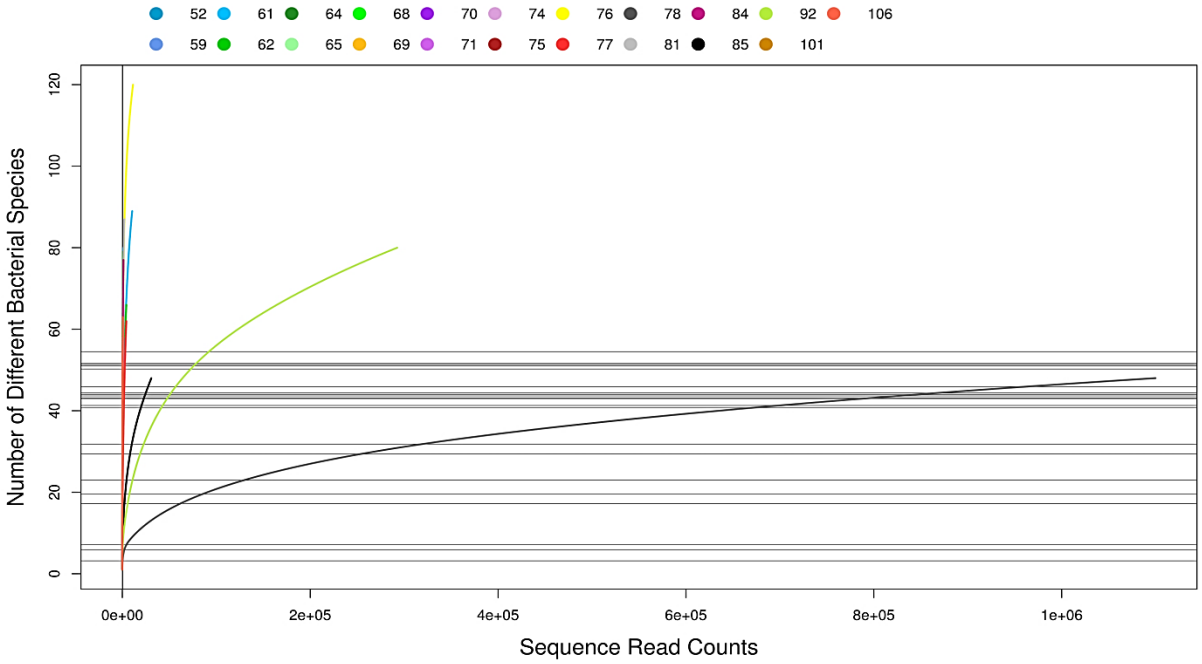

Figure 3.41 Rarefaction curve for healthy-asymptomatic cohort. Each line represents an individual participant

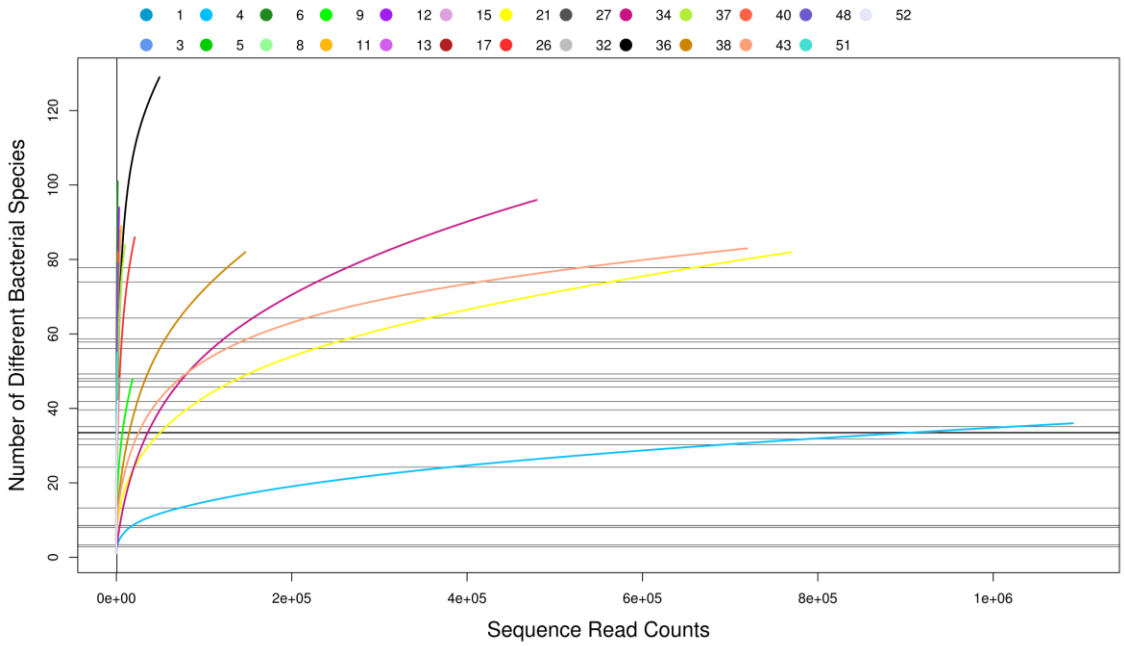

Figure 3.42 Rarefaction curve for HIV-positive cohort. Each line represents an individual participant

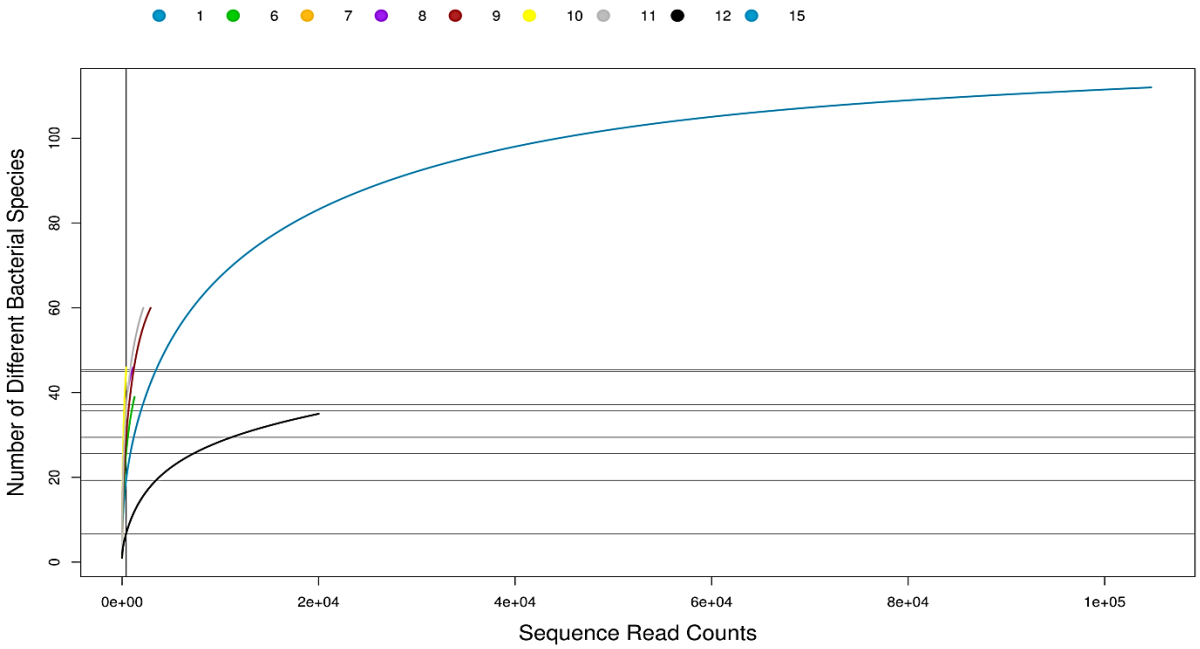

Figure 3.43 Rarefaction curve for recurrent BV cohort. Each line represents an individual participant

Table 3.61 Rarefaction curve slopes for healthy-asymptomatic cohort. Slope values for each rarefaction. Total Reads indicate the end point of the rarefaction curve. Each column represents the slope of the rarefaction curve, at Total Reads – indicated number of reads

Rarefaction Curve Slopes				
Participants	**Total Reads$_{-100}$**	**Total Reads$_{-10}$**	**Total Reads$_{-5}$**	**Total Reads$_{-1}$**
Vogue1A.01.52	0.0265	0.0240	0.0239	0.0238
Vogue1A.01.59	0.1193	0.1002	0.0995	0.0990
Vogue1A.01.61	0.0023	0.0023	0.0023	0.0023
Vogue1A.01.62	0.0054	0.0054	0.0054	0.0054
Vogue1A.01.64	0.0465	0.0419	0.0416	0.0415
Vogue1A.01.65	0.0287	0.0236	0.0234	0.0232
Vogue1A.01.68	0.0542	0.0440	0.0436	0.0432
Vogue1A.01.69	0.0301	0.0257	0.0255	0.0254
Vogue1A.01.70	0.0186	0.0177	0.0177	0.0177
Vogue1A.01.71	0.0380	0.0331	0.0329	0.0327
Vogue1A.01.74	0.0169	0.0158	0.0157	0.0157
Vogue1A.01.75	0.0233	0.0215	0.0214	0.0213
Vogue1A.01.76	0.0019	0.0019	0.0019	0.0019
Vogue1A.01.77	0.0062	0.0061	0.0061	0.0061
Vogue1A.01.78	0.0000	0.0000	0.0000	0.0000
Vogue1A.01.81	0.0097	0.0093	0.0093	0.0093
Vogue1A.01.84	0.0206	0.0194	0.0193	0.0193
Vogue1A.01.85	0.0006	0.0006	0.0006	0.0006
Vogue1A.01.92	0.0001	0.0001	0.0001	0.0001
Vogue1A.01.101	0.0415	0.0358	0.0355	0.0353
Vogue1A.01.106	0.0402	0.0333	0.0330	0.0328

Table 3.62 Rarefaction curve slopes for HIV-positive cohort. Slope values for each rarefaction curve. Total Reads indicate the end point of the rarefaction curve. Each column represents the slope of the rarefaction curve, at Total Reads – indicated number of reads

Rarefaction Curve Slopes				
Participants	**Total Reads$_{-100}$**	**Total Reads$_{-10}$**	**Total Reads$_{-5}$**	**Total Reads$_{-1}$**
Vogue1B.01.01	0.0048	0.0047	0.0047	0.0047
Vogue1B.01.03	0.0077	0.0076	0.0076	0.0076
Vogue1B.01.04	0.0000	0.0000	0.0000	0.0000
Vogue1B.01.05	0.0196	0.0173	0.0172	0.0171
Vogue1B.01.06	0.0178	0.0166	0.0166	0.0165
Vogue1B.01.08	0.0042	0.0041	0.0041	0.0041
Vogue1B.01.09	0.0010	0.0010	0.0010	0.0010
Vogue1B.01.11	0.0032	0.0032	0.0032	0.0032
Vogue1B.01.12	0.0123	0.0121	0.0121	0.0121
Vogue1B.01.13	0.0070	0.0069	0.0069	0.0069
Vogue1B.01.15	0.0038	0.0038	0.0038	0.0038
Vogue1B.01.17	0.0566	0.0527	0.0526	0.0524
Vogue1B.01.21	0.0000	0.0000	0.0000	0.0000
Vogue1B.01.26	0.0011	0.0011	0.0011	0.0011
Vogue1B.01.27	0.0187	0.0173	0.0173	0.0172
Vogue1B.01.32	0.0024	0.0023	0.0023	0.0023
Vogue1B.01.34	0.0001	0.0001	0.0001	0.0001
Vogue1B.01.36	0.0005	0.0005	0.0005	0.0005
Vogue1B.01.37	0.0017	0.0017	0.0017	0.0017
Vogue1B.01.38	0.0002	0.0002	0.0002	0.0002
Vogue1B.01.40	0.0153	0.0148	0.0148	0.0147
Vogue1B.01.43	0.0000	0.0000	0.0000	0.0000
Vogue1B.01.48	0.0224	0.0214	0.0214	0.0213
Vogue1B.01.51	0.0146	0.0129	0.0128	0.0127
Vogue1B.01.52	0.0081	0.0079	0.0079	0.0079

Table 3.63 Rarefaction curve slopes for recurrent BV cohort. Slope values for each rarefaction curve. Total Reads indicate the end point of the rarefaction curve. Each column represents the slope of the rarefaction curve, at Total Reads – indicated number of reads

Rarefaction Curve Slopes				
Participants	**Total Reads$_{-100}$**	**Total Reads$_{-10}$**	**Total Reads$_{-5}$**	**Total Reads$_{-1}$**
Vogue1B2.01.06	0.0001	0.0001	0.0001	0.0001
Vogue1B2.01.07	0.0104	0.0095	0.0095	0.0094
Vogue1B2.01.08	0.0449	0.0305	0.0298	0.0293
Vogue1B2.01.09	0.0072	0.0063	0.0062	0.0062
Vogue1B2.01.10	0.0051	0.0048	0.0048	0.0048
Vogue1B2.01.11	0.0495	0.0407	0.0402	0.0399
Vogue1B2.01.12	0.0073	0.0070	0.0069	0.0069
Vogue1B2.01.15	0.0005	0.0005	0.0005	0.0005

Univariate Analyses

A statistically significant association between our viral groups and Shannon's Diversity Index was detected; where mean viral diversity was greater in Group II compared to the remaining viral groups (Table 3.64). No other statistically significant associations were found, but based on effect size, trends suggested associations between our viral groups, and age, BMI, lifetime history of Gonorrhea, HPV-16, HPV-35, HPV-52, HPV-53, HPV-56, HPV-70, HPV-71, HPV-83, number of HPV types, CD4+ cell count, CD4+ nadir, highest VL ever, VL, duration of antiretroviral medication, duration of HIV infection, HBV surface antibody test results and HCV antibody test results (Table 3.64, 3.65).

Table 3.64 Univariate analyses of viral groups using one-way ANOVA. P-values adjusted via Benjamini Hochberg; * $p < 0.001$. Effect size of one-way ANOVA: F-Statistic (small: 0.1, medium: 0.25, large: 0.4)

Univariate Analyses of Viral Groups	
One-Way ANOVA	
Variable	**F-Statistic**
Shannon's Diversity Index*	18.50
BMI	1.32
Age	0.80
HIV Characteristics:	
CD4 (mm^3)	1.65
CD4 Nadir (mm^3)	1.03
Duration of Antiretroviral Medications	0.84
Duration of HIV Infection (months)	0.74
Highest Viral Load Ever (copies/mL)	0.49
Number HPV Types	0.48
Viral Load (copies/mL)	0.46

Table 3.65 Univariate analyses of viral groups using fisher's exact test. Effect size of Fisher's exact test: Cramer's V (small: 0.1, medium: 0.3, large: 0.5)

Univariate Analyses of Viral Groups	
Fisher's Test	
Variable	**Cramer's V**
HPV-35	0.69
HPV-53	0.69
HPV-83	0.69
HPV-56	0.62
Lifetime History of Gonorrhea	0.61
HPV-16	0.59
HPV-70	0.55
HPV-52	0.55
HPV-71	0.55
Positive HBV sAb Test Result	0.53
Positive HCV Antibody Test Result	0.52

Viral species abundance was viewed in relation to demographic, behavioural and clinical variables via the ALDEx2 package. Only one association was seen between study arm, and *Enterobacteria phage N15*, where this viral species was significantly more abundant in HIV-positive women, compared to healthy-asymptomatic women and women with recurrent BV (Table 3.66). No other associations were seen between viral species and our descriptive characteristics.

Table 3.66 ALDEx2. Kruskal Wallis analyses adjusted via Benjamini Hochberg principle. Number of reads for each study arm; healthy-asymptomatic (1A), HIV-positive (1B) and recurrent BV (1B2)

ALDEx2					
Variable	**Viral Species**	**1A**	**1B**	**1B2**	**P-Value**
Study Arm	*Enterobacteria phage N15*	0	105	0	<0.01

Cohort Comparison

In comparing our three populations of women; healthy-asymptomatic and HIV-positive women, and women with recurrent BV; we found that our HIV-positive cohort was significantly

older than our healthy-asymptomatic and recurrent BV cohort, and that more women were of Aboriginal, South Asian, Black and other origins (Table 3.67). Significantly more women in our HIV-positive cohort were also multiparous, and significantly more women in our healthy-asymptomatic and recurrent BV cohorts were nulliparous (Table 3.68).

Table 3.67 Comparison of demographic characteristics. Healthy-asymptomatic (1A), HIV-positive (1B), and recurrent BV (1B2) cohorts. Continuous variables are reported as mean ± SD, range, and effect size of one-way ANOVA: F-Statistic (small: 0.1, medium: 0.25, large: 0.4). Categorical variables are reported as %, and effect size of Fisher's exact test: Cramer's V (small: 0.1, medium: 0.3, large: 0.5)

Demographic Characteristics					
	1A	**1B**	**1B2**	**P-Value**	**Effect Size**
Age	28 ± 5.9 (20 – 42)	36 ± 7.1 (23 – 48)	29 ± 7.1 (21 – 42)	<0.01*	10.37
BMI	24.9 ± 6.4 (16.7 – 39.8)	26.0 ± 6.8 (17.8 – 42.9)	20.3 ± 3.3 (15.5 – 27.2)	1.00	2.55
Ethnicity	Caucasian: 52.4% Asian: 42.9% South Asian: 4.8%	Caucasian: 40.0% Aboriginal origins: 24.0% South Asian: 4.0% Black: 16.0% Other: 16.0%	Caucasian: 75.0% Asian: 25.0%	<0.001*	0.73

Table 3.68 Comparison of clinical characteristics. Healthy-asymptomatic (1A), HIV-positive (1B), and recurrent BV (1B2) cohorts. Categorical variables are reported as %. Effect size of Fisher's exact test: Cramer's V (small: 0.1, medium: 0.3, large: 0.5)

Clinical Characteristics					
	1A	**1B**	**1B2**	**P-Value**	**Effect Size**
Pregnancy History	Nulliparous: 71.4% Multiparous: 19.1%	Nulliparous: 16.0% Multiparous: 76.0%	Nulliparous: 62.5% Multiparous: 37.5%	<0.001*	0.58
Form of Contraception	None: 19.1% Hormonal: 28.6% Barrier: 38.1%	None: 16.0% Hormonal: 24.0% Barrier: 48.0%	None: 37.5% Hormonal: 25.0% Barrier: 25.0%	0.47 0.92 0.52	0.18 0.05 0.16
Use of Condoms (past 48 hours)	4.8%	8.0%	25.0%	0.27	0.23
Antimicrobial Use (in the past three months)	0%	40.0%	50.0%	<0.01*	0.53
(Non) Prescription Drug Use (past two months)	81.0%	80.0%	62.5%	1.00	0.08

We also saw a statistically significant difference in symptoms, more women in our recurrent BV population experienced abnormal discharge in the past two weeks and past 48 hours, and irritation and discomfort in the past 48 hours (Table 3.69). Significantly more women in our HIV-positive, and recurrent BV cohorts used antimicrobials in the past three months, compared to our healthy-asymptomatic cohort (Table 3.68). There were no statistically significant differences across the remaining demographics, clinical and behavioural characteristics across cohorts (Table 3.67-3.74). Of the diversity statistics calculated, none were significantly different between the healthy-asymptomatic, HIV-positive and recurrent BV cohort (Table 3.75). Viral groups between the three cohorts were not significantly different from each other (Table 3.76).

Table 3.69 Comparison of vaginal symptoms. Healthy-asymptomatic (1A), HIV-positive (1B) and recurrent BV (1B2) cohorts. Categorical variables are reported as %. Effect size of Fisher's exact test: Cramer's V (small: 0.1, medium: 0.3, large: 0.5)

Vaginal Symptoms					
	1A	**1B**	**1B2**	**P-Value**	**Effect Size**
Symptoms (past 48 hours)	Abnormal discharge: 0% Abnormal odor: 0% Irritation or discomfort: 9.5%	Abnormal discharge: 16.0% Abnormal odor: 12.0% Irritation or discomfort: 4.0%	Abnormal discharge: 62.5% Abnormal odor: 25.0% Irritation or discomfort: 62.5%	<0.05*	0.55
				1.00	0.30
				<0.05*	0.30
Symptoms (past 2 weeks)	Abnormal discharge: 0% Abnormal odor: 0% Irritation or discomfort: 14.3% Other: 0%	Abnormal discharge: 24.0% Abnormal odor: 16.0% Irritation or discomfort: 4.0% Other: 4.0%	Abnormal discharge: 75.0%% Abnormal odor: 25.0%% Irritation or discomfort: 62.5% Other: 0%	<0.01*	0.59
				1.00	0.55
				0.06	0.30
				1.00	0.30
Painful Intercourse	38.1%	12.0%	50.0%	0.79	0.37

Table 3.70 Comparison of behavioural characteristics. Healthy-asymptomatic (1A), HIV-positive (1B), and recurrent BV (1B2) cohorts. Categorical variables are reported as %. Effect size of Fisher's exact test: Cramer's V (small: 0.1, medium: 0.3, large: 0.5)

Behavioural Characteristics					
	1A	**1B**	**1B2**	**P-Value**	**Effect Size**
Tampon Use	Never: 33.3% Sometimes: 14.3% Every period, part of the time: 33.3% Every period, exclusively: 19.1%	Never: 36.0% Sometimes: 12.0% Every period, part of the time: 16.0% Every period, exclusively: 36.0%	Never: 12.5% Sometimes but not for every period: 12.5% Every period, part of the time: 50.0% Every period, exclusively: 25.0%	1.00	0.17
Tampon Use (past month)	47.6%	44.0%	75.0%	1.00	0.21
Use of Feminine Products	23.8%	24.0%	12.5%	1.00	0.10
Use of Feminine Products (past 48 hours)	9.5%	4.0%	12.5%	1.00	0.51
Illicit Substance Use	Never: 52.4% Past: 19.1% Current: 28.6%	Never: 24.0% Past: 28.0% Current: 48.0%	Never: 25.0% Past: 75.0% Current: 0%	0.51	0.53
Alcohol Consumption	None: 14.3% Occasional drink: 52.4% 2-3 drinks/week: 23.8% Daily: 4.8%	None: 44.0% Occasional drink: 44.0% 2-3 drinks/week: 8.0% Daily: 4.0%	None: 12.5% Occasional drink: 25.0% 2-3 drinks/week: 50.0%	0.25	0.48
Smoker	Never: 81.0% Past: 4.8% Current:14.3%	Never: 32.0% Past: 20.0% Current: 48.0%	Never: 62.5% Past: 12.5% Current: 12.5%	0.21	0.41

Table 3.71 Comparison of sexual activity characteristics. Healthy-asymptomatic (1A), HIV-positive (1B), and recurrent BV (1B2) cohorts. Continuous variables are reported as mean ± SD, range. Categorical variables are reported as %, and effect size of Fisher's exact test: Cramer's V (small: 0.1, medium: 0.3, large: 0.5). Number of sexual partners (past year) analyzed as categorical variable.

Sexual Activity Characteristics					
	1A	**1B**	**1B2**	**P-Value**	**Effect Size**
Sexual Partners	Male: 90.5% Female: 4.8% Both: 4.8%	Male: 100%	Male: 75.0% Female: 12.5%	1.00	0.21
Number of Sexual Partners (past year)	1.5 ± 1.6 (0 – 6)	1.0 ± 0.7 (0 – 3)	1.1 ± 0.4 (1 – 2)	1.00	0.21
Vaginal intercourse (past 48 hours)	Yes:23.8% No: 76.2%	Yes: 16.0% No: 84.0%	Yes: 37.5% No: 50.0%	1.00	0.21
Frequency of Oral Sex	Never: 19.1% Ever: 81.0%	Never: 48.0% Ever: 52.0	Never: 12.5% Ever: 75.0%	1.00	0.32
Frequency of Anal Sex	Never: 100.0%	Never: 92.0% Weekly: 4.0% Monthly: 4.0%	Never: 75.0% Weekly: 0% Monthly: 12.5% Other: 0%	1.00	0.22
Frequency of Sex Toy use	Never: 66.7% Daily: 0% Weekly: 19.1% Twice a month: 0% Monthly: 4.8% Other: 9.5%	Never: 84.0% Weekly: 4.0% Monthly: 8.0% Other: 4.0%	Never: 25.0% Weekly: 0% Monthly: 12.5% Other: 50.0%	0.79	0.39

Table 3.72 Comparison of genital infection history. Healthy-asymptomatic (1A), HIV-positive (1B), and recurrent BV (1B2) cohorts. Continuous variables are reported as mean ± SD, range, and effect size of one-way ANOVA: F-Statistic (small: 0.1, medium: 0.25, large: 0.4). Categorical variables are reported as %, and effect size of Fisher's exact test: Cramer's V (small: 0.1, medium: 0.3, large: 0.5).

Genital Infection History					
	1A	**1B**	**1B2**	**P-Value**	**Effect Size**
Bacterial Vaginosis	28.6%	56.0%	75.0%	1.00	0.34
Lifetime	1.3 ± 4.4 (0 – 20)	1.4 ± 1.6 (0 – 6)	6.5 ± 6.9 (0 – 20)	0.21	5.93
In the past year	0.3 ± 1.3 (0 – 6)	0.4 ± 0.8 (0 – 3)	2.8 ± 3.4 (0 – 10)	0.06	7.57
In the past two months	0	0.1 ± 0.4 (0 – 2)	0.6 ± 0.7 (0 – 2)	0.10	6.89
Yeast	71.4%	80.0%	87.5%	1.00	0.14
Urinary Tract Infections	42.9%	56.0%	50.0%	1.00	0.12
Trichomoniasis	4.8%	20.0%	0%	1.00	0.27
Genital Warts	4.8%	32.0%	12.5%	1.00	0.33
Genital Herpes	9.5%	12.0%	0%	1.00	0.14
Chlamydia	4.8%	20.0%	12.5%	1.00	0.21
Gonorrhea	0%	20.0%	0%	1.00	0.34
Syphilis	0%	4.0%	0%	1.00	0.15

Table 3.73 Comparison of Nugent scores. Healthy-asymptomatic (1A), HIV-positive (1B), and recurrent BV (1B2) cohorts. Categorical variables are reported as %. Effect size of Fisher's exact test: Cramer's V (small: 0.1, medium: 0.3, large: 0.5)

Nugent Scores					
	1A	**1B**	**1B2**	**P-Value**	**Effect Size**
Inconsistent for BV (0 – 3)	61.9%	52.0%	12.5%	1.00	0.31
Intermediate BV (4 – 6)	14.3%	16.0%	62.5%		
Consistent for BV (7 – 10)	19.1%	32.0%	12.5%		

Table 3.74 Comparison of community state types. Healthy-asymptomatic (1A), HIV-positive (1B), and recurrent BV (1B2) cohorts. Categorical variables are reported as %. Effect size of Fisher's exact test: Cramer's V (small: 0.1, medium: 0.3, large: 0.5)

Community State Types						
CST	**Dominant OUT**	**1A**	**1B**	**1B2**	**P-Value**	**Effect Size**
Type I	*L. crispatus* dominated	23.8%	20.0%	25.0%	1.00	0.24
Type II	*L. gasseri* dominated	0%	0%	0%		
Type III	*L. iners* dominated	19.1%	12.0%	37.5%		
Type 1VA	Mixed Dominant Phenotype: Mixed Dominant Phenotype: *Gardnerella* subgroup B, *Bifidobacterium* sp., *Prevotella amnii, Shigella boydii, and L.delbrueckii*	14.3%	16.0%	12.5%		
Type IVC	*Gardnerella* A dominated	14.3%	24.0%	12.5%		
Type IVD	*Gardnerella* C dominated	14.3%	24.0%	12.5%		
Type V	*L. jensenii* dominated	14.3%	4.0%	0%		

Table 3.75 Comparison of diversity statistics. Healthy-asymptomatic (1A), HIV-positive (1B), and recurrent BV (1B2) cohorts. Continuous variables are reported as mean ± SD, range. Effect size of one-way ANOVA: F-Statistic (small: 0.1, medium: 0.25, large: 0.4)

Diversity Statistics					
	1A	**1B**	**1B2**	**P-Value**	**Effect Size**
Shannon's Diversity	2.39 ± 0.98 (0.11 – 3.35)	2.02 ± 0.84 (0.43 – 3.15)	2.19 ± 0.94 (0.28 – 3.06)	1.00	0.93
Pielou's Evenness	0.56 ± 0.23 (0.03 – 0.77)	0.47 ± 0.19 (0.10 – 0.76)	0.53 ± 0.24 (0.08 – 0.75)	1.00	1.11
Species Richness	75.00 ± 17.90 (52 - 131)	79.76 ± 21.76 (40 - 136)	66.50 ± 28.98.54 (38 - 131)	1.00	1.18
Good's Coverage	0.97 ± 0.03 (0.89 – 1.00)	0.99 ± 0.01 (0.95 – 1.00)	0.99 ± 0.01 (0.97 – 1.00)	0.06	7.56

Table 3.76 Comparison of viral groups. Healthy-asymptomatic (1A), HIV-positive (1B), and recurrent BV (1B2) cohorts. Categorical variables are reported as %. Effect size of Fisher's exact test: Cramer's V (small: 0.1, medium: 0.3, large: 0.5)

Viral Groups					
Group	**1A**	**1B**	**1B2**	**P-Value**	**Effect Size**
I	19.1%	12.0%	12.5%	1.00	1.24
II	33.3%	12.0%	37.5%		
III	28.6%	44.0%	23.0%		
IV	4.8%	12.0%	0%		
V	14.3%	8.0%	12.5%		
VI	0%	8.0%	0%		
VII	0%	4.0%	2.5%		

4 Discussion

4.1 Bacteriome Discussion

Overall Findings

This project aimed to characterize the vaginal bacteriome of women with recurrent BV, with the goal of obtaining predictive vaginal bacteriome profiles, which will lay the foundation for better diagnostic tools and targeted treatments. In our cohort of 26 women with recurrent BV, bacterial profiles clustered into six CSTs. Four OTUs were detected in all samples, corresponding to *L. crispatus, L. iners, L. jensenii,* and *G. vaginalis* Group A. 12 women fell into clusters dominated by *Lactobacillus* species, and the remaining 14 women fell into non-*Lactobacillus* dominated clusters. Interestingly, women who had Nugent scores inconsistent for BV did not all fall into *Lactobacillus* dominated clusters. Two women had profiles dominated by *Klebsiella pneumoniae,* and *G. vaginalis* Group A. However, we saw that all women who had Nugent scores consistent for BV had non-*Lactobacillus* dominated profiles. Although Nugent scores are the gold standard for BV diagnosis, these results indicate CSTs are more illustrative because unlike the Nugent score, which focuses on a subset of bacterial species, CSTs provide a comprehensive representation of bacteria present within the vaginal microbiome.

When comparing CSTs with demographic, clinical and behavioural characteristics of our cohort, trends appeared suggesting associations between several variables and clusters. We saw trends between our BV-associated CSTs, CST III, IVA, IVC and IVD, and lifetime history of trichomoniasis, oral sex frequency, and hormonal contraception. Whereas, number of sexual partners in the past year, no method of contraception, abnormal discharge in the past 48 hours and Shannon's Diversity Index were suggested to be associated with our *L. crispatus* dominated CST; CST I. These findings are suggestive of the role these activities could play in the etiology of recurrent BV, both in traditional *G. vaginalis* dominated cases and non-traditional cases.

We also saw an association between our BV-associated CSTs, and *L. iners*, *G. vaginalis* Group A, *G. vaginalis* Group B, *G. vaginalis* Group C, *G. vaginalis* Group D, *Atopobium vaginae*, *Clostridia sp*. BVAB2, *Megasphaera sp*. type 1, *Eubacterium cellulosolvens*, and *Prevotella timonensis*. These results support the pathogenetic role these bacterial species play in recurrent BV. This is further portrayed by the association seen between number of BV episodes in the past year, and *G. vaginalis* Group A, *G. vaginalis* Group C and *Clostridia sp*. BVAB2, as well as between number of BV episodes in the past two months, and *G. vaginalis* Group A, and *G. vaginalis* Group C.

This data to date gives us a much broader, more nuanced understanding of the organism clusters seen in women with recurrent BV.

Study Population

Twenty-six women were recruited into our study. These women were non-pregnant, HIV-negative, not menopausal, between the ages 20-47 years old (mean age 32 years old), and had a history of recurrent BV. Women were included who may not have had four clinically documented cases of BV, but experienced troublesome BV and BV-like symptoms over a significant period of time. Self-reported questionnaires indicated that most women in our cohort reported a history of BV, but not all; their histories were verified through lab documents and chart reviews. This inconsistency may be due to misinterpretation, where several women mistook their symptoms as yeast. Also, some women may not have been symptomatic on the day of the study visit, but all women in our population had a verified BV history.

On review of our results, it became apparent that although women were originally enrolled filling the criteria for recurrent BV, some histories were not consistent once verified through chart reviews. In addition, on the day of the visit, some women were not very

symptomatic. We sought to enrol women with a history of recurrent BV, and did not include nor exclude women on the basis of Nugent score alone, as there are many women who were not getting clinical examinations conducted at the time of each of their previous episodes of presumed BV. In addition, it should be noted, that the Nugent score is dependent on the presence of *Lactobacillus*, *G. vaginalis*, and *Mobiluncus* species[58]. Past research has shown us not only is *G. vaginalis* not a definitive diagnostic marker for BV [16,19,39,40,59–63], but there is also a great deal of complexity and diversity associated with BV [9,16,19,41,42]. Other forms of dysbiosis can consequently be missed with Nugent's score alone.

Sampling Depth

Our metagenomic process was highly robust resulting in a total of 742 isotigs, assembled from 127,768 reads, and based on a >55% cut-off identity to the *cpn*60 gene target a total of 122 unique *cpn*60 OTUs were identified in our cohort. In each of our samples, Good's coverage was greater than 99% (range: 0.99 - 1.00) (Table 3.8). This implies that for every 100 additional sequence reads, less than one additional OTU would be discovered. In addition, Chao1 estimates for each sample were close to the actual number of OTUs detected, therefore providing further support that sampling was sufficient and samples are representative of the communities seen in the vaginal environment (Table 3.8).

Furthermore, rarefaction curves provided additional evidence of sufficient sampling depth. When sampling depth is sufficient, rarefaction curves plateau, reaching a slope of zero. This indicates that with further sampling, there is a low probability of discovering more species and the rarest species remain to be sampled. When rarefaction curves do not reach this plateau state, this indicates a large fraction of diversity remains to be discovered. Each rarefaction curve represents an individual participant in the cohort (Figure 3.5). Visually, we can identify curves which plateau, and curves which do not. It appears curves clustered to the left of Figure 3.5, are

still rising and do not approach a slope of zero. This indicates that not all participants in our study have sufficient sampling depth, and further sampling may be required for sampling to be representative of the vaginal environment. However, this may be a graphing artifact, where participants with fewer reads have the curvature of their rarefaction curves obscured by the large range of the x-axis. As it was difficult to visually determine if each curve was reaching a plateau, the slope was calculated at five different positions in each curve: the total reads of each participant, and the total reads less 1, 5, 10, and 100 read(s) (Table 3.9). Looking at the slope at each of these points, we see that as we reach the total reads for each participant, the slope is decreasing and reaching zero; something we cannot see visually. Overall, the diversity indices and the rarefaction curves indicate that sampling depth was sufficient in our study, and our samples are representative of the vaginal microbiota of women with recurrent BV.

Relationship with Nugent Scores

Past literature has associated *Lactobacillus* species with health, and their predominance in the vaginal environment is thought to promote a healthy microbiota [13,38–40,47]. However, *L. iners* has been seen at high levels in both women with and without BV [9,41,42]. As indicated earlier, *L. iners* is suggested to be a transitional bacterial species, pushing a healthy microbiota into an unhealthy state [9,16,19,41]. Hierarchical clustering showed that our participants clustered into six groups. Twelve women fell into *Lactobacillus* dominated clusters, where half of these women had profiles dominated by *L. iners*, and fell into CST III. Of the remaining six women, five fell into CST I, the *L. crispatus* dominated cluster and one woman fell into CST II, the *L. gasseri* dominated cluster. Fourteen women fell into non-*Lactobacillus* clusters, CST IVA, CST IVC, and CST IVD. As indicated earlier, CST IVA was a mixed dominance group, where profiles were dominated by *G. vaginalis* Group B*, Prevotella amnii, Shigella boydii, Bifidobacterium breve, Bifidobacterium dentium*, and *L. delbrueckii*. CST IVC and CST IVD were dominated by *G. vaginalis* Group A and *Megasphaera micronuciformis*, and *G. vaginalis* Group C,

respectively. CST I and CST II were denoted as clusters associated with health, due to the abundance of *Lactobacillus* species. Whereas, CST III, IVA, IVC and IVD were seen as BV-associated clusters due to the presence of bacterial species, which past literature has linked to BV and a perturbed vaginal microbiota (Figure 3.1, 3.2).

Results of the Mollicutes and *Ureaplasma* specific PCR assays showed in our "healthy" CSTs, 3/5 women were positive for Mollicutes, and 2/5 were negative. Of the three positive Mollicutes samples, two were also positive for *U. parvum* and one was positive for *U. urealyticum*. Four of the 19 women in our BV-associated CSTs were women negative for Mollicutes. Of these women, three were also negative for *Ureaplasma* species, while one was positive for *U. urealyticum*. As indicated earlier, *Ureaplasma* are Mollicutes, thus the result seen here may be due to the lack of robustness of the Mollicutes PCR compared to the Ureaplasma assay. The remaining 15 women in our BV-associated CSTs were positive for Mollicutes, where 12 were positive for *U. parvum* and three were positive for *U. urealyticum.* We could not conclude whether these positive Mollicutes samples also contained *Mycoplasma* species. It possible that they contain *Mycoplasma (M.) hominis*, or *M. genitalium*, but we cannot confirm this with the assays we conducted. A *Mycoplasma* specific assay would have to be conducted (Table 3.7, Figure 3.3, 3.4).

Two of six women who had Nugent scores inconsistent for BV, did not fall into *Lactobacillus* dominated clusters. One woman's profile mostly consisted of *Klebsiella Pneumoniae* and fell into the mixed dominance cluster (CST IVA), and another woman fell into the *G. vaginalis* Group A dominated cluster (CST IVC). The remaining four women had profiles dominated by *L. crispatus*, *gasseri* or *iners*. Women with Nugent scores for intermediate BV fell into *Lactobacillus* and non-*Lactobacillus* dominated clusters. Eight women had profiles dominated by *Lactobacillus* species, three by *L. crispatus*, and five by *L. iners*. Two women fell

into the mixed dominant heterogeneous cluster, and three women had profiles dominated by *G. vaginalis* Group A. Of the two women in the mixed dominance group, one profile was dominated by *Escherichia coli* and *G. vaginalis* Group A, and the other *by Actinobacteria* sp. Notably, the one woman who fell into CST III had a bacterial profile largely dominated by not only *L. iners,* but *Megasphaera* sp. type 1 as well. All women with Nugent scores consistent for BV fell into non-*Lactobacillus* dominated clusters; CST IVA: 3, CST IVC: 2, CST IVD: 2. One woman in CST IVC had, in addition to *G. vaginalis* Group A, a mixture of *G. vaginalis* Group C, *Atopobium vaginae*, *Clostridia sp* BVAB2 and BVAB3, *Megasphaera sp.* type 1, and *Prevotella timonensis*, in lower abundances. The three woman in CST IVA had diverse profiles; one profile was dominated by a mixture of *Actinobacteria sp., Clostridia sp.* BVAB2, *Atopobium vaginae, L. iners*, and *Prevotella amnii,* another by *G. vaginalis* Group C, *Clostridia sp*. BVAB2, *Prevotella timonensis* and other *Prevotella* species, and the third by *Escherichia coli*, *L. gasseri, L. iners* and *Streptococcus sp*. We can see that women with Nugent scores inconsistent for BV and for intermediate BV fell into five and four distinct clusters, respectively; clusters which were *Lactobacillus* and non-*Lactobacillus* dominated. Those women with scores consistent for BV fell into three distinct clusters, all non-*Lactobacillus* dominated (Figure 3.1). We anticipated that women with Nugent scores inconsistent for BV would have *Lactobacillus* dominated profiles, and women with Nugent scores for intermediate BV or consistent for BV would have *L. iners* and non-*Lactobacillus* dominated profiles, respectively. Interestingly, our clusters did not distribute the way we would have thought between Nugent scores, though not entirely different from past findings.

Gajer et al. found similar findings where high Nugent scores were associated with CST IVB, the cluster with higher proportions of *Atopobium, Prevotella, Parvimonas, Sneathia, Gardnerella, Mobiluncus* or *Peptoniphilus*, whereas low Nugent scores were associated with CST IVA, the cluster with modest proportions of *Lactobacillus* species and low proportions of

other anaerobic bacteria like *Anaerococcus*, *Corynebacterium*, *Finegoldia* or *Streptococcus* [40]. In addition, Ravel et al. found women with Nugent scores consistent with BV were most often associated with group IV, a heterogeneous group with high proportions of strictly anaerobic bacteria [39]. However, high Nugent scores were also seen in other *Lactobacillus* dominated groups. Women with low and intermediate Nugent scores were mostly seen in *Lactobacillus* dominated groups I, II, III and V but were also present in group IV [39]. However, our findings are in line with previous research done by our own team, which characterized the vaginal microbiome of healthy-asymptomatic women [47]. Women with Nugent scores consistent for BV were only observed in CST IVA, CST IVC and CST IVD, where Nugent scores inconsistent for BV and for intermediate BV were observed in all CSTs within the cohort [47].

The relative abundance of bacterial species present in our cohort was viewed in relation to Nugent scores, and 19 OTUs were found to be significantly associated with Nugent scores (Table 3.12). The relative abundance of *L. crispatus*, *L. gasseri*, *L. jensenii*, *L. reuteri*, *Clostridium saccharoperbutylacetonicum*, and *L. iners* were greater in those women with Nugent scores inconsistent for BV, than those for intermediate BV and consistent for BV. This was not entirely surprisingly as *Lactobacillus* species have been noted in literature to be associated with health [13,38–40,47]. As indicated previously, past research done by our team looking at healthy-asymptomatic women found approximately 75% of women in the population fell into the three CSTs dominated by *Lactobacillus* species, where 50% had vaginal microbiotas dominated by *L. crispatus.* Our team also saw a relationship between *Lactobacillus* species and Nugent scores, where *Lactobacillus* species were present at a lower abundance in intermediate and consistent for BV Nugent scores [47]. Ravel et al. found a similar correlation between *Lactobacillus* species and low Nugent scores [39]. The relationship between *L. iners* and low Nugent scores was interesting as *L. iners* is a suggested transitional bacterial species. It is possible that these women will be pushed into the direction of a perturbed microbiome, and that *L. iners* is priming

the environment for BV-associated bacteria. Also interesting was the association between *Clostridium saccharoperbutylacetonicum* and low Nugent scores, a relationship which has not been noted in literature before. There was no information on this bacterial species in relation to the vaginal microbiome, nor the human microbiome. Research has shown that this bacterial species is a producer of butanol, where it converts sugars (glucose) into lactic acid, and when there is an abundance of lactic acid in the environment it then converts this substrate into butanol. This process would initially decrease the pH, similar to the production of lactic acid by *Lactobacillus* species. But as *Clostridium saccharoperbutylacetonicum* uses lactic acid, this would in turn increase the pH and remove the protective effect on the vaginal environment, thus making the environment hospitable to BV-associated bacteria [192]. Changes in pH, Nugent scores, bacterial species, and metabolite levels could be followed longitudinally to assess this possible mechanism.

Several bacterial species in our cohort were also significantly associated with Nugent scores intermediate and consistent for BV; *Clostridia sp.* BVAB2, *Atopobium vaginae*, *G. vaginalis* Group C, *G. vaginalis* Group A, *G. vaginalis* Group B, *G. vaginalis* Group D, *Prevotella timonensis*, *Megasphaera* sp. type 1, *Peptoniphilus harei*, *Prevotella buccalis*, *Prevotella bergensis*, *Clostridium genomosp* BVAB3, and *Prevotella bivia* (Table 3.12). BV has been noted to be complex and diverse in the literature, and we see our findings are in line with previous research as these bacterial species have been suggested to be associated with BV [16,19,41,42,47,62,113,116,119,193]. Ravel et al. noted a relationship between high Nugent scores and *Aerococcus, Anaeroglobus, Anaerotruncus, Atopobium, Coriobacteriaceae_2, Dialister, Eggerthella, Gardnerella, Gemella, Megasphaera, Mobiluncus, Parvimonas, Peptoiphilus, Prevotella, Porphyomonas, Prevotellaceae_1, Prevotellaceae_2, Ruminococcaceae,* and *Sneathia*, which is similar to our findings [39]. Similar findings were also seen in our team's past research, where the abundance of *G. vaginalis* Group B, *Clostridia sp.* BVAB2, *Actinobacteria*,

Proteobacteria and *Streptococcus* species were observed in Nugent scores intermediate and consistent for BV [47]. Bacterial species we detected in our cohort were not all seen in Ravel et al. and our team's past study as both studies did not aim to recruit a recurrent BV population, and these women are largely healthy-asymptomatic [39,47]. Thus, significant species we saw in our own study may not have been present in these related studies or not at a high enough abundance to make correlations.

Relationship with Community State Types

On the basis of the bacterial species which made up our CSTs, our six CSTs were grouped into two categories. CST III, CST IVA, CST IVC and CST IVD were designated the BV-associated CSTs due to presence of BV-associated bacteria within these clusters, and CST I was designed the "healthy" CST due to presence of *L. crispatus;* previously documented as healthy bacteria in the context of the vaginal microbiome. Relative abundances of bacterial species which were higher in our healthy CST were *L. crispatus*, *jensenii*, *gasseri*, *Clostridium saccharoperbutylacetonicum*, *L. salivarius*, *L. reuteri*, and *Staphylococcus hominis* (Table 3.14). This was similar to the bacterial species we saw to be significantly associated with low Nugent scores, and not entirely surprising as *Lactobacillus* species have been associated with health. The relationship between *Staphylococcus hominis* and the vaginal microbiota has not been described in literature in detail. *Staphylococcus hominis* is described as a generally harmless commensal part of the human body. It may also have probiotic potential as it produces bacteriocins [194], potentially promoting a healthy environment in this manner.

CSTs associated with BV were associated with *L. iners*, *G. vaginalis* Group A, *G. vaginalis* Group B, *G. vaginalis* Group C, *G. vaginalis* Group D, *Atopobium vaginae*, *Clostridia sp*. BVAB2, *Megasphaera sp*. type 1, *Eubacterium cellulosolvens*, and *Prevotella timonensis* (Table 3.14). This was similar to the bacterial species significantly associated with Nugent

scores intermediate and consistent for BV. These results are supported by past literature, which have linked their presence to BV [16,19,41,42,47,62,113,116,119,193].

As mentioned earlier, the Nugent score is reliant on the abundance of three bacterial species: *Lactobacillus*, *G. vaginalis*, and *Mobiluncus* species [58]. Past research has shown us that recurrent BV is a great deal more complex and diverse [9,16,19,41,42].The same was seen in our own findings, where several bacterial species were seen to be associated with BV; some in line with past studies, while others novel. As evident by our findings, there was not one dominant bacterial species driving recurrent BV. Recurrent BV appears to be one of diversity. CSTs were more illustrative than Nugent scores alone, because they were a more comprehensive representation of bacteria present, compared to the Nugent score which focuses on certain bacterial species. CSTs emphasized the notion that recurrent BV may not be a single profile but an umbrella term where multiple distinct profiles reside. CSTs are based on bacterial dominance, but there is also a need to look at the relative abundance of bacterial species, as bacterial species at a lower abundance could be etiological drivers of BV. Identifying bacterial profiles of women, and understanding the patterns and diversity may lead to more targeted BV treatments in the future.

BV-associated Community State Types Trends

We were also interested in determining whether distinct patterns of behaviours, demographic or clinical characteristics were associated with our CSTs. Effect size suggested trends between our BV-associated CSTs (III, IVA, IVC, and IVD), and oral sex, use of hormonal contraception, lifetime history of trichomoniasis, and an increasing number of BV episodes experienced in the past two months and year (Table 3.10).

Prior studies have indicated that bacterial species, such as *G. vaginalis*, found within the vaginal microbiome have also been identified within the oral microbiome [88,195]. Oral sex has been noted to transmit pathogens [196,197], and past research has identified a relationship between receptive oral sex and the vaginal flora, where a greater number of current episodes of receptive oral sex was significantly associated with patterns of unstable flora [198]. Tcharmouoff et al. found similar findings in their pilot study [199]. Of the 17 women enrolled, nine women who have sex with women (WSW) practiced receptive oral sex in the previous four weeks, of which six had BV. This was compared to the eight participants who did not practice receptive oral sex, where none of the women had BV. In their larger prospective study of 256 women, there was a significant association between receptive oral sex and BV [199]. Of the 111 women who practiced receptive oral sex in the previous four weeks, 41 (36.9%) had BV, whereas, of the 145 women did not practice receptive oral sex, 14 (9.7%) had BV [199]. Recent research has not found supporting results, and has indicated no association between oral sex and BV [53,80]. However, in another study later that year, Marrazzo, Thomas, Fiedler, et al. noted a dose response relationship between the risk of BV acquisition and higher reported episodes of receptive oral-vulvovaginal sex [112]. They found a 21% increase in the risk of BV for each additional 10 oral sex acts in last three months [112]. More recently, Marrazzo et al. also detected BV-associated bacteria, *G. vaginalis* and *Leptotrichia*/*Sneathia* in oral swab of patients. *G*. *vaginalis* and *Leptotrichia*/*Sneathia* were detected in 62% and 13% of oral swabs, respectively [88]. Furthermore, G.vaginalis was more likely to be detected in the oral cavity among women who acquired BV, whereas *L. crispatus* was not [88]. Oral sex has also been linked to the transmission of syphilis, gonorrhea and herpes, and HIV, Chlamydia and HPV [196,200]. In our study, we found a positive association between oral sex and our BV-associated CSTs. Literature has suggested that oral sex may transmit genital pathogens [196,197]. Therefore, with the trends we see in our own study, oral sex could potentially transmit BV-associated pathogens in the vaginal microbiota of women, resulting in BV.

A couple of studies have noted no associations between oral contraceptives and BV [53,201], but more recent research has indicated that hormonal contraception provides a protective effect against BV [52,79,84–87]. To reiterate, estrogen may increase the glycogen content of epithelial cells, which is an essential substrate for the generation of lactic acid by *Lactobacillus* species, and in turn this may promote a healthy vaginal environment [84]. Several studies have noted a similar protective effect of oral hormonal contraceptives [79,84,87]; suggesting this protective effect was due to estrogen. Another study conducted by Bradshaw et al. further supported the protective effect of hormonal contraception as well suggesting that not any hormonal contraceptive would offer this protective effect [85]. Univariate analyses showed lower rates of incident BV with use of oestrogen-containing contraception (combined oral contraceptive pill or ring) in the past three months. This association was not seen with use of any hormonal contraceptive (including progesterone-only methods) [85]. This was not in line with our findings, where we instead found that hormonal contraception was associated with our BV-associated CSTs. However, two women in our cohort were using progestin-containing hormonal contraceptives, which are suggested to not confer the same protective effect which estrogen-containing contraceptives are suggested to have. Further research is needed to understand this relationship.

We also found trends which suggest an association between our BV-associated CSTs and lifetime history of trichomoniasis. Trichomoniasis is marked by an increased volume of vaginal discharge which can be malodorous, green or yellow in color, and frothy in appearance, the presence of an anaerobic parasitic organism, *Trichomonas (T.) vaginalis*, as well as an increase in the vaginal pH, greater than 4.5 [14,20]. The resulting pH change may result in a hospitable environment for pathogenic bacteria. Previous research supports this notion. CST IV, as defined by Ravel et al., has been found to be significantly associated with an eight fold increase in the odds of detecting *T. vaginalis*, compared to the remaining *Lactobacillus*

dominated CSTs [39,202]. Furthermore, there was a positive association seen between *T. vaginalis* and *Mycoplasma*, *Parvimonas*, *Sneathia* and other bacterial species present in CST IV [202]. We did not detect *T.* vaginalis in our cohort, but our findings support a relationship between trichomoniasis and BV-associated CSTs, as has been described by the above results. Colonization of *T. vaginalis* resulting in trichomoniasis may create a basic environment within the vaginal microbiome, allowing further disruption by BV-associated anaerobes, which may result in BV.

G. vaginalis Group A, *G. vaginalis* Group C and *Clostridia sp.* BVAB2 have been indicated in past research to be associated with BV [16,19,41,42,47,62,113,119]. *Clostridia sp.* BVAB2 has been found to be a highly specific indicator of BV, and seen in over 80% of women with BV [62,113]. Srinivasan et al. had similar findings; 95% of their sequence reads could be accounted for by 24 taxa, BVAB2 and *G. vaginalis* being two of them [41]. Srinivasan et al., and Vitali et al. similarly found that *G. vaginalis* was present in approximately 97% of women with BV [41,42]. These findings are supported by Li et al. who saw a higher prevalence and abundance of *G. vaginalis* in women with BV [16]. Past research by our own team also saw that *Clostridia sp.* BVAB2 was more abundant in samples with Nugent scores intermediate for BV and consistent for BV, than samples which were inconsistent for BV [47]. These past studies are in line with findings of the present study. Our BV-associated CSTs, CST III, IVA, IVC and IVD were associated with an increasing number of BV episodes experienced in the past two months and year. Specifically, *G. vaginalis* Group A and *G. vaginalis* Group C were found to be significantly associated with the number of episodes of BV in the past two months. *G. vaginalis* Group A and *G. vagianlis* Group C were 1.7 and 3 times more abundant among women with one or more BV episodes in the past two months, respectively. Also, *G. vaginalis* Group A, *G. vaginalis* Group C and *Clostridia sp.* BVAB2, which are present in our BV-CSTs, were found to be significantly associated with the number of episodes of BV in the past year (Table 3.15). *G. vaginalis* Group

A, *G. vaginalis* Group C and *Clostridia sp.* BVAB2 were 1.4, 2.6 and 2.7 times more abundant among women with one or more BV episodes in the past year, respectively. Given the link between BV and these bacterial species, it was not surprising that women with profiles consisting of *G. vaginalis* Group A, *G. vaginalis* Group C and *Clostridia sp.* BVAB2 would experience a greater number of BV episodes.

The positive associations between oral sex, use of hormonal contraception, lifetime history of trichomoniasis, and an increasing number of BV episodes experienced in the past two months and year, and our BV-associated CSTs suggests that these activities could play a role in the etiology/pathogenesis of recurrent BV, perhaps introducing pathogenic bacteria or making the environment hospitable to BV-associated bacteria, causing a shift in the microbiota towards BV.

'Healthy' Community State Types Trends

Trends in our data also suggested an association between our "healthy" CST (CST I), and Shannon's Diversity Index, abnormal discharge in the past 48 hours, no method of contraception used, and number of sexual partners in the past year (Table 3.10). In comparison to healthy asymptomatic women, past research has indicated that women with BV have more diverse bacterial profiles, where bacterial species are more prevalent and abundant [9,16,19,41,113,119]. Shannon's Diversity Index was positively associated with CST I in our study, indicating that CST I was associated with greater diversity. In our cohort, abnormal discharge in the past 48 hours was also found to be positively associated with CST I. Marrazzo, Thomas, Fiedler, et al., and Marrazzo, Thomas, Agnew, et al. found that recent change in vaginal discharge was linked to BV [53,112]. Furthermore, lack of contraception use and number of sexual partners were positively associated with CST I. As indicated previously, there is evidence that having higher numbers of sexual partners increases one's risk of BV [26,52,56,80–83]. As mentioned

earlier, the alkaline property of semen is suggested to cause a loss of vaginal acidity for hours after intercourse and may allow for the overgrowth of anaerobic bacteria and/or affect the maintenance of H_2O_2-producing lactobacilli, resulting in an imbalance of the vaginal environment and subsequently the acquisition of BV [80,82]. Additionally, males may be carriers of BV-associated organisms, as studies have isolated BV-associated organisms from the male genital tract [83,95–97]. However, Li et al. suggested a protective antibacterial effect of prostasomes (prostate-derived organelles in semen) or prostasome-derived proteins [86]. This could potentially lead to an inhospitable environment for BV-associated organisms, and allow *Lactobacillus* species to colonize. Furthermore, studies have noted the protective effect of hormonal contraceptives [79,84,85,87], as well as condom use [83,84]. There is conflicting research suggesting that an increasing number of sexual partners, combined with lack of contraceptive use, may be either detrimental or protective to the vaginal environment. Either these activities create a hospitable environment and allow BV-association organisms to colonize the vaginal microbiota, or create an inhospitable environment preventing the introduction of BV-associated organisms. Further research is necessary to more clearly define this relationship to BV. At first glance, these activities would not be thought to be associated with a "healthy" CST, instead more to be linked to BV-associated CSTs. These associations could be an artifact of a small sample size, which may not exist in a larger sample size. Moreover, our cross-sectional design has allowed us to take a snapshot of these participants' vaginal microbiomes in a moment of time. Shifts within the vaginal environment would not be captured, and it is likely that these women were captured in a moment when their vaginal microbiomes were "healthy", either through treatment or natural shifts. This effect could also be exacerbated by our small sample size.

Shannon's Diversity Index Trends

As noted earlier, past research has indicated a relationship between BV and increased bacterial diversity. Lifetime history of chlamydia was the only clinical variable to be significantly

associated with Shannon's Diversity Index, where mean diversity was higher in women who had a lifetime history of chlamydia ($p<0.01$) (Table 3.11). Although there was a positive relationship between diversity, and indirectly BV, and lifetime history of Chlamydia, it is unclear whether BV predisposes the vaginal environment for Chlamydia, or vice versa. Women with BV have been shown to be at an increased risk for the acquisition of Chlamydia [203–206]. However, this association could be a result of the two conditions having similar risk factors [207]. Chlamydia may cause a disruption in the vaginal environment, allowing for the colonization of other bacteria, leading to increased diversity. Or perhaps an already diverse vaginal environment is susceptible to infection by Chlamydia. In addition, bacterial diversity and *Lactobacillus* dominance are usually not analogous terms. Research has shown that women with BV have more diverse vaginal environments colonized by an array of bacterial species, whereas healthy-asymptomatic women have less diverse vaginal environment dominated by *Lactobacillus* species. Recently, *Lactobacillus* has illustrated protective effects against Chlamydia, where its production of lactic acid is able to mediate Chlamydia inhibition [208–211]. This previous research illustrating a relationship between bacterial diversity and Chlamydia is supported by our own findings.

No other significant associations were seen, but trends suggested positive associations between Shannon's Diversity Index and use of illicit substances and alcohol, use of antimicrobials in the past three months, lifetime history of yeast, and number of BV episodes in the past year and lifetime. Trends also suggested a negative association between use prescription and non-prescription drugs in the past two months, and Shannon's Diversity Index (Table 3.11).

Not many studies have analyzed the relationship between BV, and illicit substance and alcohol use. Research has looked at the risk of sexually transmitted diseases (STDs) among injection drug users (IDUs), suggesting that this population engages in riskier sexual behaviours

which puts them at a higher risk for STDs [212–214]. Sexual activity is a suggested risk factor for BV, thus indirectly illicit substance use and alcohol consumption could potentially be linked to BV as well. However, past research has found that although this population engages in risky sexual behaviours, this does not correlate with a higher prevalence of STDs. Poulin et al. who examined the prevalence of STDs among 96 current female drug users in Quebec, Canada saw that IDUs were more likely to have more than five heterosexual partners than non-IDUs, and more likely to report commercial sexual partners in the previous six months [215]. The prevalence of STD prevalence rates were 6.3% (6/96) in IDUs and 4.6% (7/152) in non-IDUs [215]. Similar findings were found by Latka et al. who found that half their study cohort reported more than two sexual partners in the past six months, and less than 1/5 reported consistent condom use, and the prevalence of chlamydia and gonorrhea was 5.2% and 2%, respectively [212]. However, Bachmann et al. found that 61% of IDUs in the population had BV [216]. Plitt et al. found similar results, of 112 participants who were IDUs, 56% were positive for BV, and 5% were intermediate for BV [214]. As past literature has indicated a relationship between BV, and bacterial diversity [9,16,19,41,113,119], the positive association between substance use and Shannon's Diversity seen in our study, is in line with past studies. Few studies are present that examine this relationship, thus further research needs to be conducted in order to define and examine this association.

Our cohort consists of women with recurrent BV, many using antimicrobials for BV treatment, in addition to other antimicrobials to treat yeast, UTI and HSV infections. As noted earlier, antimicrobials can disrupt the vaginal microbiome, which can lead to colonization of BV-associated bacteria and increased diversity within the vaginal environment. Moreover, as mentioned earlier, many women may not be responding to current treatments as the BV recurrence rate is 30% [14,60]. This recurrence rate increases the longer women are off treatment regimens [20,64,65]. This indicates that although these women were on antimicrobials, which are

targeted towards anaerobic organisms with the goal to reset the vaginal environment to *Lactobacillus*-dominance, in many women BV will re-occur and increased diversity may occur within the vaginal microbiota. This was supported by our findings where there was a positive relationship between use of antimicrobials within the past three months, and Shannon's Diversity Index.

We also saw a positive relationship between lifetime history of yeast, and Shannon's Diversity Index. Vitali et al. found that the vaginal microbiota was not associated with dysbiosis in women with candidiasis [217]. Instead these women had vaginal microbiotas dominated by a diversity of *Lactobacillus* species. There was a shift in these women, from H_2O_2 producing *Lactobacillus* species, *L. acidophilus, L. gasseri,* and *L. vaginalis*, to non- H_2O_2 *Lactobacillus* species, *L. iners.* In addition, there was also an increase seen in the relative abundance of *Lactobacilli* in women with candidiasis compared to uninfected women [217]. As noted before, *L. iners* is suggested to be a transitional bacterial species. Past studies have also supported that it is a poor producer of H_2O_2. Thus its dominance in the vaginal microbiota could permit the growth of disruptive bacterial species at higher proportions, and lead to BV and increased diversity in the vaginal microbiota [9,121]. This is supported by past studies, which have seen the presence of candida species in women with BV [218–220]. Rivers et al. saw that 33% of women diagnosed with BV were also colonized by yeast species [220]. Other studies have seen that 74% of women with recurrent BV had at least one vaginal culture positive for yeast [218], whereas others saw that in commercial sex workers in Kenya, 27% of those who were symptomatic with vulvovaginal candidiasis also had BV [219]. We found a positive relationship between the lifetime history of yeast and Shannon's Diversity Index, and past research seems to support this finding where yeast and increased diversity are linked.

As indicated above, there was a positive relationship seen in our study between number of BV episodes within the past two months and year, and our BV-associated CSTs, as well as with the relative abundance of *G. vaginalis* Group A, *G. vaginalis* Group C and *Clostridia sp.* BVAB2 (Table 3.15). This aligns with our additional association between Shannon's Diversity Index, and number of BV episodes within the past year and lifetime. Furthermore, as mentioned multiple times, diversity has been linked with BV in the past, thus we would anticipate that women who experienced a greater number of BV episodes would also have diverse profiles.

Women in our cohort were taking non-prescription drugs such as vitamin supplements for general health, and probiotics, as well as prescription drugs such as hormonal contraceptives. As indicated previously, several studies have also noted the protective effect of hormonal contraceptives [79,84,85,87]. In addition to probiotics introducing *Lactobacillus* species into the vaginal environment, these prescriptions could cause an environment dominated by *Lactobacillus* and necessary substrates, which in turn would decrease the diversity of the vaginal environment. However, in our cohort we saw that women who used hormonal contraceptives and probiotics were found in both low and high Shannon's Diversity index categories. Further research is needed to elucidate this relationship between use of prescription and non-prescription drugs in the past two months, and Shannon's Diversity Index.

Comparison to Healthy-Asymptomatic Cohort

Past research was done by our group to characterize the vaginal microbiome of 310 healthy asymptomatic women. We have a power of 55% to detect differences between this population, and our own of women with recurrent BV. As a consequence, results were more descriptive. We saw that our recurrent BV population had significantly fewer women with Nugent scores inconsistent for BV, and more Nugent scores consistent for BV and intermediate BV, compared to the healthy-asymptomatic population (Table 3.22). Furthermore, we also saw a

statistically significant difference in the CSTs present in each population; 50.3% (156/310) of profiles were dominated by *L. crispatus* and fell into the CSTI cluster in our healthy asymptomatic cohort, whereas in our recurrent BV cohort 77% (20/26) of women fell into BV-associated CST clusters (III, IVA, IVC, IVD) (Table 3.23). This supports that our analysis is able to define healthy populations and dysbiotic populations.

When recruiting women into the healthy-asymptomatic population, it was done so that the population would be representative of the British Columbia (BC) Lower Mainland. We see that this population and our recurrent BV population were not significantly different from each other across demographic characteristics, which suggests that our recurrent BV population was also representative of the BC Lower Mainland (Table 3.16). Across the behavioural and clinical characteristics, the majority were not significantly different between the two populations (Table 3.17-3.21). Unsurprisingly, significantly more women with recurrent BV used antimicrobials in the past three months and experienced vaginal symptoms in the past 48 hours and 2 weeks. This reflects our recruitment, and inclusion criteria, where we sought to enrol women with BV and BV-like symptoms into the recurrent BV population, and by definition, BV symptoms were not expected in our healthy-asymptomatic population. In addition, one of the inclusion criteria for our healthy-asymptomatic population was that women did not use antimicrobials in the past three months.

Significantly more women with recurrent BV reported substance use (within three months, or three months prior to the study visit). As indicated previously, there were not many studies which looked at the relationship of the vaginal microbiome and substance use, however two studies were present which indicated a high prevalence of BV in IDUs [214,216]. We see our study supports this association, as more women report substance use in our recurrent BV

populations compared to our healthy-asymptomatic population. However, explanations for the potential relationship between BV and substance use are less clear.

Past research has indicated that women with recurrent BV have more diverse bacterial profiles than healthy asymptomatic women; where bacterial species are more prevalent and abundant [9,16,19,41,113]. Fredricks et al. noted that the mean number of bacterial phylotypes seen in women with BV was 12.6 compared to 3.3 seen in women without BV [113]. Shannon's Diversity Index, Pielou's evenness, and species richness (Chao1) was slightly higher in our population of women with recurrent BV, however we did not find a statistical difference between the diversity statistics (Table 3.24).

Relative abundances of bacterial species differed between our two populations of healthy-asymptomatic women, and women with recurrent BV. *Propionibacterium granulosum, Streptococcus constellatus, Streptococcus agalactiae, G. vaginalis* Group B, and *L. crispatus* were more abundant in the healthy-asymptomatic women versus our recurrent BV population (Table 3.13). *Propionibacterium granulosum*, and *Streptococcus constellatus* have not been described in literature pertaining to the vaginal microbiome. *Propionibacterium granulosum* has been linked to the skin, and has a role in acne. Although *Streptococcus constellatus* has not been specifically noted in literature to play a role in the vaginal environment, *Streptococcus* species have. Past research has noted the presence of *Streptococcus* species in healthy-asymptomatic women, more specifically within heterogeneous CSTs, dominated by multiple bacterial species [39,40,47,217]. *Streptococcus* species are also lactic acid producers like *Lactobacillus* species, demonstrating their functional redundancy within the vaginal microbiota, able to conserve the overall community function [39]. *Streptococcus agalactiae*, also known as Group B Streptococcus (GBS) was also seen to be more abundant in healthy-asymptomatic women in our study. GBS is a known vaginal colonizer, and past studies have documented that

GBS is present in the vagina and rectum of approximately 10-30% of women [221]. Although not necessarily harmful in non-pregnant women, it increases the frequency of premature labour and perinatal transmission, resulting in high neonate mortality and morbidity [222]. Leclair et al. noted that the prevalence of GBS was significantly increased in women with vaginitis; where 18% of women without vaginitis had positive GBS cultures, compared to 33% and 34% of women with common vaginitis (vaginal candidiasis, bacterial vaginosis and trichomoniasis) and inflammatory vaginitis, respectively [221]. Leclair et al. also noted that women with desquamative inflammatory vaginitis had a fivefold increase in the risk of being GBS-positive [221]. These findings may not align with our own, but Leclair et al. looked at a slightly different population that ours, one not completely made up of women with recurrent BV [221]. However, they do note an increased risk of GBS and disruptions in the vaginal flora. And research has illustrated the presence of GBS in healthy women, thus its presence here is expected, but we would have anticipated its relative abundance to be higher in our population of women with recurrent BV.

Furthermore, *G. vaginalis* Group B was seen to be more abundant in our healthy-asymptomatic population. Previous studies have also detected *G. vaginalis* in the vaginal microbiome of healthy and asymptomatic women [16,19,39,40,59–63]. However, this result was unexpected. Previous work done by our group investigated the distribution of *G. vaginalis* subgroups in women with and without BV. Group B was significantly more abundant in women with BV than any other subgroup, and more abundant in women with BV compared to a normal vaginal microbiome [63]. Moreover, we saw *L. crispatus* was more abundant in healthy-asymptomatic women, and this is supported by a vast amount of literature that notes its role in the healthy vaginal microbiome [9,13,38,39,41,42,47].

Relative abundances of *L. plantarum*, *Streptococcus devriesei, Actinobacteria sp., Clostridia sp* BVAB2, and *G. vaginalis* Group C were higher in women with recurrent BV, versus

healthy-asymptomatic women (Table 3.13). *L. plantarum* is a known probiotic, used to establish a healthy vaginal microbiome [223], making its higher abundance in women with recurrent BV unexpected. Also, there is no literature describing a relationship between *Streptococcus devriesei* and the vaginal environment. However as noted earlier, literature illustrates a relationship with *Streptococcus* species and the vaginal environment, where their lactic acid production may serve to conserve the overall function of the vaginal microbiota [39]. Its suggested relationship with healthy-asymptomatic women also makes its abundance in women with recurrent BV unexpected. However, the higher abundances of *Actinobacteria sp., Clostridia sp* BVAB2, and *G. vaginalis* Group C in women with recurrent BV aligns with past literature which has noted that these bacterial species are associated with BV [16,19,41,42,47,62,113,193].

Addressing Hypotheses

We hypothesized that utilizing *cpn*60 sequencing methods, we would obtain predictive vaginal microbiome profiles for women with recurrent bacterial vaginosis as the foundation for diagnostic tools. We anticipated that certain distinct profiles would be more prevalent in this population compared to the healthy-asymptomatic population; more specifically that fewer women in our population would possess profiles dominated by *Lactobacillus* species, and more women would possess distinct profiles dominated by *G. vaginalis* species and heterogeneous profiles. This was based on Gajer et al. who noted that some taxa in CST IVB, *Atopobium, Gardnerella, Mobiluncus, Prevotella,* and *Sneathia*, have previously been associated with BV [40,62,113,114]. Another study conducted by Mehta et al. found that the proportion of women clinically diagnosed with BV was lower in CSTs containing a higher relative abundance of *Lactobacillus* species [168]. Even a previous study conducted by our own group had indicated associations between BV+ Nugent scores with CST IVA, IVC and IVD, whereas BV- Nugent scores were associated with *Lactobacillus*-dominated CST I, III, and V [47]. Determining distinct profiles specific to women with recurrent BV would allow for more targeted treatment. Our findings

support this hypothesis, as we saw six distinct bacterial profiles in our population. In comparison to the healthy-asymptomatic population, CST III, CST IVA, CST IVC and CST IVD were significantly more abundant ($p<0.05$). These profiles were heterogeneous and dominated by *G. vaginalis* species, as well as *L. iners*.

We also hypothesized that each distinct profile would be associated with its own set of behavioural, clinical and demographic variables. Literature has made several suggestions on the relationship between BV, bacterial species, and behavioral, clinical and demographic characteristics [26,41,52,53,56,62,79–91,93,94,100,107,112–114]. We hoped to find certain patterns that would improve understanding of vaginal dysbiosis, and be predictive of this genital condition. Our findings support this hypothesis, as we saw several trends suggesting an association between CSTs and oral sex, use of hormonal contraception, lifetime history of trichomoniasis, number of BV episodes experienced in the past two months and year, Shannon's Diversity Index, abnormal discharge in the past 48 hours, no method contraception used and number of sexual partners in the past year.

Limitations

There were several limitations in our study. Recruitment of women into this study was difficult, as it was hard to capture women when they were experiencing symptoms. Most women with this condition are seen in the community and treated prior to coming to the sub-speciality referral clinic or research clinic. As a result we did not reach our target sample size of 50 women. We recruited 26 women, which reduced the power to detect differences between CSTs, as well as compared to our healthy-asymptomatic population. Trends seen here may be an artifact of a small sample size. A larger sample size would be needed to study the diversity seen with recurrent BV. Secondly, on the day of assessment some women were asymptomatic although having a history of recurrent BV. As it was a difficult study to recruit for these women

remained in our population, but we can argue that these asymptomatic women should be removed from our study population, or should be analyzed as a different subset. Also, we had a cross-sectional design, which looked at women with recurrent BV at a single time-point which does not capture how their vaginal microbiome may change over the course of this condition. Furthermore, *cpn*60 was unable to detect *Mycoplasma* and *Ureaplasma* species due to their lack of a *cpn*60 gene. We see that Mollicutes PCR is not always robust, and does not detect the identities of specific *Mycoplasma* species. It was able to indicate "yes" or "no" to the presence of Mollicutes, where a positive Mollicutes result could mean only positive *Ureaplasma*. Moreover, in one participant we saw a negative Mollicutes result yet a positive *Ureaplasma* result. This could be due to several reasons; technical failure of the Mollicutes specific PCR or due to sequence variation, or Mollicutes were present below the detection threshold. These suggest that the Mollicutes PCR may not be as robust as the *Ureaplasma* assay. Although this type of analysis did not allow us to compare bacterial proportions, we still gained insight into the presence or absence of these specific bacterial taxa.

Future Directions

In the future we would want to increase our sample size, and look at a larger population of women with recurrent BV. This would give us the power to detect differences between our CSTs, and for a comparison to our healthy-asymptomatic population. With the sample size we had, we saw that recurrent BV was quite diverse. A larger sample size would allow us to better define this diversity. This would also allow us to re-analyze the trends we saw in our current study, and determine whether the associations observed were artifacts of a small sample size or in fact statistically significant. With further research into this population, we would be able to better define the associations we saw. Moreover, future research needs to be conducted to determine the dynamics of recurrent BV, and how the vaginal microbiome shifts throughout this condition. There are several longitudinal studies published, but no culture-independent studies

characterize the vaginal microbiome over time to determine how recurrent BV transitions. In addition, studying the transcriptome and metabolome of women with recurrent BV would determine not only what bacterial species are present in the vaginal microbiota, but also provide information on function and what metabolites they are producing. We may determine functional redundancy within the vaginal microbiome, where multiple bacterial species are present but they working towards the same goal.

Conclusion

Overall, we sought to characterize the vaginal microbiomes of women with recurrent BV to determine distinct bacterial profiles and descriptive variables predictive of this condition. 26 women with a history of recurrent BV were enrolled from various clinics in Vancouver BC, and were representative of the Lower Mainland. The *cpn*60 universal target gene was used to characterize the vaginal microbiome of these women. Using this gene target overcomes limitations of the 16S rRNA target, being highly discriminatory and able to provide sufficient species-level identification and uniform interspecies variability, yield higher numbers of OTUs at different cut-off points, as well as identify subgroups within *G. vaginalis.*

Using the *cpn*60 universal target gene, along with hierarchal clustering, we were able to identify six distinct microbiome clusters within our population. In this modest sample of women with recurrent BV we saw that although all women had a history of recurrent BV, not all had formal clinical diagnosis of BV on the day of sampling by Nugent's score. However, the *cpn*60 generated profiles show a high diversity of microbial species, and were much more illustrative than Nugent scores alone. The ability of *cpn*60 based testing to subdivide these women into different community state types promises to allow direction of therapy and prediction of natural course of disease.

Trends were found between our CSTs, and several descriptive characteristics in our population. CST III, CST IVA, CST IVC and CST IVD were designated the BV-associated CSTs due to the presence of BV-associated bacteria within these clusters. Trends suggested an association between these clusters and oral sex, use of hormonal contraception, lifetime history of trichomoniasis, and an increasing number of BV episodes experienced in the past two months and year. CST I, the *L. crispatus* dominated cluster, was associated with Shannon's Diversity Index, abnormal discharge in the past 48 hours, no method contraception used and number of sexual partners in the past year. Furthermore, specific bacterial species were found to be associated with number of BV episodes in the past two months and year. Associations between Shannon's Diversity Index, and lifetime history of chlamydia, use of illicit substances and alcohol, use of antimicrobials in the past three month, lifetime history of yeast, number of BV episodes in the past year and lifetime, and use of prescription and non-prescription drugs in the past two months, were seen. As the etiology of recurrent BV remains a topic of debate, our trends suggest novel associations not only between diversity and descriptive variables, but also between CSTs, and descriptive variables, which points the direction future research needs to take. Our ability to subdivide these women into CSTs, and then relate these CSTs to behavioural, clinical and demographic variables has important consequences. This illustrates that recurrent BV may consist of several CSTs, each CST with a distinct profile of descriptive variables, indicating that not all women with recurrent BV are the identical.

Our research project has important consequences for the improvement of women's health, and for our understanding of the distinction between vaginal health and disease. Not only are our findings important for vaginal microbiome research, but they illustrate the advantages of *cpn*60 in microbiome research. In the future, recruitment of more women and sequencing more samples will help further define the bacterial communities which exist in the vaginal microbiome of women with recurrent BV.

4.2 Virome Discussion

Clinical Cohorts

Our clinical cohorts provided a breadth of demographic, behavioural and clinical information about the population, and as a result were not comparable to the cohorts analyzed in the sparsely available literature of the vaginal virome. With regards to sample size, fewer healthy-asymptomatic women and more HIV-positive women were analyzed in our study compared to previous reports [140,156,162], and no studies have analyzed the viral species within women with recurrent BV. Fifty-four women were analyzed in our study. Twenty-one women were non-pregnant, HIV-negative, not menopausal, had a regular menstrual cycle (~28 days), between the ages 20 - 42 years (mean 28 years), did not use either systemic or topical antimicrobial therapy within the prior month, and were deemed healthy and asymptomatic through clinical assessment. Twenty-five women were not pregnant, HIV-positive, not menopausal, and between the ages 23 - 48 (mean 37 years). Eight women were not pregnant, not menopausal, HIV-negative, between the ages 21 - 42 (mean 29 years), and had a history of recurrent BV.

Our three cohorts were significantly different from each other on the basis of demographics. Analyses illustrated that our HIV-positive cohort was significantly older than our healthy-asymptomatic and recurrent BV cohort, and that more women were of Aboriginal, South Asian, Black and other origins. Also, significantly more women in our HIV-positive cohort were multiparous, and significantly more women in our healthy-asymptomatic and recurrent BV cohorts were nulliparous. We also saw a statistically significant difference in symptoms, expectedly more women in our recurrent BV population experienced abnormal discharge in the past two weeks and past 48 hours, and irritation and discomfort in the past 48 hours. Also as expected, significantly more women in our HIV-positive, and recurrent BV cohorts used antimicrobials in the past three months, compared to our healthy-asymptomatic cohort where

part of the exclusion criteria was use of systemic or topical antimicrobials in the prior month. There were no statistically significant differences across the remaining demographics, clinical and behavioural characteristics across cohorts (Table 3.67-3.74). Of the diversity statistics calculated, none were significantly different between the healthy-asymptomatic, HIV-positive and recurrent BV cohort (Table 3.75). Viral groups between the three cohorts were not significantly different from each other either (Table 3.76).

Analysis of our these three different cohorts will allow us to not only characterize the vaginal virome and identify viral species in this environment, but also determine viral profiles associated with health and disease. We can determine what the underlying “healthy normal’ virome profiles are, what the spectrum of coinfection looks like in HIV positive women, as well as understanding the diversity behind recurrent BV, and piecing together its etiology.

Sampling Depth

In metagenomic studies, in general, sampling depth can be a concern as we wish to capture samples, which are representative of the vaginal environment [224–226]. To ensure sufficient depth of viral sampling, Good’s coverage equation was applied [227]. The mean good’s coverage for our three cohorts was 98% (range: 0.89 - 1.00) (Table 3.58-3.60). This implies that for every 100 additional sequence reads, less than two additional species would be discovered. In addition, Chao1 estimates for each sample were close to the actual number of species detected, therefore providing further support to suggest that sampling was sufficient, and samples are representative of the communities seen in the vaginal environment (Table 3.58-3.60).

Furthermore, rarefaction curves also suggested sufficient sampling depth. As indicated earlier, when sampling depth is sufficient, rarefaction curves plateau, reaching a slope of zero. Each rarefaction curve represents an individual participant (Figure 3.41-3.43). In Figure 3.41-

3.43, we can identify curves which plateau, and curves which do not. Curves clustered to the left of Figure 3.41-3.43 appear to be still rising and do not approach a slope of zero. This suggests that not all participants in our study have sufficient sampling depth, and further sampling may be required for sampling to be representative of the vaginal environment. However, this similar graphing artifact was seen in the earlier bacteriome section, where participants with fewer reads have the curvature of their rarefaction curves obscured by the large range of the x-axis. As it was difficult to visually determine if each curve was reaching a plateau, the slope was calculated at five different positions in each curve: the total reads of each participant, and the total reads less 1, 5, 10, and 100 read(s) (Table 3.61-3.63). Looking at the slope at each of these points, we see that as we reach the total reads for each participant, the slope is decreasing and reaching zero; something we cannot see visually. Overall, the diversity indices and the rarefaction curves provide evidence to suggest that sampling depth was sufficient in our study, and our samples may be representative of the vaginal virome of women.

In addition to sufficient sampling depth, we were able to provide proof of principle regarding the classification tool used; Taxonomer [189]. A subset of women in our HIV-positive population, are also enrolled in another study conducted by our team to look at HPV types in this population (Table 3.56). Thus, we have HPV data available for those enrolled women. HPV type information from our alternative study was compared to the classified HPV types, and their abundances in this study. Although this does not hold true for all women and all their HPV types, women who were positive for a HPV type in the alternative study, also have a relatively high abundance in that specific HPV type in our own study. Through these comparisons we are able to illustrate confidence in our results.

Vaginal Virome Findings

As mentioned previously, the majority of research on the vaginal microbiome aims to define the bacterial communities which occupy this environment. There is limited research which seeks to define viral species; in fact currently there are no studies whose primary goal is to describe the vaginal virome. However, there are few studies which aimed to define HPV types within the vaginal microbiome and one exploratory study which aimed to define the microbial communities in the human microbiome. Past research conducted by Wylie et al. sampled 102 healthy-asymptomatic individuals from five major body sites; nose, skin, mouth, stool and vagina [140]. The vaginal component of this study illustrated a dominance of papillomaviruses, where 37.5% of subjects carried one or more *Alphapapillomavirus*. Subjects carried as many as four distinct *Alphapapillomavirus*, including oncogenic types HPV-16, and HPV-18. They also suggested a dynamic relationship between the viral and bacterial species which occupy the vaginal environment. *Alphapapillomavirus* were more common in an individual's sample with more bacterial diversity; whose vaginal microbiome showed less than 85% *Lactobacillus*-dominance and had an increase in anaerobic bacteria [140]. This research coincides with Ma et al. who sought to define HPV types present in the human microbiome [162]. Research conducted by this group saw that the prevalence of HPV in the vaginal microbiome of healthy women was 41.5%, and that more than 50% of their subjects were coinfected with two to three HPV types. Most HPV types identified in the vagina belonged to the genus *Alphapapillomavirus*. Moreover, of those types that have been previously characterized to be high-risk, ten were found in nine women in their cohort; who did not have clinical disease (Ma 2014). Unlike these past studies, we found that the prevalence of HPV in our healthy-asymptomatic population was higher. HPV was detected in all women in our healthy-asymptomatic cohort. Fifty-one different HPV types were detected, and women had at least four distinct *Alphapapillomavirus* with low risk HPV-89 being the most abundant. Women in our study had on average 17.9 (4 – 30) different HPV types, which was higher than previous reports

[140,162]. Although not in line with past research entirely, we saw that similarly to past research oncogenic types were also detected in our population; HPV-33, HPV-45, HPV-51, HPV-52, and HPV-56. Also, we saw that most HPV types identified belonged to the *Alphapapillomavirus* genus.

An alternative study conducted by Ameur et al. aimed to profile the vaginal microbiome of 20 HIV-positive women to determine the spectrum of coinfection [156]. Forty-six different HPV types were present in this cohort, with 5 to 21 HPV types present per woman. The most abundant type present in this population was HPV-58, followed by a number of other high-risk HPV types [156]. Women who are infected with HIV are at a higher risk for opportunistic infection due to their suppressed immune system, and can be infected by HSV, HPV, and/or CMV among viral and bacterial pathogens [154–156]. As consequence, it would be expected that the prevalence of HPV would be higher in HIV-positive women versus healthy-asymptomatic women [156]. In our cohort of HIV-positive women, we saw that HPV was detected in all women, and women had at least four distinct *Alphapapillomavirus*. In total, 78 different HPV types were present, and on average, women had 21.2 HPV types (4 – 41), including oncogenic types 18, 35, 33, 45, 51, 52, 56. HPV-62 was the most abundant. In comparison to past research, we detected more HPV types in total, and more HPV types were seen per women. Although not similar to past research, we see that a greater number of different HPV types and oncogenic HPV types were seen in our HIV-positive population compared to our healthy-asymptomatic population, as well as having a higher average of HPV types per participant; which is in line with research that indicates that the HPV would be more prevalent in a HIV-positive population than healthy.

Unlike these past studies, we did not limit our characterization of the vaginal virome to DNA viruses and HPV types, but we sought to complete a comprehensive identification of all viral and phage species present in the vaginal microbiome. Seven distinct viral groups were

detected in this study (Figure 3.9). 22% (12/54) of women fell into viral groups I and IV; strictly dominated by the family *Papillomaviridae* and its species. 60% (32/54) of women fell into mixed dominance clusters Groups II and III, consisting of *Alphapapillomavirus 3* and *Coccolithovirus*, and *Alphapapillomavirus 8*, phage species and chlorella viruses, respectively. The remaining 18% (10/54) of women fell into Groups V, VI and VII, dominated by *Lactobacillus* phage species, human herpesvirus 4, and *Gammaretrovirus* and *Podoviridae* species, respectively (Table 3.53). The majority of women in our healthy-asymptomatic and recurrent BV cohort fell into viral group II, whereas the majority of women in our HIV-positive cohort fell into viral group III (Table 3.54, 3.55, 3.57). Our three cohorts were not significantly different for each other in regards to viral groups (Table 3.76). Six species were detected in all samples, and corresponded to *Alphapapillomavirus 3*, *6* and *10*, human herpesvirus 4, *Coccolithovirus*, and unclassified *Papillomaviridae* species. Unique species which were detected in all participants of the healthy-asymptomatic cohort included *Alphapapillomavirus 14*, *Glypta fumiferanae ichnovirus*, human herpesvirus 6A, human mastadenovirus D, *Lymphocryptovirus*, *Pandoravirus salinus*, and *Staphylococcus virus G1*. *Alphapapillomavirus 9* was also seen in all participants of the HIV-positive population, and in the recurrent BV cohort, *Alphapapillomavirus 9*, *14*, *Glypta fumiferanae ichnovirus*, human mastadenovirus D, *Lymphocryptovirus*, *Pandoravirus salinus*, and unclassified *Siphoviridae* species were also detected in all participants. We also saw an association between *Enterobacteria phage N15*, and our HIV-positive cohort, suggesting a role in HIV pathogenicity or coinfection. Many of these species and families detected in our cohorts, and viral groups have not been previously linked to the vaginal environment nor is their clinical relevance to humans described, making their detection here novel. Several species were present consistently throughout each cohort, and were unique to that population. These species could represent what is “normal” for that population, and only detected in that specific cohort. More research is required to further characterize these viral groups, understand the relationship between these viral species to these cohorts, and to each other, and their clinical significance.

Viewing all three cohorts together, distinct viral patterns could not be seen within each bacteriome CST, as well as when *Lactobacillus* dominated, *L. iners* dominated and heterogeneous profiles were compared (Figure 3.13, 3.14). However, in viewing cohorts separately we saw that the majority of healthy-asymptomatic women with *L. crispatus* dominated profiles (CST I) had viral profiles dominated by *Papillomaviridae* species (viral groups I and IV) (Figure 3.18). This pattern was specific to *L. crispatus* dominated profiles, and was not seen in other *Lactobacillus*-dominated profiles which also fell into viral groups II and V (Figure 3.19). However, in our recurrent BV population we saw the opposite trend. *Lactobacillus*-dominated profiles (CST I) were dominated by phage species and fell into viral groups V and VII, and did not have an abundance of *Papillomaviridae*. In comparison, *L. iners* dominated profiles (CST III) had a greater abundance of *Papillomaviridae* (Group II and III), and heterogeneous profiles (CST IVA, IVC, IVD) were the most abundant in *Papillomaviridae* (Group I – III). No other patterns were seen within bacteriome CSTs, as well as between *Lactobacillus*-dominated, *L. iners* dominated, and heterogeneous profiles in our cohorts (Figure 3.37, 3.38). The difference seen between CST I in our healthy-asymptomatic and recurrent BV populations, may be due to the fact that these women are not similar. Although falling into CST I, denoted the “healthy” CST, these women in the recurrent BV population do have a history of recurrent BV unlike those within CST I in the healthy-asymptomatic cohort. They may also be in a transition phase with a ‘normal’ profile in the bacteriome but the virome revealing the instability in the microbiome.

Past literature aligns with the pattern we detected in our recurrent BV cohort, however not with the pattern we detected in our healthy-asymptomatic cohort. As indicated earlier, Wylie et al. found *Alphapapillomavirus* were more common in an individual's sample with more bacterial diversity, and suggested a dynamic relationship between viral and bacterial species [140]. Brotman et al. also detected the same pattern within their population of women enrolled in a

douching cessation study. 45% and 12% of women with bacterial profiles which fell into *L. crispatus* dominated CST I and *L. gasseri* CST II, respectively, had HPV-positive samples [107]. Whereas, women who bacterial profiles dominated by *L. iners* (CST III) and low-*Lactobacillus* dominance (CST IV), had higher proportions of positive HPV samples; 72% and 71% [107]. Past research has shown a relationship between douching and vaginal dysbiosis [26,52,93,94,103,104]; thus the relationship Brotman et al 2014 describes here between CSTs and HPV is similar to the pattern we detected in our own cohort of women with recurrent BV. Our findings with our cohort of women with recurrent BV align with past research conducted in healthy-asymptomatic women, and women who engage in douching practices. However, our findings within the healthy-asymptomatic population are opposite to what literature has described, and we do not see this trend with our HIV-positive population. Wylie et al., and Brotman et al. describe that *Alphapapillomavirus* and HPV are associated with bacterial diversity and certain CSTs, however both studies do not indicate which HPV types or *Alphapapillomavirus* exactly [107,140]. As we see with bacterial species, *Lactobacillus* and *G. vaginalis*, different subtypes can have different roles and functions [9,16,19,26,27,41,63,72]. Specific papillomaviruses may be associated with specific CSTs and bacterial diversity. In our study, we see that *Alphapapillomavirus* 1-14 have been detected in our HIV-positive population; however the same types were not seen within our healthy-asymptomatic and recurrent BV cohorts. In our healthy-asymptomatic cohort, *Alphapapillomavirus 4, 11* and *12* were not detected, and in our recurrent BV cohort, *Alphapapillomavirus 2, 4, 5, 8* and *13* were not detected. Distinct papillomaviruses could be either driving the patterns or taking advantage of the dybiosis. Exact mechanisms behind this relationship are unknown although it is postulated that metabolites of *Lactobacillus* species, specifically lactic acid, may be effective against HPV infection [107]. This coincides with the above mentioned findings, which reported higher prevalence of HPV types and *Alphapapillomavirus* in women with a low relative abundance of *Lactobacillus* species, as well as findings by Clarke et al., which reported higher HPV detection, risk of multiple infections, low-grade intraepithelial

lesions, and *Chlamydia trachomatis* infection among women with elevated pH [228]. Further research is necessary to understand this relationship.

Viral profiles were viewed in relation to Nugent scores, and in our healthy-asymptomatic cohort it appeared that a trend between Nugent scores and phage species existed (Figure 3.20). Women with Nugent scores consistent for BV had more profiles with phage species, compared to women with Nugent scores for intermediate BV and inconsistent for BV, where *Papillomaviridae* species were more abundant. This was in line with the relationship we saw above with bacterial diversity and *Lactobacillus* dominance, and *Papillomaviridae*. However, this pattern was not seen in our recurrent BV population and HIV-positive population, where phage and *Papillomaviridae* species were detected in women with Nugent scores consistent and inconsistent for BV (Figure 3.25, 3.39). We anticipated that more women with recurrent BV would have viral profiles dominated by phage species, and thus fall into viral group V due to past literature suggesting that the reduction of *Lactobacillus* species in women with recurrent BV may be a result of phage species [229]. However of the five viral group V profiles in our populations, one was present in our recurrent BV population, two were present in our HIV-positive population and three in our healthy-asymptomatic population (Figure 3.16). Past research shows a significant difference between the percentage of lysogenic lactobacilli isolated from women with and without BV. Approximately 50% of lysogenic lactobacilli were isolated from women with BV, whereas 30% were isolated from women without BV. Kilic et al. found that isolated vaginal lactobacilli phage could infect a broad range of lactobacilli including *L. crispatus*, *L. jensenii*, *L. gasseri*, *L. fermentum*, and *L. vaginalis* [229]. They suggested that the lysogenic *Lactobacillus* strains may be a source of potentially infectious phages [229]. These findings align with our own, where we see an association between *Lactobacillus* phage species and high Nugent scores.

As patterns emerged within our healthy-asymptomatic and recurrent BV cohort, it was surprising that no patterns were seen in our HIV-positive population between viral groups, and bacteriome CSTs, Nugent scores, and between *Lactobacillus* dominated, *L. iners* dominated and heterogeneous profiles (Figure 3.23-3.25). Viral profiles were viewed in relation to viral loads. Suppressed women fell into Groups I - VII; 69% (11/16) of women had profiles which fell into Groups I - IV, viral groups with varying abundances of *Papillomaviridae*, and 31% (5/16) of women had profiles which fell into non-*Papillomaviridae* groups (Groups V - VII) (Figure 3.26). Women with low level viral loads and unsuppressed women had profiles which fell into Groups I - IV; predominately Group III. Groups I – IV are dominated by *Papillomaviridae* to varying degrees (Figure 3.27, 3.28). We saw a pattern of *Papillomaviridae* abundance with low level and unsuppressed viral loads, suggesting coinfection with HPV types in HIV positive women which is in line with past literature [156,165]. As mentioned earlier, literature has described a relationship between bacterial diversity and *Papillomaviridae*, where *Papillomaviridae* was more prevalent in women who had greater bacterial diversity. We saw similar results in our group of women with low level viral loads; as Shannon's Diversity Index increased (indicated greater bacterial diversity), participants showed a greater *Papillomaviridae* dominance (Figure 3.29). This trend was not seen in suppressed and unsuppressed viral load groups (Figure 3.30, 3.31). And no trends between viral species and Shannon's Diversity Index for bacterial species were seen in the absence of *Papillomaviridae*. In the absence of *Papillomaviridae*, women with low level viral loads showed a pattern of *Siphoviridae* dominance with increasing viral loads; this pattern was not seen in the unsuppressed viral group (Figure 3.32-3.34). The relationship between *Siphoviridae*, and viral load and HIV has not been described in literature. Further research is needed to provide clarity on this trend.

Human Papillomavirus, Human Herpesvirus and Human Immunodeficiency Virus

We were interested in the prevalence of specific viruses, HHV, HPV and HIV in our populations (Figure 3.21, 3.35, 3.40). Overall, we found that in comparison to HIV and HHV, *Alphapapillomavirus* were the most abundant in all three cohorts. As mentioned earlier, *Alphapapillomavirus* were detected in all women in our healthy-asymptomatic cohort as well as the HIV-positive cohort. The same was seen in our recurrent BV cohort. With regards to *Betapapillomavirus* and *Gammapapillomavirus*, these were not as prevalent as *Alphapapillomavirus* in any of the populations. In the healthy-asymptomatic population, the prevalence of *Betapapillomavirus* was 5% (1/21), and the prevalence of *Gammapapillomavirus* was 19% (4/21). The prevalence of *Betapapillomavirus* was higher in our HIV-positive and recurrent BV cohorts, where 32% (8/25) and 50% (4/8) of the population had at least one *Betapapillomavirus*, respectively. The prevalence of *Gammapapillomavirus* in the HIV-positive population was similar to our healthy-asymptomatic population at 20% (5/25), and lower in our recurrent BV cohort at 12.5% (1/8). The alpha genus is predominantly associated with mucosal infections, commonly infecting the anogenital tract and oral cavity, whereas *beta*- and *Gammapapillomavirus* are predominantly associated with cutaneous tumours; skin cancer [191].

After papillomaviruses, HHV4 (also known as Epstein-Barr Virus: EBV) was the most abundant and prevalent in all women in all three populations. This is in line with current literature which states that HHV4 is highly prevalent worldwide, and known to infect and persist for life in >90% of young adults without causing disease [191,230]. EBV has been known to cause infectious mononucleosis (“mono”), Burkitt's lymphoma, as well as other diseases in immunocompromised patients [191,230]. In addition to HHV4, other HHVs detected in our populations included HSV-1 and -2, HHV5, HHV6 and HHV8. Similar to HHV4, infection of HHV6 (part of the *Roseolovirus* genus) is ubiquitous affecting the 90% of the adult population in developed countries. Infection with HHV6 also results in chronic viral [231–233]. This coincided with

our findings; we saw that HHV6 was detected in all populations. HHV6 was prevalent in all participants of the healthy-asymptomatic cohort, in 88% (22/25) of women in the HIV-positive cohort, and 87.5% (7/8) of women in the recurrent BV cohort. HSV-1 and -2 were only detected in our HIV-positive population, where the prevalence was 20% (5/25). As noted before, HSV is the second most prevalent sexually transmitted viral infection worldwide, and furthermore a relationship between BV and HSV has been noted. Therefore, it was unexpected that HSV-1 and -2 were not detected in our healthy-asymptomatic and recurrent BV cohort. However, the prevalence of HSV within the HIV-positive population in our study coincides with past research, which specifies that not only is HSV a suggested risk factor for HIV, but women who are HIV positive also have a higher prevalence of HSV [155,156].

HHV5 also known as *Cytomegalovirus* was detected in all three populations. The prevalence of HHV5 was 19% (4/21) in the healthy-asymptomatic cohort, 36% (9/25) in the HIV-positive cohort, and 25% (2/8) in the recurrent BV cohort. HHV5 prevalence in our study was slightly lower than literature reports. However, our findings are similar to past research which indicates that CMV is prevalent not only in the immunocompromised population, but present in those who are immunocompetent and asymptomatic [146,147]. Moreover, CMV was more prevalent in our HIV-positive population, as was anticipated, as individuals who are infected with HIV are at a higher risk for reactivation of infections due to their cell mediated immune suppression [154–156]. Also coinciding with our findings, past research has indicated a relationship between BV and CMV. To reiterate, a study conducted by Ross et al. 2005, showed that women with BV had increased rates of CMV seroprevalence and seroconversion.

HHV8 is part of the *Rhadinovirus* genus, and also known as *Kaposi's sarcoma-associated herpesvirus*. It is the etiologic agent associated with Kaposi sarcoma, primary effusion lymphoma, and multicentric Castleman's disease [234,235]. In our study, HHV8 was only

detected in the healthy-asymptomatic and HIV-positive populations. The prevalence in our healthy-asymptomatic population was 29% (6/21), and 20% (5/25) in our HIV-positive population. Our findings in the healthy-asymptomatic cohort coincide with an older study, where HHV8 prevalence in the United States ranged from 0-20% [236], however more recent work reports lower prevalence rates in the general population; unlike other human herpesviruses, HHV8 is not widespread in the general population [234,237,238]. However, research reports higher prevalence rates are seen with immunocompromised individuals, along with men who have sex with men, and different geographical regions [234,236,237,239,240].

A low abundance of reads classified as HIV were detected in our HIV-positive cohort and our healthy-asymptomatic cohort. The prevalence of these reads in the HIV-positive and healthy-asymptomatic cohorts were 48% (10/21) and 36% (9/25), respectively. Reads corresponding to HIV in our healthy-asymptomatic cohort ranged from 1 – 4 per participant, with individuals on average having 1.5 reads. The detection of HIV classified reads in our healthy-asymptomatic population may be a result of, although unlikely, contamination or misclassification. However, detection of these low abundant HIV reads argues that we should increase our biologically relevant frequency threshold to four, removing these low abundant reads and thus correcting for this erroneous result. We also saw that reads corresponding to HIV in the HIV-positive population were not overly abundant. Reads corresponding to HIV in HIV-positive cohort ranged from 1 – 27 per participant, with individuals on average having 5.9 reads. The low detection of HIV could be the result of HIV being a RNA virus. In our study RNA viruses were not abundant; RNA viruses made up 0.13% of the viral species detected in this study whereas DNA viruses made up 96.5% of viral species. As indicated in our limitations, compared to the generation of DNA libraries, the efficiency of generating cDNA libraries was much less thus leading to underrepresentation of the RNA fraction in our population. Furthermore, if ribosomes passed through the centrifugation and filtration steps in our protocol,

the abundance of ribosomal RNA (rRNA) could have diluted the presence of viral RNA. Also, it is well known that RNA is not as stable as DNA, and thus can be subjected to degradation over the experimental process. Although, it is more likely that low detection of reads classified as HIV is a result of the majority of our population have suppressed HIV viral loads (16/25). Overall, HIV should not be detected in our healthy-asymptomatic cohort, and it is possible that its detection here may be a result of misclassification or cross-contamination. As mentioned in our limitations, viral identification was not validated to ensure that viral classifications were accurate; secondary analyses are required. We do see that relative to other cohorts, as expected HIV was more abundant in our HIV-positive cohort.

Relationship to Demographic, Clinical and Behavioural Variables

A statistically significant association between our viral groups and Shannon's Diversity Index was detected (Table 3.64). Our viral groups were significantly different from each other based on mean viral diversity. Viral group II; dominated by *Alphapapillomavirus 3* and *Cocolithovirus*; was the most diverse, and Group IV; dominated by *Alphapapillomavirus 3*; was the least diverse. No other statistically significant associations were found, but based on effect size, trends in our data suggested associations between our viral groups, and age, BMI, lifetime history of Gonorrhea, HPV-16, HPV-35, HPV-52, HPV-53, HPV-56, HPV-70, HPV-71, HPV-83, number of HPV types, CD4+ cell count, CD4+ nadir, highest VL ever, VL, duration of antiretroviral medication, duration of HIV infection, HBV surface antibody test results and HCV antibody test results (Table 3.64, 3.65).

Viral diversity within the vaginal environment has not been described; however literature is present on bacterial diversity present within the vaginal microbiota. As previously illustrated, in comparison to healthy asymptomatic women, past research has indicated that women with BV have more diverse bacterial profiles, where bacterial species are more prevalent and

abundant [9,16,19,41,113,168]. We would anticipate then that women with recurrent BV had more viral profiles which fell into Group II, and would be more diverse than women in the healthy-asymptomatic cohort who would have profiles which fell into Group IV, and would be less diverse. However, viral diversity was not significantly different between cohorts, and we did not see a relationship between viral and bacterial diversity. This does coincide with past research. Spear et al. found that bacterial diversity was not different between HIV-positive and HIV-negative populations, which suggests viral diversity would not differ between the two populations [241]. Further research is needed to define the consequences of viral diversity on the vaginal microbiome.

HPV-35, HPV-53 and HPV-83 were detected in one individual each, and those women fell into viral groups II, VI and IV, respectively. HPV-52 and HPV-71 were detected in one individual, who fell into Group I. A higher percentage of women in Group I and VI were seen to be positive for HPV-16, compared to other viral groups. Also, a higher percentage of women in Group I were seen to be positive for HPV-56, and a higher percentage of women in Group I and III were seen to be positive for HPV-70, compared to other viral groups. Moreover, women in Group I had a greater number of HPV types on average (4.3 types), whereas women in Group V had zero HPV types. Group I, II, III and IV were either dominated by *Alphapapillomavirus* or were mixed dominance groups containing *Alphapapillomavirus*, making their association with HPV types more likely, and anticipated. Group VI was dominated by HHV4, and only seen in our HIV-positive population. HPV-16 and HPV-53 are carcinogenic, and possibly carcinogenic [165,190]. Past research has shown a number of high-risk HPV types detected in the vaginal microbiome of HIV-positive women, and states that HPV is seen to be higher in HIV-positive individuals [156]. And to reiterate again, HIV individuals are more susceptible to infections due to immune suppression. Thus, the trend detected here between viral group VI, and HPV-16 and HPV-53 are supported by past work. Furthermore, in addition to Group I, IV and VII, a greater

percentage of women in Group VI had a lifetime history of Gonorrhea compared to other viral groups. Additionally, Group VI had a greater percentage of women with positive HBV sAB and HCV antibody test results. Groups II and VII, and I also had a greater percentage of women with positive HBV sAB and HCV antibody test results, respectively. These trends too, between HBV and HCV, and viral group VI, are supported by literature. Not all trends were supported by literature as viral communities have not been described in relation to many of the trends we saw here, and more research is needed to define these relationships.

On average, women who fell into viral group VII were older (mean age: 36.1 years) whereas those who fell into Group I were younger (mean age: 28.5 years). Women in Group IV had a higher BMI on average (mean BMI: 31 kg/m^2), whereas those in Group II had the lowest BMI (mean: 21.6 kg/m^2). With regards to HIV characteristics, women in Group I had the lowest mean CD4+ cell count (290 mm^3) and mean CD4+ nadir (90 mm^3), whereas women in Group VII had the highest mean CD4+ cell count (870 mm^3) and mean CD4+ nadir (320 mm^3). Women in Group VII also had the shortest mean duration of antiretroviral use (858 days) and shortest mean duration of HIV infection (121 months), whereas women in Group II had the longest mean duration of antiretroviral use (2945 days) and longest mean duration of HIV infection (211.7 months). In addition, on average women in Group VII had the lowest 'highest viral load ever' (63,600 copies/mL), and women in Group II had the highest 'highest viral load ever' (256,616 copies/mL). According to the mean viral load at study visit calculated for each viral group, women in Group II, V, VI and VII were suppressed, women in Group I had an low level viral load (78 copies/mL), and women in Group III and IV were unsuppressed. Women in Group III had the highest mean viral load (9781 copies/mL).

This coincides with the pattern we described earlier between greater *Papillomaviridae* abundance, and low level and unsuppressed viral loads. Additionally, past research has

suggested a relationship between CD4+ cell count, HIV viral load and HPV types. Kang et al. detected a relationship between a set of high-risk HPV types and viral loads >400 copies/mL [242]. However, this trend was not seen when a different set of high-risk HPV types were analyzed; illustrating the differential characteristics of HPV types. In addition, Kang et al. reported an association between CD4+ cell counts and HPV clearance [242]. CD4+ cell counts >350 cells/µl were associated with a higher probability of HPV clearance [242]. Other studies have seen the same relationship between high HPV prevalence and low CD4+ cell counts [243–247]. More recently, Menon et al. saw a strong association between HPV-53, multiple high risk infections, and CD4+ cell count <200 cells/µl [248]. Low CD4+ cell counts of <200 cell/µl were seen to be a significant predictor of HPV-16 and HPV-53. The mean high-risk HPV infections was significantly higher in women with a CD4+ cell count of <200 cells/µl compared to women with a CD4+ cell count >200 cells/µl. Also, the prevalence of HPV-16 in women with CD4+ cell count of <200 cells/µl was 26.9% compared to 37.5 % in women with CD4+ cell count >200/µl [248]. Furthermore, studies have seen a relationship between viral load and HPV. Fife et al. saw that with long lasting HIV suppression <50 copies/mL for two years, the risk of HPV detection went from 62% to 39% [245]. Konopnicki et al. supported these results [247]. Overall, this coincides with our own findings where *Papillomaviridae* dominated viral groups I, II and III had low CD4+ cell counts and high viral loads on average. In our HIV-positive population, half our cohort fell into viral group III and the remaining viral profiles were split between the other viral groups, with one woman in Group VII. Trends should be taken cautiously with low numbers within each of these viral groups.

Addressing Hypotheses

We hypothesized that each CST will exhibit distinct and separate viral patterns, similar to the diversity defined in the CSTs of the bacteriome. This was based on the bacterial diversity seen in the vaginal microbiome, as well as on Brotman et al. who identified a differential

distribution of HPV-positive samples across CSTs [107]. CST I and CST II; dominated by *L. crispatus* and *L. gasseri*, respectively, had lowered proportions of HPV-positive samples, compared to CST III and IV; dominated by *L. iners* and low-*Lactobacillus* groups, respectively. In addition, women who clustered into CST III and CST IVB were seen to have a higher proportion of high risk HPV-positive samples [107]. Literature [9,16,19,41,42,113,168] has illustrated the diversity in microorganism populations and we anticipated the same diversity in our viral populations. Our findings support this hypothesis, as we saw seven distinct viral groups in our population. Group I and IV were strictly dominated by the viral family *Papillomaviridae*, Group II and III were mixed dominance clusters consisting of *Papillomaviridae* and phage species, and Group V, VI and VII were dominated by *Lactobacillus* phage species, human herpesvirus 4, and *Gammaretrovirus* and *Podoviridae*, respectively.

Additionally, we hypothesized, that the diversity in viral patterns could be correlated with demographic, behavioral and clinical data. Additionally, as seen with the vaginal bacteriome, we mentioned that certain bacterial patterns could be correlated with our metadata, and we anticipated the vaginal virome will also exhibit correlations between viral diversity and metadata. Our findings support this hypothesis, as we saw several trends suggesting an association between viral groups, and age, BMI, lifetime history of Gonorrhea, HPV-16, HPV-35, HPV-52, HPV-53, HPV-56, HPV-70, HPV-71, HPV-83, number of HPV types, CD4 count, CD4 nadir, highest VL ever, VL, duration of antiretroviral medication, HBV surface antibody test results and HCV antibody test results.

Limitations

There were several limitations in our study. This was a preliminary look at the vaginal virome of three different cohorts of women, and a small subset of available viral samples was chosen for analyses. As a result of our small sample size, we did not have the power to detect

differences between viral groups. Trends seen here may be an artifact of a small sample size. A larger sample size would be needed to analyze viral communities with the vaginal environment, and its associations with descriptive characteristics. Secondly, we had a cross-sectional design, which looked at women at a single time-point. Wylie et al., Brotman et al., and Ma et al. sampled subjects over several time points to analyze the stability of the papillomaviruses overtime [107,140,162]. As indicated earlier, sampling at a single time point does not capture the dynamics of vaginal microbiome, especially with regards to women with recurrent BV. As we saw in our study, women may have presented in a transition phase with a 'normal' profile in the bacteriome but the virome revealed the instability in the microbiome. It is important to define the viral communities within the vaginal microbiome as well as microbiome stability. Third, we were limited in our ability to detect RNA viruses. Of the viral and phage species we detected in all three populations, 0.13% were RNA viruses compared to the 96.5% which were DNA viruses. Compared to the generation of DNA libraries, the efficiency of generating cDNA libraries was much less thus leading to underrepresentation of the RNA fraction in our population. Furthermore, if ribosomes passed through the centrifugation and filtration steps in our protocol, the abundance of rRNA could have diluted the presence of viral RNA. Also, it is well known that RNA is not as stable as DNA, and thus can be subjected to degradation over the experimental process; however this was less likely in our experiments as RNA was reverse transcribed into cDNA. Fourth, our study had lack of quantification and was more qualitative in nature, as several aspects in our study normalized our data. For one, vaginal swab material was not identical for each participant, resulting in greater concentrations of sample for some participants over others. Also, during PCR the amount of amplified product does not necessarily reflect the amount of template which was initially present in the reaction. As the number of PCR cycles increases, amplification efficiency decreases eventually resulting in the 'plateau effect'. This can be caused by a number of things; shortage of primer or nucleotide substrate, deactivation of *Taq* DNA polymerase, re-annealing of amplified DNA. The plateau effect results in similar

amounts of amplified product being generated after a number of PCR cycles, regardless of the initial template amount. Thus, sequences present at a lower abundance and higher abundances in our study could be reflected as being present within the vaginal virome at the same level after a number of cycles of PCR. Moreover, by removing bacterial, human and large cellular components, as well as removing reads detected in our negative controls from samples, species detected in the vaginal environment where the absolute amount could be low, can now be present at a greater abundance. Fifth, viral identification was not validated to ensure that viral classifications were accurate. It is possible that when reads were classified, they were 'hitting' the same gene or region of the genome and thus may not accurate. This was highlighted by our detection of HIV in our healthy-asymptomatic population. Further validation via bioinformatics algorithm or viral PCR is required to ensure classifications are accurate. Lastly, thresholds were not used when classifying reads to viral species. Taxonomer uses kmer weight to determine which viral species likely corresponds to each read. Thresholds are needed to decide whether this is an accurate classification or whether there is not enough weight to confidently classify a read.

Future Directions

In the future we would want to increase our sample size, and look at a larger population of healthy-asymptomatic women, HIV-positive women, and women with recurrent BV, as well as sequence deeper. This would give us the power to detect differences between our viral groups. With the sample size and sequence coverage we had, we saw that the vaginal virome was quite diverse. A larger sample size and increased coverage would allow us to better define this diversity, and viral clusters. This would also allow us to re-analyze the trends we saw in our current study, and determine whether the associations observed were artifacts of a small sample size or in fact statistically significant. With further research, we would be able to better define the associations we saw. Furthermore, we would be interested in looking at the vaginal

virome over time to assess the stability of viral species detected in this study. In our current study, we determined if any correlations existed between our viral species and bacteriome CSTs, and found no significant associations or trends to suggest a relationship. However, this is not to say that a relationship between viral species and lower abundance bacterial species does not exist. Bacteriome CSTs consist of bacterial species, which are dominant in the bacterial profile of an individual, and do not speak to the lower abundance bacterial species which are present. In the future, we would be interested in determining whether a relationship between our viral species and these lower abundant bacterial species exists. Moreover, due to the qualitative nature of our study we want to step away from relative abundances, and quantitative analyses, and cluster our viral species based on 'presence and absence' data. These resulting clusters will be related to our descriptive characteristics. Additionally, in the future, classification thresholds need to be established such that we are confident in the classification of each read. In order to be confident our classifications, proper validation of viral identification is also required. As indicated earlier, alternative bioinformatics algorithm, use of multiple classification software, or confirming the identification of viruses through PCR could provide a secondary check-point to valid the viral species each read has been assigned to.

Conclusion

Overall, we sought out to characterize the vaginal virome to not only define the viral species present in this environment, but to shed some light on this under research area. Our population was made up of 21 healthy-asymptomatic women, 25 HIV-positive women and eight women with recurrent BV. Our three cohorts were not significantly different across the majority of demographic, behavioural and clinical variables, other than age, ethnicity and pregnancy history. In addition to species that were detected in all participants from all three cohorts, in each cohort unique species were also detected in all participants. This may represent the "normal" vaginal virome for that specific clinical cohort. We had an interest in the prevalence of

specific viruses, HHV, HPV and HIV in our populations. *Alphapapillomavirus* were the most abundant across all cohorts and this was followed by *Lymphocryptovirus* (predominantly human herpesvirus 4). The majority of prevalence in our populations aligned with the current literature. Viral profiles clustered in seven distinct viral groups.

Within our healthy-asymptomatic cohort, the majority of women with *L. crispatus* dominated profiles had viral profiles dominated by *Papillomaviridae* species. This was specific to CST I, and not seen in all *Lactobacillus*-dominated profiles. However, the opposite trend was seen in the recurrent BV population. *Lactobacillus*-dominated profiles (CST I) were dominated by phage species, in comparison to *L. iners* dominated profiles (CST III), and heterogeneous profiles (CST IVA, IVC, IVD) which had a greater abundance of *Papillomaviridae*. Different patterns were seen in healthy-asymptomatic and recurrent BV cohorts, illustrating differences in viral communities between health and disease states. A relationship was also detected between *Lactobacillus* phage dominated viral group V and high Nugent scores, once again furthering our understanding between the vaginal health and disease. Within our HIV-positive population, there appeared to be a relationship between *Papillomaviridae* abundance and bacterial diversity, and *Siphoviridae* abundance and HIV plasma viral loads. These relationships were only detected in the low level viral load group, and not within the suppressed and unsuppressed viral load groups.

Viral groups were significantly associated with Shannon's Diversity Index, and trends suggested a relationship between our viral groups, and age, BMI, lifetime history of gonorrhea, HPV-16, HPV-35, HPV-52, HPV-53, HPV-56, HPV-70, HPV-71, HPV-83, number of HPV types, CD4+ cell count, CD4+ nadir, highest VL ever, VL, duration of antiretroviral medication, duration of HIV infection, HBV surface antibody test results and HCV antibody test results. We also saw an association between *Enterobacteria phage N15*, and our HIV-positive cohort. These findings

are suggestive of an interaction between certain viral patterns and descriptive variables, however further research is needed to explore these novel relationships and their importance to the clinical cohorts studied here.

The majority of research on the vaginal microbiome characterizes bacterial species which occupy this environment. To our knowledge, this is the first study of its kind to conduct a comprehensive characterization of the vaginal virome of specific cohorts of women. Thus most of our reported findings are novel and provide the first stepping stone to describing the vaginal virome.

4.3 Concluding Statement

To conclude, not only did our findings illustrate the advantages of *cpn*60 sequencing, but indicated the importance of viral characterization in microbiome research. Recruitment of more women and sequencing more samples will aid in further defining the bacterial and viral communities which exist within the vaginal microbiome. Future research is required to further describe these relationships detected here, as well as their consequences for vaginal health and disease. Our research project better defines the vaginal microbiome as a whole, and has important consequences for the improvement of women's health.

Bibliography

1. Walker Jr B, Kassim K, Stokes LD. The Microbiome: A Contributor to Health and Disease. *J Health Care Poor Underserved*. 2015;26(1):62-72.
2. Prince AL, Antony KM, Ma J, Aagaard KM. The Microbiome and Development : A Mother's Perspective. *Semin Reprod Med*. 2014;77030(212):14-22.
3. Bray J. Canadian Microbiome Initiative. 2008.
4. Sender R, Fuchs S, Milo R. Commentary Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. 2016.
5. Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. *Nat Rev Genet*. 2012;13(4):260-270.
6. Arrieta M-C, Stiemsma LT, Dimitriu PA, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci Transl Med*. 2015;7(307).
7. Huttenhower C, Gevers D, Knight R, et al. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486(7402):207-214.
8. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The Human Microbiome Project. *Nature*. 2007;449(October):804-810.
9. Lamont RF, Sobel JD, Akins R a., et al. The vaginal microbiome: New information about genital tract flora using molecular based techniques. *BJOG An Int J Obstet Gynaecol*. 2011;118:533-549.
10. MacPhee RA, Hummelen R, Bisanz JE, Miller WL, Reid G. Probiotic strategies for the treatment and prevention of bacterial vaginosis. *Expert Opin Pharmacother*. 2010;11(18):2985-95.
11. Huang B, Fettweis JM, Brooks JP, Jefferson KK, Buck G a. The Changing Landscape of the Vaginal Microbiome. *Clin Lab Med*. 2014;34(4):747-761.
12. Sweet RL. Gynecologic conditions and bacterial vaginosis: implications for the non-pregnant patient. *Infect Dis Obstet Gynecol*. 2000;8(3-4):184-90.
13. Zhou X, Brown CJ, Abdo Z, et al. Differences in the composition of vaginal microbial communities found in healthy Caucasian and black women. *ISME J*. 2007;1:121-133.
14. Public Health Agency of Canada. Section 2 - Primary Care and Sexually Transmitted Infections - Canadian Guidelines on Sexually Transmitted Infections. May 2013.
15. Workowski K, Berman S, Prevention C for DC and. Sexually transmitted diseases treatment guidelines, 2010. *Morb Mortal Wkly Rep*. 2010;59(RR-12):1-110.
16. Ling Z, Kong J, Liu F, et al. Molecular analysis of the diversity of vaginal microbiota associated with bacterial vaginosis. *BMC Genomics*. 2010;11:488.
17. Gillet E, Meys JF, Verstraelen H, et al. Bacterial vaginosis is associated with uterine cervical human papillomavirus infection: a meta-analysis. *BMC Infect Dis*. 2011;11(1):10.
18. Mitchell C, Manhart LE, Thomas KK, Agnew K, Marrazzo JM. Effect of Sexual Activity on Vaginal Colonization With Hydrogen Peroxide-Producing Lactobacilli and Gardnerella vaginalis. *Sex Transm Dis*. 2011;38(12):1137-1144.
19. van de Wijgert JHHM, Borgdorff H, Verhelst R, et al. The Vaginal Microbiota: What Have We Learned after a Decade of Molecular Characterization? *PLoS One*. 2014;9(8):e105998.
20. van Schalkwyk J, Yudin MH. Vulvovaginitis : Screening for and Management of Trichomoniasis, Vulvovaginal Candidiasis, and Bacterial Vaginosis. *J Obs Gynaecol Can*. 2015;37(3):266-274.
21. Dominguez-Bello MG, Costello EK, Contreras M, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A*. 2010;107(26):11971-5.
22. Aroutcheva A, Gariti D, Simon M, et al. Defense factors of vaginal lactobacilli. *Am J*

Obstet Gynecol. 2001;185(2):375-379.
23. Atassi F, Servin AL. Individual and co-operative roles of lactic acid and hydrogen peroxide in the killing activity of enteric strain Lactobacillus johnsonii NCC933 and vaginal strain Lactobacillus gasseri KS120.1 against enteric, uropathogenic and vaginosis-associated pathog. *FEMS Microbiol Lett*. 2010;304(1):29-38.
24. Lambert J a, John S, Sobel JD, Akins R a. Longitudinal analysis of vaginal microbiome dynamics in women with recurrent bacterial vaginosis: recognition of the conversion process. *PLoS One*. 2013;8(12):e82599.
25. Martín R, Suárez JE. Biosynthesis and degradation of H2O2 by vaginal lactobacilli. *Appl Environ Microbiol*. 2010;76(2):400-405.
26. Hawes SE, Hillier SL, Benedetti J, et al. Hydrogen peroxide-producing lactobacilli and acquisition of vaginal infections. *J Infect Dis*. 1996;174(5):1058-1063.
27. Antonio MA, Hawes SE, Hillier SL. The identification of vaginal Lactobacillus species and the demographic and microbiologic characteristics of women colonized by these species. *J Infect Dis*. 1999;180(6):1950-6.
28. O'Hanlon DE, Moench TR, Cone R a. In vaginal fluid, bacteria associated with bacterial vaginosis can be suppressed with lactic acid but not hydrogen peroxide. *BMC Infect Dis*. 2011;11(1):200.
29. O'Hanlon DE, Moench TR, Cone R a. Vaginal pH and microbicidal lactic acid when lactobacilli dominate the microbiota. *PLoS One*. 2013;8(11):1-9.
30. Mardh PA. The vaginal ecosystem. *Am J Obs Gynecol*. 1991;165(4 Pt 2):1163-1168.
31. Hillier SL. Diagnostic microbiology of bacterial vaginosis. *Am J Obstet Gynecol*. 1993;169(2):455-459.
32. Schleifer KH, Ludwig W. Phylogeny of the Genus Lactobacillus and Related Genera. *Syst Appl Microbiol*. 1995;18(4):461-467.
33. Du Plessis EM, Dicks LMT. Evaluation of random amplified polymorphic DNA (RAPD)-PCR as a method to differentiate Lactobacillus acidophilus, Lactobacillus crispatus, Lactobacillus amylovorus, Lactobacillus gallinarum, Lactobacillus gasseri, and Lactobacillus johnsonii. *Curr Microbiol*. 1995;31(2):114-118.
34. Falsen E, Pascual C, Sjödén B, Ohlén M, Collins MD. Phenotypic and phylogenetic characterization of a novel Lactobacillus species from human sources: description of Lactobacillus iners sp. nov. *Int J Syst Bacteriol*. 1999;49 Pt 1:217-21.
35. Rodriguez Jovita M, Collins MD, Sjödén B, Falsen E. Characterization of a novel Atopobium isolate from the human vagina: description of Atopobium vaginae sp. nov. *Int J Syst Bacteriol*. 1999;49 Pt 4:1573-6.
36. Dimitonova SP, Bakalov BV, Aleksandrova-Georgieva RN, Danova ST. Phenotypic and molecular identification of lactobacilli isolated from vaginal secretions. *J Microbiol Immunol Infect*. 2008;41(6):469-477.
37. Hill JE, Goh SH, Money DM, et al. Characterization of vaginal microflora of healthy, nonpregnant women by chaperonin-60 sequence-based methods. *Am J Obstet Gynecol*. 2005;193(3):682-692.
38. Zhou X, Bent SJ, Schneider MG, Davis CC, Islam MR, Forney LJ. Characterization of vaginal microbial communities in adult healthy women using cultivation-independent methods. *Microbiology*. 2004;150:2565-2573.
39. Ravel J, Gajer P, Abdo Z, et al. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A*. 2011;108 Suppl:4680-4687.
40. Gajer P, Brotman RM, Bai G, et al. Temporal dynamics of the human vaginal microbiota. *Sci Transl Med*. 2012;4(132):132ra52.
41. Srinivasan S, Hoffman NG, Morgan MT, et al. Bacterial communities in women with bacterial vaginosis: High resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. *PLoS One*. 2012;7(6):e37818.

42. Vitali B, Cruciani F, Picone G, Parolin C, Donders G, Laghi L. Vaginal microbiome and metabolome highlight specific signatures of bacterial vaginosis. *Eur J Clin Microbiol Infect Dis*. 2015;34(12):2367-2376.
43. Burton JP, Cadieux PA, Reid G. Improved understanding of the bacterial vaginal microbiota of women before and after probiotic instillation. *Appl Environ Microbiol*. 2003;69(1):97-101.
44. Hyman RW, Fukushima M, Diamond L, Kumm J, Giudice LC, Davis RW. Microbes on the human vaginal epithelium. *Proc Natl Acad Sci U S A*. 2005;102:7952-7957.
45. Chaban B, Links MG, Jayaprakash TP, et al. Characterization of the vaginal microbiota of healthy Canadian women through the menstrual cycle. *Microbiome*. 2014;2(1):23.
46. Zhou X, Hansmann M a., Davis CC, et al. The vaginal bacterial communities of Japanese women resemble those of women in other racial groups. *FEMS Immunol Med Microbiol*. 2010;58:169-181.
47. Albert AYK, Chaban B, Wagner EC, et al. A Study of the Vaginal Microbiome in Healthy Canadian Women Utilizing cpn60-Based Molecular Profiling Reveals Distinct Gardnerella Subgroup Community State Types. *PLoS One*. 2015;10(8):e0135620.
48. McGregor JA, French JI. Bacterial vaginosis in pregnancy. *Obstet Gynecol Surv*. 2000;55(5 Suppl 1):S1-19.
49. Sweet RL, Gibbs RS. *Infectious Diseases of the Female Genital Tract*. 3rd ed.; 1997.
50. Pyles RB, Vincent KL, Baum MM, et al. Cultivated vaginal microbiomes alter HIV-1 infection and antiretroviral efficacy in colonized epithelial multilayer cultures. *PLoS One*. 2014;9(3).
51. Allsworth JE, Peipert JF. Prevalence of bacterial vaginosis. *Obs Gynecol*. 2007;109(1):114-120.
52. Koumans EH, Sternberg M, Bruce C, et al. The Prevalence of Bacterial Vaginosis in the United States, 2001-2004; Associations With Symptoms, Sexual Behaviors, and Reproductive Health. *Sex Transm Dis*. 2007;34(11):864-869.
53. Marrazzo JM, Thomas KK, Agnew K, Ringwood K. Prevalence and Risks for Bacterial Vaginosis in Women Who Have Sex With Women. *Sex Transm Dis*. March 2010:1.
54. Kenyon C, Colebunders R, Crucitti T. The global epidemiology of bacterial vaginosis: A systematic review. *Am J Obstet Gynecol*. 2013;209(6):505-523.
55. Public Health Agency of Canada. Vaginal Discharge (Bacterial Vaginosis, Vulvovaginal Candidiasis, Trichomoniasis) - Section 4 Management and Treatment of Specific Syndromes - Canadian Guidelines on Sexually Transmitted Infections. 2013.
56. Verstraelen H, Verhelst R, Vaneechoutte M, Temmerman M. The epidemiology of bacterial vaginosis in relation to sexual behaviour. *BMC Infect Dis*. 2010;10(1):81.
57. Amsel R, Totten PA, Spiegel CA, Chen KC, Eschenbach D, Holmes KK. Nonspecific vaginitis. Diagnostic criteria and microbial and epidemiologic associations. *Am J Med*. 1983;74(1):14-22.
58. Nugent RP, Krohn M a., Hillier SL. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. *J Clin Microbiol*. 1991;29(2):297-301.
59. Eschenbach DA. Bacterial vaginosis and anaerobes in obstetric-gynecologic infection. *Clin Infect Dis*. 1993;16:S282-S287.
60. Sobel JD. Vaginitis. *N Engl J Med*. 1997;337:1896-1903.
61. Moodley P, Connolly C, Sturm a W. Interrelationships among human immunodeficiency virus type 1 infection, bacterial vaginosis, trichomoniasis, and the presence of yeasts. *J Infect Dis*. 2002;185:69-73.
62. Fredricks DN, Fiedler TL, Thomas KK, Oakley BB, Marrazzo JM. Targeted PCR for Detection of Vaginal Bacteria Associated with Bacterial Vaginosis. *J Clin Microbiol*. 2007;45(10):3270-3276.

63. Paramel Jayaprakash T, Schellenberg JJ, Hill JE. Resolution and characterization of distinct cpn60-based subgroups of gardnerella vaginalis in the vaginal microbiota. *PLoS One*. 2012;7(8):1-11.
64. Bradshaw C, Morton AN, Hocking J, et al. High recurrence rates of bacterial vaginosis over the course of 12 months after oral metronidazole therapy and factors associated with recurrence. *J Infect Dis*. 2006;193(11):1478-1486.
65. Sobel R, Sobel JD. Metronidazole for the treatment of vaginal infections. *Expert Opin Pharmacother*. 2015;16(7):1109-1115.
66. Swidsinski A, Mendling W, Loening-Baucke V, et al. Adherent biofilms in bacterial vaginosis. *Obstet Gynecol*. 2005;106(5):1013-1023.
67. Patterson JL, Stull-Lane A, Girerd PH, Jefferson KK. Analysis of adherence, biofilm formation and cytotoxicity suggests a greater virulence potential of Gardnerella vaginalis relative to other bacterial-vaginosis-associated anaerobes. *Microbiology*. 2010;156(Pt 2):392-9.
68. Patterson JL, Girerd PH, Karjane NW, Jefferson KK. Effect of biofilm phenotype on resistance of Gardnerella vaginalis to hydrogen peroxide and lactic acid. *Am J Obstet Gynecol*. 2007;197(2):170.e1-7.
69. Machado A, Jefferson KK, Cerca N. Interactions between Lactobacillus crispatus and bacterial vaginosis (BV)-associated bacterial species in initial attachment and biofilm formation. *Int J Mol Sci*. 2013;14(6):12004-12.
70. Hardy L, Jespers V, Dahchour N, et al. Unravelling the Bacterial Vaginosis-Associated Biofilm: A Multiplex Gardnerella vaginalis and Atopobium vaginae Fluorescence In Situ Hybridization Assay Using Peptide Nucleic Acid Probes. Al-Ahmad A, ed. *PLoS One*. 2015;10(8):e0136658.
71. Marrazzo JM. Relationship of Specific Vaginal Bacteria and Bacterial Vaginosis Treatment Failure in Women Who Have Sex with Women. *Ann Intern Med*. 2008;149(1):20.
72. Schellenberg JJ, Paramel Jayaprakash T, Withana Gamage N, et al. Gardnerella vaginalis Subgroups Defined by cpn60 Sequencing and Sialidase Activity in Isolates from Canada, Belgium and Kenya. Landay A, ed. *PLoS One*. 2016;11(1):e0146510.
73. Ness RB, Hillier S, Richter HE, et al. Can known risk factors explain racial differences in the occurrence of bacterial vaginosis? *J Natl Med Assoc*. 2003;95(3):201-212.
74. Peipert JF, Lapane KL, Allsworth JE, Redding C a, Blume JD, Stein MD. Bacterial vaginosis, race, and sexually transmitted infections: does race modify the association? *Sex Transm Dis*. 2008;35(4):363-367.
75. Rajamanoharan S, Low N, Jones SB, Pozniak a L. Bacterial vaginosis, ethnicity, and the use of genital cleaning agents: a case control study. *Sex Transm Dis*. 1999;26:404-409.
76. Royce RA, Jackson TP, Thorp JM, et al. Race/ethnicity, vaginal flora patterns, and pH during pregnancy. *Sex Transm Dis*. 1999;26(2):96-102.
77. Goldenberg RL, Klebanoff MA, Nugent R, Krohn MA, Hillier S, Andrews WW. Bacterial colonization of the vagina during pregnancy in four ethnic groups. *Am J Obstet Gynecol*. 1996;174(5):1618-1621.
78. Klatt TE, Cole DC, Eastwood DC, Barnabei VM. Factors Associated with Recurrent Bacterial Vaginosis. *J Reprod Med*. 2010;55(1-2):55-61.
79. Gallo MF, Macaluso M, Warner L, et al. Bacterial vaginosis, gonorrhea, and chlamydial infection among women attending a sexually transmitted disease clinic: a longitudinal analysis of possible causal links. *Ann Epidemiol*. 2012;22(3):213-20.
80. Cherpes TL, Hillier SL, Meyn LA, Busch JL, Krohn MA. A delicate balance: risk factors for acquisition of bacterial vaginosis include sexual activity, absence of hydrogen peroxide-producing lactobacilli, black race, and positive herpes simplex virus type 2 serology. *Sex Transm Dis*. 2008;35(1):78-83.

81. Fethers K, Twin J, Fairley CK, et al. Bacterial Vaginosis (BV) Candidate Bacteria: Associations with BV and Behavioural Practices in Sexually-Experienced and Inexperienced Women. Ratner AJ, ed. *PLoS One*. 2012;7(2):e30633.
82. Smart S, Singal A, Mindel A. Social and sexual risk factors for bacterial vaginosis. *Sex Transm Infect*. 2004;80(1):58-62.
83. Fethers KA, Fairley CK, Hocking JS, Gurrin LC, Bradshaw CS. Sexual risk factors and bacterial vaginosis: a systematic review and meta-analysis. *Clin Infect Dis*. 2008;47(11):1426-35.
84. Bradshaw C, Vodstrcil LA, Hocking JS, et al. Recurrence of bacterial vaginosis is significantly associated with posttreatment sexual activities and hormonal contraceptive use. *Clin Infect Dis*. 2013;56(6):777-786.
85. Bradshaw C, Walker J, Fairley CK, et al. Prevalent and Incident Bacterial Vaginosis Are Associated with Sexual and Contraceptive Behaviours in Young Australian Women. *PLoS One*. 2013;8(3):e57688.
86. Li XD, Wang CC, Zhang XJ, et al. Risk factors for bacterial vaginosis: Results from a cross-sectional study having a sample of 53,652 women. *Eur J Clin Microbiol Infect Dis*. 2014;33:1525-1532.
87. Marconi C, Duarte MTC, Silva DC, Silva MG. Prevalence of and risk factors for bacterial vaginosis among women of reproductive age attending cervical screening in southeastern Brazil. *Int J Gynaecol Obstet*. 2015;131(2):137-41.
88. Marrazzo JM, Fiedler TL, Srinivasan S, et al. Extravaginal reservoirs of vaginal bacteria as risk factors for incident bacterial vaginosis. *J Infect Dis*. 2012;205(10):1580-8.
89. Muzny C a, Sunesara IR, Kumar R, et al. Characterization of the vaginal microbiota among sexual risk behavior groups of women with bacterial vaginosis. *PLoS One*. 2013;8(11):e80254.
90. Bradshaw C, Walker SM, Vodstrcil LA, et al. The influence of behaviors and relationships on the vaginal microbiota of women and their female partners: The WOW health study. *J Infect Dis*. 2014;209(10):1562-1572.
91. Forcey DS, Vodstrcil LA, Hocking JS, et al. Factors Associated with Bacterial Vaginosis among Women Who Have Sex with Women: A Systematic Review. Graham SM, ed. *PLoS One*. 2015;10(12):e0141905.
92. Vodstrcil LA, Walker SM, Hocking JS, et al. Incident bacterial vaginosis (BV) in women who have sex with women is associated with behaviors that suggest sexual transmission of BV. *Clin Infect Dis*. 2015;60(7):1042-53.
93. Brotman RM, Klebanoff MA, Nansel TR, et al. A longitudinal study of vaginal douching and bacterial vaginosis - A marginal structural modeling analysis. *Am J Epidemiol*. 2008;168(2):188-196.
94. Klebanoff MA, Nansel TR, Brotman RM, et al. Personal hygienic behaviors and bacterial vaginosis. *Sex Transm Dis*. 2010;37(2):94-9.
95. Hillier SL, Rabe LK, Muller CH, Zarutskie P, Kuzan FB, Stenchever MA. Relationship of bacteriologic characteristics to semen indices in men attending an infertility clinic. *Obstet Gynecol*. 1990;75(5):800-4.
96. Serour F, Samra Z, Kushel Z, Gorenstein A, Dan M. Comparative periurethral bacteriology of uncircumcised and circumcised males. *Genitourin Med*. 1997;73(4):288-90.
97. Manhart LE, Khosropour CM, Liu C, et al. Bacterial vaginosis-associated bacteria in men: association of Leptotrichia/Sneathia spp. with nongonococcal urethritis. *Sex Transm Dis*. 2013;40(12):944-9.
98. Hilber AM, Francis SC, Chersich M, et al. Intravaginal practices, vaginal infections and HIV acquisition: systematic review and meta-analysis. *PLoS One*. 2010;5(2):e9119.
99. Brown JM, Hess KL, Brown S, Murphy C, Waldman AL, Hezareh M. Intravaginal

practices and risk of bacterial vaginosis and candidiasis infection among a cohort of women in the United States. *Obstet Gynecol*. 2013;121(4):773-80.
100. Masese L, Baeten JM, Richardson B a, et al. Incident Herpes Simplex Virus Type 2 Infection Increases the Risk of Subsequent Episodes of Bacterial Vaginosis. *J Infect Dis*. 2013;209:1-5.
101. Esber A, Moyo P, Munjoma M, et al. Cessation of intravaginal practices to prevent bacterial vaginosis: a pilot intervention in Zimbabwean women. *Sex Transm Infect*. 2015;91(3):183-188.
102. Pavlova SI, Tao L. In Vitro Inhibition of Commercial Douche Products Against Vaginal Microflora. *Infect Dis Obstet Gynecol*. 2000;8:99-104.
103. Ness RB, Hillier SL, Richter HE, et al. Douching in relation to bacterial vaginosis, lactobacilli, and facultative bacteria in the vagina. *Obstet Gynecol*. 2002;100(4):765-772.
104. Schwebke JR, Desmond R a, Oh MK. Predictors of bacterial vaginosis in adolescent women who douche. *Sex Transm Dis*. 2004;31(7):433-436.
105. Alcaide ML, Mumbi M, Chitalu N, Jones DL. An intervention to decrease intravaginal practices in hiv-infected women in Zambia: a pilot study. *J Assoc Nurses AIDS Care*. 2013;24(3):219-26.
106. van De Wijgert JH, Mason PR, Gwanzura L, et al. Intravaginal practices, vaginal flora disturbances, and acquisition of sexually transmitted diseases in Zimbabwean women. *J Infect Dis*. 2000;181(2):587-94.
107. Brotman RM, Shardell MD, Gajer P, et al. Interplay between the temporal dynamics of the vaginal microbiota and human papillomavirus detection. *J Infect Dis*. 2014;210(11):1723-1733.
108. Calzolari E, Masciangelo R, Milite V, Verteramo R. Bacterial Vaginosis and Contraceptive Methods. *Int J Gynaecol Obs*. 2000;70(3):341-346.
109. Shoubnikova M, Hellberg D, Nilsson S, Mårdh P a. Contraceptive use in women with bacterial vaginosis. *Contraception*. 1997;55(6):355-8.
110. Baeten JM, Nyange PM, Richardson BA, et al. Hormonal contraception and risk of sexually transmitted disease acquisition: results from a prospective study. *Am J Obstet Gynecol*. 2001;185(2):380-5.
111. Madden T, Grentzer JM, Secura GM, Allsworth JE, Peipert JF. Risk of bacterial vaginosis in users of the intrauterine device: a longitudinal study. *Sex Transm Dis*. 2012;39(3):217-22.
112. Marrazzo JM, Thomas KK, Fiedler TL, Ringwood K, Fredricks DN. Risks for Acquisition of Bacterial Vaginosis Among Women Who Report Sex with Women: A Cohort Study. Myer L, ed. *PLoS One*. 2010;5(6):e11139.
113. Fredricks DN, Fiedler TL, Marrazzo JM. Molecular identification of bacteria associated with bacterial vaginosis. *N Engl J Med*. 2005;353(18):1899-1911.
114. Fredricks DN. Molecular methods to describe the spectrum and dynamics of the vaginal microbiota. *Anaerobe*. 2011;17(4):191-195.
115. Verhelst R, Verstraelen H, Claeys G, et al. Cloning of 16S rRNA genes amplified from normal and disturbed vaginal microflora suggests a strong association between Atopobium vaginae, Gardnerella vaginalis and bacterial vaginosis. *BMC Microbiol*. 2004;4:16.
116. Verstraelen H, Verhelst R, Claeys G, Temmerman M, Vaneechoutte M. Culture-independent analysis of vaginal microflora: the unrecognized association of Atopobium vaginae with bacterial vaginosis. *Am J Obstet Gynecol*. 2004;191(4):1130-2.
117. Trama JP, Pascal KE, Zimmerman J, Self MJ, Mordechai E, Adelson ME. Rapid detection of Atopobium vaginae and association with organisms implicated in bacterial vaginosis. *Mol Cell Probes*. 2008;22(2):96-102.
118. Bradshaw C, Tabrizi SN, Fairley CK, Morton AN, Rudland E, Garland SM. The

association of Atopobium vaginae and Gardnerella vaginalis with bacterial vaginosis and recurrence after oral metronidazole therapy. *J Infect Dis*. 2006;194(6):828-836.
119. Menard J, Fenollar F, Henry M, Bretelle F, Raoult D. Molecular Quantification of Gardnerella vaginalis and Atopobium vaginae Loads to Predict Bacterial Vaginosis. *Clin Infect Dis*. 2008;47(1):33-43.
120. Verstraelen H, Swidsinski A. The biofilm in bacterial vaginosis: implications for epidemiology, diagnosis and treatment. *Curr Opin Infect Dis*. 2013;26(1):86-9.
121. Jespers V, van de Wijgert J, Cools P, et al. The significance of Lactobacillus crispatus and L. vaginalis for vaginal health and the negative effect of recent sex: a cross-sectional descriptive study across groups of African women. *BMC Infect Dis*. 2015;15(1):115.
122. Hill JE, Seipp RP, Betts M, et al. Extensive Profiling of a Complex Microbial Community by High-Throughput Sequencing Extensive Profiling of a Complex Microbial Community by High-Throughput Sequencing. *Society*. 2002;68(6):3055-3066.
123. Links MG, Dumonceaux TJ, Hemmingsen SM, Hill JE. The Chaperonin-60 Universal Target Is a Barcode for Bacteria That Enables De Novo Assembly of Metagenomic Sequence Data. *PLoS One*. 2012;7(11).
124. Hill JE, Penny SL, Crowell KG, Goh SH, Hemmingsen SM. cpnDB: A chaperonin sequence database. *Genome Res*. 2004;14:1669-1675.
125. Huse SM, Ye Y, Zhou Y, Fodor A a. A core human microbiome as viewed through 16S rRNA sequence clusters. *PLoS One*. 2012;7(6):1-12.
126. Staley JT, Konopka A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol*. 1985;39:321-46.
127. Woo PCY, Lau SKP, Teng JLL, Tse H, Yuen KY. Then and now: Use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clin Microbiol Infect*. 2008;14(10):908-934.
128. Nossa CW, Oberdorf WE, Yang L, et al. Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. *World J Gastroenterol*. 2010;16(33):4135-4144.
129. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing Artifacts on 16s rRNA-based studies. *PLoS One*. 2011;6(12).
130. Links MG, Demeke T, Gräfenhan T, Hill JE, Hemmingsen SM, Dumonceaux TJ. Simultaneous profiling of seed-associated bacteria and fungi reveals antagonistic interactions between microorganisms within a shared epiphytic microbiome on Triticum and Brassica seeds. *New Phytol*. 2014;202:542-553.
131. Hamady M, Knight R. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Res*. 2009;19(7):1141-1152.
132. Conlan S, Kong HH, Segre JA. Species-Level Analysis of DNA Sequence Data from the NIH Human Microbiome Project. *PLoS One*. 2012;7(10).
133. Hummelen R, Fernandes AD, Macklaim JM, et al. Deep Sequencing of the Vaginal Microbiota of Women with HIV. Bereswill S, ed. *PLoS One*. 2010;5(8):e12078.
134. Vermette CJ, Russell AH, Desai AR, Hill JE. Resolution of phenotypically distinct strains of Enterococcus spp. in a complex microbial community using cpn60 universal target sequencing. *Microb Ecol*. 2010;59:14-24.
135. Wang GCY, Wang Y. Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. *Appl Environ Microbiol*. 1997;63(12):4645-4650.
136. Schellenberg J, Links MG, Hill JE, Hemmingsen SM, Peters GA, Dumonceaux TJ. High-Throughput Next Generation Sequencing. *High-Throughput Next Gener Seq Methods Appl Methods Mol Biol*. 2011;733.
137. Schellenberg J, Links MG, Hill JE, et al. Pyrosequencing of the chaperonin-60 universal target as a tool for determining microbial community composition. *Appl Environ Microbiol*.

2009;75:2889-2898.
138. Norman JM, Handley S a, Parkes M, et al. Disease-Specific Alterations in the Enteric Virome in Article Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. *Cell*. 2015;160(3):447-460.
139. Abeles SR, Pride DT. Molecular bases and role of viruses in the human microbiome. *J Mol Biol*. 2014;426(23):3892-3906.
140. Wylie KM, Mihindukulasuriya K a, Zhou Y, Sodergren E, Storch G a, Weinstock GM. Metagenomic analysis of double-stranded DNA viruses in healthy adults. *BMC Biol*. 2014;12:71.
141. Handley S a., Thackray LB, Zhao G, et al. Pathogenic simian immunodeficiency virus infection is associated with expansion of the enteric virome. *Cell*. 2012;151(2):253-266.
142. Virgin HW. The virome in mammalian physiology and disease. *Cell*. 2014;157(1):142-150.
143. Thurber R V, Haynes M, Breitbart M, Wegley L, Rohwer F. Laboratory procedures to generate viral metagenomes. *Nat Protoc*. 2009;4(4):470-483.
144. Money DM, Arikan YY, Remple V, et al. Genital tract and plasma human immunodeficiency virus viral load throughout the menstrual cycle in women who are infected with ovulatory human immunodeficiency virus. *Am J Obstet Gynecol*. 2003;188:122-128.
145. Ross S a, Novak Z, Ashrith G, et al. Association between genital tract cytomegalovirus infection and bacterial vaginosis. *J Infect Dis*. 2005;192:1727-1730.
146. Sekhon HS, Press RD, Schmidt W a., Hawley M, Rader A. Identification of cytomegalovirus in a liquid-based gynecologic sample using morphology, immunohistochemistry, and DNA real-time PCR detection. *Diagn Cytopathol*. 2004;30(6):411-417.
147. Public Health Agency of Canada. Cytomegalovirus. 2011.
148. Chandler SH, Handsfield HH, McDougall JK. Isolation of multiple strains of cytomegalovirus from women attending a clinic for sexually transmitted diseases. *J Infect Dis*. 1987;155(4):655-660.
149. Staras SA, Dollard SC, Radford KW, Flanders WD, Pass RF, Cannon MJ. Seroprevalence of Cytomegalovirus Infection in the United States, 1988-1994. *Clin Infect Dis*. 2006;43(9):1143-1151.
150. Freeman Jr RB. The “Indirect” Effects of Cytomegalovirus Infection. *Am J Transplant*. 2009;9(11):2453-2458.
151. Virgin HW, Wherry EJ, Ahmed R. Redefining chronic viral infection. *Cell*. 2009;138(1):30-50.
152. Mougin C, Bernard C, Schaal JP, Laurent R, Carbillet JP, Lab M. Detection of human cytomegalovirus in female and male anogenital lesions with papillomavirus-associated dysplastic changes. *Cancer Detect Prev*. 1995;19(4):316-24.
153. Pattanapanyasat K, Thakar MR. CD4+ T cell count as a tool to monitor HIV progression & anti-retroviral therapy. *Indian J Med Res*. 2005;121(4):539-549.
154. Public Health Agency of Canada. Chlamydial Infections - Section 5 - Management and Treatment of Specific Infections - Canadian Guidelines on Sexually Transmitted Infections. 2013.
155. Steben M, Money D, Wong T, et al. Genital herpes: gynaecological aspects. *J Obstet Gynaecol Can*. 2008;30(207):347-361.
156. Ameur A, Meiring TL, Bunikis I, et al. Comprehensive profiling of the vaginal microbiome in HIV positive women using massive parallel semiconductor sequencing. *Sci Rep*. 2014;4:4398.
157. Fife KH, Fortenberry JD, Ofner S, Katz BP, Morrow RA, Orr DP. Incidence and prevalence of herpes simplex virus infections in adolescent women. *Sex Transm Dis*.

2006;33(7):441-444.
158. Patrick DM, Dawar M, Cook DA, Krajden M, Ng HC, Rekart ML. Antenatal Seroprevalence of Herpes Simplex Virus Type 2 (HSV-2) in Canadian Women. *Sex Transm Dis*. 1999;28(7):424-428.
159. Public Health Agency of Canada. Canadian Guidelines on Sexually Transmitted Infections - Section 5 Management and Treatment of Specific Infections - Genital Herpes simplex virus (HSV) Infections. 2013.
160. Fleming DT, Leone P, Esposito D, et al. Herpes virus type 2 infection and genital symptoms in primary care patients. *Sex Transm Dis*. 2006;33(7):416-421.
161. Allsworth JE, Lewis V a, Peipert JF. Viral sexually transmitted infections and bacterial vaginosis: 2001-2004 National Health and Nutrition Examination Survey data. *Sex Transm Dis*. 2008;35(9):791-6.
162. Ma Y, Madupu R, Karaoz U, et al. Human papillomavirus community in healthy persons, defined by metagenomics analysis of human microbiome project shotgun sequencing data sets. *J Virol*. 2014;88(9):4786-97.
163. Money DM, Roy M, Provencher DM, Steben M, Murphy KJ, Bryson P. Canadian Consensus Guidelines on Human Papillomavirus Canadian Consensus Guidelines on. 2007;29(8).
164. Moore R a, Ogilvie G, Fornika D, et al. Prevalence and type distribution of human papillomavirus in 5,000 British Columbia women - implications for vaccination. *Cancer Causes Control*. 2009;20:1387-96.
165. Public Health Agency of Canada. Human papillomavirus (HPV) infections - Section 5 Management and Treatment of Specific Infections. 2014.
166. Ogilvie GS, van Niekerk DJ, Krajden M, et al. A randomized controlled trial of Human Papillomavirus (HPV) testing for cervical cancer screening: trial design and preliminary results (HPV FOCAL Trial). *BMC Cancer*. 2010;10(1):111.
167. Mao C, Hughes JP, Kiviat N, et al. Clinical findings among young women with genital human papillomavirus infection. *Am J Obstet Gynecol*. 2003;188(August 1997):677-684.
168. Mehta SD, Donovan B, Weber KM, et al. The vaginal microbiota over an 8- To 10-year period in a cohort of HIV-infected and HIV-uninfected women. *PLoS One*. 2015;10(2):e0116894.
169. Van Kuppeveld FJM, Van der Logt JTM, Angulo AF, et al. Genus- and species-specific identification of mycoplasmas by 16S rRNA amplification. *Appl Environ Microbiol*. 1992;58(8):2606-2615.
170. Watson HL, Blalock DK, Cassell GH. Variable antigens of Ureaplasma urealyticum containing both serovar-specific and serovar-cross-reactive epitopes. *Infect Immun*. 1990;58(11):3679-3688.
171. Champely S, Ekstrom C, Dalgaard P, Gill J, Weibelzahl S, De Rosario H. Package “pwr” Basic Functions for Power Analysis. 2016.
172. Money D. The laboratory diagnosis of bacterial vaginosis. *Can J Infect Dis Med Microbiol*. 2005;16(2):77-79.
173. Oksanen J, Blanchet FG, Friendly M, et al. Package “vegan” Community Ecology Package. 2016.
174. Kindt R. BiodiversityR: Package for Community Ecology and Suitability Analysis. 2016.
175. Marcon E, Herault B. Package “entropart” Entropy Partitioning to Measure Diversity. 2016.
176. R Core Team. R: The R Stats Package. 2016.
177. Meyer D, Zeileis A, Gerber F, Friendly M. Visualizing Categorical Data: Package vcd. 2015.
178. Torchiano M. Package “effsize” Efficient Effect Size Computation. 2016.
179. Gloor GB, Macklaim JM, Fernandes AD. Displaying Variation in Large Datasets: Plotting

a Visual Summary of Effect Sizes. *J Comput Graph Stat*. 2016;25(3):971-979.
180. Roberts DW. Package “labdsv” Ordination and Multivariate Analysis for Ecology. 2016.
181. Li L, Victoria JG, Wang C, et al. Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. *J Virol*. 2010;84(14):6955-65.
182. Shan T, Li L, Simmonds P, Wang C, Moeser A, Delwart E. The Fecal Virome of Pigs on a High-Density Farm. *J Virol*. 2011;85(22):11697-11708.
183. Zhang W, Li L, Deng X, Kapusinszky B, Pesavento P a, Delwart E. the Fecal Virome of Cats in an Animal Shelter. *J Gen Virol*. 2014:2553-2564.
184. Zúñiga S, Sola I, Moreno JL, Sabella P, Plana-Durán J, Enjuanes L. Coronavirus nucleocapsid protein is an RNA chaperone. *Virology*. 2007;357(2):215-27.
185. Simons JF, Hutchison SK. Metagenomics analysis using the Genome Sequencer^{TM} FLX system. *Nat Methods*. 2007;(8):1-3.
186. Applied Biosystems. Essentials of Real Time PCR About Sequence Detection Chemistries TaqMan Chemistry. *ReVision*. 2005:1-8.
187. Wang D, Urisman A, Liu Y-T, et al. Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol*. 2003;1(2):E2.
188. Illumina. *Nextera XT DNA Library Preparation Guide*.; 2015.
189. Flygare S, Simmon K, Miller C, et al. Taxonomer: an interactive metagenomics analysis portal for universal pathogen detection and host mRNA expression profiling. *Genome Biol*. 2016;17(111).
190. World Health Organization. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans Human Papillomaviruses. 2007;(90).
191. World Health Organization. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. 2012;100B.
192. Yoshida T, Tashiro Y, Sonomoto K. Novel high butanol production from lactic acid and pentose by Clostridium saccharoperbutylacetonicum. *J Biosci Bioeng*. 2012;114(5):526-30.
193. Srinivasan S, Liu C, Mitchell CM, et al. Temporal variability of human vaginal bacteria and relationship with bacterial vaginosis. *PLoS One*. 2010;5(4):e10197.
194. Sung C, Kim BG, Kim S, Joo HS, Kim PI. Probiotic potential of Staphylococcus hominisi MBBL 2-9 as anti-Staphylococcus aureus agent isolated from the vaginal microbiota of a healthy woman. *J Appl Microbiol*. 2010;108(3):908-916.
195. Dewhirst FE, Chen T, Izard J, et al. The human oral microbiome. *J Bacteriol*. 2010;192(19):5002-17.
196. Saini R, Saini S, Sharma S. Oral sex, oral health and orogenital infections. *J Glob Infect Dis*. 2010;2(1):57-62.
197. Ballini A, Cantore S, Fatone L, et al. Transmission of nonviral sexually transmitted infections and oral sex. *J Sex Med*. 2012;9(2):372-84.
198. Schwebke JR, Richey CM, Weiss HL. Correlation of behaviors with microbiological changes in vaginal flora. *J Infect Dis*. 1999;180(5):1632-1636.
199. Tchamouroff SE, Panja SK. The association between receptive cunnilingus and bacterial vaginosis. *Sex Transm Infect*. 2000;76(2):144-145.
200. Edwards S, Carne C. Oral sex and transmission of non-viral STIs. *Sex Transm Infect*. 1998;74(2):95-100.
201. Fethers KA, Fairley CK, Morton A, et al. Early sexual experiences and risk factors for bacterial vaginosis. *J Infect Dis*. 2009;200(11):1662-70.
202. Brotman RM, Bradford LL, Conrad M, et al. Association between Trichomonas vaginalis and vaginal bacterial community composition among reproductive-age women. *Sex Transm Dis*. 2012;39(10):807-12.
203. Wiesenfeld HC, Hillier SL, Krohn MA, Landers D V, Sweet RL. Bacterial vaginosis is a strong predictor of Neisseria gonorrhoeae and Chlamydia trachomatis infection. *Clin*

Infect Dis. 2003;36(5):663-8.
204. Brotman RM, Ravel J, Cone RA, Zenilman JM. Rapid fluctuation of the vaginal microbiota measured by Gram stain analysis. *Sex Transm Infect*. 2010;86(4):297-302.
205. Allsworth JE, Peipert JF. Severity of bacterial vaginosis and the risk of sexually transmitted infection. *Am J Obstet Gynecol*. 2011;205(2):113.e1-6.
206. Center for Disease Control and Prevention. Bacterial Vaginosis - 2015 STD Treatment Guidelines. 2015.
207. Nilsson U, Hellberg D, Shoubnikova M, Nilsson S, Mårdh PA. Sexual behavior risk factors associated with bacterial vaginosis and Chlamydia trachomatis infection. *Sex Transm Dis*. 1997;24(5):241-246.
208. Gong Z, Luna Y, Yu P, Fan H. Lactobacilli Inactivate Chlamydia trachomatis through Lactic Acid but Not H2O2. *PLoS One*. 2014;9(9):e107758.
209. Mastromarino P, Di Pietro M, Schiavoni G, Nardis C, Gentile M, Sessa R. Effects of vaginal lactobacilli in Chlamydia trachomatis infection. *Int J Med Microbiol*. 2014;304(5-6):654-61.
210. Rizzo A, Fiorentino M, Buommino E, Donnarumma G, Losacco A, Bevilacqua N. Lactobacillus crispatus mediates anti-inflammatory cytokine interleukin-10 induction in response to Chlamydia trachomatis infection in vitro. *Int J Med Microbiol*. 2015;305(8):815-27.
211. Nardini P, Ñahui Palomino RA, Parolin C, et al. Lactobacillus crispatus inhibits the infectivity of Chlamydia trachomatis elementary bodies, in vitro study. *Sci Rep*. 2016;6(29024).
212. Latka M, Ahern J, Garfein RS, et al. Prevalence, incidence, and correlates of chlamydia and gonorrhea among young adult injection drug users. *J Subst Abuse*. 2001;13(1):73-88.
213. Tyndall MW, Patrick D, Spittal P, Li K, O'Shaughnessy M V, Schechter MT. Risky sexual behaviours among injection drugs users with high HIV prevalence: implications for STD control. *Sex Transm Infect*. 2002;78(Suppl 1):i170-5.
214. Plitt SS, Garfein RS, Gaydos C a, Strathdee S a, Sherman SG, Taha TE. Prevalence and correlates of chlamydia trachomatis, neisseria gonorrhoeae, trichomonas vaginalis infections, and bacterial vaginosis among a cohort of young injection drug users in Baltimore, Maryland. *Sex Transm Dis*. 2005;32(7):446-453.
215. Poulin C, Alary M, Bernier F, Ringuet J, Joly JR. Prevalence of Chlamydia trachomatis, Neisseria gonorrhoeae, and HIV infection among drug users attending an STD/HIV prevention and needle-exchange program in Quebec City, Canada. *Sex Transm Dis*. 1999;26(7):410-420.
216. Bachmann LH, Lewis I, Allen R, et al. Risk and prevalence of treatable sexually transmitted diseases at a Birmingham substance abuse treatment facility. *Am J Public Health*. 2000;90(10):1615-1618.
217. Vitali B, Pugliese C, Biagi E, et al. Dynamics of vaginal bacterial communities in women developing bacterial vaginosis, candidiasis, or no infection, analyzed by PCR-denaturing gradient gel electrophoresis and real-time PCR. *Appl Environ Microbiol*. 2007;73(18):5731-5741.
218. Redondo-Lopez V, Meriwether C, Schmitt C, Opitz M, Cook R, Sobel JD. Vulvovaginal candidiasis complicating recurrent bacterial vaginosis. *Sex Transm Dis*. 1990;17(1):51-53.
219. McClelland RS, Richardson BA, Hassan WM, et al. Prospective study of vaginal bacterial flora and other risk factors for vulvovaginal candidiasis. *J Infect Dis*. 2009;199(12):1883-90.
220. Rivers CA, Adaramola OO, Schwebke JR. Prevalence of bacterial vaginosis and vulvovaginal candidiasis mixed infection in a southeastern american STD clinic. *Sex*

Transm Dis. 2011;38(7):672-4.
221. Leclair CM, Hart AE, Goetsch MF, Carpentier H, Jensen JT. Group B Streptococcus. *J Low Genit Tract Dis*. 2010;14(3):162-166.
222. Li J, McCormick J, Bocking A, Reid G. Importance of vaginal microbes in reproductive health. *Reprod Sci*. 2012;19(3):235-42.
223. Vicariotto F, Mogna L, Piano M Del. Effectiveness of the Two Microorganisms Lactobacillus fermentum LF15 and Lactobacillus plantarum LP01, Formulated in Slow-release Vaginal Tablets, in Women Affected by Bacterial Vaginosis. *J Clin Gastroenterol*. 2014;48:S106-S112.
224. Schloss PD, Handelsman J. Toward a Census of Bacteria in Soil. *PLoS Comp Biol*. 2006;1(7).
225. Huang B, Alexander Taylor Harper D, Zhan R. Test of sampling sufficiency in palaeontology. *GFF*. 2014;136(1):105-109.
226. Lee H, Han S, Kim M, et al. In-depth analysis of sampling optimization methods. In: Sanchez MI, Ukraintsev VA, eds. ; 2016.
227. Good IJ. the Population Frequencies of Species and the Estimation of Population Parameters. *Biometrika*. 1953;40(3-4):237-264.
228. Clarke MA, Rodriguez AC, Gage JC, et al. A large, population-based study of age-related associations between vaginal pH and human papillomavirus infection. *BMC Infect Dis*. 2012;12:33.
229. Kiliç AO, Pavlova SI, Alpay S, Kiliç SS, Tao L. Comparative study of vaginal Lactobacillus phages isolated from women in the United States and Turkey: prevalence, morphology, host range, and DNA homology. *Clin Diagn Lab Immunol*. 2001;8(1):31-9.
230. Government of Canada PHA of C. Epstein-Barr Virus. 2010.
231. Caserta MT, Mock DJ, Dewhurst S. Human herpesvirus 6. *Clin Infect Dis*. 2001;33(6):829-33.
232. Hill JA, Sedlak RH, Jerome KR. Past, present, and future perspectives on the diagnosis of Roseolovirus infections. *Curr Opin Virol*. 2014;9:84-90.
233. Agut H, Bonnafous P, Gautheret-Dejean A. Laboratory and clinical aspects of human herpesvirus 6 infections. *Clin Microbiol Rev*. 2015;28(2):313-35.
234. Ablashi D V, Chatlynne LG, Whitman JE, Cesarman E, Cesarman E. Spectrum of Kaposi's sarcoma-associated herpesvirus, or human herpesvirus 8, diseases. *Clin Microbiol Rev*. 2002;15(3):439-64.
235. Dittmer DP, Damania B, Silverberg M, et al. Kaposi sarcoma–associated herpesvirus: immunobiology, oncogenesis, and therapy. *J Clin Invest*. 2016;126(9):3165-3175.
236. Chatlynne LG, Ablashi D V. Seroepidemiology of Kaposi's sarcoma-associated herpesvirus (KSHV). *Semin Cancer Biol*. 1999;9(3):175-185.
237. National Institutes of Health. Guidelines for Prevention and Treatment of Opportunistic Infections in HIV-Infected Adults and Adolescents Human Herpesvirus-8 Disease. 2013.
238. Silverberg MJ, Lau B, Achenbach CJ, et al. Cumulative Incidence of Cancer Among Persons With HIV in North America. *Ann Intern Med*. 2015;163(7):507.
239. Cannon MJ, Dollard SC, Black JB, et al. Risk factors for Kaposi's sarcoma in men seropositive for both human herpesvirus 8 and human immunodeficiency virus. *AIDS*. 2003;17(2):215-22.
240. Bagni R, Whitby D. Kaposi's sarcoma-associated herpesvirus transmission and primary infection. *Curr Opin HIV AIDS*. 2009;4(1):22-26.
241. Spear GT, Sikaroodi M, Zariffard MR, Landay AL, French AL, Gillevet PM. Comparison of the diversity of the vaginal microbiota in HIV-infected and HIV-uninfected women with or without bacterial vaginosis. *J Infect Dis*. 2008;198(8):1131-40.
242. Kang M, Cu-Uvin S. Association of HIV Viral Load and CD4 with HPV Detection and Clearance in HIV Infected Women Initiating HAART. *HIV Med*. 2012;13(6):372-378.

243. Clarke B, Chetty R. Postmodern cancer: the role of human immunodeficiency virus in uterine cervical cancer. *Mol Pathol*. 2002;55(1):19-24.
244. Heard I, Palefsky JM, Kazatchkine MD. The impact of HIV antiviral therapy on human papillomavirus (HPV) infections and HPV-related diseases. *Antivir Ther*. 2004;9(1):13-22.
245. Fife KH, Wu JW, Squires KE, Heather Watts D, Andersen JW, Brown DR. Prevalence and Persistence of Cervical Human Papillomavirus Infection In HIV-Positive Women Initiating Highly-Active Antiretroviral Therapy. *J Acquir Immune Defic Syndr*. 2009;51(3):274-282.
246. Paramsothy P, Jamieson DJ, Heilig CM, et al. The effect of highly active antiretroviral therapy on human papillomavirus clearance and cervical cytology. *Obstet Gynecol*. 2009;113(1):26-31.
247. Konopnicki D, Manigart Y, Gilles C, et al. Sustained viral suppression and higher CD4+ T-cell count reduces the risk of persistent cervical high-risk human papillomavirus infection in HIV-positive women. *J Infect Dis*. 2013;207(11):1723-9.
248. Menon SS, Rossi R, Harebottle R, Mabeya H, Vanden Broeck D. Distribution of human papillomaviruses and bacterial vaginosis in HIV positive women with abnormal cytology in Mombasa, Kenya. *Infect Agent Cancer*. 2016;11(17).

Appendix

Appendix A

CHILDREN'S & WOMEN'S HEALTH CENTRE OF BRITISH COLUMBIA

WOMEN'S HEALTH RESEARCH INSTITUTE AT BC WOMEN'S

Subject Information and Consent Form

Vaginal Microbiome Group Initiative (Vogue) – Study 1B2

Principal Investigator: **Dr. Deborah Money**, Professor, MD, FRCSC
UBC Department of Obstetrics and Gynaecology
Women's Health Research Institute
[redacted]

Researcher: **Research Coordinator**
Women's Health Research Institute, BC Women's Hospital and Health Centre
[redacted]

Emergency Telephone Number: Dr. Deborah Money, [redacted]

Sponsor(s): **Canadian Institutes of Health Research (CIHR)**
Genome British Columbia (Genome BC)

INTRODUCTION
You are being invited to take part in this research study because you are experiencing symptoms of a condition known as vulvovaginitis (inflammation or infection of the vulva and vagina) and are scheduled for a planned pelvic exam.

YOUR PARTICIPATION IS VOLUNTARY
Your participation is entirely voluntary, so it is up to you to decide whether or not to take part in this study. Before you decide, it is important for you to understand why the research is being done and what it will involve. This consent form will tell you about the study and why the research is being done, what it will involve, and the possible benefits, risks and discomforts to help you decide whether or not you wish to take part.

If you wish to participate, you will be asked to sign this form. If you do decide to take part in this study, you can choose to withdraw at any time without giving any reasons for your decision.

If you do not wish to participate, you do not have to provide any reason for your decision not to participate, nor will you lose the benefit of any medical care to which you are entitled or are presently receiving.

Please take time to read the following information carefully and to discuss it with your family and your doctor before you decide.

BACKGROUND AND PURPOSE OF THIS STUDY

Changes to the communities of bacteria that naturally exist inside the vagina can have negative effects for women. These effects include: increased risk of being infected with sexually transmitted infections, preterm birth, problems becoming pregnant, early pregnancy loss, and infections that may significantly impact quality of life. However, researchers and doctors still do not know very much about what kinds of bacteria and how many of these bacteria are normally present in a healthy vagina. A better understanding of what types of bacteria are in the vagina is needed in order to identify the subtle imbalances and shifts in bacterial populations that are "healthy" and maintain reproductive health, versus "unhealthy" that have negative effects and increase the chances of disease. As part of our study, we want to study the vaginal bacteria present in a sample of 50 women with recurrent vulvovaginitis between the ages of 18 and 49.

The purpose of our study is to identify the different types and numbers of bacteria living in the vagina using new, highly specific DNA-based methods. We want to learn which bacteria in the vagina are associated with health and which are associated with disease. We hope the information gained through our study will help us to develop tests and therapy to diagnose and treat abnormal vaginal bacteria before they lead to greater health problems for women.

WHO CAN PARTICIPATE IN THIS STUDY?

Women presenting to a research or medical clinic for a pelvic examination who:

1. Have a history of recurrent vulvovaginitis (defined as at least four episodes in 12 months with symptoms that may include pruritus (itchiness), burning, pain, discharge, or dyspareunia (painful sexual intercourse).
2. Are 18 to 49 years of age.
3. Have an adequate comprehension of the English language to sign written informed consent.
4. Are not currently pregnant.
5. Are not menopausal.
6. Are HIV negative

WHO SHOULD NOT PARTICIPATE IN THIS STUDY?

Women who have one or more of the following exclusion criteria SHOULD NOT participate:

1. Do not have a history of recurrent vulvovaginitis.
2. Are younger than 18 years of age or older than 49 years of age.
3. Are menopausal.
4. Are not able to provide written informed consent.
5. Are currently pregnant.
6. Are HIV positive

WHAT DOES THIS STUDY INVOLVE FOR YOU?

If you decide to take part in this study, and sign this consent form, you can expect the following:

- If you attend a clinic for gynaecological services, you will be scheduled for a *speculum exam* as part of your visit. A speculum will be inserted into the vagina to open the vaginal canal in order to see the cervix (which is at the end of the vagina). Swabs will then be taken by gently brushing against the surface of the vagina. These samples will be taken as indicated for a planned exam to test for infections. These samples are unrelated to the study and will be collected as part of your scheduled examination regardless of whether you participate in the study or not.
    - For women who have consented to the study, we will take four additional swab samples from your vagina. The collection of additional vaginal swabs will be taken at the same time as your planned swabs and will take a minimal amount of time (1-2 minutes). The collection of the additional samples should not add any discomfort to your examination.

- If you attend the Women's Health Research Institute (WHRI) Research Clinic at BC Women's Hospital and consent to the study, you will also undergo a speculum exam. Four study samples will then be taken by gently brushing a swab against the surface of the vagina. In addition, vaginal or cervical swabs will be taken to test for infections if the nurse or physician examining you feels that it is required. These samples will only be taken if it is clinically indicated. If there are any abnormalities noted during the exam, you will be referred to your regular family physician for appropriate medical follow-up. The maximum amount of time required for the test and collection of study samples is 20 minutes.

- It is possible that study staff may contact you in the future to request further samples to do more testing directly related to this study. Whether or not you agree to provide any additional samples in the future will be completely voluntary and you do not have to provide any reason for your decision if you decide not to provide the additional samples. If you do agree to provide additional samples in the future, this will be reviewed with you in a separate consent form.

- Either prior to or following your exam, a researcher will ask you questions related to your medical and sexual history. This interview may be conducted in-person or over the phone. You do not have to answer any questions you do not feel comfortable answering. Answering the questions should only take 15-20 minutes of your time. It is possible that study staff may need to contact you in the future by telephone or email to request further information regarding your medical or sexual history.

- You will also be asked to complete an online survey that asks questions about knowledge, attitudes and behaviors that may affect vaginal health. The Research Coordinator will provide you with an ID number to access the survey. No personal information that could be used to identify you, such as your name or date of birth, will be collected or stored online in this survey.

- Once you have completed the survey, the data is sent to a secure facility at the University of Guelph, where it will be analyzed by members of the study team. You may skip over any questions you do not feel comfortable answering. Answering the online questions

should take 10-30 minutes of your time. If you would prefer to complete the questionnaire using a paper form, this can be provided to you.

- One of the four vaginal swabs we collect will be transported to the clinical laboratory at the BC Women's and Children's Hospital Laboratory for Gram Stain analysis. This involves a lab technician taking your vaginal sample and making a smear on a glass slide using color stains and then viewing your sample using a microscope to examine the different bacteria that are present in the vagina.

- The remaining three vaginal swabs we collect will be stored frozen at -20°C or -80°C as required and transported to the BC Women's Research Laboratory where they can only be accessed by researchers directly involved with our study. The identities of the subjects from which the swab samples were obtained will be kept strictly confidential and can only be accessed by study researchers. Samples will be stored at the BC Women's Research Laboratory and shipped to our partner laboratories at the University of Saskatchewan and the University of Western Ontario for genomic sequencing. The sequencing results will be used to identify different types of bacteria and other microorganisms that are present in the sample. Your DNA or tissue will not be analyzed in any of these samples as this study's objective is to study bacteria and other microorganisms.

- If any unused samples are left after the study testing has been done, we will ask your permission to store your samples at the BC Women's Research Laboratory or one of our partner laboratory sites for a maximum of 25 years for future research related to this study or similar studies looking at the vaginal microbiome. This will be reviewed with you in a separate consent form.

POTENTIAL RISKS AND BENEFITS

There are no research-related risks associated with this study. The collection of additional swabs for the purpose of this study will require a very small amount of additional time (approximately 1-2 minutes) over your planned examination, or up to a maximum of 20 minutes for a WHRI research clinic visit. In some cases, women experience minor discomfort when vaginal swabs are collected. We will minimize the inconvenience and potential discomfort by collecting all study swabs at the same time.

There are no direct benefits to participating in this study. However, you are possibly benefiting women in the future by helping us determine what types of vaginal bacteria are associated with health and disease in women.

NEW FINDINGS

You will be told of any new information learned during the course of the study that might cause you to change your mind about staying in the study. At the end of the study, you will be provided with the overall results of the study; however, we will not provide individual results.

WHAT HAPPENS IF YOU DECIDE TO WITHDRAW YOUR CONSENT?

Your participation in this research is entirely voluntary. You are under no obligation to be included in this study. You may withdraw from this study at any time. If you decide to enter the

study and to withdraw at any time in the future, there will be no penalty or loss of benefits to which you are otherwise entitled, and your future medical care will not be affected. If you choose to withdraw, your samples and collected data will be destroyed.

COSTS AND REIMBURSEMENTS
The study doctor will not receive any money for your participation in this study. There is no cost to you for participating in this study. You will, however, receive a $20 honorarium for participating in this study.

CONFIDENTIALITY
Your confidentiality will be respected. However, research records and health or other source records identifying you may be inspected in the presence of the Principal Investigator or her designate by representatives of Health Canada and the UBC Research Ethics Boards for the purpose of monitoring the research. No information or records that disclose your identity will be published without your consent, nor will any information or records that disclose your identity be removed or released without your consent unless required by law.

You will be assigned a unique study number as a subject in this study. Only this number will be used on any research-related information collected about you during the course of this study, so that your identity [i.e. your name or any other information that could identify you] as a subject in this study will be kept confidential. Information that contains your identity will remain only with the Principal Investigator (Dr. Deborah Money) and/or designates at the Women's Health Research Institute in Vancouver. The list that matches your name to the unique study number that is used on your research-related information will be retained at the Women's Health Research Institute in a locked filing cabinet in a locked room, and will not be removed or released without your consent unless required by law.

Your rights to privacy are legally protected by federal and provincial laws that require safeguards to insure that your privacy is respected and also give you the right of access to the information about you that has been provided to the sponsor and, if need be, an opportunity to correct any errors in this information. Further details about these laws are available on request to your study doctor.

Signing this consent form in no way limits your legal rights against the sponsor, investigators, or anyone else.

WHO TO CONTACT IF YOU HAVE QUESTIONS OR CONCERNS ABOUT YOUR RIGHTS AS A SUBJECT DURING THE STUDY
If you have any questions or concerns about this study, please contact **Dr. Deborah Money** at [redacted]

If you have any questions or concerns regarding your rights as a research subject, please call the **Research Subject Information Line** in the University of British Columbia (UBC) Office of Research Services at [redacted] or [redacted] You may also email your questions or concerns to [redacted]

Subject Information and Consent Form

Vaginal Microbiome Project Team (Vogue) – Study 1B2

Please indicate the following by marking the boxes below:

- I have read and understood the subject information and consent form.
- I have had sufficient time to consider the information provided and to ask for advice if necessary.
- I have had the opportunity to ask questions and have had satisfactory responses to my questions.
- I understand that the information collected will be kept confidential and that the result will only be used for scientific objectives.
- I understand that my participation in any research study is voluntary and that I am completely free to refuse to participate or to withdraw from this study at any time without changing in any way the quality of care that I receive.
- I understand that I am not waiving any of my legal rights as a result of signing this consent form.
- I understand that there is no guarantee that this study will provide any benefits to me.
- I understand study staff may contact me in the future by telephone or email if they have further questions related to my medical or sexual history that can not be found in my chart.
- I understand study staff may contact me in the future by telephone or email to request additional vaginal swab samples for further testing directly related to this study. I understand that providing additional samples in the future is voluntary and I can choose not to provide the samples without giving any reason for my decision.
- I have read this form and I freely consent to participate in this study.
- I have been told that I will receive a dated and signed copy of this consent form for my records.
- I hereby consent to participate in the study.

Name of Subject (Please print)	***Name of Subject***	***Date***

Name of Person conducting consent *(Please print)*	***Signature of Person conducting consent***	***Date***

The person who may be contacted about this research is: **Dr. Deborah Money**, who can be reached at [redacted]

Appendix B

CHILDREN'S & WOMEN'S HEALTH CENTRE OF BRITISH COLUMBIA

WOMEN'S HEALTH RESEARCH INSTITUTE AT BC WOMEN'S

Subject Information and Consent Form for Optional Tissue Banking

Vaginal Microbiome Group Initiative (Vogue) – Study 1B2

Principal Investigator: **Dr. Deborah Money**, Professor, MD, FRCSC
UBC Department of Obstetrics and Gynaecology
Women's Health Research Institute
[redacted]

Researcher: **Research Coordinator**
Women's Health Research Institute, BC Women's Hospital and Health Centre
[redacted]

Emergency Telephone Number: Dr. Deborah Money, [redacted]

Sponsor(s): **Canadian Institutes of Health Research (CIHR)**
Genome British Columbia (Genome BC)

INTRODUCTION

In addition to the main part of the research study, you are being invited to participate in this optional study because the unused samples you provided as part of the main study could be used to continue our research in this area. As more information becomes available on the types of bacteria that live in the vagina, we would like to have the opportunity to do further laboratory studies on your unused tissues/samples. You can choose to participate in the main study without participating in this optional part of the study.

Additionally, if you are interested in hearing about new research studies, you can choose to give us permission to store your information in a secure database and contact you again about future opportunities to participate in our research.

This consent form will explain why we wish to keep your unused tissues/samples and what will happen to your tissues/samples after they are collected. Once you understand why we wish to bank (store) your unused samples, and if you agree to participate, you will be asked to sign this consent form. You will be given a copy of this form to keep for your records.

If you decide to take part in this optional part of the study you are free to withdraw at any time without giving any reasons for your decision.

If you do not wish to participate, you do not have to provide any reason for your decision, nor will you lose the benefit of any medical care to which you are entitled or are presently receiving.

Please take time to read the following information carefully and to discuss it with your family and your doctor if you wish before you decide.

BACKGROUND
You have already consented to participate in our main study, which will study what types of bacteria and other micro-organisms are in the vagina and how they contribute to women's health and disease.

PURPOSE
The reason we want to bank your unused tissue samples (if there are any of your samples remaining after the main study testing has been completed) is because this field of research is changing very rapidly and new discoveries are made often. If we keep your unused samples, we will be in a position to apply new tests that are not available today but might be available in the future. Of course, any new test that we would apply to your samples would be geared towards the same research goal: determining what types of micro-organisms are present in the vagina of women.

Examples of new tests may include:
- new technologies to identify specific types of bacteria
- new technologies to identify other micro-organisms in your vagina, such as viruses and fungi
- new technologies to analyze the chemicals produced by bacteria in your vagina
- new technologies to analyze the immune response produced by your body towards the bacteria present in your vagina

We also understand some women are interested in women's health issues and women's health research. Creation of a database of women who are interested in being involved in future research will allow us to inform them about upcoming studies for which they may be eligible.

WHO CAN PARTICIPATE?
The optional parts of the study are open to all subjects enrolled in the main part of the study.

WHAT DOES THIS PART OF THE STUDY INVOLVE?
The storage of your tissue samples does not involve any more of your time or the collection of any additional samples or information from you.

STUDY PROCEDURES / TISSUE BANKING DETAILS
- Your vaginal swabs (tissues) will have been collected as part of the main study (the time, and potential risks and discomforts this may involve is described in the main consent form).
- Any unused tissue samples will be stored at -80° C in a secure facility at the BC Women's Research Laboratory or at one of our partner laboratory sites at the University of

Saskatchewan (Hill Lab) or the University of Western Ontario (Reid Lab).

- The individual study participant's DNA or tissue will not be analyzed in any of these samples as this study's objective is microbial analysis.
- The stored tissue samples will have none of your personally identifying information on them. Your sample will be identified with your main study ID number only.
- Your name, as a part of the main study, is on a password-protected electronic master list kept on a password-protected computer network in a locked research office at the Women's Health Research Institute.
- Dr. Deborah Money (BC Women's Research Laboratory), Dr. Janet Hill (Hill Lab) or Dr. Gregor Reid (Reid Lab) will be the custodian of your tissue samples for as long as they are stored.
- Zahra Pakzad, the WHRI Researcher for this project, is the custodian of the study master list at the Women's Health Research Institute.
- It is possible that tests yet to be developed or identified, but which are related to the tests being done in the main study could become available in the future. It is in the event such as this that the research team requests your consent to bank your tissue samples for an indefinite amount of time not to exceed 25 years.
- Should the research team want to conduct any future tests on your samples that are **NOT** directly related to the main study research or similar research studies of the vaginal microbiome, this would only be done after obtaining ethical approval from the UBC Children's and Women's Health Centre Research Ethics Board. This Board aims to help protect the rights of research subjects. Only after ethical approval had been received from the UBC Children's and Women's Health Centre Research Ethics Boards would you be invited to participate. You could then choose whether or not to give your consent**.**
- Only Investigators who are part of the main study or collaborating with the study investigators will have access to your banked samples.
- Contact information will be collected from women who are interested in hearing about future research and will be kept in a password-protected electronic master list. The WHRI Researcher, Zahra Pakzad, will be the custodian of the master list.
- If at any time you wish to have your tissue samples destroyed, you can contact Dr. Deborah Money at [redacted], and she will personally ensure your samples are destroyed as you have requested.

POTENTIAL RISKS AND BENEFITS

There are no known risks associated with this optional study.

There are no direct benefits to participating in this study. However, we hope that by participating you are possibly benefiting future women by advancing current knowledge of the micro-organisms present in the vagina and how they contribute to women's health and disease.

ALTERNATIVES TO PARTICIPATION IN THIS PART OF THE STUDY

The alternative is to not participate. Choosing to not participate will have no effect on your participation in the main study.

NEW FINDINGS
You will be told of any new information learned during the course of the study that might cause you to change your mind about participating in the study.

COSTS AND REIMBURSEMENTS
There are no personal expenses or remunerations as a result of participating in this part of the study. The study doctor will not receive any money for your participation in this optional part of the study.

CONFIDENTIALITY
Your confidentiality will be respected. No information that discloses your identity will be released or published without your specific consent to the disclosure. All information that is obtained will be dealt with in a confidential manner. The information will be entered into a data file. This data will be identified by code-number instead of by your name. Only the researchers will have access to the code. However, research records and medical records identifying you may be inspected in the presence of the Investigator or his/her designate, Health Canada, and the UBC Research Ethics Boards for the purpose of monitoring the research. However, no records that identify you by name or initials will be allowed to leave the Investigators' offices.

Signing this consent form in no way limits your legal rights against the sponsor, investigators, or anyone else.

WHO TO CONTACT IF YOU HAVE QUESTIONS OR CONCERNS ABOUT YOUR RIGHTS AS A SUBJECT DURING THE STUDY
If you have any questions or concerns about this study, please contact **Dr. Deborah Money** at [redacted]

If you have any questions or concerns regarding your rights as a research subject, please call the **Research Subject Information Line** in the University of British Columbia (UBC) Office of Research Services at [redacted] or [redacted] You may also email your questions or concerns to [redacted]

Optional Subject Information and Consent Form for Tissue Banking

Vaginal Microbiome Project Team – Study 1B2

Please indicate the following by marking the boxes below:

- I have read and understood the subject information and consent form.
- I have had sufficient time to consider the information provided and to ask for advice if necessary.
- I have had the opportunity to ask questions and have had satisfactory responses to my questions.
- I understand that the information collected will be kept confidential and that the result will only be used for scientific objectives.
- I understand that my participation in any research study is voluntary and that I am completely free to refuse to participate or to withdraw from this study at any time without changing in any way the quality of care that I receive.
- I understand that I am not waiving any of my legal rights as a result of signing this consent form.
- I understand that there is no guarantee that this study will provide any benefits to me.
- I have read this form and I freely consent to participate in this study.
- I have been told that I will receive a dated and signed copy of this consent form for my records.

Please tick boxes:

□ I hereby consent to allow the researchers to store my samples for future studies for a period of time not to exceed 25 years.

□ I hereby consent to allow the researchers to add my name and contact information to a list of people interested in being contacted about future research studies.

Name of Subject (Please print)	***Signature of Subject***	***Date***
Name of Person conducting consent *(Please print)*	**Signature of Person conducting consent**	***Date***

Appendix C

PELVIC EXAM FINDINGS

a. External genital exam performed □ not performed □

□ Normal

□ Abnormal → □ warts/condylomas
□ ulcer
□ evidence of female circumcision
□ dermatological abnormality:
specify if possible:_____
□ other, specify: _____

b. Speculum exam performed □ not performed □

Vaginal Appearance:

□ Normal

□ Abnormal → Please specify _____

Vaginal Discharge:

□ Normal

□ Abnormal → Please specify: Colour:_____
Consistency: _____
Volume: _____

Cervical Appearance:

□ Normal

□ Abnormal (e.g. mucopurulent cervicitis) → Please specify _____

d. Wet Mount: performed □ not performed □

pH: lower□ normal □ elevated □

range (number): _____

e. HPV Positive: Yes ☐ No ☐ Unknown ☐ Type (if available) _____

HPV vaccine: Yes ☐ No ☐ Date: _ _ / _ _ _ / _ _ _

f. Most Recent Pap Results Date: _ _ / _ _ _ / _ _ _ _

Result (if available) _____

g. Other: _____

Comments:_____

Verbal Consent Obtained Prior to Sample Collection: ☐ Yes ☐ No

Verbal Consent / Pelvic Performed by _____ Research Staff Signature

CLINICAL CHECKLIST

DATE exam performed: _ _ / _ _ _ / _ _ _ _

TIME exam performed: _ _:_ _

	ITEM	YES	NO	Specify Date/Time if different from above
1.	External Genital Exam performed	☐	☐	__ __/__ __ __/__ __ _____:_____
3.	Speculum Exam performed	☐	☐	
A.	Gram Stain Study Swab collected	☐	☐	
B.	Study Tubes collected: Number of tubes: _____ Samples collected: A ☐ B ☐ C ☐ D ☐	☐	☐	
C.	Chlamydia / Gonorrhoea Swabs collected (not required; not included in study kit)	☐	☐	
D.	*Culture Swab collected:* *(not required; not included in study kit)*	☐	☐	
E.	*Herpes Swab collected:* *(not required; not included in study kit)*	☐	☐	
F.	*Trichomonas Swab collected:* *(not required; not included in study kit)*	☐	☐	
G.	*Wet Mount (optional)*	☐	☐	

Obtained by _____ Research Staff Signature

LAB RESULTS (to be completed post-visit). Please attach any documentation.

Report Date (dd/mm/year): __ __/ __ __/ __ __ __ __

Result

Gonorrhea: **+ □ / - □**
Done □
Not Done □
Not Processed □

Chlamydia: + □ / - □
Done □
Not Done □
Not Processed □

Herpes: + □ / - □
Done □
Not Done □
Not Processed □

Trichomonas: + □ / - □
Done □
Not Done □
Not Processed □

Gram Stain:
Nugent's Score (0-10) _____
Is the Nugent score:
Consistent with BV □
Intermediate BV □
Not Consistent with BV □
N/A no bacterial cells □

Interpretation:_____
_____
_____

Vaginal Swab:
Was Culture done: Yes □ / No □
Result:_____
_____

Obtained by _____ Research Staff Signature

Appendix D

INCLUSION CRITERIA

	YES	NO	n/a
Subject has vulvovaginitis	☐	☐	
Subject is HIV negative	☐	☐	
Subject is between 18 and 49 years of age	☐	☐	
Subject is not menopausal	☐	☐	
Subject is not currently pregnant	☐	☐	

Does subject meet all inclusion criteria?

Yes ☐ ***No*** ☐

Confirmed by: _____ Research Staff Signature

INFORMED CONSENT PROCESS

	YES	NO	n/a
1. Was the study "Informed Consent" form presented in the subject's language of preference? If English is not language of preference, list language preference_____	☐	☐	
2. If English is *not* the subject's language of preference, was a translator present when the Informed Consent forms were read and discussed?	☐	☐	☐
3 Does the subject understand the study procedures and agrees to participate in the study by giving written informed consent.	☐	☐	
4. Was the subject allowed to ask questions medical in nature?	☐	☐	
5. Have all the subject's questions about the study been answered? List questions asked _____ _____ If NO, please comment below:	☐	☐	
6. Does the subject understand that her participation in this study is voluntary? If NO, please comment below:	☐	☐	
8. Was the subject given a signed copy of the informed consent form?	☐	☐	

The Informed Consent form has been read in its entirety by the subject. Discussions have been conducted and the subject's questions have been answered by the Investigator/ RN/ Research staff member. The subject has signed the Informed Consent form prior to having any study procedures performed.

Yes ☐ ***No*** ☐

Date consent signed __ __/__ __ __/__ __ Time consent signed _____ (24-hour clock)

Performed by: _____ Research Staff Signature

Attach additional pages if needed.

DEMOGRAPHICS

AGE: _____

Height: _____ cm ☐ inches ☐ BMI(calculated during data entry): _____ kg/m²

Weight: _____ kg ☐ lbs ☐

Ethnicity: ☐ White / Caucasian
- ☐ Black / African Canadian
- ☐ Hispanic
- ☐ Asian
- ☐ South Asian
- ☐ Aboriginal / First Nations / Métis / Inuit _____
- ☐ Other (specify mother/father ethnicities) _____

Marital Status: ☐ Single ☐ Married/Common Law ☐ Other (specify)_____

Highest Education level attained:
- ☐ Did not complete high school
- ☐ High school diploma
- ☐ Some Post-secondary
- ☐ Post-secondary**/**Undergraduate Degree (Bachelor's)
- ☐ Graduate Degree (e.g. Master's, Ph.D)
- ☐ Other: _____

First 3 Digits of Postal Code: _____

Collected by: _____ Research Staff Signature

GENERAL MEDICAL HISTORY

Does participant have any known significant <u>current</u> or <u>chronic</u> disease?

☐ No ☐ Yes, please complete the following:

System	Diagnosis
Respiratory (e.g. asthma)	
Gastrointestinal (e.g. celiac, inflammatory or irritable bowel syndrome)	
Musculoskeletal (e.g. arthritis)	
Genitourinary (not incl. infections).	
Allergies/ Autoimmune Disorders	
Cardiovascular	
Other	

Collected by: _____ Research Staff Signature

GENITAL INFECTION HISTORY

Has the subject ever been diagnosed with one of the conditions listed below?

□ Uncertain □ No □ Yes, please complete the following list:

	Yes	No	Not Sure	Number of Infections			Treatment for most recent infection:
				Past 2 months	Past 1 Year	Lifetime	
Bacterial Vaginosis	☐	☐	☐				None □ Prescription □ Over the counter □ Natural products □
Yeast Infection Candida	☐	☐	☐				None □ Prescription □ Over the counter □ Natural products □
UTI	☐	☐	☐				None □ Prescription □ Over the counter □ Natural products □
Trichomoniasis	☐	☐	☐				None □ Prescription □ Over the counter □ Natural products □
Genital Warts Condylomas	☐	☐	☐				None □ Prescription □ Over the counter □ Natural products □
Genital Herpes	☐	☐	☐				None □ Prescription □ Over the counter □ Natural products □
Chlamydia	☐	☐	☐				None □ Prescription □ Over the counter □ Natural products □
Gonorrhea	☐	☐	☐				None □ Prescription □ Over the counter □ Natural products □
Syphillis	☐	☐	☐				None □ Prescription □ Over the counter □ Natural products □

Obtained by: _____ Research Staff Signature

ANTIMICROBIAL USE

Apart from the responses above and excluding antiretroviral combinations, has the subject taken any antimicrobials in the past 3 months? ☐ Yes ☐ No

(*This includes: oral medication, topical medication, and intravaginal medication*)

If YES, complete the following section:

ANTIMICROBIAL USE				
Drug name	**Date started**	**Date stopped**	**Dose/freq**	**Reason for antibiotic treatment**

Collected by: _____ Research Staff Signature

PRESCRIPTION/NON-PRESCRIPTION DRUG USE

Is this patient currently taking, or has this patient taken any prescription/non-prescription drugs, <u>including probiotic supplements or herbal remedies</u> and excluding antiretroviral combinations, in the past two months?

☐ Yes ☐ No

If YES, complete the following section:

PRESCRIPTION/NON-PRESCRIPTION DRUG USE				
Drug name	**Date started**	**Date stopped**	**Dose/freq**	**Reason for drug treatment**

Collected by: _____ Research Staff Signature

REPRODUCTIVE HEALTH

When was your last menstrual period? (1st day of LMP) __ __/__ __ __/__ __ __ __

Do you have a "normal" menstrual cycle? (i.e. period every 3-5 weeks) ☐ Yes ☐ No

In the past year, how often did you use tampons during your periods?

- □ Never
- □ Sometimes but not for every period
- □ Every period / Part of the time
- □ Every period / Exclusively

In the past year, how often did you use a menstrual cup (Instead, DivaCup) during your periods?

- □ Never
- □ Sometimes but not for every period
- □ Every period / Part of the time
- □ Every period / Exclusively

Pregnancy History :

G_____ T_____ P_____ SA_____ TA_____ L_____

Have you noticed any of the following vaginal symptoms?

	Past 48 hours	Past 2 weeks
Abnormal discharge	☐ Yes ☐ No	☐ Yes ☐ No
Abnormal odor	☐ Yes ☐ No	☐ Yes ☐ No
Irritation or discomfort	☐ Yes ☐ No	☐ Yes ☐ No
Other (please describe)	☐ Yes ☐ No	☐ Yes ☐ No

Collected by: _____ Research Staff Signature

USE OF FEMININE HYGIENE PRODUCTS

Do you use douche products? □ Yes □ No

If YES, how often:

☐ Daily ☐ Monthly

☐ A few times per week ☐ Every few months

☐ A few times a month

If YES, what product(s) do you use? _____

Have you douched in the past 48 hours? ☐ Yes ☐ No

Do you use feminine wipes or genital deodorant products? □ Yes □ No

If YES, how often:

☐ Daily ☐ Monthly

☐ A few times per week ☐ Every few Months

☐ A few times a month

If YES, what product(s) do you use? _____

Have you used these products in the past 48 hours? ☐ Yes ☐ No

Collected by: _____ Research Staff Signature

SAFER SEX PRACTICES

Since your last menstrual period, what method of contraception have you used?

(*Check all that apply)*

□ NA – Current partner is a female
□ NA – not sexually active
□ None / Withdrawal / Rhythm method
□ Hormonal *(if used, fill out section below with most recent product)*
□ Surgical Sterilization: □ Subject □ Partner
□ Barrier:
- □ Male Condom
- □ Female Condom
- □ Sponge
- □ Diaphragm
- □ Other, specify _____
- □ Copper IUD
- □ Spermicide
- □ Abstinence

Hormonal contraceptive use	**Current Use**	**Total # yrs Used**
Progestin only pills	□	_____
Estrogen/Progestin Combination pills	□	_____
Nuvaring vaginal ring	□	_____
Mirena IUD	□	_____
Depo Provera injection	□	_____
Ortho Evra patch	□	_____
Implanon implant	□	_____
Emergency contraceptive pill	□	_____
Other Hormonal, specify _____	□	_____

If Not Known, Record Product Name: _____

Comments:

_____
_____
_____
_____
_____

Obtained by: _____ Research Staff Signature

SEXUAL ACTIVITY

Are your sexual partners: ☐ Male, ☐ Female, or ☐ Both?
Have you had vaginal intercourse in the past 48 hours? ☐ Yes ☐ No
Number of partners you have had vaginal intercourse with in the *past year*: _____
Number of partners you have had vaginal intercourse with in the *past 2 months*: _____
If you have had vaginal intercourse with a male partner in the past 2 months, was your most recent male partner circumcised? ☐ Yes ☐ No
Do you experience any pain or discomfort during vaginal intercourse? ☐ Yes ☐ No
If YES, how often? _____ (%)

How often do you engage (receive) in oral sex?

☐ Never
☐ Daily
☐ Weekly
☐ Twice Per Month
☐ Monthly ☐ Other (Specify) _____

Have you had oral sex in the past 48 hours? ☐ Yes ☐ No

How often do you engage in anal sex?

☐ Never
☐ Daily
☐ Weekly
☐ Monthly ☐ Other (Specify) _____

Have you had anal sex in the past 48 hours? ☐ Yes ☐ No

How often do you use sex toys?

☐ Never
☐ Daily
☐ Weekly
☐ Monthly ☐ Other (Specify) _____

Have you used a sex toy in the past 48 hours? ☐ Yes ☐ No
What kind of sex toy(s) do you use: _____
Are they penetrative? ☐ Yes ☐ No
Do you and your partner(s) use the same toys? ☐ Yes ☐ No

Obtained by: _____ Research Staff Signature

HISTORY OF SUBSTANCE USE

Drug Use

Please assess current and/or historical use of the following substances.
Enter number of years used.
Amount: 1= Occasionally 2= once or twice a week 3= about once daily 4= more than once daily.

Substance	Never Used	Currently Using (within past 3 months)		Previously Used (prior to last 3 months)	
		Amount	Years (duration)	Amount	Years (duration)
Heroin – Inhaled	☐				
Heroin – IV	☐				
Heroin – other	☐				
Cocaine - inhaled	☐				
Cocaine – IV	☐				
Cocaine – other	☐				
Crack – all methods	☐				
Crystal meth – inhaled	☐				
Crystal meth – IV	☐				
Crystal meth – other	☐				
THC/Marijuana	☐				
Opiates/Opioids	☐				
Benzodiazapines	☐				
Methadone	☐				
Other, specify	☐				
Other, specify	☐				

Current Alcohol Use

☐ None ☐ Occasional drink ☐ 2-3 Drinks per week ☐ daily → ____ of drinks per day

Has subject ever had an alcohol abuse problem?

☐ No
☐ Currently
☐ Historically → _____ # of years,
If past when did she stop?__ __/__ __ __/__ __ __ __

Tobacco Use

☐ Never smoked
☐ Current Smoker → Average cigarettes per day _____ # of years _____
☐ Past Smoker
→ Average cigarettes per day _____ # of years _____
→ Quit how long ago _____ (years, months, or date)

Obtained by: _____Research Staff Signature

Appendix E

Chaperonin-60 (*cpn*60)

Primers	Primer Sequences
H729/ H730	**H729** (M13 sequencing primer underlined) 5'-CGCCAGGGTTTTCCCAGTCACGACGAIIIIGCIGGIGAYGGIACIACIAC-3' **H730** (M13 sequencing primer underlined) 5'-AGCGGATAACAATTTCACACAGGAYKIYKITCICCRAAICCIGGIGCYTT-3'
H1594/ H1595	**H1594** (M1340F sequencing primer underlined) 5'-CGCCAGGGTTTTCCCAGTCACGACGACGTCGCCGGTGACGGCACCACCAC-3' **H1595** (M1348R sequencing primer underlined) 5'-AGCGGATAACAATTTCACACAGGACGACGGTCGCCGAAGCCCGGGGCCTT-3'

Primer Set for *cpn*60

Reagents	Final concentration
10X PCR buffer	1X
$MgCl_2$	2 - 2.5 mM
Forward primer (H729 or H1594)	400 nM (20 pmol/reaction)
Reverse primer (H730 or H1595)	400 nM (20 pmol/reaction)
dNTPs	200 µM
Taq polymerase	2.5 units/reaction
Template	Variable
Water	to 50 µL

*cpn*60 reactions contained following

Step	Stage	Number of Cycles	Temperature (°C)	Time (min:sec)
Initial Denaturation	1	1	94	5:00
Amplification	2	40	94 * 72	0:30 0:30 0:45
Final Extension	3	1	72	10:00

*cpn*60 reactions performed in 0.2mL thin-wall strip tubes using the Eppendorf Mastercyler or a BioRad iCycler using the following conditions. *For H729/H730: annealing temp is 50°C, and for H1594/H1595: annealing temp is 57°C

qPCR & qRT-PCR

Escherichia coli

Primers and Probe	Primers and Probe Sequences
*uid*A forward primer: 784F	5'-GTGTGATATCTACCCGCTTCGC-3'
uidA reverse primer: 866R	5'-GAGAACGGTTTGTGGTTAATCAGGA-3'
uidA probe: EC807	5'-FAM-TCGGCATCCGGTCAGTGGCAGT-BHQ1

Primers and probe set used

Reagents	1x Reaction (µl)	Final Concentration
ABI Taqman-Universal Mastermix (2X)	10	1X
784F (10 µM)	0.8	0.4 µM
866R (10 µM)	0.8	0.4 µM
EC807 (10 µM)	0.2	0.05 µM
Template	2	-
H_2O up to	20	-

Reaction contained following. Volumes needed for a 20 µl reaction volume.

Step	Stage	Number of Cycles	Temperature (°C)	Time (min:sec)
Reverse Transcription	1	1	50	2:00
Initial Denaturation	2	1	95	10:00
Amplification	3	40	95 60	0:15 1:00

Reactions performed in MicroAmp Fast Optical 96-Well Reaction Plate using the Applied Biosystems 7500 Real-Time PCR System using the following conditions.

Human

Primers and Probe	Primers and Probe Sequences
RNase P Forward Primer	5'-AGATTTGGACCTGCGAGCG-3'
RNase P Reverse Primer	5'-GAGCGGCTGTCTCAACAAGT-3'
RNase P Probe	5'-NED-TCTGACCTGAAGGCTC-MGBNFQ-3'

Primer and probe set used.

Reagents	1X Reaction (µl)
ABI Taqman-Universal Mastermix (2X)	10
RNase P Forward	0.8
RNase P Reverse	0.8
RNase P Probe	0.2
Template	2
PCR Water	6.2

Reaction contained following. Volumes needed for reaction volume of 20 µl

Step	Stage	Number of Cycles	Temperature (°C)	Time (min:sec)
Reverse Transcription	1	1	50	2:00
Initial Denaturation	2	1	95	10:00
Amplification	3	40	95 60	0:15 1:00

Reactions performed in MicroAmp Fast Optical 96-Well Reaction Plate using the Applied Biosystems 7500 Real-Time PCR System using the following conditions.

Adenovirus

Primers	Primer Sequences
Ad2-F (Forward Primer)	5’- CCAGGACGCCTCGGAGTA-3’
Ad2-R (Reverse Primer)	5’- AAACTTGTTATTCAGGCTGAAGTACGT -3’
Ad4-F (Forward Primer)	5’-GGACAGGACGCTTCGGAGTA-3’
Ad4-R (Reverse Primer)	5’-CTTGTTCCCCAGACTGAAGTAGGT-3’

Primer Set

Reagents	1X Reaction (µl)
2x SYBR Green Master Mix	10
Ad2-F	0.25
Ad2-R	0.25
Ad4-F	0.25
Ad4-R	0.25
Template	1
PCR Water	8

Reaction contained following. Volumes needed for reaction volume of 20 µl

Step	Stage	Number of Cycles	Temperature (°C)	Time (min:sec)
Initial Denaturation	1	1	94	0:20
Amplification	2	40	94 60	0:03 0:30

Reactions performed in MicroAmp Fast Optical 96-Well Reaction Plate using the Applied Biosystems 7500 Real-Time PCR System using the following conditions.

Enterovirus

Primers	Primer Sequences
Verstrepen-F (Forward Primer)	5’-CCCTGAATGCGGCTAATCC-3’
Watzinger-R (Reverse Primer)	5’-ARATTGTCACCATAAGCAGCCA-3’

Primer Set

Reagents	Volume (µl)
Forward Primer	100
Reverse Primer	100
Probe	100
Water	93

20X Primer/Probe Mix

Reagents	1X Reaction (µl)
PCR grade water	9
TaqMan Fast Virus 1-Step Master Mix (4x)	5
20X Primer/Probe Mix	1
Template	5

Enterovirus reactions contained following

Step	Stage	Number of Cycles	Temperature (°C)	Time (min:sec)
Reverse Transcription	1	1	50	5:00
Initial Denaturation	2	1	95	0:20
Amplification	3	40	95 60	0:03 0:30

Enterovirus reactions performed in MicroAmp Fast Optical 96-Well Reaction Plate using the Applied Biosystems 7500 Real-Time PCR System using the following conditions

Appendix F

Quant-iTTM high-sensitivity DNA assay kit using a Qubit® fluorometer

(Invitrogen, Carlsbad, CA, USA)

Load Microplate with Working Solution

1. Equilibrate assay components to room temperature.

2. Make the working solution by diluting Quant-iT™ dsDNA HS reagent 1:200 in Quant-iT™ dsDNA HS buffer and mix well.

3. Load 199 µL of working solution in each Qubit® Assay tube; 190 µL for Quant-iT™ dsDNA HS standards

Add Samples and Mix Well

4. Add 10 µL of each of the Quant-iT™ dsDNA HS standards to separate tubes and mix well.

5. Add 1 µL of each unknown DNA sample to separate tubes and mix well.

Qubit® Fluorometer

6. On the Home screen of the Qubit® Fluorometer, select dsDNA High Sensitivity as the assay type. The “Read standards” screen is displayed. Press Read Standards to proceed.

7. Insert the tube containing Standard #1 into the sample chamber, close the lid, then press Read standard. When the reading is complete (~3 seconds), remove Standard #1. 8. Repeat 7 for Standard #2

9. Insert a sample tube into the sample chamber, close the lid, then press Read tube. When the reading is complete (~3 seconds), remove the sample tube. The instrument displays the results on the assay screen.

10. To calculate concentration of your original sample, choose “calculate sample concentration” and indicate the sample volume added to the assay tube (1 µL)

11. Repeat steps 9-10 until all samples have been read.

Appendix G

Round a/b Random Amplification Protocol
ROUND A

Klenow RT Denature	
Reagents	**Volume (µl)**
Sample	8.5
10mM dNTPs	4
200uM Sol primer A	0.5
Added Total (to sample)	4.5
Running Total	13
Thermocycler conditions	
Temperature	**Time**
75ºC (denature) *	5 minutes
25ºC (annealing)	Infinite
Klenow RT 1st strand	
Reagents	**Volume (µl)**
40 U/ul Rnase Inhibitor	1
0.1M DTT	1
5x FS Buffer	4
200U/ul SSIII	1
Added Total	7**
Running Total	20
Thermocycler conditions	
Temperature	**Time**
25ºC (annealing)	5 minutes
50ºC (FS Extension)	60 minutes
70ºC (heat inactivation)	15 minutes
4ºC (storage)	Infinite
Klenow SS3 Inactivation	
Reagents	**Volume (µl)**
10x NEBuffer 2	2.35
200uM Sol Primer A	0.1
Added Total	2.45
Running Total	22.45

Klenow SS3 Inactivation	
Thermocycler conditions	
Temperature	**Time**
95ºC (denature) 4ºC (storage)	2 minutes Infinite
Klenow RT 2nd Strand	
Reagents	**Volume (µl)**
5U/ul Klenow Fragment Added Total Running Total	1 1 23.45
Thermocycler conditions	
Temperature	**Time**
37ºC (2nd strand synthesis) 75ºC (heat inactivation) 4ºC (storage)	60 minutes 20 minutes Infinite

* 95ºC for DNA samples, and 75ºC for RNA samples
** skip this step for DNA samples, add 7 µl water instead, and skip Klenow RT 1st Strand thermocycler conditions

ROUND B

Reagents	Volume (µl)
Round A Template	5
Water	37.8375
10X PCR Buffer for KlenTaq LA	5
10 mM dNTP	1
Sol Primer B (100uM)	1
KlenTaq LA Polymerase Enzyme	0.1625

Round B Random Amplification Reagents

Step	Cycles	Temperature	Time
Denature	1	94ºC	4 minutes
Annealing	1	68ºC	5 minutes
Amplification	30	94ºC	30 seconds
		50ºC	1 minute
		68ºC	1 minute
Final Elongation	1	68ºC	2 minute
Hold	NA	10ºC	Infinite

Thermocycler conditions for Round B Random Amplification

Appendix H

Ampure XP Bead Cleaning

1. Bring AMPure beads to room temperature before use
2. Prepare fresh 70% Ethanol
   Note: Prepare enough for two washes = 400 µl = (X samples = 20% pipetting errors)
3. Transfer 49.5 µl* of samples to hard-shell round plate
4. Invert AMPure bead bottle 10X to resuspend beads
5. Add 89.1 µl** of AMPure beads (1.0x** sample volume) to each well, and gently pipette up and down 10x to mix
6. Incubate at room temperature for 5 minutes
7. Place plate on magnetic stand until supernatant has cleared (at least 2 minutes)
8. Aspirate supernatant and discard
9. Wash beads with 100 µl of 70% ethanol. Gently aliquot, without disturbing the pellet
10. Incubate at room temperature on magnetic stand for 30 seconds
11. Aspirate wash supernatant and discard
12. Repeat wash steps 8 to 10
13. Ensure all ethanol wash has been removed. If some liquid remains, use a new tip to remove from well
14. Allow beads to air-dry for <5 minutes on the magnetic stand. Avoid over-drying pellet
15. Remove the plate from magnetic stand and add 40 µl of water, pipette up and down 10x to mix
16. Incubate at room temperature for 2 minutes
17. Place plate on magnetic stand until supernatant has cleared (at least 2 minutes)
18. Transfer 39 µl eluant by pipette to labelled tubes/plate, careful to avoid disturbing the pellet

*in second wash use 48 µl instead

**in second wash use 48 µl instead (1.0x sample volume)

Appendix I

BPMI Digestion

Reagents	Volume(µl)
Round B Product	30
Water	7
10x NEBuffer 3.1	5
BpmI (2U/µl)	5

BpmI Digestion Reagents

Temperature	Time
37ºC	2 hours
65ºC	20 minutes
4ºC	Infinite

BpmI Digestion Thermocycler conditions.

Appendix J

NEBNext Ultra DNA Library Preparation (Illumina)

NEBNext End Prep

1. Mix the following components in a sterile nuclease-free tube:

• End Prep Enzyme Mix	3.0 µl
• End Repair Reaction Buffer (10X)	6.5 µl
• Fragmented DNA	55.5 µl
Total volume	65 µl

2. Mix by pipetting followed by a quick spin to collect all liquid from the sides of the tube.
3. Place in a thermocycler, with the heated lid on, and run the following program:

    30 minutes @ 20°C

    30 minutes @ 65°C

    Hold at 4°C

Adaptor Ligation

! If DNA input is < 100 ng, dilute the NEBNext Adaptor for Illumina (provided at 15 µM) 10-fold in 10 mM Tris-HCl or 10 mM Tris-HCl with 10 mM NaCl to a final concentration of 1.5 µM, use immediately.

1. Add the following components directly to the End Prep reaction mixture and mix well:

• Blunt/TA Ligase Master Mix	15 µl
• NEBNext Adaptor for Illumina* !	2.5 µl
• Ligation Enhancer	1 µl
Total volume	83.5 µl

2. Mix by pipetting followed by a quick spin to collect all liquid from the sides of the tube.
3. Incubate at 20°C for 15 minutes in a thermal cycler.
4. Add 3 µl of USER™ Enzyme to the ligation mixture from Step 3.
5. Mix well and incubate at 37°C for 15 minutes

Size Selection/Cleanup of Adaptor-ligated DNA

1. Vortex AMPure XP Beads to resuspend.
2. Add 13.5 µl dH_2O to the ligation reaction for a 100 µl total volume.

3. Add 37.5 µl of resuspended AMPure XP Beads to the 100 µl ligation reaction. Mix well by pipetting up and down at least 10 times.
4. Incubate for 5 minutes at room temperature.
5. Quickly spin the tube and place the tube on an appropriate magnetic stand to separate the beads from the supernatant. After the solution is clear (about 5 minutes), carefully transfer the supernatant containing your DNA to a new tube **(Caution: do not discard the supernatant)**. Discard the beads that contain the unwanted large fragments.
6. Add 20 µl resuspended AMPure XP Beads to the supernatant, mix well and incubate for 5 minutes at room temperature.
7. Quickly spin the tube and place it on an appropriate magnetic stand to separate the beads from the supernatant. After the solution is clear (about 5 minutes), carefully remove and discard the supernatant that contains unwanted DNA. Be careful not to disturb the beads that contain the desired DNA targets **(Caution: do not discard beads)**.
8. Add 200 µl of 80% freshly prepared ethanol to the tube while in the magnetic stand. Incubate at room temperature for 30 seconds, and then carefully remove and discard the supernatant.
9. Repeat Step 8 once.
10. Air dry the beads for 5 minutes while the tube is on the magnetic stand with the lid open. Caution: Do not overdry the beads. This may result in lower recovery of DNA target.
11. Remove the tube/plate from the magnet. Elute the DNA target from the beads into 17 µl of 10 mM Tris-HCl or 0.1 X TE. Mix well on a vortex mixer or by pipetting up and down. Incubate for 2 minutes at room temperature.
12. Quickly spin the tube and place it on a magnetic stand. After the solution is clear (about 5 minutes), transfer 15 µl to a new PCR tube for amplification.

PCR Enrichment of Adaptor Ligated DNA

1. Mix the following components in sterile strip tubes:

• Adaptor Ligated DNA Fragments	15 µl
• NEBNext Q5 Hot Start HiFi PCR Master Mix	25 µl
• Index Primer/i7 Primer	5 µl
• Universal PCR Primer/i5 Primer	5 µl
Total volume	50 µl

2. PCR cycling conditions:

Initial Denaturation:	30 seconds @ 98°C

Denaturation Annealing/Extension (4-12 cycles):	10 seconds @ 98°C
	75 seconds @ 65°C
Final Extension:	5 minutes @ 65°C
	Hold 4°C ∞

**NEB suggest 4 PCR cycles for 1 µg DNA input 7-8 cycles for 50 ng, and 12 for 5 ng DNA input. Further optimization of PCR cycle number may be required.*

Cleanup of PCR Amplification

1. Vortex AMPure XP Beads to resuspend.
2. Add 45 µl of resuspended AMPure XP Beads to the PCR reactions (~ 50 µl). Mix well by pipetting up and down at least 10 times.
3. Incubate for 5 minutes at room temperature.
4. Quickly spin the tube and place it on an appropriate magnetic stand to separate beads from supernatant. After the solution is clear (about 5 minutes), carefully remove and discard the supernatant. Be careful not to disturb the beads that contain DNA targets **(Caution do not discard beads)**.
5. Add 200 µl of 80% ethanol to the PCR plate while in the magnetic stand. Incubate at room temperature for 30 seconds, and then carefully remove and discard the supernatant.
6. Repeat Step 5 once.
7. Air dry the beads for 5 minutes while the PCR plate is on the magnetic stand with the lid open. Caution: Do not overdry the beads. This may result in lower recovery of DNA target.
8. Remove the tube/plate from the magnet. Elute DNA target from beads into 33 µl 0.1X TE. Mix well by pipetting up and down at least 10 times. Quickly spin the tube and incubate at room temperature for 2 minutes.
9. Place the sample on an appropriate magnetic stand to separate beads from supernatant. After the solution is clear (about 5 minutes), carefully transfer 28 µl supernatant to a new PCR tube. Libraries can be stored at –20°C.

Appendix K

Agilent Technology 2100 Bioanalyzer:

Preparing the Gel-Dye Mix

1. Allow the DNA dye concentrate and DNA gel matrix to equilibrate to room temperature for 30 minutes.
2. Vortex the blue-capped DNA dye concentrate for 10 seconds and spin down. Make sure the DMSO is completely thawed.
3. Pipette 25 µl of the blue-capped dye concentrate into a red-capped DNA gel matrix vial. Store the dye concentrate at 4 °C in the dark again.
4. Cap the tube, vortex for 10 seconds. Visually inspect proper mixing of gel and dye.
5. Transfer the gel-dye mix to the top receptacle of a spin filter.
6. Place the spin filter in a microcentrifuge and spin for 15 minutes at room temperature at 2240 g ± 20 % (for Eppendorf microcentrifuge, this corresponds to 6000 rpm).
7. Discard the filter according to good laboratory practices. Label the tube and include the date of preparation.

Loading the Gel-Dye Mix

1. Allow the gel-dye mix to equilibrate to room temperature for 30 minutes before use. Protect the gel- dye mix from light during this time.
2. Take a new DNA chip out of its sealed bag and place the chip on the chip priming station.
3. Pipette 9.0 µl of the gel- dye mix at the bottom of the well marked **G**.
4. Set the timer to 60 seconds, make sure that the plunger is positioned at 1 ml and then close the chip priming station. The lock of the latch will click when the Priming Station is closed correctly.
5. Press the plunger of the syringe down until it is held by the clip.
6. Wait for exactly 60 seconds and then release the plunger with the clip release mechanism.
7. Visually inspect that the plunger moves back at least to the 0.3 ml mark.
8. Wait for 5 seconds, then slowly pull back the plunger to the 1 ml position.
9. Open the chip priming station.
10. Pipette 9.0 µl of the gel-dye mix in each of the wells marked G.

Loading the Marker

1. Pipette 5 µl of green-capped DNA marker into the well marked with the ladder symbol and into each of the 12 sample wells.

Loading the Ladder and the Samples

1. Pipette 1 µl of the yellow-capped DNA ladder in the well marked with the ladder symbol.
2. In each of the 12 sample wells pipette 1 µl of sample (used wells) or 1 µl of deionized water (unused wells).
3. Place the chip horizontally in the adapter of the IKA vortex mixer and vortex for 60 seconds at 2400 rpm.

**Make sure that the run is started within 5 minutes.*

Inserting a Chip in the Agilent 2100 Bioanalyzer

1. Open the lid of the Agilent 2100 bioanalyzer.
2. Check that the electrode cartridge is inserted properly and the chip selector is in position (1).
3. Place the chip carefully into the receptacle. The chip fits only one way.
4. Carefully close the lid. The electrodes in the cartridge fit into the wells of the chip.
5. The 2100 expert software screen shows that you have inserted a chip and closed the lid by displaying the chip icon at the top left of Instrument context. Label samples, and click start

Appendix L

Size Selection

Agarose Gel

1. Weigh out 3 grams of UltraPure LMP Agarose
2. Mix with 150 mL of TBE Buffer (0.5x)
3. Microwave for 75 seconds or more, till the agarose powder has dissolved
4. Add 15 µl of Gel Red to mixture, and swirl
5. Microwave for 20 seconds or more, till solution turns faint pink colour
6. Let solution cool slightly before pouring on to plate (do not forget to add well comb to plate before pouring solution)
7. Allow 30 minutes for gel to polymerize

Gel Electrophoresis

1. Fill gel holder with 0.5x TBE Buffer
2. Place gel in gel holder, and remove well comb
3. Add 2 µl of O'GeneRuler 100 bp plus Ladder to single well
4. Add 6 µl of 6x Orange Loading Dye to samples, pipette up and down
5. Add samples to wells
6. Run gel at 120V for 1.5 hours*, or till samples have migrated ⅔ down the plate

*Different voltage and runtimes were tested for optimal separation of DNA fragments

DNA Excision

1. Weigh empty 2 mL tube
2. Place gel on Dark Reader blue light transilluminator (Dark Reader® Technology, Clare Chemical Research) to visualize bands
3. Excise each sample at desired bp range and place in empty tube

DNA Purification (QIAquick Gel Extraction Kit; Qiagen)

1. Weigh the gel slice in a colorless tube. Add 3 volumes Buffer QG to 1 volume gel (100 mg gel ~ 100 µl). The maximum amount of gel per spin column is 400 mg. For >2% agarose gels, add 6 volumes Buffer QG.
2. Incubate at 50°C for 10 min (or until the gel slice has completely dissolved). Vortex the tube every 2–3 min to help dissolve gel. After the gel slice has dissolved completely,

check that the color of the mixture is yellow (similar to Buffer QG without dissolved agarose). If the color of the mixture is orange or violet, add 10 µl 3 M sodium acetate, pH 5.0, and mix. The mixture turns yellow.

3. Add 1 gel volume isopropanol to the sample and mix.
4. Place a QIAquick spin column in a provided 2 ml collection tube or into a vacuum manifold. To bind DNA, apply the sample to the QIAquick column and centrifuge for 1 min or apply vacuum to the manifold until all the samples have passed through the column. Discard flow-through and place the QIAquick column back into the same tube. For sample volumes of >800 µl, load and spin/apply vacuum again.
5. If DNA will subsequently be used for sequencing, in vitro transcription, or microinjection, add 500 µl Buffer QG to the QIAquick column and centrifuge for 1 min or apply vacuum. Discard flow-through and place the QIAquick column back into the same tube.
6. To wash, add 750 µl Buffer PE to QIAquick column and centrifuge for 1 min or apply vacuum. Discard flow-through and place the QIAquick column back into the same tube. Note: If the DNA will be used for salt-sensitive applications (e.g., sequencing, blunt ended ligation), let the column stand 2–5 min after addition of Buffer PE. Centrifuge the QIAquick column in the provided 2 ml collection tube for 1 min to remove residual wash buffer.
7. Place QIAquick column into a clean 1.5 ml microcentrifuge tube.
8. To elute DNA, add 50 µl Buffer EB (10 mM Tris·Cl, pH 8.5) or water to the center of the QIAquick membrane and centrifuge the column for 1 min. For increased DNA concentration, add 30 µl Buffer EB to the center of the QIAquick membrane, let the column stand for 1 min, and then centrifuge for 1 min. After the addition of Buffer EB to the QIAquick membrane, increasing the incubation time to up to 4 min can increase the yield of purified DNA.

Appendix M

Ethanol Precipitation

Linear Acrylamide
3M NAOAC (pH 4.6)
100% Ethanol
70% Ethanol
1.5 or 2ml tubes

Materials required for ethanol precipitation.

- Add 5µl Linear Acrylamide to each sample
- Add 25 µl NaOAc to each sample (10% of final volume)
- Add 700 µl of 100% Ethanol to each sample (2.5X of final volume including regents added above)
- Homogenize sample and reagents; vortex, spin down briefly and place in box in -80ºC freezer overnight
- Prepare 70% Ethanol and place in -20ºC fridge overnight
- Next day, thaw samples
- Centrifuge for 30 mins at 17,000 g(rcf) at 4ºC
    - White precipitate should form at bottom; this is the DNA
- Discard supernatant, make sure not to disturb the pellet
- Add 1ml of 70% Ethanol to each sample, do not homogenize
- Centrifuge again with above conditions
- Remove supernatant, being very careful not to disturb the pellet
- In BSC: place diapers on bottom and rest tubes upside on racks for 5-10 mins
    - No more than 10 mins
    - If ethanol still present, pipette out
- Warm EB Buffer at 42-50ºC for 5 mins
- Resuspend samples in 25 µl of EB buffer
    - Make sure to homogenize
- Store samples at -80ºC