A study of the genital microbiotas of Black South African women and men: associations with human papillomavirus and HIV infections

---

A Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY (PhD)

by Research in

Medical Virology

Division of Medical Virology | Department of Pathology

Faculty of Health Sciences

UNIVERSITY OF CAPE TOWN

by

David Harris Onywera

Supervisors: Dr. Tracy L. Meiring and Prof. Anna-Lise Williamson

October 2018

Declaration

I, David Harris Onywera, hereby declare that this research project is a product of my original work, with no plagiarism, and has not been submitted for any diploma or degree in any other university or college. This thesis has been submitted to the *Turnitin* module (similarity and originality checking software) and I confirm that my supervisor has seen my report and any concerns revealed by such have been resolved with my supervisor.

Signature: **Date:**

Signed by candidate Sat., 20th October 2018

....

"…if I settle on the far side of the sea, even there your hand will guide me…" (Psa. 139.9-10, NIV)

Dedication

To those who have suffered, are suffering, and will suffer from infectious diseases, science will continue providing hope.

To my great friend, Bethwel Kiplagat Tanui, you didn't last forever to see this dream that we had crafted during our undergraduate life. Even though I am now done again as it was planned, my writings here do not give me the solace that I still seek. Even if there were confetti rains to brace this occasion, I wouldn't put on a smile.

They often say, "It's either you publish or you perish", yet even the most recent publications continue to rapidly perish through time – for it is undoubtedly recognised that there is meaningless in everything! As I continued to write this section, I continued to brood over, "So, is the alleged recency and relevance really worth it in the long-term?" I then remembered my MSc thesis, the toil I put in it, and yet it lost its recency the moment I started drafting it. Socrates, son of Sophroniscus, would've also upheld my argument. Throughout my life, I have realised that there is only one gem on Earth that can never lose its recency and relevance, and that is my splendid family. To my precious and supportive father, mother, siblings, nieces and nephews, you ever believed in me, even when I occasionally found my reasoning scattering and felt like I was doing nothing during my studentship. No matter the light-years separating us, you have always been there to refresh my well-being and give me a salacious life.

To my unborn family, I undertook my PhD studies in order to grant you a better life; a life that I desired during my childhood. It might have sounded stupid when I wrote this, but I knew that time would come to my rescue.

To the incomprehensible Creator of the universe, I wouldn't be who I am today without Your will and relentless love. I have faith and still believe that You kept me in sight amid the raging storms (I know), making me fearlessly take heights that many thought I wouldn't endure. Your faithfulness and promises revived and carried me during those difficult and extremely trying moments, wherein some, including the best warriors, chose to either quit or commit suicide! *You're always on my side... Thank You for being such a good Father… I know You are for me, I know You are with me... I will rest in Your promises, because all Your promises are yes and amen* (♪♫♬)…

Acknowledgements

Whether I have added new knowledge or not to the field…

I had a great passion for advancing my career, but had it not been for the South Africans who consented to participate in the HPV Couples Cohort Study, I would not have toured the Microbiome Planet.

I express my profound gratitude to Dr. Tracy L. Meiring and Prof. Anna-Lise Williamson for introducing me to the microbiome research and funding the project. Through your supervision, I wrote and customised many scripts, mostly in Python language. I commanded the friendly Python to constrict the sequenced data, which in turn confessed all the meaningful information (some beyond the scope of this thesis) that it was hiding!

A usted la Dra. Julia Ponomarenko, debo agradecer. Estoy muy agradecido por los valiosos entrenamientos y retiros bioinformática que me presentaron durante mis seis-mes de investigación en el Centro de Regulación Genómica (CRG) de Barcelona (Cataluña), una vez en España.

Moreover, I would like to acknowledge other research advisors, Drs. Zizipho Z. A. Mbulawa, Luca Cozzuto, and Ms. Sarah Bonnin. I am indebted to my external examiners (professors Janneke van de Wijgert, Jennifer E. Balkus, and Gilbert G. G. Donders) for their assessments and ensuring that this thesis is in the right shape.

A very special thanks goes out to my friends for their research ideas, moral, emotional, and spiritual support. With special mentions to Shivan Chetty, Henry Munyanduki, Alltalents Murahwa, Ongeziwe “Ongie” Taku, Lamech Mwapagha, Donald Nyangahu, Jerome Wendoh, Enock Havyarimana, Patricia Otieno-Odhiambo, Millicent Omondi, Nadia Ikumi, Brejnev Muhire, Yasheel Abeeluck, Auberi Courchay, Alín Tobares, Carlos Company, Toni Hermoso, Stephen Nzeki, Fredrick Nindo, Levi Omache, Erick Auma, and Harrison Fredrick “HF”. Seth Inzaule, I still appreciate the daily devotions and the long hours we would spend, encouraging one another, at times with refined stupidity! Rehema Mukami, we shared many things in common, and the fact that you kept on inquiring whether I was done, made me proliferate my desire to finish. Harriet Wandia, you proved to be an exceptional nurse. You nursed my nihilism, regularly hacking into my system and running the right version of outlook in me. Dr. Jospeh Anejo-Okopi, you advanced from being a stranger to a friend, then to a research colleague, mentor, pastor, and father. You have always wanted me the best.

This work was financially supported in part by the South African National Research Foundation (NRF), Poliomyelitis Research Foundation (PRF), University of Cape Town (UCT) Research Cancer Initiative, Cancer Association of South Africa (CANSA), and CRG-Novartis-African Mobility Programme. I acknowledge the UCT International and Refugee Students’ Scholarship and Postgraduate Publication Incentive (PPI) Funding that I received. The UCT’s ICTS Department is accredited for its high performance computing services: http://hpc.uct.ac.za.

I express my wholehearted appreciation to my father’s family, *Konywera*, for their encouragement throughout my years of study. To my little nieces and nephews, especially Queen and Kelly, thanks for jogging my reasoning with your boundless inquiries and stories!

Finally, I would not have weaved this thesis together, without me first being weaved together by GOD. When I was at the bottom, the only place I could look at was up – because I have a senseless reason for not banking on any kind of knowledge…

…the knowledge will ultimately greet irrelevance and be ignored.

List of Symbols, Abbreviations, and Acronyms

-	Negative
%	Percentage
~	Approximately
+	Positive
<	Less Than
>	Greater Than
≤	Less Than or Equal To
≥	Greater Than or Equal To
µl	Microliter
^{0}C	Degree Celsius
ABC	Adenosine Triphosphate-Binding Cassette
Ae	Aerobic
AIDS	Acquired Immunodeficiency Syndrome
An	Anaerobic
ASCUS	Atypical Squamous Cells of Undetermined Significance
ATP	Adenosine Triphosphate
AV	Aerobic Vaginitis
BLASTn	Nucleotide Basic Local Alignment Search Tool
BMI	Body Mass Index
bp	Base Pair
BSA	Bovine Serum Albumin
BV	Bacterial Vaginosis
BVAB	Bacterial Vaginosis-Associated Bacteria
CCR5	C-C Chemokine Receptor Type 5
CD4	Cluster of Differentiation 4
CDC	Centers for Disease Control and Prevention
CAF	Central Analytical Facilities
CI	Confidence Interval
CIN	Cervical Intraepithelial Neoplasm
***cpn*60**	*Chaperonin-60*
CST	Community State Type
CVM	Cervicovaginal Microbiota
DADA2	Divisive Amplicon Denoising Algorithm 2
DIV	Desquamative Inflammatory Vaginitis
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide Triphosphate
DP	Difference between Mean Proportions
dsDNA	Double-Stranded Deoxyribonucleic Acid
DZ	Dizygotic
EDTA	Ethylenediaminetetraacetic Acid
FAn	Facultative Anaerobic
FDR	False Discovery Rate
GBS	Group B *Streptococcus*
H_2O	Water
H_2O_2	Hydrogen Peroxide
HIV	Human Immunodeficiency Virus
HMP	Human Microbiome Project
HPV	Human Papillomavirus
hr	Hour
HREC	Human Research Ethics Committee
HR-HPV	High-Risk Human Papillomavirus (also **hrHPV**)
HRT	Hormonal Replacement Therapy

HSIL	High-Grade Squamous Intraepithelial Lesion
HSV-2	Herpes Simplex Virus Type 2
IL	Interleukin
INF	Interferon
KEGG	Kyoto Encyclopaedia of Genes and Genomes
Kg/m^2	Kilogram per Square Meter
LBG	Lactobacillary Grade
LDA	Linear Discriminant Analysis
LefSe	Linear Discriminant Analyses Effect Size
LPS	Lipopolysaccharide
LSIL	Low-Grade Squamous Intraepithelial Lesion
MAe	Microaerophilic
Mb	Megabyte
mg	Milligram
MIG	Monokine Induced by Interferon Gamma
min	Minutes
MIP	Macrophage Inflammatory Protein
ml	Millilitre
mM	Millimolar
mRNA	Messenger Ribonucleic Acid
MZ	Monozygotic
ng	Nanogram
nM	Nanomolar
OTU	Operational Taxonomic Unit
PC	Principal Coordinate
PCoA	Principal Coordinate Analysis
PCR	Polymerase Chain Reaction
PGM	Personal Genome Machine
pH	Potential of Hydrogen
PICRUSt	Phylogenetic Investigation of Communities by Reconstruction of Unobserved States
PyNAST	Python Nearest Alignment Space Termination
QIIME	Quantitative Insights into Microbial Ecology
qPCR	Quantitative Polymerase Chain Reaction
r	Pearson’s Correlation
RANTES	Regulated Upon Activation, Normal T-Cell Expressed and Secreted
RCT	Randomised Clinical Trial
RDP	Ribosomal Database Project
RR	Relative Risk
rRNA	Ribosomal Ribonucleic Acid
sec	Second
SED	Sexually Enhanced Disease
SOP	Standard Operating Procedure
sp.	Species (pl. **spp.**)
STAMP	Statistical Analyses of Metagenomic (and other) Profiles
STI	Sexually Transmitted Infection
STM	Specimen Transport Medium
Taq	*Thermus aquaticus*
TBE	Tris-Borate-EDTA
TGF-β1	Transforming Growth Factor-Beta 1
TNF	Tumour Necrosis Factor
UK	United Kingdom
UVT	Ultraviolet Transmitter
USA	United States of America (also **US**)
V	Volt

VVA	Vulvovaginal Atrophy
VVC	Vulvovaginal Candidiasis
WSW	Women Who Have Sex with Women
α	Alpha
β	Beta
γ	Gamma

List of Figures

List of Tables

List of Supplemental Tables

List of Supplemental Figures

List of Equations

Table of Contents

Abstract

Persistent genital infection with oncogenic or high-risk human papillomavirus (HPV) is causally associated with cervical cancer in women and some penile cancers in men. The role of the complex genital microbiota in HPV infection has not been extensively addressed. This study characterised the genital microbiotas of heterosexually-active Black South African women and men, predominantly of the Xhosa ethnicity, recruited from a community in Cape Town, South Africa. The association of the genital microbiotas with prevalent HPV, HIV, demographic, behavioural, and clinical characteristics of the participants was examined.

In Chapter 2 the bacterial communities in cervicovaginal samples from 62 HIV-seronegative South African women were profiled by Ion Torrent PGM sequencing of the V4 hypervariable region of the bacterial 16S rRNA gene (IT-V4 method). The cervicovaginal microbiotas (CVMs) were found to cluster into three distinct community state types (CSTs): *Lactobacillus iners*-dominated CVMs (CST I (38.7%, 24/62)), unclassified *Lactobacillus*-dominated CVMs (CST II (4.8%, 3/62)), and diverse CVMs (CST III (56.5%, 35/62)) with an array of heterogeneous bacteria, predominantly the bacterial vaginosis (BV)-associated *Gardnerella*, *Prevotella*, *Sneathia*, and *Shuttleworthia*. The majority of the women had non-*Lactobacillus*-dominated CVMs. Lactobacilli are recognised as protective against sexually transmitted infections. Among the *Lactobacillus* species detected in the women, *L. iners* was the most prevalent and abundant. This species is recognised as the least protective amongst the vaginal lactobacilli. Women in CST I were more likely to be on hormonal contraception compared to women in CST III (relative risk (RR): 2.6 [95% CI 1.3-5.3]; p=0.005). Further research is required to confirm this association and to determine the biological mechanism.

Microbiome research methodologies are constantly improving and in Chapter 3 the performance of two bacterial 16S rRNA gene amplicon-based methodologies were compared. The CVMs of 19 women were characterised using the IT-V4 method (Chapter 2) and using the Illumina 16S rRNA metagenomics method (IM-V3/V4 method). The latter method involves sequencing the V3 and V4 hypervariable regions of the 16S rRNA gene on the Illumina MiSeq platform. The two methods showed a high degree of correlation (r=0.89, p<0.0001) in the average relative abundance of shared bacterial taxa. Procrustes analyses of the weighted UniFrac distances further showed a statistically consistent clustering between the two methods (M^2=0.3, p<0.0001). The IM-V3/V4 method proved to have a greater throughput, longer read-length, and lower error rates than the IT-V4 method and was therefore used in the subsequent chapters (4 and 5).

In Chapter 4, the CVMs of 87 HIV-seronegative women from the same cohort were examined using the IM-V3/V4 method. The CVMs clustered into eight CSTs. Only 23 women

(26.4%) had CVMs dominated by a single *Lactobacillus* species, this included two women (2.3%) with *L. crispatus* (CST-1), two (2.3%) with *L. jensenii* (CST-2), and 19 (21.8%) with *L. iners* (CST-3). The majority of the women (64.4% (56/87)), however, had diverse and heterogeneous CVMs (CST-8) that were associated with BV (p<0.001). The remaining women had CVMs dominated by *Aerococcus* (CST-4, 1.1% (1/87)), *Streptococcus* (CST-5, 4.6% (4/87)), *Chlamydia* (CST-6, 2.3% (2/87)), and *Corynebacterium* (CST-7, 1.1% (1/87)). Thirty seven (42.5%) and 30 (34.5%) of the women were positive for HPV and high-risk HPV, respectively, as detected by the Roche Linear Array HPV Genotyping assay. By using linear discriminant analysis (LDA), LefSe, we noted that prevalent high-risk HPV infection was associated with higher relative abundances of *Gardnerella vaginalis, Sneathia, Atopobium vaginae, Aerococcus* and *Pseudomonas veronii* (LDA score>2, p<0.05).

In the final experimental chapter, the penile microbiotas of 238 Black South African men were characterised. This is the first large-scale study of the penile microbiota of South African men. *Corynebacteriaceae* (47.2%) and *Prevotellaceae* (6.6%) were found to be the most abundant bacterial families. The penile bacterial communities clustered into six CSTs (designated 1-6). A majority of the men (53.4% (127/238)) had *Corynebacterium*-dominated microbiotas (CST-1). The remaining CSTs (2-6) had lower relative abundances of *Corynebacterium* than CST-1 and were colonised with several vaginal bacteria. The prevalences of these CSTs (2-6) in men together with their respective most abundant genera (besides *Corynebacterium*) were as follows: CST-2 (9.2%; unclassified *Clostridiales* and *Porphyromonas*), CST-3 (8.8%; *Gardnerella*), CST-4 (7.6%; *Chryseobacterium* and *Acinetobacter*), CST-5 (18.5%; unclassified *Clostridiales* and *Porphyromonas*), and CST-6 (2.5%; *Lactobacillus*). One hundred and thirty (54.6%) and 102 (42.9%) of the men were positive for HPV and high-risk HPV, respectively, as detected by the Roche Linear Array HPV Genotyping assay. Of the 130 HPV-positive men, 37 (28.5%) and 93 (71.5%) had single and multiple HPV types, respectively. Men in CST-1 were less likely to have high-risk HPV and multiple HPV infections relative to men in CSTs 2-6 (RR: high-risk HPV: 0.8 [95% CI 0.6-1.0]; p=0.027 and multiple HPV: 0.8 [95% CI 0.6-1.0]; p=0.042). LefSe revealed that prevalent HPV infection was strongly associated with higher relative abundances of *Sneathia*, *Porphyromonas*, *Prevotella*, *Dialister*, and *Campylobacter* (LDA score >3, p<0.05). The relative abundances of the latter three bacteria together with *Peptoniphilus* were strongly associated with high-risk HPV infection (LDA score >3, p<0.05). In our cohort, 88 men (37.0%) were positive for HIV. Of these, 71.6% and 60.2% were positive for HPV and high-risk HPV infection, respectively. Among the HIV-negative men (n=150), 44.7% and 32.7% were positive for HPV and high-risk HPV infection, respectively. Although HIV status did not impact the overall composition of the penile microbiotas, HIV-infected men had higher relative abundances of *Staphylococcus*, *Faecalibacterium*, *Strenotrophominas*, *Jonquetella*, *Ruminococcus*, *Roseburia*, and *Lamia* (LDA score >2, p<0.05). Men with BV-negative female sexual partners (66.5% (157/236)) had higher relative abundances of *Lactobacillus* in their penile microbiotas than men with BV-positive female partners (p=0.007). *Atopobium*, *Sneathia*, and *Saccharofermentans* were significantly more prevalent in men with BV-positive female partners than men with BV-negative partners (p<0.020).

The main limitations of our study include relatively small sample size of women, insufficient participant information such as host genetics, other STIs (e.g., herpes simplex virus) and abnormal vaginal flora (e.g., aerobic vaginitis), using a less sensitive method to diagnose BV in women, and inherent biases evident in any retrospective study. Moreover, we did not adjust for confounding factors in our analysis due to the small sample size. Despite the underscored limitations, our findings provide insight into the baseline genital microbiotas of the Black South African women and men. The associations identified in this cross-sectional study between specific microbiota members and HPV infection, particularly the association between *Sneathia* and HPV/high-risk HPV infection, identified in both women and men, are hypothesis-generating and warrant further investigation. The study forms a critical starting point for future longitudinal confirmatory association studies and studies examining these bacteria as potential biomarkers or risk factors for HPV infection.

CHAPTER 1

Literature Review and Problem Statement

1.1 Introduction

Human papillomavirus (HPV) is a small, double-stranded DNA, nonenveloped virus that can infect mucosal squamous and cutaneous epithelia in humans [1,2]. To date, over 220 HPV genotypes have been characterised [3], with at least 40 of these infecting the mucosa of the genital tract [2]. HPVs are classified as low- or high-risk depending on their oncogenicity, with low-risk HPVs causing benign lesions and high-risk HPVs linked to premalignant and malignant lesions [1,2,4,5]. Genital HPV infection is one of the most common sexually transmitted infections (STIs) in men and women [6,7]. Sexually transmitted HPV infections are usually transient [8,9] and asymptomatic with more than 90% of the detected HPVs cleared naturally within 1-2 years [9-11]. However, some HPV infections may persist, with persistence being higher in women than men [7]. Persistent genital infection with high-risk HPV may cause cervical, vaginal and vulval cancers in women [12,13] and penile cancers in men [14,15].

The incidence of HPV and associated genital cancers is higher in less developed countries compared to developed countries [16,17]. South Africa has a high burden of HPV and HPV-related cancers – particularly cervical cancer [6,16]. HPV infection is also compounded by the high burden of human immunodeficiency virus (HIV) infection in the region [6,18], with an estimated 19% of adults aged 15-49 years being infected [18]. HIV has been associated with increased incidence and reduced clearance of HPV infection [19]. HPV prevalence in a Black South African cohort was found to be at least 1.5-fold higher among HIV-infected men (76.6%) and women (74.0%) than HIV-uninfected men (50.8%) and women (36.7%) [20]. HIV-infected individuals were also more likely to have multiple HPV infections compared to HIV-uninfected individuals [20].

Several factors may contribute to the high burden of HPV and HIV in South Africa including host behaviour [21,22], host biological/immunological factors [21-25], host genetic factors, [26,27] viral factors [28,29], and other STIs [30-32]. It is also now emerging that the genital bacterial communities or microbiome may have a potential role in infection and persistence of HPV, as well as susceptibility to HIV [33-36]. Recent studies have also revealed that ethnicity influences the microbiome and may differently affect STI burden in different population groups [37-40]. Currently, information on the cervicovaginal microbiome of South African women remains scarce and there is no published literature on the penile microbiome of South African men. Therefore, the focus of this study was the characterisation of the genital microbiotas in South African women and men with the aim of investigating associations with HPV and HIV infection in this population. This review summarises the latest research findings on the cervicovaginal and penile microbiome, highlighting the need for investigation of the genital microbiome in different ethnic groups. The review also focusses on previous studies examining associations of the genital microbiome with HPV and HIV infection.

1.2 Human microbiome

The human body is inhabited by a remarkably vast array of microorganisms [37]. These microorganisms (bacteria, archaea, viruses, and fungi) together with their genomes and host environment have collectively been referred to as the microbiome. “Microbiota”, a term coined by the late Joshua Lederberg, an American molecular biologist and Nobel laureate in Physiology/Medicine, refers to a collection microorganisms that colonise a defined habitat [41]. Such habitats in the human body with a microbiome include the gut, skin, vagina, penis, ear, oral cavity, etc. These microbiomes are site-specific and vary in diversity from one individual to another [37].

In the past, information on microorganisms colonising the human body were primarily characterised by classical approaches including microscopy and culture [42-44]. These methods are costly, laborious and time-consuming, requiring skilled personnel, and specialised nutrient requirements for some bacteria [43,44]. An estimated 99% of microorganisms discovered to date cannot yet be cultured [45,46]. The characterisation of bacterial communities has however improved significantly since the emergence of culture-independent molecular-based approaches that have helped overcome the challenges associated with classical approaches. Molecular approaches such as phylogenetic microarray and deep sequencing of amplicons rely on specific bacterial genetic markers, such as the *chaperonin-60* (*cpn*60) housekeeping gene and, the more widely adopted, 16S ribosomal RNA (rRNA) gene [47,48].

The 16S rRNA gene consists of nine hypervariable regions (designated V1-V9) that are flanked by highly conserved sequences (Fig. 1.1) that enable taxonomic profiling to be conducted [47,48]. The resolution of bacterial community composition in a specified environment is dependent on which hypervariable region is chosen [49]. The 16S rRNA sequencing has the ability to detect low-abundant species in a given environment [50] and novel bacterial taxa, thereby providing a comprehensive characterisation of the microbiome [51]. Further, the 16S rRNA gene reference databases, such as Greengenes and SILVA, are comprehensive and well-curated [52,53]. For these reasons, the 16S rRNA gene has been widely adopted, allowing cross-study comparisons [54].

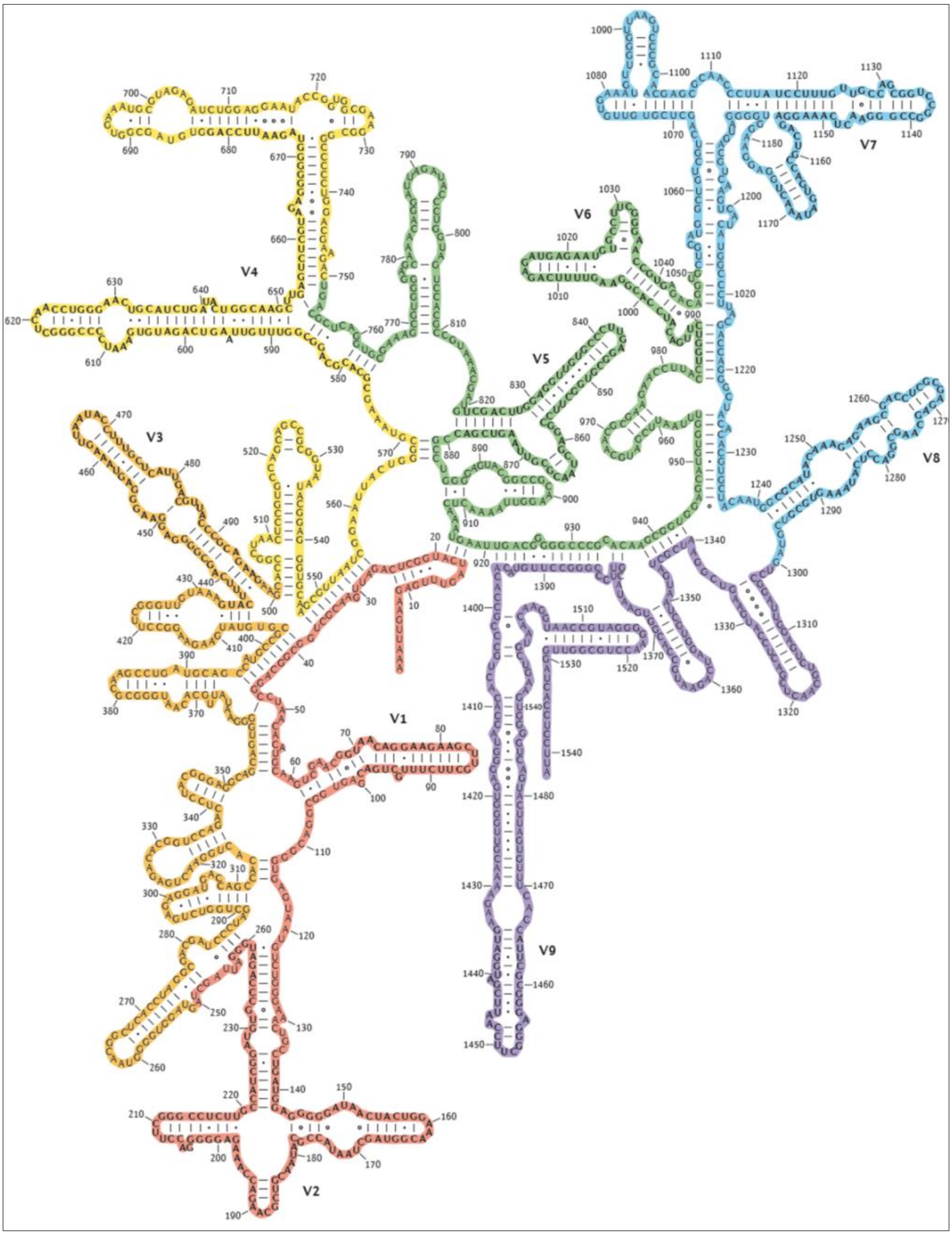


Fig. 1.1. Secondary structure of the bacterial 16S ribosomal RNA (rRNA). The figure is based on the secondary structure of the *Escherichia coli* 16S rRNA. The 16S rRNA gene is made up nine hypervariable regions (V1-V9) that form the loops of the structure separated by highly conserved sequences that form the stem structures. Figure is taken from Yarza and co-workers (2014) [47].

The *cpn*60 gene generally has a better discriminatory power and is more phylogenetically informative than 16S rRNA [55]. However, it is not widely used for a number of reasons, including the use of inosine-containing degenerate primers for amplification of the *cpn*60 gene that can affect base-pairing and introduce sequence-based primer annealing bias, under-representation of organisms with high G+C content (≥58%) (e.g., organisms in genus *Bifidobacterium*) and the need to amplify high G+C templates with a mixture of primers at a range of annealing temperatures [56].

In the context of the current study, the term "microbiota" will be used to refer to total bacteria characterised by sequencing portions of the bacterial 16S rRNA gene.

1.3 Female genital tract and its microbiome

The cervicovaginal environment is metabolically and microbiologically complex. The female genital tract is covered by epithelial cells [57-59], with columnar epithelial cell covering the endocervical mucosa and stratified squamous epithelial cells covering the ectocervix and vagina [59]. The structure of the female genital tract provides a unique microenvironment that is involved in immune defense and maintenance of cervicovaginal health by supporting the growth of endogenous *Lactobacillus* spp. [57]. Colonisation of the female genital tract with *Lactobacillus* spp. is thought to be due to the presence of glycogen that acts as the main carbon source for these communities [60-62]. It has been established that the cervicovaginal epithelia of reproductive-age women are covered by a layer of glycogen [57,58]. Glycogen is also present in vaginal secretions [61,63].

In humans, glycogen is usually synthesised from extracellular glucose through a series of coordinated enzymatic reactions, including elongation and branching of the glycogen chain as catalysed by glycogen synthase and glycogen branching enzymes [64]. Alternatively, it can be synthesised through an indirect pathway, the gluconeogenesis route, by utilising gluconeogenic precursors such as alanine [64]. There is evidence that levels of glycogen are hormonally regulated [58], with possibly oestrogen and/or progesterone driving the accumulation of glycogen on the surface of the cervicovaginal mucosa [63,65,66]. In order to be efficiently utilised as an energy source, glycogen is usually hydrolysed by α-amylase in the vaginal epithelium and secretions to maltose, maltotriose, and maltotetraose [60].

By utilising the products of hydrolysed glycogen, lactobacilli are known to produce lactic acid [60,67,68]. Lactic acid is thought to be responsible for low vaginal pH [63,68], colonisation persistence (maintenance of dominance) of *Lactobacillus* spp. [68], and to provide protection against opportunistic and pathogenic infections [65,68-71]. In addition to low pH due to lactic acid production, lactobacilli employ other mechanisms to protect against opportunistic and pathogenic bacteria. These include the production of hydrogen peroxide (H_2O_2), bacteriocins, bacteriocin-like substances and biosurfactants as well as self-aggregation and adherence to cervicovaginal receptors [69]. *In vitro* assays have suggested that the aggregation-promoting factors of the lactobacilli are protein and/or lipoprotein in nature while adherence factors are glycoproteins and carbohydrates [72]. Interference assays have shown that these factors inhibit growth of genitourinary pathogens, e.g., *Gardnerella vaginalis*, due to steric hindrance and displacement of adherent bacterial cells from vaginal cells [72].

Apart from the high abundance of *Lactobacillus* spp., other key features of the female genital tract in healthy women have long been recognised [58]. These include low vaginal pH (<4.5) [38,63,70,73] as well as a distinct lack of Gram-negative facultative and obligatory anaerobic bacteria based on classical microbiological analyses [74,75]. Much of our early understanding of the vaginal bacteria has come from studies of bacterial vaginosis (BV), the most common vaginal bacterial disorder in sexually active women [76] (refer to subsection 1.3.1). New molecular-based approaches have also recently greatly increased our understanding of the cervicovaginal microbiome (refer to subsection 1.3.3).

1.3.1 Bacterial vaginosis

BV arises from gross perturbations in the cervicovaginal milieu. It is characterised by a substantial depletion and/or displacement of *Lactobacillus* spp., which are supplanted by an overgrowth of facultative anaerobic and/or anaerobic bacteria that often include *Gardnerella*, *Prevotella*, *Mobiluncus*, *Mycoplasma*, and *Porphyromonas* [74,77,78].

BV is mainly diagnosed by the "gold standard" Nugent score [44,79,80] or the Amsel's criteria [79]. Nugent score is the most accurate method that identifies vaginal bacteria based on morphological characteristics [43,44]. It involves assessing shifts in vaginal morphotypes (Gram-positive rods/*Lactobacillus*, Gram-negative bacteria, and Gram-variable rods) under a microscope [43,81]. In the Nugent scoring system, a score of 0-3 indicates a *Lactobacillus*-dominated, vaginal environment, hence normal CVM (absence of BV). The presence of intermediate morphotypes (a mixture of *Lactobacillus*, and Gram-negative bacteria/Gram-variable rods is scored 4-6, while the absence of *Lactobacillus* and presence of Gram-negative bacterium/Gram-variable rods is assigned a score of 7-10 and is indicative of BV [43,81]. In Amsel's criteria, BV is diagnosed by the presence of at least three of the following four characteristics: 1) homogenous white-grey sticky vaginal discharge, 2) elevated vaginal pH, 3), positive Whiff test (fishy or amine odour upon additional of 10% potassium hydroxide to vaginal fluid), and 4) clue cells on wet smear [81]. BV has also been defined by a modified Amsel's criteria method that relies on the identification of any two of the four characteristics, e.g., positive Whiff test and the presence of clue cells on wet mount [81]. The presence of BV has also been defined by the presence of clue cells and >20% of clue cells on Papanicolaou (Pap) smear [81]. Other studies have also augmented BV diagnosis with quantitative real-time PCR (qPCR) that measures the load of specific bacterial species [79,82].

A systematic review of the global epidemiology of BV [76] revealed that BV is the most common vaginal disorder among reproductive-age women. Sub-Saharan Africa has the greatest burden of BV [6,76,83]. A third to two-thirds of South African women have BV [76]. A cross-sectional study of 435 adult women in Durban (South Africa) found that the prevalence of BV as assessed by Amsel's criteria was 31.0% [84]. A subsequent study on 227 HIV-negative adult women in Durban reported a higher prevalence of BV (53%) as diagnosed by Nugent score [85]. A surveillance study assessing incidental BV and other cervicovaginal infections among South African females aged 17-21 years found the overall BV prevalence on Pap smears to be 54.9% [86]. The study also reported that BV prevalence among HIV-uninfected and HIV-infected was 50.0% and 62.5%, respectively [86]. In Cape Town and Johannesburg (both in South Africa), the prevalence of BV among HIV-uninfected females aged 16-22 years as determined by Nugent score has been found to be 51.1% and 35.9%, respectively [87]. There is a probability that the prevalence of BV prevalence may be overestimated. This could be as a result of misdiagnosis of other vaginal disorders as BV sharing similar features with BV, e.g., aerobic vaginosis (AV) [88] as discussed in subsection 1.3.2. Furthermore, the prevalence of BV could be inflated as most of the studies on BV do not take in account the likelihood that some women could be having transitory dysbiosis (caused by recent unprotected sexual contact and/or menses) that will usually resolve on their own (without BV therapy). That could elucidate why placebo could achieve high cure rates (22%–25% after 1 week) as reported [89]. Therefore, the definition of BV might need i) some measure of severity and persistence and ii) elimination of misdiagnosis of other vaginal disorders related to BV.

The recommended first-line therapy for BV treatment is metronidazole (gel or oral tablets) or clindamycin cream [90]. The current antibiotic guidelines by the Centers for Disease Control and Prevention (CDC) are for treatment of women with symptomatic BV [90]. The principal goal of these guidelines is not to eliminate a known pathogen, but to relieve vaginal symptoms and signs of infections and possibly restore normal microbiota [90]. BV therapies are effective, with some tremendously achieving cure rates of 100% 1 week posttreatment [89]. In spite of the efficacy of BV therapy, some women with BV fail to respond to therapy. This could be as a result of the type of medication (clindamycin or metronidazole), mode of regimen delivery (oral or intravaginal) used [89,91], timing and duration of treatment, and type of population (high-risk or low-risk) [89]. BV recurrence rates of up to 50% 1 month posttreatment have been reported [89]. One study reported 23% and 58% recurrence rates at 1 and 12 month(s), respectively, after treatment of BV with oral metronidazole [92]. Through a study characterising the daily temporal dynamics of CVMs of 25 women with and without BV over a 10-week period [93], Ravel and colleagues (2013) showed that treatment of symptomatic BV using recommended therapy reduced the abundances of facultative and strict anaerobes and increased the abundances of *Lactobacillus*, particularly *L. iners*. They observed that the beneficial effect of BV treatment was transient as most of the bacterial communities reverted to their respective pre-treatment BV states within 2-4 weeks [93]. The recurrence of BV following treatment may be partly attributed to differences in capacity to eliminate BV-associated bacteria, specifically anaerobic Gram-negative rods by clindamycin and metronidazole. A longitudinal study examining the bacterial changes of 119 women aged 18-45 years following treatment of BV with either topical clindamycin or metronidazole observed that while both therapies resulted in a similar loss of *G. vaginalis and Mycoplasma hominis*, colonisation with *U. urealyticum* was significantly reduced with metronidazole (94% to 79%, $p<0.001$) but not clindamycin (73% to 65%, $p=0.3$) [91]. Nevertheless, clindamycin-resistant subpopulations of *Prevotella bivia* and black-pigmented *Prevotella* spp. reemerged and persisted 7 to 12 and 70 to 90 days, respectively, following therapy even among women colonised with clindamycin-susceptible strains [91]. Moreover, the current recommended therapy for BV does not prevent its recurrence since there are factors that directly impact the pathogenesis of recurrent BV. These include past history of BV and sociobehavioural factors such as sexual contact with partner having BV-associated bacteria [92].

Routine treatment of BV is not recommended in asymptomatic women [90,94]. While there may be a group of BV-asymptomatic women who may benefit from effective antibiotic therapy. A randomised, double-blind, placebo-controlled trial for BV therapy on 75 asymptomatic women found no significant differences in clinical response between treatment (metronidazole intravaginal gel) and placebo groups [94]. This could be because of the short-lived effect of BV treatment [89,92,93].

There are several risk factors for BV, including genetics [39,95], STIs [95,96], AV [97], physiological [39,93], and sociobehavioural factors [39,84,92,95,98-100] among other factors such as high glycaemic diet [101] and lack of H_2O_2-producing lactobacilli in the vagina and/or rectum [95,102]. Recent sexual behaviour has also been associated with BV. A large Australian community-based study, which sought to assess the influence of behaviour and sexual relationships on BV in 458 adult women who have sex with women (WSW) found that prevalent BV was associated with frequent cigarette use and multiple female sexual partners [98]. Of all the risk factors for BV, recent and frequent sexual contact has been identified as the strongest risk factor [39,98-100].

Two major models were initially proposed for establishment of BV: the *Lactobacillus* depletion model and the primary pathogen model (Fig. 1.2) [74]. In the *Lactobacillus* depletion model, it is suggested that a decrease in H_2O_2-producing lactobacilli creates an avenue for colonisation by facultative anaerobic bacteria responsible for BV [74]. The decrease in H_2O_2-producing lactobacilli may be due to a variety of factors. It has also been demonstrated that higher glycogen levels are associated with higher levels of protective lactobacilli and low pH [61]. Changes in glycogen and vaginal pH levels have been demonstrated to concomitantly occur with changes in hormonal levels, with high levels of progesterone causing a reduction in intracellular glycogen content through thinning of the glycogen vaginal epithelial layer [63,66]. High levels of progesterone has been associated with loss of H_2O_2-producing lactobacilli [66]. Loss of lactobacilli and presence of less acidic vaginal pH may then be followed by entry of the BV-associated bacteria [68]. *G. vaginalis*, *P. bivia*, and *Peptostreptococcus anaerobius*, which are BV-associated bacteria, have been shown to less acidify their environment (pH 4.7-6.0) compared to common vaginal *Lactobacillus* spp. (pH 3.2-4.8) [68]. Compared to healthy women, women with intermediate flora and BV have significantly reduced abundances and increased heterogeneity of *Lactobacillus* spp. [103] and, higher abundances of non-*Lactobacillus* species such as *Prevotella* spp. [82].

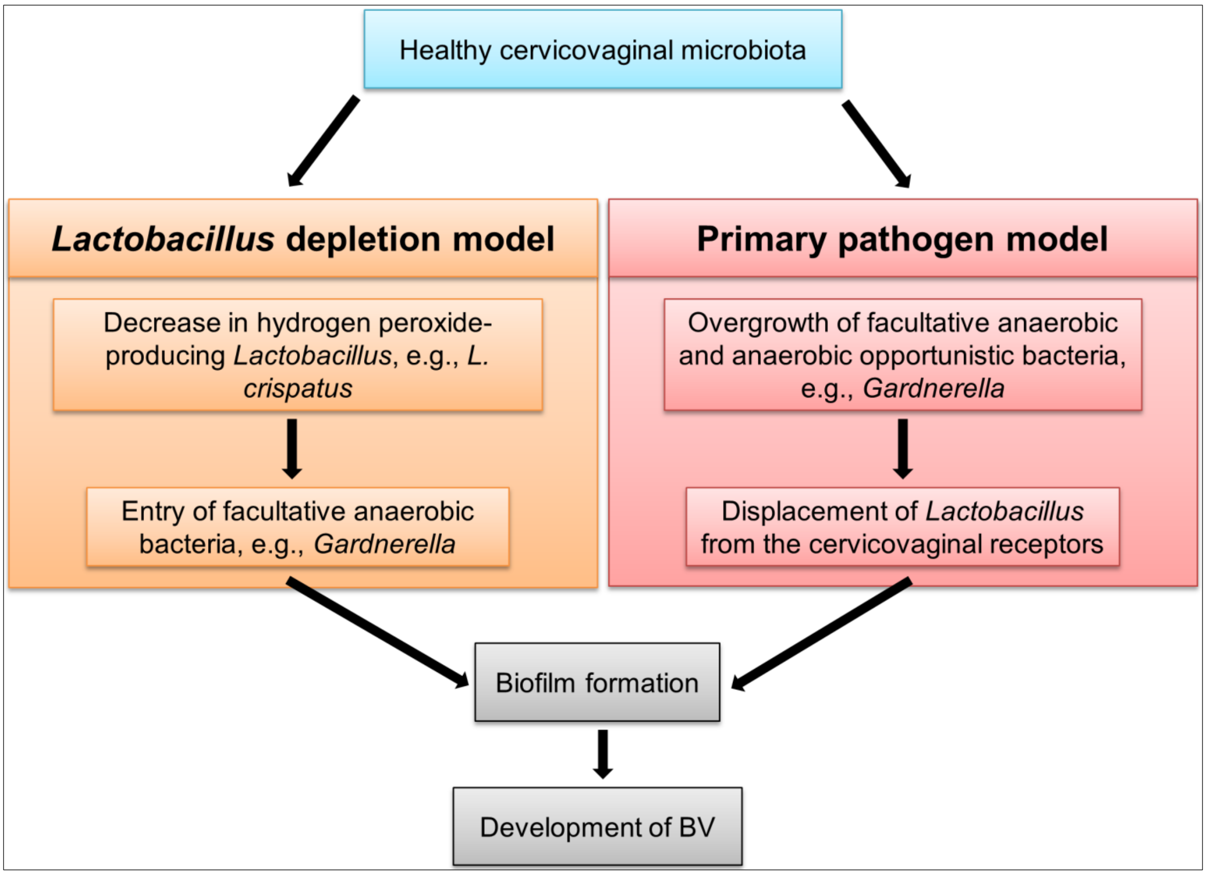


Fig. 1.2. Proposed models for establishment of BV. Under *Lactobacillus*-depletion model, BV ensues when a decrease of H_2O_2-producing *Lactobacillus* provide entry of BV-associated bacteria, e.g., *G. vaginalis*. The primary pathogen model proposes that perturbations in the cervicovaginal microenvironment causes an overgrowth of virulent BV-associated bacteria that displace *Lactobacillus* attached to the cervicovaginal receptors. In these models BV can then develop through biofilm formation. The figure is adapted from Srinivasan and Fredricks (2008) [74].

The second model is called primary pathogen model and assumes that there is an overgrowth of facultative and anaerobic bacteria that subsequently displace the lactobacilli from the cervicovaginal epithelia [74]. The mechanism of displacement may be through competition with, inhibition of and displacement of resident lactobacilli by BV-associated bacteria such as *G. vaginalis* [74,77,104-106].

These models for establishment of BV could be attributed mostly to sexual transmission of BV-associated bacteria [107-110], or the transfer of BV-associated bacteria from extravaginal reservoirs [109], or absence of H_2O_2-producing bacteria in the vagina and/or rectum [102]. BV-associated bacteria, predominantly strains of *G. vaginalis*, have been shown to be uncommon among sexually inexperienced women and colonise the genital tract of sexually active women following penile-vaginal sexual intercourse [107,108]. Increased episodes of receptive vulvovaginal, oral and anal sex have been found to increase the risk for BV acquisition [109]. Additionally, colonisation of extravaginal reservoirs (oral cavity and anus) of BV-negative women with BV-associated bacteria such as *G. vaginalis* and *Leptotrichia*/*Sneathia* spp. has been associated with BV acquisition [109]. The mechanistic principles behind each of the models for BV development are depicted in Fig. 1.2 [111].

BV, a polymicrobial infection, then develops through the formation of a vaginal biofilm [112]. A biofilm is an aggregate of microorganisms that is attached to a surface and embedded in a protective matrix of extracellular polymeric substances that these microorganism produce [113]. A biofilm often provides a habitat for other microorganisms to occupy and survive in, by increasing their social cooperation as well as their tolerance and resistance to antimicrobial compounds [113]. Hence, mucosal biofilm may be responsible for recurrent BV. In the female genital tract, *G. vaginalis* has been demonstrated to be the key player in biofilm formation and establishment of BV [114-117]. There are other bacteria found in women with BV (e.g., *Aerococcus christensenii* and *A. vaginalis*) that can form biofilms [114,116]. Detectability of biofilm has been associated with high bacterial loads of *G. vaginalis* and *Atopobium vaginae* [116]. *G. vaginalis* is known to adhere to the vaginal epithelia and form a scaffold for other bacterial species, such as *A. vaginae*, to adhere to [112,116,117]. Some of the BV-associated bacteria enhance the growth of *G. vaginalis* [117]. *G. vaginalis* has been shown to outcompete other BV-associated bacteria and some members of lactobacilli for adhering to the cervicovaginal epithelium [77,104,114]. Moreover, *G. vaginalis* isolates have been identified to vary in virulence, with cohesive form being highly prevalent, sexually transmitted, and responsible for biofilm formation and pathogenesis of BV when compared with the dispersed (planktonic) form [112,115,118]. An *in vitro* biofilm model [104] confirmed that *G. vaginalis* isolates from women with BV were more virulent than those from women without BV, and were able to displace *L. crispatus* pre-adhered to a monolayer of human cervical HeLa cells. The authors of this *in vitro* study [104] reasoned that colonisation of the vaginal epithelia by particular strains of *G. vaginalis* is essential for BV development. In addition, when the virulent *G. vaginalis* displaces the protective lactobacilli from the vaginal epithelial cells, it starts forming a biofilm that initiates further colonisation by other BV-associated bacteria [104], including *P. bivia* and *Fusobacterium nucleatum* [117]. This hypothesis supports the primary pathogen model of BV development [74]. Nelson and colleagues (2015) [119] suggested that the biogenic amines (e.g., putrescine, cadaverine and tyramine) produced by BV-associated bacteria reduce the competitive advantage of *Lactobacillus*, thereby increasing vaginal pH and enhancing predisposition to BV. A temporal study examining the changes of vaginal bacterial communities before, during, and after episodes of BV noted that prior to symptomatic BV, the genital tract is characterised mainly by the presence of strict and facultative anaerobes and elevated vaginal pH (>4.5) [93].

Emerging data suggests that the establishment of BV could be more complex than previously thought [74]. van de Wijgert (2017) [120] recapitulates that there are overlapping interrelationships between BV, vulvovaginal candidiasis (VVC, a yeast infection), STIs, and vaginal pathobiont carriage (vaginal dysbiosis with high relative abundance of bacteria with intermediate pathogenicity index, e.g., streptococci, staphylococci, or *Enterobacteriaceae*). Many of these interrelationships are bidirectional and dictated by behavioural and biological factors, including sexual transmission [120]. Therefore, BV could be caused by some of these conditions such as STIs.

Bacterial changes from normal (*Lactobacillus* predominance) to BV have been correlated with disrupted metabolic pathways [121]. A metabolomics study assessing the metabolites associated with BV identified 173 (62%) biochemicals whose levels were substantially different in women with BV [121]. Of these metabolites, 55 were present at higher levels in women with BV, e.g., inflammation-associated eicosanoid 12-hydroxyeicosatetraenoate. The rest of the metabolites (n=118), e.g., amino acids and dipeptides, occurred at lower levels [121]. BV-associated bacteria have been associated with metabolic signatures such as cadaverine, putrescine, succinate, tyramine, and deoxycarnitine that contribute to elevated vaginal pH, amine odour, vaginal discharge, and presence of clue cells [119,121]. Apart from the metabolic changes, BV and BV-associated bacteria can also alter host protein composition and levels in the cervicovaginal fluid [122,123] and interfere with the cervicovaginal epithelial architecture [122]. Such proteins, for example neutrophil elastase, neutrophil defensin-1, and kaliocin-1, either modulate host immune responses or defence against pathogens [123]. Furthermore, BV causes reversible immunological alterations that affect cervicovaginal inflammatory cytokines and immune cells, e.g., increased levels of $CD4^+$ T-cells [124], interleukin 1-beta (IL-1β) [85,103], IL-8, IL-18, and RANTES (regulated upon activation normal T-cell expressed and secreted) [103] and reduced levels of interferon gamma-induced protein 10 (IP-10) [85] and neutrophils [124]. Treatment of BV with oral metronidazole has been found to decrease levels of IL-1β, IL-8, RANTES, and $CD4^+$ T-cells [125].

With the unclear aetiology of BV and recommendation not to treat asymptomatic BV, the adverse consequences of BV will remain to be imminent. Numerous studies have implicated BV in negative reproductive health outcomes including acquisition of STIs such as HIV [126] and HPV that causes cervical cancer [12], as will be outlined (subsections 1.3.5 and 1.3.6).

1.3.2 Aerobic vaginitis

Over the last decades it became clear that alterations in the cervicovaginal milieu are not always an indication of BV or presence of communities that predispose woman to BV, but an indication of aerobic vaginitis (AV). AV is a state of abnormal CVM often mistaken for BV that was first described by Donders and colleagues in 2002 [127]. It is usually diagnosed using fresh wet mount microscopy, preferably using phase contrast technique (at 400x magnification) [75,88,127,128]. Diagnostic tests that are currently emerging are nucleic acid-based [88] and enzymatic tests [88,129]. These tests are often used in combination with the wet mount microscopy [88,129]. When using the wet mount microscopy, the AV score is usually computed using the presence of lactobacilli (designated “lactobacillary grade”, LBG: I-III), presence of inflammation, proportion of toxic leucocytes, characteristics of the microflora, and presence of immature epithelial cells [88,127]. LGB I has predominantly lactobacilli (hence, normal microflora), LGB II is devoid of lactobacillary morphotypes (with LGB IIa and LGB IIb being measure of fairly and abnormally disturbed cervicovaginal milieus, respectively), LGB III lacks lactobacilli and the cervicovaginal milieu is colonised with several coccoid aerobic bacteria that include *Escherichia coli*, *Staphylococcus aureus*, *Streptococcus agalactiae* (group B Streptococcus, GBS), and *Enterococcus faecalis* [88,127]. Interpretation of the AV score is as follows: 0-2 (no signs of AV, therefore AV is absent), 3-4 (mild AV), 5-6 (moderate AV), and 6-10 (severe AV) [88]. The severe and chronic form of AV is called desquamative inflammatory vaginitis (DIV) [88,130,131] and has been reported exclusively in White women [131]. By definition, AV score of 8-10 corresponds to DIV [75]. Clinical features of AV include vaginal epithelial inflammation, high vaginal pH (>4.5; at times as high as 6-8 [130]), parabasal cells (small round epithelial cells with a high nuclear to cytoplasmic ratio), yellowish to greenish vaginal discharge (that is thick and mucoid), and greater imbalance in local immune modulation compared to BV (increased levels of IL-1β, IL-6, and IL-8) [88,127].

Studies have reported that the prevalence of AV ranges 2-51% (non-pregnant women: 2-51% and pregnant women: 3-11%) [83,88,97,129,132,133]. AV is often detected in sexually active women [129], including African women. For example, a study that used wet mount microscopy to diagnose AV among 338 women aged 18-50 years in Kampala (Uganda) found that the prevalence of AV was 11% [83]. AV has also been found to co-occur with BV, candidiasis, and *Trichomonas vaginalis* [83,88,97,129,130]. It remains unknown if AV can transition to BV, and vice versa [127]. Due to its poor appreciation, AV was often undiagnosed in routine gynaecological examination, even in symptomatic women [88,129]. Furthermore, at times it was misdiagnosed as BV due to its shared characteristics with BV [88]. It has been acknowledged that AV was rarely diagnosed and at times misdiagnosed. Thus, these might have resulted in underestimation of the prevalence of AV and imprecise treatment of vaginal disorders [88,128].

Although not well and fully characterised, some of the risk factors for development of AV frequent douching, use of intrauterine devices, long-term use of antibiotics [88], and menopausal stage [132] to mention a few.

Investigations are now revealing that AV is implicated in adverse outcomes, such as pelvic inflammatory disease (PID), infertility, preterm birth [128], chorioamnionitis [75], foetal infections [88,128], cervical dysplasia, dyspareunia (painful sexual intercourse), STIs such as HPV, HIV, *Chlamydia trachomatis*, and *T. vaginalis* [88].

The clinical management of AV is not yet well established since like BV, its etiopathogenesis remains unclear [88,130]. Thus, treatment of AV largely depends on microscopic findings [88,130,134]. For example, hydrocortisone is used when there is an increased inflammation, oestrogen therapy in case of low maturation indexes (deficient vaginal epithelium maturation), and antimicrobials (e.g., kanamycin) when there is predominance of coccoid bacteria [88,130,134]. Antibiotics, mainly consisting of suppositories of aminoglycosides, are effective against AV and have negligible adverse effects on the women treated [133]. Antibiotics, including those for treatment of BV have varying effect on AV [128,130,131,134]. Metronidazole, for instance, is not effective against AV since it mostly targets anaerobic bacteria (mostly observed in women with BV), yet bacteria responsible for AV are aerobic enteric bacteria [130]. In contrast, clindamycin has a broad spectrum activity and therefore be effective against AV [130,131]. However, some women with chronic AV (DIV) have been reported to experience recurrent AV following treatment with topical 2% clindamycin [131]. It has been suggested that long-term therapy may be required to treat certain types of treat AV [131]. At the moment, additional research is still necessary to define the etiopathogenesis, diagnostic criteria, and optimal therapy guidelines for AV [128].

1.3.3 Cervicovaginal microbiome and community state types of healthy women

An extensive study by the Human Microbiome Project (HMP) Consortium on the ecology of human-associated microbiome has shown that the female genital tract is the least diverse microbiome habitat in the human body [37], often colonised by a single *Lactobacillus* sp. [38].

One of the first large-scale studies of the CVM using high-throughput sequencing [38] established that the CVMs of reproductive-age women can be grouped into five community state types (CSTs), even though the type and precise number of clinically/biologically relevant CSTs is still a matter of debate. The study examined the CVMs of 396 reproductive-age asymptomatic North American women of White, Black, Hispanic, and Asian ethnicity. Four of the five CSTs that were established were dominated by a single *Lactobacillus* species, namely *L. crispatus* (CST I), *L. gasseri* (CST II), *L. iners* (CST III), and *L. jensenii* (CST V), and were found in 72.6% of the women. The remaining CST (CST IV) was a diverse group with diminished abundances of lactobacilli and high abundances of strictly anaerobic bacteria, including *Sneathia*, *Atopobium*, *Gardnerella*, and *Prevotella* [38]. A subsequent retrospective case-control study further sub-divided CST IV into two groups, CST IV-A and CST IV-B [135]. CST IV-A is defined by an even composition of *Peptoniphilus*, *Corynebacterium*, *Finegoldia*, *Anaerococcus*, and *Prevotella* while CST IV-B is characterised mostly by high abundances of *Atopobium* and low abundances such as *Prevotella*, *Sneathia*, *Gardnerella*, *Parvimonas*, and *Mobiluncus* [135]. More recent studies [33,34,39,40,135-144] have confirmed these findings. The microbiological, genetic and/or biochemical, characteristics of these CSTs have been underlined in the next subsections (1.3.3.1 and 1.3.3.2). For ease of reference, the CST numbering system used by Ravel and colleagues (2011) [38] is used in the following subsections of the literature review. Here, for example, CST I refers to any CSTs dominated by *L. crispatus*. It is important to note however that CST I may be used by other authors to represent a different CST not dominated by *L. crispatus*. Several limitations to the use of CSTs include: i) they are a derived classification that focusses on the most dominant taxon or lack thereof, (ii) minority species are often not taken into account during the classification, and (iii) the abundance of many taxa is a continuous gradient and the cut-off points for separation or clustering into the various CSTs are not well defined and differ between studies. The limitations of a CST based approach are well recognised in the gut microbiome field [145].

1.3.3.1 Lactobacillus-dominated cervicovaginal microbiomes: CSTs I-III and V

The distinct compositions of the CSTs and their associated properties are shown in Fig. 1.3. CSTs dominated by *L. crispatus*, *L. gasseri*, and *L. jensenii* (Groups I, II and V, respectively Fig. 1.3) are regarded as healthy states and associated with low to relatively low vaginal pH (4.0-4.7) [38]. Vaginal microbiota dominated by *L. iners* (Group III, Fig. 1.3) are, however, seen as either healthy or transitional between healthy and BV states. *L. iners*-dominated communities tend to occur in acidic vaginal pH (pH 4.4±0.6) [38].

The *Lactobacillus* spp. can produce lactic acid [146,147] and H_2O_2 [102,148-150]. Lactic acid exists as either D- or L-isoform, with D-lactic acid having more potent antiviral activity than L-lactic acid [146]. *Lactobacillus*-dominated CVMs produce slightly higher L-lactic acid than non-*Lactobacillus*-dominated CVMs [147]. *L. crispatus*-dominated CVMs produce higher levels of both isoforms than other communities, including *L. gasseri*, *L. iners*, and *L. jensenii* [146,147]. Moreover, they produce higher levels of D- than L-lactic acid [146,147]. *L. gasseri*-dominated CVMs produce high amounts of both isoforms, with L- being higher than D-lactic acid [147]. *L. iners*-dominated CVMs produce very low levels of D-lactic acid and high levels of L-lactic acid [146,147]. The converse is true for *L. jensenii*-dominated CVMs [147]. D-lactic acid is thought to have more beneficial effect in protecting the genital tract against infections than L-lactic acid [147].

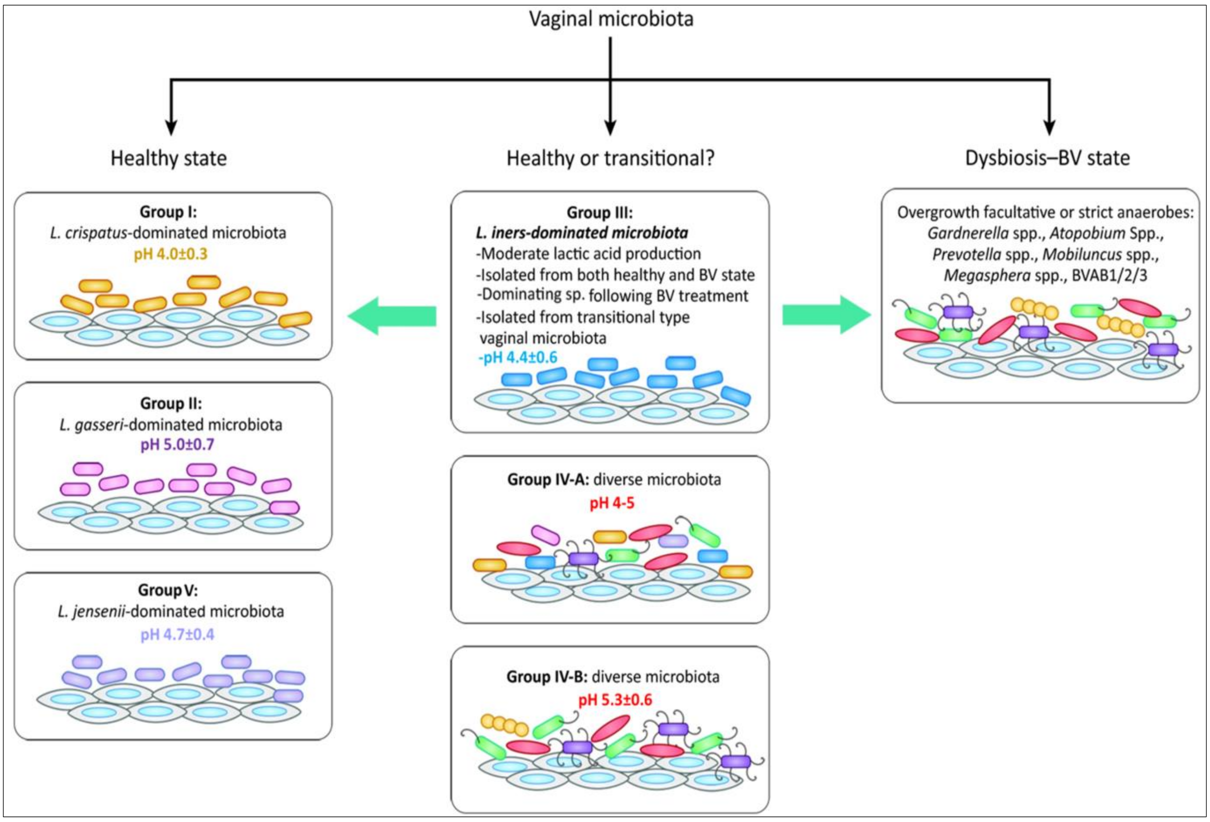


Fig. 1.3. Microbiological characteristics of vaginal bacterial communities. Community state types (CSTs) I-III and V are mainly dominated by *Lactobacillus* species and are regarded as biomarkers for vaginal health. However, CST III may also be a transitioning microbiota between healthy and dysbiotic states. CST IV-A has low abundance of *Lactobacillus* and is evenly colonised with strict anaerobic bacteria. CST IV-B has high abundance of *Atopobium* and low abundances such as *Prevotella*, *Sneathia*, *Gardnerella*, *Parvimonas*, and *Mobiluncus*. The figure is adapted from Petrova and colleagues (2017) [151].

Additionally, *Lactobacillus* species vary in their ability to produce H_2O_2. *L. iners* and *L. gasseri* are typically very low H_2O_2-producing *Lactobacillus* species compared to *L. crispatus* and *L. jensenii* [102,148,150]. In the microaerobic (hypoxic) cervicovaginal environment, H_2O_2-producing vaginal *Lactobacillus* spp. produce little or no H_2O_2 [152]. Although there is no *in vivo* evidence demonstrating that H_2O_2 produced by vaginal *Lactobacillus* spp. have antimicrobial properties at physiological conditions or in the hypoxic cervicovaginal environment [71,152,153], studies that have incorporated *in vitro* studies indicate that H_2O_2 is protective against BV [95,102,148]. This protective role could be credited to other factors that favour H_2O_2 production [152,153]. Alternatively, H_2O_2-producing vaginal *Lactobacillus* spp. could be more susceptible to inhibitory factors expressed by BV-associated bacteria, which could explain their inverse association with BV [152]. It is plausible that *Lactobacillus* such as *L. iners* with low to very low production of D-lactic acid and H_2O_2 may not be as beneficial in maintenance of a healthy CVM and protection against vaginal disorders and STIs.

It has been reported that *L. iners* is genetically and biochemically distinct from other common *Lactobacillus* species [151]. Among the lactobacilli, *L. iners* has the smallest genome (1.3 Mbp) and can encode specialised proteins e.g., iron-sulfur protein involved in resistance to oxidative stress, that enables it to adapt dysbiotic microenvironments [154]. It is able to differentially express over 10% of its genome in BV versus healthy state [155]. These include elevated expression of cholesterol-dependent cytolysin, mucin and glycerol transport and related metabolic enzymes such as glycerol-3-phosphate dehydrogenase and butyrate kinase among others when in the BV state [151,155].

Additional studies suggest that the CVMs may exhibit other differences. For example, Mirmonsef and colleagues (2014) [61] found that high glycogen levels in cervicovaginal fluid were associated with low vaginal pH (<4.5) and higher relative abundance of *L. crispatus* and *L. jensenii*, but not *L. iners* [61]. Generally low vaginal pH has been associated with lactobacilli dominance, including *L. iners*, and the opposite is also true [38,139]. Women with low vaginal pH have higher abundances of *Lactobacillus* than women with high vaginal pH [61,62]. The lack of association between glycogen levels and relative abundance of *L. iners* in the study by Mirmonsef and colleagues (2014) [61] is not surprising since *L. iners* can utilise non-glycogen carbon sources and possibly regulate vaginal pH by other metabolites besides lactate. Under dysbiotic conditions, *L. iners* can utilise glycerol, producing succinate, butyrate, and acetate; which elevate the vaginal pH thereby favouring vaginal colonisation by opportunistic and pathogenic bacteria [155]. Proteome analysis has also revealed proteins unique to specific *Lactobacillus* spp. that could be involved in their optimal survival in eubiotic and dysbiotic vaginal environments [156]. Some of these proteins, e.g., *L. iners* glycolytic enzymes glucose-6-phosphate isomerase and glyceraldehyde-3-phosphate dehydrogenase have been negatively associated with dysbiosis independent of vaginal pH and relative abundance of *L. iners* [156]. Such outlined genetic and biochemical differences make each CST functionally distinct and associated with different vaginal pH.

The CSTs, including the lactobacilli-dominated CSTs, are known to vary in their protection against colonisation by pathogens and cervical disease [33,34,140,157-159] due to their different ecological stability and resilience to perturbations [38]. Among the lactobacilli, *L. crispatus* and *L. iners* appear to be the most stable and least protective, respectively [33,42,82,140,146,149,159-161]. Furthermore, compared to other common *Lactobacillus*-dominated CVMs (CST I and CST II), *L .iners*-dominated CVMs are also more likely to transition to high Nugent score-associated *Lactobacillus*-deficient CVMs, specifically CST IV-B [162]. A cross-sectional study in a Rwandan cohort found that women with *L. crispatus*-dominated CVMs have lower prevalence of STIs, including HIV and high-risk HPV, compared to women with *L. iners*-dominated CVMs [33]. Cervical mucus with *L. crispatus*-dominated CVMs and high concentrations of D-lactic acid have been demonstrated to trap HIV-1 virions, while mucus with *L. iners* dominance and low D-lactic is considerably permissive to rapid diffusion of HIV-1 virions [146]. Furthermore, prospective observational studies have shown that *L. iners* and/or *L. gasseri*, unlike *L. crispatus*, may predispose women to abnormal CVMs [163], HIV [164], and *C. trachomatis* infection [159,165]. *In vitro* experiments have shown that strains of *L. crispatus* from healthy premenopausal women have considerably better inhibitory effect on *C. trachomatis* elementary bodies than strains of *L. gasseri* and *L. vaginalis* [158].

1.3.3.2 Non-Lactobacillus-dominated cervicovaginal microbiomes: CST IV and dysbiotic CVMs

Culture-independent studies have found that 20-33% of healthy reproductive-age women lack appreciable numbers of *Lactobacillus* spp. [38,40,144,166]. Ravel and colleagues (2011) found that such women (27.4%) had diverse CVMs (CST IV) colonised with a mixture of strict anaerobes [38] such as *Gardnerella*, *Sneathia*, *Streptococcus*, *Atopobium*, *Prevotella*, *Aerococcus*, *Leptotrichia*, *Escherichia*, *Megasphaera* and BV-associated bacteria (BVAB)-1/2/3 [38,40,82,166-168]. Most of these bacteria have been positively correlated with high Nugent scores, an indication of BV [38]. These diverse CVMs were more common in asymptomatic Black and Hispanic women compared to White and Asian women [38].

Diverse CVMs may be seen as healthy or transitional, in the absence of symptomatic BV. A 16-week longitudinal study describing the temporal dynamics of the five CSTs (I, II, III, IV-A, and IV-B) in 32 reproductive-age women [162] observed that most transitions between CSTs occurred between CST III (*L. iners*-dominated CVMs) and CST IV (diverse communities). Some diverse communities, such as Group IV-A (Fig. 1.3), are associated with low vaginal pH (4.0-5.0) [151]. Other CSTs, such as Group IV-B have been found to occur in high vaginal pH environment [151]. Diverse CSTs such as Group IV-A are able to maintain lower vaginal pH due to functional redundancy that compensates for the depletion of lactobacilli [162]. *Streptococcus*, *Aerococcus*, and *Atopobium* can produce lactic acid [162,169], which results in low pH [63,68] thereby precluding colonisation by opportunistic and pathogenic bacteria [38,162,167,169]. Thus, the observations that women without BV can be colonised with non-*Lactobacillus* spp. [82,170,171], challenge the common understanding of a healthy *Lactobacillus*-dominated CVM. To some extent, it can be acknowledged that these non-*Lactobacillus* bacterial communities may be less harmful than currently documented and constitute healthy CVMs in BV-asymptomatic women. However, this view should be considered with caution as there have been reports of healthy women non-*Lactobacillus*-dominated CVMs having asymptomatic BV [166]. Cervicovaginal milieus with CST IV-B have a high pH due to production of succinate, acetate [162], and biogenic compounds that have a less acidic pH [119]. The elevated pH has been suggested to be due to lower bacterial loads of lactic acid bacteria, including non-*Lactobacillus* species [169]. CST IV, especially CST IV-B, has been associated with BV [38,141,162,172].

Diagnosed BV (Fig. 1.3, Dysbiosis-BV state) is characterised by an overgrowth of facultative or strict anaerobes and high pH, as previously discussed (subsection 1.3.1). Many of the bacteria in diverse CVMs have been associated with BV [51,82,173,174], thus, suggesting that CST IV may be a transitional state to BV.

1.3.3.3 Uncommonly characterised cervicovaginal community state types

Less commonly reported CSTs have been defined in both healthy and diseased women. These include those dominated by an uncultivated BV-associated bacterium-1 (BVAB-1) [39,40], *Streptococcus* [40,142,143,147], *Gardnerella* [39,142,144,147,166,175], those with *L. crispatus*/*L. jensenii* codominance [40], *Sneathia* [142], *Bifidobacteriaceae*, *Mobiluncus curtisii*, *Staphylococcus*, *Lactobacillus delbrueckii*, *E. faecalis*, *E. coli* [143], *A. vaginae* [176], and *Bifidobacterium breve* [107,143] among others. *Gardnerella* and *Streptococcus* spp., for example, have been identified in sexually-inexperienced women, at times dominating their CVMs [169]. This suggests that *Gardnerella* and other non-*Lactobacillus* bacteria may also be part of a healthy CVM. On the other hand, CVMs dominated by *Fusobacterium* spp. has been associated with cervical cancer [142]. In spite of these evidences, the temporal stability and role of such non-*Lactobacillus*-dominated CSTs in genital infections and genital symptoms and cervicovaginal infections still remain underappreciated.

1.3.4 Factors affecting cervicovaginal microbiome composition

The composition of cervicovaginal microbiome is known to be influenced by multiple intrinsic and extrinsic factors, ranging from host genetics to sexual behaviours. These factors are described in detail in the subsequent subsections (1.3.4.1 and 1.3.4.2)

1.3.4.1 Ethnicity/race and cervicovaginal microbiome composition

Several studies have demonstrated that CVMs differ by ethnicity/race [38,40,95,107,123,143,144,168,177]. Compared with Caucasian women, up to about two-thirds of reproductive-age HIV-negative Black women have been observed to have *Lactobacillus*-deficient CVMs that have evidence of dysbiosis as associated with BV [33,38-40,136,138,177]. In a cross-sectional study profiling the CVMs of 396 asymptomatic reproductive-age North America women from four different ethnicities [38], it was observed that a greater majority of White (90.0%) and Asian (80.2%) women had *Lactobacillus*-dominated CVMs compared to Black (59.6%) and Hispanic (61.9%) women (Fig. 1.4) [38].

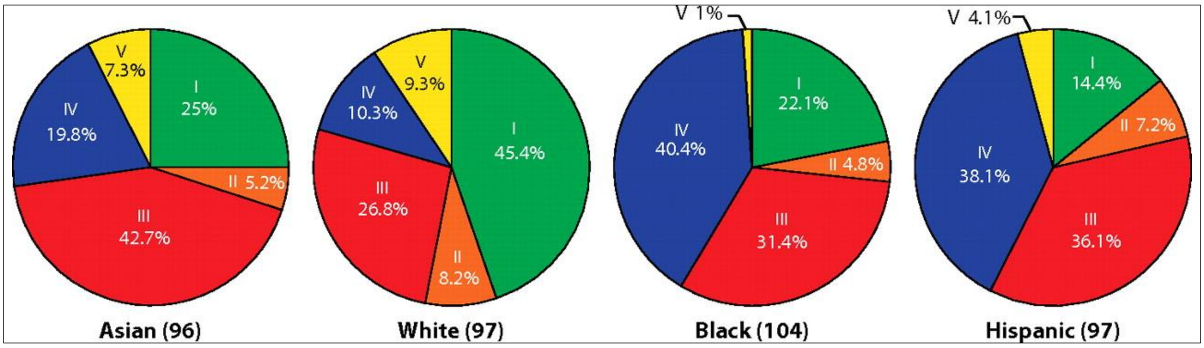


Fig. 1.4. Distribution of vaginal bacterial communities of reproductive-age women according to ethnicity. The number of women in each ethnic group (Asian, White, Black, and Hispanic) is shown in the parenthesis. The five community state types (CSTs) are numbered I-V. CST I, II, III, and V are dominated by *L. crispatus*, *L. gasseri*, *L. iners*, and *L. jensenii*, respectively. CST IV has lower abundances of lactobacilli and higher abundances of strictly anaerobic bacteria. The figure is taken from Ravel and co-workers (2011) [38].

L. crispatus-dominated CVMs (CST I) were significantly more prevalent in White women (41.9%) than in other women (Asian: 22.9%, Black: 21.9%, and Hispanic: 13.3%) [38]. *L. jensenii*-dominated CVMs (CST V) were the least frequent *Lactobacillus*-dominated CVMs and were very uncommon among Black women [38]. *L. iners*-dominated CVMs (CST III) were almost equally distributed among the four ethnic groups; White (19.3%), Asian (30.4%), Hispanic (25.9%), and Black (21.9%) [38]. The distribution of the women with diverse non-*Lactobacillus*-dominated communities CVMs (CST IV, n=108) was different among the ethnic groups. These communities were more frequent in Black (38.9%) and Hispanic (34.3%) women than White (9.3%) and Asian (17.6%) women [38]. Vaginal pH was also found to vary by ethnicity; with Black (pH 4.7±1.04) and Hispanic (pH 5.0±0.59) women having higher vaginal pH relative to White (pH 4.2±0.30) and Asian (pH 4.4±0.59) women [38]. This corroborated a previous study in North America [73] that detailed that Black women had a significantly more alkaline vaginal pH (5.3±0.7) compared to non-Black women (4.7±0.6) mostly of White and Hispanic races. Another study from the same region reported that the vaginal pH of African American women was about 0.2 higher than that of Mexican American women [177]. Thus, vaginal pH is greatly impacted by ethnicity/race [73], which is associated with the genital tract bacteria [177]. It has been identified that diverse non-*Lactobacillus*-dominated CVMs are common in cervicovaginal microenvironment that are less acidic (pH 4.7-6.2) [38,70,177], such as those of women of African and Hispanic heritage [38].

A preceding study by Zhou and colleagues (2007) [40] observed the incidence of non-*Lactobacillus*-dominated CVMs was 4.7 times higher among healthy Black women (33%) compared with Caucasian women (7%). The CVMs of Black women were mostly dominated by strict anaerobes, including *Atopobium* and phylotypes in the order *Clostridiales* [40]. The authors speculated that the different CVMs may have different community resilience; with non-*Lactobacillus*-dominated CVMs that are common among Black women, exhibiting the lowest resilience [40]. They further postulated that this could explain why Black women are more susceptible to BV and STIs than Caucasian women [40]. This reasoning can be extended to account for the high burden of STIs among African women [6], including those with asymptomatic non-ulcerative STIs and BV. It has been demonstrated that Black women are more likely to acquire BV than White, Hispanic, Asian, Native American, and multi-ethnic women [95]. Newton and colleagues (2001) [177] strongly associated African American race with incident STIs and BV compared with Mexican American race. A large community-based study performed on African women in rural Uganda [178] reported a very high prevalence of ulcerative STIs and BV, with majority being asymptomatic. For example, of the 6,356 adult women that were enrolled, 50.6% and 23.8% had BV and trichomonas, respectively [178]. About 81.1% of the women with BV (n=2,462) and 80.1% of the women trichomonas (n=1,472) were asymptomatic [178].

Several other studies [107,143,168,170,177] have confirmed that ethnicity/race influences the composition of CVMs, with White and Asian women more likely have CVMs dominated by a single or multiple lactobacilli, especially *L. crispatus* than Black women. A study describing the CVMs of 416 European and 1,268 African American women of reproductive-age group strongly correlated African American ethnicity with higher relative abundance of BV-associated bacteria [39]. Healthy women of the European ancestry had higher relative abundances of *Lactobacillus* spp. while African American women had heterogeneous non-*Lactobacillus*-dominated communities [39].

Studies solely conducted on African cohorts have reported high prevalence of diverse and heterogeneous non-*Lactobacillus*-dominated CVMs [33,87,136-138]. A study conducted among 146 young Black Durban (South Africa) HIV-negative women aged 18-23 years found that 63% of the women had non-*Lactobacillus*-dominated CVMs [136]. Nonetheless, *G. vaginalis*-dominated CVMs accounted for 45% of the two non-*Lactobacillus*-dominated CSTs that were identified [136]. The other non-*Lactobacillus*-dominated CST consisted of CVMs heterogeneously dominated by *Gardnerella*, *Prevotella*, *Megasphaera Shuttleworthia*, and *Sneathia* [136]. The prevalences of the two *Lactobacillus*-dominated CSTs detected were as follows: *L. iners* (29%) and *Lactobacillus* other than *L. iners* (most likely *L. crispatus*) (9%) [136]. A recent study conducted among 168 young women aged 16-22 from Cape Town and Johannesburg (South Africa) [87] found the prevalence of non-*Lactobacillus*-dominated CVMs to be 47% [87]. These CVMs were strong predictors of genital inflammation [87]. The prevalence of *L. crispatus*-dominated CVM (22%) was 2.5-fold higher than previously reported among young Black women in South Africa [136]. The prevalence of *L. iners*-dominated CVM (32%) was similar to published report on young South African cohort [136].

The literatures underscored herein regarding ethnic/racial variations in CVMs highlight that CVMs are strongly affected by host genetics [179,180]. A Korean cohort study comparing the CVMs of monozygotic (MZ) twins and their families (mothers and sisters) and unrelated individuals [179] showed that the CVMs of MZ twins had greater resemblance to each other than to those of unrelated individuals. The CVMs were greatly similar between MZ twins (13 pairs), followed by MZ twins versus their sisters (n=8), then MZ twins versus their mothers (n=8), and finally between the MZ twins and unrelated individuals [179]. In a more recent study investigating the heritability of CVMs and its relationship with obesity among 542 Korean females, including 222 MZ and 56 dizygotic (DZ) twins together with their families and non-relatives [180], it was identified that the majority of the heritable genital tract bacteria had significantly greater positive correlations in MZ twins than in DZ twins. In this Korean twin-family cohort, *Prevotella* had the highest heritable component (72.2%) whereas *Bifidobacterium* and *Planococcaceae* had the weakest heritability [180]. *Lactobacillus* was found to have a high heritability, at 41.2% for *L. iners* and 36.9% for *L. crispatus* [180]. Single nucleotide polymorphism association tests to associate specific human genes with vaginal bacteria in this cohort, identified the rs2069812 allele in the IL-5 gene as being associated with higher abundances of cervicovaginal *Prevotella melaninogenica* [180].

1.3.4.2 Other factors affecting cervicovaginal microbiome composition

Besides ethnicity/race, other factors affecting cervicovaginal composition include physiological [39,162,179], environmental [74,162], sociodemographic, and behavioural factors [39,74,98,99,181,182].

In terms of physiological changes, longitudinal studies on reproductive-age women have shown that CVMs are hormonally regulated and dynamic during life stages [58,169], and menses [93,162]. The human female genital tract continues to undergo significant changes as the woman ages. These include changes in vaginal cytology, pH, and vaginal secretions that affect the CVMs [58]. Fig. 1.5 summarises the biological changes in the female genital tract from pre-puberty stage to postmenopausal stage.

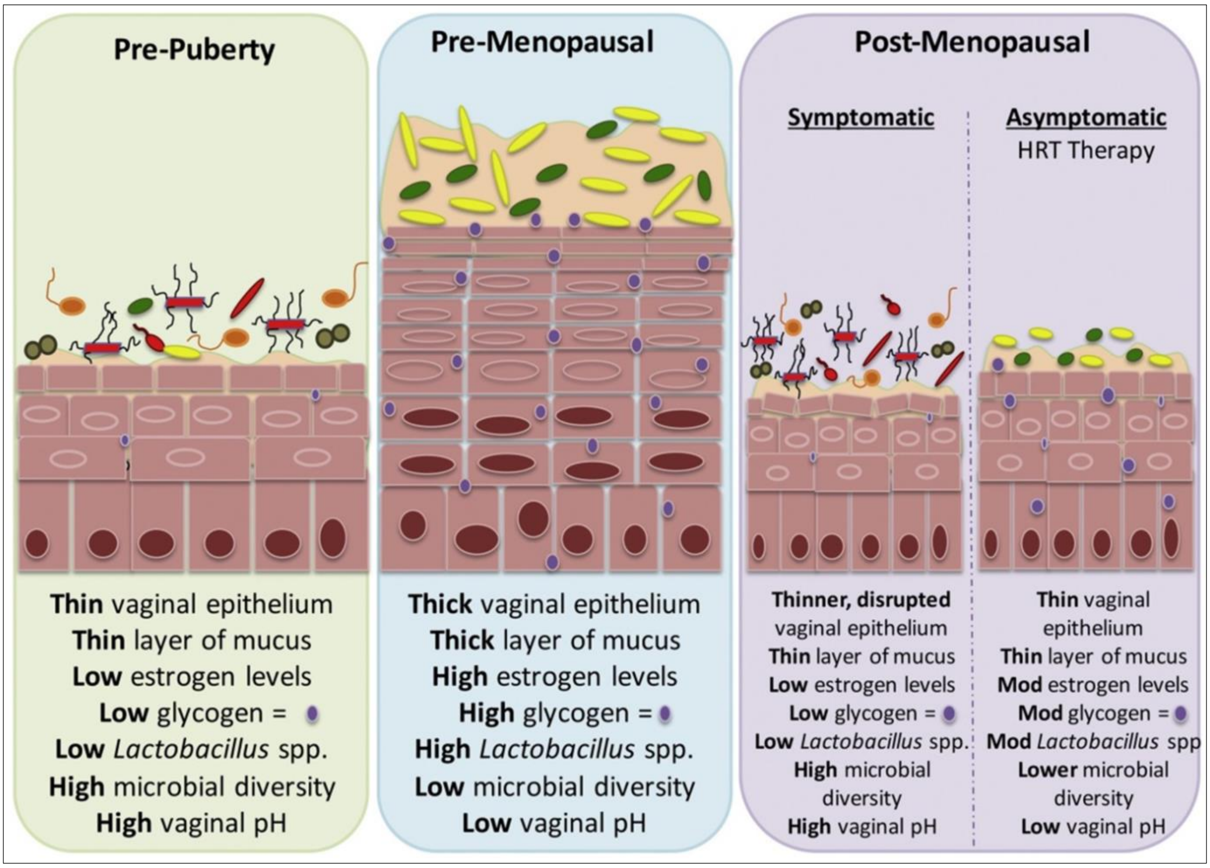


Fig. 1.5. Oestrogen-driven changes in vaginal epithelia and microbiota. In pre-pubertal stage, the low oestrogen level is associated with vaginal thinning, low glycogen content, low abundance of lactobacilli, and high vaginal pH (left panel). In premenopausal women, high oestrogen levels promote vaginal epithelial thickening, availability of glycogen, colonisation with *Lactobacillus* spp., and low vaginal pH (middle panel). As reproductive-age women enter menopausal stage, the oestrogen levels drop, which eventually results in loss of *Lactobacillus* and increased microbial diversity (right panel). These changes in microbiotas have been associated with vaginal symptoms such as vaginal dryness, atrophy, and dyspareunia in symptomatic women (left side of right-hand panel, postmenopausal). However, the CVMs of asymptomatic women are likely to be colonised by *Lactobacillus* spp. (right side of right-hand panel, postmenopausal). Systematic or topical hormonal therapy replacement (HRT) can be used to restore dominance of *Lactobacillus* spp. by moderately increasing oestrogen and glycogen levels. The figure is taken from Muhleisen and Herbst-Kralovetz (2016) [183].

Prior to puberty, oestrogen levels are low, (left panel, Fig. 1.5) and the female vaginal mucosa consists of a thin, stratified squamous epithelia (anucleated on the apex) covered with a thin glycogen layer. Usually, there is low abundance of *Lactobacillus* spp., hence high pH and microbial diversity. In reproductive-age women (middle panel, Fig. 1.5), oestrogen levels increase and the vaginal epithelium becomes thicker with a thicker layer of mucus. This thickening promotes the availability of glycogen necessary for *Lactobacillus* spp. colonisation, hence low vaginal pH. There are significantly greater abundances of *Lactobacillus* in premenopausal women than in pre-puberty and postmenopausal women. Most women experience menses between early adolescence and late premenopausal period [58,93,169,173].

There is also a strong relationship between the host's menopausal status and the composition of the CVMs [183-185]. This is because as women enter menopausal stage, the oestrogen levels reduce [58,183]. This is accompanied by thinning of vaginal epithelial and mucus layer, akin to that of females in puberty stage, and diminished glycogen content [183]; hence, loss of *Lactobacillus* and increased vaginal pH and bacterial diversity (Fig. 1.5, right panel) [58,179,184,185], but largely depends of vaginal conditions such as atrophies and dryness [58,184,186]. There are also greater abundances of *Streptococcus*, *Finegoldia*, *Gardnerella*, and *Prevotella*, among other non-*Lactobacillus* bacteria in postmenopausal women [179,180,187]. In a study examining the association between the CVM, menopausal status, and signs of vulvovaginal atrophy (VVA) in 87 women aged 35-60 years [184], CVMs with low abundances *Lactobacillus* and higher proportions of anaerobes (CST IV-A) were associated with VVA. VVA has two main effects: 1) downregulating the genes such as suprabasin and desmoglein 1 that maintain the epithelial and barrier integrity, and 2) upregulating the genes associated with inflammation e.g., complement factor B and H [186]. These changes alter *Lactobacillus*-dominated CVMs [186], hence the low predominance of *Lactobacillus* as menopause advances.

Following hormonal replacement therapy (HRT), the abundances of *Lactobacillus* greatly increased (up to 96.8%) in postmenopausal women [180]. This increase is due an increase in the glycogen content as a function of oestrogen [183]. Preliminary results by Lee and co-workers (2013) [179] suggested that the CVMs of postmenopausal women receiving HRT resemble those of premenopausal women. A review by Muhleisen and Herbst-Kralovetz [183] underscored that oestrogen-based HRT restores *Lactobacillus*-dominated CVMs in postmenopausal women and may resolve vaginal symptoms.

A longitudinal study describing the changes in the composition of CVMs in 32 reproductive-age North American women over a 16-week period showed that the stability and resilience of CVMs vary during menses, with *Lactobacillus*-dominated CVMs being more stable and resilient than diverse and heterogeneous CVMs [162]. Each CST may transiently transition to another CST but revert to initial CST at the end of menses [162]. Furthermore, the study observed that in one subject, *L. crispatus*-dominated CVM (CST I) changed to *L. iners*-dominated CVM (CST III) during menses, but reverted to CST I after menses [162].

Despite the decreased and increased abundance of *L. crispatus* and *L. iners*, respectively, the metabolome of the bacterial community in this subject remained constant during menses, with high levels of lactate being maintained during even the CST [162]. Menses may transiently lead to an elevated pH, hence development of BV [93]. Irregular vaginal bleeding has been associated with non-*Lactobacillus* bacteria, specifically *E. coli* [177].

It has also been shown that CVMs are affected by the pregnancy status [135,188]. Pregnant women have a more stable CVM with higher relative abundances of *Lactobacillus* spp., including *L. vaginalis*, compared to non-pregnant women [135,188]. Stability of the CVMs in pregnant women is perhaps due to the absence of menses, hence shifts in hormonal-associated changes in CVMs as observed in non-pregnant women [93,162,169]. A retrospective case-control prospective study on non-pregnant and pregnant women [135] further observed that communities with lower relative abundances of *Lactobacillus* (CST IV-A and CST IV-B) were less frequent in women with normal pregnancy. Pregnant women have a low prevalence of bacteria associated with BV and preterm delivery [39,135]. Increased dominance and diversity of *Lactobacillus* spp., but not *L. iners*-dominance, has been associated with term delivery [189]. The composition of CVMs during normal pregnancy is dictated by the gestational age. As pregnancy progresses, there is a concomitant increase in the relative abundances of *Lactobacillus* spp. and a decrease in relative abundances of anaerobic and strict anaerobic bacteria, namely *Gardnerella*, *Atopobium*, *Sneathia*, *Dialister*, *Ureaplasma*, *Parvimonas*, *Eggerthella*, *Gemella*, and BVAB-1 [190].

Diet may also affect the CVM. A large cohort study assessing the association of vaginal bacteria and specific dietary indices of 1,735 reproductive-age North American women [101], predominantly Black (86%), observed that high glycaemic load was associated with prevalence, progression, and persistence of BV. In a study examining the relationship between obesity and cervical *Lactobacillus* composition [187], *L. iners*-dominated bacterial communities were associated with obesity (BMI ≥ 25 kg/m^2), particularly among young women of reproductive-age without cervical dysplasia.

In terms of sociobehavioural factors, cigarette use, alcohol consumption, new and multiple sexual partners, lack of or use of specific hormonal contraceptives, and inconsistent condom use have been correlated with BV-associated bacteria and BV [39,98-100,191,192]. There is evidence that genital bacteria may be heterosexually transmitted [107,110,193,194] and that the male urogenital tract and semen microbiome may shape CVMs [105,106], as will be detailed in section 1.5. Women who engage in heterosexual sexual contact have been found to be more likely to have CVMs dominated by *L. iners* and *G. vaginalis* than other *Lactobacillus*-deficient CVMs [107]. Moreover, some vaginal products are known to have detrimental effects on CVMs. An *in vitro* study showed that Vagisil, a feminine moisturiser, inhibited the growth of *L. crispatus* at 2 hours and killed all of this bacterial isolate at 24 hours [182]. Hormonal contraception has been observed to have protective effect against BV-associated bacteria and BV [92,132,192,195,196]. For example, combined oral contraception (progestin and oestrogen combination) has been associated with increased H_2O_2-producing lactobacilli dominance and reduced BV-associated bacteria [191] whereas intrauterine devices have been found to have negative or no association with BV or BV-associated bacteria [191,192]. Unlike oestrogen-based contraceptives, progestin-based contraceptives are may lead to loss of H_2O_2-producing lactobacilli [191,192] by causing a hypoestrogenic effect in the vagina epithelia, thus thinning of the glycogen vaginal epithelial layer [66].

There are other factors, e.g., exogenous aetiological agents of STI, which influence the resident bacterial communities in the genital tract [177]. Preceding herpes simplex virus type 2 (HSV-2) infection has been independently associated with establishment of BV [95].

1.3.5 Role of cervicovaginal microbiome in health and disease

The resident cervicovaginal bacterial communities are responsible for maintaining a stable environment which could impact on woman's reproductive health. These bacteria are thought to form the first line of defence against infections [78,197]. Thus, imbalances of these communities may result in reproductive disease in both pregnant and non-pregnant women.

1.3.5.1 Impact on infant delivery

Pregnant women with CVMs that have reduced dominance and diversity of protective *Lactobacillus* spp. and increased dominance of *L. iners* may be at risk of preterm delivery [189]. A study associating early changes in CVMs and pregnancy outcome [198] observed that women reporting a prior preterm delivery and high concentrations of BV-associated bacteria, namely *Sneathia*, *Mobiluncus*, and BVAB-1 prior to 16 weeks gestation had greater odds of having a spontaneous preterm delivery. The study also noted that this adverse obstetrical outcome occurred in women with a previous preterm delivery and increasing concentrations of *Sneathia*, *Megasphaera*, and BVAB-1 through 22-24 weeks of gestation [198]. BV and most of these BV-associated bacteria have been associated with spontaneous abortion [199,200]. Abundances of *Lactobacillus* spp. are often high in pregnant women [135], especially those with term delivery [189]. Additionally, women with term delivery often have decreased abundances of anaerobes and/or strictly anaerobes such as *G. vaginalis* and *A. vaginae* [190]. Compared to Caucasian women, Black women are more likely to be colonised by BV-associated bacteria that have been associated with preterm delivery [39]. Thus, the proportion of adverse reproductive and gestational outcomes may be higher in Black women than in non-Black women, e.g., women of Hispanic and mixed race [201].

1.3.5.2 Impact on neonates

The maternal CVM may also be beneficial to neonates. The gut microbiome of neonates is initially acquired at birth and resembles those of the maternal genital tract [150,202]. A study utilising a DNA hybridisation technique observed that *L. gasseri* and *L. crispatus* were predominant in neonates' intestines and in their mother's vaginas [150]. The mode (vaginal or caesarean-section) and age (preterm or term/full term) of delivery are some of the factors that affect post-natal development of the infants' gastrointestinal microbiome. One study observed that the body sites, including the intestines (examined using meconium), of vaginally delivered neonates were colonised with their maternal vaginal *Lactobacillus* whereas the body sites of caesarean-section delivered neonates were colonised with maternal skin bacteria, comprising *Corynebacterium*, *Staphylococcus*, and *Propionibacterium* spp. [202]. A literature review on intestinal bacterial colonisation in preterm infants [203] outlined that intestines of preterm infants had abnormal bacterial colonisation. The preterm infants had low numbers of beneficial bacteria (lactobacilli and bifidobacteria), and high numbers of potentially pathogenic (enterococci and *E. coli*) and pathogenic (*Staphylococcus and Klebsiella*) bacteria [203]. There is an intricate interplay between the indigenous gut microbiome and host immune system, which is crucial in regulation of gastrointestinal homeostasis [204]. An altered relationship between the gut microbiome (colonisation with enteropathogenic bacteria) and immunity has been associated with diseases such as infectious colitis and inflammatory bowel disease [204]. Thus, the initial colonisation of the neonate's gut with maternal vaginal bacteria may contribute to gastrointestinal health and disease.

1.3.5.3 Impact on pelvic inflammatory disease

Dysbiotic CVMs may have a role in pelvic inflammatory disease (PID) [205]. A review examining the association between BV and PID based on 19 studies observed a relationship between BV and salpingitis, cervicitis, and endometritis. Although BV has found to be common in women with PID, its causality in PID has not been proven. Nevertheless, studies independently associating BV-associated bacteria and PID are scarce. Even with the limited information, particular BV-associated bacteria such as *Gardnerella*, *Ureaplasma*, and *Mobiluncus* have been associated with PID [205].

1.3.5.4 Impact on reproductive fertility

It may be acknowledged that the composition and diversity of CVM has an impact on reproductive fertility. In a recent study comparing the cervicovaginal bacterial pattern of women with infectious fertility (STI-related, n=26) versus women with non-infectious fertility (n=26), healthy (fertile) women (n=89), and female sex workers (n=54) [175], notable variations in CVMs were found. In particular, the relative abundance of *Gardnerella* was higher in female sex workers (10.8%), women with infectious (10.1%), and non-infectious (5.6%) infertility, than in fertile women (5.4%) [175]. Besides, the CVMs of women with infectious infertility had a higher microbial diversity and were also colonised with non-*Lactobacillus* bacteria, including BV associated bacteria such as *Prevotella*, *Atopobium* and *Sneathia*, and BVAB-1/2/3 compared to the CVMs of fertile women [175]. Furthermore, there was a higher prevalence of *Gardnerella*- and *Lactobacillus*-dominated CVMs in women with infectious infertility and healthy women, correspondingly [175].

The importance of *Lactobacillus* in reproductive fertility has been detailed in additional studies [185,206,207]. A study examining the prevalence and quantity of vaginal *Lactobacillus* spp. in 92 fertile and 22 postmenopausal healthy women observed that the number and diversity of *Lactobacillus* spp. was greater in fertile women than in postmenopausal women [185]. Only 3% of the fertile women and 91% of the postmenopausal women were colonised by a single *Lactobacillus* sp. [185]. A majority of the fertile women (97%) were colonised by at least two *Lactobacillus* species: 15% with 2, 70% with 3-4, 11% with 5, and 1% with 6 detectable *Lactobacillus* spp. [185]. Only 9% of the postmenopausal women were colonised by 2 *Lactobacillus* spp. [185]. Fertile women had higher abundances and prevalences of *Lactobacillus* spp. in comparison with postmenopausal women [185]. *L. iners* (82.6%), *L. crispatus* (70.7%), and *L. gasseri* (67.4%) were the most prevalent *Lactobacillus* spp. in fertile women whereas *L. gasseri* (22.7%), *L. jensenii* (22.7%), and *L. crispatus* (18.2%) were the most prevalent in postmenopausal women [185]. The study also observed an association between *Lactobacillus* species, particularly *L. iners* as well as *L. gasseri*, with other *Lactobacillus* spp. in fertile women [185]. This association was not computed in postmenopausal women because most of them were colonised by a single *Lactobacillus* sp. [185]. A longitudinal study using a murine model to investigate the effect of probiotic *Lactobacillus* (*L. plantarum* 2621) on infertility due to sperm-agglutinating *E. coli* [206] suggests that displacement of sperm-agglutinating *E. coli* by lactobacilli may restore fertility, as manifested successful fertilisation (upon mating), consistent pregnancy-related changes, and delivery of pups at the end of gestation period. *In vitro* studies have shown that vaginal probiotic with multiple lactobacilli (*L. plantarum*, *L. brevis*, and *L. salivarius*) inhibits sperm lipid peroxidation induced by reactive oxygen species, thereby maintaining sperm motility and viability [207].

Dysbiotic CVMs, manifested by Nugent scores of between 4-10, have also been found in a large proportion of infertile women (39%) [200]. A longitudinal study on infertile Danes women undergoing *in vitro* fertilisation treatment [208] noted that the rate of clinical pregnancy (as measured by ultrasound of foetal heartbeat at 7 weeks of gestation) was 3.5-

fold lower in women with abnormal vaginal flora (as depicted by elevated concentrations of *G. vaginalis* and/or *A. vaginae*) (9.1%) than in women with normal vaginal flora (32.1%). In a study of the urogenital microbiotas of infertile couples, Borovkova and colleageus (2011) [105] found that the prevalence of *Ureaplasma parvum* was higher in women whose male sexual partner had inflammatory prostatitis (80% versus 50%). In addition, their study observed that after sexual intercourse, shifts in CVMs (acquisition of bacteria such as streptococci, *Prevotella*, *Clostridium*, and *E. coli*) were more common in women with dysbiotic CVMs (≥4 Nugent score) and those with male partners having inflammatory prostatitis [105]. Borovkova and colleagues (2011) [105] hypothesised that these changes could affect fertilisation.

1.3.5.5 Impact on sexually transmitted infections

There is substantial evidence that CVMs can affect the natural history of STIs such as HIV [33,36,126,161,164,197,209], HPV [33,34,81,139,179,209,210], cytomegalovirus, gonorrhoea [209], trichomoniasis, chlamydia [84,172,209], and HSV-2 [33,209] among others [33]. A survey associating CVMs with STIs among high-risk women in Rwanda observed that women with *Lactobacillus*-deficient CVMs had a higher prevalence of HSV-2 (82.9-87.5%) than *Lactobacillus*-dominated CVMs [33]. A recent nested case-control study found that women with *L. iners*-dominated CVMs had increased risk of *C. trachomatis* infection [159]. The study also found that high abundances of *L. crispatus* in the CVM was associated with lower susceptibility to *C. trachomatis* infection [159]. Incident trichomoniasis and chlamydia have been associated with preceding BV status [84]. The impact of CVMs on HIV and HPV will be outlined separately in the next subsection (1.3.6), as this was the topic of this study.

1.3.6 Cervicovaginal microbiome and sexually transmitted infections

1.3.6.1 Cervicovaginal microbiome and HIV infection

Most studies agree that altered CVMs increase the risk of HIV acquisition [33,126,146,164,197,209] and transmission [146,211]. A meta-analysis based on 23 articles with a total of 30,739 women observed that BV was associated with 60% increased risk of HIV acquisition in HIV incidence studies [126]. Specific BV-associated bacteria such as *A. vaginae*, *P. bivia*, *P. melaninogenica*, *Veillonella montpellierensis*, *Mycoplasma*, *Aerococcus*, *Megasphaera*, *Parvimonas*, and *Sneathia*/*Leptotrichia* have been associated with increased risk for HIV [36,164]. Altered CVMs with high abundances of *G. vaginalis* has been associated with antepartum HIV-1 transmission [211]. The protective role of BV-negative microbiomes against HIV is likely due to lactobacilli. The mechanism of protection by *Lactobacillus* against HIV is postulated to be mainly due to acidification of the vaginal pH, production of lactic acid and adsorptive properties of lactobacilli [212].

HIV acquisition in women with altered and *Lactobacillus*-deficient CVMs is also enhanced by genital inflammation [24,25,136,164]. Elevated multiple proinflammatory responses have been associated with disrupted CVMs and HIV-1 replication [161]. An *in vitro* study testing the effect of cervicovaginal inflammation on HIV infectivity using cell-free HIV isolates and preinfection cervicovaginal lavage samples from 27 Black South African women aged 18-40 years [25] found that preinfection genital inflammation was associated with acquisition of low-infectivity HIV variants. High levels of genital inflammatory chemotactic cytokines e.g., IL-8, IP-10, macrophage inflammatory protein-1-alpha (MIP-1α), and MIP-1β have been shown to contribute to HIV acquisition [24]. Persistent genital inflammation, which may be due to asymptomatic STIs, increases the risk of HIV infection [24]. Proinflammatory cytokine IL-1β, for example, has been found to occur at higher concentrations in women with asymptomatic STIs and BV [85]. Genital inflammation is also perhaps due to disruption of the mucosal epithelial integrity and impaired wound healing by dysbiotic CVMs [122]. *In vitro* and *in vivo* models have demonstrated that genital inflammation is caused by proinflammatory-associated bacteria, such as *Sneathia* and *Atopobium* that trigger cytokine and chemokine production, including IL-1, IL-8, tumour necrosis factor-alpha (TNF-α), interferon-alpha (INF-α), INF-γ, and IL-12p70 [164]. High levels of MIP-1α and MIP-1β that attract CCR5-expressing cells have been associated with increased bacterial diversity [164]. CCR5-expressing cells are involved in HIV transmission [197]. BV has been associated with high $CD4^+$ T-cells [124].

CVMs have also been investigated in women who are already HIV-infected. A cross-sectional study evaluating the CVMs of 278 Nigerian women [137] revealed that HIV-infected women more likely to have *Lactobacillus*-deficient CVMs (CST IV-B and CST VI, a Proteobacteria-dominated CVM) relative to HIV-uninfected women. Another study conducted on a Rwandan cohort observed that women with *Lactobacillus*-dominated CVMs, especially *L. crispatus*, had the lowest prevalence of HIV (9.1%) compared to *Lactobacillus*-deficient CVMs (34.3-56.3% prevalence) [33]. The study further reported that the composition of the CVM affects HIV shedding [33]. HIV-infected women with *L. crispatus*-dominated CVMs were found to have significantly reduced genital HIV-1 RNA shedding [33]. Experiments have demonstrated that diverse CVMs and at times *L. iners*-dominated CVMs cannot trap and suppress HIV replication whereas other lactobacilli, including *L. crispatus* and *L. gasseri* can significantly suppress HIV-1 replication [146,161,212]. Bacteria such as *A. vaginae*, *Sneathia sanguinegens*, *Aerococcus* spp., and *Ruminococcaceae* spp. have been found to enhance HIV replication even in *Lactobacillus*-dominated CVMs, specifically of *L. iners* type [161].

Information on the impact of HIV on CVMs and BV is relatively scarce and still inconclusive. A recent longitudinal study on 22 HIV-uninfected women and 42 HIV-infected women (22 with stable infection and 20 with progressive disease) observed no association between HIV status and CVMs at baseline and follow-up over an 8-10 year period [213]. In a cohort of 82 young South African girls, Puran and co-workers (2014) [86] observed a non-significant trend towards an increased BV prevalence among HIV-infected females (62.5% in HIV-infected versus 50.0% in HIV-uninfected). Spear and colleagues (2008) [174] investigated whether HIV-status had any influence on the CVMs of 21 women in the U.S. with and without BV. Although they observed that HIV status had minimal effect on the overall bacterial diversity, they reported a trend towards increased diversity in HIV-infected women with BV relative to HIV-uninfected women with BV [174]. Furthermore, they observed that three taxa, *Propionibacterineae*, *Citrobacter*, and *Anaerococcus* were present only in HIV-infected women [174]. The differences in study population and small sample size in these studies may explain the lack of consistency in these studies and appropriately designed longitudinal studies are needed.

1.3.6.2 Cervicovaginal microbiome and genital HPV infections

Several studies, mostly on Westerners and Eurasian cohorts, have outlined an association of BV as well as cervicovaginal bacterial communities with HPV infection and HPV natural history. The findings of these studies are underscored below (1.3.6.2.1-1.3.6.2.3).

1.3.6.2.1 BV, vaginal pH and genital HPV infections

A meta-analysis of 12 published articles on 6,372 women aged 13-69 years associated prevalent BV with uterine cervical HPV infection [81]. A subsequent 6-year follow-up study on 329 Finnish women [214] associated asymptomatic BV with persistent cervical HPV infections [214]. Another temporal study was conducted in the U.S. to evaluate the association between common vaginal infections/disorders (including BV) and the natural history of HPV among women aged 15-56 years that included 756 HIV-positive women and 380 women at high risk for acquiring HIV infection [210]. The study found that preceding or current BV status was associated with prevalent HPV, incident HPV, and delayed HPV clearance, whereas no association was observed between preceding HPV infection and BV [210]. Additionally, the effect of BV on incident and prevalent HPV was not affected by HIV status. However, in a model adjusted for a combined HIV status and CD4 group, BV was significantly associated with increased risk of incident and prevalent HPV [210].

The vaginal pH is believed to be a modifier of the natural history of HPV infection. This is based on the observation that decreased vaginal acidity ($pH \geq 5.0$) has been associated with increased risk of infection with high-risk and multiple HPV infections [70].

1.3.6.2.2 Cervicovaginal bacteria, community state types and genital HPV infections

The association of cervicovaginal bacteria and CSTs with the prevalence and natural history of genital HPV infection is summarised in Table 1.1. Published evidences [137,157,179,215], mostly based on cross-sectional studies, have reported higher frequencies and abundances of particular bacterial taxa in CVMs are possible biomarkers for HPV infection, including high-risk genotypes types. A cross-sectional study assessing the microbial diversity of 32 HPV-negative and 38 HPV-positive Chinese women with normal cervical cytology [215] found that *L. gasseri* and *G. vaginalis* were more frequently detected in HPV-positive women than in HIV-negative women. Furthermore, HPV-positive women had significantly higher microbial diversity than HPV-negative women [215]. Frequencies of *L. iners* and *L. gallinarum* did not vary between the two HPV groups [215]. Compared to HPV-negative women, the CVMs of women with prevalent HPV infection have been found to have higher relative abundances of bacteria such as *Prevotella* [137,179], *Megasphaera*, *Sneathia* [137,157,179], *Fusobacterium*, *Dialister* [179], *Pseudomonas*, *Brevibacterium* [157], *Anaerococcus*, and *Arcanobacterium* [137]. In a Korean twin cohort cross-sectional study linking CVMs with host genetics [180], women infected with HPV had different microbial profiles that varied according to their menopausal status. *Lactobacillus* and *Sneathia* were dominant in premenopausal women infected with HPV (55.4% and 11.6%, respectively), while non-heritable *Prevotella* (22.7%) was the most abundant in postmenopausal women infected with HPV [180]. On the other hand, *Alkanindiges*, *Anaerococcus tetradius*, *Corynebacterium accolens*, *Finegoldia magna*, *F. nucleatum*, *Raoultella planticola*, *Halorubrum*, and *Vampirovibrio* have been found to be highly enriched in women without prevalent HPV infection [137,141,157].

In spite of a few conflicting reports [34,139,141], most CVMs results [33,34,139,141,142,179] have associated higher abundances of *Lactobacillus* and/or *Lactobacillus*-dominated CSTs with lower prevalent HPV infection. A Nigerian cohort cross-sectional study that characterised bacteria associated with prevalent high-risk HPV infection found that HPV-negative women had higher relative abundances of *L. iners* than HPV-positive women [137]. A study on North American cohort prospectively assessing the relationship between CVMs and HPV while controlling for immune status [139] found that among the *Lactobacillus* spp. (*L. crispatus*, *L. gasseri*, *L. iners*, and *L. jensenii*), only the relative abundance of *L. crispatus* was significantly associated with decreased HPV or oncogenic HPV detection. There was no association of relative abundance of genus *Lactobacillus* with prevalent HPV [139]. A Korean twin cohort cross-sectional study [179] had previously found significantly higher relative abundances of *Lactobacillus* in HPV-negative women relative to HPV-positive. Recently, a prospective study by Shannon and colleagues (2017) [141] assessing the impact of natural history of HPV on CVM and genital immune responses in 59 African/Caribbean women in Canada associated both low bacterial load and abundances of *L. gasseri* with prevalent HPV.

Table 1.1. Community state type and classified bacterial genus/species in cervicovaginal microbiota associated with HPV infection, clearance and persistence.

Cervicovaginal bacteria/CST	**Association with prevalent HPV infection**	**Association with acquisition of HPV infection**	**Association with HPV clearance or persistence**
Lactobacillus crispatus	Negative [139,142]	-	-
Lactobacillus gasseri	Positive [215] Negative [141] No [139]	-	-
Lactobacillus iners	Negative [137,139,215]	-	-
Lactobacillus jensenii	No [139]	-	-
Lactobacillus gallinarum	No [215]	-	-
Lactobacillus	No [139]	-	-
***Lactobacillus* spp.**	Negative [179]	-	-
Actinomyces	Negative [157]	-	-
Albidiferax	-	-	Positive [157]
Alkanindiges	Negative [157]	-	-
Anaerococcus	Positive [137]	-	-
Anaerococcus tetradius	Negative [141]	-	-
Arcanobacterium	Positive [137]	-	-
Atopobium vaginae	-	-	Positive [157]
Brevibacterium	Positive [157]	-	-
Corynebacterium accolens	Negative [141]	-	-
Dialister	Positive [137,179]	-	-
Gardnerella vaginalis	Positive [215]	-	-
Faecalibacterium	-	-	Positive [157]
Finegoldia magna	Negative [141]	-	-
Fusobacterium	Positive [179]	-	-
Fusobacterium nucleatum	Negative [141]	-	-
Halorubrum	Negative [137]	-	-
Megasphaera	Positive [137,157,179]	-	-
Pediococcus	Positive [157]	-	-
Peptoniphilus asaccharolyticus	Positive [137]	-	-
Peptoniphilus harei	Negative [141]	-	-
Peptostreptococcus anaerobius	Positive [137]	-	-
Prevotella	Positive [137,179]	-	-
Pseudomonas	Positive [157]	-	-
Pseudoxanthomonas	Negative [157]		
Raoultella planticola	Negative [141]	-	-
Sneathia	Positive [137,157,179]	-	-
Vampirovibrio	Negative [157]	-	-
CST I (*L. crispatus* dominated)	Negative [33,139]	-	
CST II (*L. gasseri* dominated)	-	-	Positive [34]
CST III (*L. iners* dominated)	Positive [34] Negative [33]	-	-
CST IV*/Diverse CVMs	Positive [33,34,141,214,215]	-	Negative [34]
CST IV-A	-	Positive [34]	-
CST IV-B	-	-	Positive [34]
CST IV-BV	-	-	Positive [157]

Abbreviations: HPV – human papillomavirus, CST – community state type.

*CST IV – *Lactobacillus*-deficient cervicovaginal microbiotas that are usually dominated by a mixture of strict or facultative anaerobic bacteria [38]. Some authors have subdivided this BV-like CVM depending on the array of bacteria colonising the cervicovaginal milieu: CST IV-A and CST IV-B [34] or CST IV-BV and CST IV-AB [157].

Compared to diverse CVMs (CST IV) *Lactobacillus*-dominated CVMs (CST I: *L. crispatus*-dominated, CST II: *L. gasseri*-dominated, and at times CST III: *L. iners*-dominated) have been associated with reduced prevalent HPV and high-risk HPV infection [33,34,139,179]. The study by Shannon and colleagues (2017) [141] found that CST IV was more prevalent among women with HPV infection relative to CST I and CST II. In another longitudinal study [34], women with CVMs in CST IV and CST III had higher frequencies of HPV (71% and 72%, respectively) compared with CST I (45%) and CST II (12%). Further, there was a trend towards HPV positivity in samples from women in CST IV-B and CST III relative to CST II [34]. Other published literatures [33,139] have observed an inverse relationship between CST I and prevalent of any HPV and/or high-risk HPV. Cross-sectional studies have associated increasing microbial diversity with an increasing prevalence of HPV and high-risk HPV among Rwandan [33] and Korean [179] women. There is dearth of studies on the association of cervicovaginal bacterial communities with acquisition of HPV infection (Table 1.1). A CVM/HPV interplay temporal study by Brotman and colleagues (2014) [34] on 32 African American women showed that there was a trend of an increased risk to transition to HPV positivity in women with CST IV-A (low *Lactobacillus* CVMs with several anaerobes, including *Prevotella*, *Atopobium*, *Anaerococcus*, and *Streptococcus*) compared to CST I [34].

In terms of rates of HPV remission, Brotman and colleagues (2014) [34] showed that *Lactobacillus*-dominated CVMs, CST II in particular, was associated with more rapid remission of HPV compared to CST III and non-*Lactobacillus* CVMs, specifically CST IV-B (highly colonised with *Atopobium*, *Prevotella*, and *Gardnerella*) [34]. Their study demonstrated that, of all the CSTs (CSTs I-III, IV-A, and IV-B), CST IV-B had the lowest clearance rates of HPV [34]. A prospective study characterising CVMs of Italian women with cleared and persistent high-risk HPV infection [157] found that lower relative abundances of *Albidiferax* and higher relative abundances of *Atopobium* and *Faecalibacterium* were strongly associated with HPV persistence.

The effect of CVMs against HPV infection seems to be due to the mechanisms employed by *Lactobacillus* spp., including low acidic milieu that is thought to be contributed by production lactic acid [63,68]. Vaginal pH may also be important in HPV pathogenesis. A large Costa Rican cohort prospective study associated reduced risk of HPV detection with a vaginal pH of <5.0 [70]. The relative abundance of genus *Lactobacillus* has been more significantly associated with acidic pH than the relative abundance of the individual *Lactobacillus* species [139]. Among the individual *Lactobacillus* species, the relative abundance of *L. crispatus* has been found to exhibit the highest correlation with acidic pH, followed by *L. jensenii*, *L. gasseri*, and *L. iners* [139]. Differences in *Lactobacillus* spp. may therefore explain the different effects observed on the natural history of HPV infections among women with *Lactobacillus*-dominated CVMs [34]. Apart from the low vaginal pH, other biological mechanisms such as the isoform of lactic acid have been posited as important factors utilised by *Lactobacillus* such as *L. crispatus* to foil detection of HPV infection [139]. D- and L-isoforms of lactic acid can have antiviral activity [146,212]. However, high ratio of L- to D-lactic acid as evident in *L. iners*-dominated CVMs [146,147] may predispose women to BV and alter cause loss cervical integrity [147] that can allow HPV particles to enter and infect the basal cells [4]. Differences in prevalence and remission of HPV may further be pointed to the immunological responses [141], temporal changes of the cervicovaginal milieus and stability of the CSTs [34]. A study assessing the impact of natural history of HPV on CVM and genital immune responses noted that prevalent HPV was associated with elevated levels of IP-10 and monokine induced by gamma interferon (MIG) while HPV clearance was associated with increased cervical Langerhans cells that also play a role in HIV acquisition [141].

1.3.6.2.3 Cervicovaginal bacteria, community state types and HPV-associated cervical precancer and cancer

Several studies [140,142,160,176,179,216-219], albeit relying on cross-sectional data, have examined the cervicovaginal microbiota during the various stages of HPV-associated cervical precancer and cancer (Table 1.2, ranked by the recency of the publication). Cervical precancer has been categorised according to the extent of neoplasms (dysplasias/dyskaryosis – abnormal growth of cells on the surface of the uterine cervix) [4,5] while cervical cancer according to increasing invasiveness of the malignancy (Stages I-IV) [5]. The categories of neoplastic phenotypes are cervical intraepithelial neoplasia (CIN) 1-3 [4,5]. CIN 1, CIN 2, and CIN 3 are mild, moderate, and severe dysplasia, respectively [5]. The corresponding cytology of CIN 1 is low-grade intraepithelial lesion (LSIL) while for CIN 2 and CIN 3 is high-grade intraepithelial lesion (HSIL) [5]. It is worth noting that atypical cells of undetermined significance (ASCUS) is a very mild grade of cervical abnormality as manifested by the presence of slightly abnormal squamous abnormal cells that are of uncertain precancerous nature [5].

There is rapidly evolving evidence that dysbiotic CVMs and certain taxa in the CVMs are associated with precancerous cervical lesions [140,142,160,176,179,219,220] (Table 1.2). A culture-independent case-control study assessing the association between CIN and uterine cervical microbiota of Korean HPV cohort aged 18-65 years [160] found that women with predominance of *G. vaginalis* and *A. vaginae* in tandem with a high proportion of *L. iners*/*L. crispatus* ratio had a high likelihood of having CIN. The study also showed that a synergistic effect between BV-associated bacteria, particularly *A. vaginae*, and high-risk HPV, was associated with increased risk of CIN [160]. Cervical microbiotas of women with HSIL have been observed to have lower relative abundances of *L. jensenii* and *L. coleohominis* and higher relative abundances of *S. sanguinegens*, *A. tetradius*, and *Peptostreptococcus anaerobius* relative to women with LSIL [140]. Women with normal cervical cytology have been found to have higher relative abundances of *L. crispatus* and *L. iners* than women with SIL and cervical cancer respectively [142]. However, a study examining the association between CVM and CIN [219] observed that relative abundances of *Lactobacillaceae*, *Lactobacillus*, *L. reuteri*, and several *Lactobacillus* operational taxonomic units (OTUs) were higher in women with CIN 2+ (CIN 2 and CIN 3) relative to women with CIN 1. These inconsistencies could be due to differences in study design and lack of adjustment for other confounding factors.

Table 1.2. Community state type and classified bacterial genus/species in cervicovaginal microbiota associated cervical precancer and cancer.

Study description/aim	Number of participants	Summary of the main findings	Reference
To assess the association of dietary pattern with CIN and the combined effect of dietary pattern and cervical microbiome on the risk of CIN	137 South Korean women aged 18-65 years who included 65 with CIN (50 with CIN 1 and 15 with CIN 2 or CIN 3) and 72 controls (with normal or ASCUS in Pap smear results)	▪ Women with a semi-Western diet in the high-scoring group (with a highly unbalanced dietary pattern) had a higher CIN risk than women with a semi-Western diet in the low- or medium-scoring group. ▪ CVMs dominated by *L. iners* and *A. vaginae* had a higher risk of CIN compared to *L. crispatus*-dominated CVMs. ▪ There was a synergistic effect between semi-Western diet and *A. vaginae*-dominated CVM, which increased the CIN risk.	Seo *et al.* (2016) [176]
To examine the association between CVMs and high-grade CIN (CIN 2+) and to evaluate whether the CVMs are associated with oxidative damage	430 high-risk HPV-positive North American women aged 19-50 years who included 79.1% with CIN 2+ (208 CIN 2 and 132 CIN 3 diagnosed cases) and the rest (n=90) with CIN 1 (non-cases)	▪ CVM dominated by *L. iners* and unclassified *Lactobacillus* was positively associated CIN 2+. ▪ Relative abundances of *Lactobacillaceae*, *Lactobacillus*, *L. reuteri*, and several *Lactobacillus* OTUs were higher in women with CIN 2+ than with CIN 1. ▪ DNA oxidative damage did not correlate with CVM structure.	Piyathilake *et al.* (2016) [219]
To evaluate the profile of CVMs at different stages of the natural history of cervical cancer	32 Mexican women aged 22-61 years who included 20 women without cervical lesions (50% HPV-positive), 4 with SIL (all HPV-positive), and 8 with cervical cancer (all HPV-positive)	▪ Women with SIL and cervical cancer had a significantly higher microbial diversity compared to HPV-negative women without cervical lesions. ▪ Women with normal cervical cytology had higher abundances of *L. crispatus* and *L. iners*, whilst *Sneathia* spp. and *Fusobacterium* spp. dominated the CVMs of women with SIL and cervical cancer, respectively. ▪ Cervical expression levels of TGF-β1 and IL-4 mRNA were higher in women with *Fusobacterium*-dominated CVMs.	Audirac-Chalifour *et al.* (2016) [142]
To investigate whether increasing microbiome diversity is associated with increasing CIN severity	169 women aged 18-45 years from the UK who included 20 with normal cervical cytology, 52 with LSIL, 92 with HSIL, and 5 with invasive cervical cancer	▪ Severity of CIN was associated with increasing CVM diversity and CST IV. ▪ Compared with disease-free healthy controls, the prevalence of CST IV was about 2-fold, 3-fold, and 4-fold greater in women with LSIL (21%), HSIL (27%), and invasive cervical cancer (40%), respectively. ▪ Association between CVM diversity and disease severity was observed irrespective of the HPV status. ▪ Prevalence of CST I (*L. crispatus*-dominated CVMs) was lower with increasing disease severity (CST I prevalence: 50% in normal, 42% in LSIL, 40% in HSIL, and 20% in invasive cervical cancer). ▪ CVMs of women with HSIL had higher relative abundances of *Sneathia sanguinegens*, *Anaerococcus tetradius*, and *Peptostreptococcus anaerobius* relative to women with LSIL. ▪ Women with LSIL had higher relative abundances of *L. jensenii* and *L. coleohominis* than women HSIL.	Mitra *el al.* (2015) [140]
To evaluate the association of uterine cervical microbiota with CIN	120 Korean women aged 18-65 years who included 70 women with CIN and 50 control women	▪ Women with predominance of *G. vaginalis* and *A. vaginae* in tandem with a high proportion of *L. iners*/*L. crispatus* ratio had a high likelihood of having CIN.	Oh *et al.* (2015) [160]

		▪ The highest synergistic effect was between *A. vaginae* and high-risk HPV was associated with increased risk of CIN.	
To assess the association between CVM and HPV infection	50 Korean women age 31-73 years who included 23 HPV-positive (4 with CIN) and 27 HPV-negative (1 with CIN) premenopausal women	▪ Average relative abundance of *Lactobacillus* spp. was significantly lower in groups infected with HPV or with CIN than in the premenopausal control group. ▪ *Lachnospiraceae* and *Veillonella* spp. included bacterial taxa that were in higher relative abundances in women with CIN.	Lee *et al.* (2013)
To clarify the association between these two conditions using a systematic review and meta-analysis	Over 20,500 women aged 13-78 years from different countries (South Africa, Costa Rica, Brazil, Canada, USA, UK, Italy, Netherlands, Sweden, Finland, and South Korea) representing different races/ethnicities	▪ Women with BV were significantly more likely to have CIN/SIL than women without BV.	Gillet *et al.* (2012) [220]
To determine the prevalence, risk factors, and impact on histologic changes of BV in women with cytological abnormalities of the uterine cervix	280 Dutch women aged 20-66 years who included 68 with minor lesions (no lesions or CIN 1) and 212 with more severe lesions (CIN 2 and CIN 3)	▪ The prevalence of BV was 56%. ▪ Of all the factors (age of sexual debut, current *Chlamydia trachomatis* infection, number of cigarettes currently smoked per day, and lifetime number of sexual partners) significantly associated with BV, only the latter two were identified as independent significant risk factors for BV. ▪ BV did not influence the severity of CIN or mitotic index.	Peters *et al.* (1995) [221]

Abbreviations: CVM – cervicovaginal microbiota, OTU – operational taxonomic unit, CIN – cervical intraepithelial lesion, ASCUS – atypical squamous cells of undetermined significance, SIL – squamous intraepithelial lesion, LSIL – low-grade SIL, HSIL – high-grade SIL, HPV – human papillomavirus, TGF – transforming growth factor, IL – interleukin, mRNA – messenger RNA (ribonucleic acid), BV – bacterial vaginosis, CST – community state type.

*CST IV – *Lactobacillus*-deficient cervicovaginal microbiotas that are usually dominated by a mixture of strict or facultative anaerobic bacteria [38].

Based on the available literature, Kyrgiou and colleagues (2016) have suggested that the interplay between specific CSTs such as BV-associated CST IV and HPV infection may be correlated with the development of cervical disease, e.g., SIL [216]. CST IV is a high diversity CVM [38] that has been positively associated with severity of CIN irrespective of HPV status [140]. A cross-sectional cohort study characterising the CVMs in CIN disease progression [140] found that, compared with disease-free healthy controls, the prevalence of CST IV was about 2-fold, 3-fold, and 4-fold greater in women with LSIL (21%), HSIL (27%), and invasive cervical cancer (40%), respectively. In contrast, the prevalence of communities with *L. crispatus* predominance (CST I) was decreased with increasing disease severity (CST I prevalence: 50% in normal, 42% in LSIL, 40% in HSIL, and 20% in invasive cervical cancer) [140]. In a large study on a cohort of high-risk HPV-positive women [219], there was the positive association of a cervical mucosal community dominated by *L. iners* and unclassified *Lactobacillus* with CIN 2+. A small cross-sectional pilot study evaluating the profile of cervical microbiotas and cytokines at different stages of the natural history of cervical cancer in 32 Mexican women [142] found that *Fusobacterium* spp. was predominant in CVMs of women with cervical cancer. Such women had elevated levels of transforming growth factor (TGF)-β1 and IL-4 mRNA [142]. Moreover, the study confirmed that high microbial diversity (in non-*Lactobacillus* CVMs) is associated with cervical precancer and cancer [140]. There is also a connection between BV and precancerous cervical lesions. A systematic review and meta-analysis of the association of BV and precancerous cervical lesions (mostly from cross-sectional studies) [220] observed that women with BV were significantly more likely to have CIN/SIL than women without BV.

The observed interrelationship between cervicovaginal bacteria communities, HPV, and cervical precancer/cancer has led some investigators [142,218] to suggest possible mechanisms by which cervicovaginal bacterial communities may promote the development of cervical cancer. Audirac-Chalifour and co-workers (2016) [142] proposed a mechanism in which specific bacteria in the CVMs mediate changes in cytokine expression levels to promote the development of SIL and cervical cancer (Fig. 1.6). They speculate that infection of the cervicovaginal epithelia with high-risk HPV causes the CVM to shift from *L. crispatus* to *L. iners* predominance. As the HPV-infected cells continue to transform to SIL, CVM diversity increases with the presence of pathobionts, mainly *Sneathia* and *Fusobacterium* spp. *Fusobacterium necrophorum* is present when cervical cancer finally ensues. The shift in CVM composition and diversity is due to the immunosuppressive microenvironment evoked by HPV infection, which continues to mediate the cross-talk between cytokines and CVM [142].

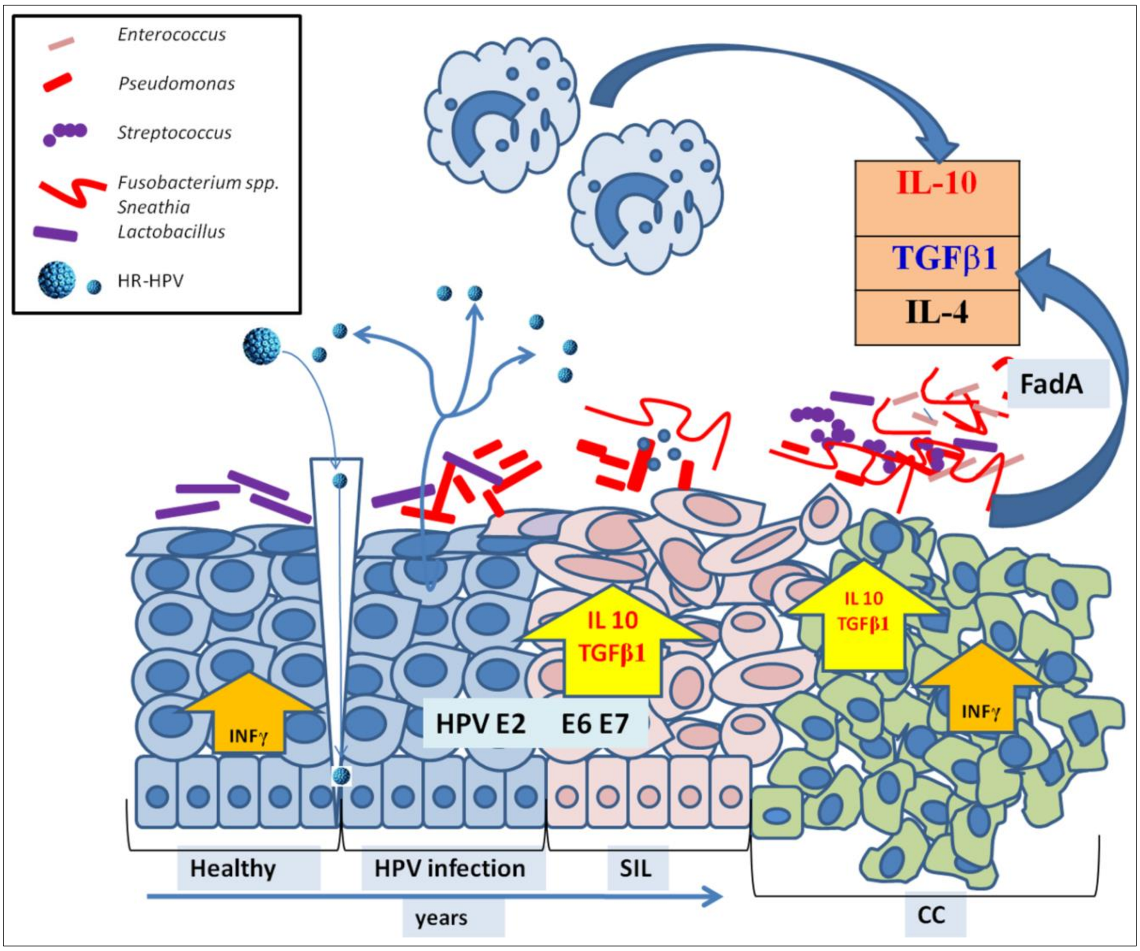


Fig. 1.6. Proposed mechanism of cervicovaginal microbiome changes during the cervical cancer development. Changes occurring to the normal cervicovaginal epithelia from the onset of infection with high-risk HPV until the development of cervical cancer through SIL are shown. Microbiome structure in each stage is denoted by the predominant bacteria. HPV infection results in loss of *Lactobacillus* and increased diversity of the CVM, with *Pseudomonas* spp. predominance. Early proteins (E2, E6, and E7) of HPV facilitate the expression of IL-10 and the occurrence of macrophages type 2. The latter is also promoted by TGF-β1, which is subsequently triggered by the CVM present. As HPV infection persists and SIL progresses to cervical cancer, the diversity of the CVM increases, with the predominance of *Sneathia*, *Fusobacterium*, *Enterococcus*, and *Streptococcus*. The diversity of the CVM increases through its toxins, particularly FadA, a virulent bacterial cell surface adhesion factor from *Fusobacterium* spp. FadA then disrupts the tight junctions and enhance a metastasis akin to colon carcinoma. The figure is taken from Audirac-Chalifour and co-workers (2016) [142].

Although all the studies associating specific CVMs and bacterial taxa with precancerous cervical lesions are cross-sectional-based and therefore do not provide evidence for causality, a review by Mitra and co-workers (2016) [218] founded on observational studies of CVM, HPV, and CIN proposed possible mechanisms through which CVMs may use to mediate cervical heath and disease, including cervical disease (Fig. 1.7). Most of these mechanisms have already been mentioned earlier in this chapter and include: 1) vaginal pH, lactic acid, and H_2O_2, 2) production of bacteriocins and biosurfactants, 3) disruption of cervicovaginal mucus and epithelial barrier, 4) induction of oxidative stress, and 5) participation of specific bacterial species and manipulation of cellular targets, e.g., survivin, tumour protein p53, etc. [218].

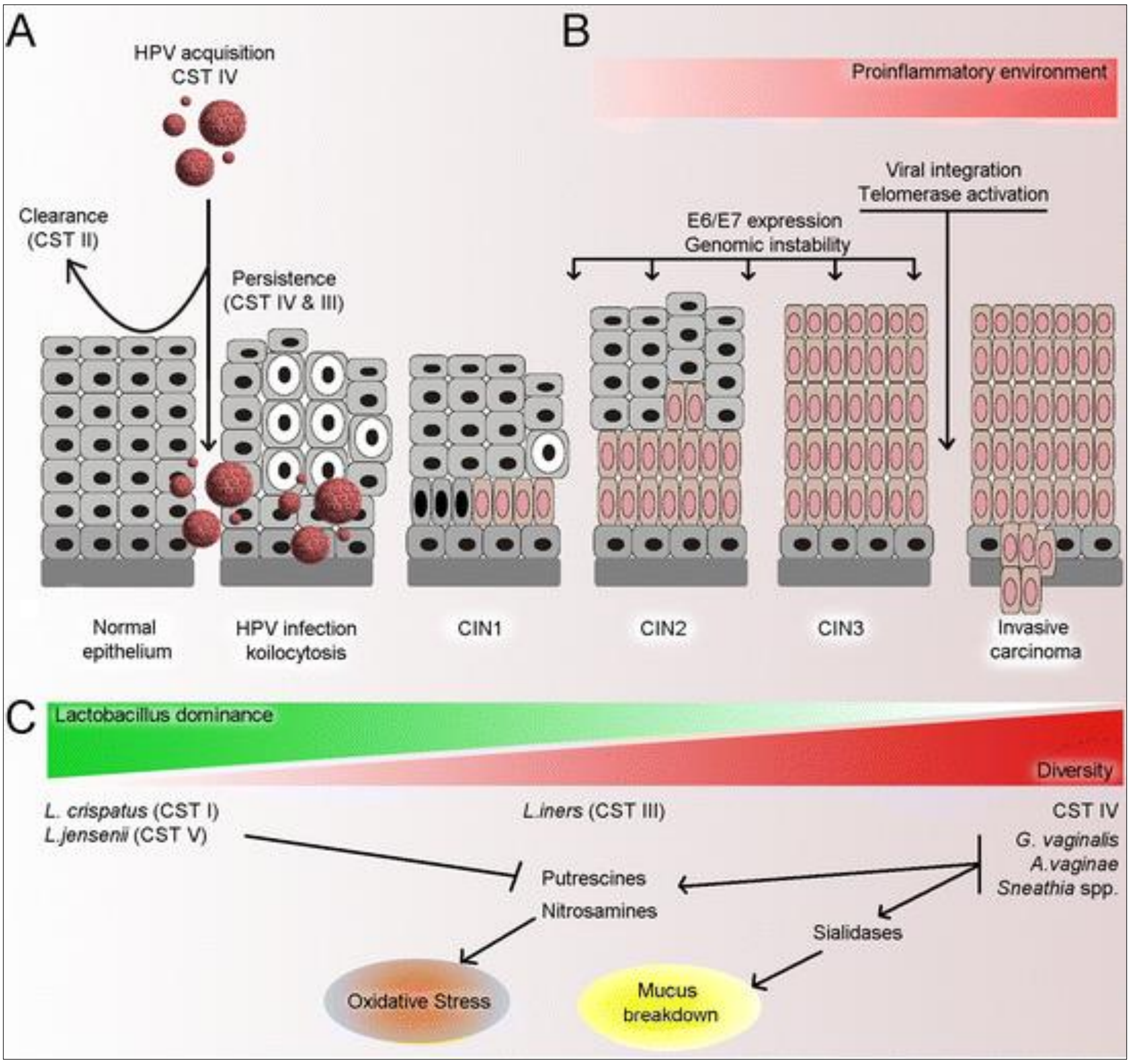


Fig. 1.7. Possible mechanisms linking cervicovaginal microbiotas with the evolution of cervical disease. a) *Lactobacillus*-deficient CVMs (CST IV, BV-like CVMs) and occasionally *L. iners*-dominated CVMs (CST III) are associated with incident and persistent HPV infection. *Lactobacillus*-dominated CVM (mostly CST II, with *L. gasseri* predominance) is associated with the fastest remission of acute HPV infection. HPV persistence may cause koilocytic dysplasia, mild lesions of the uterine cervix (CIN 1). b) Dysbiosis may evoke robust proinflammatory changes in the cervicovaginal milieu that enhance HPV-induced carcinogenesis. HPV-induced carcinogenesis is facilitated through expression of E6 and E7, host genome instability, viral integration, and telomerase activation. c) Higher bacterial diversity with *Lactobacillus* deficiency has been associated with increasing disease severity. Specific bacteria frequently found in BV-like CVMs (*Sneathia*, *Gardnerella*, and *Atopobium*) may produce sialidases that cause mucus breakdown. This may lead to loss of the epithelial integrity and production of biogenic amines (putrescines and nitrosamines) associated with oxidative stress-induced DNA damage, which is a crucial step in carcinogenesis. Specific *Lactobacillus* spp. can dampen the expression of these biogenic amines; hence reducing the risk of oxidative damage. *L iners* is known to be the least protective among the common *Lactobacillus* spp. and has a bidirectional role in health and disease, and perhaps cervical disease. The figure is taken from Mitra and co-workers (2016) [218].

The observation of diminished abundance of *Lactobacillus* in women with CIN [140,142,179,220] is growing to be a common finding in microbiome studies. It could be that cervicovaginal communities without appreciable numbers of protective *Lactobacillus* promote the persistence of HPV [34] and thereby increase the risk of HPV-induced malignancy.

The protective nature of *Lactobacillus* against HPV-induced precancerous lesions was demonstrated in a small 6-month longitudinal controlled pilot study examining the effect of the oral probiotic *Lactobacillus casei* on the clearance of HPV-induced lesions [222]. Probiotic users (n=24) had 60.0% chance of clearing HPV-related cervical abnormalities relative to the control group (n=27) who had a 30.7% chance at 6 months, although this difference was not statistically significant. After 3 months, 25.0% of the probiotic users and 7.7% of the control participants had cleared HPV infections (p=0.13). After 6 months, the proportions were 29.2% and 19.2%, correspondingly (p=0.41) [222]. Lactic acid from lactobacilli is known to exhibit some antiviral activity [223]. It can inhibit the virus by direct virus-probiotic interaction aided by adsorption or trapping mechanism. Alternatively, lactic acid bacteria can evoke host's immune antiviral responses and/or produce nonproteinaceous metabolites with inhibitory capabilities [223]. Lactic acid also acidifies the cervicovaginal environment, hence resulting in low pH [63,68] which is a key mechanism for controlling the composition of CVMs and preventing colonisation of the microenvironment with pathogens [68,69]. Elevated vaginal pH of ≥5.0 has been associated with detection of high-risk HPV and LSIL in women [70]. Lactobacilli appear to have other mechanisms used to curtail viral infections. Motevaseli and colleagues (2013) [224] showed using culture and bioassay experiments that *L. crispatus* and *L. gasseri* have cytotoxic activity on tumour cells and that this activity is independent of pH and lactic acid.

1.4 The human penis and its microbiome

The human penis mainly consists of the glans, corona, urethra, corpora cavernosa penis, corpus spongiosum and prepuce or foreskin (in uncircumcised men) (Fig. 1.8) [225-227]. The various surfaces of the penis represent different microenvironments, which differ in properties such as oxygen availability, keratinisation and wetness, which would provide suitable niches for different bacterial communities.

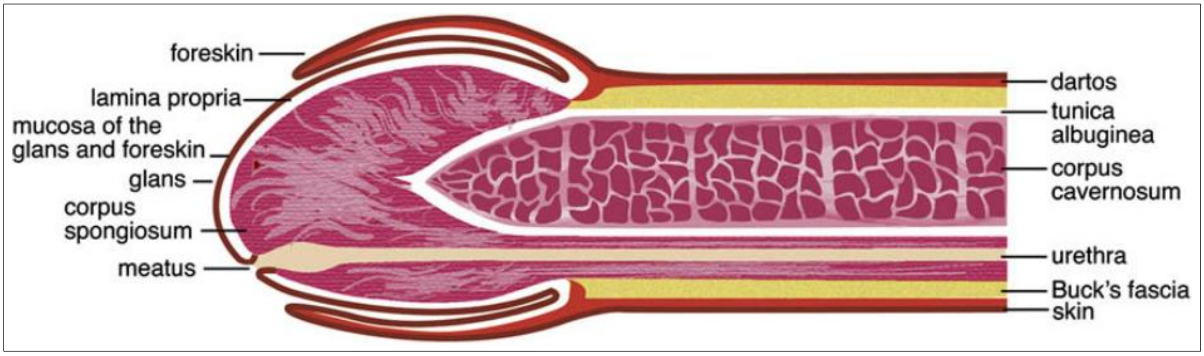


Fig. 1.8. Anatomy of the human penis. The figure is taken from Wasco and Shah (2009) [227].

The external surface of a relaxed penis is covered with “dry” keratinised squamous epithelial layer that is relatively impermeable to STIs in the absence of epithelial trauma and/or lesions [225,226]. In majority of uncircumcised men, the prepuce often covers the glans, corona (junction between the glans and penile shaft), and the tip of the penis (meatus) [225]. The subpreputial epithelium that covers the glans, corona, and inner surface of the prepuce is made up of mucosal (“wet”) epithelial cells [225,226,228]. The subpreputial space is an anoxic microenvironment [229]. The length of the prepuce of uncircumcised men has been associated with subpreputial wetness [230]. This subpreputial penile wetness of glans and coronal sulcus increases vulnerability to STIs, especially HIV [231]. Circumcision removes at least most of the mucosal preputial epithelium but leaves the dry keratinised epithelial surface [225]. Circumcision results in keratinisation of the glans penis [226]. At the penile urethral orifice, the meatus consists of keratinised squamous epithelium, which transitions into non-keratinised form in the fossa navicularis. The non-keratinised squamous epithelium further transitions into pseudostratified columnar epithelium through the urethral inner lining [225,227]. The meatus, urethral orifice, and the penile skin, including the inner surface of the foreskin, are suitable niches for microbes [110,225,226] and at times pathogens [232]. The mucosal epithelial of the penis has immunological cells that act against infections by pathogens [228].

Removal of the prepuce by medical male circumcision significantly reduces the risk of STIs, including HSV-2, syphilis, HPV, and HIV [233-239]. Three randomised clinical trials

(RCTs), conducted in South Africa [233], Kenya [237], and Uganda [238] indicated that medical circumcision reduced the rate of heterosexual HIV transmission by about 60%. Two recent systematic reviews and meta-analyses [240,241] have found that male circumcision reduces prevalent HPV by between 32-43%. The foreskin has a high prevalence of HPV compared to other penile sites in uncircumcised men [241]. The protective effect of male circumcision against HPV infection has been reported to be enhanced in glans/corona and urethra relative to sites more distal sites to the prepuce such as the penile shaft and scrotum [241]. Following the observed considerable protective effects of medical male circumcision against HIV/STIs [240-242], male circumcision has been cited as an essential and effective element of HIV/STIs prevention strategies not only in Africa, but the rest of the world too [243].

In many settings within sub-Saharan Africa, including South Africa, traditional male circumcision is common [244-249]. A systematic review by Wilcken and colleagues (2010) [244] reported that in eastern and southern Africa, traditional male circumcision is performed as a ritual to initiate the males into manhood. It is also been reported that it is preferred over voluntary medical male circumcision because it is deemed affordable and the cost is negotiable at times [246]. It is usually performed by a medically untrained provider in a non-clinical setting, with either the foreskin partially or fully (sufficiently) removed [244,245,249]. Traditional male circumcision may offer some level of protection against HIV infection. A 2-year follow-up study among low-risk men in rural Kenya [248] observed a 3-fold lower HIV incidence rate in predominantly traditionally circumcised men than uncircumcised men. A cross-sectional comparative study based on 18 demographic and health surveys conducted in sub-Saharan Africa [250], where traditional male circumcision is predominant, strongly associated circumcised with reduced risk HIV infection. However, a population-based survey of predominantly traditionally circumcised South African men found no association between circumcision and HIV infection [245]. A more recent study investigating the association between type of male circumcision (medical versus traditional) and HIV status on a Basotho cohort [247] noted that traditionally circumcised men were more likely to be HIV-infected than medically circumcised men.

The differences in the protective effect of traditional circumcision on HIV infection are likely due to variations in the age of circumcision, age of sexual debut and amount of foreskin removed during traditional circumcision [235,244-247]. These are known to vary regionally [244-246]. A study in Cape Town, which separated traditionally circumcised men by partial and full removal of the foreskin, showed that partially circumcised and uncircumcised men had the same risk for HIV infection, which was significantly greater than that of fully circumcised men [251]. Differences in HIV risk in medically and traditionally circumcised men may also be attributable to lack of HIV risk reduction counselling or formal counselling received in traditionally circumcised men [252,253] and inaccurate self-reporting of male circumcision status [249]. A cross-sectional study on a Basotho cohort comparing concordance of circumcision status by self-report with physical examination [249] found that of the men self-reporting being circumcised only half had their foreskins completely removed, while the rest were found to be either uncircumcised (23.4%) or partially circumcised (26.6%) [249]. The partial removal of the foreskin could still maintain the subpreputial penile wetness of glans and coronal sulcus akin to that of uncircumcised men. This subpreputial penile wetness has been associated with HIV infection [231]. Moreover, traditional male circumcision has been found to be more likely to have complications such as infections and delayed wound healing [244,246], which may impact the risk of STIs. In contrast, the incidences of such complications were found to be low in RCTs of medical circumcision for HIV reduction [233,237,238,242].

The penile microbiota remains largely understudied. To date, there are less than ten published papers that have used deep sequencing to examine this topic. The findings of these publications are summarised in Table 1.3, ranked by the recency of the publication. Of these studies, five have been on a Ugandan cohort [35,194,229,254,255] while the rest were on U.S. [110] and Australian [256] cohorts. The studies have focussed on changes to microbiotas following medical male circumcision [229,254] or on microbiotas of men who are circumcised and/or uncircumcised [35,110,194,256,257]. These studies have characterised the penile microbiotas using swabs from either coronal sulcus alone [35,194,229,254], coronal sulcus and glans [256], or glans, coronal sulcus, and penile shaft (hereafter referred as penile skin) [110]. Microbiotas in these studies have been targeted using the hypervariable V1-V3, V3-V5 [257], V3-V4 [35,229,256], V3-V6 [194,254], V4-V6 [110], and V6-V9 [257] of the 16S rRNA gene.

Table 1.3. Summary of findings from penile microbiota studies carried out to date.

Study description/aim	Number of participants and study design	Penile microenvironment and region of 16S rRNA targeted	Study cohort	Main finding	Predominant bacterial taxa in cross-sectional penile specimens	Reference
To investigate the impact of dual-partner antibiotic treatment of symptomatic BV on the vaginal and penile microbiota	21 HIV-negative men (including 4 circumcised) with female partners were recruited, treated twice daily for 7 days with oral metronidazole 400mg and/or 2% clindamycin, and followed for 3 weeks posttreatment	▪ Coronal sulcus and glans ▪ V3-V4 hypervariable region	Australian	▪ Dual-partner treatment of BV had immediate and prolonged effect on CVM composition. ▪ Diversity of the CVMs and prevalence and abundance of BV-associated bacteria significantly reduced following treatment. ▪ In men, the effect was immediate, with significant reduction of BV-associated bacteria, including HIV high-risk anaerobes [35]. However, the beneficial effect had waned at day 28. ▪ There was by recolonisation of the penile microbiotas with BV-associated bacteria.	▪ *Corynebacterium* and BV-associated bacteria (specifically *Finegoldia*, *Peptoniphilus*, and *Prevotella*) were highly prevalent and abundant in baseline cutaneous penile specimens.	Plummer *et al.* (2018) [256]
To assess the association between penile anaerobic bacteria, cytokines, and HIV acquisition in a case-control study	182 uncircumcised men (46 who became HIV-infected (cases) and 136 who remained uninfected (controls) after 2-year follow-up)	▪ Coronal sulcus ▪ V3-V4 hypervariable region	Ugandan	▪ High abundances of anaerobic *Prevotella*, *Dialister*, *Finegoldia*, and *Peptoniphilus* were associated with acquisition of HIV. ▪ Abundances of these anaerobes were also correlated with increased levels of chemoattractant cytokine cytokines, particularly IL-8, which can induce an inflammatory response that recruits HIV target cells to the foreskin.	▪ Not provided.	Liu *et al.* (2017) [35]
To compare the subpreputial microbiota after PrePex device placement to that of uncircumcised men	147 men (including 2 men who underwent a one week PrePex device placement and 145 uncircumcised men)	▪ Coronal sulcus ▪ V3-V6 hypervariable region (454 FLX)	Ugandan	▪ PrePex users had significantly higher total penile bacterial and anaerobic absolute abundance than circumcised men. ▪ PrePex users had higher abundances of anaerobic bacteria, mainly *Porphyromonas*, *Anaerococcus*, *Peptoniphilus*, and *Campylobacter ureolyticus*.	▪ Anaerobic bacteria (*Peptoniphilus*, *Anaerococcus*, *Prevotella*, *Porphyromonas*, and *Finegoldia*) were common in both PrePex users and uncircumcised men.	Liu *et al.* (2016) [255]
To determine the penile skin, male urethral, and	93 HIV-negative men (including 56 circumcised),	▪ Glans, coronal sulcus, and shaft ▪ V4-V6	USA	▪ Penile skin and urethral microbiotas of males with BV-positive female partners were significantly more similar to their partners' CVMs	▪ *Corynebacterium*, *Prevotella*, *Peptoniphilus*, and *Anaerococcus* were the most predominant bacteria in the	Zozaya *et al.* (2016) [110]

vaginal microbiotas of heterosexual couples with and without BV	matched with their female sexual partners who were either negative or positive for BV	hypervariable region		compared to the CVMs of non-partner women with BV. ▪ Penile skin diversities of males with BV-positive partners were significantly higher than that of males with partners without BV.	penile skin microbiotas. ▪ *Corynebacterium*, *Staphylococcus*, and *Gardnerella* were highly abundant in circumcised men. ▪ *Prevotella*, *Porphyromonas*, *Anaerococcus*, *Streptococcus*, and *Finegoldia* were highly abundant in uncircumcised men.	
To assess the relationship between penile microbiota of uncircumcised men and Nugent-BV in female partners	165 uncircumcised HIV-negative men who were enrolled in a randomised trial of medical male circumcision for HIV prevention while their female partners were enrolled into a parallel study	▪ Coronal sulcus ▪ V3-V6 hypervariable region	Ugandan	▪ Coronal sulci microbiotas clustered into 7 CSTs (designated 1-7 according to increasing bacterial density). Men in CST4-7 had higher bacterial densities compared to men in CST1-3. ▪ Men with multiple extramarital partners and female partners with Nugent-BV were more likely to be in CST4-7 than their counterparts.	▪ CST1-3 had higher prevalences and abundances of *Corynebacterium*, *L. vaginalis*, and *Staphylococcus* than CST4-7. ▪ CST4-7 had increased prevalences and abundances of BV associated bacteria such as *Porphyromonas*.	Liu *et al.* (2015) [194]
To compare the coronal sulcus microbiota of uncircumcised and circumcised men at enrolment and 1-year follow-up	156 HIV-negative men uncircumcised men (randomised to either immediate circumcision (intervention group) or circumcision delayed for 24 months (control group))	▪ Coronal sulcus ▪ V3-V6 hypervariable region	Ugandan	▪ At baseline, the prevalences of coronal sulcus bacteria were similar between intervention and control groups. ▪ Male circumcision significantly reduced the bacterial load by reducing both the prevalences and abundances of several anaerobic coronal sulcus bacteria. ▪ Aerobic *Kocuria* and facultative anaerobic *Facklamia* became prevalent after circumcision.	▪ *Prevotellaceae*, *Clostridiales Family XI*, unclassified *Clostridiales*, and *Corynebacteriaceae* were the most abundant coronal sulcus bacterial families at baseline. ▪ The most predominant baseline bacteria were: *Prevotella* spp., unclassified *Clostridiales*, and *Corynebacterium* spp. ▪ Other predominant bacteria: *Peptoniphilus*, *Anaerococcus*, *Finegoldia*, *Murdochiella*, *Porphyromonas*, and *Lactobacillus*.	Liu *et al.* (2013) [254]
To characterise the microbiota of the coronal sulcus and urine and determine the microbiota stability over three consecutive	18 adolescents (including 5 circumcised) provided baseline specimens and followed at monthly	▪ Coronal sulcus and urethra ▪ Full-length ▪ V1–V3, V3–V5 and V6–V9 hypervariable regions	USA	▪ Coronal sulci microbiota and *Staphylococcus*, *Mobiluncus*, *Prevotella*, *Dialister*, and *Anaerococcus* were relatively stable over time. ▪ BV-associated taxa such as *Prevotella*, *Atopobium*, *Mobiluncus*, *Megasphaera*, and *Gemella* were detected in coronal microbiotas of both sexually experienced and inexperienced participants. *Pseudomonas* in coronal sulcus	▪ *Corynebacterium*, *Staphylococcus*, and *Anaerococcus* were the most abundant taxa in coronal sulci microbiota. ▪ *Staphylococcus* was highly abundant in circumcised men while *Porphyromonas* and *Prevotella* were abundant in uncircumcised men.	Nelson *et al.* (2012) [257]

months	intervals over a three-month period			microbiota was less abundant than previously reported [229]. ▪ *Streptococcus*, *Lactobacillus*, and *Staphylococcus* were the most abundant bacteria in urine.		
To assess the impact of circumcision on the penile microbiota	12 HIV-negative men were randomised to either immediate circumcision (intervention) or to circumcision delayed for 24 months (controls) and were followed at 6, 12 and 24 months	▪ Coronal sulcus ▪ V3-V4 hypervariable region	Uganda	▪ Anaerobic families, *Clostridiales Family XI* and *Prevotellaceae*, were (p=0.006) were significantly abundant in pre-circumcision coronal sulci samples. ▪ Coronal sulci microbiotas were less heterogeneous after circumcision. ▪ After circumcision bacterial families consisting of predominantly anaerobic genera were significantly decreased whereas predominantly facultative anaerobic genera increased. ▪ Aerobic/facultative anaerobic family *Corynebacteriaceae* and a facultative anaerobic *Staphylococcaceae* were significantly abundant in post-circumcision coronal sulci samples. ▪ *Pseudomonadaceae* and *Oxalobacteraceae* were positively correlated, but negatively correlated with *Clostridiales Family XI*, *Prevotellaceae*, *Corynebacteriaceae*, and *Staphylococcaceae*.	▪ Of the 42 bacterial families identified, *Pseudomonadaceae* and *Oxalobacteraceae* were the most abundant, regardless of the circumcision status.	Price *et al.* (2010) [229]

Abbreviations: BV – bacterial vaginosis, HIV – human immunodeficiency virus, IL – interleukin, CVM – cervicovaginal microbiota, CST – community state type.

In most of these studies [110,194,229,254,256,257], *Corynebacterium* has been identified as a key bacterium in the penile microbiota. For example, two studies that examined the coronal sulcus microbiotas on a U.S. [257] and Ugandan [254] cohort found that *Corynebacterium* was the most and third most abundant taxa, respectively. This finding on *Corynebacterium* dominance has recently been mirrored in studies that utilised coronal sulcus/glans [256] and penile skin samples [110]. *Corynebacterium* is a Gram-negative bacterium that has been identified as a predominant skin colonizer [258]. It is known to prefer high-humidity (moist) and high-temperature niches, e.g., inguinal crease (side of the groin), although dry sites such as the buttocks have also been shown to be enriched with Gram-negative bacteria [258]. The role of *Corynebacterium* remains underappreciated since it is a fastidious bacterium that is difficult to culture [258]. Other predominant taxa identified in the penile microbiota included *Staphylococcus*, *Lactobacillus*, and BV-associated bacteria such as *Prevotella*, *Finegoldia*, *Peptoniphilus*, *Porphyromonas*, and *Gardnerella* [35,110,194,229,254,256,257]. The prevalence and abundance of all these aforementioned taxa appear to vary by the penile site sampled and/or circumcision status. For example, *Prevotella* was more abundant in pre-circumcised/uncircumcised men than in post-circumcised/circumcised men whereas *Staphylococcus* was more abundant in post-circumcised/circumcised men than in pre-circumcised/uncircumcised men [35,110,229,254,257]. *Staphylococcus*, an aerobic bacterium, is a predominant and stable skin commensal [258].

As previously outlined, the human penis has several microenvironments with different epithelia [225,228]. Each microenvironment supports different microbiotas, likely due to the differences in the availability of oxygen and moisture [35,110,194,229,232,254,255,257]. For example, a longitudinal study by Nelson and colleagues (2012) [257] characterising the coronal sulcus and distal urethra microbiotas of young adolescent men found that the two penile sites had different profiles of bacterial communities. *Corynebacterium* and *Staphylococcus* were the most abundant bacteria in coronal sulcus microbiota while *Streptococcus* and *Lactobacillus* were the most abundant bacteria in distal urethra (Fig. 1.9) [257]. Coronal sulci microbiota and *Staphylococcus*, *Mobiluncus*, *Prevotella*, *Dialister*, and *Anaerococcus* were relatively stable over time [257]. Nelson and colleagues (2012) [257] also further found that *Lactobacillus*, *Gardnerella*, *Aerococcus*, *Mycoplasma*, *Ureaplasma*, and *Veillonella* were specific for distal urethra and that the urethra microbiotas were less stable over time compared to coronal sulcus microbiotas.

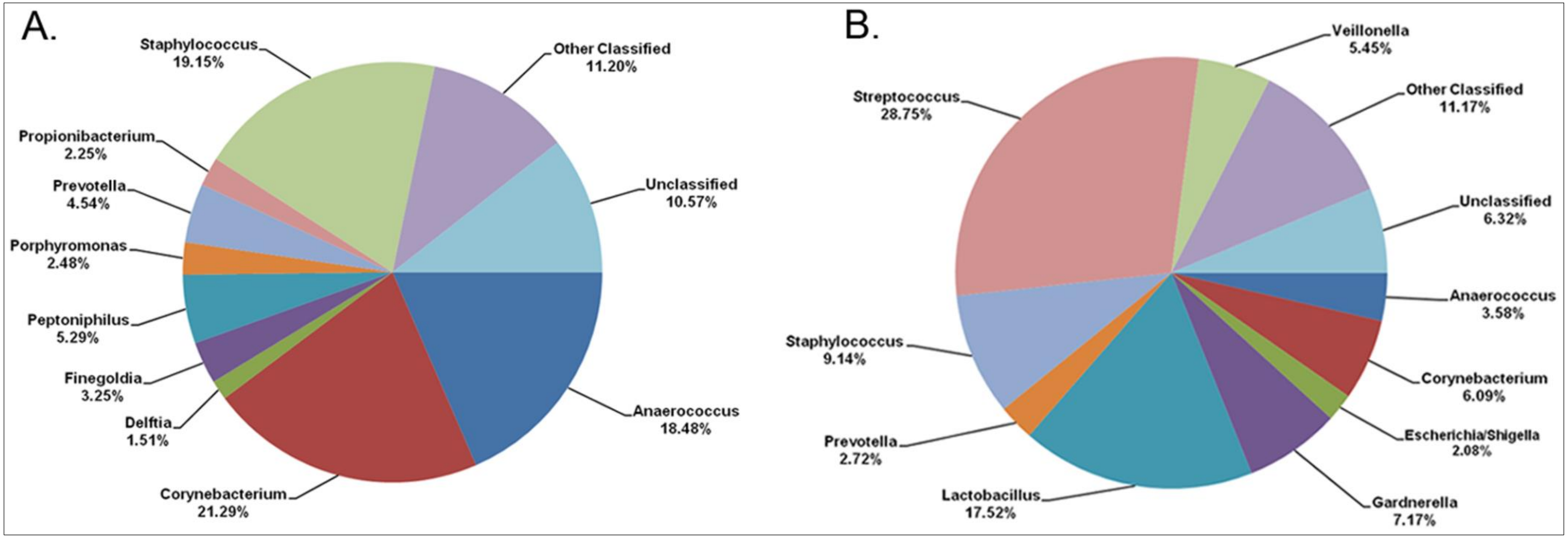


Fig. 1.9. Proportion of major bacterial taxa in a) coronal sulcus and b) urine samples. The figure is taken from Nelson and colleagues (2012) [257].

In uncircumcised men, the subpreputial space has depleted oxygen levels, and therefore tends to be occupied by anaerobic bacteria; while circumcised men have an aerobic niche that is predominantly occupied by aerobic/facultative anaerobic bacteria [110,229,254]. Circumcision has been associated with a lower bacterial load, reduced microbial diversity [254] and a decrease in anaerobic families including *Clostridiales Family XI* and *Prevotellaceae* [229]. *Kocuria* spp. (an aerobe) and *Facklamia* spp. (a facultative anaerobe) were found to be the only bacteria that were exclusively more prevalent in the coronal sulci post-circumcision [254]. Nelson and colleagues (2012) [257] further compared the composition of coronal sulci of young circumcised and uncircumcised adolescent men and found differentially abundant bacteria. They found that the coronal sulci of uncircumcised men had higher proportions of *Porphyromonas* and *Prevotella* and lower proportions of *Staphylococcus* than coronal sulci of circumcised men. The study also detected BV-associated bacteria in coronal sulci of sexually-experienced and inexperienced men [257]. *Corynebacterium* and *Staphylococcus* have been observed to be highly abundant in circumcised men [110,229]. Pathogenic bacteria have also been independently associated with uncircumcised status [232]. A study assessing the bacteria in the coronal sulci and urethra of 315 circumcised and uncircumcised South Indian men [232] observed that pathogenic bacteria such as *Klebsiella* and *Enterococcus* spp. were more likely to occur in coronal sulci and urethra of uncircumcised men than of their counterparts.

It has been postulated that changes on penile microbiotas following circumcision play a mechanistic role in decreased HIV acquisition. This is thought to be due to removal of anoxic environments following circumcision, which results in reduction of anaerobic bacterial families [229] and pathogenic bacteria, e.g., *Klebsiella* sp., *E. coli*, and *P. aeruginosa* [232]. The prepuce, glans, and coronal sulci of uncircumcised men are inhabited mainly by anaerobic bacteria that include *Prevotella*, *Finegoldia*, *Porphyromonas*, *Peptoniphilus*, and *Anaerococcus* [35,110,194,229,254-257]. The predominance of most of these bacteria appears to be enhanced by penile anaerobiosis. A study comparing the subpreputial microbiota of 2 men 7 days after PrePex circumcision device placement to that of 145 uncircumcised men in Uganda observed that the PrePex device facilitated anaerobic bacteria overgrowth, possibly by causing low oxygen tension in the subpreputial space [255]. Compared to uncircumcised men, PrePex users were found to have 6.5-fold higher absolute abundance of the total penile anaerobes, among which comprised *Porphyromonas*, *Anaerococcus*, *Peptoniphilus*, *Campylobacter ureolyticus* [255]. In uncircumcised men, the higher loads of anaerobic bacteria create proinflammatory milieus that activate Langerhans cells to present the HIV to macrophages, dendritic cells, and T-cells immune cells [229,259]. A 2-year case-control study on 182 uncircumcised heterosexual Ugandan men [35] found that 10-fold increased abundances of *Finegoldia*, *Peptoniphilus*, *Prevotella*, and *Dialister* on coronal sulci were associated with 54-63% risk of HIV seroconversion. The study also demonstrated a correlation between abundances of these anaerobes and elevated levels of chemokines, e.g., IL-8 [35]. Levels of IL-8 gradually decline following male circumcision [259]. By reducing the diversity and density of anaerobic bacteria, and proinflammatory responses, male circumcision reduces susceptibility to STIs, notably HIV.

Male circumcision has been strongly associated with reduced prevalent HPV infection [234,236,240,241], including oncogenic and multiple HPV infections [239]. Unlike in HIV, the role of male circumcision in the natural history of HPV infection is certainly not clear, as systematic reviews and meta-analyses [240,241] have found the relationship of male circumcision with genital HPV incidence and remission to be inconsistent. Although a systematic review and meta-analysis by Larke and colleagues (2011) [241] found that male circumcision was weakly associated with both reduced acquisition and increased clearance of genital HPV, a recent systematic review and meta-analysis by Zhu and colleagues (2017) [240] found no association. Currently, there are no published reports on the association between penile microbiota and HPV. Information on penile microbiotas of South African men is unavailable.

1.5 Genital microbiome sharing between heterosexual couples

Lately, there has been considerable discussion regarding the genital microbiome of couples [106,110,256] and recognition that BV-associated bacteria may be sexually transmitted between partners [106,108,109,193,194,256]. Emerging culture-independent studies are revealing that the human penis, particularly of uncircumcised men, can be a reservoir of BV-associated bacteria [35,110,229,254,256] and other bacterial pathogens [232], which are often reduced by male circumcision [110,229,254]. Medical male circumcision has also been shown to drastically reduce the risk of genital ulceration, trichomoniasis [234], cervical cancer, [236], and BV in female sexual partners [234,260]. The penile bacteria may perhaps influence BV status in the female partner [256]. Furthermore, several studies [99,100,108] suggest that persistent or recurrent BV in women could be primarily due to heterosexual exposure to BV-associated bacteria. Although many studies have been published on the human genital microbiome, molecular and microbiological evidence that definitively demonstrates the sexual exchange of BV-associated bacteria is limited.

Evidence of shared microbiome between partner pairs has been observed in urogenital microbiome [106,110,193,194,256]. Studies on this topic together with the role of male partner in shaping the CVMs are scarce [105,106,193]. The few available studies have documented heterosexual transmission of BV-associated bacteria [110,193,194]. Eren and colleagues (2011) [193] demonstrated that monogamous sexual partners have a higher likelihood of harbouring the same oligotypes of the important BV-associated *G. vaginalis*. Zozaya and colleagues (2016) [110] observed that the penile and urethral microbiotas of males, whose partners were positive for BV, were more similar to the CVMs of their female partners than to the CVMs of non-partner women with BV, regardless of the circumcision status of the male partner. This observation was absent in males whose female sexual partners were BV-negative [110]. BV-associated bacteria, mainly *Gardnerella*, *Aerococcus*, *Anaerococcus*, *Sneathia*, *Leptotrichia*, *Eggerthella*, *Dialister*, and *M. hominis* were concordant between BV-positive women and their male partners [110]. In a Ugandan cohort, Liu and colleagues (2015) [194] observed that uncircumcised men with penile microbiotas dominated by BV-associated bacteria were more likely to have BV-positive female sexual partners. Moreover, having multiple female sexual partners was associated with penile microbiotas that had diversity associated with BV [194]. This suggested that sexual intercourse directly influences the genital microbiome. One study of the genital tract microbiotas of 17 infertile couples showed significant shifts in vaginal microbiotas after sexual intercourse with male partner. Importantly, these changes included bacterial species known to be associated with BV [105]. Mändar and co-workers (2015) showed that the microbiome of semen and the vagina are in association and that a single episode of penile-vaginal intercourse without condom use resulted in a significant reduction of vaginal lactobacilli [106], although this was contradictory to a previous study that showed that only the densities of *E. coli* were significantly increased [181]. Hence, more studies are still needed to characterise genital microbiome of couples.

Recent and high frequency of unprotected receptive vaginal sex has a very strong influence on the female genital tract health and may increase the risk for BV acquisition and recurrence [39,99,100,109]. This is because of the direct impact of male penile, urethral, and semen microbiome on the female CVMs [106,110] following heterosexual transmission of bacteria, including BV-associated bacteria [115,194] and depletion of *Lactobacillus* [105]. One study observed that immediately after unprotected vaginal-penile sexual intercourse, there was significant reduction of *L. crispatus* [105] while a previous study [181] had indicated that the bacterial densities for most bacteria (including lactobacilli) were similar except for *E. coli* whose concentration was significantly increased. A study on the effect of sexual debut on vaginal microbiota of young North American women [108] found that women who engaged in sexual activity were more likely to have increased numbers of *G. vaginalis* than virgins. Furthermore, the study observed that there was neither loss of colonisation with *Lactobacillus* spp. nor gain in colonisation with other BV-associated bacteria [108]. A current longitudinal study [107] showed that although sexual intercourse did not affect the stability of CVMs, it was associated with increased diversity of *G. vaginalis*. A longitudinal study assessing the risk factors of BV among North American women aged 18-30 years associated incidental BV with sexual intercourse with uncircumcised men and receptive anal sex prior to penile-vaginal sexual intercourse [95]. Although there is epidemiological evidence that sexual exposure increases the risk of BV [98,99] including its recurrence [100], BV is not considered as an STI [99,261]. A systematic review and meta-analysis [99] reported that having new and multiple male sexual partners increased the risk of BV by 1.6-fold while history of female sexual partners increased the risk in treated female partner by 2-fold. It was further observed that consistent condom use is associated with reduced incident and recurrent BV [99].

Currently, BV remains a microbiological enigma in the human reproductive health and presents a treatment challenge due to its complex polymicrobial nature. According to an observational evidence based on five out of six RCTs with different methodologies [261], male partner treatment of BV with antibiotics does not have a beneficial effect in reducing the risk of BV in the female sexual partner. However, findings of a pilot study on the effect of dual partner treatment on BV (i.e., treatment of BV-positive female and her sexual male partner) in 21 Australian couples [256] points to a possible beneficial effect of treatment of the male sexual partner in preventing recurrence of BV in the female partner. This study observed an immediate and prolonged decrease in predominance of BV-associated bacteria, reduction in bacterial diversity, and increased predominance of *L. iners* in the CVMs three weeks posttreatment [256]. This study also showed that antibiotics for treatment of BV had a short-term effect on the cutaneous penile microbiota since BV-associated bacteria re-emerge shortly after treatment [256]. Most of the cutaneous penile microbiotas were recolonised with BV-associated bacteria three weeks posttreatment [256]. The lack of beneficial effect in reducing the risk of BV in the female sexual partner after male partner treatment (as observed in several RCTs) coupled with its short-term beneficial effect in the treated male partner, is one of the primary reasons why BV is not regarded as an STI, despite its similar epidemiology to STIs [99]. Moreover, there is no single bacterial pathogen that has been acknowledged to be responsible for its aetiology [99]. A critical review of the available literature on mechanisms and epidemiology of BV with regards to sexual behaviour suggested that BV should be considered a sexually enhanced disease (SED) [111]. This is based on two pathogenetic mechanisms that are ultimately dependent on sexual intercourse: 1) unprotected sexual intercourse alkalinises the vaginal microenvironment consequently promoting a decline of lactobacilli and emergence of BV-like communities, and 2) unprotected and protected sexual intercourse may cause mechanical transfer of perineal enteric BV-associated bacteria to the vaginal [111]. The aetiology of BV is poorly understood and therefore studies evaluating the genital microbiome sharing between couples would enhance our understanding of the pathogenesis of BV acquisition and recurrence.

1.6 Problem statement

The main goal of this research study was to characterise the genital microbiotas of Black South African women and men and explore their associations with HPV and HIV infections. In particular, the CVMs of reproductive-age women were profiled and associated with demographic, sociobehavioural, and clinical information. Possible biomarkers for HPV were evaluated. The study also predicted the metagenome functions of the CVMs. With regards to men, the study provided an overview of the penile microbiotas and potential biomarkers for HPV and HIV infections. BV-associated bacteria were also assessed in the penile microbiotas. The motivation for this genital microbiota study included the following:

There is scarcity of data on CVMs of reproductive-age Black African women. Until recently, majority of the CVM studies were being performed on Western populations and those of European ancestry. These studies [33,34,39,40,135-141,143] have identified between 3-8 CSTs, some of which have not reported in other cohorts. This is an indication that the CVMs are population-specific and may not yet be completely characterised. It is unclear why BV, HPV, and HIV are endemic among reproductive-age Black South African women. The reasons for this, including biological factors, remain largely unexplored and unexplained yet it is known that the composition and function of the CVMs have a profound effect on women's reproductive health [122]. *Lactobacillus*-deficient CVMs, which are highly prevalent among Black women, have been associated with BV, which may play a role in the natural history of HPV. Thus, this study sought to examine the CVM composition and function, and associations of the CVMs with prevalent HPV among the Black South African cohort.

The interest in penile microbiome is still new and understudied. To date, there are very few publications on the human penile microbiome. Research continues to demonstrate that bacteria in the human penile microbiome can be sexually transmitted to the female genital tract [110,193,194,256]. Owing to the to the fact that unprotected sexual intercourse influences the CVMs [99,100,107] and may cause transition to communities that are associated with BV [105], the present study also sought to characterise the penile microbiotas in relation to CVMs. Whether the risk of infection of STIs is affected by male circumcision or penile microbiotas has not been adequately examined. Thus, the penile microbiota and its associations with STIs present an interesting area of research. In South Africa, information on the penile microbiotas of men is unavailable. Traditional male circumcision is common in South Africa and may not offer the same level of protection against STIs as medical circumcision [233,234,245,247,251,262], probably due to differences in penile bacterial communities and immune responses [35,232,259]. Information of how penile microbiotas of traditionally circumcised men compares to those of medically circumcised men remains unknown. Specific members of microbiotas, e.g., *Sneathia* in the female genital tract, are known to predict certain phenotypes of disorders and diseases, e.g., high-risk HPV infection and SIL in women. Such information on potential biomarkers for infections, specifically HPV and HIV, is entirely unknown for penile microbiotas.

To address the objectives of this study, the genital bacterial communities of Black South African women and men were profiled by deep sequencing of the partial 16S rRNA gene and analysed by bioinformatics and statistical tools. Deep sequencing technology was used because of its unmatched potential to allow rapid and novel characterisation of CVMs at an unprecedented detail [38,54,136,162]. Knowledge gained from this study could be used to develop new diagnostics and therapeutics, and provide approaches aimed at improving STI and BV management not only in South Africa, but in the world.

CHAPTER 2

Characterisation of Cervicovaginal Microbiotas of Reproductive-Age Black South African Women with and without Human Papillomavirus Infection using Ion Torrent Sequencing of the V4 Region of the 16S rRNA Gene

2.1 Introduction

In this chapter, the CVMs of 62 reproductive-age South African women, with and without HPV infections, were characterised and associated with demographic, sociobehavioural, and clinical factors.

The female genital tract is regarded as a milieu for microbiota that impacts health and disease [78]. A CVM dominated by *Lactobacillus* spp. is considered to be a “healthy” state [38]. These lactobacilli are thought to have a protective role in preventing genital disease by restricting the growth of non-indigenous organisms, including STIs [33,77,146,147,263]. Gross disruption of the resident CVM causes BV [74,78], the most common vaginal disorder among reproductive-age women [76]. BV has been independently associated with adverse reproductive health outcomes [138,189,205,210,264-266], including increased acquisition of HPV [209,267] that is associated with cervical cancer [12,268]. With the exception of *L. iners*, CVMs colonised with *Lactobacillus* such as *L. crispatus* and *L. gasseri* have been shown to be protective against HPV infection [34,139] while those with BV-associated bacteria and *Lactobacillus* deficiency have been associated with HPV infection [34,215].

It is currently known that differences exist in CVM between healthy women of different ethnic/racial background [38,40]. For instance, a study characterising the CVMs of reproductive-age North American women found different prevalence of *Lactobacillus*-dominated CVMs among Asian (80.2%), White (89.7%), Hispanic (61.9%), and Black (59.6%) women [38]. Most studies have reported that 23-63% of reproductive-age Black women lack appreciable numbers of *Lactobacillus* that are credited for sustaining health [33,38-40,87,136-138]. Among women of non-Black ethnicity, this prevalence seldom reaches 40% [38-40]. However, most CVM studies have focussed on women of Caucasian, Asian, and Hispanic background [37,38,163,179,189]. The precise definition of these CVMs, apart from lacking or having low abundances of *Lactobacillus* [33,38,135,143] remains an interesting subject of discussion. Among the Black South African women with *Lactobacillus*-dominated CVMs, 59-77% of them have *L. iners*-dominated CVMs [87,136]. Of the known common *Lactobacillus* spp. colonising the female genital tract [38], *L. iners* is the least stable [162] and protective against adverse reproductive health outcomes [34,159,163,165].

There is paucity of knowledge on the CVM of Black South African women [136,138]. Differences in the CVM in different ethnic groups provided a justification for studying the CVMs of Black South African women. Further, South African women have a high prevalence of both BV [76,84-87] and STIs [6], including HPV [16] and HIV [18].

Here, we assessed the CVMs by deep sequencing the V4 hypervariable region of the 16S rRNA bacterial marker gene using the Ion Torrent Personal Genome Machine (PGM).

The Ion Torrent platform was selected for this study as it represented a potentially more cost effective and rapid technology than the other next generation sequencing platforms such as Illumina [269], particularly due to the lower initial purchase costs of the platform. The availability of an Ion Torrent PGM at the Central Analytical Facility (CAF) at the Stellenbosch University in South Africa presented a unique opportunity to evaluate this technology that had not been extensively investigated for microbiome analysis.

2.1.1 Aim and objectives

The aim of this chapter was to investigate the baseline structure of CVMs in HIV-seronegative Black South African women aged 18-44 years and determine their associations with the participants' metadata. Specific objectives included:

i. To characterise the CVMs of sexually active HIV-negative Black South African women by Ion Torrent sequencing of barcoded V4 16S rRNA amplicons,
ii. To assess the association of the established CSTs with demographic, sociobehavioural, and clinical data of these Black South African women, and
iii. To determine co-occurrence and co-exclusion bacterial patterns in the CVMs.

2.2 Materials and Methods

2.2.1 Ethics statement

This study was approved by the Human Research Ethics Committee (HREC) of the University of Cape Town (HREC references 258/2006 and 580/2014).

2.2.2 Study samples, study design and study population characteristics

This was a retrospective, cross-sectional study based on data and baseline cervical DNA samples from a study examining the natural history of genital HPV among Black heterosexual couples in Gugulethu, Cape Town, South Africa [270]. Details of enrolment and recruitment for the HPV natural history study have been described previously [270]. In brief, participants in the HPV Couples Cohort Study were recruited at the Manyanani Clinic in Empilisweni Center in Gugulethu and followed at 6-monthly interval for 24 months [270]. All study participants had provided written informed consent. The concordance of genital HPV infection in these couples together with the impact of HIV coinfection on prevalent HPV and HPV genotype-specific concordance were assessed [270].

Cervical samples were collected using Digene cervical samplers and stored in Digene specimen transport medium (STM, Qiagen, USA) at -80°C. Nucleic acids were extracted from cervical samples as previously described by Mbulawa and co-workers (2009) [270] using the MagNA Pure Compact System and the MagNA Pure Compact Nucleic Acid Isolation kit (Roche Molecular Diagnostics, Mannheim, Germany). HPV genotypes were detected by Mbulawa and Co-workers (2009) [270] using the Roche Linear Array HPV genotyping test (Roche Molecular Diagnostics, Mannheim, Germany). This test detects 37 HPV genotypes, including the oncogenic high-risk HPV types HPV−16, −18, −31, −33, −35, −39, −45, −51, −52, −56, −58 and −59; probable oncogenic high-risk HPV types HPV−26, −53, −66, −67, −68, −70, −73 and −82; and non-oncogenic low-risk HPV types HPV-6, −11, −40, 42, −54, −55, −61, −62, −64, −69, −71, −72, −81, −83, −84, −89 (HPV-CP6108) and –IS39. Only samples with positive human β-globin results were included in this study. BV was incidentally identified on Pap smears by any shift in bacterial flora suggestive of BV and/or the presence of *Gardnerella*.

To be eligible for the present study, only cervical specimens from HIV-seronegative women aged 18-44 years were considered. These samples had to have information on the HPV status and sufficient volume for microbiota analysis (≥15.0 µl of the extracted DNA). Exclusion criteria included self-reported menstruation at the time of sampling, and HIV-seropositivity. Demographic, sociobehavioural and medical information were abstracted from the HPV Couples Cohort Study [270].

The demographic, sexual and smoking, behavioural and clinical information of the 62 heterosexual Black South African women included in this study are summarised in Table 2.1. All the women were sexually active, with the majority (18/25 (72%)) of hormonal contraceptives users being on Depo-Provera. Five women (20%) were on norethisterone enanthate. Both norethisterone enanthate and Depo-Provera are injectable progestin contraceptives. Very few women (2/25 (8%)) were on oral contraceptives and the identity (oestrogen or progestin or combination) was unknown.

Of the HPV-infected women (37.1% (23/62)), the majority (65.2% (15/23)) were infected with a single HPV genotype. When stratified by HPV oncogenicity, 34.8% (8/23) and 78.3% (18/23) of the HPV-infected women were infected with low-risk and high-risk types, respectively. Only 13% (3/23) were infected with both low-risk and high-risk HPV types.

Thirteen of the women (22.0%) had abnormal cervical cytology: 6.8% with ASCUS, 11.9% with LSILs, and 3.4% with HSILs. A few women had experienced vaginal discharge and genital ulceration in the last six months, 16.1% and 3.2% respectively. Twenty two women (35.5%) were positive for BV. A majority of the women (79.0%) had never smoked cigarettes.

Table 2.1. Baseline demographic, sociobehavioural and clinical information of the 62 heterosexual Black South African women.

Characteristic	**All participants**
	(N = 62)
Age (years)	34.5 (25.8-39.0)
Age at sexual debut (years)^	18.0 (17.0-18.8)
Lifetime number of sexual partners^	2.0 (2.0-4.0)
Number of sex acts with study partner in last month^	2.0 (2.0-4.0)
Current use of hormonal contraceptives* (% (n/N))	
No	55.4 (31/56)
Yes	44.6 (25/56)
HPV (% (n/N))	
Negative	62.9 (39/62)
Positive	37.1 (23/62)
High-risk (% (n/N))	
Negative	71.0 (44/62)
Positive	29.0 (18/62)
Cervical cytology (% (n/N))	
Normal	78.0 (46/59)
ASCUS	6.8 (4/59)
LSIL	11.9 (7/59)
HSIL	3.4 (2/59)
Experienced vaginal discharge in last 6 months (% (n/N))	
No	83.9 (52/62)
Yes	16.1 (10/62)
Experienced genital ulceration in last 6 months (% (n/N))	
No	96.8 (60/62)
Yes	3.2 (2/62)
Incidental BV on Papanicolaou smear (% (n/N))	
No	64.5 (40 /62)
Yes	35.5 (22/62)
Cigarette use (% (n/N))	
Never smoked	79.0 (49/62)
Ex-smoker	1.6 (1/62)
Current smoker	19.4 (12/62)

Abbreviations: HPV – human papillomavirus, ASCUS – atypical cells of undetermined significance, LSIL – low-grade squamous intraepithelial lesion, HSIL – high-grade squamous intraepithelial lesion, BV – bacterial vaginosis.
Continuous variables are expressed as medians with interquartile ranges (IQRs, at 25th and 75th percentiles).
^Data was not available on the age at sexual debut for two women, lifetime number of sexual partners of two women and number of sexual acts with study partner in the last month of six women.
*The hormonal contraceptives included oral pill, norethisterone enanthate and Depo-Provera.

2.2.3 Bacterial V4 16S rRNA library preparation and sequencing

The hypervariable V4 region of the 16S rRNA gene was amplified using the universal PCR primers 515f (5’-GTGCCAGCMGCCGCGGTAA-3’) and 806r (5’-GGACTACHVGGGTWTCTAAT-3’) [271]. Each PCR contained 1x *Ex Taq* buffer (Takara Bio Inc., Shiga, Japan), 0.025 U *Ex Taq* polymerase, 0.8 mM dNTP mixture, 0.56 mg/ml BSA, 400 nM each primer and 100 ng cervical DNA. Each sample was amplified in 3 replicate reactions. PCR conditions were 98°C for 2 minutes, followed by 30 cycles of 98°C for 20 seconds, 50°C for 30 seconds and 72°C for 45 seconds, and a final elongation step at 72°C for 10 minutes. The triplicate samples were pooled and purified using the Agencourt AMPure XP system (Beckman Coulter, Beverly, MA, USA) according to the manufacturer’s instructions. Amplicon sizes were confirmed by electrophoresis on 1.5% TBE agarose gels and imaging with the UVT GelDoc-It™ system. The amplicons were quantified using the

Quant-iT® PicoGreen dsDNA assay (Thermo Fisher Scientific, Waltham, MA, USA) with FLUOstar OPTIMA Multi-Mode Micro Plate Reader (BMG Labtech, Ortenberg, Baden-Württemberg, Germany). Sequencing libraries were prepared using the KAPA Library Preparation kit and barcoded Adaptor kits for Ion Torrent™ platforms (KAPA Biosystems, Wilmington, MA, USA). Barcoded amplicons were pooled (24 samples per pool) in equimolar amounts and the final library sizes and concentrations assessed on a Bioanalyzer High Sensitivity Chip (Agilent Technologies, Santa Clara, CA, USA). Sequencing was performed on the Ion Torrent Personal Genome Machine (PGM) (Life Technologies, Beverly, MA, USA) at the CAF at Stellenbosch University (Stellenbosch, South Africa).

2.2.4 V4 16S rRNA gene amplicon sequence data analyses

The qualities of the raw sequenced reads were visually inspected using FastQC v0.11.2 [272]. QIIME v1.8.0 [273] with imported UPARSE (usearch7.0.1090) [274], was used to analyse and interpret the nucleotide sequence data from the CVMs. In the initial sequence pass, reads were quality-filtered and demultiplexed in QIIME using the user-defined parameters in Supplemental Table 2.1. Briefly, reads with lengths outside the 200-400-bp range, with a quality score of less than 25 (sliding window 50) and without barcodes or any mismatches in the barcode sequences were discarded. A second quality filter was performed in UPARSE with user-defined parameters (Supplemental Table 2.1). Sequences were dereplicated followed by abundance sorting and discarding singletons. OTU clustering was performed by UPARSE-OTU method that uses a greedy clustering algorithm, with binning of reads with 97% pairwise identity. This step is performed simultaneously with representative sequence picking and *de novo* chimera filtering. Representative sequences from each unique OTU cluster were picked using abundance algorithm. Additional chimeras were removed by UCHIME algorithm [275]. This was done against the Greengenes database (gg13_8 Release) [53]. Taxonomy was assigned using the Ribosomal Database Project (RDP) Naïve Bayesian Classifier [276], with the Greengenes database (gg13_8 Release) [53]. Phylogeny was inferred by aligning representative sequences to Greengenes core set using PyNAST [277]. A phylogenetic tree was built using FastTree. Further analyses were done as defined in Supplemental Table 2.1.

Diversity, rarefaction, and sample ordinations were computed in QIIME. Multiple rarefactions at different sequencing depths were performed, and rarefactions (collector's) curves plotted to evaluate the completeness of the sampling efforts. Alpha diversity was computed by chao1, observed_species, Shannon, Simpson, and PD_whole_tree metrics. Beta diversity was estimated using weighted and unweighted UniFrac distances, and Bray-Curtis dissimilarity metric. Reads from each sample were rarefied to 5,000 reads based on rarefaction curves prior to calculating beta diversity. Other biodiversity metrics including

Dominance and Shannon equitability indices were calculated using an in house script in R v3.2.2 (R Core Team 2016). An all-by-all pairwise distance matrix of UniFrac distances were generated and used to hierarchically cluster and ordinate samples. The ordinations were performed using Principal Coordinate Analysis (PCoA).

2.2.5 Identification and comparison of community state types

Hierarchical clustering and heatmap generation were performed in R v3.2.2 (R Core Team 2016). Hierarchical clustering on Bray-Curtis dissimilarity (Vegan package [278]) was done using the average neighbour algorithm.

Further analyses to detect differentially abundant bacterial taxa between the CSTs were performed by Statistical Analyses of Metagenomic (and other) Profiles (STAMP) v2.1.3 [279]. The White's non-parametric t-test (two-sided type) was used for computation. The confidence interval method used was the difference between mean proportions (DP): bootstrap. The thresholds for p-value and effect size were 0.05 and 1.0, respectively. The p-values were corrected to false discovery rate (FDR) q-values (with 0.05 as the threshold for significance) using the Benjamini-Hochberg procedure.

2.2.6 Correlational analyses of 60 cervicovaginal OTUs

To assess bacterial positive and negative relationships, Spearman's correlation of $\log_2$-transfomed counts of all OTU pairs were tested on metagenomeSeq v1.12.1 [280,281] in R v3.2.2, and the OTUs (n=60, 28.0%) with the greatest variance displayed on a correlogram.

2.2.7 Statistical methods

Statistical analyses were carried out using GraphPad Prism v6.01 (San Diego, USA). Mann-Whitney unpaired nonparametric and Chi-square/Fisher's exact tests (with two-tailed p-value) were used to examine the association of continuous and categorical variables with CSTs, as appropriate. The alpha diversity metrics of the three CSTs were compared by Kruskal-Wallis test. Two-group comparison between the alpha diversity of CST, HPV, and BV groups was computed by Mann-Whitney unpaired nonparametric test.

2.3 Results

2.3.1 Data analysis and output

Ion Torrent reads were processed and classified as described in the methods subsection 2.2.4. A total of 1,392,562 high-quality sequencing reads from 62 samples were included in the final analysis with a median of 16,453 reads per sample (range 5,343-235,897 per sample). During the quality-filtering processes more than 70% of the reads were discarded due to low quality (results not shown). Reads discarded included those with a Phred score less than 25 (sliding window 50 bases) and expected errors greater than 0.5. Chimeras, singletons, and reads that failed during sequence alignment were also discarded.

2.3.2 Taxonomic composition of the cervicovaginal microbiotas

Twelve phyla were identified in the CVMs, of which the most abundant were Firmicutes (54.7%), Actinobacteria (16.6%), Bacteriodetes (13.3%), and Fusobacteria (13.0%) (Supplemental Fig. 2.1). Of the 221 OTUs, 152 (68.8%) belonged to these four phyla. The other phyla, in decreasing relative abundances, included Terenicutes (1.2%), Proteobacteria (0.8%), Chlamydiae (0.2%), an unclassified phylum (0.1%), Cyanobacteria (0.1%), Verrucomicrobia (0.1%), Spirochaetes (<0.1%), and SR1 (0.1%).

A heatmap of the relative abundances of the predominant genera (>0.5% relative abundance) and their respective phyla identified in the CVMs of the 62 women is shown in Fig. 2.1. Above the 0.5% threshold, only 9 phyla with 43 genera were observed. *Lactobacillus* (mean relative abundance: 46.0%), *Gardnerella* (19.5%), *Prevotella* (9.7%), and *Sneathia* (9.5%) were the most prevalent and abundant genera in the phyla Firmicutes, Actinobacteria, Bacteriodetes, and Fusobacteria, respectively. Fifty eight (93.5%), 59 (95.2%), and 54 (87.1%) of all the women had *Gardnerella*, *Prevotella*, and *Sneathia*, respectively. Other bacteria such as *Shuttleworthia* (6.7%), *Dialister* (1.2%), *Atopobium* (0.5%), and *Clostridium* (0.1%) commonly found in women with BV [51] were not abundant.

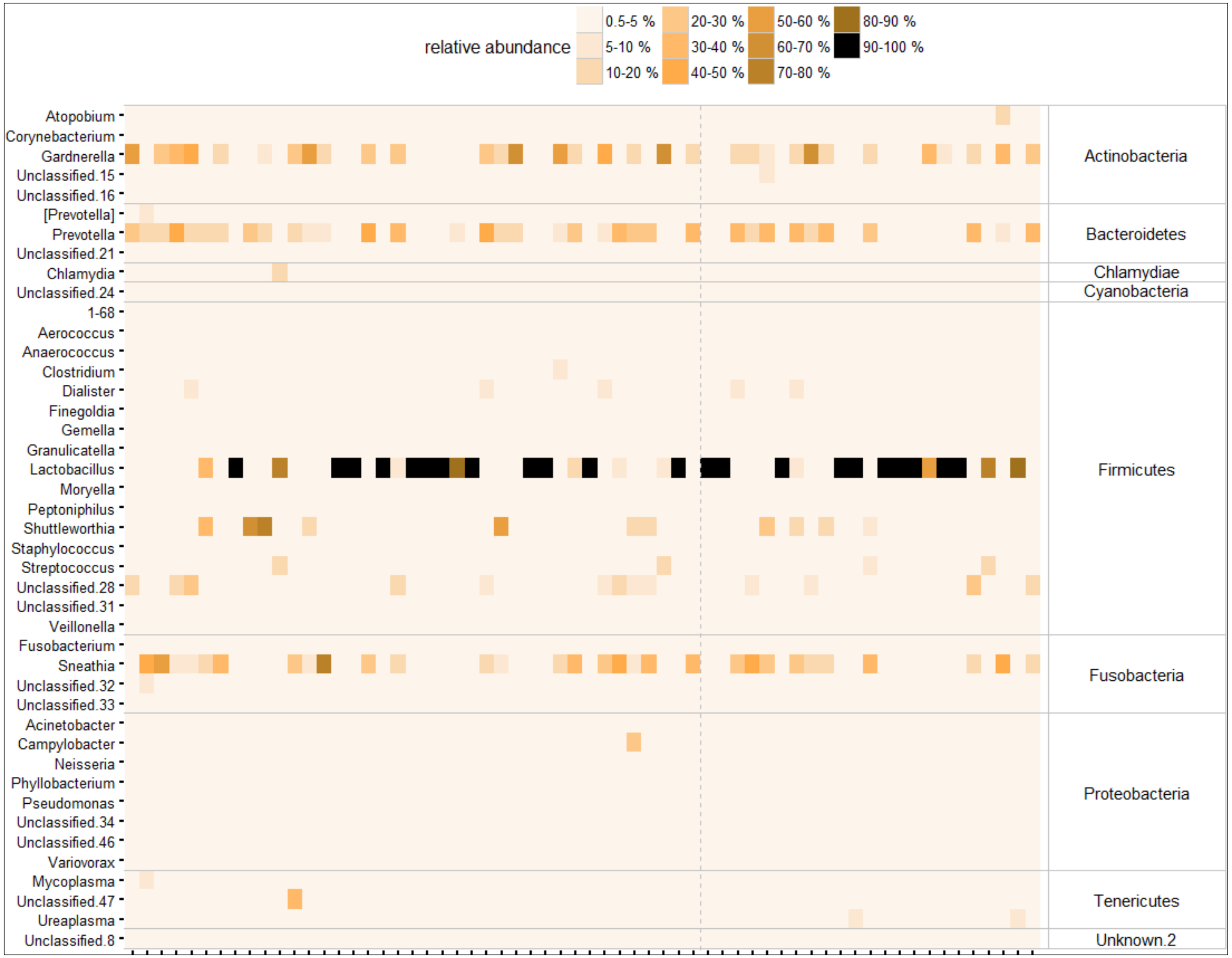


Fig. 2.1. Relative abundances of genera in the 62 women. Only bacteria that occurred at >0.5% relative abundances are shown. Each dot on the x-axis represents a participant. The vertical (dashed) line stratifies the heatmap according to HPV status (left – HPV-negative and right – HPV-positive. The horizontal (solid) lines separate the different phyla, as indicated on the right.

Analysis of the relative abundances of top 11 bacterial genera is shown in Fig. 2.2. There was a trend of an inverse relationship between *Lactobacillus* and the other bacteria, mostly BV-associated bacteria. Twenty seven women (43.5%) had CVMs with ≥50.0% relative abundances of *Lactobacillus*, which were associated with CVM reduced diversity (p<0.0001). As the relative abundance of *Lactobacillus* reduced to ≥30.0% across the women (29/62, 46.8%), the diversity of the CVMs increased (p=0.019).

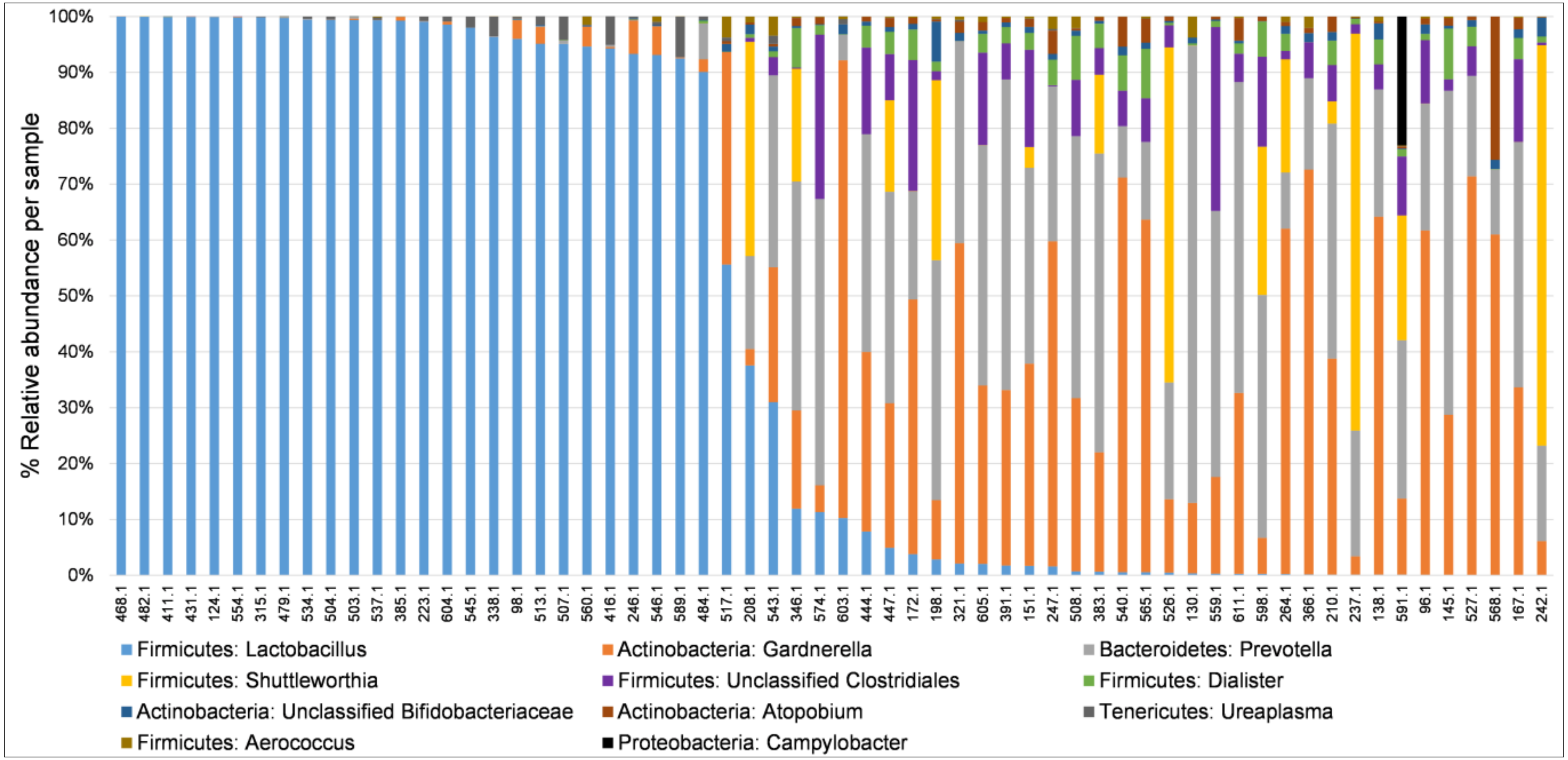


Fig. 2.2. Relative abundances of *Lactobacillus* and top 10 genera in cervicovaginal microbiotas. The samples have been ordered according to reducing relative abundances of *Lactobacillus*. Both the phyla and genera names are stated.

A total of 221 unique OTUs were identified in the 62 CVMs and ranged from 16-104 OTUs per sample. The 30 most abundant OTUs represented 97% of all the reads. The most abundant OTU was classified as *L. iners*, representing 40.6% of all the reads. This OTU was also the most prevalent OTU (100%), with a mean relative abundance of 37.6%. The next most abundant OTUs were classified as *Gardnerella* (9.8% mean relative abundance and 87.1% (54/62) prevalence), *Shuttleworthia* (6.3% mean relative abundance, 46.8% (29/62) prevalence) and *Sneathia* (5.3% mean relative abundance, 80.6% (50/62) prevalence).

A total of 26 different OTUs were classified as *Prevotella,* making this the most diverse genus. Six and eight different potential species or OTUs belonging to the genera *Gardnerella* and *Lactobacillus*, respectively, were detected. Members of lactobacilli were *L. iners*, *L. coleohominis*, *L. mucosae*, *L. ruminis*, and four unclassified *Lactobacillus* spp. It was impossible to assign species name to these *Lactobacillus* spp. due to insufficient taxonomic discrimination by the V4 16S rRNA gene. All the women harboured at least one of the eight *Lactobacillus* spp., with *L. iners* existing in all the CVMs at different relative abundances. *L. iners* accounted for 88.2% of all the lactobacilli reads. Forty seven (75.8%) of CVMs had at least two *Lactobacillus* spp., but at unequal relative abundances.

2.3.3 Cervicovaginal community state types

Hierarchical clustering of the CVMs based on the type and abundances of the bacterial taxa identified three distinct community state types, CSTs I-III (Fig. 2.3). CST I was dominated by *L. iners* and found in 24 women (38.7%). CST II was dominated by an unclassified *Lactobacillus* (*Lactobacillus.4*) and present in only three women (4.8%). Each of the specified bacteria constituted more of the bacterial relative abundance (cut-off of at least 55.3%) in the CVMs. CST III was the most common CST occurring in 35 women (56.5%). This CST was characterised by a diverse and complex array of facultative and strictly anaerobic bacteria (*Gardnerella*, *Prevotella*, *Sneathia*, *Shuttleworthia*, *Clostridium*, *Streptococcus*, *Atopobium*, *Mycoplasma*, *Dialister*, *Ureaplasma*, and a consortium of low-abundant bacteria) and very low numbers of *Lactobacillus*, including *L. iners* and three unclassified *Lactobacillus* (*Lactobacillus.1*, *Lactobacillus.3*, and *Lactobacillus.4*). While there was a continuum of relative abundances of the bacterial taxa in CST III, four sub-clusters were evident (Fig. 2.3), three sub-clusters were dominated by *Shuttleworthia* (n=4 women, 11.4%), *Gardnerella* (n=8, 22.9%), and *Sneathia* (n=5, 14.3%), and one with mixed taxa (n=18, 51.4%).

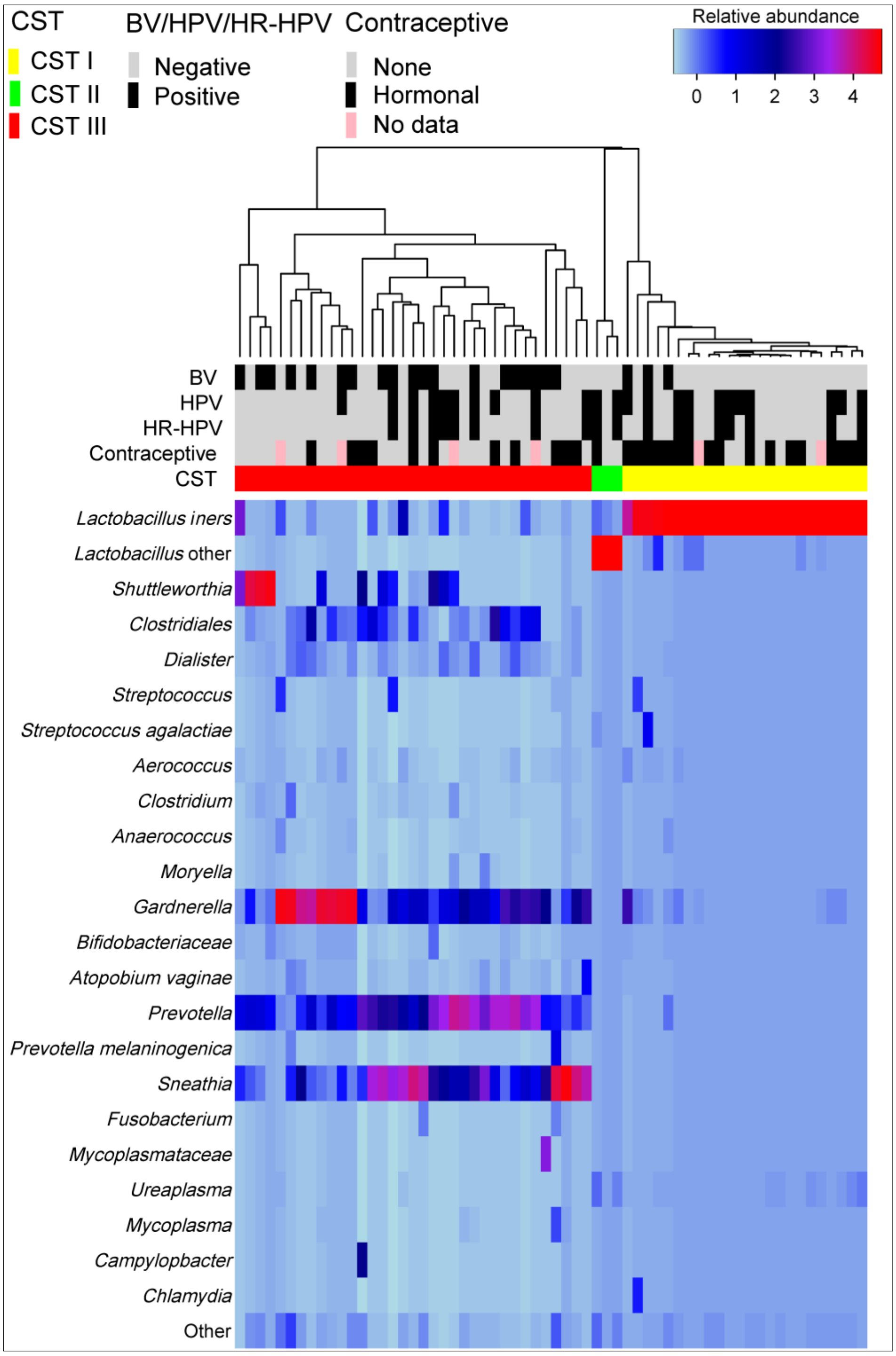


Fig. 2.3. Heatmap of the relative abundances of bacterial taxa in the cervicovaginal microbiotas (CVMS) of 62 Black South African women. Rows represent the bacterial taxa and columns the samples. The 23 most abundant taxa are displayed, with less abundant taxa grouped as “Other”. The dendrogram depicts the average linkage hierarchical clustering of the CVMs based on Bray-Curtis dissimilarity. The CVM community state types (CSTs), human papillomavirus (HPV) and high-risk human papillomavirus (HR-HPV) infection status, bacterial vaginosis (BV) findings and contraceptive usage of the women are indicated.

Next, the bacterial features that were differentially abundant between the CSTs were identified using the statistical tool STAMP [279]. Significant differences in the percentage mean proportions (at 0.95 CI) of specific bacterial taxa in the three CSTs (I-III) were found as shown in the STAMP error bar plots in Supplemental Fig. 2.2. *L. iners* and the unclassified *Lactobacillus* sp. were differentially enriched in CST I and CST II, correspondingly (Supplemental Fig. 2.2a). Comparison between CST I and CST III showed that *L. iners* (together with an unclassified *Lactobacillus* sp.) and BV-associated bacteria (*Atopobium*, unclassified *Clostridiales*, *Dialister*, *Gardnerella*, *Prevotella*, *Shuttleworthia*, and *Sneathia*) were enriched in CST I and CST III, correspondingly (Supplemental Fig. 2.2b). Comparison of CST II versus CST III showed that an unclassified *Lactobacillus* sp. and BV-associated bacteria (unclassified *Clostridiales*, *Dialister*, *Gardnerella*, *Prevotella*, and *Sneathia*) were enriched in CST II and CST III, correspondingly (Supplemental Fig. 2.2c). Next, the metadata for the women in each CST were compared.

2.3.4 Comparison of the community state types by participants' metadata

The demographic, sexual, smoking, and clinical characteristics of the women assigned to each of the three CSTs are shown in Table 2.2. The metadata for women in CST I and CST III were compared, while CST II was excluded from statistical comparisons due to the small sample size. The median age, age at sexual debut, number of sex acts with study partner in last month, cervical cytology, vaginal discharge, genital ulceration, and cigarette use were not significantly different between the women in CST I and III. A significantly greater number of women with a CVM from CST I reported hormonal contraceptive use compared to women with CST III (15/22 (68.2%) versus 9/31 (29%), p=0.005). The distribution of the contraceptives in each CST was as follows: Depo-Provera (CST I (50.0%), CST II (33.3%), CST III (19.4%)), norethisterone enanthate (CST I (18.2%), CST II (0.0%), CST III (3.2%)), and oral pill (CST I (0.0%), CST II (0.0%), CST III (6.5%)). Specifically, more women with CST I reported Depo-Provera use compared to women with CST III (11/22 (50.0%) versus 6/31 (19.4%), p=0.035). Oral pill and norethisterone enanthate contraceptive use did not differ between CST I and CST III. Incidental findings of BV on Pap smears was significantly more frequent (p=0.001) in women with CST III than CST I.

There was no significant difference in HPV infection between women in CST I and CST III. Follow-up HPV genotyping data at the 6-month visit was available for 41 (66.1%) women (Supplemental Fig. 2.3). At follow-up, 20 women remained HPV-negative, with six of these women having CST I and 14 diverse (CST III) baseline microbiotas. A total of four women acquired infection with a new HPV genotype and all had diverse baseline microbiotas (CST III). Six women (one CST I, one CST II and four CST III) cleared HPV infection. Infection with the same HPV genotype (persistent HPV infection) was observed in 6 women (3 CST I, 1 CST II and 4 CST III). The remaining 5 women (4 CST I and 1 CST III) had complex HPV infection patterns that could neither be classified as cleared, acquired nor persistent as they had a combination of these categories.

Table 2.2. Clinical, demographic, sociobehavioural, and microbiological characteristics of the women by cervicovaginal community state type.

Characteristic	**CST I**	**CST II**	**CST III**	**p-value**[#]
	(N=24, 38.7%)	**(N=3, 4.8%)**	**(N=35, 56.5%)**	
Age (years)	30.5 (22.3-37.8)	31.0 (30.0-32.0)	35.0 (28.0-40.0)	0.203
HPV infection (% (n/N))				
Any HPV type	45.8 (11/24)	66.7 (2/3)	28.6 (10/35)	0.174
Any high-risk HPV type	33.3 (8/24)	66.7 (2/3)	22.9 (8/35)	0.374
Single HPV infection	25.0 (6/24)	66.7 (2/3)	20.0 (7/35)	0.649
Multiple HPV infection	20.8 (5/24)	0.0 (0/3)	8.6 (3/35)	0.251
HPV status at 6 month visit[+] **(% (n/N))**				
Negative	42.9 (6/14)	0.0 (0/2)	56.0 (14/25)	0.431
Acquired	0.0 (0 /14)	0.0 (0/2)	16.0 (4/25)	0.277
Cleared	7.1 (1/14)	50.0 (1/2)	16.0 (4/25)	0.637
Persisted	21.4 (3/14)	50.0 (1/2)	8.0 (2/25)	0.329
Age at sexual debut (years)[^]	17.0 (16.0-18.0)	18.0 (18.0-19.0)	18.0 (17.0-19.0)	0.289
Lifetime number of sexual partners	2.0 (2.0-3.0)	2.0 (2.0-6.0)	2.0 (2.0-4.0)	0.711
Number of sex acts with study partner in last month[^]	2.0 (2.0-4.0)	2.0 (2.0-4.0)	2.0 (2.0-4.0)	0.445
Current use of hormonal contraceptives* (% (n/N))	68.2 (15/22)	33.3 (1/3)	29.0 (9/31)	**0.005**
Cervical cytology (% (n/N))				
Normal	75.0 (18/24)	33.3 (1/3)	84.4 (27/32)	0.382
ASCUS	4.2 (1/24)	33.3 (1/3)	6.3 (2/32)	1.000
LSIL	16.7 (4/24)	0.0 (0/3)	9.4 (3/32)	0.447
HSIL	4.2 (1/24)	33.3 (1/3)	0.0 (0/32)	0.429
Experienced vaginal discharge in last 6 months (% (n/N))	20.8 (5/24)	0.0 (0/3)	14.3 (5/35)	0.726
Experienced genital ulceration in last 6 months (% (n/N))	4.2 (1/24)	0.0 (0/3)	2.9 (1)	1.000
Incidental BV on Papanicolaou smear (% (n/N))	12.5 (3/24)	0.0 (0/3)	54.3 (19/35)	**0.001**
Cigarette use (% (n/N))				
Never smoked	87.5 (21/24)	100.0 (3/3)	71.4 (25/35)	0.143
Ex-smoker	4.2 (1/24)	0.0 (0/3)	0.0 (0/35)	0.407
Current smoker	8.3 (2/24)	0.0 (0/3)	28.6 (10/35)	0.098

Abbreviations: HPV – human papillomavirus, ASCUS – atypical cells of undetermined significance, LSIL – low-grade squamous intraepithelial lesion, HSIL – high-grade squamous intraepithelial lesion, BV – bacterial vaginosis, CST – community state type, CST I – *L. iners*-dominated cervicovaginal microbiotas, CST II – an unclassified *Lactobacillus* sp.-dominated cervicovaginal microbiotas, CST III – diverse cervicovaginal microbiotas.

[#] p-values are shown for comparison of each variable between CST I and CST III. Associations of continuous variables (expressed as medians with interquartile ranges (IQRs, at 25th and 75th percentiles)) and categorical variables were computed by Mann-Whitney unpaired and Chi-square/Fisher's exact tests, respectively. CST II was excluded from the statistical analyses due to the low sample number (n=3; 4.8%). Significant p-values (<0.05) are shown in **bold**.

^Data was not available on the age at sexual debut for two women (one CST I and one CST III) and number of sexual acts with study partner in the last month for two women (one CST I and one CST III).

[+] Data was available for 41 women only. Five of these women (4 CST I and 1 CST III) had "complex" HPV infection patterns.

*The hormonal contraceptives included oral pill, norethisterone enanthate and Depo-Provera.

2.3.5 Comparison of alpha diversity across CSTs, BV, HPV, and hormonal contraceptive use

Alpha diversity in the CVMs was estimated using a variety of indices, including Simpson, Dominance, Shannon Diversity and Shannon Equitability (Fig. 2.4). A higher Dominance, Shannon, and Shannon Equitability index value, and lower Simpson index value designates a higher alpha diversity.

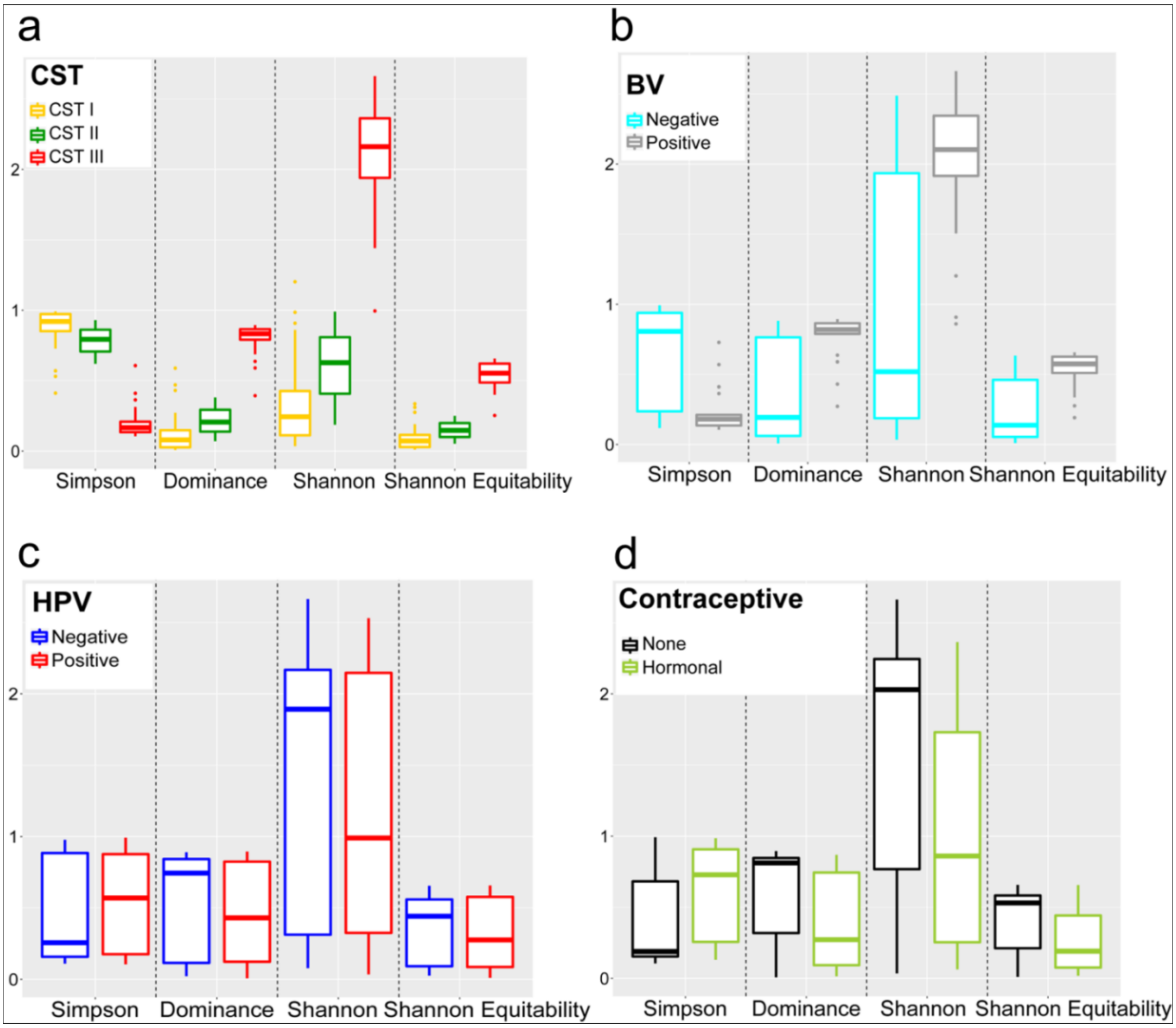


Fig. 2.4. Alpha diversity measures of cervicovaginal microbiotas (CVMs). Comparison of the alpha diversity of the CVMs grouped by a) community state type (CST), b) bacterial vaginosis (BV) status, c) human papillomavirus (HPV) infection status, and d) hormonal contraceptive usage. Each box plot is colour-coded according to the type of CST, HPV, BV, and hormonal contraception. In each plot, the box ranges from the first to the third quartile, with the median represented by the horizontal line. The whiskers extend to the smallest and largest non-outliers and outliers represented by the dots.

When grouped by CST (Fig. 2.4a), CST I (*L. iners*-dominated) and CST III (diverse group) were significantly different for all the alpha diversity indices – Simpson: 0.9 (0.8-1.0) versus 0.2 (0.1-0.2), p<0.0001, Dominance: 0.1 (0.0-0.2) versus 0.8 (0.8-0.9), p<0.0001, Shannon: 0.2 (0.1-0.5) versus 2.2 (1.9-2.4), p<0.0001, and Shannon equitability: 0.1 (0.0-0.1) versus 0.6 (0.5-0.6), p<0.0001, respectively. Thus, bacterial diversity in CST III was significantly higher than CST I. Although not compared with the others, the alpha diversity indices for CST II (with predominance of an unclassified *Lactobacillus*) were as follows –

Simpson: 0.8 (0.6-0.9), Dominance: 0.2 (0.1-0.4), Shannon: 0.6 (0.2-1.0), and Shannon equitability: 0.1 (0.1-0.3).

The alpha diversity of the CVMs of women in the BV-negative and BV-positive groups (Fig. 2.4b) differed significantly, with greater diversity in the BV-positive group (Shannon index 0.5 (0.2-1.9) versus 2.1 (1.8-2.4), p<0.0001, respectively).

When grouped by HPV status (Fig. 2.4c), no significant difference in alpha diversity was observed between the HPV-negative and HPV-positive groups, with Shannon index: 1.9 (0.3-2.2) and 1.0 (0.3-2.2), p=0.483, respectively.

Alpha diversity also varied between hormonal contraceptive users and non-users, with hormonal contraceptive users having a lower microbial diversity than non-users. The Shannon index was 0.9 (0.2-1.9) versus 2.3 (0.6-2.3), p=0.025, respectively (Fig. 2.4d).

When stratified by the type of contraceptive used, the diversities based on Shannon index were statistically different (p=0.022): 0.7 (0.2-1.7), 0.5 (0.2-1.4), 2.3 (2.3-2.4), and 2.0 (0.6-2.3) for Depo-Provera, norethisterone enanthate, oral pill, and non-users, respectively (Supplemental Fig. 2.1). Women on injectable hormonal contraceptives had CVMs with significantly lower microbial diversity than women on oral pills and hormonal contraceptive non-users.

2.3.6 Comparison of beta diversity across CSTs, BV, HPV, and hormonal contraceptive use

The PCoA of weighted UniFrac distances of the 62 CVMs is shown in Fig. 2.5. The CVMs are coloured by CST, BV, HPV and hormonal contraception status of the women. Principal coordinate (PC)-1 accounted for 64% of the variance between samples, PC2 for 10% and PC3 8%.

Beta diversity analysis of the 62 samples showed that each of the established CST (I-III) represented a highly distinct bacterial community (Fig. 2.5a). There was a clear separation of the samples according to the different CSTs. Most of the bacterial communities (37/62, 59.7%) were broadly segregated in the 3-dimensional space due to considerable differences in bacterial composition and relative abundance. All of these communities belonged to CST III. This clustering and separation was supported by the Jackknife replicates that were used to estimate the uncertainty in PCoA plots and hierarchical clustering of the CVM. The majority of the samples with *Lactobacillus*-dominated CVMs (CST I and CST II, n=25, 92.6%) exclusively clustered together in the upper right quadrant (Fig. 2.5a). Two samples from CST I did not cluster with this group likely due to presence of other bacterial taxa, e.g., *Gardnerella*, *Prevotella*, and *Aerococcus*, in these CVMs. These samples were from women with BV.

Comparison of the beta diversity estimates of BV-negative and BV-positive samples is shown in Fig. 2.5b. The 25 samples that clustered together in the upper right quadrant consisted mostly of BV-negative samples (22/25, 88.0%). Samples from BV-positive women were spread over a greater area in the plot due to their high and varying bacterial diversity. The majority of women with BV (19/22, 86.4%) were in CST III.

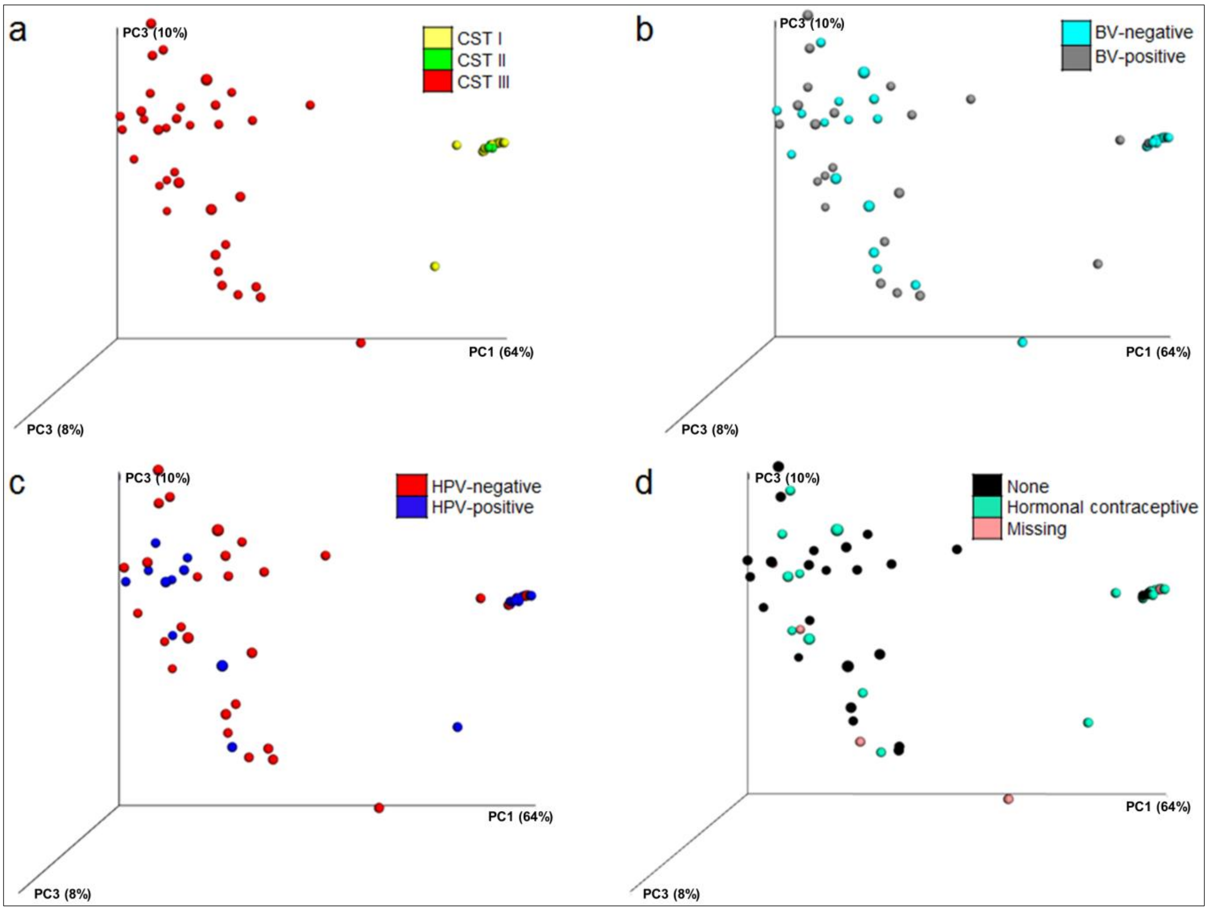


Fig. 2.5. Beta diversity of the cervicovaginal microbiotas (CVMs). Principal Coordinates Analysis (PCoA) plots of the weighted UniFrac distances of the CVMs coloured according to a) community state type (CST), b) bacterial vaginosis (BV) status, c) human papillomavirus (HPV) infection status, and d) hormonal contraception usage. The first three principal coordinate (PC) axes and the percentage variation explained by each (PC1: 64%, PC2: 10%, and PC3: 8%) are shown. Each solid point represents a bacterial community.

Clustering of the samples was independent of the HPV status (Fig. 2.5c). Of the samples that clustered together, 48% (12/25) were HPV-positive. Thus, the weighted UniFrac distances of the CVMs showed that there was no apparent influence of HPV infection.

The majority of the samples from women on hormonal contraception (16/25, 64.0%), mostly Depo-Provera (12/18, 75.0%), clustered together in the upper right quadrant (Fig.

2.5d). These samples had bacterial communities dominated by lactobacilli, the majority of which were *L. iners*-dominated.

2.3.7 Co-occurrence and co-exclusion patterns of cervicovaginal bacterial OTUs

Pairwise correlations were calculated for all pairs of OTUs identified in the CVMs. A correlation matrix of the pairwise correlations between the 60 most abundant cervicovaginal bacterial OTUs is shown in Fig. 2.6.

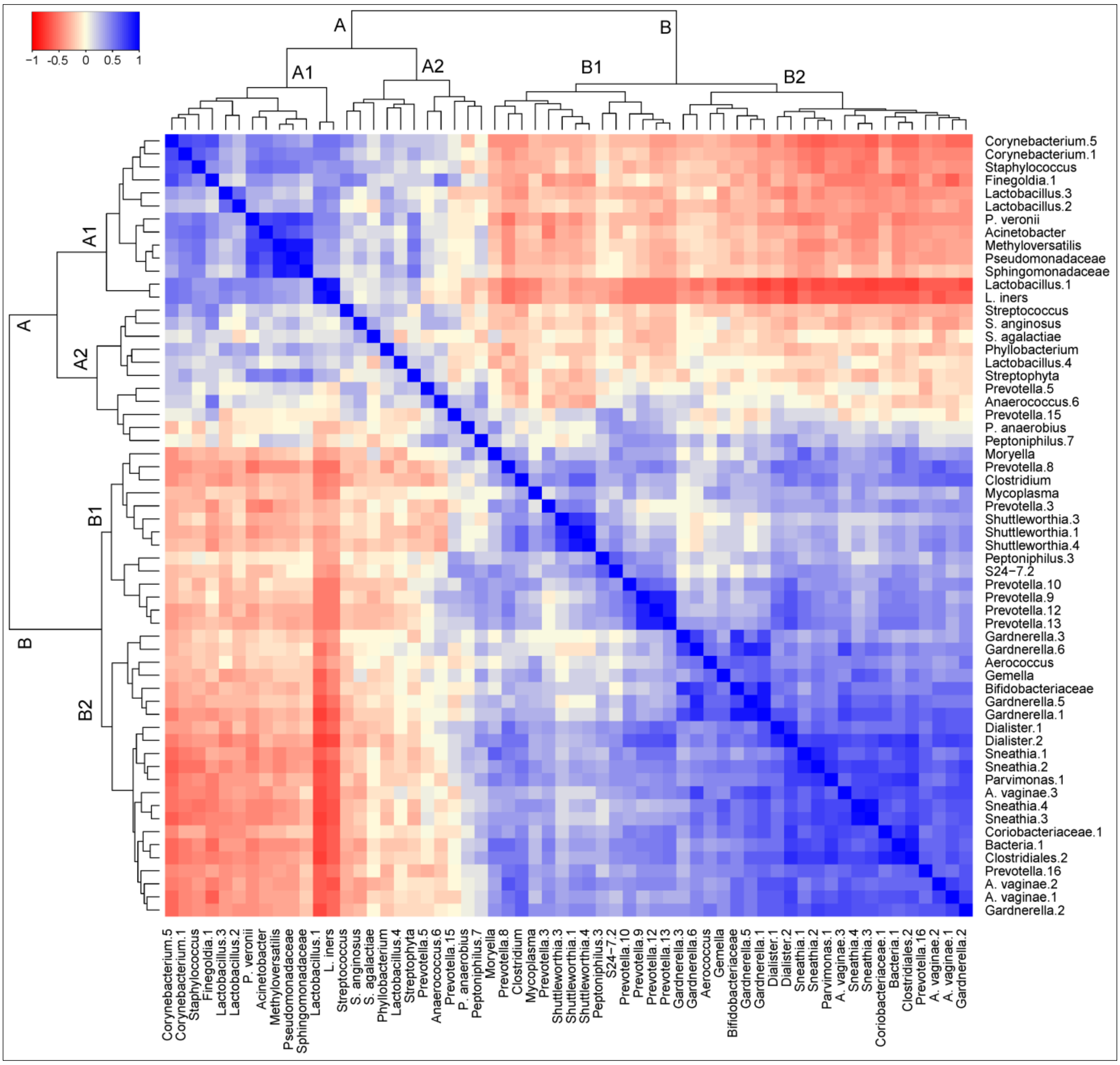


Fig. 2.6. Correlogram of 60 cervicovaginal bacterial OTUs showing co-occurrence and co-exclusion patterns. Spearman's rank correlations between OTU counts were calculated in metagenomeSeq and the samples clustered. The correlation coefficients range from -1 (red; incompatibilities, co-exclusions, or oppositional interactions) to +1 (blue; symbiotic, mutualistic, or co-occurrence interactions.

From the dendrogram, two major bacterial correlation clusters, Cluster-A (mostly with OTUs classified as *Lactobacillus* and *Streptococcus*), and Cluster-B (mostly with BV-associated bacteria), each with two sub-clusters (Cluster-A1 and Cluster-A2; and Cluster-B1 and Cluster-B2) that had different competitive interactions were observed. Cluster-A1 had *L. iners* together with three unclassified *Lactobacillus* OTUs (*Lactobacillus.1*, *Lactobacillus.2*, and *Lactobacillus.3*) and other anaerobic/facultative anaerobic bacteria, e.g., *Staphylococcus* and *Pseudomonas*; while Cluster-A2 had an unclassified *Lactobacillus* (*Lactobacillus.4*) mostly with *Staphylococcus*, *Streptococcus* and two unclassified *Prevotella* species.

Cluster-B had the vast majority of bacteria (n=36, 60.0%), most of which are either strictly anaerobic or facultative anaerobic bacteria. Anaerobes do not require oxygen for growth whereas facultative anaerobes can respire and grow in the presence or absence of oxygen [282]. Cluster-B1 mostly included *Prevotella* (six OTUs), and *Shuttleworthia* (3 OTUs), while Cluster-B2 constituted *Gardnerella* (5 OTUs), *Sneathia* (4 OTUs), *Atopobium* (3 OTUs), *Dialister* (2 OTUs), and an unclassified bacteria among other BV-associated bacteria.

Even though slight negative correlations existed between Cluster-A1 and Cluster A2, majority of the bacterial OTUs in these clusters had positive correlations. A significant majority of bacteria in these clusters, e.g., *Lactobacillus*, had negative correlations with the bacteria, e.g., *Gardnerella* and *Prevotella*, in Cluster-B. There was some overlap in the interaction of bacteria in Cluster-A2 and Cluster-B1; interactions between these clusters appeared minimally low, i.e., very little to moderate. Regardless of the correlation cluster, strong positive correlations were very common between phylogenetically related bacterial OTUs, e.g., *Lactobacillus* spp., but the extent of these interactions varied.

2.4 Discussion

2.4.1 Cervicovaginal microbiotas and community state types in Black South African women

The present study identified three core CSTs in the 62 reproductive-age Black South African women analysed. These included *L. iners*-dominated CVMs (CST I), unclassified *Lactobacillus*-dominated CVMs (CST II), and heterogeneous CVMs (CST III). In the present study the dominant *Lactobacillus* in CST II could not be classified to species level because the V4 16S rRNA region did not allow deeper taxonomic discrimination. Thus, we could not determine if the CST II in our cohort is any of the other common vaginal CSTs dominated by *Lactobacillus* species (*L. crispatus*, *L. gasseri*, and *L. jensenii*), as established by Ravel and colleagues (2011) [38]. Similar results are reported by Roesch and co-workers (2011) [283] who report classification of *Lactobacillus* OTUs up to the genus level only using Ion Torrent PGM sequencing of the V4 region of the 16S rRNA gene to characterise the vaginal microbiome of Brazilian women.

In the present study, *L. iners* was the most prevalent *Lactobacillus* (89%) among the *Lactobacillus*-dominated CVMs. *L. iners* is a key member of the CVM and has been reported in many cultivation independent studies of the CVMs of White, Hispanic, Asian, and Black women [34,38,40,82,87,135,162]. It has also been shown that among the African women colonised by lactobacilli, *L. iners* is the most predominant *Lactobacillus* sp. [33,39,40,136,137,143]. Recently, these findings have been reported in Rwanda [33,138], South Africa [87,136,138], Kenya, Tanzania [138], and Nigeria [137]. The observed predominance of *L. iners* (89%) among the *Lactobacillus* spp. in the current study is up to 30% more than that published from other African cohort studies [33,87,136-138]. These published reports found that among HIV-negative non-pregnant women, *L. iners* accounted for 59-88% of all the *Lactobacillus*-dominated CVMs. The variations in *L. iners* predominance could be attributed to the differences in behaviour and genetics within the Black population [143,284]. In contrast to these findings, *L. crispatus* was found to be the most predominant *Lactobacillus* spp. among White women in North America [38].

This study detected eight *Lactobacillus* spp. (*L. iners*, *L. coleohominis*, *L. mucosae*, *L. ruminis*, and four unclassified *Lactobacillus* spp.) in total. Many of these have been detected in two independent culture-dependent studies of premenopausal South African women with and without BV carried out by Damelin and colleagues (2011) [42] and Pendharkar and colleagues (2013) [148]. These cultivation studies found, however, that *L. crispatus* and *L. jensenii* were the most frequently occurring vaginal lactobacilli [42,148], whereas our study found *L. iners* to be the most prevalent, occurring in all the women (100%). This greatly contrasts with the cultivation studies where the prevalence of *L. iners* was 36% [42] and 75% [148]. This dissimilarity could be due to difficulty in culturing *L. iners* [265] and/or the higher sensitivity of deep sequencing technology [285]. An additional disparity with the cultivation studies was that all the women analysed in this study were colonised by at least one *Lactobacillus* spp. The presence of at least one *Lactobacillus* spp. in virtually all the CVMs, irrespective of the type of CSTs, is in line with published literature [33,38]. Ravel and co-workers (2011) [38] reasoned that this attests that the production of lactic acid is conceivably conserved in all CVMs, thus pinpointing it as the core acidifier. In line with previous literature [40], the present study also confirmed that CVMs with approximately equal dominance of two or more *Lactobacillus* spp. are absent or underrepresented in Black women.

Congruent with previous studies [33,138], the heterogeneous group (CST III) was more diverse, as reflected in the measures of community diversity: richness (number of potential species) and evenness (equality of the distribution of relative abundance of the potential species), than CST I and CST II. CST III had increased richness and evenness of the bacterial communities. The diversity was characterised by multiple bacteria that occurred at low abundance. These results are in agreement with previous findings that found most non-*Lactobacillus*-dominated CVMs are more diverse than *Lactobacillus*-dominated CVMs [33,34,38,135-137]. For example, in a study examining the CVMs of 174 Rwandan female sex workers [33], *Lactobacillus*-deficient CVMs were consistently observed to have higher community richness (a median of 14-18 genera per sample) compared with two *Lactobacillus*-dominated CVMs (with a median of 5 and 6 per sample). *Lactobacillus*-deficient CVMs were also shown to be significantly more even than their counterparts [33]. In our study, we also observed that CST III exhibited intracluster variation and could be further classified into four sub-clusters, depending on bacterial dominance: CST III-*Shuttleworthia*, CST III-*Gardnerella*, CST III-*Sneathia*, and CST III-mixed, which lacked a clear dominance. The interpretation of *Shuttleworthia* should however be considered with caution since it has been noted that *Shuttleworthia* is likely a misclassification and is probably BVAB-1 [286]. Since sequences for BVAB-1 may not be included in Greengenes database, this would result in its misclassification. *Gardnerella*-dominated CVMs have previously been reported among Black South African women [136] and other sub-Saharan Africa women, e.g., African Surinamese and Ghanaian women [143]. The prevalence of CST III was high (57%) in this population that we studied. This is comparable to the prevalence (63%) reported by Anahtar and colleagues (2015) [136] in a culture-independent study among South African women. Compared with White women, a majority of women of African ancestry are known to have diverse and heterogeneous CVMs [38-40,143] that have bacteria usually associated with BV [51,82]. In our study, all BV-positive women were in CST III. The overall prevalence of BV in our cohort was high (36%) and within the range (31-63%) that has been reported among HIV-negative South African women that has been reported among South African women [76,84,87].

L. iners was present in low abundance in all CST III CVMs. Unlike many *Lactobacillus* spp., *L. iners* can occur with BV-associated bacteria [51,138,287] and at times uninterrupted by *Gardnerella* [77]. The detection of *L. iners* in diverse CVMs is not unusual [143]. *L. iners* has been consistently isolated in women with and without vaginal syndromes, and those with intermediate flora [42,82,148] or CSTs transitioning to healthy or dysbiotic states [162]. Some authors have suggested that CVMs vary in resilience and that some lactobacilli, including *L. iners*, have inefficient colonisation resistance to opportunistic and pathogenic bacteria [40,148] or have better tolerant and survival phenotypes even in perturbed milieus [82,148]. There is compelling omics evidence supporting the second explanation. Genomics have shown that the genome of *L. iners* emerged as a result of rapid evolutionary events that endowed it with specialised adaptation capabilities even in dysbiotic milieu [154]. Meta-transcriptomics have strengthened these facts, e.g., *L. iners* is able to differentially express over 10% of its genome in dysbiotic state, which is characterised by elevated expression of cholesterol-dependent cytolysin, mucin and glycerol transport and related metabolic enzymes such as glycerol-3-phosphate dehydrogenase [155]. Moreover, it can utilise glycerol as a carbon source under dysbiotic conditions, producing acids like succinate, which elevate the vaginal pH thereby favouring vaginal colonisation by BV-associated bacteria [155]. In the current study, *L. iners* could either be playing a protective or non-protective role. *L. iners* can predispose women to an aberrant CVM or BV [163] and has been associated with *C. trachomatis* infection [159,165]. Temporal dynamics have demonstrated that *L. iners*-dominated CVM is more likely to transition to non-*Lactobacillus*-dominated CVM than to *Lactobacillus*-dominated CVM [162]. There is a possibility that some of the women with CST III were transitioning to either a healthy or dysbiotic state. *L. iners* can disrupt *G. vaginalis* biofilms [149], thereby restoring a healthy CVM. Alternatively, the diverse CVMs (CST III) could be a variant of healthy CVM that is unreported or very uncommon in other populations.

Among the BV-associated bacteria, *Gardnerella* and *Prevotella* were the most predominant, occurring in at least 94% of the CVMs. *Gardnerella* and *Prevotella* are known to be ubiquitous and mostly predominant in CVMs with BV [33,51,82]. All women in CST II and CST III had *Prevotella*. *Gardnerella* occurred in 85.2% of *Lactobacillus*-dominated CVMs (CST I and CST II) and in all the CVMs in CST III. There was sequence heterogeneity of *Gardnerella* in the CVMs with six OTUs classified as *Gardnerella*. *G. vaginalis* has been observed have different sequence variants [193]. Our study could not assess whether the *Gardnerella* detected were the virulent or non-virulent forms [104,115,118]. These *Gardnerella* OTUs may be sexually transmitted [107,115,193] or from extravaginal reservoirs of vaginal bacteria [109].

2.4.2 Association of hormonal contraception with cervicovaginal Lactobacillus predominance

This study observed that hormonal contraception use was associated with *Lactobacillus*-dominated cervicovaginal microbiotas in reproductive-age black South African women. This extends previous observations. Observational studies have shown that hormonal contraception has a protective effect against BV [92,129,192,195,196]. A meta-analysis of the association of hormonal contraception with BV [195] found that women using hormonal contraceptives had a significantly reduced risk of incident, prevalent and recurrent BV compared to non-users. Hormonal contraception reduces the risk of BV by 10-30% [196]. A culture-independent study documenting the relationship between ethnicity and CVMs of women in Netherlands [143] also found that a significant number of women not using hormonal contraceptives had *G. vaginalis*-containing CVMs that are deficient in *Lactobacillus* and harbour an array of other anaerobic BV-associated bacteria, including *Atopobium*, *Sneathia*, *Leptotrichia*, *Prevotella*, *Megasphaera*, *Mobiluncus*, and BVAB-1. In contrast to this, a study by Anahtar and colleagues (2015) [136] found no association between CVMs and hormonal contraception (mostly progestin) among asymptomatic young South African women. A study by Lennard and colleagues (2018) [87] found that hormonal contraception was independently associated with CST, especially the one with diverse bacterial community which was highly predictive of genital inflammation among young South African women. A recent study by Ferreira and colleagues (2018) [123] on Brazilian cohort found that, although not statistically significant, women on hormonal contraception 12 months prior to the study were more likely to have normal microbiota compared to those without history of hormonal contraception. A large cohort study by Brooks and colleagues (2017) [191] on predominantly African American and Caucasian women found that women on combined oral contraceptives (combination of an oestrogen and a progestin) had significantly higher abundances of H_2O_2-producing *Lactobacillus*, specifically *L. crispatus* and *L. jensenii*, and lower abundances of BV-associated bacteria compared to women who used condoms, depot medroxyprogesterone acetate, or levonorgestrel-releasing intrauterine system. These observations are in line with a previous study on predominantly reproductive-age South African women that inversely associated combined oral contraception with BV [84].

The benefits of hormonal contraception are also seen in women with HPV infections. In a population-based, 5-year longitudinal study to assess the determinants of HPV clearance in 227 HPV-positive Colombian women aged 13-85 years with normal cytology [8], 23% and 7% of HPV infections still persisted after 1 year and 5 years, respectively. Women who had ever used oral contraceptives had faster clearance rates than their counterparts [8]. A non-*Lactobacillus*-dominated CVM colonised with BV-associated bacteria such as *Gardnerella*, *Prevotella*, *Megasphaera*, and *Atopobium* has been identified as risk factor for persistence of HPV infection [157]. Therefore, HPV remission and reduced risk of BV due to contraceptives may be attributable to the impact of hormonal contraceptives on CVMs [195,196]. Although it is known that the type of hormonal contraception may differentially impact the CVM [191], our study grouped all hormonal contraceptive methods together due to the small number of women on contraception (45% (25/56)). In our study, this may not have a pronounced impact on the results since most of the women in our study on hormonal contraception (at least 92% (23/25)) used known progestin contraceptives (Depo-Provera and norethisterone enanthate) while the rest were on either oestrogen or combined (oestrogen and progestin) contraceptives.

A number of possible mechanisms may underlie the reduced BV risk seen in women using hormonal contraceptives. It has been hypothesised that oestrogen may reduce BV risk by increasing the amount of free glycogen available in epithelial cells for conversion by

Lactobacillus into lactic acid [61]. The production of lactic acid by *Lactobacillus* is known to be protective against colonisation by opportunistic and pathogenic bacteria, including BV-associated bacteria [71]. Progestin, on the other hand, has a different biological mechanism for reducing BV risk. Progesterone contraceptives create a systemic hypoestrogenic environment [66], reducing menstruation. *G. vaginalis* is known to sequester iron for growth from compounds such as haemoglobin [288]. Levels of *G. vaginalis* have been observed to increase during menses [173]. Therefore, the reduction of menstruation by progestin reduces the availability of iron-rich haemoglobin thereby preventing establishment of *G. vaginalis* and dysbiotic CVMs. Further, Gajer and co-workers (2012) [162] have observed that during menses the microbiotas lose their stability and may transition to BV-associated microbiotas such as CST III reported herein. Among the common *Lactobacillus* spp. colonising the female genital tract, the relative abundance of *L. iners* has been found not to differ considerably with different amounts of glycogen, even in low glycogen concentrations [61]. This is due to its ability to also source for energy from non-glycogen saccharides [155]. *L. iners* can disrupt *G. vaginalis* biofilms [149]. Progestin-induced amenorrhea neither increases lactobacilli abundance nor affects the abundance of *L. iners* (CST I), but has been associated with significant reduction of total bacterial load and proinflammatory responses [289]. All the progestin-based effects may reduce the transition to disturbed CSTs, hence BV.

2.4.3 Co-occurrence and co-exclusion profiles of bacterial OTUs

The correlational analyses provided evidence for synergism (co-occurrence) and antagonism (co-exclusion) of the cervicovaginal bacterial communities. These ecological patterns have been elucidated to act concurrently to structure the microbiota compositions [290]. Positive correlations suggest niche sharing, similar nutrient requirements and or metabolic co-dependencies. In contrast, negative correlations are indications for niche filtering and or competition for growth nutrients [51,290,291].

In the present study, most BV-associated bacteria in Cluster-B tended to ecologically co-occur and appeared to have negative relationships with bacteria in the other Cluster-A, mostly *Lactobacillus*. The relative abundances of *Lactobacillus* and *Gardnerella* and BV-associated bacteria revealed the inverse relationship of *Lactobacillus* with these bacteria, as observed in other studies [51,77,173]. Co-occurrence of *Gardnerella* with *Prevotella* and other BV-associated bacteria parallels previous findings [38,169]. Moreover, phylogenetically related bacterial phylotypes, e.g., *Prevotella* spp., had strong positive correlations, which can be assumed to be due to the high-level of resource overlap [290]. Variations in positive correlations of more closely phylogenetically-related bacterial species, e.g., the different *Atopobium* spp., or *Gardnerella* spp., illuminates the existence of diverse bacterial genetic profiles at species level [193,292] plausibly with different virulent competencies; these strains have dissimilar phenotypic behaviours in cervicovaginal health and disease [115,292].

The co-occurrence and co-exclusion profiles show that it is implausible that a genital bacterium can ecologically exist and function in isolation [51]. As stated by Ravel and co-workers (2011) [38], the precise relevance of these bacterial interactions remains undefined; therefore, subject to further investigations.

2.4.4 Summary

The CVMs of a majority of reproductive-age HIV-seronegative Black South African women (57%) are not dominated by *Lactobacillus*, the bacteria assumed to constitute a healthy CVM. These women had diverse and heterogeneous bacterial communities that were positively associated with BV. BV in turn has been associated with HIV infection [126]. Several studies have shown that *Lactobacillus*-deficient CVMs are associated with higher HIV and STI acquisition [33,164,165]. Vaginal microbiome may then be a contributing factor to the high burden of HIV and HPV in the Black South African women population. Zhou and colleagues (2007) [40] speculated that differences in the CVMs of Caucasian and Black women might partly explain the racial differences in the risk of BV and STIs, with Black women being more susceptible to BV and STIs than Caucasian women since they have non-*Lactobacillus*-dominated CVMs. Although it is known that genetic factors have influence on CVMs [179,180], we did not examine the genetic background of the women in our cohort to ascertain whether they were truly of African descent and did not have genes from other races.

Hormonal contraception, and in particular progestin-based contraception, was associated with *L. iners*-dominated CVMs. Other researchers [143] have also shown that polybacterial *G. vaginalis*-containing CVMs are associated with non-hormonal contraception. This suggests that hormonal contraception merits further investigation as an intervention, e.g., probiotic to maintain or replenish the dominance of *Lactobacillus* and to protect against BV.

In contrast to other studies [33,34,141,179], we did not find any association of CST, alpha diversity or beta diversity with HPV in this chapter. There may be no association or, more likely, this was due to shortcomings of the methodology used in this chapter. This is not limited to the absence of negative and positive controls to optimise our analyses.

The high prevalence of BV in our cohort might not be accurate since BV in the parent study (HPV Couples Cohort Study) was not diagnosed using the “gold standard” Nugent score [44,79,80] or the Amsel’s critera [44,79], but using Pap smear that inspected the presence of *Gardnerella* and/or bacterial shifts indicative of BV. Although the Pap smear method has a high specificity (93-94%) for BV diagnosis and may be an accurate test when positive, the sensitivity of the method (43-49%) is significantly lower than that of the Gram stain method [80,293]. We also did not take into account the possibility of these women having AV, which at times is misdiagnosed for BV [88].

It is worth specifying that potential confounders such as age, sexual behaviour, and smoking may be present in our study and that there was no adjustment for these in the comparisons presented herein (specifically the association between hormonal contraception and *Lactobacillus* dominance) due to the small sample size. This represents a limitation of the analyses. In addition, several limitations associated with the V4 16S rRNA Ion Torrent methodology were identified in this chapter. These include: i) the loss of high number of reads (>70%) during quality-filtering, and ii) the relatively short V4 16S rRNA region, which resulted in lack of classification of several bacterial taxa, including four of the eight detected *Lactobacillus* OTUs. Failure to confidently assign species name to some bacteria was a major limitation since it impeded accurate comparison of our results to other published studies. Because of the limitations associated with the V4 16S rRNA Ion Torrent methodology, we therefore decided to recharacterise the CVMs of 19 randomly selected samples included in this chapter with a different method. The new method involved sequencing of a longer region of the 16S rRNA gene (V3 and V4 hypervariable regions), and the use of a higher throughput next generation sequencing platform, Illumina MiSeq. The comparison of the CVM profiles for the 19 samples is presented in the next chapter.

CHAPTER 3

Comparison of Ion Torrent Sequencing of the Bacterial V4 16S rRNA Region and Illumina Sequencing of the V3-V4 16S rRNA Region for Characterisation of Cervicovaginal Microbiotas

3.1 Introduction

Most microbiome studies rely on next generation sequencing of the partial 16S rRNA gene. The bacterial profiles generated by these studies may however be biased by a number of experimental and computational factors. These potentially include the DNA extraction method [294], the primers and 16S rRNA region targeted [49,257,295], the sequencing platform [49,295], the sequencing quality [49,296], and data analysis tools [296,297]. This chapter sought to compare the performances of two 16S rRNA metagenomics methods that differed in primer choice and sequencing technology used. This allowed for an examination of the impact of these factors on the cervicovaginal microbiota profiles generated.

The 16S rRNA gene is made up of nine hypervariable regions (designated V1-V9) [48]. The choice of which region of the 16S rRNA gene to target for microbiome studies is complex as the phylogenetic resolution achieved with each region will differ for the different taxa within the community. There may also be mismatches to the bacterial universal 16S rRNA primers in certain taxa. *In silico* analyses using databases of available 16S rRNA sequences have suggested optimal primers and variable region for capturing as many taxa as possible at genera level [49,298,299]. The optimal choice will, however, also be influenced by the specific microbial community being profiled. In the cervicovaginal microbiota, for example, species level classification of lactobacilli may be of greater priority than capturing all the less abundant taxa present. Regions of the 16S rRNA gene previously used in examining the CVM include the V1-V2 [162], V1-V3 [169], V2-V3 [179], V3-V4 [51], V4 [51,136], V4-V6 [110], V6, and V6-V9 [54] regions. There is no agreement on the optimal variable region for cervicovaginal microbiome studies [300]. It is however known that the universal primer 27F, used in studies examining the V1-V2 and V1-V3 regions, has mismatches to members of the family *Bifidobacteriaceae* which includes *G. vaginalis* [74], a common member of the CVM [38,39,51,54,135].

The different hypervariable regions are known to provide different taxonomic resolution and different estimates of taxa abundance due to differences in amplification performance of the primers. For example, experimental studies have found that pyrosequenced V6-V9 and V1-V3 datasets underestimate *Prevotella* and fail to detect *Gardnerella*, respectively, while pyrosequenced V1-V3 and V3-V5 datasets overestimate *Corynebacterium* [257]. Additional studies have found that the relative abundances of taxa estimated using V4 datasets are comparable to V4-V5 datasets but different from V3-V4 and V7-V8 datasets [49]. In addition to hypervariable V2 region, some studies have suggested that the hypervariable V4 region gives the lowest error rates when assigning taxonomy than regions like V6 [276,301]. Srinivasan and co-workers (2012) [51] described and validated a classification approach for pyrosequenced V4 sequences from vaginal bacteria, which achieved species-level classification of 99% of their sequence reads. Longer regions such as V2 and V3-V4 may however have more phylogenetic information, hence may be more reliable than shorter regions such as V6 [49,276]. However, there are certain scenarios where shorter regions perform comparable to longer sequences. For example, a study comparing the composition and structure of CVMs by different targeted metagenomics method and using the cervicovaginal microbiome 16S rRNA database found that phylogenetic read placement of 454 V6 and V6-V9 16S rRNA datasets were very similar [54]. Apart from using combined primer regions, for instance, V3-V4, longer reads with greater quality can be generated via paired-end sequencing that increases the accuracy of taxonomic classification [49,302-304].

Next generation sequencing platforms, such as Ion Torrent and Illumina, differ in output, read-length, read-quality and error rates that may impact microbiome profiling [269,304,305]. A single run on the Ion Torrent PGM with the latest chemistry is expected to yield up to 5.5 million reads of up to 400-bp in length. The Illumina MiSeq yields significantly greater output and length, with 44-50 million reads in a single run with v3 reagents and paired-end reads of 300-bp each. The platforms further differ in the type of sequencing errors and biases expected [269,304,305], with the Ion Torrent datasets having higher error rates than equivalent datasets from the Illumina platform [304]. For example, the average error rate in the reads from the Ion Torrent PGM platform is 1.5 and 1.4 errors per 100 bases for reads in the forward and reverse orientation, respectively, while those from the Illumina MiSeq platform is 0.9 errors per 100 bases [305].

In Chapter 2, the CVMs were profiled using the V4 region (~250-bp) sequenced on the Ion Torrent PGM platform (hereafter abbreviated as the IT-V4 method). Classification assignment for some bacteria could not be achieved to lower taxonomic levels possibly due to short read-lengths. Furthermore, the quality of the IT-V4 reads was relatively low with the result that many reads were discarded during quality-filtering. To address these shortcomings, 19 cervical samples, previously sequenced with the IT-V4 method, were profiled using the V3-V4 16S rRNA region (~460-bp) sequenced on Illumina MiSeq platform with paired-end 300-bp read-lengths (hereafter abbreviated as the IM-V3/V4 method). The microbial community profiles generated using the two methods were then compared. The IM-V3/V4 method was expected to increase throughput and produce longer high-quality reads to enable efficient taxonomic classification.

3.1.1 Aim and objectives

The aim of this chapter was to evaluate the complementarity of the taxonomic profiles of a subset of 19 specimens characterised by Ion Torrent sequencing of the V4 16S rRNA region (IT-V4) and Illumina sequencing of the V3-V4 16S rRNA regions (IM-V3/V4). Specific objectives included:

i. To compare the proportion of IT-V4 and IM-V3/V4 read sequences confidently assigned taxonomy at the different classification ranks (i.e., from phylum to genus),
ii. To compare diversity of the IT-V4 versus IM-V3/V4 datasets using Shannon index and UniFrac distances,
iii. To assess if there are differentially abundant taxa in IT-V4 versus IM-V3/V4 16S rRNA datasets, and
iv. To evaluate the extent of concordance between CVMs profiled by IT-V4 versus IM-V3/V4 datasets.

3.2 Materials and Methods

3.2.1 16S rRNA metagenomics library preparation, sequencing, and data analysis for the IT-V4 method

In order to compare the performance of the two 16S rRNA metagenomics methods, we randomly selected 19 (30.6%) samples from the 62 cervical samples that were sequenced on the Ion Torrent PGM platform (Life Technologies, Beverly, MA, USA) as described in Chapter 2. All the 19 samples were nucleic acids that had previously been extracted [270] using the MagNA Pure Compact System and the MagNA Pure Compact Nucleic Acid Isolation kit (Roche Molecular Diagnostics, Mannheim, Germany), thus, no new DNA extraction was performed for the IT-V4 and IM-V3/V4 methods. These 19 samples included 47.4% (n=9) and 57.9% (n=11) that were positive for HPV and BV, respectively. Only 21.1% (n=4) of the samples were positive for both HPV and BV. Details of the amplification and sequencing of the bacterial 16S rRNA gene (V4) are described in subsection 2.2.3. The V4 16S rRNA dataset from the 19 samples was analysed using bioinformatics pipelines as previously described (subsection 2.2.4).

3.2.2 16S rRNA metagenomic library preparation and sequencing for the IM-V3/V4 method

The hypervariable V3-V4 16S rRNA region of the bacterial gene was targeted by universal bacterial primer 319F (5’-CCTACGGGNGGCWGCAG-3’) and 806R (5’-GACTACHVGGGTATCTAATCC-3’) as previously described [306]. The primers were fused with Illumina overhang adapter sequences (forward overhang: 5’ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3’ and reverse overhang: 5’ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3’). The 16S rRNA amplicons for sequencing were prepared according to the 16S rRNA metagenomics protocol for MiSeq

System (Illumina, San Diego, CA, USA) [307], with minor modifications. The amplicon PCR step was different from the 16S rRNA metagenomics protocol for MiSeq System (Illumina, San Diego, CA, USA) that uses 2x KAPA HiFi HotStart ReadyMix (KAPA Biosystems, Wilmington, MA, USA). Here, amplicon PCR was performed using TaKaRa *Ex Taq*® Hot Start Version (Takara Bio Inc., Shiga, Japan) and identical thermocycling conditions as described in Chapter 2 (subsection 2.2.3). The Takara *Ex Taq* DNA polymerase has a 3'-to-5' proofreading exonuclease, thereby providing high sensitivity, high efficiency, and high yield [308]. The presence of amplicon (~560-bp) was confirmed by agarose gel electrophoresis. Pooled triplicate samples (amplicons) were run alongside 100-bp DNA ladder (50.0 µg/ml) on 1.5% ethidium bromide-stained agarose gel in Tris borate EDTA buffer at 100V for 2 hrs. The bands were visualized on the BioRad Molecular Imager® Gel Doc™ XR+ Imaging System and imaged using the Image Lab™ Software v5.21 (Bio Laboratories Inc., Hercules, CA, USA). The amplicons were purified using Agencourt AMPure XP System (Beckman Coulter, Beverly, MA, USA) and their concentrations estimated as previously described in subsection 2.2.3.

Next, Index PCR of the purified products was performed using KAPA HiFi HotStart ReadyMix PCR Kit (KAPA Biosystems, Wilmington, MA, USA) according to the 16S rRNA metagenomics protocol for MiSeq System (Illumina, San Diego, CA, USA) [307]. Each PCR reaction consisted of Nextera XT Index Primer 1 (N7xx), Nextera XT Index Primer 2 (S5xx), 1x KAPA HiFi HotStart Ready Mix, nuclease free water, and 100 ng DNA template. PCR conditions included: an initial denaturation at 95°C for 3 min followed by 8 cycles of denaturation at 95°C for 30 sec, annealing at 55°C for 30 sec, and extension at 72°C for 30 sec. A final extension was performed at 72°C for 5 min. Each sample library was analysed by gel electrophoresis. The libraries were then purified, eluted in low Tris-EDTA buffer, and quantified as previously described (subsection 2.2.3). All the libraries were pooled in equimolar concentrations and the resultant library quantified using a Bioanalyzer High Sensitivity Chip (Agilent Technologies, Santa Clara, CA, USA). The library was sequenced on the Illumina MiSeq using a paired-end 300-bp protocol and v3 reagents by Macrogen Inc. (Seoul, South Korea).

3.2.3 V3-V4 16S rRNA gene amplicon sequence data analyses

The qualities of the Illumina raw sequenced reads were visually inspected using FastQC v0.11.2 [272]. The data was processed and analysed using QIIME v1.8.0 [273] and UPARSE (usearch8.0.1616) [274] as described in Chapter 2, but with specific modifications as summarised in Supplemental Table 3.1. Raw forward and reverse reads were merged, quality-filtered, globally-trimmed to 400-bp length, and dereplicated in UPARSE. Specifically, the usearch *merge_fastq* command was used to merged and filter the reads. A maximum expected error (*fastq_maxee, E*) of 1.0 was used to filter the forward and reverse reads followed by merging the reads using a minimum overlap (*min_overlap*) of 32 bases. The first 30 nucleotides which were of relatively poor quality and partly included the oligonucleotide primers (17 bases) were stripped from the reads using the *stripleft* parameter. All the merged reads were truncated to 400 bases (using *trunclen*) to exclude regions with deteriorating quality. The usearch command, *fastq_filter*, using the *minlen* and *fastq_maxee* (1.0) parameters, were used to discard poor quality merged reads and those that were shorter than 400 bases. Reads were further processed using the parameters shown in Supplemental Table 3.1 for clustering into OTUs, picking representative sequences, assigning taxonomy using the RDP Naïve Bayesian Classifier [276] (with the Greengenes database (gg13_8 Release) [53]), and creating a phylogenetic tree and OTU table.

Alpha and beta diversities were measured using metrics implemented in QIIME as outlined in Chapter 2 (subsection 2.2.4). A rarefaction curve for each sample was generated at multiple sequence depths in order to determine a sufficient subsampling depth necessary to estimate microbial diversity. Based on these curves, a subsampling of 12,161 reads per sample was used for beta diversity analysis. PCoA of the beta diversity measures were used to visualize the variations in microbial community composition according to participants' metadata. EMPeror v0.9.60 [309] was used to visualize the PCoA plots.

3.2.4 Comparison of bacterial communities by Ion Torrent V4 and Illumina V3-V4 16S rRNA sequenced amplicons

The OTU richness, relative abundances and distributions of the bacterial taxa in IT-V4 and IM-V3/V4 datasets were computed. The fractions of reads classified at five taxonomic ranks (phylum to genus) and unique taxa to each method were assessed. Concordance between the two 16S rRNA metagenomics methods was evaluated at the genus level. Here, all species were collapsed to their respective genera. The genera that were shared between the two methods were then identified and their relative abundances in each sample calculated. The relative abundances of shared genera were then averaged for each method and their linear relationships computed using Pearson's correlation.

Procrustes analysis [310] was done on UniFrac distance matrices with 1,000 random samplings using the Monte Carlo algorithm to compute goodness (closeness) of fit (M^2) and p-values. The analysis was visualized using PCoA. For the calculation of UniFrac distance, the number of reads per sample was rarefied to 5,000 for IT-V4 method and 12,161 for IM-V3/V4 method. The transformed Procrustes matrices were visualized using EMPeror v0.9.60 [309]. The goodness of fit between the two sets of coordinate being superimposed was measured as M^2 by summing over the residuals. The M^2 value ranges from 0 for identical matrices to 1 for completely dissimilar matrices. A value less than 0.3 is considered a good fit (Jenva Kopylov, Personal Communication). The extent of significance of the Procrustes matrix correlations were measured by 1,000 Monte Carlo permutations. The number of times a random permutation resulted in a better clustering of the points than the actual data was also assessed by the *Count_Better* function in QIIME.

3.3 Results

3.3.1 Data analysis and output

A total of 440,122 270-bp and 1,046,720 400-bp reads from IT-V4 and IM-V3/V4 method, respectively, were analysed after filtering. These clustered into 172 and 108 OTUs for IT-V4 and IM-V3/V4 method, respectively. The alpha diversity as measured by the Shannon index was statistically different (p=0.0003) between the two methods, with a median of 1.6 (0.3-2.2) for IT-V4 and 2.1 (1.3-2.5) for IM-V3/V4.

3.3.2 Taxonomic classification of the reads and relative abundances of the bacterial taxa in Ion Torrent V4 and Illumina V3-V4 16S rRNA libraries

The fractions of the reads assigned at five taxonomic ranks (phylum to genus) were calculated in order to assess if longer reads (V3-V4) were better classified to deeper taxonomic levels than shorter reads (V4). At phylum, class, and order level, 99.95% reads were assigned for Ion Torrent V4 16S rRNA sequenced amplicons versus 100.00% for Illumina V3-V4 16S sequenced amplicons (Fig. 3.1). At family and genus level, 96.09% and 92.19% of the Ion Torrent V4 reads were classified versus 99.79% and 95.85% for Illumina V3-V4 16S rRNA reads, respectively. At the genus level, the majority of the unclassified reads from the IT-V4 (41.5%) and IM-V3/V4 (41.0%) datasets were from order *Clostridiales* and family *Flavobacteriaceae*, respectively.

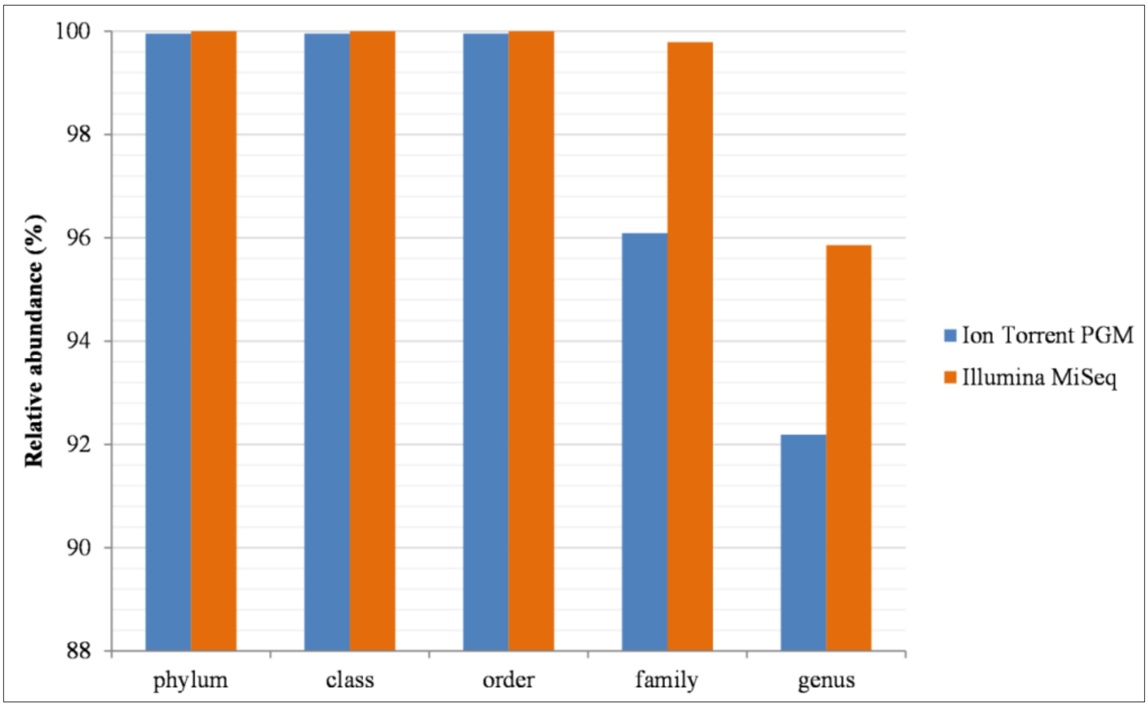


Fig. 3.1. Proportion of classified Ion Torrent V4 and Illumina V3-V4 16S rRNA sequenced reads at different taxonomic ranks. For better visualization, the y-axis (% relative abundance) scale starts at 88%.

These analyses were followed by identifying taxa shared between and unique to each of these targeted metagenomics methods. At the phylum level, the IT-V4 method detected 11 phyla namely Firmicutes (relative abundance: 39.1%), Actinobacteria (22.3%), Bacteriodetes (19.0%), Fusobacteria (16.5%), Terenicutes (2.3%), Chlamydiae (0.7%), Proteobacteria (0.1%), and four low-abundant (<0.1%) phyla (unclassified bacteria.1, unclassified bacteria.2, Cyanobacteria, and Verrucomicrobia). The IM-V3/V4 method detected 8 phyla namely Firmicutes (44.9%), Bacteriodetes (23.2%), Fusobacteria (16.5%), Actinobacteria (9.8%), Chlamydiae (1.9%), Terenicutes (1.6%), TM7 (1.2%), and Proteobacteria (0.9%). The most abundant phyla (Firmicutes, Actinobacteria, Bacteriodetes and Fusobacteria) were shared by both methods.

At the genera level, a total of 95.2% of the reads in the IT-V4 dataset and 87.2% of the reads in the IM-V3/V4 dataset were classified to the same genera (Fig. 3.2). Ninety eight and 55 genera were detected by the IT-V4 and IM-V3/V4 methods, respectively. Thirty six genera were found to occur in both IT-V4 and IM-V3/V4 methods, with *Lactobacillus*, *Gardnerella*, *Prevotella*, *Sneathia*, and *Shuttleworthia* being the most abundant shared genera.

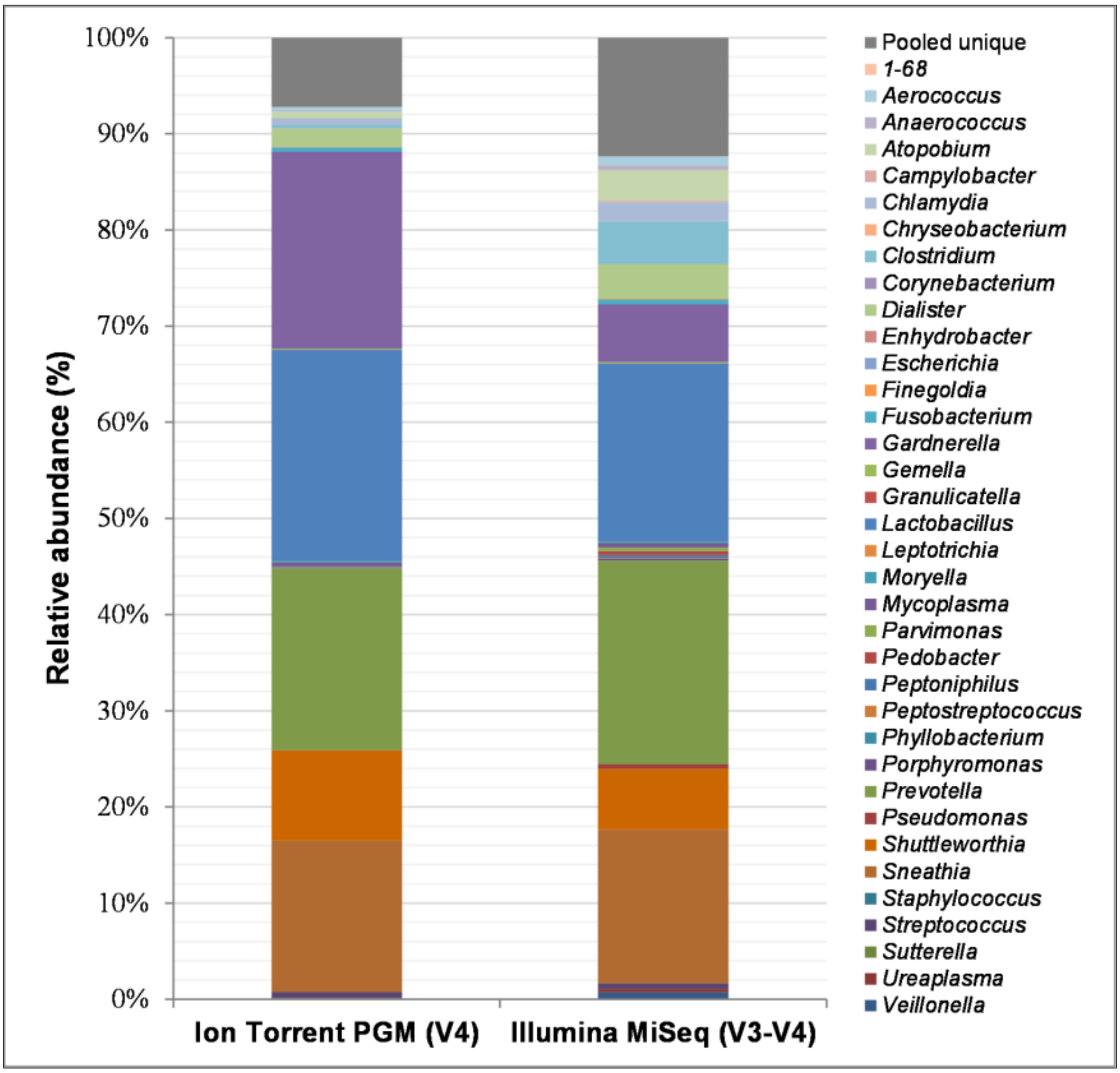


Fig. 3.2. Average relative abundance of genera shared between and unique to Ion Torrent V4 and Illumina V3-V4 16S rRNA amplicon sequencing methods. V4 and V3-V4 16S rRNA amplicons were used in Ion Torrent PGM and Illumina MiSeq, respectively. The relative abundances of the genera across the samples were averaged to get the respective per-method relative abundances. For each 16S rRNA metagenomics method, all unique genera were pooled together (as “Pooled unique”).

The IT-V4 method generated 43 more unique genera (n=62) than the IM-V3/V4 method (n=19). These unique genera represented only 4.8% of the total reads in the IT-V4 dataset. This proportion was 8.0% fewer than the proportion of the reads unique to the IM-V3/V4 method (12.8%) (Fig. 3.2, Pooled unique). The genera detected by IT-V4 and IM-V3/V4 in each of the 19 paired samples are shown in Fig. 3.3. There was a high sample-to-sample agreement between the IT-V4 and IM-V3/V4 method, but *Clostridium* and *Gardnerella* were differentially abundant taxa that were highly enriched in IM-V3/V4 and IT-V4 datasets, respectively. Moreover, there was a high variability in the pooled unique genera. Generally, the pooled unique taxa were more predominant with IM-V3/V4 datasets compared to IT-V4 datasets.

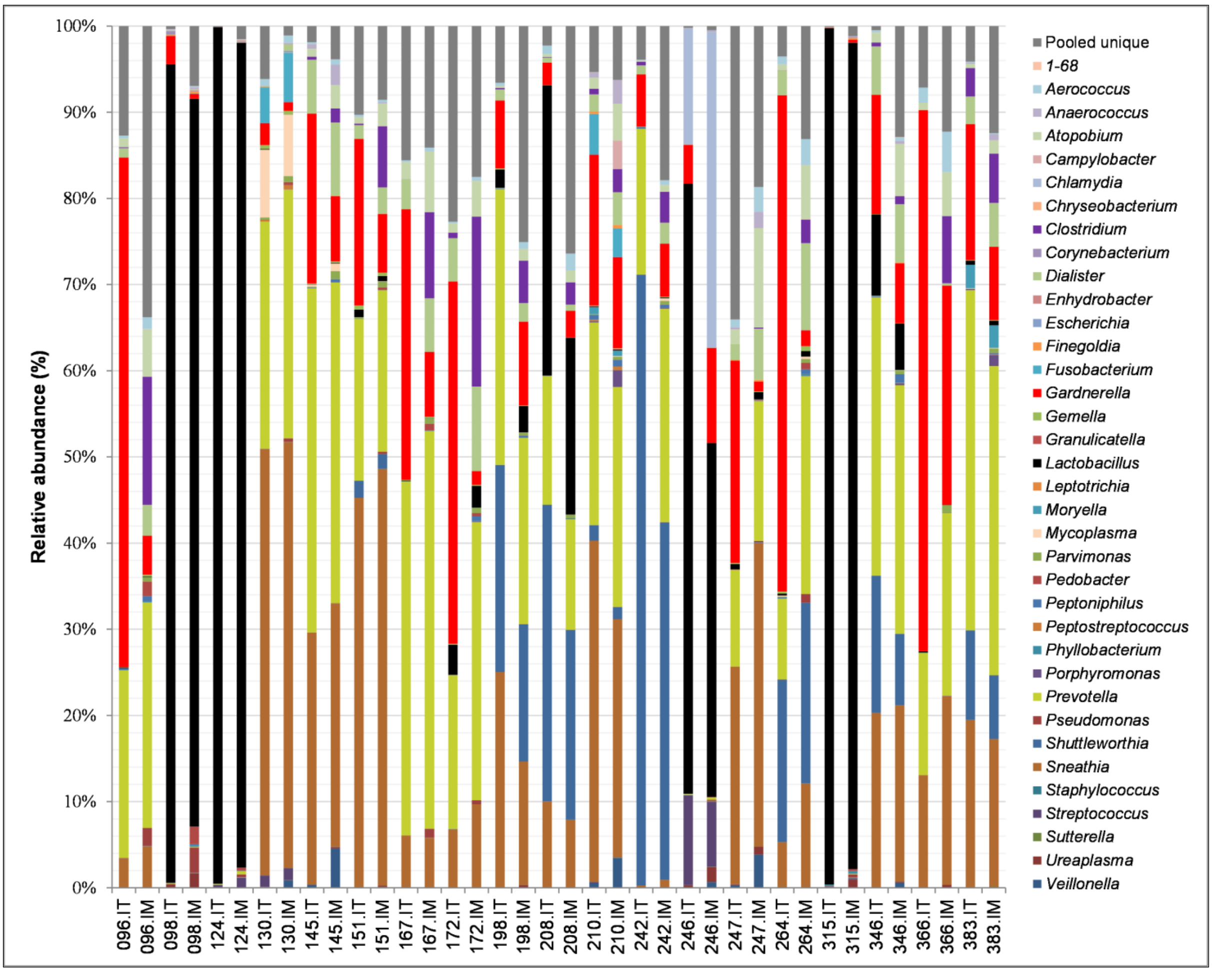


Fig. 3.3. Relative abundances (%) of bacterial taxa at genus-level in 19 matched samples sequenced by Ion Torrent and Illumina MiSeq methods. V4 and V3-V4 16S rRNA amplicons were sequenced on Ion Torrent PGM and Illumina MiSeq, respectively. All unique genera to each 16S rRNA metagenomics method were pooled together (as "Pooled unique"). Numbers in the x-axis represent the sample IDs and different amplification and sequencing methods; IT: Ion Torrent V4 16S rRNA dataset and IM: Illumina V3-V4 16S rRNA dataset.

Due to the observed intra-sample variability in certain genera between IT-V4 and IM-V3/V4 methods, the average relative abundances of the 36 genera that were found to occur in both methods were then compared and correlated using Pearson's correlations. The relative abundances of these genera in the two methods were plotted against another (Fig. 3.4a). The correlation was high between IT-V4 and IM-V3/V4 datasets (r=0.89, p<0.0001).

From the scatterplot (Fig. 3.4a), genera such as *Gardnerella* and *Clostridium* were the most extreme outliers with genera *Lactobacillus*, *Sneathia*, *Prevotella*, *Atopobium* being lesser outliers. The relative abundances of each of these bacteria across the 19 samples in the two 16S rRNA metagenomics methods were plotted and the correlations computed. Although the average relative abundance of each of the bacteria with the IT-V4 and IM-V3/V4 datasets had a positive correlation, some of the bacteria, e.g., *Gardnerella*, were more abundant in one dataset compared to the other, thereby supporting previous findings (Figs. 3.2 and 3.3).

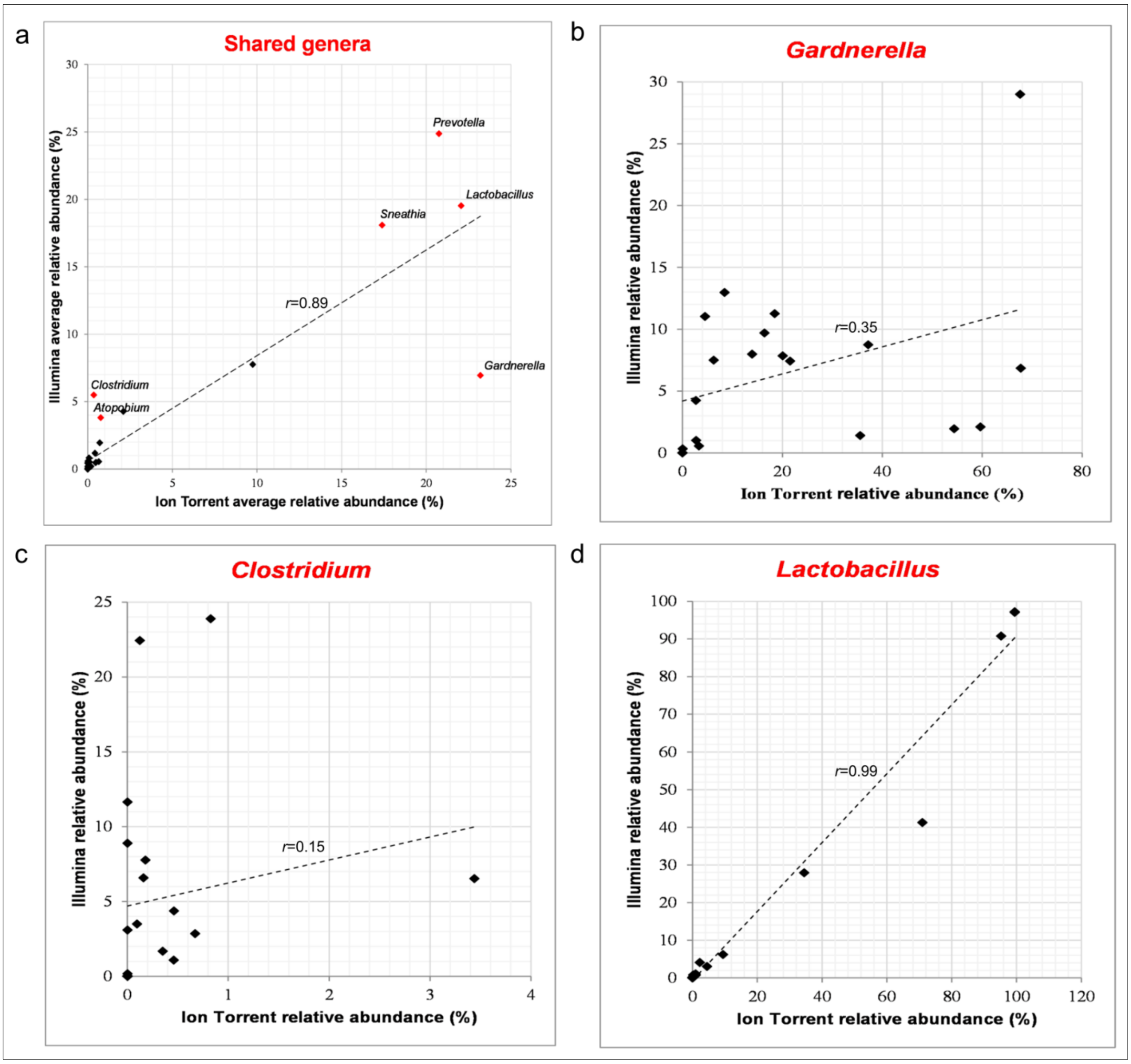


Fig. 3.4. Comparison of the relative abundances of genera as detected by Ion Torrent V4 and Illumina V3-V4 16S rRNA amplicon sequencing methods. a) Thirty six shared genera. Taxa that had noticeable variations between the two 16S rRNA metagenomics methods are highlighted with red dots and labelled, b) *Gardnerella*, c) *Clostridium*, and d) *Lactobacillus*. The results of the other bacteria (*Sneathia*, *Atopobium*, and *Prevotella*) are shown in Supplemental Fig. 3.1. The dashed line (trendline) is the line of identity (1=1). The relative abundance of each of the six highlighted genera from each method was compared across the 19 samples. Pearson's correlation (*r*) values are shown on the trendline.

Correlation was low for *Gardnerella* (*r*=0.35, p=0.147; Fig. 3.4b) and *Clostridium* (*r*=0.15, p=0.531; Fig. 3.4c). *Gardnerella* was identified in 18 samples (94.7%) by the IT-V4 method and in all the 19 (100.0%) samples by the IM-V3/V4 method. The median relative abundance was higher (16.4% (range: 0.0-67.2%)) in the IT-V4 dataset than in the IM-V3/V4 dataset (7.4% (range: 0.0-29.0%)), p=0.004. *Clostridium* was identified in 11 samples (57.9%) by the

IT-V4 method and in all the samples (100.0%) by the IM-V3/V4 method. The median relative abundance was 0.1% (range: 0.0-3.4%) in the IT-V4 dataset and 3.1% (range: 0.0-23.9) for IM-V3/V4, p<0.0001.

The correlation in the relative abundances for *Sneathia* (*r*=0.96, p<0.0001), *Atopobium* (*r*=0.91, p<0.0001), and *Prevotella* (*r*=0.90, p<0.0001) was high (Supplemental Fig. 3.1). The correlation for *Lactobacillus* was also high (*r*=0.99, p<0.0001; Fig. 3.4d). Among the *Lactobacillus* members, *L. iners* was the most abundant species. *L. iners* was classified to species level with both methods. *L. coleohominis*, *L. ruminis*, and *L. mucos*ae were taxonomically identified only in the IT-V4 dataset while *L. reuteri* was only identified in IM-V3/V4 dataset. Because lactobacilli are thought to be keystone cervicovaginal bacteria, we sought to identify the species of the unclassified *Lactobacillus* members in the IT-V4 and IM-V3/V4 datasets using the web-based BLASTn (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The IT-V4 dataset had three unclassified *Lactobacillus* OTUs (that we named *Lactobacillus.1*, *Lactobacillus.2*, and *Lactobacillus.3*) that could not be identified using BLASTn. On the other hand, the IM-V3/V4 dataset had four unclassified *Lactobacillus* OTUs (that we named *Lactobacillus.1*, *Lactobacillus.2*, *Lactobacillus.3*, and *Lactobacillus.4*). BLASTn identified *Lactobacillus.1*, *Lactobacillus.2*, and *Lactobacillus.4* in the IM-V3/V4 dataset as *L. jensenii*, *L. crispatus*, and *L. gasseri*, respectively.

3.3.3 Hierarchical clustering and beta diversity analyses of Ion Torrent V4 and Illumina V3-V4 16S rRNA datasets

To further test the level of concordance between the two methods, hierarchical clustering was performed using Bray-Curtis dissimilarity matrix and average linkage method. The heatmap of the relative abundances of the major bacterial taxa (30 for IT-V4 dataset and 31 for IM-V3/V4 dataset) in the 19 women are shown in Fig. 3.5. Two identical community groups were detected by both 16S rRNA metagenomics methods. These included *L. iners*-dominated CVMs (21.1%, 4/19) and diverse CVMs (78.9%, 15/19). In the *L. iners*-dominated CVMs, *L. iners* constituted a considerable proportion of the bacterial relative abundance (≥71.1% and ≥41.9% for IT-V4 and IM-V3/V4 dataset, respectively, in the heatmap). Samples in each of these clusters were concordant between the two 16S rRNA metagenomics methods. Supporting our previous results highlighted in this chapter, a quick inspection shows that some taxa, such as *Gardnerella* and *Clostridium*, were differentially represented by the two 16S rRNA metagenomics methods. The clustering of samples within the diverse bacterial communities varied slightly between the two datasets.

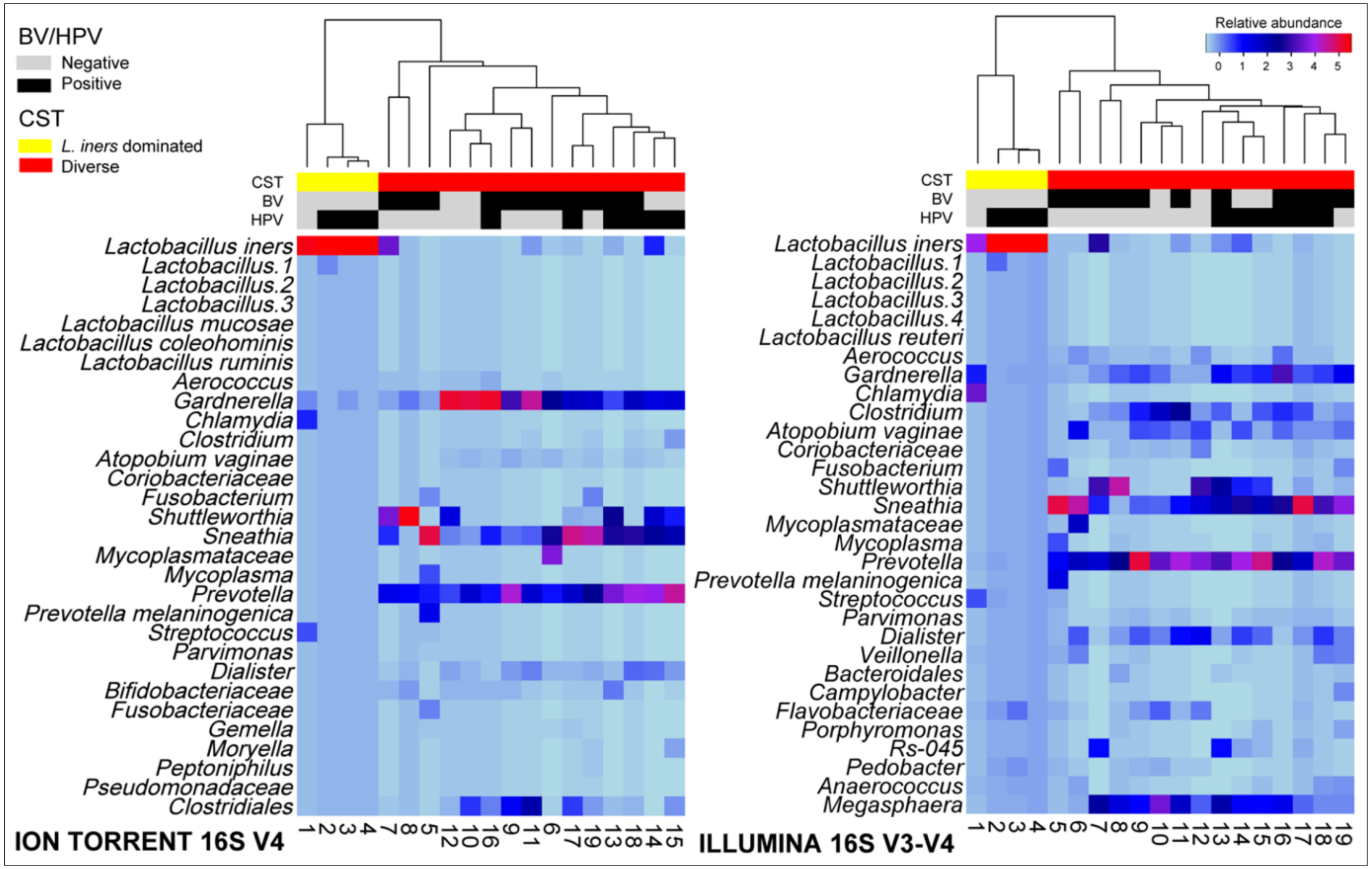


Fig. 3.5. Heatmap of the relative abundances of bacterial taxa in 19 cervicovaginal microbiotas characterised by Ion Torrent V4 and Illumina V3-V4 16S rRNA amplicon sequencing methods. Except for the *Lactobacillus* OTUs, only bacterial taxa with ≥0.05% and ≥0.16% relative abundances are shown for the IT-V4 and IM-V3/V4 datasets, respectively. Whereas the unclassified *Lactobacillus* OTUs (*Lactobacillus.1*, *Lactobacillus.2*, and *Lactobacillus.3*) in the Ion Torrent V4 16S rRNA could not be identified by BLASTn, three of the four unclassified *Lactobacillus* OTUs in the Illumina V3-V4 16S rRNA dataset were identified: as *L. jensenii* (*Lactobacillus.1*), *L. crispatus* (*Lactobacillus.2*), and *L. gasseri* (*Lactobacillus.4*). Each column represents a study participant with a specific ID (designated 1-19 along the x-axis). The dendrogram is based on average linkage hierarchical clustering of the Bray-Curtis dissimilarity matrix. Stronger intensity of red in the heatmap indicates a higher proportion of each taxon in each community while lower intensity towards light blue indicates a lower proportion. The community state type (CST), human papillomavirus (HPV) infection and bacterial vaginosis (BV) status of the women are indicated.

To test whether the beta diversities generated from the IT-V4 and IM-V3/V4 datasets are similar, Procrustes analysis of the weighted UniFrac distances was performed. This was done by aligning the first three principal coordinates from the two 16S rRNA metagenomics analyses (Fig. 3.6).

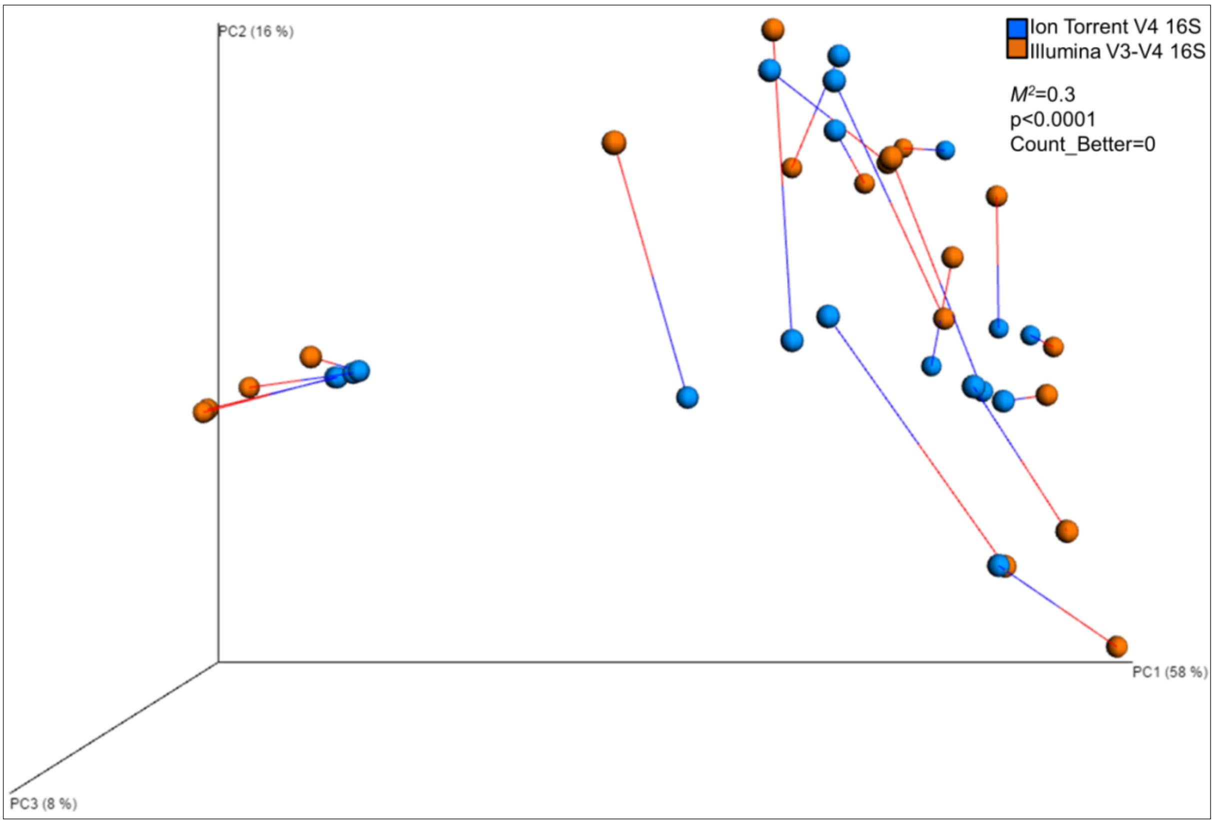


Fig. 3.6. Procrustes analysis of weighted UniFrac distances of Ion Torrent V4 and Illumina V3-V4 16S rRNA amplicon sequencing methods. V4 and V3-V4 16S rRNA amplicons from 19 samples were sequenced on Ion Torrent PGM and Illumina MiSeq platform. The first three principal coordinate axes and the percentage variation explained by each indicated are shown (PC1: 58%, PC2: 16%, and PC3: 8%). Samples are indicated by coloured circles with bars connecting the corresponding sample characterised by the Ion Torrent V4 method (blue) and Illumina V3-V4 method (orange).

Variations in the dataset are explained by the principal coordinate axes (PC1: 58%, PC2: 16%, and PC3: 8%). Each paired sample was connected by a bar, with a lower distance indicating a greater concordance. The goodness of fit between the two datasets was statistically significant (M^2=0.3, p<0.0001 based on 1000 Monte Carlo iterations), indicating that the sample clustering patterns overlapped very well between the two methods. Moreover, the *Count_Better* value was 0, indicating that none of the random permutations returned better clustering and that the observed fit of the data points was significant and not by chance.

3.4 Discussion

3.4.1 Comparison of throughput, taxonomic classification of reads, and bacterial taxa from Ion Torrent V4 and Illumina V3-V4 16S rRNA libraries

The aim of this chapter was to assess the complementarity of Ion Torrent V4 and Illumina V3-V4 16S rRNA datasets for profiling bacterial cervicovaginal communities. We had previously sequenced the V4 16S rRNA libraries on an Ion Torrent PGM platform since the technical specifications of this sequencing technology included low-cost, high-speed (2 hours), and high throughput capability (sequence yield per run of up to 1Gb on 318 chip) [269]. In our current study, the IM-V3/V4 method yielded 2.4-fold more quality-filtered reads than the IT-V4 method. The majority of the reads in the IT-V4 dataset (>70%) were discarded due to poor quality. Several studies [49,297,305,311] that have used Ion Torrent, 454 and Illumina technologies have noted that stringent quality-filtering may sometimes discard >50% of the raw reads. For example, a study by Salipante and colleagues (2014) [305], that was carried out subsequent to the study presented in Chapter 2, compared the performance of Illumina MiSeq and Ion Torrent PGM platform for 16S rRNA-based bacterial community profiling using mock community and human-derived microbial specimens found that >64.4% of the Ion Torrent-generated reads were discarded after data-filtering. In the case of Illumina, only 5.6% of the total dataset was discarded [305]. The authors suggested that Ion Torrent datasets may require less stringent data-filtering parameters, such as during trimming, although this may affect the bacterial community profiles [305] as substitution errors and indels are common in untrimmed bases [312]. Such error-prone reads may lead to an overestimation of microbial diversity [296]. Ion Torrent PGM is more prone to errors compared to Illumina MiSeq [269,304,305,312]. This could also have resulted from the amplifying enzymes: the Takara *Ex Taq* DNA polymerase used in IM-V3/V4 method is known to have a higher fidelity than the typical *Taq* polymerase used in IT-V4 method [308]. The high rate of indels [269,312], substitution errors [312], and premature read truncation (hence, shorter reads with variable lengths) [304] associated with the Ion Torrent platform may explain the low number of quality-filtered reads generated by the IT-V4 method [305]. It has been shown that, compared to Illumina MiSeq dataset, less reads may be retained from the Ion Torrent PGM dataset after quality-filtering and chimera checking due to poor quality of the reads [295]. Further, Illumina has a superior throughput to most deep sequencing technologies, including Ion Torrent [54,269,303,304]. Merging the Illumina MiSeq paired-end reads into contigs not only increases read-length, but also quality [49,302-304].

In terms of OTUs, the IT-V4 method identified a greater number of OTUs (172 OTUs) in the 19 samples compared to the IM-V3/V4 method (108 OTUs). A previous study demonstrated that shorter regions (V4 16S rRNA amplicons) produced lower numbers of OTUs compared to longer regions (e.g., V4-V5 16S rRNA amplicons) when the same sequencing platform was used to compare [303]. The authors argue that shorter reads have less sequence information and are, thus, less likely to be discriminated as distinct OTUs [303]. The reads generated by the IT-V4 method were about 1.5 times shorter than the V3-V4 16S rRNA reads generated by IM-V3/V4 method. The observed high number of OTUs with the shorter region (IT-V4 method) is likely due to sequence errors resulting in spurious OTUs [269,304,305]. This is substantiated by the low relative abundance of many of the OTUs in the IT-V4 dataset. There is also a possibility that some of these OTUs were chimeras that missed detection. It has been shown that chimeras present in short sequences, especially from closely related bacteria, pose a more difficult challenge to detect and remove than in longer sequences using the typical chimera detection algorithms [303].

More IM-V3/V4 reads were unambiguously assigned taxonomy than IT-V4 reads. The difference in classification was more evident at lower taxonomic ranks. While about 100% of IM-V3/V4 reads were assigned taxonomy at family level, only 96% of the IT-V4 reads were assigned at this level. At genus level, the same trend was noted (96% versus 92%). Taxonomic assignment is more dependent on the choice of primers and 16S rRNA region than on the sequencing technology [302]. Error rates in taxonomic classification have been associated with the choice and length of 16S rRNA sub-region, with the hypervariable V3 and/or V4 sub-region(s) having the most robust performance than other sub-regions and configurations [313]. Interestingly, a study by Claesson and co-workers (2010) [49] that compared 454 and Illumina technologies for profiling the gut microbiota using different configurations in the 16S rRNA gene showed that the V3-V4 region performed the worst compared to single regions such as V4 and tandem regions such as V4-V5. Nonetheless, *in silico* predictions had posited that the V3-V4 configuration had the best taxonomic classification accuracies regardless of the sequencing technology, Illumina paired-end and 454 sequencing [49]. In accordance with this and previous studies [49,301,302], longer reads (V3-V4 from paired-end sequencing) were better for taxonomy assignments than shorter reads (V4 single-end sequencing). Higher sequencing error rates in Ion Torrent PGM [304,305] could have also contributed to the low classification rates. Erroneous reads are less likely to have high confidence hits in the Greengenes database. It is important to note that variation in the detectability of the different *Lactobacillus* spp. could potentially be due to amplification bias and/or misclassification of *Lactobacillus* spp. Regardless of the 16S rRNA metagenomics method, unclassified taxa could represent PCR/sequencing errors (including taxa-specific bias) [54,269,303,305], inability of the Greengenes database to discriminate the sequence from others [303], or even novel bacterial candidates. The CVM has been shown to include many novel bacteria of the order *Clostridiales* [51].

Overall, the present study showed that there was significant correlation (*r*=0.89, p<0.0001) between the relative abundances of the bacterial taxa as determined by the two methods. The relative abundances of most of the shared genera were strongly correlated between the two 16S rRNA metagenomics methods, e.g., the correlation value for *Lactobacillus* and *Sneathia* was 0.99 and 0.96, respectively. For other community members, the relative abundances were different. For example, *Gardnerella* and *Clostridium* were significantly more enriched in IT-V4 and IM-V3/V4 datasets, respectively. The lack of classification of order *Clostridiales* sequences (in the IT-V4 datasets) to genus level could have resulted in the disparate relative abundance estimates of *Clostridium* in the two methods. Disparate relative abundance estimates were also evident in individual samples and therefore underscore the existence of bias for some organisms with different 16S rRNA metagenomics methods [49,269,302,305]. For example, a previous 16S rRNA metagenomics validation study on synthetic mock community found *Propionibacterium* was nearly absent with Illumina V4 16S rRNA amplicon sequencing method but at 1.5% and 2.4-2.9% relative abundance with 454 and Illumina V4-V5 16S rRNA amplicon sequencing methods [303]. Brooks and co-workers (2015) [314] observed bias in certain bacteria when they assessed mock vaginal samples using 454 V1-V3 16S rRNA datasets. *L. iners* and *S. amnii* were overestimated whereas *G. vaginalis* and *S. agalactiae* were underestimated by their method [314]. As the actual microbiota compositions in the 19 samples were unknown and no internal controls were run, it is not possible to say which method overrepresented or underrepresented the bacterial relative abundances. Further evaluation using mock communities, matched primers, and sequencing platform would be needed to delineate the direction of the revealed bias.

3.4.2 Community state types and beta diversity analyses of Ion Torrent V4 and Illumina V3-V4 16S rRNA datasets

The cervicovaginal microbiotas of the 19 women included in this comparison clustered into two CSTs (*L. iners*-dominated and diverse communities) regardless of the method used.

None of the cervicovaginal microbiotas of the 19 women were dominated by *L. crispatus or L. jensenii*. However, the clustering within the diverse communities from the IT-V4 dataset was slightly different from the IM-V3/V4 dataset. The method used for hierarchical clustering is dependent on the bacterial composition and abundance. The variation in clustering for some samples could have resulted from amplification and sequencing bias as well as differences in the classification of 16S rRNA reads. This is not the first study observing some variations in the CSTs of samples sequenced with different methodologies. A study comparing the composition and structure of CVMs assessed using 454 V6 and V6-V9 16S rRNA datasets found that different CSTs were established for some samples [54]. These results, together with ours, are reflective of the existence of amplification and sequencing bias. Fouhy and colleagues (2016) [295] assessed the effect of DNA extraction (repeat bead beating versus modified Qiagen DNA extraction method), primer choice (V4-V5, V1-V2, and V1-V2 degenerate primers), and sequencing platform (Ion Torrent PGM versus Illumina MiSeq) using mock communities. The authors found that the different regions of the 16S rRNA gene yielded different microbial composition and abundance; with the V4-V5 Ion Torrent PGM combination being the only approach that detected all of the 20 expected mock community species [295]. Furthermore, the samples clustered according to the sequencing platform used, with the exception of V4-V5 primer that was less affected by the sequencing technology [295]. In another study comparing the effect of hypervariable region(s) (V4, V6–V8, and V7–V8) and sequencing platform (454 versus Illumina MiSeq) choice [302], Tremblay and colleagues (2015) found that the V4 Illumina MiSeq combination produced the highest similarity with the expected taxonomic results. In the present study, different 16S rRNA regions (of different lengths) and sequencing platforms were used, which could have resulted in different classification efficiencies.

Finally, to determine whether the beta diversity derived from the IT-V4 dataset was consistent with IM-V3/V4 dataset, Procrustes analyses of weighted UniFrac distances were performed. Procrustes analyses demonstrated a reproducible statistically significant correlation between the beta diversity estimations of the IT-V4 and IM-V3/V4 methods. This could be ascribed to the fact that beta diversity estimates were performed at sufficient and even sampling depth across the samples in each amplicon sequencing method. Congruent with this, Caporaso and co-workers (2012) [311] demonstrated that such beta diversity estimations are less dependent on subsampling effort. However, sufficient sequencing depth may be necessary for comprehensive analyses [311]. Sample-to-sample deviation in clustering on the PCoA plot may also be attributed to differences in read-length and quality of sequencing platforms [315]. Nelson and colleagues (2014) [303], however, observed that the choice of the 16S rRNA region had a greater influence on the beta diversity estimates than the type of sequencing technology utilised [303]. While each of these two factors can independently impact on beta diversity, the present study design could not evaluate them independently.

3.4.3 Summary

Overall, this small comparative study of two targeted metagenomics methodologies showed reasonable complementarity between IT-V4 and IM-V3/V4 libraries. The Illumina-based method, however, had longer read-length (due to the 300-bp paired-end sequencing), higher throughput, and lower error rates (hence, less spurious OTUs and higher percentage of reads retained after data-filtering) than the Ion Torrent PGM method. Moreover, the IM-V3/V4 method had more reads that were classified at the various taxonomic levels (phylum to genus). The advantages provided by IM-V3/V4 method supported its utility in the subsequent experiments described in chapters 4 and 5. Given the demonstrated impact of the primer choice and sequencing platform on the outcome of microbiome diversity, there is need to include internal controls, such as mock communities, in order to optimise and validate the microbiome analyses, thereby mitigating any potential bias.

CHAPTER 4

Characterisation of Cervicovaginal Microbiotas of Reproductive-Age Black African Women with and without Human Papillomavirus Infection using Illumina Sequencing of the V3-V4 Region of the 16S rRNA Gene

4.1 Introduction

In this chapter, the CVMs of 91 reproductive-age South African women with and without HPV infections were characterised and associated with demographic, sociobehavioural, and clinical factors. The metagenome functions of the CVMs were also predicted. This study was an expansion of the cohort from Chapter 2 but using a high-throughput Illumina sequencing approach which was more reliable as demonstrated in Chapter 3. The sample size was increased from 62 to 91 with a view to achieve enough statistical power that allows association between CVMs and participants' variables to be confidently deduced.

Black South African women have a high burden of BV [76,84-86] and HPV, including high-risk HPV [16] that cause cervical cancer [12]. Even though it is emerging that BV and *Lactobacillus*-deficient CVMs are associated with higher HPV prevalence [33,34,81,141,157,215], there are no published studies that have examined the association between the CVMs and HPV infection in South African women. Literature is also rapidly expanding to show that specific taxa in the CVM may be potential biomarkers in HPV infection and natural history [137,157,179]. While common *Lactobacillus* spp. such as *L. crispatus* and *L. gasseri* are inversely associated with HPV infection [34,139], *Sneathia*, and *G. vaginalis* have been positively associated with HPV infection [179,215] as well as precancerous cervical lesions [140,142,160,176]. None of these studies were conducted on the high HPV-burden population in South Africa.

Evidence from large cohort studies demonstrates that ethnicity/race influences CVMs [38-40,143,177]. The majority of the CVM studies have been performed on Western populations and those of European ancestry [38,39,143,144]. A few studies have shown that many African women have CVMs that are deficient in *Lactobacillus* species and characterised by diverse and heterogeneous populations of mainly BV-associated bacteria [33,39,87,136,143]. This diversity remains largely unexplored and unexplained especially among Black South African women. Black women have also been shown to have less acidic vaginal pH (>4.5) than White, Asian and/or Hispanic women [38,70,73]. Less acidic vaginal pH is associated with *Lactobacillus*-deficient CVMs [38,61,62,68]. Less acidic vaginal pH may not be protective against vaginal infections. Elevated vaginal pH (≥5), for example, was associated with a 10-20% increased risk of HPV infection in a large population-based study of premenopausal Costa Rican women [70]. Mitra and co-workers (2016) [218] proposed that this may be related to the susceptibility of the HPV E5 oncoprotein to acidic pH [316].

Knowledge on the role of all the cervicovaginal bacterial taxa is not yet comprehensive. Only a few studies have examined the functions of the CVMs [87,122,155,175,219,317]. Cervicovaginal metagenome composition and function may have a profound effect on the cervicovaginal milieu, such as maintaining the vaginal epithelial integrity [122]. Soluble compounds from *G. vaginalis* culture, for example, have been found to impair wound-healing capacity in the female genital tract [122]. Mucosal and epithelial impairment caused by particular CVMs has been suggested as a possible mechanism through which CVMs mediate cervical disease, including HPV-related precancerous cervical lesions [218].

Overall, the motivation for this study was due to the scarcity of data on CVMs of reproductive-age Black African women and the lack of consistent association between CVMs and HPV infection [34,81,139-141,210,215,221,267].

4.1.1 Aim and objectives

The aim of this chapter was to investigate the composition of CVMs of reproductive-age Black South African women (aged 18-44 years) and the associations with HPV infections and other demographic, sociobehavioural, and clinical factors. Specific objectives included:

i. To characterise the CVMs of reproductive-age HIV-seronegative Black South African women by sequencing the V3-V4 16S rRNA amplicons on the Illumina MiSeq platform,
ii. To assess the association of the established CSTs with demographic, sociobehavioural, and clinical information of reproductive-age Black South African women,
iii. To detect potential bacterial markers for prevalent HPV and high-risk HPV in the women, and
iv. To predict the functional capabilities of the bacterial communities in the CVMs using the V3-V4 16S rRNA dataset.

4.2 Materials and Method

4.2.1 Ethics statement

This information has been provided in the Materials and Methods section (subsection 2.2.1) in Chapter 2.

4.2.2 Study samples, study design and study population characteristics

The study samples, study design including inclusion and exclusion criteria, as well as nucleic acid isolation and HPV genotyping were as previously outlined (subsection 2.2.2). Here, cervical specimens from 91 HIV-seronegative Black South African women aged 18-44 years were considered for microbiota and HPV investigation.

Eighty seven women (95.6%) were finally included in the study. Participants excluded (4.4%, 4/91) were those whose samples had <12,161 reads for 16S rRNA metagenomics analyses. The cohort baseline characteristics of the 87 sexually active women are summarised in Table 4.1.

Table 4.1. Baseline demographic, sexual and smoking, behavioural, and clinical information of the 87 heterosexually-active Black South African women.

Characteristic	All participants (N = 87)
Age (years)	32.0 (25.0-39.0)
Age at sexual debut (years)^	18.0 (16.0-19.0)
Number of lifetime sexual partners	3.0 (2.0-4.0)
Number of sex acts with study partner in last month^	4.0 (2.0-8.0)
Current use of hormonal contraceptives* (% (n/N))	
No	57.5 (42/73)
Yes	42.5 (31/73)
HPV (% (n/N))	
Negative	57.5 (50/87)
Positive	42.5 (37/87)
High-risk (% (n/N))	
Negative	65.5 (57/87)
Positive	34.5 (30/87)
Cervical cytology (% (n/N))	
Normal	75.6 (62/82)
ASCUS	13.4 (11/82)
LSIL	11.0 (9/82)
Experienced vaginal discharge in last 6 months (% (n/N))	
No	84.6 (66/78)
Yes	15.4 (12/78)
Experienced genital ulceration in last 6 months (% (n/N))	
No	97.7 (85/87)
Yes	2.3 (2/87)
Incidental BV on Papanicolaou smear (% (n/N))	
No	56.3 (49/87)
Yes	43.7 (38/87)
Cigarette use (% (n/N))	
Never smoked	64.4 (56/87)
Ex-smoker	3.4 (3/87)
Current smoker	32.2 (28/87)

Abbreviations: HPV – human papillomavirus, ASCUS – atypical cells of undetermined significance, LSIL – low-grade squamous intraepithelial lesion, BV – bacterial vaginosis.

Continuous variables are expressed as medians with interquartile ranges (IQRs, at 25th and 75th percentiles).

^Data was not available on the age at sexual debut for one woman, lifetime number of sexual partners of one woman, and number of sexual acts with study partner in the last month of two women.

*The hormonal contraceptives included oral pills, norethisterone enanthate, Depo-Provera, and steroids.

Of the 87 women, 37 (42.5%) were positive for HPV infection. Among the HPV-positive women, 51.4% (19/37) had single HPV infections. A total of 45.9% (17/37) and 81.1% (30/37) of the women were infected with low- and high-risk types, respectively. Of these, 10 women (27.0% (10/37)) were infected with both low- and high-risk HPVs, respectively. When stratified according to high-risk HPV status, 30 women (34.5%) were high-risk HPV-positive. The median age of high-risk HPV-negative and high-risk HPV-positive group was 35 years (IQR 27.0-40.3) and 28.0 years (IQR 23.5-35.5), respectively. High-risk HPV-positive women were significantly more likely to be younger (p=0.007) and have LSIL cervical cytology (p=0.053) compared with high-risk HPV-negative women. More high-risk HPV-negative women had normal cervical cytology as compared to high-risk HPV-positive women (83.6%

(46/55) versus 59.3% (16/27), p=0.016). None of the other selected variables (age at sexual debut, number of lifetime sexual partners, number of sexual acts with study partner in the last month, hormonal contraception, ASCUS cervical cytology, vaginal discharge and genital ulceration in the last 6 months, BV status, and cigarette use) differed statistically between the two high-risk HPV groups. These results were found to be consistent when the women were stratified into HPV-negative and HPV-positive groups (results not shown), with LSIL cervical cytology being more prevalent in HPV-positive than HPV-negative women (p=0.003).

Follow-up HPV genotyping data at the 6-month visit was available for 63.2% (55/87) of the women. The HPV data at the 6-month visit is shown in Supplemental Fig. 4.1. Twenty six women (26/87 (78.8%)) remained HPV-negative at the 6-month follow-up. A total of 7 women acquired a new HPV genotype infection. Specific HPV genotypes were cleared in 4 women. Two women had persistent infection with the same HPV genotype. Sixteen women had complex HPV genotyping results. In this study, this category included the following definitions: i) women who had cleared certain HPV genotype infections but had other persistent genotypes, ii) women that had cleared certain HPV genotype infections but had acquired new genotypes, and iii) a combination of i and ii.

4.2.3 Illumina MiSeq V3-V4 16S rRNA metagenomics library preparation and sequencing

The amplification, amplicon purification and quantification, and library sequencing protocol used is described in Chapter 2 (subsection 2.2.3) and 3 (subsection 3.2.2).

4.2.4 PCR and sequencing controls

Positive and negative controls were also run concomitantly with the experimental samples. Two mock bacterial communities, HM-782D and HM-783D (BEI Resources, Manassas, VA, USA), comprising of genomic DNA from 20 bacterial strains, were used as positive controls (Supplemental Table 4.1). HM-782D contained equimolar concentrations (100,000 rRNA operons per organism per μl) of each of the 20 bacteria. In HM-783D, the concentrations of these bacteria were staggered (ranging from 1,000-1,000,000 rRNA operons per organism per μl). Nuclease free water was used as a negative control. Digene STM was included as an extraction control to check for possible contaminants. This was a mock DNA extraction from Digene STM. The nuclease free water and Digene STM controls were sequenced despite no visible amplicons on agarose gels.

4.2.5 V3-V4 16S rRNA gene amplicon sequence data analyses

The raw sequenced data were processed and analysed using QIIME v1.8.0 [273] and UPARSE (usearch8.0.1616) [274] as previously described (Chapter 3 – subsection 3.2.3) and according to the criteria summarised in Supplemental Table 3.1. The relative abundance of the most abundant non-mock bacteria (0.0222%) in the even mock community (HM-782D) was used as the filtering threshold for the whole OTU table. All controls (two mock communities, two nuclease free water, and Digene STM) and sequences of samples that failed alignment were excluded from the final OTU table.

Alpha diversity was measured using the observed_species, chao1, Shannon, Shannon equitability, Simpson, Dominance, and PD_whole_tree metrics and compared as described previously (subsection 2.2.4). Read count per sample was rarefied to 12,161 reads prior to calculation of beta diversity as described in subsection 3.2.3.

4.2.6 Hierarchical clustering of cervicovaginal microbiota and identification of bacterial ecological interactions

To identify the CSTs, hierarchical clustering (average neighbour linkage) was performed based on Bray-Curtis dissimilarity indexes calculated using the Vegan R package [278]. Only taxa with relative abundances greater than or equal to 0.33% were included in the heatmap together with multiple variables (BV, HPV, and high-risk HPV statuses).

Differentially abundant bacteria between the established CSTs were identified using a two-sided White’s nonparametric statistical t-test [318] on the STAMP software v2.1.3 [279] using the parameters underlined in Chapter 2 (subsection 2.2.5).

Positive and negative correlations between the counts of all OTUs in the CVMs were assessed by metagenomeSeq v1.12.1 [280,281] as described in subsection 2.2.6. Only the correlations of the OTUs with the greatest variance (n=40, 33.3%) were plotted on a correlogram.

4.2.7 Identification of potential biomarkers for HPV infection and BV

Differentially abundant bacterial taxa in women with and without HPV and BV were assessed using the Linear Discriminant Analysis (LDA) effect size (LefSe) algorithm v1.0 [319]. The significance level (p-value) for Kruskal-Wallis and pairwise the Wilcoxon signed-rank tests was set at 0.05. Discriminant features were identified at a logarithmic LDA score threshold of 2.0 unless indicated.

4.2.8 Prediction of metagenome function

The metagenome function was predicted using Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) v1.1.0 [317] and the Kyoto Encyclopedia of Genes and Genomes (KEGG) module abundances analysed and visualized using LefSe v1.0 [319] and STAMP v2.1.3 [279].

4.2.9 Statistical analyses

Statistical analyses were done with STATA v13.0 (StataCorp, College Station, TX, USA) and R v3.2.2 (R Core Team 2016). Associations between the established CSTs and participants' metadata were calculated using Fisher's exact/Chi-square tests for categorical variables and Mann-Whitney unpaired nonparametric test for continuous variables, with statistical significance at p<0.05 (two-tailed).

The alpha diversity between the CSTs was compared using Mann-Whitney unpaired nonparametric test, with p<0.05 showing significance. More than two-group comparison of the alpha diversity measures of the CSTs was computed by Kruskal-Wallis test.

4.3 Results

4.3.1 Validation and optimisation of microbial analyses using PCR and sequencing controls

This subsection summarises the procedure that was used to ensure an accurate representation of microbial diversity and relative abundance. Cross-contamination between specimens and contaminants from other sources was assessed by use of two negative controls (Digene STM extraction control and a nuclease free water control). These controls allow for identification of contaminants in the kits used during library preparation and sequencing (DNA extraction, amplification, and sequencing kits) and any contaminations that may arise during processing of the samples. The impact of PCR and sequencing artefacts was additionally assessed by the inclusion of two mock bacterial communities as controls.

Overall, the total number of quality-filtered sequencing reads from the library of 95 samples (91 cervicovaginal samples and four controls samples) was 5,940,510. The average number of quality-filtered reads per cervicovaginal sample was 63,558. For the nuclease free water and STM extraction controls, the total number of quality-filtered reads was only 103 and 533, respectively. The predominant taxa in nuclease free water control were *L. iners* (28 of 103 reads (27.2%)) and *Sneathia* (14/103 (13.6%)). In the Digene STM extraction control *Flavobacteriaceae* (232/533 (43.5%)) and *P. veronii* (109/533 (20.5%)) were the most abundant taxa. The total number of reads (level of contaminants) in the nuclease free water and extraction control differed from the average reads per cervicovaginal sample by at least 99%. This indicated that contaminants were nearly absent in the sequenced library and unlikely to affect our analysis.

The bacteria in the two mock communities (even HM-782D and staggered HM-783D) were also assessed in order to evaluate the extent to which the observed community composition deviated from the expected composition. Both mock communities contained 20 bacterial species (the composition of the two mock communities is shown in Supplemental Table 4.1). The number of quality-filtered reads in the HM-782D and HM-783D mock controls was 90,071 and 66,062, respectively. Nineteen of the 20 mock bacteria were successfully amplified and sequenced (Fig. 4.1). One mock bacterial species, *S. aureus,* was not detected as a separate species. *S. aureus* is very closely related to another mock species, *Staphylococcus epidermidis* [320,321] and, it is likely that the 16S V3-V4 reads from both species were classified as *S. epidermis*. Taxonomic assignment for some species, such as those that were misclassified (i.e., *Clostridium beijerinckii*, *Listeria monocytogenes*, and *Neisseria meningitidis*) had to be manually curated. Here, the representative sequences of such bacteria were aligned against the bacterial sequences in the 16S rRNA database using the web-based BLASTn (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Misclassification of certain species and failure to assign others a species name (e.g., *Streptococcus mutans and Streptococcus pneumoniae*) has been demonstrated to be a limitation of the Greengenes 16S database [303,322].

HM-782D contained equimolar concentrations (100,000 operons per organism per μl) of each of the 20 bacteria. In HM-783D, the concentrations of the bacteria were staggered (ranging from 1,000-1,000,000 operons per organism per μl). Fig. 4.1 shows the observed distributions of the mock bacteria in the sequence data from HM-782D and HM-783D. As expected, the bacterial species distributions in were more even HM-782D than in HM-783D.

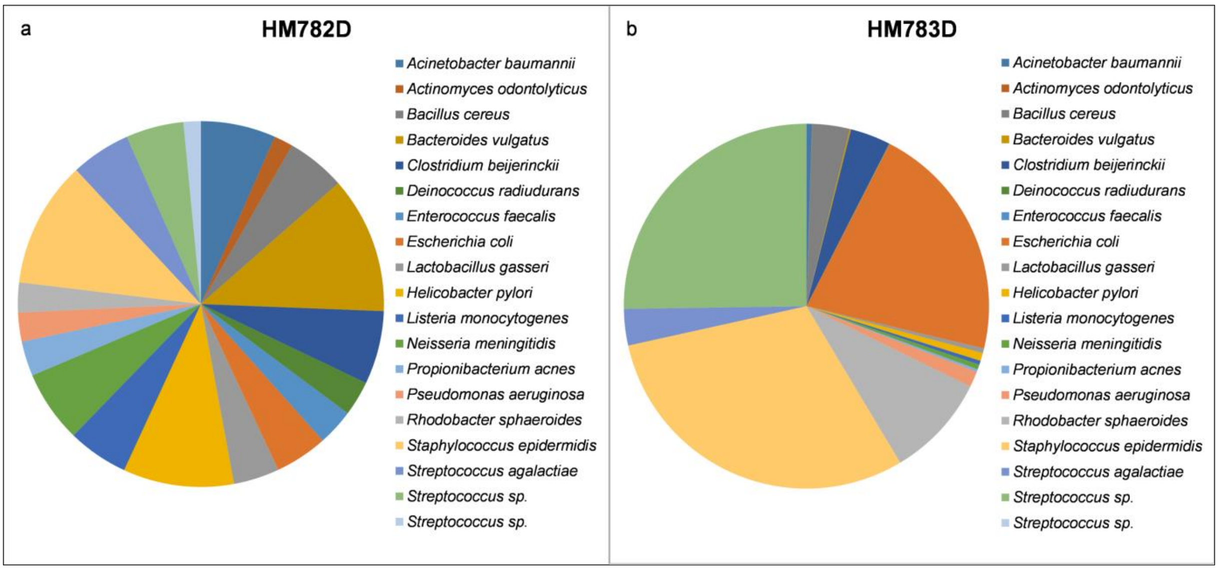


Fig. 4.1. Observed distribution of bacteria in even (HM-782D) and staggered (HM-783D) mock bacterial communities.

The relative abundance of the bacteria was then evaluated to assess the extent of contamination. The mock bacteria in HM-782D and HM-783D accounted for 99.89% and 99.84% of the total reads in each sample, respectively. Non-mock bacteria (false positive) were present at very low levels (0.11% of reads in HM-782D and 0.16% in HM-783D). A total of 26 and 34 non-mock bacteria (OTUs) were detected in HM-782D and HM-783D, with most of these represented by a single read only (50% (13/26) and 62% (21/34), respectively). The most abundant non-mock bacteria in HM-782D and HM-783D were *L. iners* (0.02%) and *Sneathia* (0.03%), respectively. The relative abundance of *L. iners* in HM-782D was assumed to represent the level of cross-contamination and was used as a threshold to control for spurious OTUs in the cervicovaginal samples as described by Narrowe and co-workers (2015) [323]. Filtering of the data for the cervicovaginal samples included removal of any OTUs with less than 0.02% relative abundance and removal of singletons. Discarding low-abundant OTUs is recommended as they are likely to represent false positives (Robert C. Edgar, Personal Communication).

To compare the observed relative abundances to the expected (theoretical) relative abundances, all the detected mock species were collapsed to their respective genus. This resulted in 17 genera. A total of 12 genera (*Propionibacterium*, *Deinococcus*, *Enterococcus*, *Acinetobacter*, *Clostridium*, *Neisseria*, *Streptococcus*, *Lactobacillus*, *Staphylococcus*, *Listeria*, *Escherichia*, and *Bacillus*) did not have pronounced differences (of ≥40%) between the theoretical and observed relative abundances (Fig. 4.2). The genera, *Bacteroides* and *Helicobacter* were however considerably overrepresented (by ≥40%) while *Actinomyces*, *Pseudomonas*, and *Rhodobacter* were underrepresented (by at least 40%) (Fig. 4.2).

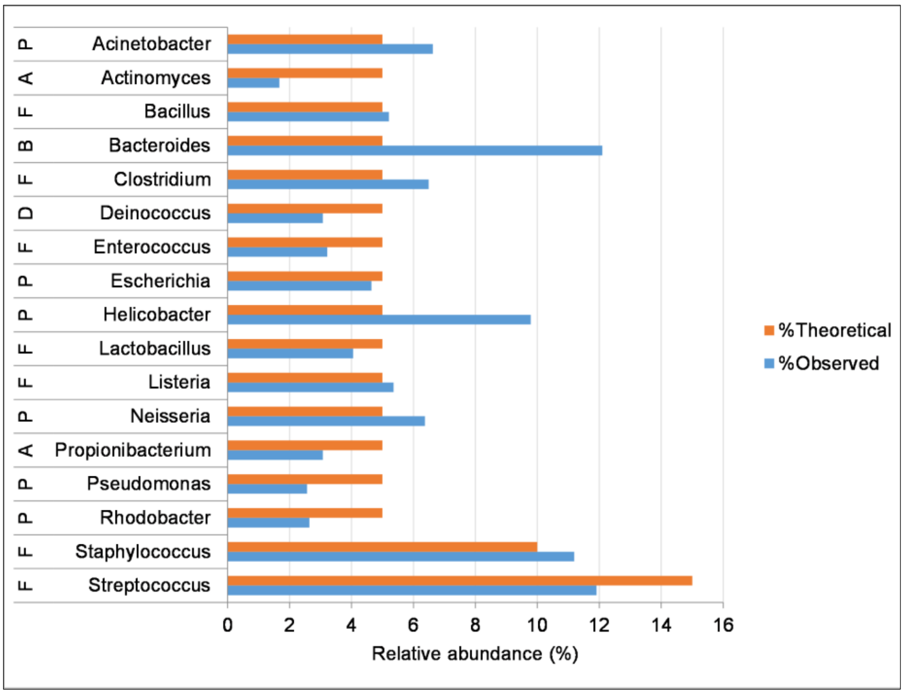


Fig. 4.2. Relative abundances of the mock genera in even mock community (HM-782D). The phyla are indicated on the left, with A representing Actinobacteria, B - Bacteroidetes, D - Deinococcus-Thermus, F - Firmicutes and P - Proteobacteria.

Similar results were obtained with the staggered community HM-783D (Fig. 4.3), although with slight variations. The mock bacterial genera *Actinomyces*, *Pseudomonas*, *Staphylococcus*, *Streptococcus*, *Enterococcus*, *Deinococcus*, *Escherichia*, and *Propionibacterium* did not have pronounced differences between the theoretical and observed relative abundances. Genera *Bacteroides*, *Helicobacter*, *Acinetobacter*, *Neisseria*, *Lactobacillus*, *Clostridium*, *Listeria*, and *Bacillus*, however, were considerably overrepresented (by ≥40%) and *Rhodobacter* was underrepresented by at least 40%.

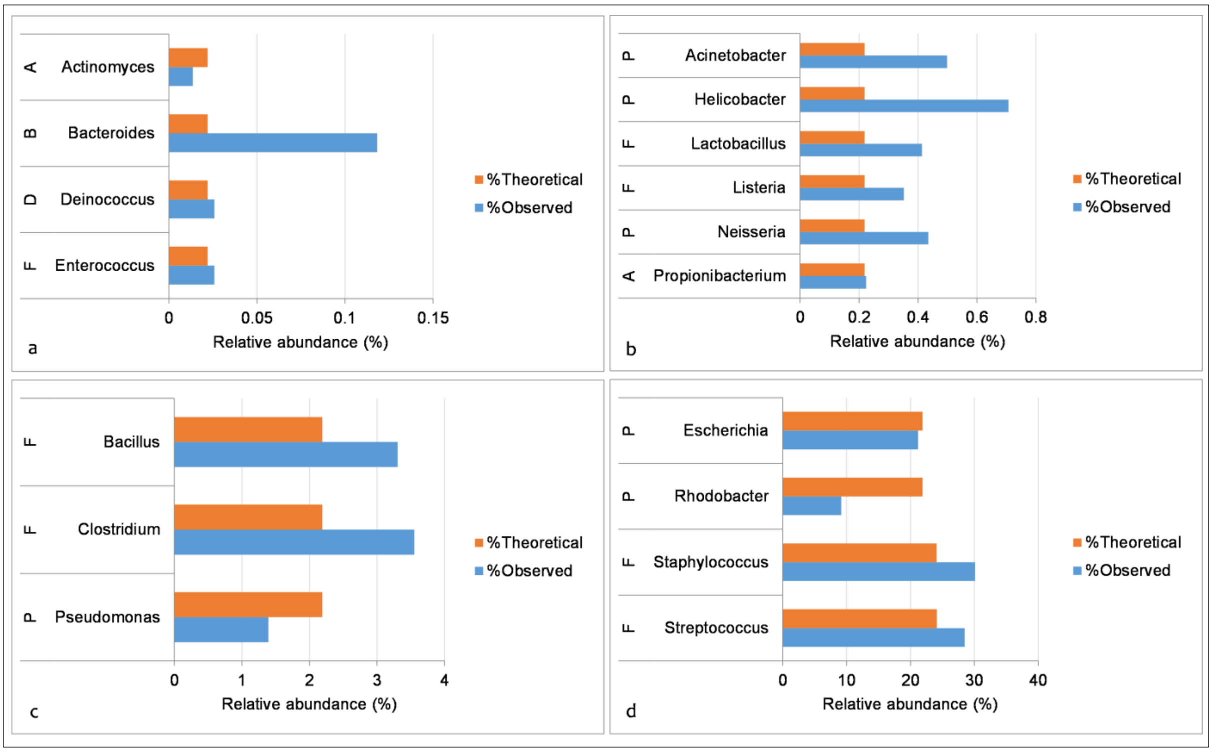


Fig. 4.3. Relative abundances of the mock genera in staggered mock community (HM-783D). The bacteria have been separated by the operon count in the mock community: a) 1,000 operons, b) 10,000 operons, c) 100,000 operons, and d) 1,000,000 operons. The phyla are indicated on the left, with A representing Actinobacteria, B - Bacteroidetes, D - Deinococcus-Thermus, F - Firmicutes and P - Proteobacteria.

To evaluate the number of false positives and negatives, abundance thresholds of 1%, 0.022%, and 0.01% were tested (Julia Ponomarenko, Personal Communication). Filtering of OTUs at a relative abundance of 1% resulted in the absence of false positives and negatives in both HM-782D and HM-783D, at the genus level. At 0.022% threshold, no false positive or negatives were detected in HM-782D; however, for HM-783D, one false negative (*Actinomyces*) was detected. Three genera (*Sneathia*, *Prevotella*, and *Shuttleworthia*) were false positives in HM-783D at the 0.01% abundance threshold. The minimal to absent false positives and negatives (in terms of reads) and the fact that OTU filtering at 0.02% relative abundance was performed, shows that our microbiota results are reliable.

4.3.2 Data analysis and output

A total of 5,694,432 high-quality 400-bp reads were finally analysed (≥12,161 reads per sample) from 87 samples. Samples excluded (n=4) were those with less than 12,161 reads. This cut-off was selected based on rarefaction curves of alpha diversity that showed that for each sample, 12,161 reads could provide a reasonable coverage (subsampling depth) of microbial diversity (Supplemental Fig. 4.2). The rarefaction analysis of observed_species and Shannon indices are shown in Supplemental Fig. 4.2a and Supplemental Fig. 4.2b, respectively. The rarefaction curves of all the samples had plateaued at 12,161 reads when using the Shannon and Simpson Index. The plots of observed_species, chao1, and PD_whole_tree metrics for many samples did not however plateau at the cut-off point. Metrics such as observed_species and chao1 did not plateau because they are sensitive to rare (low-abundant) species, which are usually incorporated in their calculations [315,324].

4.3.3 Bacterial prevalence and relative abundance in the cervicovaginal microbiotas

A total of 120 non-spurious OTUs were identified in the 87 samples (27-85 OTUs per sample). OTUs with less than 0.02% relative abundance were excluded based on the analysis of the controls in subsection 4.2.4. The OTUs were taxonomically classified by comparison to the Greengenes 16S rRNA database (v13.8). Grouping of the classified OTUs into their respective taxonomic ranks showed a total of eight bacterial phyla and 58 genera in the 87 CVM samples. Table 4.2 is a summary of the overall relative abundance and prevalence of the eight phyla detected, the corresponding major genera in each phylum and the most abundant OTUs (potential species) in these genera.

Firmicutes was the most abundant phylum, represented by 45 (37.5%) of the 120 OTUs. *Lactobacillus* was the major genus in this phylum, represented by six OTUs. The most abundant OTU (classified as *L. iners*) was found in this genus and was present in all the samples. Other *Lactobacillus* OTUs (*L. jensenii*, *L.* crispatus, *L. gasseri*, *L. reuteri*, and an unclassified *Lactobacillus* sp.) were present at significantly lower relative abundance and prevalence. Bacteroidetes was the second most abundant phylum, with *Prevotella* being the major genus in this phylum. *Prevotella* was the most diverse of the 58 genera, with a total of 21 OTUs. Among these, *Prevotella.4* was the most abundant. Actinobacteria and Fusobacteria were the next most abundant phyla, with *Gardnerella* and *Sneathia*, respectively, being the most abundant genera. Three OTUs were detected in each of these genera, with *G. vaginalis.3* and *Sneathia.3* being the most abundant. Surprisingly, the *G. vaginalis.3* and *Sneathia.3* OTUs were present in all the samples (Table 4.2).

Table 4.2. Relative abundance and prevalence of bacterial phyla, major genera in these phyla and the most abundant OTUs in these genera.

Bacterial description			Relative abundance (%)	Prevalence (% (n/N))
Phylum	**Genus**	**OTU**		
Firmicutes			49.7	100.0 (87/87)
	Lactobacillus		24.5	100.0 (87/87)
		L. iners	20.4	100.0 (87/87)
		L. jensenii	1.9	44.8 (39/87)
		L. crispatus	1.2	46.0 (40/87)
		L. gasseri	0.1	19.5 (17/87)
		L. reuteri	0.04	19.5 (17/87)
		Lactobacillus	0.7	34.5 (30/87)
Bacteroidetes			18.3	100.0 (87/87)
	Prevotella		16.2	100.0 (87/87)
		Prevotella.4	5.1	94.3 (82/87)
Actinobacteria			13.1	100.0 (87/87)
	Gardnerella		9.1	100.0 (87/87)
		G. vaginalis.3	7.2	100.0 (87/87)
Fusobacteria			11.8	100.0 (87/87)
	Sneathia		10.8	100.0 (87/87)
		Sneathia.3	7.4	100.0 (87/87)
Terenicutes			2.5	96.6 (84/87)
	Unclassified *Mycoplasmataceae*		1.1	42.5 (37/87)
		Unclassified *Mycoplasmataceae*	1.1	42.5 (37/87)
Chlamydiae			2.1	42.5 (37/87)
	Chlamydia		2.1	42.5 (37/87)
		C. trachomatis	2.1	42.5 (37/87)
Proteobacteria			2.0	100.0 (87/87)
	Unclassified *Rickettsiales*		0.5	21.8 (19/87)
		Unclassified *Rickettsiales*	0.5	21.8 (19/87)
TM7			0.5	41.4 (36/87)
	Unclassified *Rs-045*		0.5	41.4 (36/87)
		Unclassified Rs-045	0.5	41.4 (36/87)

Abbreviations: OTU – operational taxonomic unit.

The relative abundances of the top 21 genera across the 87 women are shown in Fig. 4.4. Women with higher relative abundances of *Lactobacillus* had lower relative abundances of the other genera (mostly *Prevotella*, *Sneathia*, *Shuttleworthia*, and *Gardnerella*) (Fig. 4.4), most of which have been associated with BV [51,74]. Women with lower relative abundances of *Lactobacillus* had a higher microbial diversity, characterised by the presence of polymicrobial non-*Lactobacillus* communities. The co-occurrence and co-exclusion patterns of the bacterial OTUs were therefore compared in the next subsection.

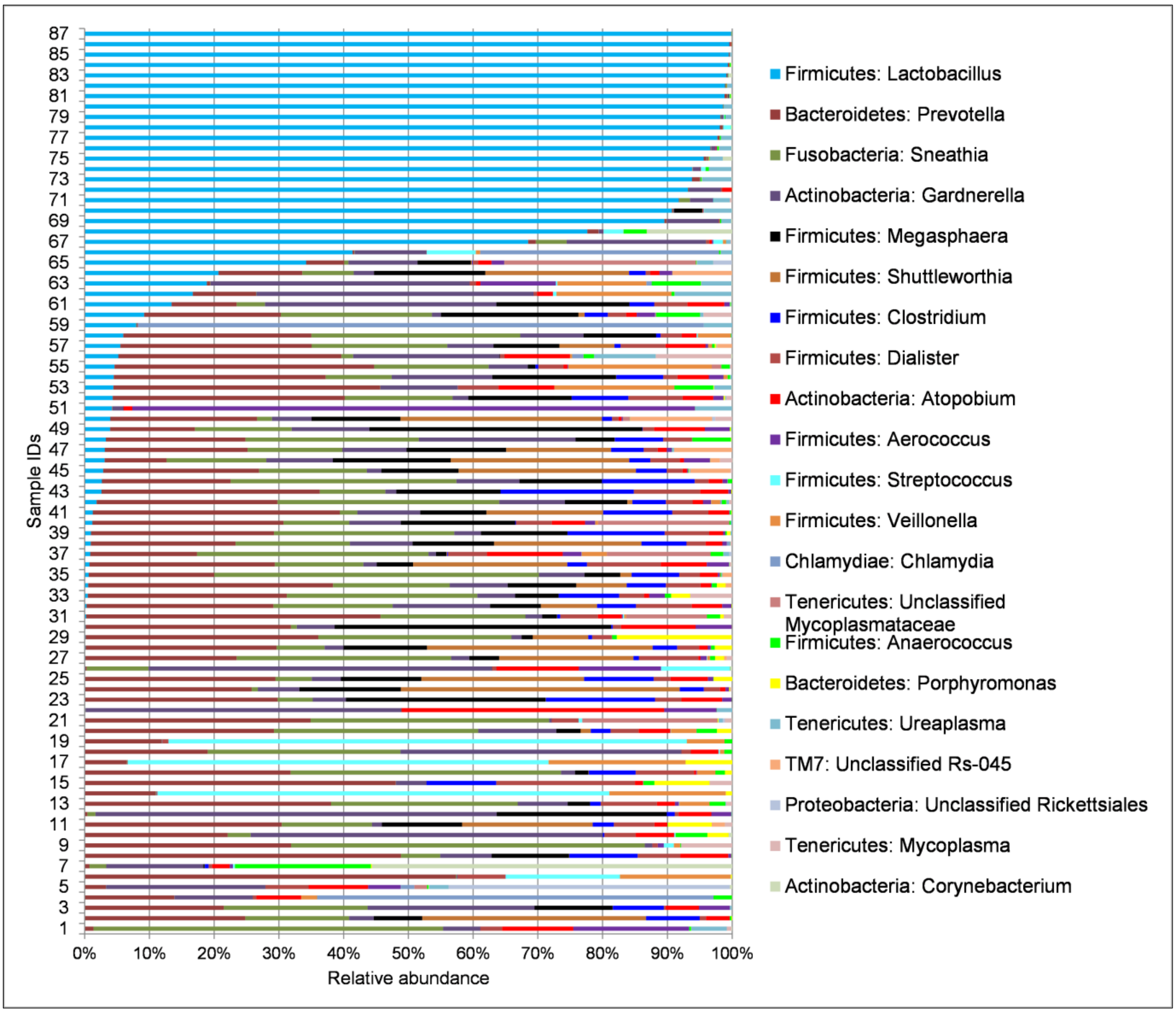


Fig. 4.4. Relative abundances of *Lactobacillus* and top 20 genera in cervicovaginal microbiotas. The samples have been ordered from the top according to reducing relative abundances of *Lactobacillus*. Both the phyla and genera names are stated.

4.3.4 Co-occurrence and co-exclusion patterns of bacterial OTUs

To gain deeper insight into the ecological relationships (oppositional and symbiotic) responsible for community architecture, pairwise correlational analyses between the abundances of all bacterial OTUs were computed. A correlogram of the pairwise correlations between the top 40 bacterial OTUs with the greatest variance is shown in Fig. 4.5.

The dendrogram shows two major clusters, designated A and B. Cluster-A, comprising 40.0% of the bacterial OTUs, mostly included *Lactobacillus* spp. (25.0%, 4/16). OTUs in Cluster-B included several bacteria associated with BV and several members in the genus *Prevotella* (29.2%, 7/24). There was a tendency of bacterial OTUs from closely related taxa having positive correlations. Each of the two major clusters could be divided into two sub-clusters (A1 and A2, and B1 and B2).

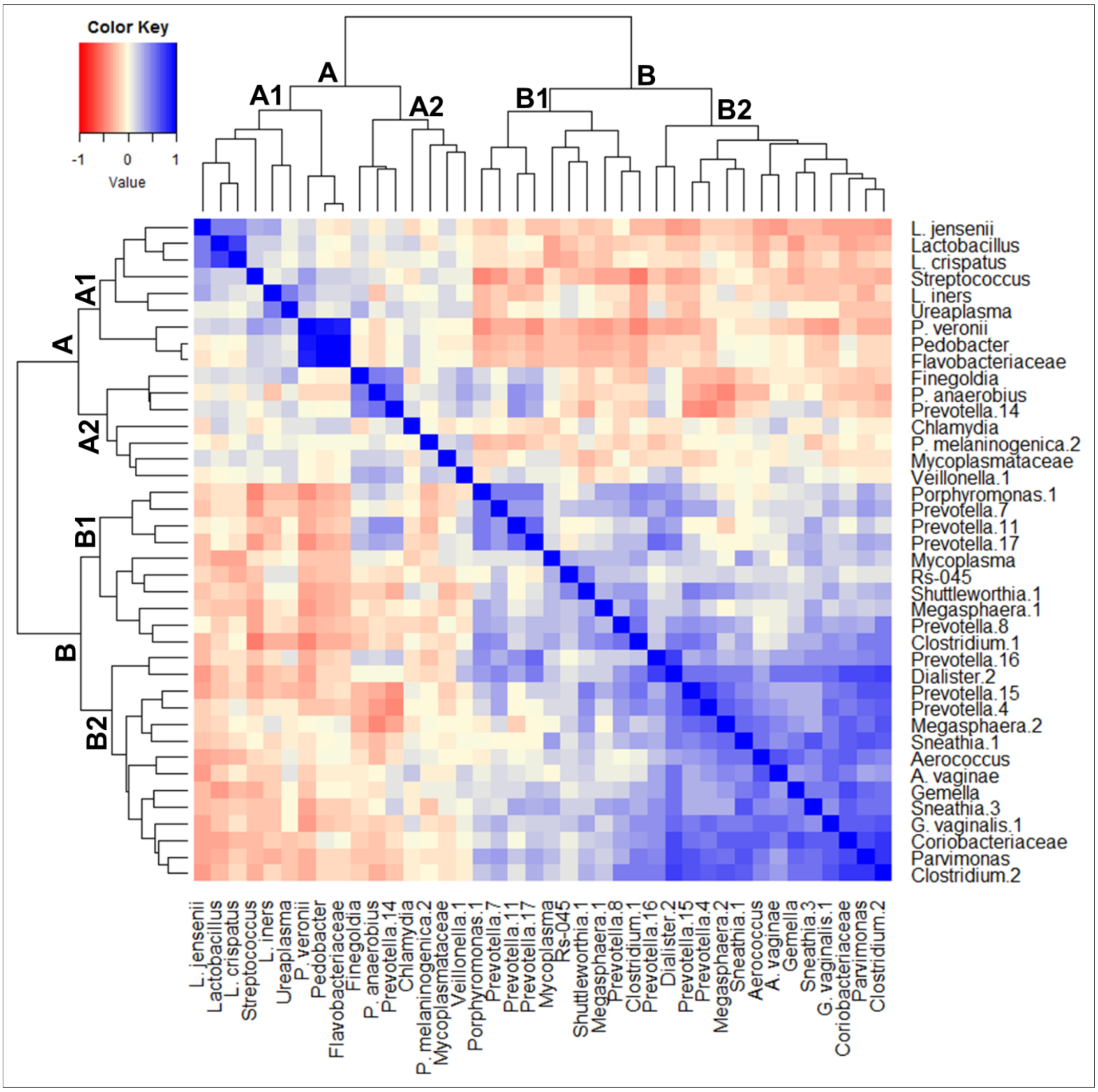


Fig. 4.5. Correlogram of bacterial OTUs showing co-occurrence and co-exclusion patterns. These were computed by Spearman's rank correlation between OTUs. The correlation coefficients range from -1 (red; co-exclusions relationships) to +1 (blue; co-occurrence relationships). High negative and positive coefficient values indicate strong correlations. The blue diagonal line represents correlations of +1. White shows absence of bacterial relationships. Only the top 40 bacterial OTUs with the greatest variance are shown.

Sub-cluster A1 constituted lactobacilli (*L. jensenii*, *L. crispatus*, *L. iners*, and an unclassified *Lactobacillus*), *Streptococcus*, *Ureaplasma*, *P. veronii*, *Pedobacter*, and *Flavobacteriaceae*. There was a strong positive correlation among *Lactobacillus* spp., with the greatest between *L. crispatus* and the unclassified *Lactobacillus* species. *L. iners* had the lowest correlation with the other lactobacilli and high correlation with *Ureaplasma*. *P. veronii*, *Pedobacter*, and *Flavobacteriaceae* had strong positive correlations among themselves. Sub-cluster A2 comprised of 9 bacterial OTUs. Among these, *Finegoldia*, *P. anaerobius*, and

Prevotella.14 had the strongest positive correlations among themselves. Inter-sub-cluster correlations between A1 and A2 were generally weaker than their respective intra-sub-cluster correlations. Interestingly, each *Lactobacillus* spp. (sub-cluster A1) had a positive correlation, although very weak, with *Prevotella.14* (sub-cluster A2).

Sub-cluster B1 comprised 10 bacteria OTUs, with *Porphyromonas.1*, *Prevotella.7*, *Prevotella.11*, and *Prevotella.17* having strong positive correlations among themselves. The other OTUs (*Mycoplasma*, *Rs-045*, *Shuttleworthia*, *Megasphaera*, *Prevotella.8*, and *Clostridium.1*) in this sub-cluster co-occurred together. Sub-cluster B2 had 14 bacterial OTUs, all which exhibited strong positive correlations among themselves, than with members in sub-cluster B1. These OTUs included *Prevotella.16*, *Dialister.2*, *Prevotella.15*, *Prevotella.4*, *Megasphaera.2*, *Sneathia.1*, *Aerococcus*, *A. vaginae*, *Gemella*, *Sneathia.3*, *G. vaginalis.1*, *Coriobacteriaceae*, *Parvimonas*, and *Clostridium.2*. The results also indicated a slight interaction gradient between sub-cluster A2 and B1.

4.3.5 Clustering of the cervicovaginal microbiotas into distinct community state types

The CVMs were hierarchically clustered by their composition and relative abundances of bacterial taxa. A heatmap of the relative abundances of the 32 major taxa in the 87 women and clustering of the CVMs is shown in Fig. 4.6.

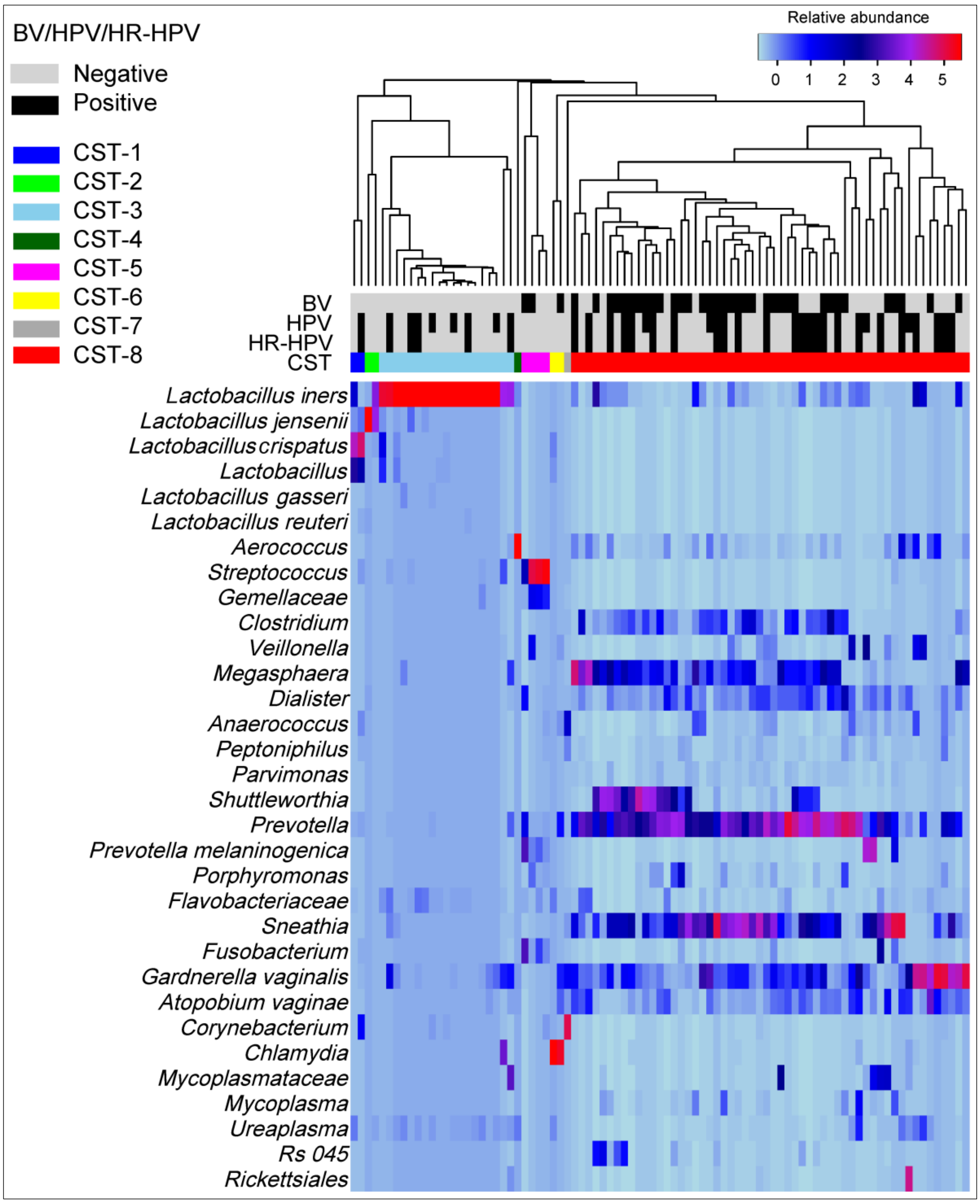


Fig. 4.6. Heatmap of the relative abundances of bacterial taxa found in cervicovaginal microbiotas of 87 South African reproductive-age women. Rows represent the bacterial taxa (sorted by classification) and columns the samples. The colour key for the relative abundances is indicated in the upper right corner. The human papillomavirus (HPV) and high-risk human papillomavirus (HR-HPV) infection status as well as bacterial vaginosis (BV) findings of the women are indicated. The dendrogram based on average linkage hierarchical clustering of the Bray-Curtis dissimilarity matrix is shown and was used to define the eight community state types (CSTs). Except for *L. gasseri* and *L. reuteri*, each with 0.03% relative abundance, the other 30 bacterial taxa included had ≥0.33% relative abundance.

The CVMs were clustered into eight CSTs (designated CST 1-8) based on the Bray-Curtis index. This is a beta diversity metric that takes into account the abundances of the shared taxa in the samples. The percentage relative abundance of the most abundant OTUs in each of the CSTs is provided in Table 4.3 (CSTs 1-7) and Table 4.4 (CST-8).

Table 4.3. Percentage relative abundance of the most predominant potential species (OTUs) in CSTs 1-7.

CST	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Most abundant Bacteria	*L. crispatus*	*L. jensenii*	*L. iners*	*Aerococcus* spp.	*Streptococcus* spp.	*C. trachomatis*	*Corynebacterium* spp.
Median (IQR)	39.9 (37.3-42.5)	72.3 (48.7-95.9)	88.7 (80.8-95.8)	86.6 (86.6-86.6)	33.3 (12.6-48.6)	70.0 (60.8-78.8)	22.3 (22.3-22.3)
Mean (Range)	39.9 (37.3-42.5)	72.3 (48.7-95.9)	82.4 (33.9-98.5)	86.6 (86.6-86.6)	31.5 (9.1-50.3)	70.0 (60.8-78.8)	22.3 (22.3-22.3)

Abbreviations: CST – community state type, IQR – interquartile range.

Table 4.4. Percentage relative abundance of the most predominant potential species (OTUs) in CST-8.

Bacteria	***Sneathia*** **sp.**	***G. vaginalis***	***Prevotella*** **sp.**	***Megasphaera*** **sp.**	***Shuttleworthia*** **sp.**
Median (IQR)	10.1 (1.6-19.4)	4.6 (1.9-9.3)	3.1 (0.0-15.5)	4.2 (0.0-12.6)	0.02 (0.0-14.0)
Mean (Range)	11.4 (0.01-49.2)	9.7 (0.1-49.9)	7.9 (0.0-29.9)	7.3 (0.0-37.2)	6.9 (0.0-41.3)

Abbreviations: CST – community state type, IQR – interquartile range.

CSTs 1-3 were dominated by *Lactobacillus* spp. and included 26.4% (n=23) of the women. In the heatmap, *Lactobacillus* spp. accounted for a high proportional abundance (≥41.2%) in each of these CSTs compared to other bacteria: CST-1 was dominated by *L. crispatus* (≥43.2% mean relative abundance), CST-2 by *L. jensenii* (≥ 49.2%), and CST-3 by *L. iners* (≥41.2%). The largest *Lactobacillus*-dominated CST was CST-3 (*L. iners*), which was found in 19 women (21.8%). CST-1 (*L. crispatus*) and CST-2 (*L. jensenii*) were each found in only 2 (2.3%) women. There was no CST dominated by *L. gasseri*.

CSTs 4-8 were not dominated by *Lactobacillus* spp. CST-8 was the largest group occurring in 64.4% (n=56) of the women. These women had diverse CVMs, mostly comprising a continuum of facultative and obligate BV-associated anaerobes such as *G. vaginalis*, *Prevotella*, *Sneathia*, *Atopobium*, *Shuttleworthia*, *Clostridium*, *Megasphaera*, and *Dialister*. The remaining CSTs occurred at a low prevalence, with CST-4 (dominated by *Aerococcus*), CST-5 (*Streptococcus*), CST-6 (*Chlamydia*), and CST-7 (*Corynebacterium*) occurring in only one, four, two and one women, respectively.

The reduction of the complex compositional data by clustering of the CVMs into CSTs allowed for statistical comparisons as discussed in the following subsection.

4.3.6 Comparison of characteristics of the women with cervicovaginal microbiotas in the community state types 3 and 8

The characteristics of the women with CVMs in CST-3 and CST-8 were statistically compared and are shown in Table 4.5. Women with CVMs from the remaining CSTs (CST-1, CST-2, and CST-4 to CST-7) were excluded from this comparison as these groups had low numbers (≤4 women).

Table 4.5. Comparison of the demographic, sociobehavioural, and clinical characteristics of the women with cervicovaginal microbiotas belonging to community state type-3 (CST-3) (*L. iners*-dominated) and CST-8 (diverse).

Characteristic	**CST-3**	**CST-8**	**p-value**[#]
	(N = 19, 25.3%)	**(N = 56, 74.7%)**	
Age (years)	29.0 (23.0-40.0)	32.5 (25.3-39.0)	0.531
HPV infection (% (n/N))			
Any HPV type	42.1 (8/19)	50.0 (28/56)	0.552
Any high-risk HPV type	26.3 (5/19)	42.9 (24/56)	0.201
Single HPV infection	15.8 (3/19)	26.8 (15/56)	0.535
Multiple HPV infection	26.3 (5/19)	23.2 (13/56)	0.765
HPV status at 6 month visit (% (n/N))			
Negative	56.3 (9/16)	38.2 (13/34)	0.231
Acquired	12.5 (2/16)	11.8 (4/34)	1.000
Cleared	0.0 (0/16)	11.8 (4/34)	0.292
Persisted	6.3 (1/16)	2.9 (1/34)	0.542
Complex[+]	25.0 (4/16)	35.3 (12/34)	0.467
Age at sexual debut (years)[^]	17.0 (15.0-19.0)	17.0 (16.0-18.0)	0.768
Lifetime number of sexual partners[^]	3.0 (2.0-5.0)	3.0 (2.0-4.0)	0.998
Number of sex acts with study partner in last month[^]	4.0 (2.0-7.0)	4.0 (2.0-8.0)	0.397
Current use of hormonal contraceptives* (% (n/N))	44.4 (8/18)	32.6 (15/46)	0.375
Cervical cytology (% (n/N))			
Normal	77.8 (14/18)	73.1 (38/52)	0.694
ASCUS	0.0 (0/18)	19.2 (10/52)	0.055
LSIL	22.2 (4/18)	7.7 (4/52)	0.190
HSIL	0.0 (0/18)	0.0 (0/52)	1.000
Experienced vaginal discharge in last 6 months (% (n/N))	10.5 (2/19)	16.0 (8/50)	0.715
Incidental BV on Papanicolaou smear (% (n/N))	0.0 (0/19)	62.5 (35/56)	**<0.0001**
Sexually transmitted infections			
Chlamydia[&]	42.1 (8/19)	41.1 (23/56)	0.937
Trichomoniasis[&]	10.5 (2/19)	5.4 (3/56)	0.596
Cigarette use (% (n/N))			
Never smoked	63.2 (12/19)	64.3 (36/56)	0.930
Ex-smoker	5.3 (1/19)	3.6 (2/56)	1.000
Current smoker	31.6 (6/19)	32.1 (18/56)	0.964

Abbreviations: HPV – human papillomavirus, ASCUS – atypical cells of undetermined significance, LSIL – low-grade squamous intraepithelial lesion, HSIL – high-grade squamous intraepithelial lesion, BV – bacterial vaginosis, CST-3 - *L. iners*-dominated cervicovaginal microbiotas, CST-8 – diverse cervicovaginal microbiotas.

[#] p-values are shown for comparison of each variable between CST-3 and CST-8. Associations of continuous variables (expressed as medians with interquartile ranges (IQRs, at 25th and 75th percentiles)) and categorical variables were computed by Mann-Whitney unpaired and Chi-square/Fisher's exact tests, respectively. Significant p-values (<0.05) are shown in **bold**.

^Data was not available on the age at sexual debut for one woman (CST-8), number of lifetime sexual partner for one woman (CST-8), and number of sexual acts with study partner in the last month for two women (CST-8).

[+] Women who had various combinations of acquired, cleared and/or persistent HPV infections at the 6-month follow-up.

*The hormonal contraceptives included oral pills, norethisterone enanthate, Depo-Provera, and steroids.

[+] The "complex" HPV status consisted of women that had cleared one HPV genotype but were infected with new genotypes and/or had other genotypes that had persisted at the 6-month follow-up.

[&] Chlamydia was detected by deep sequencing while trichomoniasis screened by wet microscopy.

None of the women with CVMs in CST-3 had BV ($p<0.0001$) or ASCUS ($p=0.055$). The prevalence, clearance, and persistence of HPV infections did not statistically differ between women with CVMs in CST-3 and CST-8. The other participants' metadata were not significantly different between CST-3 and CST-8.

Chlamydia trachomatis infection was the only bacterial STI that was detected in the sequence data, though the women in our cohort could have had other bacterial STIs but were at levels too low to detect by the targeted metagenomics method. About 42.5% of the women (37/87) had *C. trachomatis,* with *a* mean relative abundance of *C. trachomatis* of 4.9% (range: 0.001-78.8%). Its prevalence did not vary between women in CST-3 and CST-8. Trichomoniasis occurred at low prevalence in CST-3 and CST-8. One woman in each of these two CSTs was coinfected with trichomoniasis and *C. trachomatis*.

4.3.7 Comparison of alpha diversity across CSTs, BV, HPV, and high-risk HPV groups

Differences in community composition of the CSTs were assessed by comparing alpha diversity (within-sample diversity) metrics (Fig. 4.7a). The median alpha diversity of the two major CSTs (CST-3 and CST-8) was significantly different, with CST-8 having a greater diversity than CST-3. The median indices for CSTs-3 versus CST-8 were – Simpson: 0.8 (0.7-0.9) versus 0.1 (0.1-0.2), $p<0.0001$, Dominance: 0.2 (0.1-0.3) versus 0.9 (0.8-0.9), $p<0.0001$, Shannon: 0.5 (0.3-0.9) versus 2.3 (2.1-2.5), $p<0.0001$, and Shannon equitability: 0.1 (0.1-0.2) versus 0.6 (0.5-0.6), $p<0.0001$, respectively. The alpha diversities of the CSTs were also determined by rank abundance of the bacterial community populations (Supplemental Fig. 4.3). Notably, the rank abundance plots showed that CST-3 and CST-8 differed in their community compositions, with CST-8 being richer, more even and having less bacterial dominance than CST-3. Additional analysis by STAMP highlighted that 2 and 16 bacterial taxa were more enriched in CST-3 and CST-8, respectively (Supplemental Fig. 4.4). In CST-3, *L. iners* and *L. crispatus* were in greater relative abundance than in CST-8. In CST-8, BV-associated bacteria (*Atopobium*, *Gardnerella*, *Aerococcus*, *Clostridium*, *Dialister*, *Megasphaera*, *Prevotella*, *Shuttleworthia*, *Sneathia*, and *Veillonella*) were more abundant than in CST-3.

The alpha diversity of the CVMs also significantly varied with BV status (Fig. 4.7b), with BV-positive women (Shannon index: 2.3 (2.1-2.5)) having a higher bacterial diversity than BV-negative women (1.6 (0.6-2.2)), $p<0.0001$.

The alpha diversity of the CVMs of women with HPV (Fig. 4.7c) and high-risk HPV (Fig. 4.7d) infection did not differ significantly ($p<0.05$) to that in CVMS from uninfected women. The median Shannon index of the CVMs was 2.2 (0.9-2.5) for HPV-negative and 2.1 (1.7-

2.4) for HPV-positive women, p=0.747. The phylogenetic diversity, as measured by mean PD_whole_tree in QIIME, was 5.9±1.1 for HPV-negative and 5.9±0.7 for HPV-positive, p=0.808. The median Shannon index was 2.1 (0.9-2.4) for high-risk HPV-negative and 2.2 (1.8-2.5) for high-risk HPV-positive women, p=0.198. The phylogenetic diversity mean values were 5.9±1.0 for high-risk HPV-negative women and 5.9±0.7 for high-risk HPV-positive, p=0.982.

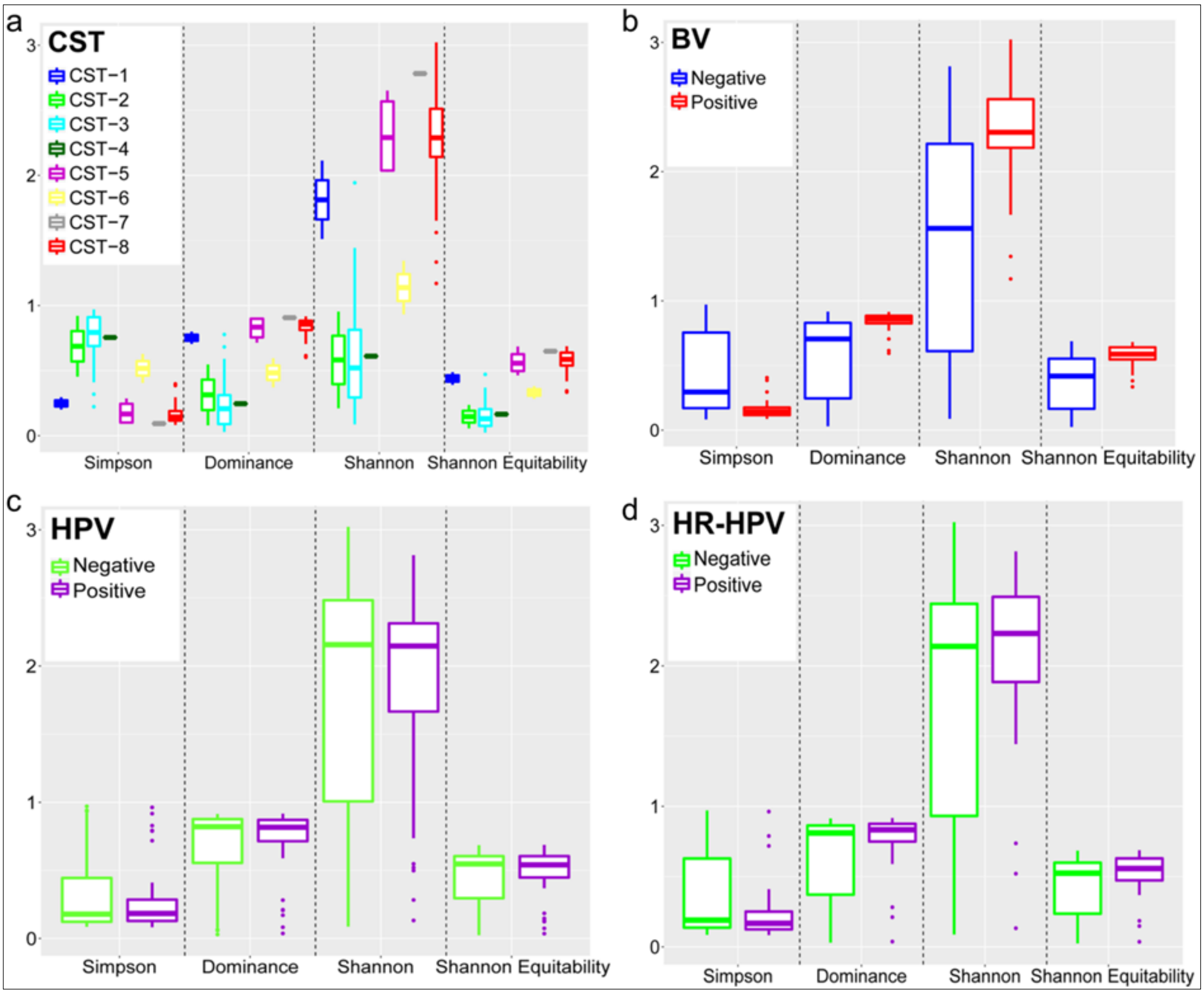


Fig. 4.7. Alpha diversity measures of cervicovaginal microbiotas (CVMs). Comparison of the alpha diversity of CVMs grouped by a) community state type (CST), b) bacterial vaginosis (BV) status, c) human papillomavirus (HPV) infection status and d) high-risk (HR)-HPV infection status. In each plot, the box ranges from the first to the third quartile, with the median represented by the horizontal line. The whiskers extend to the smallest and largest non-outliers and outliers are represented by the dots.

4.3.8 Comparison of beta diversity across CSTs, BV, HPV, and high-risk HPV groups

The between sample diversity (beta diversity) was computed using weighted UniFrac distances and compared across the CSTs, BV status and HPV infection groups. The PCoA of the weighted UniFrac distance matrix of the 87 CVMs is shown in Fig. 4.8, with CVM samples coloured by the various categories.

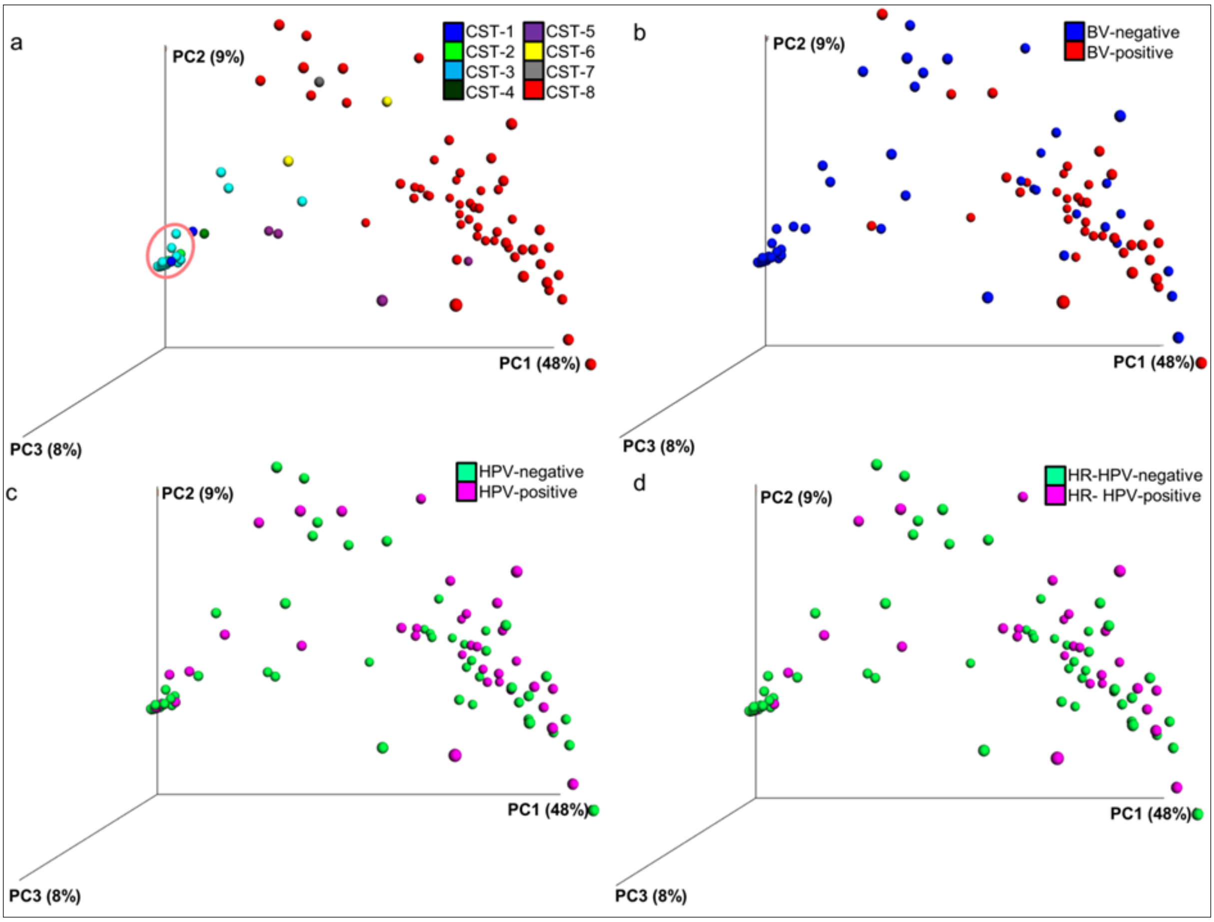


Fig. 4.8. Beta diversity of the cervicovaginal microbiotas (CVMs). Principal Coordinates Analysis (PCoA) plots of the weighted UniFrac distances of the CVMs coloured according to a) community state type (CST), b) bacterial vaginosis (BV) status, c) human papillomavirus (HPV) infection status, and d) high-risk (HR)-HPV infection status. The first three principal coordinate (PC) axes and the percentage variation explained by each (PC1: 48%, PC2: 9%, and PC3: 8%) are shown. Each solid point represents a bacterial community.

PCoA of UniFrac distances showed that the CSTs were spatially segregated and therefore compositionally and phylogenetically distinct from each other. A few samples (21.8%, 19/87) clustered tightly together (pink-circled cluster in Fig. 4.8a). This cluster consisted of lactobacilli-dominated communities: one CST-1, 2 CST-2, and 16 CST-3. The remaining four *Lactobacillus*-dominated samples (1 CST-1 and 3 CST-3) were segregated from this cluster. This is likely due to the presence of non-*Lactobacillus bacteria* at very low relative abundances in these samples.

The majority of the CVM samples from BV-negative women were distinctively segregated in the 3D space from CVM samples from BV-positive women (Fig. 4.8b). Samples from BV-negative women that were not segregated had phylogenetically diverse BV-associated bacteria such as *Gardnerella*, *Atopobium*, *Prevotella*, *Shuttleworthia* and *Clostridium*.

There was no distinct segregation of the CVMs on the PCoA plot by HPV status (Fig. 4.8c) or high-risk HPV infection status (Fig. 4.8d). The majority of the *Lactobacillus*-dominated samples (73.7%, 14/19) that clustered tightly together were HPV-negative.

4.3.9 Potential biomarkers for HPV and high-risk HPV infection

The comparison of the CVMs of HPV or high-risk HPV-infected women and -uninfected women indicated that overall community composition (as measured by alpha and beta diversity) did not differ significantly. Next, LefSe analysis was used to detect if specific bacterial taxa were differentially abundant between CVMs from women with and without HPV and high-risk HPV infection. Bacterial taxa identified as differentially abundant using LefSe were termed potential or possible "biomarkers", using the definition of Segata and colleagues (2011) [319]. High-risk HPV-positive group (34.5% (30/87) comprised of women with any of the high-risk HPV genotypes regardless of infection with any of the low-risk HPV genotypes whereas high-risk HPV-negative group (65.5% (57/87)) consisted of women that were negative for HPV and those positive for low-risk HPV.

When comparing the CVMs of high-risk HPV-infected and -uninfected women, 30 bacterial taxa were detected as significant potential biomarkers for high-risk HPV. The LDA scores and phylogenetic relationships of these taxa are shown in Fig. 4.9. At the genera level the CVMs of high-risk HPV-infected women had significantly greater relative abundances (LDA score <-2.0, p<0.05) of *Gardnerella*, *Sneathia*, *Atopobium*, *Aerococcus*, and *Pseudomonas* than women without HPV infection. At species level, the strongest association (LDA score <-4.0) with high-risk HPV infection was found with *G. vaginalis. A. vaginae* and *P. veronii* (LDA score <-2.0, p<0.05) was also significantly enriched in high-risk HPV-infected women.

Haemophilus and *Phyllobacterium* were the only genera enriched in women without high-risk HPV infection. *Lactobacillus* was not detected as being differentially abundant between women with and without HPV or high-risk HPV infections.

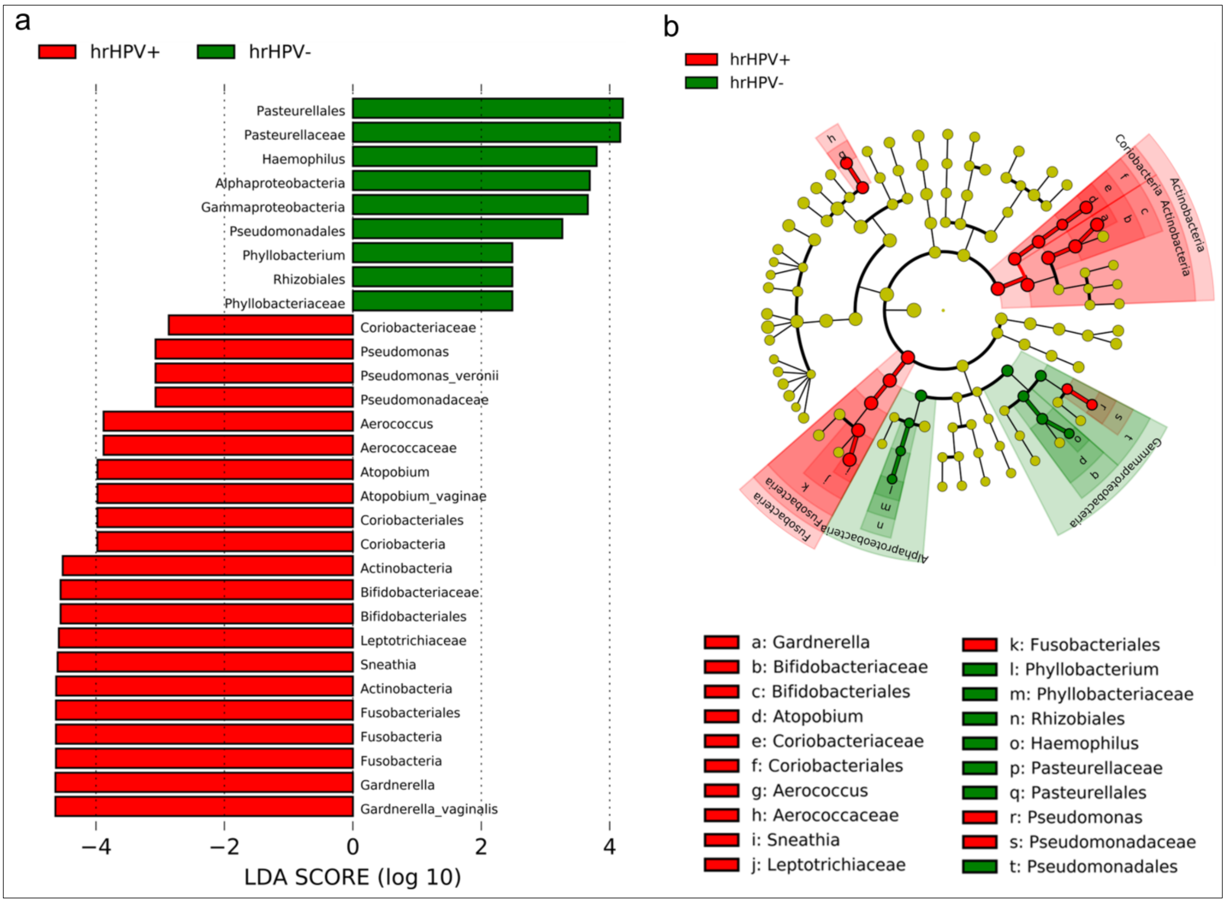


Fig. 4.9. Potential biomarkers for high-risk HPV by LefSe. a) Histogram of differentially abundant taxa in cervicovaginal microbiotas of women with and without high-risk (hr)-HPV infections, and b) A six-level cladogram with a taxonomic hierarchical structure. Each coloured solid represents a taxon and its diameter is proportional to the taxon's relative abundance. Red and green solids represent statistically significant taxon ranks in hrHPV-positive and hrHPV-negative group, respectively. Only features with logarithmic LDA scores >2.0 or <-2.0 are shown.

A total of 28 bacterial taxa were found to be differentially abundant between the CVMs of HPV-negative (57.5% (50/87)) and HPV-positive (42.5% (37/87)) women. The LDA scores and phylogenetic relationships of these taxa of these are shown in Supplemental Fig. 4.5. At the genera level HPV-positive women had higher relative abundances of *Gardnerella*, *Sneathia*, and *Aerococcus* than HPV-negative women. *Campylobacter*, *Haemophilus*, and *Pseudomonas*, were the strongest potential biomarkers for women without HPV infection.

4.3.10 Potential biomarkers for bacterial vaginosis

LefSe was also used to identify differentially abundant bacterial taxa in women with and without BV. The specific taxa ranks that were differentially abundant (LDA score >2.0 or LDA score <-2.0, p<0.05) in women with and without BV are shown in Fig. 4.10. *L. iners* and *L. reuteri* were found to be highly enriched in women without BV. Women who were positive for BV (43.7% (38/87)) had significantly higher relative abundances (LDA score <-4.0, p<0.05) of genera *Sneathia*, *Megasphaera*, *Clostridium*, *Atopobium*, and *Gardnerella* than women without BV (56.3% (49/87)). Most of these bacteria have been associated with BV [38,141]. Among the genera detected, *Staphylococcus* was the most abundant in the CVMs of women without BV. We observed that *L. iners* was more abundant in women without BV than in women with BV.

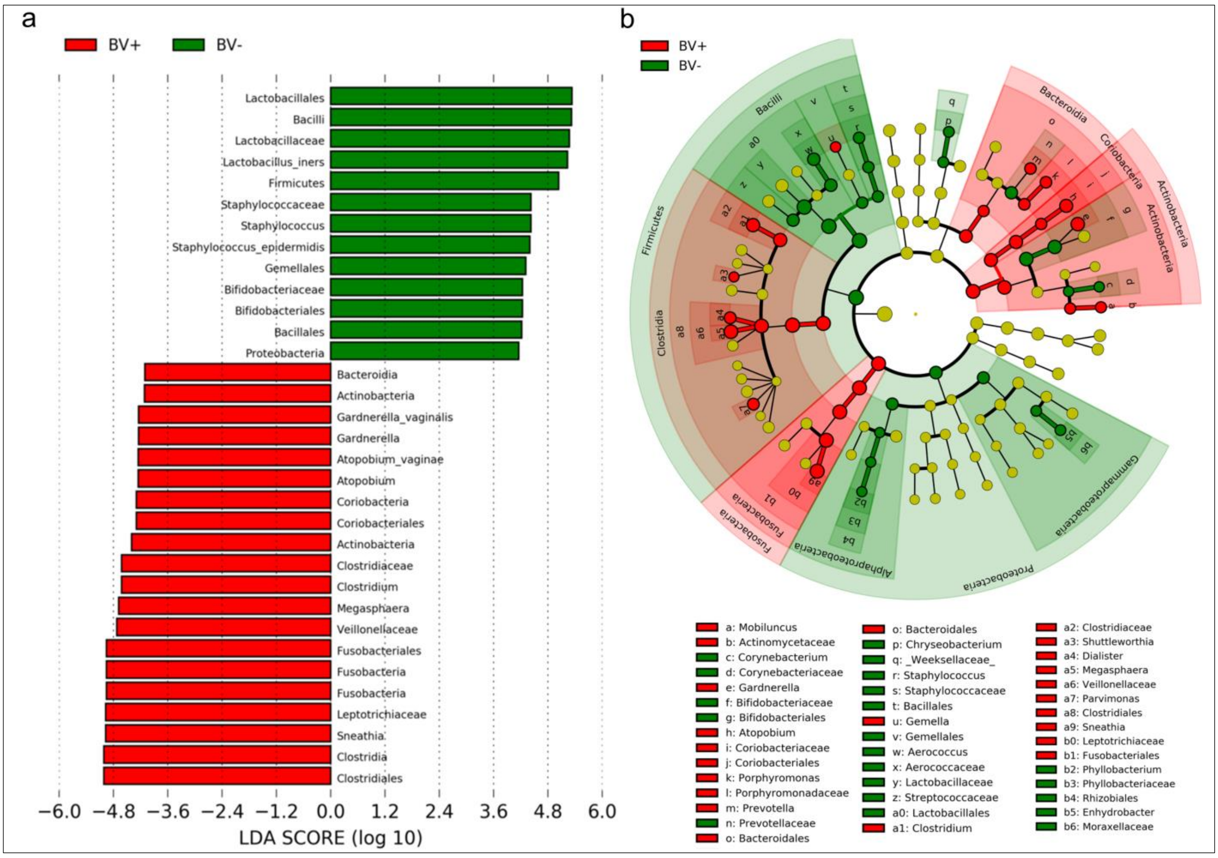


Fig. 4.10. Potential biomarkers for bacterial vaginosis by LefSe. a) Histogram of differentially abundant taxa in cervicovaginal microbiotas of women with and without bacterial vaginosis (BV), and b) A six-level cladogram with a taxonomic hierarchical structure. Each coloured solid represents a taxon and its diameter is proportional to the taxon’s relative abundance. Red and green solids represent statistically significant taxon ranks in women with and without BV, respectively. For purposes of better visualization, only features with logarithmic LDA scores >4.0 or <-4.0 are shown in the histogram. For the cladogram, all features with logarithmic LDA scores >2.0 or <-2.0 are shown.

4.3.11 Predicted metagenome functions across CSTs, BV and high-risk HPV

Functional profiling of the bacterial communities was predicted using PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) [317] as described in the Material and Methods (subsection 4.2.8). This software uses evolutionary modelling to predict the potential functional composition of bacterial communities using marker gene (such as 16S rRNA) sequence data and a reference database. The KEGG (Kyoto Encyclopaedia of Genes and Genomes) Module database was used for functional annotation. Metagenome functional profiles were generated for the CVMs (in CST-3 and

CST-8), BV-positive and BV-negative women as well as women high-risk HPV-positive and high-risk HPV-negative women.

The relative abundances of the predicted KEGG functional modules in CST-3 and CST-8 were compared with STAMP [279]. The KEGG modules found to be significantly enriched are shown in Fig. 4.11. A significant enrichment of transport permeases, including ATP-binding cassette (ABC) transporters, was found in CST-8 compared to CST-3.

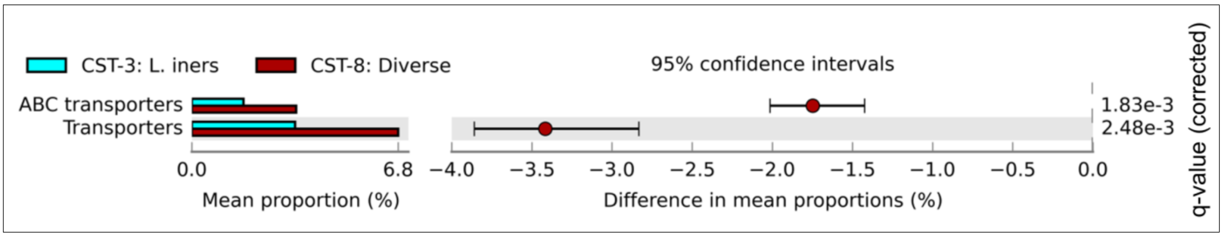


Fig. 4.11. Extended error bar plots depicting enriched KEGG functional modules in two community state types. Only functional modules identified by PICRUSt with greater than zero difference between their percentage proportions in CST-3 and CST-8 with an effect size ≥1.0 and q-value ≤0.05 are shown. The statistical comparison was carried out in STAMP. The p-values (adjusted by Benjamini Hochberg correction to account for False Discovery Rates), effect size and confidence intervals (0.95, DP: bootstrap) computed by White's non-parametric t-test (two-sided type) are shown. The difference in mean proportions and 95% confidence intervals are shown on the bar plots.

The relative abundances of the KEGG functional modules in CTS-3 and CST-8 were also compared with LefSe [319]. A total of 13 functional and metabolic modules were differentially abundant (LDA scores >3.0 or <-3.0, p<0.05) as indicated in Fig. 4.12. The relative abundances of transporters and ABC transporters were significantly higher in CST-8 than in CST-3 (Fig. 4.12), as was found when using STAMP (Fig. 4.11). The relative abundances of the transporters in CST-3 and CST-8 CVMs are shown in (Supplemental Fig. 4.6). Additional modules identified by LefSe analysis to be more abundant in CST-8 than in CST-3 included those involved in bacterial chemotaxis, bacterial motility, and flagella assembly (Fig. 4.12). The modules for ribosome biogenesis and energy metabolism were less abundant in CST-8 compared to CST-3.

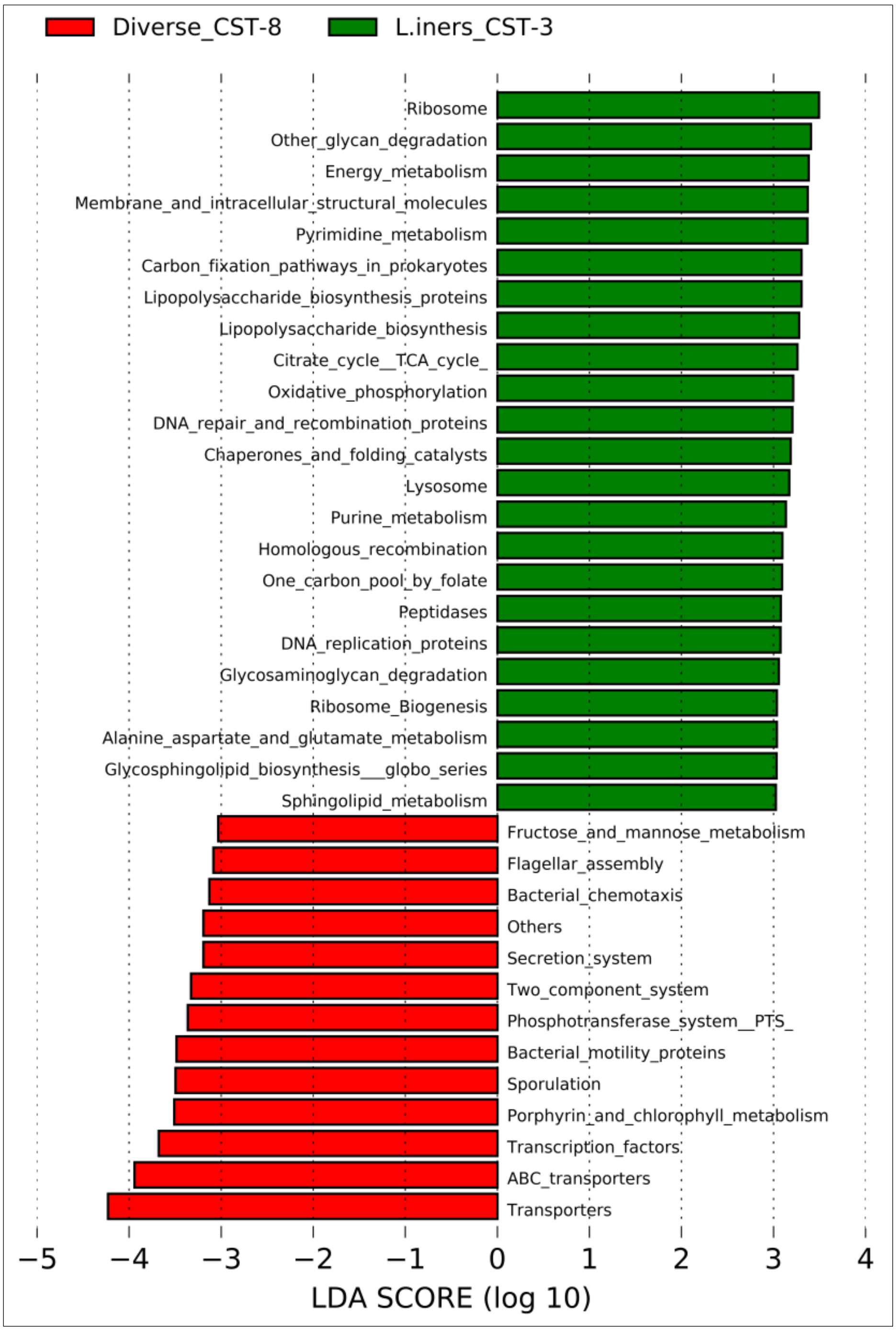


Fig. 4.12. Histogram of the differentially enriched KEGG pathway and metabolic functional modules in community state types (CSTs) CST-3 and CST-8. Only modules with a logarithmic LDA score >3.0 or <-3.0 are shown.

KEGG functional contents in BV-positive versus BV-negative women were also compared (Supplemental Fig. 4.7). CVMs from women with BV were enriched with modules such as transporters, transcriptional factors, sporulation, fructose and mannose metabolism, porphyrin metabolism, and phosphotransferase system that were also found in CVMs of women in CST-8.

The CVMs of women with and without high-risk HPV were found to be enriched in 11 and 20 functional modules, respectively (Fig. 4.13). Sporulation and transcriptional factors were the most significantly enriched functional modules in high-risk HPV-positive women. Membrane and intracellular structural molecules were the most enriched modules in women without HPV infection.

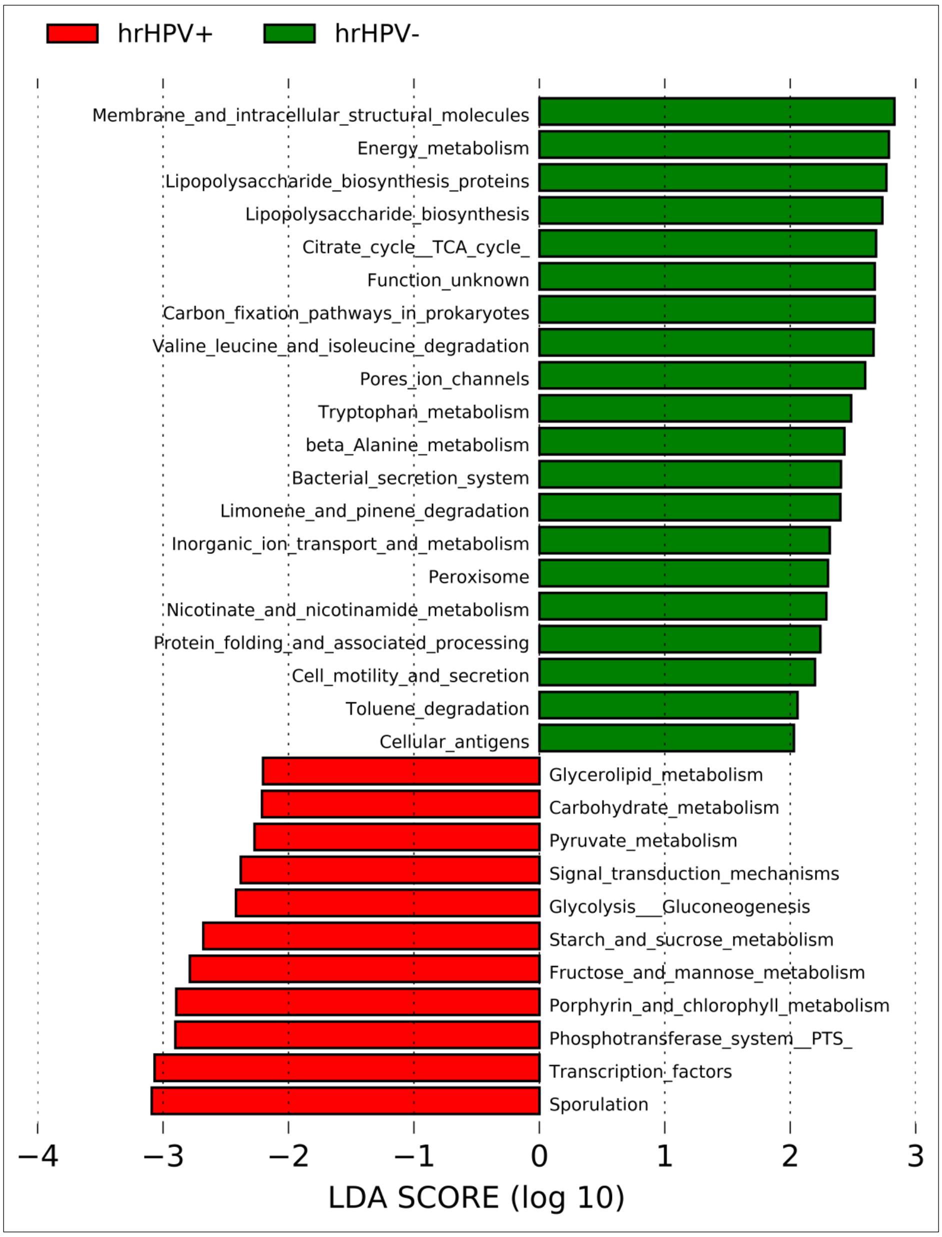


Fig. 4.13. Histogram of the differentially enriched KEGG pathway and metabolic modules in the cervicovaginal microbiotas (CVMs) of high-risk HPV-positive (hrHPV+) and high-risk HPV-negative (hrHPV-) women. Only modules with logarithmic LDA scores >2.0 or <-2.0 are shown.

Several functional modules were found to be significantly differentially abundant in all three comparisons (CST, BV and high-risk HPV). Sporulation and transcription factors were enriched in the CVMs of CST-8, BV-positive and high-risk HPV-positive women. Membrane and intracellular structural molecules, lipopolysaccharide (LPS) biosynthesis, amino acid metabolism, TCA cycle, and protein folding modules were enriched in CVMs of CST-3, BV-negative and HPV-negative women.

4.4 Discussion

4.4.1 Benchmarking 16S rRNA metagenomics analyses

Most 16S rRNA metagenomics studies have overlooked the effects of reagent contaminants, sample cross-contamination, PCR and sequencing artefacts on microbiome analyses; yet it is known that such artefacts are inherent and can cause a significant overestimation of diversity [274,315,325]. For this reason, inclusion of controls in each run is strongly recommended [303,315,322]. In this study, nuclease free water and extraction controls were concurrently run with the CVM samples throughout the experimental and computational processes to assess the presence and level of contaminating bacteria. These negative controls contained insignificant traces of bacterial contaminants and were in agreement with the findings of Salter and co-workers (2014) [325]. This indicated that the level of sample contamination was negligible.

Mock bacterial communities were additionally included in our study to estimate the levels of contamination, rate of PCR and sequencing errors, and classification errors. The mock communities included in this study performed similarly compared to previous surveys [295,297,305]. For example, a marked overrepresentation and underrepresentation of certain mock genera, such as the overrepresentation of *Bacteriodetes*, was observed and are comparable to findings by Cozzuto and co-workers [297]. Variations between the theoretical and observed relative abundances are expected [302,314] and could be due to primer bias [49,295,301,302], amplification artefacts associated with multiple primer combinations [49], and/or bias in sequencing technology [295,302,305]. Misclassification of certain species (e.g., *N. cinerea*) and failure to assign others a species name (e.g., *S. mutans*) has been demonstrated to be a limitation of the Greengenes database [303,322]. Of the 20 bacteria in the mock communities, we were able to amplify, sequence, and detect nineteen. *S. aureus* was the only mock bacteria not detected. This is likely due to the fact that *S. epidermidis* and *S. aureus* share 99.4% identity in their 16S rRNA V3 and V4 nucleotide sequences, and their reads were likely clustered together as a single OTU [320,321] and misclassified as *S. epidermidis.* This might explain why the observed relative abundance of genus *Staphylococcus* (even mock: 11.2%, staggered mock: 30.1%) was close to the expected relative abundance (even mock: 10%, staggered mock: 24.096%), which solely relied on the counts of *S. epidermidis*.

Low levels of contaminants or non-mock bacteria were observed in the mock communities, with *L. iners,* at a relative abundance 0.02%, being the most abundant in HM-

782D. Low-level contamination in mock communities has been reported elsewhere [274,297,314,323] and may be due to a variety of factors such as PCR cross-contamination [303,323], reagent contaminants [325], carryover contamination amid sequencing (0.1% rate) [303], unknown sequencing errors [305,315], and index misassignments (0.06% rate) [303]. Non-mock taxa in our study could have also emerged as chimeras from the most abundant taxa [274,295,303]. Because such error rates and artefacts can impact the validity of the perceived diversity, quality-filtering is as an essential step in metagenomics [302,303,315]. To date, there is no universal bioinformatics strategy for filtering background signals using mock communities [303]. In this study, the approach described by Narrowe and co-workers (2015) [323] was modified and included removal of singletons and filtering OTUs at a threshold of 0.02% abundance. The 0.02% threshold was the relative abundance of the most predominant non-mock bacteria (*L. iners*) detected in the even mock community in our study. A caveat of filtering OTUs using relative abundance error is that it might cause low discovery rates of rare OTUs [323]. In our study, false positives and negatives were absent at genus level in HM-782D and very low in HM-783D, when filtering at 1%, 0.022% and 0.01% abundances. This very low rate/absence of false positives and negatives coupled with the other assessments of the bacteria in the controls ensured that the diversity and relative abundance of the bacteria in the CVMs in our study are reliable.

4.4.2 Composition and community state types of *Lactobacillus*-dominated cervicovaginal microbiotas of the Black South African women

In this chapter the CVMS of 87 women were characterised and eight CSTs identified based on bacterial community composition. A total of three CSTs were found to be dominated by *Lactobacillus* species. The prevalence of *Lactobacillus*-dominated CSTs in the 87 women was relatively low (26%). Among the *Lactobacillus*-dominated CSTs, the *L. iners*-dominated group (CST-3) was the most prevalent while *L. crispatus* (CST-1) and *L. jensenii* (CST-2) were less common. Several studies have shown that among the lactobacilli, *L. iners* is most likely to colonise the CVMs of reproductive-age HIV-negative African women and African American women [33,38-40,51,87,137,138]. These studies, and two recent studies in South Africa [87,136], have found the prevalence of *L. iners*-dominated CVMs to range between 31-65% among HIV-negative African women. Of the 87 women in our study, 22% had *L. iners*-dominated CVMs. This prevalence is less than what has been previosuly documented and what we reported in Chapter 2 (39%). Compared to other lactobacilli, *L. iners* is known to be both less stable [77] and protective [159,163], and capable of surviving in dysbiotic cervicovaginal environments [82,155] because of its genomic plasticity [155]. Unlike the other common *Lactobacillus* spp., *L. iners* colonisation is less dependent on glycogen (the main carbon source in the vagina)

[61], whose breakdown leads to production of lactic acid that is responsible for low vaginal pH [63,68,121]. Comparative meta-transcriptomics has shown that *L. iners* can also metabolise non-glycogen carbon sources, producing short-chain fatty acids that are known to increase vaginal pH [121,155]. Such acids include succinate, one of the biomarkers for BV [121,155]. Additionally, the metabolite profile of *L. iners* has been shown to be intermediate between that of *L. crispatus*/*L. jensenii* and BV-associated bacteria [121]. *L. iners*-dominated CVMs have also been associated with STIs, such as HIV, HPV, HSV-2, and *C. trachomatis* [33,146,159,165]. A study, for example, that characterised the CVMs of 93 women in Netherlands contact-traced by *C. trachomatis*-positive sexual partners [165] found that women with *C. trachomatis* infection were more likely to have CVMs dominated by *L. iners* (or diverse anaerobic bacteria) than by *L. crispatus*.

The present study observed a very low prevalence of CST-1 (*L. crispatus*) (2.3%). This is much lower than reported in other studies of African women, including two South African studies, where the prevalence of *L. crispatus*-dominated CVMs ranged between 7-22% [33,38,40,87,136,138]. *L. crispatus* is thought to be the most protective *Lactobacillus* species against BV [33,163,172] and STIs such as HPV [33,34,139], HIV [33,146,161], HSV-2 [33], chlamydia [158,159,165,172], and other bacterial and viral STIs [33]. It is typically a H_2O_2-producing *Lactobacillus* species [102,148] and produces high amounts of both isoforms of lactic acid, especially D-lactic acid [146,147]. Cervicovaginal milieus colonised with *L. crispatus* are known to have the lowest vaginal pH [38].

A low prevalence of CST-2 (*L. jensenii*) (2.3%) was observed. This was consistent with other studies that have found its prevalence to be up to 1% among African cohorts [33,38,40,136,138]. *L. gasseri* and *L. reuteri* were detected as low-abundant *Lactobacillus* spp. in the CVMs, but not as distinct CSTs. *L. gasseri*-dominated CVMs are very uncommon among African women (<5%) [38,40] and have not been reported among South African women [87,136,138]. *L. gasseri*-dominated CVMs have been associated with more rapid clearance of HPV than other CSTs, including *L. crispatus*-dominated [34]. In agreement with previous studies in African women [40], the prevalence of CVMs with multiple lactobacilli dominance at abundance ratios of at least one third to half was rare in our cohort (4.6%). Genetic or behavioural differences in women may explain the paucity of certain *Lactobacillus* spp. dominating the CVMs of African women [38-40,51,180].

4.4.3 Composition and community state types of non-Lactobacillus-dominated cervicovaginal microbiotas of the Black South African women

A total of five non-*Lactobacillus*-dominated CSTs (CST 4-8) were detected in the CVMs of the 87 women. CST-8 was the most prevalent among all the women (64%) and among the five non-*Lactobacillus*-dominated CSTs (88%). The CVMs in CST-8 were diverse and heterogeneous communities with many BV-associated bacteria. The other low prevalence non-*Lactobacillus*-dominated CSTs were dominated by *Aerococcus* (CST-4), *Streptococcus* (CST-5), *Chlamydia* (CST-6), and *Corynebacterium* (CST-7). The structures of each of these CSTs and their implications are discussed below:

4.4.3.1 Diverse and heterogeneous non-Lactobacillus-dominated cervicovaginal microbiotas of the Black South African women

CST-8 was the most common CST (64%) among the 87 women. It consisted of diverse and heterogeneous communities with low relative abundances of *Lactobacillus* spp. and high relative abundances of many BV-associated taxa such as *Gardnerella*, *Sneathia*, *Prevotella*, *Megasphaera*, *Atopobium*, *Shuttleworthia*, and *Clostridium*. This CST is similar to CST IV as defined by Ravel and co-workers (2011) [38]. The predominance of CST-8 confirmed previous findings [38-40,136] that have reported a high prevalence (up to about 63%) of BV-associated diverse CVMs among African cohorts. The predominance of these BV-associated bacteria (especially *Gardnerella*, *Sneathia*, *Shuttleworthia*, and *Prevotella*) in the present study is similar to our previous observations (Chapter 2).

Analysis of the alpha diversity of CVMs in CST-8 indicated greater richness and evenness, i.e., more diversity, than those in *Lactobacillus*-dominated CSTs. The high richness and evenness of the diverse communities herein reinforces earlier findings [38,40,135] that have observed low *Lactobacillus*-dominance and increased evenness and richness of BV-associated bacteria in these CVMs. Increased community diversity has been associated with BV [33,141,326]. Expectedly, a strong association between CST-8 and BV was observed, although not all women (38%) in this CST had BV as detected by cytology. A study by Gajer and co-workers (2012) [162] showed that diverse CVMs may persist without manifestation of high Nugent scores indicative of BV [162]. The evidence that some women with diverse CVMs were BV-negative provides more support to the perception that lactobacilli may not be the only biomarkers for health [327]. It is apparent that each diverse community is unique and that diverse CSTs, including CST-8, are clusters of CVMs with a continuum of different community compositions [33,38-40,135]. It is increasingly being documented that some diverse non-*Lactobacillus*-dominated communities may constitute normal CVMs [38-40,51,173,326], particularly among Black women [38-40,51]. Such CVMs do however tend to occur in less acidic milieus (pH 4.7-6.2) [38,177] and to vary in stability and resilience [40,54,135,162,188]. Further, increased community diversity has been correlated with altered metagenome functions [122], increased immune activation [122,136,138], host proteome alterations [328], disruption of cervicovaginal epithelial integrity, and impaired wound-healing capacity [122]. Adverse reproductive health outcomes associated with BV-associated diverse communities include heightened genital proinflammatory responses [136,164,199], preterm delivery, abrupt and persistent postpartum microbiota disturbance [188], detectability and delayed clearance of HPV [33,34,137,141,157,179,214], intraepithelial lesions/neoplasms [140,160], and cervical carcinogenesis [142], and increased susceptibility to HIV [36,164], HSV-2, and other STIs [33,165,172]. In this study, all women with ASCUS cytology were in CST-8. Diverse CVMs have been found to be more common in women with ASCUS than in women with normal cervical cytology [140]. No association of CST-8, however, with HPV and high-risk HPV infection was observed as discussed later.

L. iners was present in all the CVMs of women in CST-8, with a mean relative abundance of 2.6%. Studies have shown that other *Lactobacillus* spp. such as *L. crispatus*, are rarely or never found in BV-associated diverse CVMs [33,39,40,110,138,139,162,163], such as CST-8. The detectability of *L. iners* in all the diverse CVMs in our study offers a glimpse into its role in these complex ecologies. Proteomics, metatranscriptomics, and epithelial adhesion assays have shown that *L. iners* can survive in a non-*Lactobacillus*-dominated environment [77,82,155,156]. Conceivably, some of these communities might have been transitioning between a *Lactobacillus*-dominated state and diverse CVMs via *L. iners* [54,135,162,188].

4.4.3.2 Other non-Lactobacillus-dominated cervicovaginal microbiotas of the Black South African women

The other CSTs (4-7) detected could not fit into any of the five CSTs (I-V) established by Ravel and co-workers (2011) [38]. They were *Lactobacillus*-deficient, did not constitute highly diverse and heterogeneous microbiotas, and were dominated by *Aerococcus* (CST-4), *Streptococcus* (CST-5), *Chlamydia* (CST-6) or *Corynebacterium* (CST-7). These bacteria are, however, usually not in such high abundances that they exclusively dominate the CVMs [33,38,39,51,137,326]. Hence, the presence of CSTs 4-7 shows that the CVMs, especially of African women, are still not yet completely characterised. Additional CSTs have been reported in several studies [39,40,142] for example, Proteobacteria-dominated CVMs were found in Nigerian women [137]. *Streptococcus*-dominated CVMs have been reported in one study of White women [147].

In the current study, one woman had CST-4 (CST dominated by *Aerococcus*) and four women had CST-5 (CST dominated by *Streptococcus*). The overall prevalence of *Aerococcus* and *Streptococcus* in the 87 women was high (78.2% (n=68) and 71.3% (n=62), respectively), but at a relatively low mean relative abundance (2.2%, and 1.6%, respectively). Both *Aerococcus* and *Streptococcus* are lactic acid bacteria in the same order (*Lactobacillales*) as *Lactobacillus*. Several lines of evidence indicate that these lactic acid producing bacteria, similar to *L. iners*, occur in both healthy and dysbiotic states and may have a bidirectional roles in genital health and disease. *Aerococcus* has been found in high relative abundances in healthy African women [39], but also as a predominant bacteria in diverse CVMs and associated with BV [38,326]. *Streptococcus* is a predominant cultivable non-*Lactobacillus* vaginal bacterium among South African women [42] and common in women with depleted abundances of lactobacilli [38,157,162]. *Streptococcus* and *Aerococcus* have been reported as two of the bacteria populating the cervicovaginal milieu following depletion of *Lactobacillus* in women who cleared HPV infections [157]. A recent study in South Africa strongly associated *Aerococcus* with genital inflammation *in vivo*, particularly in women with diverse CVMs [136]. Further, it has been shown that during menses, *L. gasseri*-dominated CVMs can be transiently supplanted by *Streptococcus*-dominated CVMs without a corresponding reduction in levels of lactic acid [162]. It might be crucial to investigate the identity of the *Streptococcus* in our cohort as *Streptococcus agalactiae* (GBS) has associated with AV [75,88,127,133].

Two women (2%) in our study had CVMs dominated by *Chlamydia* (CST-6). There is no previous report in the literature of *Chlamydia* dominating the CVMs of women. The species identified in the present study is *C. trachomatis*, which is a pathogenic bacterium, frequently known to have infected women currently diagnosed with infectious infertility [175]. The *Chlamydia*-dominated CVMs in our study are likely to be full-blown *C. trachomatis* infection and not a CST. In this study, chlamydia was the only bacterial STI detected by deep sequencing. The prevalence and mean relative abundance of *Chlamydia* as determined by sequencing of the V3-V4 regions of the 16S rRNA gene, was 42.5% (n=37) and 2.1%, respectively. The detection of *C. trachomatis* using 16S rRNA sequencing is probably not an accurate method for determining the prevalence of *C. trachomatis* as the broad range PCR assay is not sensitive for *C. trachomatis*. An accepted diagnostic assay for *C. trachomatis* would need to be performed in order to accurately assess its prevalence in the women and to determine the concordance between these 16S rRNA sequencing results and the diagnostic assay. Lennard and colleagues (2018) [87] recently identified *C. trachomatis* as one of the most common STIs among young Black South African women, with its prevalence (43%) in young women (16-22 years) in Cape Town being 2.5-fold higher than in women of similar age group in Johannesburg (17%). *C. trachomatis* has also been cited as the most prevalent STI among young women (18-23 years) in Durban (13%) [136]. Several studies have proposed that high *C. trachomatis* prevalence may be due to various behavioural and physiologic factors [177], including elevated vaginal pH [70], lack of appreciable numbers of protective *Lactobacillus* such as *L. crispatus* [158], and/or predominance of indole-producing bacteria such as *Porphyromonas asaccharolytica* and *Peptoniphilus harei* that enhance chlamydial colonisation and survival during tryptophan starvation (orchestrated by host IFN-γ-mediated immune response) [172].

The study also detected a single CVM dominated by *Corynebacterium* (CST-7). The prevalence and mean relative abundance of *Corynebacterium* in all 87 CVMs was 72.4% (n=63) and 0.59%, respectively. *Corynebacterium* has been found in up to moderate abundances in women with diverse CVMs [33,143,162], and BV-positive European women [39]. One study found that it was highly prevalent (90%), with minimal effect on clustering [33]. We observed that women without BV had higher relative abundances of *Corynebacterium* than women with BV. In contrast to this, higher relative abundances of *Corynebacterium* have been associated with BV in women of European ancestry [39]. A study of HPV infection and CVMs of African and Caribbean women living in Canada [141] observed higher relative abundances of *Corynebacterium accolens* in HPV-negative women. No association with HPV or high-risk HPV infection was however observed in this study or in the study by Di Paola and co-workers (2017) [157]. Further studies will be needed characterise role of *Corynebacterium* in CVMs in African women.

4.4.4 Associations of predominant community state types with demographic, sociobehavioural, and clinical characteristics of the women

CST-3 (with *L. iners* dominance) and diverse and heterogeneous CVMs (CST-8) were the most common CSTs in our cohort. Among the participant characteristics examined, only incidental BV was statistically different between CST-3 and CST-8. All the women with BV were in CST-8. Similar to our finding on association between CST-8 and BV, Gautam and colleagues (2015) [138] reported an association of BV with diverse and heterogeneous CVMs among multi-ethnic African women from Kenya, Tanzania, Rwanda, and South Africa. BV is often common in women with *Lactobacillus*-deficient and diverse and heterogeneous CVMs [38,87]. In our study the overall prevalence of BV was 44% (among the Cape Town women). The prevalence of BV has been shown to be higher among women from Cape Town (51%) compared to women from Johannesburg (36%) [87]. In terms of cervical cytology, we observed a trend towards an association of CST-8 with ASCUS. All the women with ASCUS were in CST-8. There are very limited studies that have examined and found associations of CVMs with ASCUS. A study by Mitra and co-workers (2015) investigating the relationship between CIN disease progression and diversity of CVMs [140] observed that the prevalence of diverse and heterogeneous CVMs was 2-fold more in women with ASCUS (19%) than in women with normal cervical cytology (10%). The prevalence of *L. iners*-dominated CVMs (CST III) was 2.7-fold less in women with normal cervical cytology compared to women with ASCUS (15%) [140]. Despite the small numbers of women with ASCUS (n=7, 15.2%) in the study by Mitra and co-workers (2015) [140], their results suggest that there may be a relationship between increased CVM diversity and abnormal cervical cytology, with the severity being associated with *Lactobacillus*-dominated CVMs, particularly *L. crispatus*-dominated CVMs (CST I) [140]. In our analyses, we had excluded other *Lactobacillus*-dominated CVMs (with *L. crispatus* and *L. jensenii* dominance) due to low numbers and only used the *L. iners*-dominated CVMs as the reference group. Thus, our analyses are not comparable to those of Mitra and co-workers (2015) [140].

The other participant information such as hormonal contraception and sexual behaviour did not vary between the two CSTs, findings that were both consistent and inconsistent with previous studies [33,34,136,138,139,179,215]. For instance, we did not find any association between the CVMs (CST-3 and CST-8) and cigarette smoking. This was in contrast to a pilot study that found smokers were less likely to have *Lactobacillus*-dominated CVMs [329]. Additionally, there was no difference between the CSTs with history of vaginal discharge in the last six months, unlike a previous study [138] that found that women with diverse CVMs were more likely to report vaginal discharge than women with *L. iners*-dominated CVMs. The absence of statistical difference between CST-3 and CST-8 in terms of abnormal cervical cytology (LSIL and HSIL) is in agreement with Gautam and colleagues [138]. Similar to CVM studies [179,215], including those on African or African American cohort [33,136,138,139], we did not find any difference in prevalent HIV, HPV, high-risk HPV, and bacterial STIs between CST-3 and CST-8. There was also no association between alpha or beta diversity of the CVMs with HPV infection. Women with HPV infection have been reported to have increased bacterial diversity (manifested by the presence of diverse and heterogeneous CVMs) than women without HPV infection [141,179,215]. However, other researchers [34] have found no difference in prevalent HPV and high-risk HPV (including its clearance) between *L. iners*-dominated CVMs and diverse CVMs. The lack of association and/or conflicting results could be due to differences in study population including small number of women with certain characteristics (e.g., SIL), study design, and study methodology.

4.4.5 Co-occurrence and co-exclusion patterns of bacterial OTUs

Spearman's correlations between the 40 OTU pairs with the greatest variance revealed both co-occurrence and co-exclusion relationships existed in the CVMs, similar to the findings in Chapter 2, subsection 2.3.7. Such relationships are crucial for assembly of bacterial communities including diverse ones such as those found in BV [51,290]. Phylogenetically and functionally related taxa tend to co-occur [330]. Both inter- and intra-taxonomy relationships are known to occur in microbial communities [290]. In this study, *Lactobacillus* and *Streptococcus*, all in order *Lactobacillales*, coexisted together Cluster-A1; hence, an indication of high degree of metabolic resource overlap. The lactobacilli co-occurrence patterns observed in this study were similar to that observed by Ravel and co-workers [38], Anahtar and co-workers (2015) [136], and McClelland and co-workers [36], with *L. iners* having the least positive correlation and *L. jensenii* the greatest positive correlation with other *Lactobacillus* members. Compared to *L. jensenii*, *L. iners* had lower negative correlations with BV-associated bacteria that were mostly in Cluster-B. These findings have been observed elsewhere [36,51] and underscore that *L. iners* is less resilient to perturbations and more tolerant of dysbiotic states than other common *Lactobacillus* spp. [110,162].

Positive correlations, such as those seen between *Lactobacillus* spp., may reveal analogous ecological traits such as metabolic interdependencies [51,290]. The widespread overlap of these positive relationships, especially in Cluster-B, provides evidence for functional redundancy in CVMs [51]. As a result, such communities are less affected by major perturbations since loss of one function is compensated by others. Metabolite exchanges due to cross-feeding plasticity can support community survival amid scarcity of resources [290]. Functional redundancy of BV-associated bacteria [51] may also be credited for the recurrence and persistence of BV in some women.

Negative correlations reflect environmental filtering, similar resource competition, and metabolic shifts [38,51,290,330]. Bacterial OTUs in Cluster-A, e.g., *Lactobacillus* spp., had a strong negative correlation with OTUs in Cluster-B that mostly had BV-associated bacteria such as *Gardnerella*, *Sneathia* and *Shuttleworthia*. A previous study on the temporal variability of CVM and relationship with BV [173] similarly observed an inverse relationship between higher quantities of *Lactobacillus*, *L. jensenii* and *L. crispatus*, and lower quantities of BV-associated bacteria, mostly *G. vaginalis*. Other investigators [36,136] have confirmed these observations. Lactobacilli are protective [33,34,77,146,160,187,224] and use a variety of proposed mechanisms to prevent colonisation by BV-associated bacteria as well as by opportunistic and pathogenic bacteria [38,54,77,158].

Bacterial interactions can be used to explain the onset of pathogenesis or recovery from a disease or disorder [330]. Transition of bacterial communities from one community state type to another requires that bacteria co-occur with one another, with minimal or without co-exclusion properties. In our study, although clusters A and B exhibited co-exclusion patterns, sub-clusters A2 and B1 appeared to harbour the transitional bacteria between clusters A and B. There were minimal overlap of relationships, that is, low negative to low positive correlations between bacterial OTUs in sub-cluster A2 and sub-cluster B1. Bacterial taxa in sub-clusters A2 and B1, such as *Finegoldia*, *Prevotella*, *Clostridium*, *Dialister*, and *Mycoplasma* among others, seemed to coexist. This insight suggests that it is possible for communities with bacteria in sub-cluster A2 to transition to communities with bacteria in sub-cluster B1, and vice versa.

4.4.6 Associations of the bacterial taxa with HPV

Given the mounting importance of cervicovaginal bacterial taxa in HPV infection [137,139,160,179,215] and the high burden of HPV in South Africa [16,20,270,331], the present study sought to investigate the association of the relative abundances of bacterial taxa with prevalent HPV among reproductive-age HIV-seronegative Black South African women. This is the first study to assess the bacterial biomarkers for genital HPV and high-risk HPV in South Africa. In this study, higher relative abundances of *Sneathia*, *G. vaginalis*, *Atopobium vaginae*, *P. veronii*, and *Aerococcus* were strongly correlated with high-risk HPV infections. *Sneathia*, *G. vaginalis* and *Aerococcus* were also associated with any HPV infection.

Of all the published putative bacteriological markers for HPV/high-risk HPV infections, *Sneathia* is the most widely detected potential biomarker and has been reported in Nigerian [137], Italian [157] and Korean [179] cohort studies. *Sneathia* has also been associated with SIL in HPV-positive women [140,142] and genital inflammation [87,136].

G. vaginalis and *A. vaginae* have been observed to populate the vaginal milieu following depletion of protective *Lactobacillus* in women with persistent HPV infection [157]. It has been reported that women with persistent HPV infection have higher relative abundances of *Atopobium* than women who clear HPV infections [157]. *Atopobium* and *Gardnerella* have been associated with BV-linked clue cells [51] and increased risk of CIN [160]. *Gardnerella*, the likely key driver of biofilm formation in the vagina [114,115,117], has been associated with preterm delivery [188] plus heightened immune activation, disruption of epithelial integrity, and impaired wound-healing [122]. It would be of interest to investigate if this activity of *G. vaginalis* is related to an increased risk for HPV infection. The presence of microwounds in the female genital tract allows HPV to access the basal epithelial cells of the cervical transformation zone [4]. The sialidase-encoding gene from *G. vaginalis* has been identified as a potential biomarker for HPV persistence [157]. *A. vaginae* is thought to augment the development of biofilm development since it is predominant in *G. vaginalis*-dominated biofilm [116]. Moreover, high bacterial load of *A. vaginae* has been positively associated with the occurrence of biofilm [116]. *Sneathia*, *Atopobium*, and *Gardnerella*, potential biomarkers for high-risk HPV in the current study were found to co-occur in the correlation analysis. Ravel and co-workers (2011) [38] also report co-occurrence of these taxa, but the exact significance of this relationship remains to be unmasked.

The observation that genus *Pseudomonas* is a potential biomarker for prevalent high-risk HPV infection in our cohort is in agreement with an Italian [157] cohort study. The identity of the *Pseudomonas* spp. in the Italian cohort study [157] was not identified. The species *P.*

veronii was identified as a potential biomarker of high-risk HPV in our study. Species in the genus *Pseudomonas* are increasingly being identified as possible biomarkers for HPV/high-risk HPV infection. A small pilot study observed higher abundances of *Pseudomonas oleovorans* in HPV-positive women with normal cervical cytology [142]. Surprisingly, we found that *Pseudomonas* was not a possible biomarker for prevalent HPV infection. Other possible biomarkers for HPV/high-risk HPV such as *Aerococcus* and *Gardnerella* that we detected have not been reported in cross-sectional studies elsewhere.

There were a number of possible biomarkers for HPV/high-risk HPV, e.g., *Pseudoxanthomonas*, *Megasphaera*, *Anaerococcus*, and *Dialister,* that have been published [137,157,179] but were not identified as potential biomarkers in our study. A more recent study on African/Caribbean women by Shannon and co-workers (2017) [141] also evaluated the relative abundances of specific bacteria in the CVMs by HPV status. In contrast to our findings, they detected neither *Gardnerella, Sneathia* nor *Atopobium* at greater relative abundances in the CVMs of women with HPV infections. These differences could be due to differences in the study design and population.

In high-risk HPV-negative women, greater relative abundances of *Haemophilus* and *Phyllobacterium* were found. Of note, *Lactobacillus* and *L. iners* were not found to be differentially abundant. Di Paola and colleagues (2017) [157], similarly found that *Lactobacillus* was not differentially abundant in the CVMs of Italian women with and without HPV/high-risk HPV infection. Several studies, however, have reported a greater relative abundance of lactobacilli, including *L. iners*, in the CVMs of women without HPV/high-risk HPV [137,141,179]. They observed greater relative abundances of *L. gasseri*, *F. nucleatum*, and *C. accolens* in women without HPV infections [141]. Similarly, the Korean [179] and Nigerian [137] studies found that women without HPV/high-risk HPV infections had greater relative abundances of *Lactobacillus* spp., which included *L. iners* and an unclassified *Lactobacillus* species. In our study, the relative abundance of lactobacilli was found not to be significantly different in women with and without HPV infection. This may be due to the low abundances of *Lactobacillus* spp., such as *L. gasseri*, in the study population.

Longitudinal data on the HPV infection outcomes at 6-month was available for 55 (63.2%) of the women included in this study. These were categorised as negative, cleared, persistent, acquired, or complex HPV infections (Supplemental Fig. 4.1). However, due to the complexity of HPV infections and the small sample sizes in each of the multiple categories of HPV infection, an analysis of the baseline bacterial taxa and longitudinal HPV infection outcomes were not possible.

4.4.7 Predicted functional cervicovaginal metagenome functions and their implications

Metagenomic functional profiles of the CVMs were predicted from the V3-V4 16S rRNA data using PICRUSt [317]. The relative abundances of the predicted metagenome functions of the cervicovaginal bacterial communities in CST-3 and CST-8 (Fig. 4.11 and Fig. 4.12), BV-negative and BV-positive women (Supplemental Fig. 4.7), as well as high-risk HPV-negative and high-risk HPV-positive women (Fig. 4.13) were then compared. Several differences in metagenome function or KEGG pathways were found in all three comparisons. Only the most significant differences that were found in common between these groups are discussed here. It is also important to note that the metagenome function abundances are extrapolated from the abundances in the 16S rRNA data. The functions found to be significantly enriched in each group, therefore, provides insight into the genomic capacities of the predominant bacterial taxa within that group.

Several KEGG pathways were found to be enriched in CVMs belonging to both the CST-8 and BV-positive groups; including transporters, fructose and mannose metabolism, sporulation, and transcription factors. Among these transporters were the most significantly enriched functional module, with ABC transporters, in particular, enriched in CST-8. ABC transporters are membrane proteins that facilitate the transport of a variety of substrates across the cell membrane, ranging from sugars, amino acids, xenobiotics and vitamins to polysaccharides, peptides and proteins. The ability to facilitate the uptake and export of a wide range of substrates enables their involvement in several cellular processes such as nutrient uptake, secretion of cellular waste, osmotic stress, lipid transport and macromolecular transport during biogenesis [332]. ABC transporters are also known to contribute to antimicrobial drug resistance [333]. Genes encoding proteins involved in transport, including ABC-type transporters, are abundant in the genomes of many *G. vaginalis* and *Sneathia* strains [334,335], predominant bacteria in CST-8 and BV-positive CVMs and the driver of community diversity and biofilm formation [114-116]. Interestingly, several recent studies have indicated a potential importance for ABC transporters in strains of *G. vaginalis* associated with BV and in *G. vaginalis* biofilms. Strains of *G. vaginalis* associated with symptomatic BV were found in two studies to uniquely encode genes for ABC transporters that are absent in the genomes of *G. vaginalis* strains not associated with BV [336,337]. It has furthermore been found that transcripts encoding ABC transporters are upregulated in *G. vaginalis* biofilm cells compared to planktonic cells [118]. Cells in biofilms are likely to encounter restricted availability of nutrients and the elevated levels of ABC transporter proteins may facilitate greater nutrient uptake and survival under these conditions. It is also possible that ABC transporters are involved in the shuttling of non-glycogen polysaccharides for metabolism and export of cell-surface glycoconjugates for biofilm and O-antigen formation [338,339].

Fructose and mannose metabolism was enriched in CST-8, BV-positive and high-risk HPV-positive CVMs groups, with starch and sucrose metabolism additionally enriched in the CVMs of BV-positive women. Fructose and starch are less abundant carbohydrate sources in the vagina [335]. *G. vaginalis*, a predominant member of CVMs in CST-8, BV-positive and HPV-positive women, has the ability to metabolise both fructose and starch [335]. *L. iners*, the predominant member of CST-3 CVMs, in contrast, does not have the genetic capacity to ferment fructose [335,340]. In a comparison of metabolites in cervicovaginal lavage fluid from 40 women with BV and 20 women without BV, fructose was found to be significantly lower in BV-positive samples [121]. This may be due to its metabolism by *G. vaginalis*.

These findings are supported by a recent metaproteomic comparison of *G. vaginalis*-dominated and *L. iners*-dominated CVMs carried out Zevin and co-workers 2016 [122]. Proteins involved in transport, catabolism, and membrane transport functions were more abundant in communities dominated with *G. vaginalis* than *L. iners*. Examples of differentially abundant proteins included MalE-type ABC sugar transport system periplasmic component, and α-1,4-glucan phosphorylase, an enzyme that degrades starch and glycogen, and fructose-1,6-bisphosphate aldolase. The authors [122] argued that these significant enrichments possibly enable *G. vaginalis* to outmatch lactobacilli competency for the uptake of extracellular saccharide.

Using PICRUSt analysis, our study showed that metagenome functions of CVMs of women with CST-8, BV, and high-risk HPV distinctively differed from those without such dysbiosis and disease. Several of these metagenomics functions were found to be enriched or in low relative abundance in all three groups. Metagenome modules such as transcription factors, sporulation and fructose and mannose metabolism were enriched, while LPS biosynthesis, membrane and intracellular structural molecules, energy metabolism, and citrate cycle were not abundant in all these three groups. These results show a functional interconnection between the diverse community state types (CST-8) and BV and high-risk HPV. These predicted metagenome functions have the capability to alter the host's biochemical and immune responses [122,136,138,141,142,328,341] and have a profound effect on women's reproductive health [122].

4.4.8 Summary

In this cross-sectional study, CVMs without *Lactobacillus* dominance (CSTs 4-8) were found to be common (74% prevalence) among reproductive-age HIV-seronegative Black South African women. CST-8, represented by diverse and heterogeneous bacterial communities with bacterial genera such as *Sneathia*, *Gardnerella*, *Prevotella*, *Megasphaera*, and *Shuttleworthia*, was the most prevalent CST in the cohort (64%). Only a few women (26% (23/87)) had *Lactobacillus*-dominated CVMs: CST-1 (*L. crispatus*), CST-2 (*L. jensenii*), and CST-3 (*L. iners*). Among these, the *L. iners*-dominated CVMs (CST-3) were the most predominant (83% (19/23)). *L. gasseri*-dominated CVMs were not present and remain unreported among Black South African women [87,136,138]. *L. iners* and *G. vaginalis* were ubiquitously present in all the CVMs.

CST-8 was found to have different metagenome functions compared to *L. iners*-dominated CVMs (CST-3). Transporters (including ABC transporters), flagella-mediated chemotaxis-directed motility, and bacterial motility proteins enriched in the diverse CVMs (CST-8) may be used by these BV-associated communities to enhance their survival and/or cause dysbiosis. Published literatures underscore that chemotaxis is important for directing the flagellar motility to the site of pathogenesis [342] while flagellar-mediated motility is crucial for expediting bacterial infection, invasion through self-induced phagocytosis, bacterial penetration between cell-cell junctions, post-infection dispersal [342], and differentiation of bacterial planktonic form into mature and robust biofilm architecture [342-344]. ABC transporter genes are often highly expressed in *G. vaginalis* biofilm cells [118]. Our findings are based on predictive analysis. Thus, the biochemical and immunological effects of non-*Lactobacillus*-dominated CVMs still requires a detailed study. To examine these, the relationship between such CVMs with human genital tract metabolites, proteomes, and cytokines should be examined using immunological and omics studies.

The CVMs of women with high-risk HPV infection were found to have higher relative abundances of *G. vaginalis, Sneathia, A. vaginae, Aerococcus* and *P. veronii*. Several of these have previously been reported as possible biomarkers for HPV or high-risk HPV infection [137,157,179]. Investigations into the role these potential bacterial biomarkers play in HPV/high-risk infection are warranted. The roles of these potential biomarkers can be evaluated in longitudinal studies with a larger cohort and/or in murine model of genital HPV/high-risk HPV infection.

Higher relative abundances of *Sneathia*, *Atopobium* and *Gardnerella* are more common in African women than White women [39,40,51,345] and have been associated with BV [326] and found in cases of recurrent BV [173]. In our Black South African cohort, women with BV were found to have higher relative abundances of these bacteria. In summary, the study suggests that there is a link between diverse CVMs, BV, and HPV, but longitudinal studies will be needed to unravel the processes involved.

The main limitations of this study include the relatively small sample size, use of phenotypic trait (Black skin colour) to define women of African ancestry, AV may have been overlooked and that the metagenomic method used may underestimate AV-associated microorganisms, lack of a standard method to diagnose BV, and inadequate information on female sexual behaviour. As a result, it was not possible to examine whether some of the observed associations were confounded by the missing information. Our results on metagenome functions as predicted using PICRUSt [317] were not devoid of other conditions (genital tract infections and syndromes) that might have obscured the exact functions. For example, when investigating the metagenome functions of CVMs of women with and without HPV, we did not exclude women with BV despite the fact that this could have influenced the results. Therefore, the aforementioned limitations should be addressed in future studies.

CHAPTER 5

Characterisation of Penile Microbiotas of Black South African Men with and without Human Papillomavirus and HIV Infections using Illumina Sequencing of the V3-V4 Region of the 16S rRNA Gene

5.1 Introduction

In this chapter, the penile microbiotas of 282 Black South African men with and without HPV and HIV infections were profiled. Penile bacteria associated with BV in women were also evaluated.

The best evidence to date for a potential role of the penile microbiome in STIs and HIV acquisition have come from studies examining circumcision [35,259]. Male circumcision is protective against STIs [234,235], such as HPV [240,241,262] and HIV [242] in men. Male circumcision has been found to alter the penile microbiome by significantly reducing its bacterial diversity and load [229,254]. It eliminates the anoxic microenvironment, thereby leading to loss of oxygen-intolerant bacteria (anaerobes) and gain in bacteria that can thrive in aerated penile microenvironment (facultative anaerobes and aerobes) [229,254]. Specific penile anaerobic bacteria have been identified as independent risk factors for HIV infection in men [35]. Thus, knowing the oxygen requirements of the penile bacteria is vital. Pathogenic bacteria and dysbiosis in the penile microbiota, characterised by the presence of BV-associated anaerobic bacteria such as *Prevotella,* have been identified as key risk factors for HIV acquisition in uncircumcised men [35,232]. Several recent studies [346-353] have showed that specific genital bacteria, which are more prevalent or abundant in uncircumcised men, could stimulate local immune responses that enhance epithelial inflammation and HIV target cell recruitment. This suggests that HIV acquisition could be linked to proinflammatory anaerobic bacteria in the penile bacterial microbiota. No research has been published on penile microbiotas and their relationship with HPV and HIV in South Africans.

Not only is this study the first to characterise the penile microbiota of Black South African men, but also the first study to ever examine the penile microbiotas using a cohort of predominantly traditionally circumcised men. In many southern African cultures, male circumcision is often performed by a medically untrained, traditional provider in a non-clinical setting; with the foreskin either partially or fully removed [244-246,249,251]. This type of circumcision may not be as protective as medical circumcision [234,242,260,262] as suggested by its inconsistent protective effect against HIV [245,247,248,250,251]. Men with uncircumcised and partially circumcised penises have been observed to have greater risk of acquiring STIs compared to fully circumcised men [251].

The beneficial effect of circumcision has also been observed in the female sexual partners of circumcised men [236,260], where it has been associated with reduced risk of vaginal infections, genital symptoms like BV [260], and cervical cancer in women whose male partner has a history multiple female sexual partners [236]. Consistent with the findings in chapters 2 and 4, it is increasingly becoming apparent that Black South African women have diverse CVMs that are strongly associated with BV [136,138]. Observational studies [105,107,181] have suggested that there is a relationship between unprotected sexual intercourse and imbalanced CVMs. The human penis, especially of an uncircumcised male, is a potential vector for sexual transmission of BV [194]. Thus, unprotected sexual intercourse may considerably alter CVMs [105,106] by transferring BV-associated bacteria from the penis to the female genital tract [106,193,229,257]. A small study assessing the effect of sexual activity on the CVMs of young women noted that women who engaged in unprotected penile-vaginal sex were more likely to have CVMs dominated with *L. iners* or *G. vaginalis* compared to women who reported no sex [107]. Moreover, the study associated sexual activity with multiple clades of *G. vaginalis* (GV1, GV2, G3, and GV4) [107]. Sexually transmitted penile bacteria may cause the resident bacteria in the female genital tract to shift to those associated with BV [105]. Sexually-linked GV4, for example, has been positively associated with incident and prevalent BV [107]. Because of the highly prevalent diverse and dysbiotic CVMs among Black South African women, the present study also assessed the prevalence and relative abundance of BV-associated bacteria in penile microbiotas.

5.1.1 Aim and objectives

The aim of this study was to characterise the penile microbiotas of Black South African men aged 20-67 years using V3-V4 16S rRNA Illumina MiSeq libraries and assess their relationship with HPV and HIV infection. Specific objectives included:

i. To compare the penile microbiotas of Black South African men, most of whom were traditionally circumcised, to those reported in published studies,
ii. To assess the association, if any, between the penile microbiotas and prevalent HPV infection,
iii. To assess the impact of HIV status on the penile microbiotas,
iv. To assess the predominance of BV-associated bacteria in penile microbiotas, and
v. To detect potential biomarkers in the penile microbiotas for HPV, high-risk HPV, HIV, and BV (in female sexual partners).

5.2 Materials and Methods

5.2.1 Ethics statement

This information has been provided in the Materials and Methods section (subsection 2.2.1) in Chapter 2.

5.2.2 Study samples, study design and study population characteristics

The parent study, HPV Couples Cohort Study, has been described in detail elsewhere [270]. The penile specimens from heterosexually-active Xhosa-speaking adult men aged 19-

67 years were previously collected by swabbing of the penile shaft, foreskin (if uncircumcised), and glans using dry Digene swabs. These were placed in Digene Specimen Transport Medium (STM, Qiagen, Gaithersburg, Inc., USA) and frozen at -80°C until DNA extraction. The present study relied on a total of 282 stored baseline specimens and their abstracted information from the parent study. These were randomly selected samples with baseline information on HPV and HIV and sufficient volume (≥15.0 µl) for microbiota investigation. Nucleic acid isolation and HPV genotyping were performed as previously described (subsection 2.2.2).

Two hundred and thirty eight men (84.4%) were finally included in the study. Participants excluded (15.6%, 44/282) were those whose samples had <13,014 reads for 16S rRNA metagenomics analyses. The baseline characteristics of the 238 men are summarised in Table 5.1.

Table 5.1. Baseline demographic, sexual and smoking, behavioural, and clinical information of the 238 heterosexually-active Black South African men.

Characteristic	**Participants**
	(N = 238)
Age (years)	36.0 (30.0-44.0)
Age at sexual debut (years)#^	17.0 (15.0-18.0)
Number of lifetime sexual partners	6.0 (3.0-13.0)
Number of sexual acts with study partner in last month#^	4.0 (2.0-10.0)
Current use of condom (% (n/N))	
No	35.9 (79/220)
Yes	64.1 (141/220)
HPV (% (n/N))	
Negative	45.4 (108/238)
Positive	54.6 (130/238)
High-risk HPV (% (n/N))	
Negative	57.1 (136/238)
Positive	42.9 (102/238)
HIV (% (n/N))	
Negative	63.0 (150/238)
Positive	37.0 (88/238)
HIV-positive men on antiretroviral therapy	13.3 (10/75)
CD4$^+$ T-cell count if HIV-positive (cells/µl)#^	334 (232-478)
HIV viral load if high-risk HIV-positive (copies/ml)#^	4.1 $\log_{10}$ (3.7-4.7)
Circumcised (% (n/N))	
No	5.7 (13/228)
Yes	94.3 (215/228)
Sexual partner's incidental BV on Papanicolaou smear (% (n/N))	
Negative	66.5 (157/236)
Positive	33.5 (79/236)
Cigarette use (% (n/N))	
Never smoked	16.9 (40/237)
Ex-smoker	16.9 (40/237)
Current smoker	66.2 (157/237)

Abbreviations: HPV – human papillomavirus, HIV – human immunodeficiency virus, BV – bacterial vaginosis.

#Continuous variables are expressed as medians with interquartile ranges (IQRs, at 25th and 75th percentiles).

^Data was not available on the age at sexual debut for three men, lifetime number of sexual partners of four men, number of sexual acts with study partner in the last month of six men, CD4$^+$ T-cell count of one man and viral load measurement for thirty one men.

The median age of the men was 36.0 years. All the men were sexually active, with 54.6% and 37.0% of them being positive for HPV and HIV infections, respectively. Of the HPV-positive men, 28.5% (37/130) and 71.5% (93/130) were infected with single and multiple HPV types, respectively. About 42.9% of the men were positive for high-risk HPV infections. All the circumcised men in our study (94.3%) were traditionally circumcised. Information on incidental BV of the female sexual partners was available for 99.2% (236/238) of the men, of which 33.5% were in relationships where the female partner was BV-positive.

5.2.3 Illumina MiSeq V3-V4 16S rRNA metagenomics library preparation and sequencing

The bacterial DNA was amplified by targeting the hypervariable V3-V4 region of the 16S rRNA gene using the protocol described in subsection 3.2.2. The 282 penile samples were concurrently run with internal controls: negative controls (two nuclease free water and Digene STM extraction controls), and positive controls (mock communities: two HM-782D (even, low concentration) and one HM-783D (staggered, low concentration) (BEI Resources, Manassas, VA, USA)), whose bacterial compositions were previously detailed in Chapter 4 (Supplemental Table 4.1). DNA amplicons were electrophoresed as detailed in subsection 3.2.2 and the gel bands were visualized using the BioRad Molecular Imager® Gel Doc™ XR+ Imaging System. Photos were taken using the Image Lab™ Software v5.21 (Bio Laboratories Inc., Hercules, CA, USA). The amplicons were purified and quantified as previously described (subsection 2.2.3). However, quantification was done on the SpectraMax i3x MultiMode Plate Reader and SoftMax Pro Software v6.5.1 (Molecular Devices, Sunnyvale, CA, USA). The purified V3-V4 16S rRNA libraries from the 282 samples and six internal controls were sequenced in three runs on the Illumina MiSeq using a paired-end 300-bp protocol and v3 reagents by the Center for Genomic Regulation (CRG, Barcelona, Spain).

5.2.4 V3-V4 16S rRNA gene dataset analyses

The qualities of the raw sequenced reads were assessed by FastQC v0.11.2 [272]. The sequenced reads were assessed by mothur v1.37.6 [354] using its standard operating procedure (SOP) guidelines [355], with slight modifications. The mothur pipeline had been validated as previously described [297]. The forward and reverse reads were assembled into contiguous reads. The sequences were screened and those with <439-bp minimum length, >466-bp maximum length, and >4 ambiguous base calls removed. To reduce the computational time, unique sequences in the fasta-formatted file were selected and then aligned against a reference alignment, SILVA v119 (www.arb-silva.de/), using the Needleman algorithm with kmer searching of 8 ksize (nucleotide substring), -2 gap opening and -1 gap extension penalties. The kmer searching was used because it is faster and more reliable than blast and suffix tree template searching methods. Sequences with homopolymeric run longer than 12 bases were removed. The alignments were then filtered to remove gaps as they do not have any genetic information. The sequences were then pre-clustered by abundance using a pseudo-single linkage algorithm to remove erroneous sequences with >4 nucleotide mismatches. For additional filtering, singletons were also discarded. Potential chimeras were removed by a *de novo* method using the UCHIME algorithm [275]. Non-chimeric sequences were assigned taxonomy using the Wang approach of 8 kmers implemented by RDP Classifier [276]. The “trainset9_032012.pds.fasta” and “trainset9_032012.pds.tax” were used as the RDP database sequence and taxonomy files, respectively. The cut-off for bootstrap confidence score for taxonomic assignment was 80%. Lineages of chloroplastic, mitochondrial, archaeal, eukaryotic, and other non-bacterial sequences were then removed. This was followed by assessment of the sequencing errors using the mock communities in order to determine the reliability of the sequencing procedure [355]. First, all bacterial sequences were extracted from our mock communities using the *get.groups* command. These sequences were then searched for errors against the reference sequences of mock communities (“HMP_MOCK.v35.fasta”) using the *seq.error* command. Error rate was then calculated using the formula below (Equation 5.2) and expressed as a percentage:

$$Error\ rate = \frac{Sum\ of\ mismatches\ to\ reference}{Sum\ of\ bases\ in\ query}$$

OTUs were defined at a 0.03 cut-off phylogenetic distance using the opticlust clustering algorithm. The quality (completeness) quality of sampling the bacterial communities was measured using Good’s coverage estimate [356]. The coverage calculator in mothur showed that at the recommended 92% Good’s coverage (Luca Cozzuto, Personal Communication), the majority of the OTUs (potential species), including the low-abundant ones, could be sampled. This coverage was equivalent to a subsampling depth of 13,014 reads (per sample) that showed that it could reliably capture the microbiota diversity. The OTU table was therefore normalised by rarefying at 13,014 reads counts per sample.

Alpha diversity analyses of the various metadata categories (HPV, high-risk HPV, HIV, CD4, CST, and partner’s BV) were performed using a custom script in R v3.2.2 (R Core Team 2016). Beta diversity was computed using Vegan R package v2.4.3 [278] and phyloseq v1.20.0 [357]. The alpha diversity was computed using Simpson, Dominance, Shannon, and Shannon Equitability indices, while beta diversity was computed using the UniFrac distance matrix were represented using 2D PCoA plots.

5.2.5 Prevalence and relative abundance of penile bacteria and hierarchical clustering of penile microbiotas

The prevalence and relative abundances of the bacteria in the penile microbiotas calculated in mothur was summarised at each taxonomic level using customised Python scripts. This was done at the phylum, family, and genus taxonomic ranks.

To cluster the penile microbial communities into CSTs according to bacterial community composition and abundance, the Vegan R package [278] was used. Here, the average neighbour linkage method based on Bray-Curtis dissimilarity index was used. In the heatmaps, participants' information such as HIV, HPV, high-risk HPV, partners' BV, and CST information were included.

5.2.6 Putative aerotolerance profile of the most abundant penile microbiota families

The aerotolerance (oxygen requirements) of the 40 most abundant families in the penile microbiotas were accessed from published penile microbiome literature [229,254] and extensive literature searches on the PubMed database (https://www.ncbi.nlm.nih.gov/pubmed/). The overall oxygen requirement of each family was then assigned from the oxygen profiles of the genera in that family. For families with genera with different oxygen requirements were considered to have mixed aerotolerance profiles, e.g., *Mae/Fan*. For biochemically uncharacterised or taxonomically unclassified families the oxygen requirement was designated as "unidentified". Finally, to examine the most to least common aerotolerance profile of the 40 most abundant families, we grouped all families with similar oxygen requirements together. For example, *Flavobacteriaceae* and *Moraxellaceae* were grouped together as "aerobic" while *Veillonellaceae* and *Clostridiales Incertae Sedis XI* were grouped as "anaerobic". We then measured the overall prevalence of aerotolerance profiles of each of these grouped families.

5.2.7 Co-occurrence and co-exclusion patterns of bacterial families

Positive and negative correlations between the counts of bacterial families in the penile microbiotas were assessed using metagenomeSeq v1.12.1 [280,281]. The families assessed included the eleven most abundant families in the penile microbiotas (Supplemental Table 5.1) and two less abundant families (*Pseudomonadaceae* and *Oxalobacteraceae*) that have been previously found to have a positive correlation by Price and colleagues (2010) [229]. A correlogram depicting the correlations between these families was plotted using the *plotCorr* function in the metagenomeSeq v1.12.1 [280,281].

5.2.8 Identification of potential biomarkers for penile HPV and HIV infections and partner's BV

To detect differentially abundant taxa that may act as potential biomarkers for HPV and HIV infections in men, and BV in female sexual partners, LefSe v1.0 [319] was used. The level of statistical significance (p-value) was set at 0.05 while the threshold for discriminative features based on the logarithmic LDA score was set at 2.0.

5.2.9 Statistical analyses

GraphPad Prism Software v6.01 and R v3.2.2 (R Core Team 2016) was used to perform statistical comparisons. Comparison of the HPV and CST groups with participant categorical and continuous metadata was computed by Fisher's exact/Chi-square and Mann-Whitney unpaired nonparametric tests, respectively. A two-tailed p-value of <0.05 was used as the level of significance. Relative risk (RR) with corresponding 95% confidence intervals (CI) were used to measure the magnitude of associations.

The alpha diversity indices of the various metadata categories (CST, HPV, high-risk HPV, HIV, CD4, and BV) were compared using Mann-Whitney unpaired and Kruskal-Wallis nonparametric tests. For beta diversity comparisons, the Adonis nonparametric test with 999 permutations was used. The distance matrix was used to calculate the effect size (R^2 value) that showed the extent of variation explained by the metadata category. A p-value of <0.05 was used to measure the statistical significance.

5.3 Results

5.3.1 Validation and optimisation of microbial analyses using PCR and sequencing controls

From all the 288 sequenced samples (including the internal controls and penile microbiome samples), a total of 9,406,780 high-quality sequencing reads were obtained. Of these, 139 and 1,009 reads were obtained for the two nuclease free water controls and 360 from the extraction control (Digene STM). The staggered mock community yielded 53,638 reads and the two even mock communities 51,821 and 104,070 reads. The average of number of reads from the penile samples (n=282) was 32,609 reads per sample (range: 14-123,966 reads). The total number of bacterial reads in the two nuclease free water differed from the average reads per penile sample by 96.9% and 99.6% while that of the extraction control was 98.9%. Thus, this showed that the level of contaminants in the sequenced libraries was very low. The predominant taxa in nuclease free water controls were *Pseudomonas* (62 of 139 reads (44.6%) and 505 of 1,009 reads (50.0%)) and *Delftia* (27/139

(19.4%) and 167/1,009 (16.6%)). In the extraction control, *Pseudomonas* (82/360 (22.8%)) and *Cloacibacterium* (81/360 (22.5%)) were the most abundant taxa.

For every 95 samples to be sequenced, the mothur SOP [355] recommends inclusion of a mock community for estimation of the sequencing errors. In the present study, the expected genera (n=17) in the three mock communities accounted for between 98.22-99.72% of the total reads in the corresponding mock community samples. Non-mock genera (n=34) represented only 0.79-1.78% of the reads in the two even mock communities and 0.28% of reads in the staggered mock community. The predominant non-mock genera in the each of the two even mock communities were an unclassified *Bacillus* (94/407 reads (23.1%) and 512/1,852 reads (27.6%)) and an unclassified *Lactobacillales* (82/47 (20.1%) and 314/1,852 (17.0%)). In the staggered mock community, an unclassified *Clostridiaceae* (20/150 (13.3%)) and a bacteria of an unclassified genus (17/150 (11.3%)) were the most abundant genera.

The performance of the negative and positive controls was comparable to those used in Chapter 4. Using the mock communities, we found that the sequencing error as revealed by mothur was relatively low (0.042%). This sequencing error and the low level of contamination were acceptable (Luca Cozzuto and Patrick D. Schloss, Personal Communications) and showed that the library preparation and sequencing steps for all the 288 samples were reliable. Furthermore, we performed rarefaction and rarefied our diversity analyses at 13,014 reads; thus, ensuring that sequencing errors did not affect the results (Patrick D. Schloss, Personal Communication).

5.3.2 Data analysis and output

The penile microbiotas from 238 samples were finally characterised using 8,805,830 quality-filtered non-chimeric sequences (~460-bp), with a median read count of 32,345 reads per sample (mean: 36,999 reads, range: 13,014-123,966 reads per sample). The remaining 44 samples with low read counts (<13,014 per sample) were excluded based on rarefaction plots (results not shown) and Good's coverage estimate as described in methods subsection 5.2.4. Reads were processed and analysed as described in subsection 5.2.4. All the read sequences were assigned taxonomy up to the genus level by the RDP Classifier implemented in mothur.

5.3.3 Distributions of various taxa in the penile microbiotas and their putative oxygen requirements

A total of 23 phyla at varying relative abundances were identified in the 238 penile microbiotas (Table 5.2). The most abundant phyla (>1% relative abundance) were Actinobacteria, Firmicutes, Bacteroidetes, Proteobacteria, and Fusobacteria.

Table 5.2. Relative abundances of the 23 phyla detected in penile microbiotas of 238 heterosexual Black South African men.

Phylum	Relative abundance (%)	Phylum	Relative abundance (%)
Actinobacteria	55.98	**Verrucomicrobia**	<0.01
Firmicutes	21.84	**SR**[#]	<0.01
Bacteriodetes	14.33	**Acidobacteria**	<0.01
Proteobacteria	5.46	**Planctomycetes**	<0.01
Fusobacteria	1.36	**Elusimicrobia**	<0.01
Unclassified Phylum	0.85	**Chlamydiae**	<0.01
Terenicutes	0.05	**Gemmatimonadetes**	<0.01
Synergistetes	0.04	**OD**[#]	<0.01
TM[#]	0.03	**Armatimonadetes**	<0.01
Chloroflexi	0.03	**Nitrospira**	<0.01
Deinococcus-Thermus	0.01	**Chlorobi**	<0.01
Spirochaetes	0.01		

[#] Bacterial candidate (putative) phyla. Organisms in these phyla have only been characterised by environmental sequenced-based studies [358].

Since published reports have shown that the penis can be colonised by different bacterial families [110,229,254], we further studied the bacterial families in our cohort and evaluated their oxygen requirements. Bacteria can be categorised according to their oxygen requirements for respiration and growth. Aerobic (*Ae*) bacteria require oxygen for respiration and growth, while anaerobic (*An*) bacteria do not [282]. Facultative anaerobic bacteria (*FAn*) can survive in the presence or absence of oxygen, although growth activity in oxygen-free environment is usually slower [282]. Microaerophilic (*MAe*) bacteria grow in the presence of oxygen but are sensitive to high oxygen concentrations [359]. Our study detected 210 bacterial families. The top 40 most abundant families, together with their respective oxygen requirements, are summarised in Supplemental Table 5.1. The most abundant families (at ≥3.5% relative abundance) and their oxygen requirements included *Corynebacteriaceae* (47.19%, *FAn*), *Prevotellaceae* (6.56%, *An*), unclassified *Clostridiales* (5.61%, unidentified), *Porphyromonadaceae* (4.94%, *An*), *Staphylococcaceae* (4.57%, *FAn*), *Bifidobacteriaceae* (3.88%, *An*/*FAn*), and *Lactobacillaceae* (3.81%, *MAe*/*FAn*). Of the top 40 bacterial families, *FAn* bacteria were the most dominant (53.39%) followed by *An* bacteria (18.16%) and *Ae* (9.86%). The relative abundances of families with oxygen requirement categorised as *An*/*FAn* and *MAe* were 5.66% and, 3.98%, respectively. Families with unidentified oxygen requirement had an overall relative abundance of 7.33%. The relative abundances of families with *MAe* (0.18%) and *MAe*/*Ae* (0.26%) were very low.

A total of 650 genera were detected. The distributions of these genera in the 23 phyla were different. Among the abundant phyla (>1% relative abundance), phylum Proteobacteria comprised the most genera (n=241, 37.1%), followed by Firmicutes (n=148, 22.8%), Actinobacteria (n=131, 20.2%), Bacteriodetes (n=73, 11.2%), and Fusobacteria (n=9, 1.4%). Fig. 5.1 shows a relative abundance heatmap of the most abundant genera (≥0.08%) grouped into their respective phyla. Only 50 of the 650 genera had ≥0.08% relative abundance and were distributed among 6 phyla (Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria, Proteobacteria, and bacteria with an unclassified phylum). The most abundant genera were *Corynebacterium* (47.12%), *Prevotella* (6.50%), unclassified *Clostridiales* (5.61%), *Porphyromonas* (4.85%), *Staphylococcus* (4.39%), *Lactobacillus* (3.81%), *Gardnerella* (3.78%), *Chryseobacterium* (2.44%), *Acinetobacter* (2.27%), and *Negativicoccus* (1.86%).

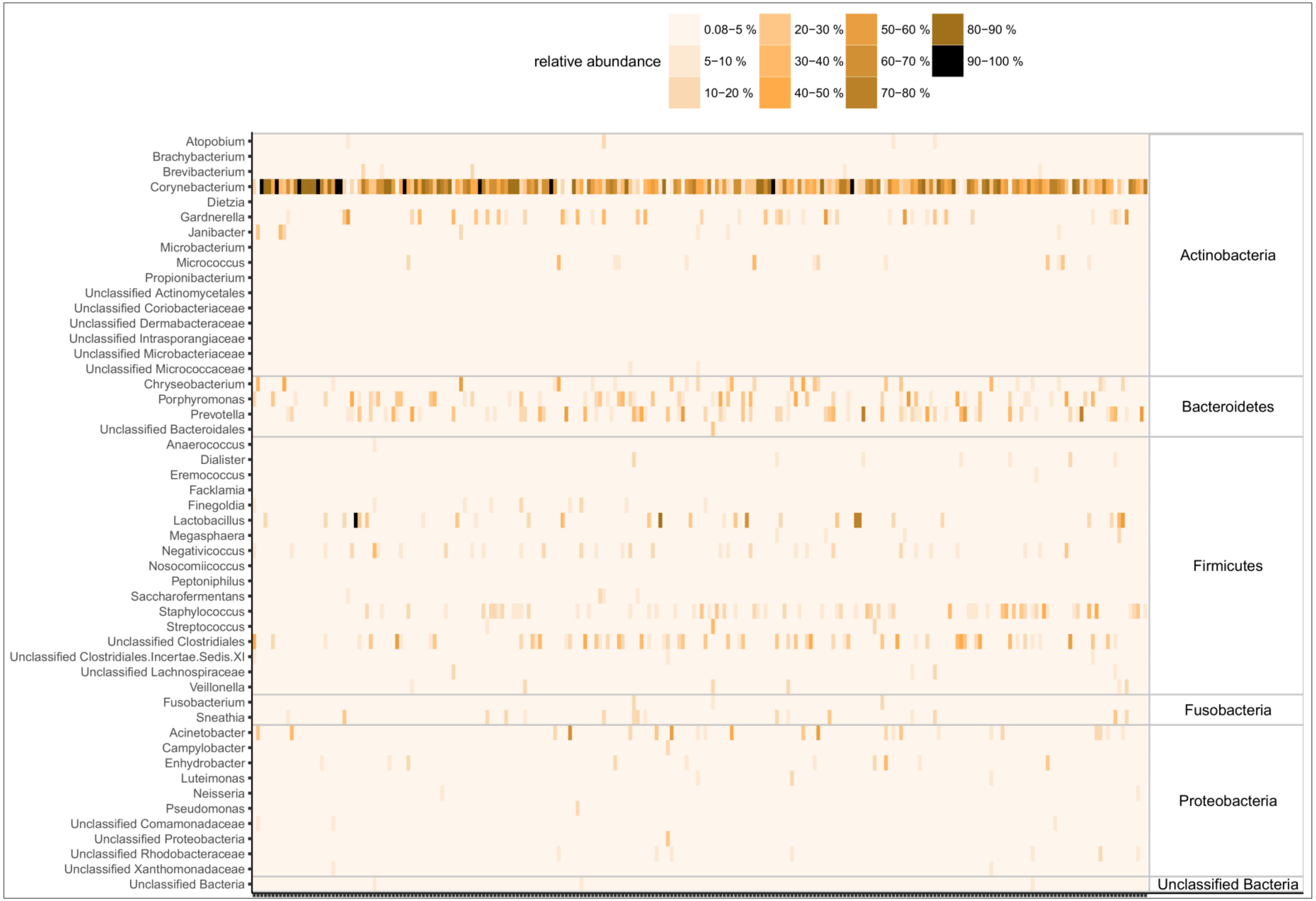


Fig. 5.1. Relative abundances of genera in the penile microbiotas of 238 Black South African men. Genera with a relative abundances ≥0.08% are shown. Columns represent individual penile samples and rows genera. The horizontal lines separate the different phyla, indicated on the right.

5.3.4 Co-occurrence and co-exclusion patterns of bacterial families

To assess the potential relationships between the bacterial families in our study, we performed correlations between the most abundant bacterial families detected in the penile microbiotas using Spearman's rank correlation. Two families, *Pseudomonadaceae* and *Oxalobacteraceae,* that were not abundant in our study, were also included in the analysis as they have been shown to be highly positively correlated with each other in a previous penile microbiome study [229].

The correlogram depicting the observed relationships between the bacterial families (positive and negative) is shown in Fig. 5.2. Two major clusters of bacterial families (designated as A and B) were evident in the correlogram, as indicated by the first bifurcation of the clustering dendrogram (Fig. 5.2).

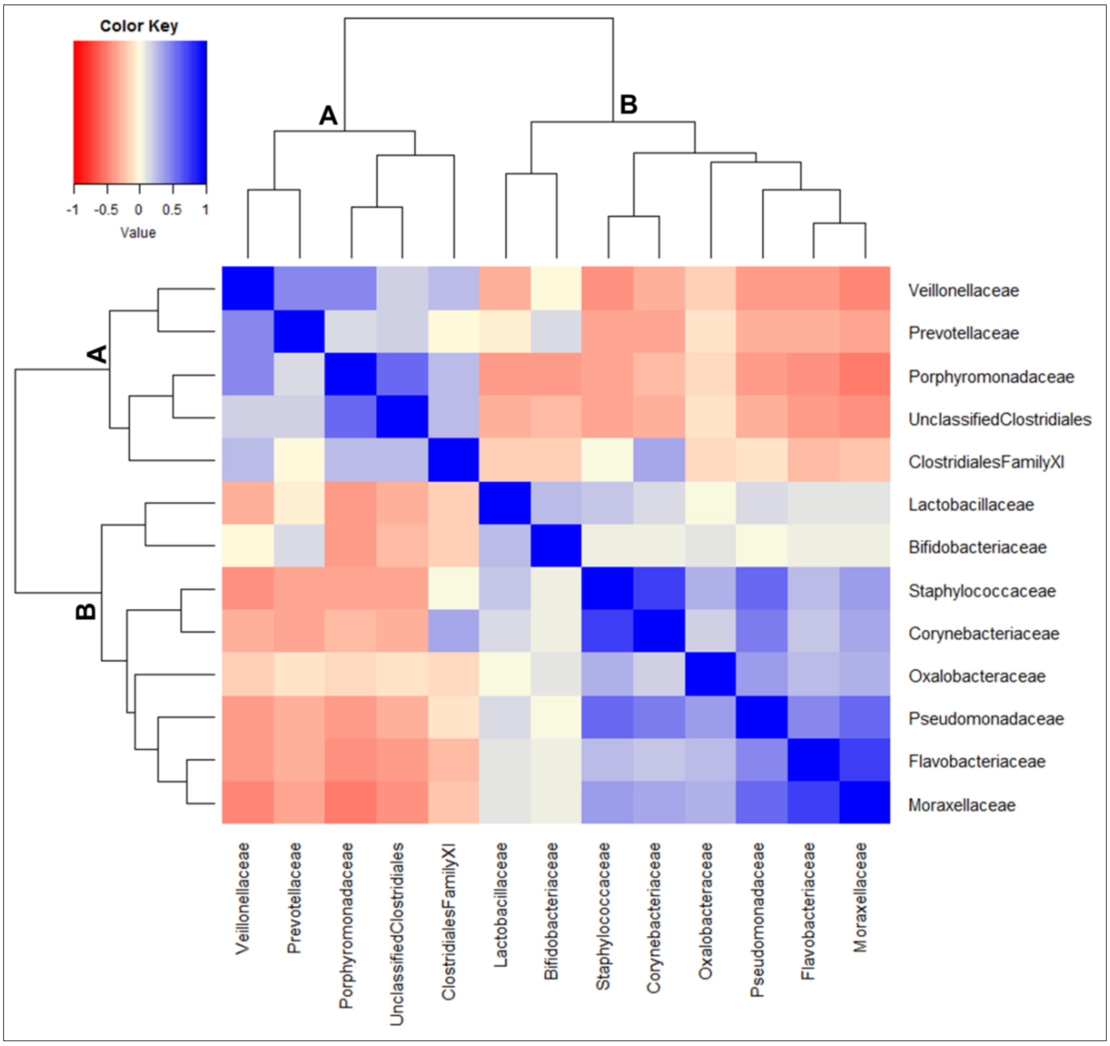


Fig. 5.2. Correlogram of 13 bacterial families showing co-occurrence and co-exclusion patterns. These were computed by Spearman's rank correlation between the families. The correlation coefficients range from -1 (red; co-exclusions relationships) to +1 (blue; co-occurrence relationships), hence, high negative and positive coefficient values indicate strong correlations. The blue diagonal line represents correlations of +1. White shows absence of bacterial relationships.

Cluster-A comprised of *Veillonellaceae*, *Prevotellaceae*, *Porphyromonadaceae*, unclassified *Clostridiales*, and *Clostridiales Incertae Sedis XI*. These bacterial families were positively correlated with one another, an indication of of niche sharing and metabolic resource overlap. Generally, these bacterial families were negatively correlated with the eight bacterial families in Cluster-B, which included the most abundant family *Corynebacteriaceae as well as Moraxellaceae*, *Flavobacteriaceae*, *Pseudomonadaceae*, *Oxalobacteraceae*, *Staphylococcaceae*, *Bifidobacteriaceae*, and *Lactobacillaceae*, although family *Clostridiales Incertae Sedis XI* (in Cluster-A) was positively correlated with family *Corynebacteriaceae* (in Cluster-B). Negative interactions between families in Cluster-A and Cluster-B suggest that these bacteria may be competing against one another and may therefore co-exclude one another in the penile niche. The bacterial families in Cluster-B were positively correlated with one another, with *Bifidobacteriaceae* and *Lactobacillaceae* exhibiting the weakest positive interactions (<0.4 correlation coefficients) with the other families in this cluster. These two bacterial families clustered together following the first bifurcation of Cluster-B. Although *Bifidobacteriaceae* and *Prevotellaceae* were in different clusters (Cluster-B and Cluster-A, respectively), they were positively correlated with one another.

5.3.5 Establishment of penile microbiota community state types

To define the CSTs, the penile samples were hierarchically clustered based on the microbiota composition. The clustering and heatmap of the relative abundance of the bacterial taxa across the 238 penile samples is shown in Fig. 5.3. Based on the Bray-Curtis dissimilarity index, the penile bacterial communities clustered into six CSTs. These CSTs were numbered from 1-6.

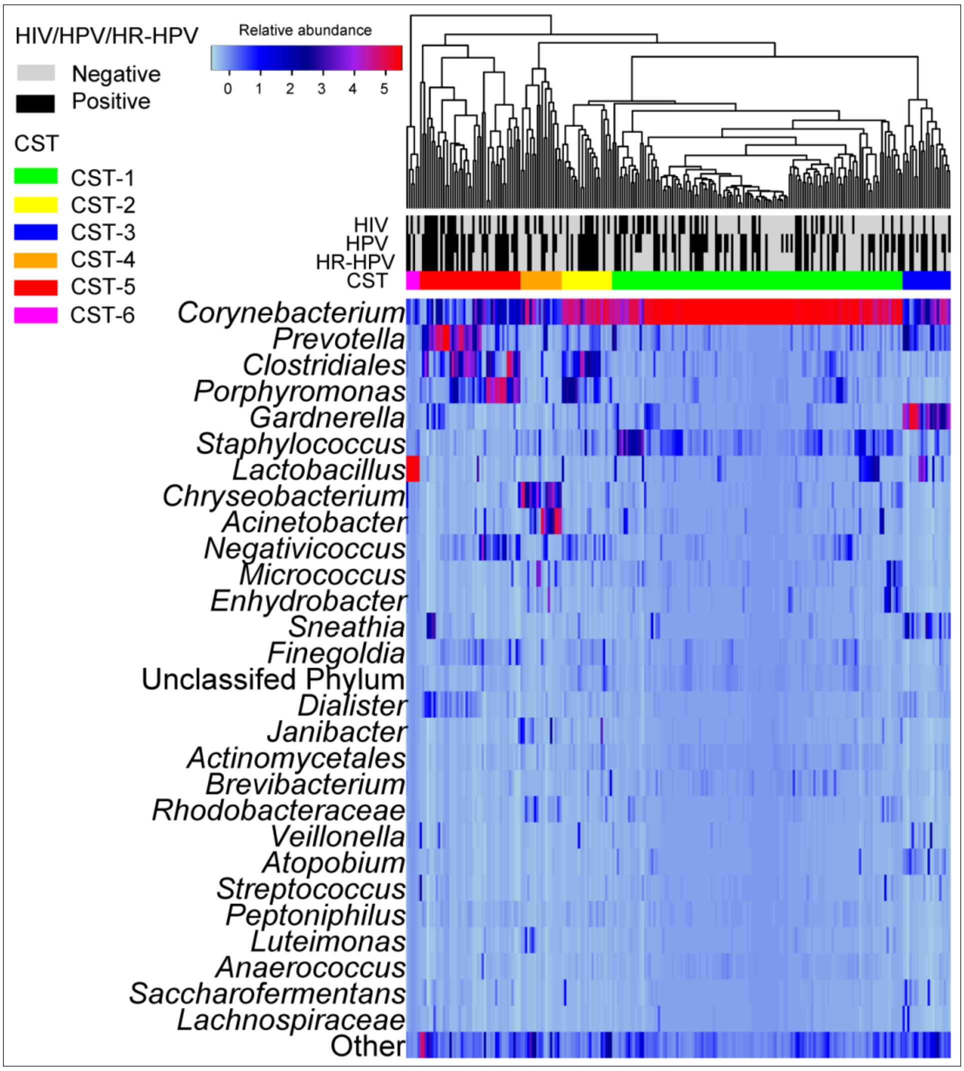


Fig. 5.3. Heatmap of the relative abundances of bacterial taxa in 238 penile microbiotas. Taxa names of bacteria are shown on the left of the heatmap. The "Other" comprised of pooled bacteria at <0.31% relative abundance each (n=622, total relative abundance: 6.49%). Rows represent the bacterial taxa and columns the samples. The colour key for the relative abundances is indicated in the upper right corner. The human immunodeficiency virus (HIV), human papillomavirus (HPV) and high-risk human papillomavirus (HR-HPV) infection status of the men are indicated. The dendrogram based on average linkage hierarchical clustering of the Bray-Curtis dissimilarity matrix is shown and was used to define the six community state types (CST-1 through -6).

Additional descriptions of these CSTs, including their prevalence, and the relative abundance of the most abundant bacterial taxa are summarised in Table 5.3. The relative abundance of *Corynebacterium* in each CST is also indicated in Table 5.3, as this was found to the most abundant genera in the penile microbiotas. The relative abundance of this genus decreased from CST-1 through to CST-6. The most prevalent CST was CST-1 and was dominated by *Corynebacterium*, with this bacteria making up most of the bacterial relative abundance (≥46.4% for hierarchical clustering). CSTs 2-5 were found in 44.1% (105/238) of the men and were characterised by diverse mixed populations of bacteria and lower relative abundances of *Corynebacterium* than CST-1. The most abundant genera (in decreasing relative abundances) in these CSTs were as follows: *Corynebacterium*, unclassified *Clostridiales*, and *Porphyromonas* in CST-2, *Gardnerella* and *Corynebacterium* in CST-3, *Chryseobacterium*, *Corynebacterium*, and *Acinetobacter* in CST-4, and *Prevotella*, unclassified *Clostridiales*, *Corynebacterium*, and *Porphyromonas* in CST-5. The remaining CST, CST-6, was the least prevalent and was dominated by *Lactobacillus* with very low relative abundance of *Corynebacterium*. In subsequent analyses, these CSTs (1-6) were examined individually or in groups based on the dominance of *Corynebacterium* or the diversity of the microbial communities in the CSTs. The specific CST groupings were: *Corynebacterium*-dominated (CST-1), non-*Corynebacterium*-dominated (grouped CST-2, -3, -4, -5, and -6, CSTs 2-6), low-diversity (CST-1 and CST-6), and diverse communities (grouped CST-2, -3, -4, and -5, CSTs 2-5).

Table 5.3. Bacterial description of the six established community state types and their prevalences among the 238 heterosexual Black South African men.

Community state type (CST)	Dominant bacterial taxa (% median relative abundance)	Prevalence of CST (% (n/N))
CST-1	*Corynebacterium* (69.5 (59.5-81.2)) *Staphylococcus* (3.4 (1.5-8.8))	53.4% (127/238)
CST-2	*Corynebacterium* (41.4 (37.9-48.1)) Unclassified *Clostridiales* (12.9 (4.3-22.8)) *Porphyromonas* (8.3 (2.3-23.9))	9.2% (22/238)
CST-3	*Gardnerella* (29.2 (23.3-37.7)) *Corynebacterium* (24.6 (11.2-32.6))	8.8% (21/238)
CST-4	*Chryseobacterium* (22.0 (12.3-34.2)) *Corynebacterium* (16.3 (7.8-29.6)) *Acinetobacter* (15.2 (1.2-36.7))	7.6% (18/238)
CST-5	*Prevotella* (20.1 (2.5-35.7)) Unclassified *Clostridiales* (16.4 (2.1-30.8)) *Corynebacterium* (11.7 (2.8-20.6)) *Porphyromonas* (8.8 (2.7-24.9))	18.5% (44/238)
CST-6	*Lactobacillus* (73.4 (64.1-91.05)) *Corynebacterium* (9.0 (6.4-13.9))	2.5% (6/238)

The relative abundance was expressed as median with interquartile ranges (IQRs, at 25th and 75th percentiles).

Next, the differentially abundant bacterial taxa were determined in *Corynebacterium*-dominated microbiotas (CST-1) versus non-*Corynebacterium*-dominated high-diversity microbiotas (CSTs 2-5) using linear discriminant analysis. We noted that 100 genera were differentially abundant in CST-1 and CSTs 2-5 (Supplemental Table 5.2). CST-1 and CSTs 2-5 had 51 and 49 enriched genera, respectively. These results highlight that BV-associated bacteria (e.g., *Prevotella*, *Porphyromonas*, *Gardnerella*, *Dialister*, *Finegoldia*, *Mobiluncus*, and *Mycoplasma*) occur in penile microbiotas and that these taxa are more predominant in *Corynebacterium*-deficient microbiotas than in *Corynebacterium*-dominated microbiotas.

5.3.6 Alpha diversity of the penile microbiotas

To assess the within-sample taxonomic diversities, the alpha diversity indices of the penile bacterial communities were calculated. The alpha diversity across CSTs, HPV status, high-risk HPV status, HIV status, CD4^{+} T-cell count, and partner's BV status were then compared and are shown in Fig. 5.4.

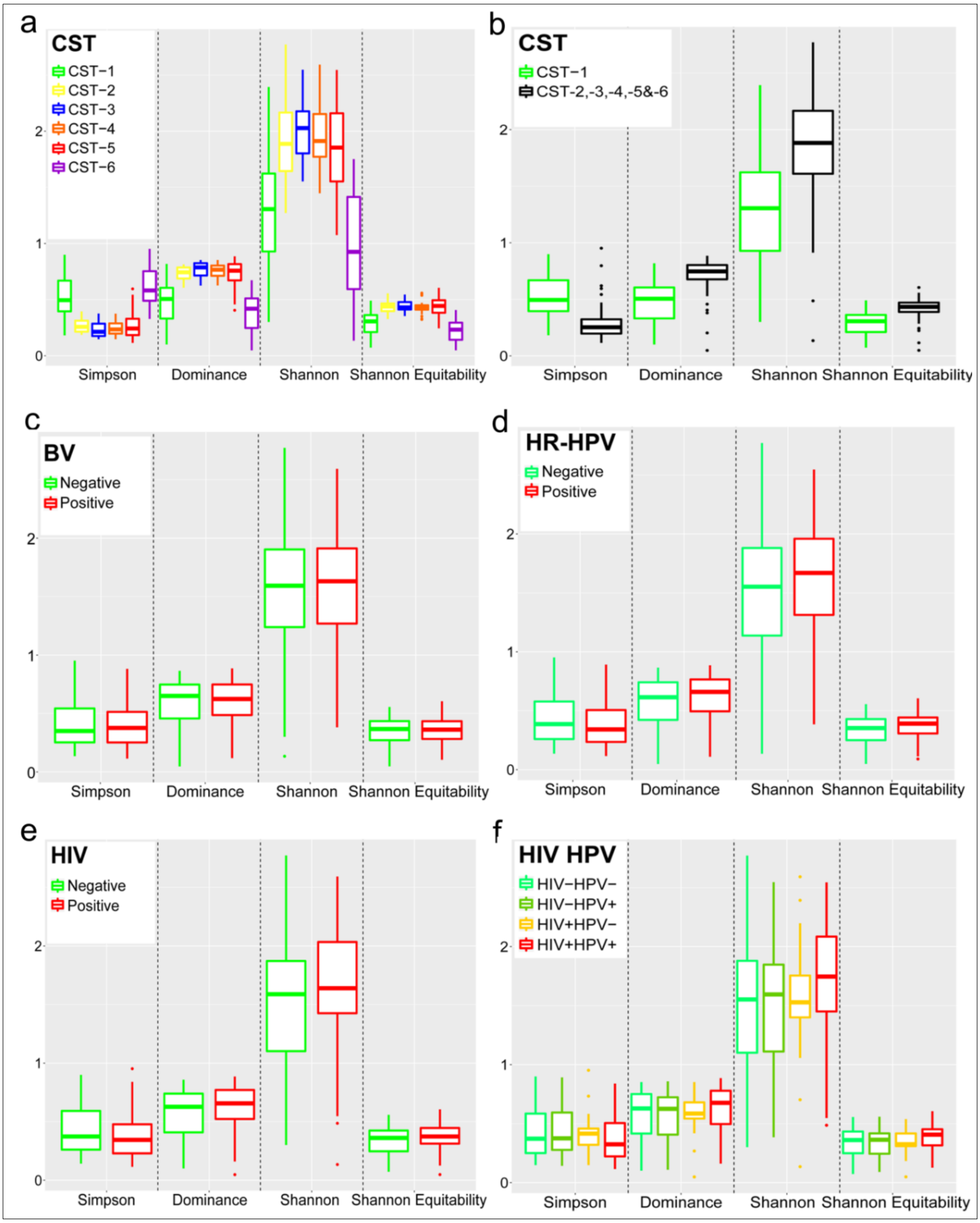


Fig. 5.4. Alpha diversity measures of penile microbiotas. Comparison of the alpha diversity of penile microbiotas grouped by a) community state type (CST), b) *Corynebacterium*-dominance in CST, c) partner's bacterial vaginosis (BV) status, d) high-risk (HR)-HPV infection status, e) human immunodeficiency virus (HIV) infection status, and f) HIV and HPV infection status. In each plot, the box ranges from the first to the third quartile, with the median represented by the horizontal line. The whiskers extend to the smallest and largest non-outliers and outliers are represented by dots.

CST-1 and CST-6 were the least diverse of the six CSTs (Fig. 5.4a). Alpha diversity of the highly prevalent *Corynebacterium*-dominated microbiotas (CST-1) was significantly lower than that for CST-2, -3, -4, and -5 (all $p<0.0001$). Alpha diversity of CTS-6 was significantly lower than CSTs 2, 3, 4, and 5 (all $p<0.001$). The alpha diversities of CST-1 and CST-6 were not significantly different ($p=0.213$).

The alpha diversity of CST-1 was significantly lower than that of non-*Corynebacterium*-dominated CSTs (CSTs 2-6) ($p<0.0001$) (Fig. 5.4b). The alpha diversities of these two groups (CST-1 versus CSTs 2-6) were statistically different ($p<0.0001$) – Simpson index: 0.5 (0.4-0.7) versus 0.3 (0.2-0.3), Dominance index: 0.5 (0.3-0.6) versus 0.7 (0.7-0.8), Shannon index: 1.3 (0.9-1.6) versus 1.9 (1.6-2.2), and Shannon Equitability index: 0.3 (0.2-0.4) versus 0.4 (0.4-0.5).

The alpha diversities of penile microbiotas of men whose female sexual partners were BV-negative were not statistically different from those whose partners were BV-positive (Shannon: 1.6 (1.2-1.9) versus 1.6 (1.2-1.9), respectively ($p=0.681$)) (Fig. 5.4c).

The median alpha diversities of the penile microbiotas of men with and without high-risk HPV infection were similar (Shannon: 1.6 (1.3-2.0) versus 1.5 (1.1-1.9), respectively ($p=0.149$)) (Fig. 5.4d). Similar non-significant results ($p=0.296$) were obtained for HPV-negative (1.5 (1.1-1.9)) versus high-risk HPV-positive (1.6 (1.3-1.9)) groups (Supplemental Fig. 5.1a).

The alpha diversities were not significantly different between men with and without HIV infection (Fig. 5.4e), although there was a trend towards increased microbiota diversity in HIV-positive men compared to HIV-negative men (Shannon: 1.6 (1.4-2.0) versus 1.6 (1.1-1.9) respectively, ($p=0.050$). The alpha diversities of the HIV-positive men with ≤350 $CD4^+$ T-cell count and those with >350 $CD4^+$ T-cell count were not significantly different (Shannon: 1.6 (1.4-2.1) versus 1.7 (1.4-2.0), respectively ($p=0.873$)) (Supplemental Fig. 5.1b).

Next, the diversity in the penile microbiota of men coinfected with both HIV and HPV (HIV+HPV+, 71.6% (63/88)), HIV-positive men without HPV infection (HIV+HPV-, 28.4% (25/88%)), HIV-negative men with HPV infection (HIV-HPV+, 44.7% (67/150)), and men without HIV and HPV coinfections (HIV-HPV-, 55.3% (83/150)) were compared. Penile microbiotas of HIV+HPV+ men had statistically higher alpha diversities (Shannon: 1.7 (1.4-2.1) than HIV-HPV- men (Shannon: 1.6 (1.1-1.9), $p=0.037$) and HIV-HPV+ men (Shannon: 1.6 (1.1-1.9) $p=0.045$) (Fig. 5.4f). Alpha diversities of penile microbiotas of HIV+HPV- men (Shannon: 1.5 (1.3-1.8) did not significantly vary from HIV+HPV+ men ($p=0.182$), HIV-HPV- men ($p=0.828$), and HIV-HPV+ men ($p=0.875$). Alpha diversities were similar between HIV-HPV+ men and HIV-HPV- men (p=0.918). HIV-positive men with high-risk HPV infections (60.2% (53/88)) had significantly higher alpha diversity than HIV-negative men with high-risk HPV (32.7% (49/150): Shannon: 1.7 (1.5-2.1) versus 1.6 (1.1-1.9), respectively (p=0.025), results not shown.

5.3.7 Beta diversity of the penile microbiotas

Beta diversities of the penile microbiotas were computed using UniFrac distances and compared across CSTs, HPV, high-risk HPV, HIV, CD4$^+$ T-cell count, and partner's BV status groups. The resultant 2D PCoA plots of clustering based on weighted UniFrac distances are shown in Fig. 5.5. The first two components (PC1 and PC2) of PCoA plots explained the highest percentages of variations in bacterial communities as indicated in Fig. 5.5.

PCoA of the beta diversity of the penile microbiotas confirmed that the six established CSTs were distinct (p=0.001) and clustered according to the unique bacterial communities in each (Fig. 5.5a). Communities with high relative abundances of *Corynebacterium* (CST-1) had lower beta diversities and clustered together. Conversely, communities with lower relative abundances *Corynebacterium* and colonised with other bacteria were spread out from the *Corynebacterium*-dominated samples (Fig. 5.5b).

No distinct segregation of the penile bacterial communities by partner's BV status (Fig. 5.5c) was observed. The beta diversities of men with and without BV sexual partners were statistically similar (p=0.495).

The beta diversity of the communities in high-risk HPV-uninfected men was significantly lower than high-risk HPV-infected men (p=0.006) (Fig. 5.5d). About 63.8% (81/127) of the men without high-risk HPV infections had *Corynebacterium*-dominated penile microbiotas that clustered together (Fig. 5.5a-b). When examining the penile microbiotas of men with and without any HPV infection, however, no distinct segregation was observed (p=0.072) (Supplemental Fig. 5.2a).

The penile bacterial communities of men with and without HIV infection showed no distinct separation, indicating that HIV did not impact the clustering of the communities (p=0.129) (Fig. 5.5e). Similar beta diversities were observed in communities from HIV-positive men with ≤350 CD4$^+$ T-cell count and those with >350 CD4$^+$ T-cell count (p=0.129) (Supplemental Fig. 5.2b).

Coinfection with HIV and HPV (Fig. 5.5f) or high-risk HPV (results not shown) did not affect the clustering of the bacterial communities (HPV/HIV coinfection: p=0.123 and high-risk HPV/HIV coinfection: p=0.081).

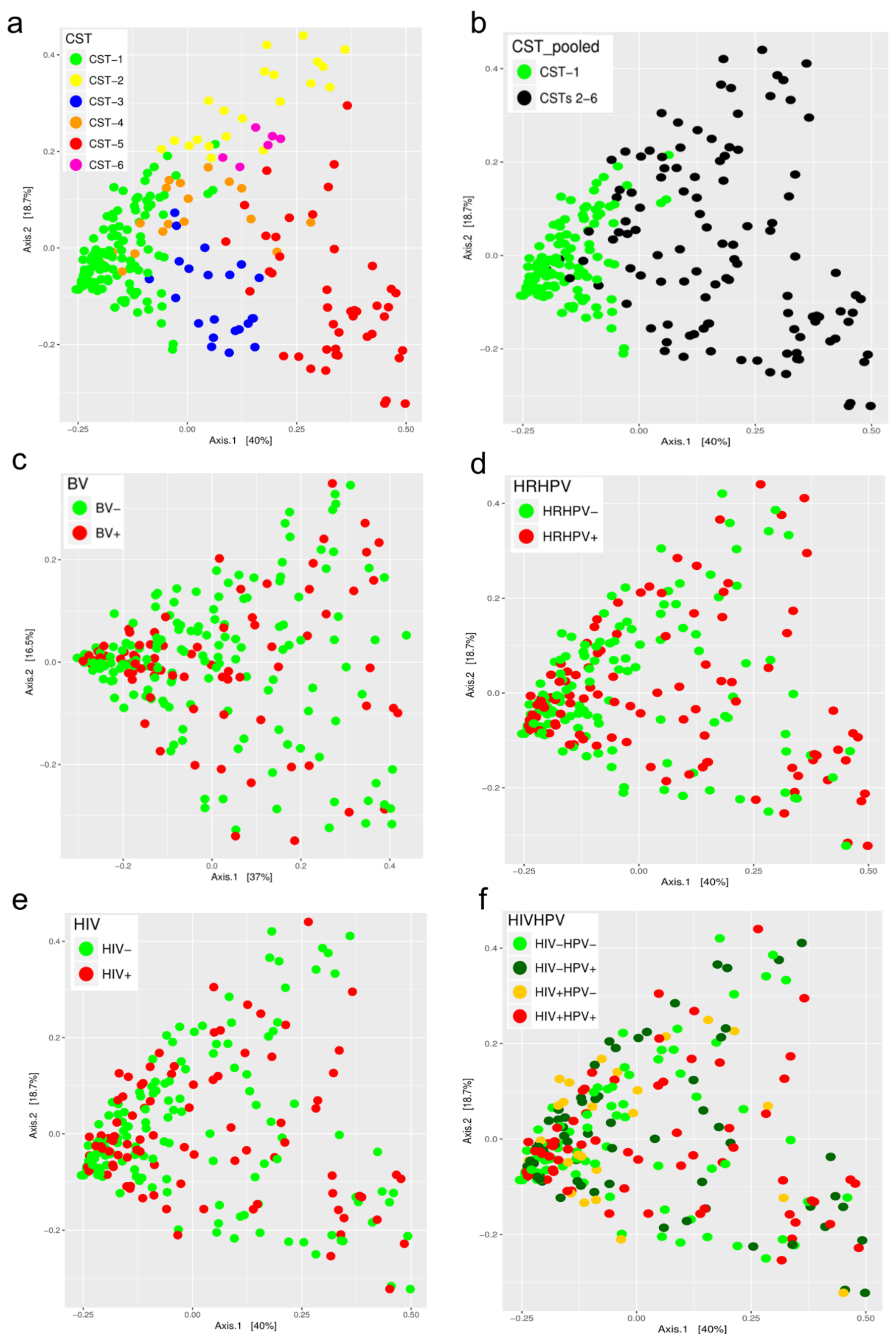


Fig. 5.5. Beta diversity of the penile microbiotas. Principal Coordinates Analysis (PCoA) plots of the weighted UniFrac distances of the penile microbiotas coloured according to a) community state type (CST), b) *Corynebacterium*-dominance in CST, c) partners' bacterial vaginosis (BV) status, d) high-risk (HR)-HPV infection status, e) human immunodeficiency virus (HIV) infection status, and f) HIV and HPV infection status. The first two principal coordinate axes of variations and the percentage variation explained by each (Axis.1: 40% and Axis.2: 18.7% (a-f) are shown. Each solid point is a bacterial community.

5.3.8 Comparison of characteristics of the men with Corynebacterium-dominated and non-Corynebacterium-dominated penile microbiotas

To investigate whether the highly prevalent *Corynebacterium*-dominated penile microbiotas (CST-1) were different from non-*Corynebacterium*-dominated penile microbiotas (pooled CSTs 2-6), statistical associations with participants' metadata were assessed. The results of comparison of CST-1 versus CSTs 2-6 according to participants' metadata are shown in Table 5.4.

Table 5.4. Comparison of the demographic, behavioural, and clinical characteristics of men with *Corynebacterium*-dominated and non-*Corynebacterium*-dominated penile microbiotas.

Characteristic	***Corynebacterium*-dominated microbiotas (CST-1)**	**Non-*Corynebacterium*-dominated microbiotas (CSTs 2-6)**	**p-value$^{\#}$**
	(N = 127, 53.4%)	**(N = 111, 46.6%)**	
Age (years)	37.0 (31.0-44.0)	36.0 (30.0-44.0)	0.597
Age at sexual debut (years)$^{\wedge}$	17.0 (15.0-18.0)	16.0 (15.0-18.0)	0.975
Number of lifetime sexual partners	6.0 (3.0-12.0)	7.0 (3.0-14.0)	0.693
Number of sexual acts with study partner in last month$^{\wedge}$	4.0 (2.0-10.0)	4.0 (2.3-10.0)	0.763
Current use of condom (% (n/N))	64.7 (77/119)	63.4 (64/101)	0.837
HPV (%(n/N))			
Any HPV type	51.2 (65/127)	58.6 (65/111)	0.254
Any high-risk HPV type	36.2 (46/127)	50.5 (56/111)	**0.027**
Single HPV infection	18.1 (23/127)	12.6 (14/111)	0.243
Multiple HPV infection	33.1 (42/127)	45.9 (51/111)	**0.042**
HPV status at 6 month visit (% (n/N))			
Negative	33.3 (25/75)	28.6 (16/56)	0.561
Acquired	17.3 (13/75)	21.4 (12/56)	0.555
Cleared	16.0 (12/75)	8.9 (5/56)	0.234
Persisted	1.3 (1/75)	3.6 (2/56)	0.576
Complex^{+}	32.0 (24/75)	37.5 (21/56)	0.512
HIV (% (n/N))	33.9 (43/127)	40.5 (45/111)	0.287
CD4^{+} T-cell count (cells/µl)$^{\wedge}$	348.5 (240.5-479.8)	308 (217.0-484.5)	0.471
HIV viral load in HIV-positive men (copies/ml)$^{\wedge}$	4.0 (3.1-4.7)	4.2 (3.8-4.8)	0.376
Circumcised (% (n/N))	93.6 (117/125)	95.1 (98/103)	0.616
Sexual partner's incidental BV on Papanicolaou smear (% (n/N))	33.9 (43/127)	33.0 (36/109)	0.893
Cigarette use (% (n/N))			
Never smoked	15.7(20/127)	18.2 (20/110)	0.618
Ex-smoker	17.3 (22/127)	16.4 (18/110	0.844
Current smoker	66.9 (85/127)	65.5 (72/110)	0.811

Abbreviations: HPV – human papillomavirus, HIV – human immunodeficiency virus, BV – bacterial vaginosis, CST – community state type.

[#] p-values are shown for comparison of each variable between CST-1 and CSTs 2-6. Associations of continuous variables (expressed as medians with interquartile ranges (IQRs, at 25th and 75th percentiles)) and categorical variables were computed by Mann-Whitney unpaired and Chi-square/Fisher's exact tests, respectively. Significant p-values (<0.05) are shown in **bold**.

^Data was not available on the age at sexual debut for three men (three with *Corynebacterium*-dominated microbiotas), lifetime number of sexual partners of four men (one with *Corynebacterium*-dominated microbiota and three with non-*Corynebacterium*-dominated microbiotas), number of sexual acts with study partner in the last month of six men (three with *Corynebacterium*-dominated microbiotas and three with non-*Corynebacterium*-dominated microbiotas), CD4^{+} T-cell count of one man (with *Corynebacterium*-dominated microbiota) and viral load measurement for thirty one men (seven with *Corynebacterium*-dominated microbiotas and twenty four with non-*Corynebacterium*-dominated microbiotas).

*The "complex" HPV status consisted of men that had cleared one HPV genotype but were infected with new genotypes and/or had genotypes that had persisted at the 6-month follow-up.

Among the factors assessed, only high-risk HPV and multiple HPV infection statistically differed between the two penile microbiota groups (CST-1 and CSTs 2-6). Men with

Corynebacterium-dominated microbiotas had a lower risk for being infected with high-risk or multiple HPV than men with non-*Corynebacterium*-dominated microbiotas (RR: high-risk HPV: 0.8 [95% CI 0.6-1.0]; p=0.027 and multiple HPV: 0.8 [95% CI 0.6-1.0]; p=0.042).

Since men in CST-1 had less prevalent high-risk HPV and multiple HPV infections compared to men in CSTs 2-6 (Table 5.4), we next compared the prevalence of HPV and high-risk HPV, and HIV infections in CST-1 versus each of the CSTs (with sufficient numbers of men, n≥20) (Table 5.5). Men in CST-5 were significantly more likely to be infected with HPV or high-risk HPV than men in CST-1 (Table 5.5, p=0.026 and p=0.004, respectively). No statistical difference in HIV prevalence in the CSTs was observed, although a trend towards a higher HIV prevalence in CST-5 compared to CST-1 was observed (Table 5.5, p=0.057).

Table 5.5. Association of selected community state types with prevalent HPV, high-risk HPV, and HIV infections.

CST	HPV prevalence (% (n/N))	RR (95% CI) (p-value[#])	High-risk HPV prevalence (% (n/N))	RR (95% CI) (p-value[#])	HIV prevalence (% (n/N))	RR (95% CI) (p-value[#])
CST-1	51.2 (65/127)	Ref	36.2 (46/127)	Ref	33.9 (43/127)	Ref
CST-2	54.5 (12/22)	1.1 (0.5-2.4) 0.771	54.5 (12/22)	1.9 (0.9-4.1) 0.104	40.9 (9/22)	1.2 (0.6-2.8) 0.522
CST-3	57.1 (12/21)	1.2 (0.6-2.7) 0.613	42.9 (9/21)	1.3 (0.6-2.8) 0.560	19.0 (4/21)	0.5 (0.2-1.4) 0.177
CST-5	70.5 (31/44)	1.8 (1.1-3.3) **0.026**	61.4 (27/44)	2.1 (1.3-3.6) **0.004**	50.0 (22/44)	1.6 (1.0-2.7) 0.057

Abbreviations: CST – community state type, HPV – human papillomavirus, HIV – human immunodeficiency virus, RR – relative risk, CI – confidence interval, Ref – reference.

[#]p-values were computed using computed Chi-square test. Significant p-values (<0.05) are shown in **bold**.

5.3.9 Differential bacterial relative abundance and potential biomarkers for HPV, high-risk HPV and HIV infections

In order to gain more insight into the bacterial taxa accounting for the different penile microbiotas in men with and without HPV, high-risk HPV, and HIV infections, LefSe analysis was applied. High-risk HPV-positive group comprised of men with any of the high-risk HPV genotypes regardless of infection with any of the low-risk HPV genotypes whereas high-risk HPV-negative group consisted of men that were negative for HPV and those positive for low-risk HPV.

Potential biomarkers for HPV and high-risk HPV were first assessed irrespective of HIV status. A total of 61 bacterial features were significantly different by HPV status at LDA score >2.0 or <-2.0 and p<0.05 (Supplemental Fig. 5.3b). The most significant classified genera that were found to be significantly more abundant in men with HPV infection (54.6% (130/238)) than in men without infection (45.4% (108/238)), were BV-associated bacteria, such as *Sneathia*, *Porphyromonas*, *Prevotella*, *Peptoniphilus*, and *Dialister*, and other genera, such as *Jonquetella,* (LDA score <-3.0, p<0.05) (Supplemental Fig. 5.3a).

Pseudomonas was the only genus, at LDA score >3.0, p<0.05, that was found in higher relative abundances in men without HPV infection than in men with HPV infection.

A total of 78 bacterial features for high-risk HPV were detected at LDA scores >2.0 or <-2.0 (results not shown). The most significant of these potential biomarkers for high-risk-HPV infection (LDA score <-3.0, p<0.05) are shown in Fig. 5.6. As was observed for HPV infection, *Prevotella*, *Dialister*, *Peptoniphilus*, and *Campylobacter* were more abundant in the penile microbiotas of men with high-risk HPV infection (42.9% (102/238)) than uninfected men (57.1% (136/238)). In high-risk HPV-negative men, *Corynebacterium*, *Micrococcus*, *Sanguibacter*, and *Brevibacterium* were highly abundant. In both HPV and high-risk HPV-infected men, families *Campylobacteraceae*, *Prevotellaceae*, and *Veillonellaceae* were found to be more abundant than in men without infection.

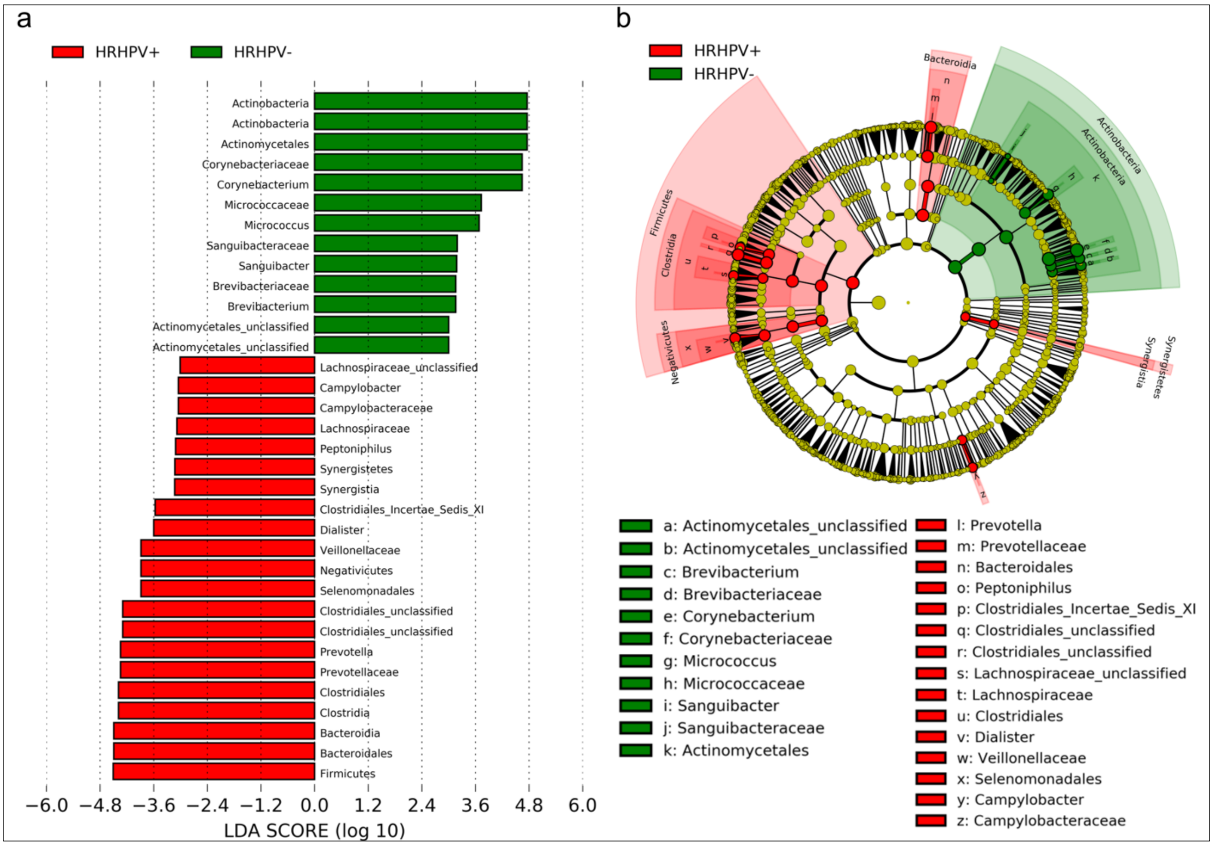


Fig. 5.6. Potential biomarkers for high-risk HPV by LefSe. a) Histogram of differentially abundant taxa in penile microbiotas of men with and without high-risk (HR)-HPV infections, and b) A six-level cladogram with a taxonomic hierarchical structure. Each coloured solid represents a taxon and its diameter is proportional to the taxon's relative abundance. Red and green solids represent statistically significant taxon ranks in HR-HPV-positive and HR-HPV-negative group, respectively. Only differentially abundant features at logarithmic LDA scores >3.0 or <-3.0 are shown.

Potential biomarkers for HPV and high-risk HPV were additionally assessed in HIV-negative men only (63.0% (150/238)). When only HIV-negative men were considered, genus *Jonquetella* in the family *Synergistaceae* was more abundant in HPV-positive men (44.7%

(67/150)) than in HPV-negative men (55.3% (83/150)). Among men without HPV infection, *Brevibacterium* in the family *Brevibacteriaceae*, *Pseudomonas* in the family *Pseudomonadaceae*, *Haemphilus*, *Kocuria*, unclassified *Flavobacteriaceae*, and *Granulicatella* were observed to be more abundant than in men with HPV infection (Supplemental Fig. 5.4a). In the context of high-risk HPV infection, the genera that were more abundant in men with high-risk HPV infection (32.7% (49/150)) than in men without high-risk HPV infection (67.3% (101/150)) included *Jonquetella*, as well as unclassified *Clostridiales*, unclassified *Campylobacteraceae*, *Paenalcaligenes*, *Pimelobacter*, unclassified *Porphyromonadaceae*, and *Blastomonas* (Supplemental Fig. 5.4b). Similar to what was found in men without HPV, *Brevibacterium, Kocuria*, unclassified *Flavobacteriaceae*, and *Granulicatella* were more abundant in men without high-risk HPV infections than in men with high-risk HPV infections. A few additional features were differentially abundant, including genera of unclassified *Corynebacteriaceae*, *Eubacterium*, *Brachybacterium*, and *Micrococcus*.

Bacterial taxa that were statistically different between men with and without HIV infections were also evaluated. These included 31 differentially abundant bacterial features in the penile microbiotas. Twenty one of these features were found in men who were infected with HIV (37.0% (88/238)) while the rest of the taxa (n=10) were found in men who were negative for HIV infection (63.0% (150/238)) (Fig. 5.7). Men with HIV infections had higher relative abundances of *Staphylococcus*, *Faecalibacterium*, *Strenotrophominas*, *Jonquetella*, *Ruminococcus*, *Roseburia*, *Pseudochrobactrum*, and *Lamia* than men without HIV infections. On the other hand, the relative abundances of the family Actinobacteria and the genera *Propionibacterium* and *Nosocomiicoccus* were greater in HIV-negative men than HIV-positive men. None of the predominant bacterial families in the penile microbiotas (*Corynebacteriaceae*, *Prevotellaceae* and unclassified *Clostridiales*) were found to differ significantly in relative abundance between HIV-infected and -uninfected men.

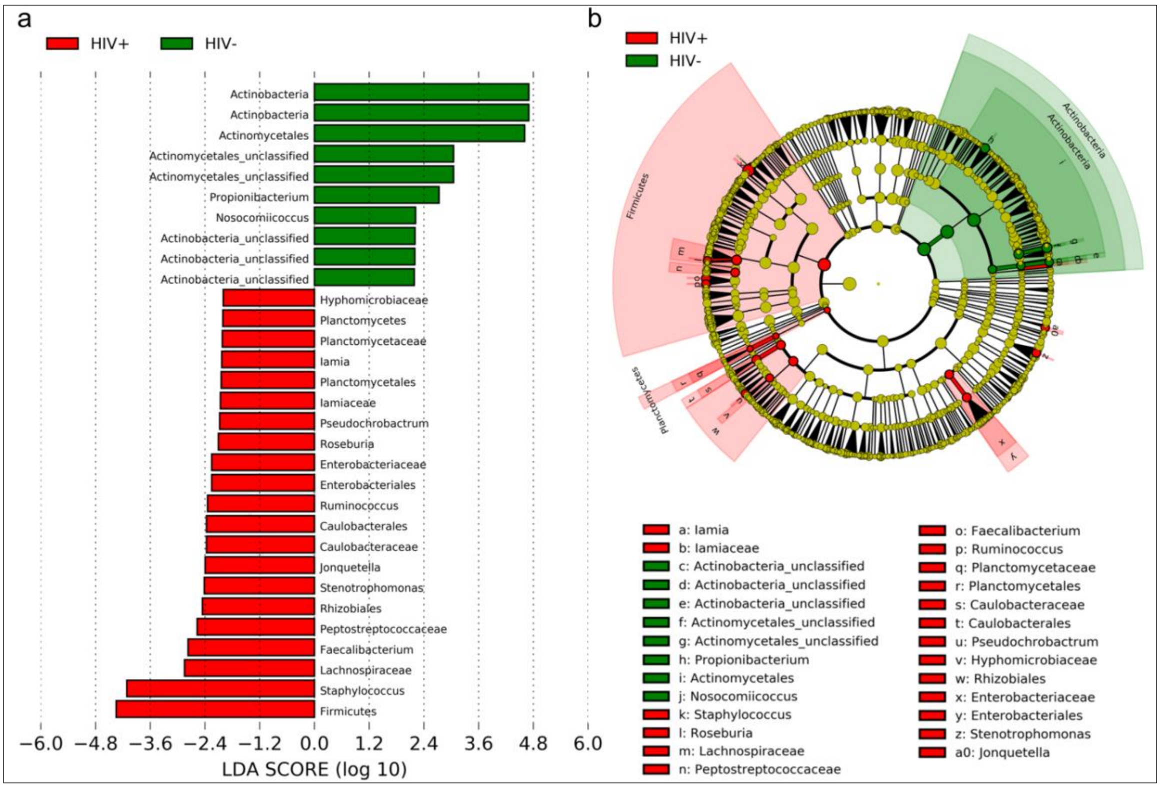


Fig. 5.7. Potential biomarkers for HIV infection by LefSe. a) Histogram of differentially abundant taxa in penile microbiotas of men with and without human immunodeficiency virus (HIV) infection, and b) A six-level cladogram with a taxonomic hierarchical structure. Each coloured solid represents a taxon and its diameter is proportional to the taxon's abundance. Red and green solids represent statistically significant taxon ranks in men with and without HIV infection, respectively. Only differentially abundant features at logarithmic LDA scores >2.0 or <-2.0 are shown.

The relative abundances of the predominant penile bacterial families stratified by HIV infection status are shown in Supplemental Table 5.1. *Corynebacteriaceae*, *Prevotellaceae*, and unclassified *Clostridiales* were found to be the most abundant families regardless of the HIV status. Each of these families occurred at relative abundances of between 5.2% and 49.7%.

5.3.10 Investigation of the impact of partner's bacterial vaginosis status on the penile microbiota

The penile microbiotas of the men were also assessed according to the BV status of their female sexual partners. Fig. 5.8 is a heatmap of the penile microbial communities stratified by BV status of the female sexual partner. Of the 238 men finally included in the analyses, partner BV status was not available for two men (0.8%). About a third (33.5%, 79/236) of the men were in a relationship where the female partner was BV-positive (hereafter referred to as a BV-positive relationship). Of these men, 54.4%, 8.9%, 13.9%, 5.1%, and 17.7% were in CST-1 to CST-5, respectively. None of the men in BV-positive relationship was found to be in CST-6 (*Lactobacillus*-dominated penile microbiotas). The prevalences of the CSTs among men whose female sexual partners were BV-negative were as follows: CST-1 (53.5%), CST-2 (8.9%), CST-3 (5.7%), CST-4 (8.9%), CST-5 (19.1%), and CST-6 (3.8%). Partners' BV status did not statistically differ between men with penile CST-1 (dominated by *Corynebacterium*) and those with other CSTs (non-*Corynebacterium*-dominated microbiotas, CSTs 2-6, p=0.893) or diverse CSTs (CSTs 2-5, p=0.862).

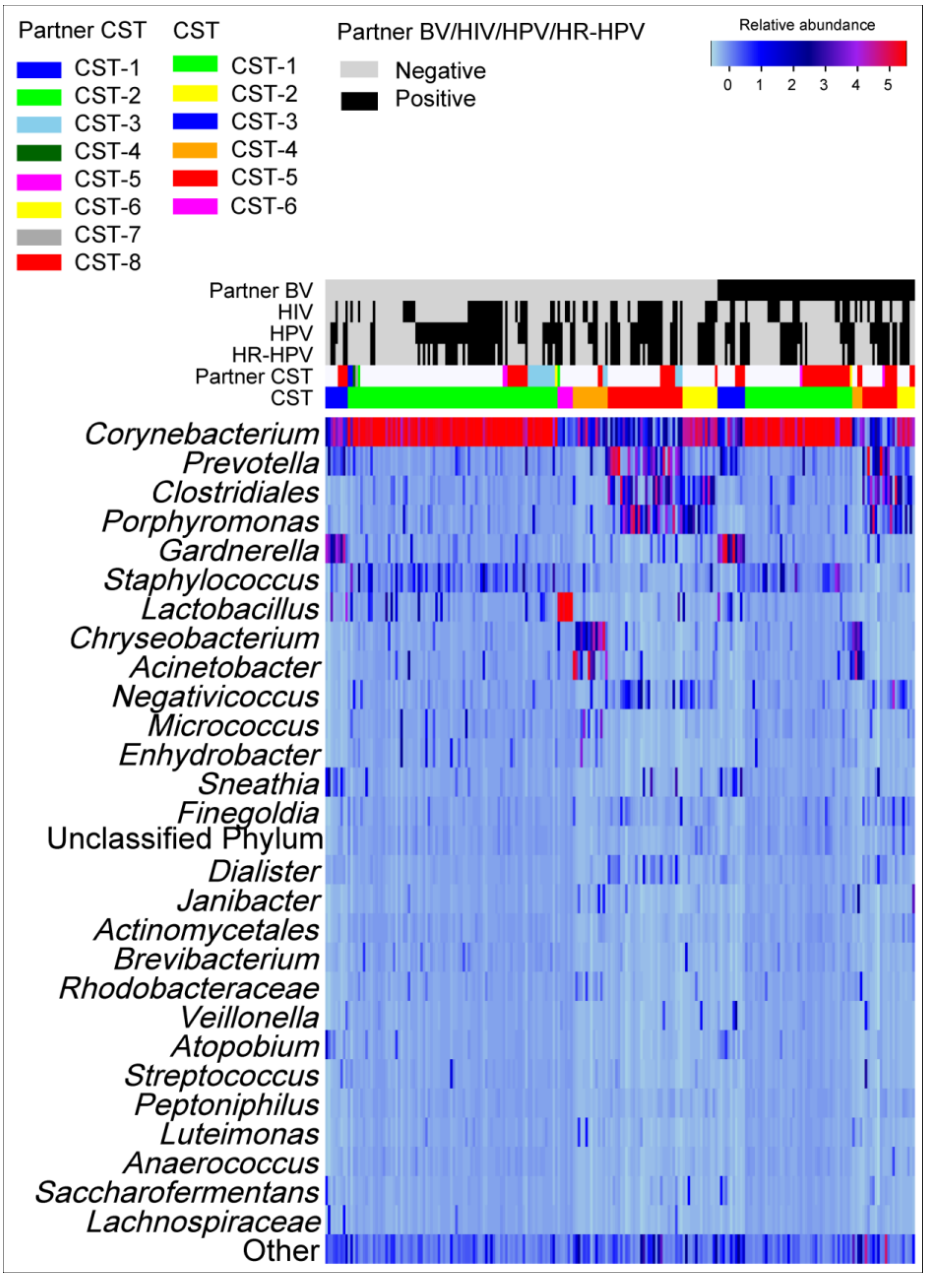


Fig. 5.8. Heatmap of the relative abundances of bacterial taxa in 236 penile microbiotas. Taxa names of bacteria at ≥0.16% relative abundance are shown on the left of the heatmap. The "Other" comprised of pooled bacteria (n=622, total relative abundance: 6.21%). Rows represent the bacterial taxa and columns the samples. Samples were clustered by the BV status of the female sexual partners. The colour key for the relative abundances is indicated in the upper right corner. The human immunodeficiency virus (HIV), human papillomavirus (HPV), high-risk human papillomavirus (HR-HPV) infection status of the men, and female partner's BV status and community state type (CST) are indicated. The CSTs in women were defined in Chapter 4 (subsection 4.3.5).

One hundred and eight men (45.4%) had female sexual partners whose CSTs were characterised in Chapter 2 and/or 4. The associations of male CSTs (CST-1 (dominated by *Corynebacterium*), and CSTs 2-5 (diverse penile microbiotas), with female CSTs (*Lactobacillus*-dominated, non-*Lactobacillus*-dominated, CST-3 (dominated by *L. iners*) and CST-8 (diverse CVMs)) were explored using Chi-square/Fisher's exact test (Supplemental Table 5.3). Men with highly diverse penile microbiotas (CSTs 2-5) were more likely to have female partners with non-*Lactobacillus*-dominated (RR: 1.8 [95% CI 1.0-3.4], 0.030) or diverse CVMs (1.9 [95% CI 1.0-3.6], 0.022) compared to men with low-diversity CSTs: *Corynebacterium*- and *Lactobacillus*-dominated penile microbiotas (Supplemental Table 5.3). Although not statistically significant, men with *Corynebacterium*-dominated microbiotas were more likely to have female partners with *Lactobacillus*-dominated (1.4 [95% CI 1.0-1.9], p=0.082) and *L. iners*-dominated CVMs (1.5 [95% CI 1.0-2.1], p=0.054) compared to men with diverse microbiotas. All men with *Lactobacillus*-dominated CSTs had female partners with *Lactobacillus*-dominated CVMs.

The prevalences and relative abundances of bacteria associated with normal Nugent scores (specifically *Corynebacterium*, *Lactobacillus*, and *Staphylococcus*) and BV Nugent scores (for example, *Prevotella* and *Treponema*) in men in BV-positive and BV-negative relationships are shown in Table 5.6. These penile bacteria had previously been uniquely associated with female partner Nugent score using indicator analysis [194]. We also included *Atopobium*, a BV-associated bacterium [51]. Amongst the normal Nugent score indicators, we found that *Lactobacillus* was significantly more abundant in the penile microbiotas of men in BV-negative than in BV-positive relationships (p=0.007) (Table 5.6). The BV Nugent score-associated bacteria that were statistically more prevalent in men in BV-positive relationships compared to BV-negative relationships were: *Atopobium* (p=0.017), *Saccharofermentans* (p=0.003), *Prevotella* (p=0.045), and *Sneathia* (p=0.003). In terms of relative abundance, only *Lactobacillus* was statistically different men in BV-negative versus BV-positive relationships.

Table 5.6. Comparison of the prevalence and relative abundance of common normal and BV Nugent score-associated bacteria in penile microbiotas of men in relationships with BV-negative and BV-positive women.

Bacteria	**Prevalence**			**Relative abundance***		
	BV-negative female sexual partner (N = 157, 66.5%) n (%)	**BV-positive female sexual partner (N = 79, 33.5%) n (%)**	**p-value#**	**BV-negative female sexual partner (N = 157, 66.5%) median (IQR)**	**BV-positive female sexual partner N = 79, 33.5%) median (IQR)**	**p-value#**
BV Nugent score indicator						
Atopobium	85 (54.1)	56 (70.9)	**0.013**	0.04 (0.01-0.47)	0.06 (0.02-0.38)	0.154
Dialister	137 (87.3)	65 (82.3)	0.304	0.14 (0.02-0.76)	0.18 (0.29-1.14)	0.390
Gardnerella	125 (79.6)	69 (87.3)	0.143	0.11 (0.02-2.27)	0.30 (0.04-2.67)	0.057
Mobiluncus	49 (31.2)	32 (40.5)	0.156	0.02 (0.01-0.06)	0.03 (0.01-0.08)	0.528
Peptostreptococcus	39 (24.8)	23 (29.1)	0.482	0.02 (0.01-0.04)	0.02 (0.01-0.03)	0.883
Porphyromonas	154 (98.1)	77 (97.5)	1.000	0.47 (0.08-4.55)	0.33 (0.07-3.11)	0.291
Prevotella	157 (100.0)	77 (97.5)	0.111	1.12 (0.23-5.35)	1.61 (0.24-7.36)	0.706
Saccharofermentans	77 (49.0)	55 (69.6)	**0.003**	0.04 (0.01-0.38)	0.06 (0.02-0.32)	0.549
Sneathia	86 (54.8)	59 (74.7)	**0.003**	0.05 (0.01-0.43)	0.12 (0.02-1.29)	0.361
Treponema	20 (12.7)	6 (7.6)	0.234	0.01 (0.00-0.02)	0.02 (0.01-0.04)	0.457
Unclassified *Clostridiales*	157 (100.0)	79 (100.0)	1.000	0.51 (0.12-4.63)	0.86 (0.10-4.44)	0.796
Unclassified *Clostridiales Family XI*	115 (73.2)	63 (79.7)	0.274	0.07 (0.01-0.33)	0.07 (0.01-0.28)	0.757
Normal Nugent score indicator						
Corynebacterium	157 (100.0)	79 (100.0)	1.000	47.44 (22.22-70.89)	48.97 (19.65-70.03)	0.900
Lactobacillus	128 (81.5)	65 (82.3)	0.889	0.16 (0.03-0.96)	0.06 (0.01-0.42)	**0.007**
Staphylococcus	157 (100.0)	78 (98.7)	0.335	1.67 (0.45-5.03)	1.96 (0.29-5.20)	0.672

Abbreviations: BV – bacterial vaginosis.

*Only participants who had the bacterial taxa in consideration were included in the computation.

#p-values are shown for the comparison of BV-negative versus BV-positive relationship groups. Prevalence and relative abundance (expressed as medians with interquartile ranges (IQRs, at 25th and 75th percentiles)) were compared with the Chi-square/Fisher's exact test and Mann-Whitney unpaired test, respectively. Significant p-values (<0.05) are shown in **bold**.

Differentially abundant bacterial features between the penile microbiotas of men with and without partners with BV were also compared using LefSe. These differentially abundant features could be BV indicators in the female sexual partners. A total of 23 differentially abundant features were identified (Fig. 5.9). Family *Lactobacillaceae*, including genus *Lactobacillus*, were found to be more abundant in men whose female sexual partners were BV-negative relative to men whose female sexual partners were BV-positive. The genera found to be significantly more abundant in men whose female sexual partners were BV-positive included *Illumatobacter*, *Providencia*, and unclassified *Campylobacterales* (LDA score <-3.0, p<0.05). None of the commonly known BV-associated bacteria, for example, *Gardnerella*, *Prevotella*, *Atopobium*, *Shuttleworthia*, *Sneathia*, *Clostridium*, *Mobiluncus*, *Porphyromonas*, *Ureaplasma*, and *Mycoplasma*, were identified as differentially abundant features, supporting the previous finding (Table 5.6).

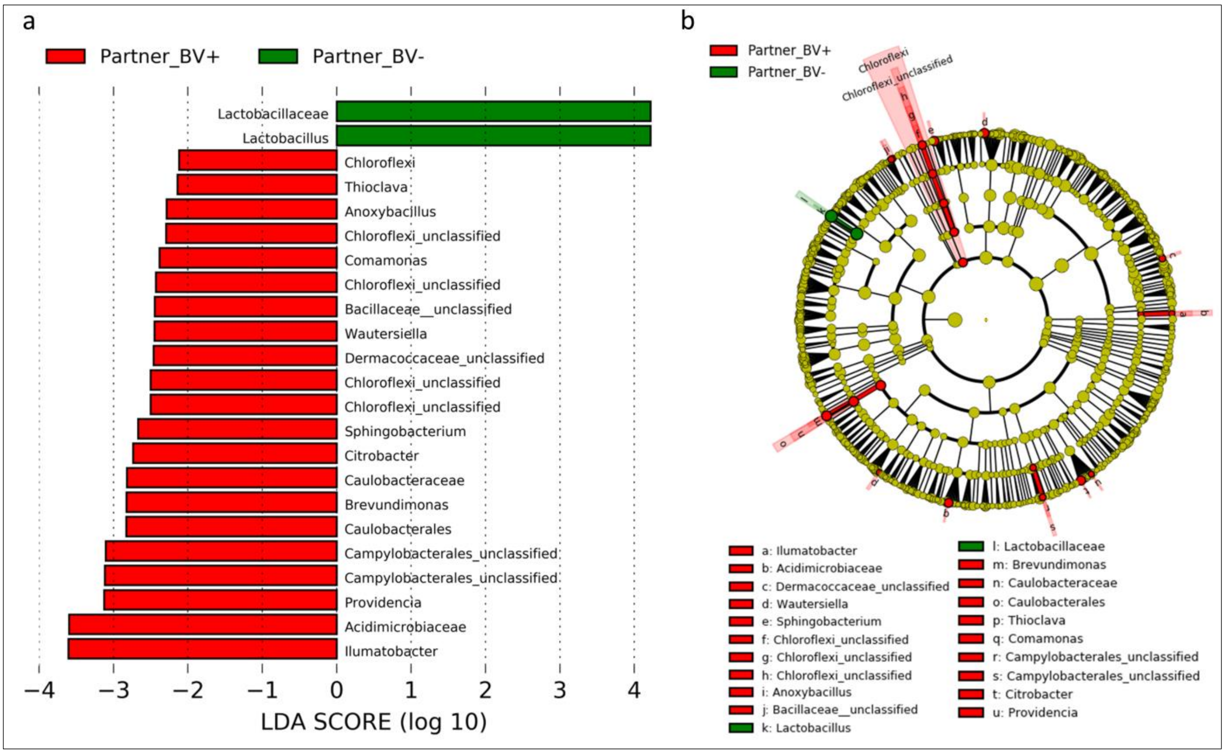


Fig. 5.9. Potential penile biomarkers for BV in female sexual partners by LefSe. a) Histogram of differentially abundant taxa in penile microbiotas of men with and without female partners with BV, b) A six-level cladogram with a taxonomic hierarchical structure. Each coloured solid represents a taxon and its diameter is proportional to the taxon's relative abundance. Red and green solids represent statistically significant taxon ranks in men with female sexual partners who were with (Partner_BV+) and without BV (Partner_BV-), respectively. Only features with logarithmic LDA scores >2.0 or <-2.0 are shown.

5.4 Discussion

5.4.1 Characterisation of the penile skin bacterial communities of Black South African adult men

In this chapter the penile microbiotas of 238 Black South African men, who were mostly (94%) traditionally circumcised, were examined. Sampling of the penile skin microbiota was done by swabbing of the glans, coronal sulci (if present) and shaft of the penis. To date only one other study [110] has used a similar sampling method. The study, by Zozaya and colleagues (2016) [110], examined the penile skin microbiotas (swabs from the glans, the coronal sulcus, and the shaft of the penis) of 56 medically circumcised and 37 uncircumcised men. Other studies have examined the microbiotas of the coronal sulci and glans [256], coronal sulci alone [35,229,254,259], or urinary tract [257,360].

A total of 23 bacterial phyla, 210 families and 650 genera were detected in the penile microbiotas of the 238 South African men. This is a larger array of bacterial phyla, families and genera than reported in previous studies of the penile skin [110], coronal sulci and glans [256], coronal sulci [229], and urinary tract [257,360]. Apart from the penile sampling method, the larger number of taxa in our study could be attributed to genetic background, sociobehavioural factors, large sample size, and heterogeneity in study methodologies. For example, the available penile microbiota studies have relied on different hypervariable region of the 16S rRNA gene sequenced on either Illumina MiSeq or 454 FLX platforms despite the findings that microbiota profile is dependent on the choice of the 16S rRNA region [49,257] and sequencing technology [49,361]. It has been noted that pyrosequenced V1-V3 and V3-V5 datasets do overestimate penile *Corynebacterium* while V6-V9 dataset underestimates *Prevotella* [257]. Even more strikingly, there is evidence that pyrosequenced V1-V3 dataset fails to capture *Gardnerella* [257]. Therefore, there is a probability that some of these methodologies might have missed to disclose a number of penile bacterial taxa.

The two most abundant phyla were Actinobacteria (56.0%) and Firmicutes (21.8%), while the two most abundant families were *Corynebacteriaceae* (47%) and *Prevotellaceae* (7%). These families have been found to be among the most abundant families on the penile skin [110] and coronal sulci [229,254]. In the study by Zozaya and colleagues (2016) [110], *Corynebacteriaceae* (46%) and *Prevotellaceae* (15%) were the most abundant families, among both uncircumcised (*Corynebacteriaceae*: 32% and *Prevotellaceae*: 21%) and circumcised (55% and 11%, respectively) men [110]. Compared to this study [110], the relative abundance of *Corynebacteriaceae* in our cohort is intermediate to that of circumcised and uncircumcised men. However, the relative abundance of *Prevotellaceae* was considerably lower than that described by Zozaya and colleagues (2016) [110].

Prevotellaceae has been found to occur in lower abundances after circumcision [229]. In our study, the vast majority of the men were circumcised. In a study of coronal sulci microbiota [229], the most abundant families were *Pseudomonadaceae* and *Oxalobacteraceae*. In the present study, the relative abundances of *Pseudomonadaceae* and *Oxalobacteraceae* were low (<0.15%), similar to observation by Zozaya and colleagues (2016) [110]. These observations suggest that the different milieus on the penile exhibit differential colonisation by the bacterial families.

Our study sought to evaluate the prevalence and relative abundance of the bacterial genera inhabiting the penis. The most abundant genera were *Corynebacterium* (47.1%), *Prevotella* (6.5%), unclassified *Clostridiales* (5.6%), *Porphyromonas* (4.9%), *Staphylococcus* (4.4%), *Lactobacillus* (3.8%), and *Gardnerella* (3.8%). These taxa were represented by at least three-quarters of all the read sequences. We observed that *Corynebacterium*, an abundant skin bacteria [258], was the most prevalent (100%) and abundant (47.1%) genus. The relative abundance of *Corynebacterium* in our study is similar to that (45.9%) observed on the penile skin by Zozaya and colleagues (2016) [110]. Further, Zozaya and colleagues (2016) [110] reported that the relative abundance of *Corynebacterium* was lower in uncircumcised men (31.7%) compared to circumcised men (55.2%). Thus, the relative abundance of *Corynebacterium* in our cohort of mostly traditionally circumcised men was intermediate to these published results, and higher than uncircumcised men [110]. Greater abundances of *Corynebacterium* have also been reported in coronal sulci microbiota post-circumcision and in circumcised men [229,254]. For example, medical male circumcision was found to increase the abundances of *Corynebacterium* in the coronal sulci microbiota of Ugandan men [229]. Apart from colonising the coronal sulci, *Corynebacterium* can also colonise the semen microbiotas of men in appreciable abundances [106,362].

The genera cited above (i.e., *Prevotella* and unclassified *Clostridiales* among others) have been observed in penile skin [110] and coronal sulci microbiotas [35,194,254,257]. The relative abundance of *Prevotella* in our study (6.5%) is similar to that (7.2%) reported in small cohort study by Plummer and colleagues (2018) [256] that obtained penile swabs from the coronal sulci and glans of predominantly uncircumcised men. However, the relative abundance of *Prevotella* in our cohort was 2.3-fold lower than that (15.0%) observed in the study by Zozaya and colleagues (2016) [110] and lower than what they reported on circumcised (11.4%) and uncircumcised men (20.5%). Our observation of the predominance of an unclassified genus in the order *Clostridiales* is not new. Sequences in order *Clostridiales* are uncommonly classified to lower taxonomic levels and have been reported to be predominant in coronal sulci microbiota [194,254]. Thus, future microbiome studies should be conducted using newly developed bioinformatics tools such as Divisive Amplicon

Denoising Algorithm 2 (DADA2) [363] that can allow accurate classification of sequences to deeper taxonomic ranks and discovery of sequence variations [363]. In our study, the relative abundance of *Porphyromonas* (4.9%) was 4-10-fold higher than has been previously reported in penile skin microbiota studies [110,256]. The relative abundance of *Staphylococcus* (4.4%) was within the range (3-5%) that has been reported in studies with a similar sample type [110,256], although its relative abundance was not predominantly high as reported in a coronal sulci microbiota study (19.2%) [257]. *Staphylococcus* is a predominant skin commensal [258]. *Lactobacillus* was also detected in the penile microbiotas of Black South African men at a relative abundance of 4%. *Lactobacillus* has been observed to colonise the penile milieus of uncircumcised and uncircumcised men, at abundances ranging between 2-8% [110,254]. It has been found to be a predominant taxon in urine and semen microbiotas of men [106,257,362] and more common in men whose female sexual partners are BV-negative [194]. *Gardnerella*, a predominant urine bacterium [257], occurred at a relatively higher relative abundance (3.8%) than has been previously described (1.3-2.1%) [110,256].

Finally, we observed a number of genera that were of low prevalence and/or relative abundance in the penile microbiotas. Examples of these included *Pseudomonas*, *Rubrobacter*, *Schwartzia*, *Dokdonella*, *Brooklawnia*, *Trueperella*, *Mitsuokella*, and *Collinsella*. For example, the relative abundance of *Pseudomonas* in our cohort was 0.1%. This observation is analogous to earlier studies [110,256,257] that have found *Pseudomonas* at relative abundances of <0.1%. Our study also observed bacteria such as *Chryseobacterium*, *Janibacter*, *Oligella*, and *Gemmiger* that have never been reported in penile microbiota studies. Since the clinical and reproductive health implication of these rare and/or low-abundant bacteria in the penile microbiota remains either unknown or poorly defined, further research into their biological roles is warranted.

5.4.2 Oxygen tolerance of the most abundant penile microbiota families

Male circumcision reduces the diversity and density of several anaerobic bacteria [254] due to the elimination of the anoxic environment that supports their colonisation [229]. We therefore examined the oxygen requirements of the 40 most abundant bacterial families in the penile microbiotas. In total, these 40 families had a mean proportional abundance of 98.8%, and thus represent the majority of the penile microbiota families. These bacterial families had differing needs for oxygen that have been described in subsections 5.2.6 and 5.3.3. The mean proportional abundance of *FAn* families (n=7) was over 50% of the penile microbiotas, with the family *Corynebacteriaceae* contributing 47%. The next most abundant group was *An* families (18.2%). High predominance of *FAn* and lower predominance of *An* families have been significantly associated with post-circumcision status [229]. The high predominance of *FAn* bacteria in our cohort could therefore be due to circumcision as the majority of the men were circumcised. *Ae* families (n=12) were about 2-fold less abundant than *An* families (n=7). It is unclear why *Ae* bacteria were less abundant than *An* bacteria in these traditionally circumcised men, but it could be that circumcision does not have a significant impact on aerobic bacteria. Liu and co-workers (2013) [254] noted, for example, only slight increases in *Ae* bacteria following male circumcision. Since the men were traditionally circumcised, it is also possible that those that were partially circumcised influenced the results. Moreover, misreporting of male circumcision status is common among traditionally circumcised men, with at least 20% of the men believing that they are circumcised, yet they are not as determined by physical examination [249]. We also observed abundant families, e.g., unclassified *Clostridiales* and unclassified *Actinomycetales*, whose oxygen requirements remain unknown. Therefore, such families should be explored further. Akin to the observations by Price and colleagues [229], we noted that families with *MAe* and mixed oxygen requirements, such as *Ruminococcaceae* and *Campylobacteraceae*, were not common.

5.4.3 Co-occurrence and co-exclusion patterns of selected bacterial families in the penile microbiotas

The most abundant families in the penile microbiotas were, in order of decreasing relative abundances, *Corynebacteriaceae*, *Prevotellaceae*, unclassified *Clostridiales*, *Porphyromonadaceae*, and *Staphylococcaceae*. Among these, strong positive correlations were observed between *Corynebacteriaceae* and *Staphylococcaceae*, both facultative anaerobes. The positive correlations highlight a possibility of cooperative interaction through metabolic resource overlap [290]. These two families in turn had strong negative correlations with the anaerobic *Prevotellaceae* and *Porphyromonadaceae*, as well as the unclassified *Clostridiales*. Negative correlations indicate the presence of competition for resources [290,330] and subniche differentiation [330]. This suggests that *Corynebacteriaceae* and *Staphylococcaceae* might be the main competitors against the other three predominant families.

5.4.4 Community state types of the penile microbiotas

A cluster analysis of the community profile revealed that the penile bacterial communities of the Black South men could be represented by six distinct community state types (CSTs) with different bacterial diversities. Over 50% of the men had *Corynebacterium*-dominated penile microbiotas (CST-1). Six of the 238 men (2.5%) had penile microbiotas dominated by *Lactobacillus* (CST-6). The remaining 44.1% of the men (CSTs 2-5) had diverse communities, neither dominated by *Corynebacterium* nor

Lactobacillus. CSTs 2-5 had significantly lower relative abundances of *Corynebacterium* and *Staphylococcus* and higher relative abundances of several taxa, particularly BV-associated bacteria compared to CST-1 (Supplemental Table 5.2). Comparisons of the alpha diversities of the communities in the CSTs indicated that CST-1 and CST-6 were significantly less diverse than CSTs 2-5. Comparisons of beta diversities of the penile microbiotas revealed that all the six CSTs were unique, with CST-1 having the lowest beta diversity.

The penile skin microbiota study by Zozaya and colleagues (2016) [110] did not establish CSTs, precluding any comparisons with our results. A study by Liu and co-workers (2015) [194] established seven CSTs (designated 1-7) with different bacterial densities in the coronal sulci penile microbiotas of 165 uncircumcised HIV-negative. CST1-3 had lower total bacterial load and higher bacterial densities of *Corynebacterium*, *Lactobacillus*, and *Staphylococcus* than CST4-7. CST4-7 had higher bacterial densities of BV-associated bacteria such as *Prevotella*, *Porphyromonas*, and *Peptoniphilus* [194]. The reduced relative abundances of *Corynebacterium*, *Lactobacillus*, and *Staphylococcus*, and predominance of BV-associated bacteria in the diverse microbiotas (CSTs 2-5) in our study is congruent with CST4-7 in the study by Liu and co-workers (2015) [194]. The prevalence of these diverse penile microbiotas (CSTs 2-5) was, however, 5.3% more than that reported for CST4-7 (38.8%) [194]. To date, our research remains the only study that has established distinct penile CSTs from multiple penile sites.

5.4.5 Penile microbiota community state types and HPV and HIV infections

This study found a positive association of non-*Corynebacterium*-dominated microbiotas (CSTs 2-6) with high-risk and multiple HPV infections. Men with *Corynebacterium*-dominated penile skin microbiotas (CST-1) in turn had less high-risk HPV and less multiple HPV infections. While no previous studies have been carried out to examine associations of penile microbiotas with HPV, *Corynebacterium* is known to be abundant in post-circumcised penile swabs [110,254] and circumcision is associated with protection against penile HPV infection [234,236,240,241,262,364]. One trial of male circumcision for HIV/STI prevention associated circumcision with reduced prevalent HPV in penile shaft and coronal sulci [364]. The reduction in prevalent HPV following circumcision is more pronounced at the glans/corona than sites more distal to the prepuce [241]. Although there is weak to no evidence of association of male circumcision with reduced HPV incidence and increased remission [240,241], a randomised trial associated male circumcision with low incidence of multiple high-risk HPV, and rapid clearance of high-risk HPV infections [262]. A large observational study on a Hawaiian cohort found a significantly lower prevalence of glans/corona HPV in circumcised men (29%) than in uncircumcised men (46%) [239]. The biological mechanism for this protective effect is unknown. The association between non-*Corynebacterium*-dominated microbiotas and high-risk HPV could be due to either an increased susceptibility to high-risk HPV genotypes or a delayed remission of these genotypes by such CSTs (diverse CSTs). In women diverse BV-like communities have been described to have the lowest clearance rates [34]. *C. accolens* has been found in greater relative abundances in HPV-negative women compared to HPV-positive women [141]. Thus, penile CSTs with lower relative abundances of *Corynebacterium* and higher relative abundances of specific bacteria, including BV-associated bacteria, should be investigated as a risk factor for HPV/high-risk HPV infection.

Among the diverse CSTs, CST-5 (*Corynebacterium*-deficient with higher relative abundances of *Prevotella*, unclassified *Clostridiales* and *Porphyromonas*) was associated with HPV and high-risk HPV infections when compared to CST-1. In women, diverse CVMs are unstable and less resilient to subtle microbiome perturbations [54,135,162] and have been associated with incident, prevalent, and persistent HPV [33,34,141,214], and genital intraepithelial lesions [140]. The different CSTs of women have been found to have different roles in HPV etiopathogenesis [33,34]. The diverse and dysbiotic CVMs have been associated with epithelial damage of the female genital tract [122]. Cellular damage serves as an avenue for HPV to access the basal lamina and form a reservoir of infection [4]. In men therefore, future studies should investigate if CSTs 2-5, which are diverse and appear dysbiotic, also potentially cause epithelial damage and increase the risk of HPV infection.

There was no significant association between non-*Corynebacterium*-dominated or *Corynebacterium*-dominated CSTs and HIV. There was, however, a trend towards increased HIV infection in men in CST-5 compared to men in CST-1. HIV and HPV coinfected men, furthermore, were found to have significantly more diverse penile microbiotas than HIV-uninfected men with or without HPV infection. There are no published studies that have investigated the penile microbiome of men who are already infected with HIV. It is now recognised that men with such diverse CSTs may be predisposed to STIs, such as HIV [35,232]. A recent prospective study by Liu and colleagues (2017) [35] on penile microbiotas of 182 uncircumcised Ugandan men showed that penile anaerobic dysbiosis is a risk factor for acquisition of HIV. The authors demonstrated that the men who seroconverted (25.3%) had significantly higher abundances of anaerobic *Prevotella*, *Dialister*, *Peptoniphilus*, and *Finegoldia* at baseline than men who remained HIV-negative [35]. Moreover, these bacteria were found to trigger proinflammatory responses that are thought to facilitate infection by HIV [35]; hence, further supporting the strong connection between certain genital anaerobic bacteria and inflammatory cytokines [136,164]. In our study, CST-5 had reduced relative abundances of *Corynebacterium* (mean: 12.9%) and was dominated by several BV-associated anaerobic bacteria and/or proinflammatory bacteria such as *Prevotella* (23.0%), unclassified *Clostridiales* (18.5%), and *Porphyromonas* (16.3%). In addition, it had other bacteria in appreciable relative abundances: *Negativicoccus* (4.8%), *Dialister* (2.6%), *Finegoldia* (2.2%), *Sneathia* (1.9%), *Gardnerella* (1.5%), and *Fusobacterium* (1.1%). High concentrations of BV-associated bacteria such as *Parvimonas*, *M. hominis*, *Sneathia*/*Leptotrichia*, and *Megasphaera* have been associated with increased risk of HIV acquisition in African women [36]. BV-associated bacteria were found to be more prevalent and abundant in diverse CSTs such as CST-5 than in low-diversity CSTs, specifically CST-1 (with *Corynebacterium* dominance) and CST-6 (with *Lactobacillus* dominance). This observation has also been reported in a penile microbiota study on a Ugandan cohort [194]. Penile anaerobic dysbiosis and/or the predominance of proinflammatory associated bacteria in CST-5 may potentially contribute to the greater prevalence of multiple HPV, high-risk HPV, and HIV infections in men in CST-5.

5.4.6 Potential biomarkers for HPV and HIV infections in men

In the analysis of the taxa in the penile microbiotas associated with HPV infections, post-circumcision associated families *Pseudomonadaceae* and *Corynebacteriaceae* [110,229] were the most significant potential biomarkers for absence of penile HPV and high-risk HPV infection, respectively. The specific genera in each of these families were *Pseudomonas* (for *Pseudomonadaceae*) and *Corynebacterium* (for *Corynebacteriaceae*). HPV-positive men had higher relative abundances of BV-associated bacteria, including *Sneathia*, *Porphyromonas*, *Prevotella*, *Peptoniphilus*, *Dialister*, and other bacteria such as *Jonquetella* and *Cetobacterium* than HPV-negative men. In men with high-risk HPV infections, *Prevotella*, *Peptoniphilus, Dialister*, and *Campylobacter* (LDA score <-3.0, $p<0.05$) were more abundant than in men without high-risk HPV.

To our knowledge, no studies have investigated the associations between penile bacterial taxa and HPV infection. The observations of *Sneathia*, *Prevotella*, *Dialister*, and *Peptoniphilus* as potential biomarkers for either HPV or high-risk HPV are in agreement with CVM studies [137,157,179] that identified these taxa as potential biomarkers for HPV in women. Similar to findings regarding CVMs in Chapter 4 subsection 4.3.9, *Pseudomonas* exhibited a higher relative abundance in HPV-negative men than in HPV-positive HIV-uninfected men.

In men with HIV infections, genera *Staphylococcus*, *Faecalibacterium*, *Stenotrophomonas*, *Jonquetella*, and *Ruminococcus* were the strongest possible biomarkers for HIV infection. Although we did not identify the bacteria to species level, several members of the identified genera have been associated with HIV infection. A pathogenic strain of *S. aureus* has been found to be more common in HIV-positive than in HIV-negative men [232]. *Stenotrophomonas maltophilia*, an opportunistic and multiantibiotic resistant pathogen that is capable of forming biofilms, is increasingly recognised in immunocompromised patients (e.g., those with chronic HIV infection and AIDS) and has been strongly associated with high fatality/case ratio [365]. *Stenotrophomonas* has been recently reported in very low abundances in penile microbiotas of HIV-negative men [256]. We observed that men without HIV infection were highly enriched with *Propionibacterium* and *Nosocomiicoccus*.

Some of the potential biomarkers that we detected are understudied taxa (e.g., in phyla Synergistia). Examples of these taxa include, *Jonquetella* and *Pyramidobacter* in family *Synergistaceae*. These were possible biomarkers for HPV or high-risk HPV, and HIV infections (Fig. 5.7 and Supplemental Fig. 5.4). Despite their inadequate characterisation [366], Synergistes/Synergistetes, which includes members of *Synergistaceae*, have been implicated in periodontitis and wound infections [366,367].

5.4.7 Penile microbiotas and female sexual partners' cervicovaginal microbiotas and bacterial vaginosis status

Following the emerging evidence suggesting that BV-associated bacteria may be sexually transmitted [110,193,194,257], even by circumcised men [194,260], we assessed the predominance of BV-associated bacteria in the penile microbiotas. In line with previous literature [35,194,254,256], several BV-associated bacteria were present in the penile microbiotas. These were further found to be significantly enriched in men with diverse penile microbiotas (CSTs 2-5). These included *Finegoldia*, *Atopobium*, *Gardnerella*, *Sneathia*, *Shuttleworthia*, *Mycoplasma*, *Mobiluncus*, *Porphyromonas*, and *Prevotella*, most of which have been associated with Nugent-BV score in female sexual partners [194]. This finding indicates that BV-associated bacteria are not only confined to the female genital tract, but can inhabit the human penis too. Our result mirrors previous reports that have found that the penile skin can be colonised by BV-associated bacteria [110,193,194,256,257].

Next, associations of the penile microbiotas with female sexual partners' cervicovaginal CSTs and BV status were examined. Men with highly diverse CSTs (2-5), with a deficiency of *Corynebacterium*/*Lactobacillus*, were more likely to have female sexual partners with BV-associated CVMs, .i.e., CVMs not dominated by *Lactobacillus*. The reverse was true for men with *Corynebacterium*/*Lactobacillus*-dominated microbiotas. Compared to men with *Corynebacterium*-dominated microbiotas (CST-1), men with diverse microbiotas had a trend towards having female sexual partners with *Lactobacillus*- or *L. iners*-deficient and diverse CVMs. CSTs with significantly reduced populations of *Corynebacterium* and *Lactobacillus* have been associated with Nugent-BV in female partners [194]. We noted that all the men with *Lactobacillus*-dominated microbiotas (CST-6) had female partners with *Lactobacillus* or *L. iners*-dominated CVMs. Our observations suggest existence of complementarity between penile and cervicovaginal CSTs. Recently, it has been shown that a decrease in prevalence and abundance of BV-associated bacteria (synonymous with a decrease in microbiota diversity) in men leads to a decrease in cervicovaginal microbiota diversity in their female sexual partner [256].

In spite of the association between penile and cervicovaginal CSTs, there was no association of *Corynebacterium*-dominated (CST-1) or diverse penile microbiotas (CST 2-5), with female sexual partners' BV status. This is in contrast to a previous study on a Ugandan cohort [194] that found that highly diverse coronal sulci microbiotas (pooled CST4-7 in their case) were associated with female partner Nugent-BV.

Specific penile bacteria were, however, found to be associated with the presence or absence of BV in the female sexual partner. Similar to other investigators [194], we observed that higher relative abundance of *Lactobacillus* in the penile microbiotas of men was significantly associated with BV-negative status in the female sexual partners. *Lactobacillus* is a protective resident bacterium in the CVM [33,34,146]; thus, penile *Lactobacillus* is highly expected to be a normal Nugent score indicator [194]. *Lactobacillus* has been identified as one of major genera shared among heterosexual couples [106]. Therefore, penile colonisation by *Lactobacillus* may be driven by unprotected sexual intercourse with women without BV [105,110,194]. In contrast to Liu and colleagues (2015) [194], *Corynebacterium* and *Staphylococcus* were not found in our study to be potential biomarkers for BV-negative status in the female sexual partner and occurred in similar relative abundances in men in BV-positive and BV-negative relationships.

Our study also observed an association of prevalence of specific BV-associated bacteria in the penile microbiotas in men in BV-positive relationships. The prevalences of *Atopobium*, *Sneathia*, and *Saccharofermentans* in the penile microbiotas were significantly positively associated with BV in the female sexual partner. Penile *Sneathia* and *Saccharofermentans* have been previously associated with female partner Nugent-BV score [194], while the association of partners' BV with the prevalence *Atopobium* has not been reported elsewhere. The difference in relative abundances of the BV-associated bacteria (including *Atopobium*, *Dialister*, *Gardnerella*, *Mobiluncus*, *Porphyromonas*, *Prevotella*, *Sneathia*, *Treponema*, and unclassified *Clostridiales*) in men with BV-negative versus BV-positive partner did not reach statistical significance. Our present study identified penile bacteria that have not been associated with BV to be more abundant in men in BV-positive relationships than in men with BV-negative relationships. These include *Illumatobacter*, *Providencia*, *Brevundimonas*, *Sphingobacterium*, and *Comamonas*. Some of these penile bacteria included genera whose members have been associated with clinical diseases. For example, *Brevundimonas diminuta* can cause urinary infections in cancer patients [368] while *Citrobacter diversus* can cause sepsis and meningitis in neonates and immunocompromised individuals [369]. Therefore, there is need to characterise the species of the penile microbiota.

Future longitudinal and large-scale studies on heterosexual couple microbiotas are also needed in order to find associations possibly missed in cross-sectional [194,370] and relatively small cohort longitudinal analyses [371]. Such temporal studies should examine the resilience and ecological permissiveness of penile microbiotas to BV-associated bacteria with respect to male circumcision status, unprotected sexual intercourse, STIs/HIV, and female partner's BV status among other variables.

5.4.8 Summary

To date, this is the largest cross-sectional study using a culture-independent approach to detail the bacterial communities of the penile skin. This is also the first study to characterise the penile microbiotas of traditionally circumcised men and to examine the associations of these microbiotas with HPV infection. We found that family *Corynebacteriaceae* and genus *Corynebacterium* were the most abundant bacterial taxa in the penile microbiotas of Black South African men. Six CSTs were established, with over 50% of the men having *Corynebacterium*-dominated penile microbiotas.

Corynebacterium-deficient penile microbiotas, mostly dominated by BV-associated bacteria, were associated with prevalent high-risk and multiple HPV infections in Black South African men. The relative abundances of specific penile bacteria were identified as possible biomarkers for HPV (the most significant being *Prevotella*, *Porphyromonas*, *Dialister*, and *Sneathia*), high-risk HPV (*Prevotella*, *Dialister*, *Peptoniphilus*, and *Campylobacter*), and HIV infection (*Staphylococcus*, *Faecalibacterium*, *Stenotrophomonas*, and *Jonquetella*). These potential biomarkers may have both diagnostic and therapeutic use, but first, the specific species or strains involved, the molecular mechanisms of these potential biomarkers in pathogenesis of HPV infection, and their role in urogenital infections should be investigated.

Among the possible biomarkers for penile HPV infection identified in this study, *Sneathia* in particular should be studied in detail. *Sneathia* was identified as a strong possible biomarker for prevalent HPV infection among the Black South African women and men (chapters 4 and 5). Furthermore, other published studies on Nigerian [137], Italian [157], and

Korean [179] cohorts have identified *Sneathia* as a possible biomarker for HPV/high-risk HPV infection. Several studies have demonstrated the pathogenic potential of specific genital *Sneathia* species and point to the possible mechanisms by which it may play a role in HPV acquisition and persistence. For example, *S. sanguinegens* has been associated with increased genital inflammation in women, i.e., elevated levels of IL-1α, IL-1β, and IL-8 [164]. *S. amnii* is also a genital bacterium, identified in women and is highly similar to *S. sanguinegens* [334]. *S. amnii* has genes that encode proteins implicated in its invasion, adherence to cervical cells and cytotoxicity [334].

In this chapter, an initial exploration of the genital microbiotas in heterosexual couples was also undertaken. In the South African cohort studied, men with highly diverse microbiotas (CSTs 2-5) were more likely to have female partners with non-*Lactobacillus*-dominated or diverse CVMs. *Lactobacillus* was more abundant in men in BV-negative relationships. The prevalences of BV-associated *Saccharofermentans*, *Atopobium*, and *Sneathia* were significantly higher in BV-positive relationships than in BV-negative relationships. The relative abundances of these taxa, and other BV-associated bacteria examined, were, however, not statistically different.

A limitation of the present study is the lack of adjustment for potential confounders, including age of men, sex acts per month, presence of other STIs, host genetic factors, host immunity, and cigarette use to name a few. In addition, BV in the female sexual partners was diagnosed using a non-standard approach (Pap smear), which has been demonstrated not to perform well when compared to Gram stain diagnosis [80,293]. Thus, it is possible that this might have affected the results on the association between penile microbiotas and female sexual partners BV status.

In summary, the findings of the chapters presented herein (chapters 2-5) have added to our understanding of the genital microbiotas of Black men and women. The associations of human genital microbiotas and specific bacterial taxa with HPV and HIV infections are hypothesis-generating and following further investigations may have implications on i) personalised microbiota-based diagnostics, and ii) probiotics that promote reproductive health and prevent disorders and diseases in this population.

References

1. Conway MJ, Meyers C: Replication and assembly of human papillomaviruses. *J. Dent. Res.* 2009, **88**:307-317.
2. Fernandes JV, Galvão de Araújo J, Araújo de Fernandes TA: Biology and natural history of human papillomavirus infection. *Open Access Journal of Clinical Trials* 2013, **5**:1-12.
3. International Human Papillomavirus Reference Center: Reference clones: human papillomavirus reference clones. Karolinska Institutet; 2018. Accessed on 28 September 2018. Available from: http://www.nordicehealth.se/hpvcenter/reference_clones/.
4. Doorbar J, Quint W, Banks L, Bravo IG, Stoler M*, et al.*: The biology and life-cycle of human papillomaviruses. *Vaccine* 2012, **30 Suppl 5**:F55-70.
5. Janicek MF, Averette HE: Cervical cancer: prevention, diagnosis, and therapeutics. *CA: Cancer J. Clin.* 2001, **51**:92-114.
6. Kenyon C, Buyze J, Colebunders R: Classification of incidence and prevalence of certain sexually transmitted infections by world regions. *Int. J. Infect. Dis.* 2014, **18**:73-80.
7. Van Doornum G, Prins M, Juffermans L, Hooykaas C, van den Hoek J*, et al.*: Regional distribution and incidence of human papillomavirus infections among heterosexual men and women with multiple sexual partners: a prospective study. *Genitourin. Med.* **70**:240–246.
8. Molano M: Determinants of clearance of Human Papillomavirus infections in Colombian women with normal cytology: a population-based, 5-year follow-up study. *Am. J. Epidemiol.* 2003, **158**:486-494.
9. Giuliano AR, Lee J-H, Fulp W, Villa LL, Lazcano E*, et al.*: Incidence and clearance of genital human papillomavirus infection in men (HIM): a cohort study. *Lancet* 2011, **377**:932-940.
10. Cutts FT, Franceschi S, Goldie S, Castellsague X, Sanjose SD*, et al.*: Human papillomavirus and HPV vaccines: A review. *Bull. World Health Organ.* 2007, **038414**:719-726.
11. Stanley M: Pathology and epidemiology of HPV infection in females. *Gynecol. Oncol.* 2010, **117**:S5-10.
12. Nobbenhuis MAE, Walboomers JMM, Helmerhorst TJM, Rozendaal L, Remmink AJ: Relation of human papillomavirus status to cervical lesions and consequences for cervical-cancer screening: a prospective study. *Lancet* 1999, **354**:20-25.
13. Walboomers JMM, Jacobs MV, Michele MM, Bosch FX, Kummer JA*, et al.*: Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J. Pathol.* 1999, **189**:12-19.
14. Husein-El Ahmed H, Canadas-De la Fuente GA: Squamous-cell carcinoma of the penis with human papillomavirus. *N. Engl. J. Med.* 2016, **374**:164.
15. Do HTT, Koriyama C, Khan Na, Higashi M, Kato T*, et al.*: The etiologic role of human papillomavirus in penile cancers: a study in Vietnam. *Br. J. Cancer* 2013, **108**:229-233.
16. Bruni L, Barrionuevo-Rosas L, Albero G, Serrano B, Mena M*, et al.*: Human papillomavirus and related diseases in the world. Summary report 19 April 2017. ICO Information Centre on HPV and Cancer (HPV Information Centre); 2017.
17. Bruni L, Barrionuevo-Rosas L, Albero G, Serrano B, Mena M*, et al.*: Human papillomavirus and related diseases in the world. Summary report 17 April 2018. ICO Information Centre on HPV and Cancer (HPV Information Centre); 2018.
18. Statistics South Africa: Mid-year population estimates 2018. In *Statistical Release P0302*. 2018. Accessed on 23 July 2018. Available from: http://www.statssa.gov.za/publications/P0302/P03022018.pdf.
19. Mbulawa ZZA, Marais DJ, Johnson LF, Coetzee D, Williamson A-L: Impact of human immunodeficiency virus on the natural history of human papillomavirus genital infection in South African men and women. *J. Infect. Dis.* 2012, **206**:15-27.
20. Mbulawa ZZ, Coetzee D, Williamson AL: Human papillomavirus prevalence in South African women and men according to age and human immunodeficiency virus status. *BMC Infect. Dis.* 2015, **15**:459.
21. Mbulawa ZZ, Johnson LF, Marais DJ, Coetzee D, Williamson AL: The impact of human immunodeficiency virus on human papillomavirus transmission in heterosexually active couples. *J. Infect.* 2013, **67**:51-58.
22. Remschmidt C, Kaufmann AM, Hagemann I, Vartazarova E, Wichmann O*, et al.*: Risk factors for cervical human papillomavirus infection and high-grade intraepithelial lesion in women aged 20 to 31 years in Germany. *Int. J. Gynecol. Cancer* 2013, **23**:519-526.

23. Mbulawa ZZ, Marais DJ, Johnson LF, Boulle A, Coetzee D, *et al.*: Influence of human immunodeficiency virus and CD4 count on the prevalence of human papillomavirus in heterosexual couples. *J. Gen. Virol.* 2010, **91**:3023-3031.
24. Masson L, Passmore JA, Liebenberg LJ, Werner L, Baxter C, *et al.*: Genital inflammation and the risk of HIV acquisition in women. *Clin. Infect. Dis.* 2015, **61**:260-269.
25. Selhorst P, Masson L, Ismail SD, Samsunder N, Garrett N, *et al.*: Cervicovaginal inflammation facilitates acquisition of less infectious HIV variants. *Clin. Infect. Dis.* 2017, **64**:79-82.
26. Coloccini RS, Dilernia D, Ghiglione Y, Turk G, Laufer N, *et al.*: Host genetic factors associated with symptomatic primary HIV infection and disease progression among Argentinean seroconverters. *PLoS ONE* 2014, **9**:e113146.
27. Wu H, Zhang J: miR-124 rs531564 polymorphism influences genetic susceptibility to cervical cancer. *Int. J. Clin. Exp. Med.* 2004, **7**:5847–5851.
28. Kiwanuka N, Laeyendecker O, Quinn TC, Wawer MJ, Shepherd J, *et al.*: HIV-1 subtypes and differences in heterosexual HIV transmission among HIV-discordant couples in Rakai, Uganda. *AIDS* 2009, **23**:2479-2484.
29. Lai CH, Chao A, Chang CJ, Chao FY, Huang HJ, *et al.*: Host and viral factors in relation to clearance of human papillomavirus infection: a cohort study in Taiwan. *Int. J. Cancer* 2008, **123**:1685-1692.
30. Kim HS, Kim TJ, Lee IH, Hong SR: Associations between sexually transmitted infections, high-risk human papillomavirus infection, and abnormal cervical Pap smear results in OB/GYN outpatients. *J. Gynecol. Oncol.* 2016, **27**:e49.
31. Fleming DT, Wasserheit J: From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection. *Sex. Transm. Infect.* 1999, **75**:3–17.
32. Silva J, Cerqueira F, Medeiros R: *Chlamydia trachomatis* infection: implications for HPV status and cervical cancer. *Arch. Gynecol. Obstet.* 2014, **289**:715-723.
33. Borgdorff H, Tsivtsivadze E, Verhelst R, Marzorati M, Jurriaans S, *et al.*: *Lactobacillus*-dominated cervicovaginal microbiota associated with reduced HIV/STI prevalence and genital HIV viral load in African women. *ISME J.* 2014:1-13.
34. Brotman RM, Shardell MD, Gajer P, Tracy JK, Zenilman JM, *et al.*: Interplay between the temporal dynamics of the vaginal microbiota and human papillomavirus detection. *J. Infect. Dis.* 2014, **210**:1723-1733.
35. Liu CM, Prodger JL, Tobian AAR, Abraham AG, Kigozi G, *et al.*: Penile anaerobic dysbiosis as a risk factor for HIV infection. *mBio* 2017, **8**.
36. McClelland RS, Lingappa JR, Srinivasan S, Kinuthia J, John-Stewart GC, *et al.*: Evaluation of the association between the concentrations of key vaginal bacteria and the increased risk of HIV acquisition in African women from five cohorts: a nested case-control study. *Lancet Infect. Dis.* 2018.
37. The Human Microbiome Project Consortium: Structure, function and diversity of the healthy human microbiome. *Nature* 2012, **486**:207-214.
38. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, *et al.*: Vaginal microbiome of reproductive-age women. *Proc. Natl. Acad. Sci. USA* 2011, **108**:4680-4687.
39. Fettweis JM, Brooks JP, Serrano MG, Sheth NU, Girerd PH, *et al.*: Differences in vaginal microbiome in African American women versus women of European ancestry. *Microbiology* 2014, **160**:2272-2282.
40. Zhou X, Brown CJ, Abdo Z, Davis CC, Hansmann MA, *et al.*: Differences in the composition of vaginal microbial communities found in healthy Caucasian and Black women. *ISME J.* 2007, **1**:121-133.
41. Marchesi JR, Ravel J: The vocabulary of microbiome research: a proposal. *Microbiome* 2015, **3**:31.
42. Damelin LH, Paximadis M, Mavri-Damelin D, Birkhead M, Lewis Da, *et al.*: Identification of predominant culturable vaginal *Lactobacillus* species and associated bacteriophages from women with and without vaginal discharge syndrome in South Africa. *J. Med. Microbiol.* 2011, **60**:180-183.
43. Mehdinejad M, Khosravi AD, Yazdizadeh H, Afshari P: Bacteriological study of vaginal discharge of pregnant women using Gram stain smear and culture. *Afr. J. Microbiol. Res.* 2011, **5**:1994-1998.
44. Krohn MA, Hillier SL, Eschenbach DA: Comparison of methods for diagnosing bacterial vaginosis among pregnant women. *J. Clin. Microbiol.* 1989, **27**:1266-1271.

45. Pande S, Kost C: Bacterial unculturability and the formation of intercellular metabolic networks. *Trends Microbiol.* 2017, **25**:349-361.
46. Vartoukian SR, Palmer RM, Wade WG: Strategies for culture of 'unculturable' bacteria. *FEMS Microbiol. Lett.* 2010, **309**:1-7.
47. Yarza P, Yilmaz P, Pruesse E, Glockner FO, Ludwig W*, et al.*: Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* 2014, **12**:635-645.
48. Baker GC, Smith JJ, Cowan DA: Review and re-analysis of domain-specific 16S primers. *J. Microbiol. Methods* 2003, **55**:541-555.
49. Claesson MJ, Wang Q, O'Sullivan O, Greene-Diniz R, Cole JR*, et al.*: Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res.* 2010, **38**:e200-e200.
50. Shah N, Tang H, Doak TG, Ye Y: Comparing bacterial communities inferred from 16s rRNA gene sequencing and shotgun metagenomics. *Pac. Symp. Biocomput.* 2011:165-176.
51. Srinivasan S, Hoffman NG, Morgan MT, Matsen FA, Fiedler TL*, et al.*: Bacterial communities in women with bacterial vaginosis: high resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. *PLoS ONE* 2012, **7**:e37818-e37818.
52. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ*, et al.*: An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 2012, **6**:610-618.
53. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL*, et al.*: Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 2006, **72**:5069-5072.
54. Smith BC, McAndrew T, Chen Z, Harari A, Barris DM*, et al.*: The cervical microbiome over 7 years and a comparison of methodologies for its characterization. *PLoS ONE* 2012, **7**:e40425-e40425.
55. Hill JE, Penny SL, Crowell KG, Goh SH, Hemmingsen SM: cpnDB: a chaperonin sequence database. *Genome Res.* 2004, **14**:1669-1675.
56. Hill JE, Town JR, Hemmingsen SM: Improved template representation in *cpn60* polymerase chain reaction (PCR) product libraries generated from complex templates by application of a specific mixture of PCR primers. *Environ. Microbiol.* 2006, **8**:741-746.
57. Anderson DJ, Marathe J, Pudney J: The structure of the human vaginal stratum corneum and its role in immune defense. *Am. J. Reprod. Immunol.* 2014, **71**:618-623.
58. Farage M, Maibach H: Lifetime changes in the vulva and vagina. *Arch. Gynecol. Obstet.* 2006, **273**:195-202.
59. Turyk M, Golub T, Wood N, Hawkins J, Wilbanks G: Growth and characterization of epithelial cells from normal human uterine ectocervix and endocervix. *In Vitro Cell Dev. Biol.* 1986, **25**:544-556.
60. Spear GT, French AL, Gilbert D, Zariffard MR, Mirmonsef P*, et al.*: Human α-amylase present in lower-genital-tract mucosal fluid processes glycogen to support vaginal colonization by *Lactobacillus*. *J. Infect. Dis.* 2014, **210**:1019-1028.
61. Mirmonsef P, Hotton AL, Gilbert D, Burgad D, Landay A*, et al.*: Free glycogen in vaginal fluids is associated with *Lactobacillus* colonization and low vaginal pH. *PLoS ONE* 2014, **9**:e102467.
62. Mirmonsef P, Modur S, Burgad D, Gilbert D, Golub ET*, et al.*: Exploratory comparison of vaginal glycogen and *Lactobacillus* levels in premenopausal and postmenopausal women. *Menopause* 2015, **22**:702-709.
63. Mirmonsef P, Hotton AL, Gilbert D, Gioia CJ, Maric D*, et al.*: Glycogen levels in undiluted genital fluid and their relationship to vaginal pH, estrogen, and progesterone. *PLoS ONE* 2016, **11**:e0153553.
64. Adeva-Andany MM, Gonzalez-Lucan M, Donapetry-Garcia C, Fernandez-Fernandez C, Ameneiros-Rodriguez E: Glycogen metabolism in humans. *BBA Clin.* 2016, **5**:85-100.
65. Ginkel PD, Soper DE, Bump RC, Dalton HP: Vaginal flora in postmenopausal women: the effect of estrogen replacement. *Infect. Dis. Obstet. Gynecol.* 1993, **1**:94-97.
66. Miller L, Patton DL, Meier A, Thwin SS, Hooton TM*, et al.*: Depomedroxyprogesterone-induced hypoestrogenism and changes in vaginal flora and epithelium. *Obstet. Gynecol.* 2000, **96**:431-439.
67. Boskey ER, Cone RA, Whaley KJ, Moench TR: Origins of vaginal acidity: high D/L lactate ratio is consistent with bacteria being the primary source. *Hum. Reprod.* 2001, **16**:1809-1813.
68. Boskey ER, Telsch KM, Whaley KJ, Moench TR, Cone RA: Acid production by vaginal flora *in vitro* is consistent with the rate and extent of vaginal acidification acid production by vaginal

flora *in vitro* is consistent with the rate and extent of vaginal acidification. *Infect. Immun.* 1999, **67**:5170-5175.

69. Boris S, Barbés C: Role played by lactobacilli in controlling the population of vaginal pathogens. *Microbes Infect.* 2000, **2**:543-546.
70. Clarke MA, Rodriguez AC, Gage JC, Herrero R, Hildesheim A*, et al.*: A large, population-based study of age-related associations between vaginal pH and human papillomavirus infection. *BMC Infect. Dis.* 2012, **12**:33.
71. O'Hanlon DE, Moench TR, Cone RA: In vaginal fluid, bacteria associated with bacterial vaginosis can be suppressed with lactic acid but not hydrogen peroxide. *BMC Infect. Dis.* 2011, **11**:200-200.
72. Boris S, Suárez JE, Vázquez F, Barbés C: Adherence of human vaginal lactobacilli to vaginal epithelial cells and interaction with uropathogens. *Infect. Immun.* 1998, **66**:1985-1989.
73. Stevens-Simon C, Jamison J, McGregor JA, Douglas JM: Racial variation in vaginal pH among healthy sexually active adolescents. *Sex. Transm. Dis.* 1994, **21**:168-172.
74. Srinivasan S, Fredricks DN: The human vaginal bacterial biota and bacterial vaginosis. *Interdiscip. Perspect. Infect. Dis.* 2008, **2008**:750479-750479.
75. Donders GGG: Definition and classification of abnormal vaginal flora. *Best Pract. Res. Cl. Ob.* 2007, **21**:355-373.
76. Kenyon C, Colebunders R, Crucitti T: The global epidemiology of bacterial vaginosis: a systematic review. *Am. J. Obstet. Gynecol.* 2013, **209**:505-523.
77. Castro J, Henriques A, Machado A, Henriques M, Jefferson KK*, et al.*: Reciprocal interference between *Lactobacillus* spp. and *Gardnerella vaginalis* on initial adherence to epithelial cells. *Int. J. Med. Sci.* 2013, **10**:1193-1198.
78. Ma B, Forney LJ, Ravel J: Vaginal microbiome: rethinking health and disease. *Annu. Rev. Microbiol.* 2012, **66**:371-389.
79. Sha BE, Chen HY, Wang QJ, Zariffard MR, Cohen MH*, et al.*: Utility of amsel criteria, nugent score, and quantitative PCR for *Gardnerella vaginalis*, *Mycoplasma hominis*, and *Lactobacillus* spp. for diagnosis of bacterial vaginosis in human immunodeficiency virus-infected women. *J. Clin. Microbiol.* 2005, **43**:4607-4612.
80. Tokyol C, Aktepe OC, Cevrioglu AS, Altindis M, Dilek FH: Bacterial vaginosis: comparison of Pap smear and microbiological test results. *Mod. Pathol.* 2004, **17**:857-860.
81. Gillet E, Meys JFA, Verstraelen H, Bosire C, De Sutter P*, et al.*: Bacterial vaginosis is associated with uterine cervical human papillomavirus infection: a meta-analysis. *BMC Infect. Dis.* 2011, **11**:10-10.
82. Zozaya-Hinchliffe M, Lillis R, Martin DH, Ferris MJ: Quantitative PCR assessments of bacterial species in women with and without bacterial vaginosis. *J. Clin. Microbiol.* 2010, **48**:1812-1819.
83. Donders GG, Gonzaga A, Marconi C, Donders F, Michiels T*, et al.*: Increased vaginal pH in Ugandan women: what does it indicate? *Eur. J. Clin. Microbiol. Infect. Dis.* 2016, **35**:1297-1303.
84. Abbai NS, Reddy T, Ramjee G: Prevalent bacterial vaginosis infection - a risk factor for incident sexually transmitted infections in women in Durban, South Africa. *Int. J. STD AIDS* 2016, **27**:1283-1288.
85. Masson L, Arnold KB, Little F, Mlisana K, Lewis DA*, et al.*: Inflammatory cytokine biomarkers to identify women with asymptomatic sexually transmitted infections and bacterial vaginosis who are at high risk of HIV infection. *Sex. Transm. Infect.* 2016, **92**:186-193.
86. Puran AC, Adler D, Wallace M, Bennie T, Phuti A*, et al.*: Incidental findings of bacterial vaginosis and other infections in Papanicolaou smears of HIV-infected and HIV-uninfected adolescent females in South Africa. *J. AIDS HIV Res.* 2014, **6**:172-176.
87. Lennard K, Dabee S, Barnabas SL, Havyarimana E, Blakney A*, et al.*: Microbial composition predicts genital tract inflammation and persistent bacterial vaginosis in adolescent South African women. *Infect. Immun.* 2018.
88. Donders GGG, Bellen G, Grinceviciene S, Ruban K, Vieira-Baptista P: Aerobic vaginitis: no longer a stranger. *Res. Microbiol.* 2017, **168**:845-858.
89. Koumans EH, Markowitz LE, Hogan V, CDC BV Working Group: Indications for therapy and treatment recommendations for bacterial vaginosis in nonpregnant and pregnant women: a synthesis of data. *Clin. Infect. Dis.* 2002, **35**:S152-172.
90. Workowski KA, Bolan GA, Centers for Disease Control and Prevention: Sexually transmitted diseases treatment guidelines, 2015. *MMWR Recomm. Rep.* 2015, **64**:1-137.

91. Austin MN, Beigi RH, Meyn LA, Hillier SL: Microbiologic response to treatment of bacterial vaginosis with topical clindamycin or metronidazole. *J. Clin. Microbiol.* 2005, **43**:4492-4497.
92. Bradshaw CS, Morton AN, Hocking J, Garland SM, Morris MB, *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**:1478–1486.
93. Ravel J, Brotman RM, Gajer P, Ma B, Nandy M, *et al.*: Daily temporal dynamics of vaginal microbiota before, during and after episodes of bacterial vaginosis. *Microbiome* 2013, **1**:29-29.
94. Schwebke JR: Asymptomatic bacterial vaginosis: response to therapy. *Am. J. Obstet. Gynecol.* 2000, **183**:1434-1439.
95. 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**:78-83.
96. Chehoud C, Stieh DJ, Bailey AG, Laughlin AL, Allen SA, *et al.*: Associations of the vaginal microbiota with HIV infection, bacterial vaginosis, and demographic factors. *AIDS* 2017, **31**:895-904.
97. Jahic M, Mulavdic M, Nurkic J, Jahic E, Nurkic M: Clinical characteristics of aerobic vaginitis and its association to vaginal candidiasis, trichomonas vaginitis and bacterial vaginosis. *Med. Arch.* 2013, **67**:428-430.
98. Bradshaw CS, Walker SM, Vodstrcil LA, Bilardi JE, Law M, *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**:1562-1572.
99. 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**:1426-1435.
100. Bradshaw CS, Vodstrcil LA, Hocking JS, Law M, Pirotta M, *et al.*: Recurrence of bacterial vaginosis is significantly associated with posttreatment sexual activities and hormonal contraceptive use. *Clin. Infect. Dis.* 2013, **56**:777-786.
101. Thoma ME, Klebanoff MA, Rovner AJ, Nansel TR, Neggers Y, *et al.*: Bacterial vaginosis is associated with variation in dietary indices. *J. Nutr.* 2011, **141**:1698-1704.
102. Antonio MA, Rabe LK, Hillier SL: Colonization of the rectum by *Lactobacillus* species and decreased risk of bacterial vaginosis. *J. Infect. Dis.* 2005, **192**:394–398.
103. Campisciano G, Zanotta N, Licastro D, De Seta F, Comar M: *In vivo* microbiome and associated immune markers: new insights into the pathogenesis of vaginal dysbiosis. *Sci. Rep.* 2018, **8**:2307.
104. Castro J, Alves P, Sousa C, Cereija T, Franca A, *et al.*: Using an *in-vitro* biofilm model to assess the virulence potential of bacterial vaginosis or non-bacterial vaginosis *Gardnerella vaginalis* isolates. *Sci. Rep.* 2015, **5**:11640.
105. Borovkova N, Korrovits P, Ausmees K, Türk S, Jõers K, *et al.*: Influence of sexual intercourse on genital tract microbiota in infertile couples. *Anaerobe* 2011, **17**:414-418.
106. Mändar R, Punab M, Borovkova N, Lapp E, Kiiker R, *et al.*: Complementary seminovaginal microbiome in couples. *Res. Microbiol.* 2015, **166**:440-447.
107. Vodstrcil LA, Twin J, Garland SM, Fairley CK, Hocking JS, *et al.*: The influence of sexual activity on the vaginal microbiota and *Gardnerella vaginalis* clade diversity in young women. *PLoS ONE* 2017, **12**:e0171856.
108. Mitchell CM, Fredricks DN, Winer RL, Laura K: Effect of sexual debut on vaginal microbiota in a cohort of young women. *Obstet. Gynecol.* 2013, **120**:1306-1313.
109. Marrazzo JM, Fiedler TL, Srinivasan S, Thomas KK, Liu C, *et al.*: Extravaginal reservoirs of vaginal bacteria as risk factors for incident bacterial vaginosis. *J. Infect. Dis.* 2012, **205**:1580-1588.
110. Zozaya M, Ferris MJ, Siren JD, Lillis R, Myers L, *et al.*: Bacterial communities in penile skin, male urethra, and vaginas of heterosexual couples with and without bacterial vaginosis. *Microbiome* 2016, **4**:16.
111. Verstraelen H, Verhelst R, Vaneechoutte M, Temmerman M: The epidemiology of bacterial vaginosis in relation to sexual behaviour. *BMC Infect. Dis.* 2010, **10**:81.
112. Verstraelen H, Swidsinski A: The biofilm in bacterial vaginosis: implications for epidemiology, diagnosis and treatment. *Curr. Opin. Infect. Dis.* 2013, **26**:86–89.
113. Flemming HC, Wingender J, Szewzyk U, Steinberg P, Rice SA, *et al.*: Biofilms: an emergent form of bacterial life. *Nat. Rev. Microbiol.* 2016, **14**:563-575.

114. Alves P, Castro J, Sousa C, Cereija TB, Cerca N: *Gardnerella vaginalis* outcompetes 29 other bacterial species isolated from patients with bacterial vaginosis, using in an *in vitro* biofilm formation model. *J. Infect. Dis.* 2014, **210**:593-596.
115. Swidsinski A, Doerffel Y, Loening-Baucke V, Swidsinski S, Verstraelen H*, et al.*: *Gardnerella* biofilm involves females and males and is transmitted sexually. *Gynecol. Obstet. Invest.* 2010, **70**:256-263.
116. Hardy L, Jespers V, Dahchour N, Mwambarangwe L, Musengamana V*, et al.*: Unravelling the bacterial vaginosis-associated biofilm: a multiplex *Gardnerella vaginalis* and *Atopobium vaginae* fluorescence in situ hybridization assay using peptide nucleic acid probes. *PLoS ONE* 2015, **10**:e0136658.
117. 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**:12004-12012.
118. Castro J, Franca A, Bradwell KR, Serrano MG, Jefferson KK*, et al.*: Comparative transcriptomic analysis of *Gardnerella vaginalis* biofilms vs. planktonic cultures using RNA-seq. *NPJ Biofilms Microbiomes* 2017, **3**:3.
119. Nelson TM, Borgogna JL, Brotman RM, Ravel J, Walk ST*, et al.*: Vaginal biogenic amines: biomarkers of bacterial vaginosis or precursors to vaginal dysbiosis? *Front. Physiol.* 2015, **6**:253.
120. van de Wijgert JHHM: The vaginal microbiome and sexually transmitted infections are interlinked: consequences for treatment and prevention. *PLoS Med.* 2017, **14**:e1002478.
121. Srinivasan S, Morgan MT, Fiedler TL, Djukovic D, Hoffman NG*, et al.*: Metabolic signatures of bacterial vaginosis. *mBio* 2015, **6**.
122. Zevin AS, Xie IY, Birse K, Arnold K, Romas L*, et al.*: Microbiome composition and function drives wound-healing impairment in the female genital tract. *PLoS Pathog.* 2016, **12**:e1005889.
123. Ferreira CST, da Silva MG, de Pontes LG, Dos Santos LD, Marconi C: Protein content of cervicovaginal fluid is altered during bacterial vaginosis. *J. Low. Genit. Tract. Dis.* 2018.
124. Giraldo PC, de Carvalho JB, do Amaral RL, da Silveira Goncalves AK, Eleuterio J, Jr.*, et al.*: Identification of immune cells by flow cytometry in vaginal lavages from women with vulvovaginitis and normal microflora. *Am. J. Reprod. Immunol.* 2012, **67**:198-205.
125. Rebbapragada A, Howe K, Wachihi C, Pettengell C, Sunderji S*, et al.*: Bacterial vaginosis in HIV-infected women induces reversible alterations in the cervical immune environment. *J. Acquir. Immune Defic. Syndr.* 2008, **49**:520-522.
126. Atashili J, Poole C, Ndumbe PM, Adimora AA, Smith JS: Bacterial vaginosis and HIV acquisition: a meta-analysis of published studies. *AIDS* 2008, **22**:1493-1501.
127. Donders GG, Vereecken A, Bosmans E, Dekeersmaecker A, Salembier G*, et al.*: Definition of a type of abnormal vaginal flora that is distinct from bacterial vaginosis: aerobic vaginitis. *BJOG* 2002, **109**:34-43.
128. Han C, Wu W, Fan A, Wang Y, Zhang H*, et al.*: Diagnostic and therapeutic advancements for aerobic vaginitis. *Arch. Gynecol. Obstet.* 2015, **291**:251-257.
129. Wang ZL, Fu LY, Xiong ZA, Qin Q, Yu TH*, et al.*: Diagnosis and microecological characteristics of aerobic vaginitis in outpatients based on preformed enzymes. *Taiwan J. Obstet. Gynecol.* 2016, **55**:40-44.
130. Donders GG, Ruban K, Bellen G: Selecting anti-microbial treatment of aerobic vaginitis. *Curr. Infect. Dis. Rep.* 2015, **17**:477.
131. Sobel JD, Reichman O, Misra D, Yoo W: Prognosis and treatment of desquamative inflammatory vaginitis. *Obstet. Gynecol.* 2011, **117**:850-855.
132. Vieira-Baptista P, Tavares S, Lima-Silva J, Beires J, Donders G: Prevalence and epidemiology of aerobic vaginitis and bacterial vaginosis in Portuguese women. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 2016, **206**:e119-e120.
133. Tansarli GS, Kostaras EK, Athanasiou S, Falagas ME: Prevalence and treatment of aerobic vaginitis among non-pregnant women: evaluation of the evidence for an underestimated clinical entity. *Eur. J. Clin. Microbiol. Infect. Dis.* 2013, **32**:977-984.
134. Tempera G, Furneri PM: Management of aerobic vaginitis. *Gynecol. Obstet. Invest.* 2010, **70**:244-249.
135. Romero R, Hassan SS, Gajer P, Tarca AL, Fadrosh DW*, et al.*: The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. *Microbiome* 2014, **2**:4-4.

136. Anahtar MN, Byrne EH, Doherty KE, Bowman BA, Yamamoto HS, *et al.*: Cervicovaginal bacteria are a major modulator of host inflammatory responses in the female genital tract. *Immunity* 2015, **42**:965-976.
137. Dareng EO, Ma B, Famooto AO, Akarolo-Anthony SN, Offiong RA, *et al.*: Prevalent high-risk HPV infection and vaginal microbiota in Nigerian women. *Epidemiol. Infect.* 2016, **144**:123-137.
138. Gautam R, Borgdorff H, Jespers V, Francis SC, Verhelst R, *et al.*: Correlates of the molecular vaginal microbiota composition of African women. *BMC Infect. Dis.* 2015, **15**:86-86.
139. Reimers LL, Mehta SD, Massad LS, Burk RD, Xie X, *et al.*: The cervicovaginal microbiota and its associations with human papillomavirus detection in HIV-infected and HIV-uninfected women. *J. Infect. Dis.* 2016, **214**:1361-1369.
140. Mitra A, MacIntyre DA, Lee YS, Smith A, Marchesi JR, *et al.*: Cervical intraepithelial neoplasia disease progression is associated with increased vaginal microbiome diversity. *Sci. Rep.* 2015, **5**:16865.
141. Shannon B, Yi TJ, Perusini S, Gajer P, Ma B, *et al.*: Association of HPV infection and clearance with cervicovaginal immunology and the vaginal microbiota. *Mucosal Immunol.* 2017.
142. Audirac-Chalifour A, Torres-Poveda K, Bahena-Roman M, Tellez-Sosa J, Martinez-Barnetche J, *et al.*: Cervical microbiome and cytokine profile at various stages of cervical cancer: a pilot study. *PLoS ONE* 2016, **11**:e0153274.
143. Borgdorff H, van der Veer C, van Houdt R, Alberts CJ, de Vries HJ, *et al.*: The association between ethnicity and vaginal microbiota composition in Amsterdam, the Netherlands. *PLoS ONE* 2017, **12**:e0181135.
144. Albert AY, Chaban B, Wagner EC, Schellenberg JJ, Links MG, *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**:e0135620.
145. Knights D, Ward TL, McKinlay CE, Miller H, Gonzalez A, *et al.*: Rethinking "enterotypes". *Cell Host Microbe* 2014, **16**:433-437.
146. Nunn KL, Wang YY, Harit D, Humphrys MS, Ma B, *et al.*: Enhanced trapping of HIV-1 by human cervicovaginal mucus is associated with *Lactobacillus crispatus*-dominant microbiota. *mBio* 2015, **6**:e01084-01015.
147. Witkin SS, Mendes-Soares H, Linhares IM, Jayaram A, Ledger WJ, *et al.*: Influence of vaginal bacteria and D- and L-lactic acid isomers on vaginal extracellular matrix metalloproteinase inducer: implications for protection against upper genital tract infections. *mBio* 2013, **4**.
148. Pendharkar S, Magopane T, Larsson P-G, de Bruyn G, Gray GE, *et al.*: Identification and characterisation of vaginal lactobacilli from South African women. *BMC Infect. Dis.* 2013, **13**:43-43.
149. Saunders S, Bocking A, Challis J, Reid G: Effect of *Lactobacillus* challenge on *Gardnerella vaginalis* biofilms. *Colloids Surf. B Biointerfaces* 2007, **55**:138-142.
150. Song Y-L, Kato N, Matsumiya Y, Liu C-X, Kato H, *et al.*: Identification of and hydrogen peroxide production by fecal and vaginal lactobacilli isolated from Japanese women and newborn infants. *J. Clin. Microbiol.* 1999, **37**:3062-3064.
151. Petrova MI, Reid G, Vaneechoutte M, Lebeer S: *Lactobacillus iners*: friend or foe? *Trends Microbiol.* 2017, **25**:182-191.
152. Tachedjian G, O'Hanlon DE, Ravel J: The implausible "*in vivo*" role of hydrogen peroxide as an antimicrobial factor produced by vaginal microbiota. *Microbiome* 2018, **6**:29.
153. O'Hanlon DE, Lanier BR, Moench TR, Cone RA: Cervicovaginal fluid and semen block the microbicidal activity of hydrogen peroxide produced by vaginal lactobacilli. *BMC Infect. Dis.* 2010, **10**:120-120.
154. Macklaim JM, Gloor GB, Anukam KC, Cribby S, Reid G: At the crossroads of vaginal health and disease, the genome sequence of *Lactobacillus iners* AB-1. *Proc. Natl. Acad. Sci. USA* 2010, **108**:205-209.
155. Macklaim JM, Fernandes AD, Bella JMD, Hammond J-A, Reid G, *et al.*: Comparative meta-RNA-seq of the vaginal microbiota and differential expression by *Lactobacillus iners* in health and dysbiosis. *Microbiome* 2013, **1**:12-12.
156. Borgdorff H, Armstrong SD, Tytgat HL, Xia D, Ndayisaba GF, *et al.*: Unique insights in the cervicovaginal *Lactobacillus iners* and *L. crispatus* proteomes and their associations with microbiota dysbiosis. *PLoS ONE* 2016, **11**:e0150767.
157. Di Paola M, Sani C, Clemente AM, Iossa A, Perissi E, *et al.*: Characterization of cervico-vaginal microbiota in women developing persistent high-risk human papillomavirus infection. *Sci. Rep.* 2017, **7**:10200.

158. Nardini P, Nahui Palomino RA, Parolin C, Laghi L, Foschi C, *et al.*: *Lactobacillus crispatus* inhibits the infectivity of *Chlamydia trachomatis* elementary bodies, *in vitro* study. *Sci. Rep.* 2016, **6**:29024.
159. van Houdt R, Ma B, Bruisten SM, Speksnijder A, Ravel J, *et al.*: *Lactobacillus iners*-dominated vaginal microbiota is associated with increased susceptibility to *Chlamydia trachomatis* infection in Dutch women: a case-control study. *Sex. Transm. Infect.* 2018, **94**:117-123.
160. Oh HY, Kim BS, Seo SS, Kong JS, Lee JK, *et al.*: The association of uterine cervical microbiota with an increased risk for cervical intraepithelial neoplasia in Korea. *Clin. Microbiol. Infect.* 2015, **21**:674 e671-679.
161. Pyles RB, Vincent KL, Baum MM, Elsom B, Miller AL, *et al.*: Cultivated vaginal microbiomes alter HIV-1 infection and antiretroviral efficacy in colonized epithelial multilayer cultures. *PLoS ONE* 2014, **9**:e93419.
162. Gajer P, Brotman RM, Guoyun B, Sakamoto J, Schütte UME, *et al.*: Temporal dynamics of the human vaginal microbiota. *Sci. Transl. Med.* 2012, **4**:132ra152-132ra152.
163. Verstraelen H, Verhelst R, Claeys G, De Backer E, Temmerman M, *et al.*: Longitudinal analysis of the vaginal microflora in pregnancy suggests that *L. crispatus* promotes the stability of the normal vaginal microflora and that *L. gasseri* and/or *L. iners* are more conducive to the occurrence of abnormal vaginal microflora. *BMC Microbiol.* 2009, **9**:116-116.
164. Gosmann C, Anahtar MN, Handley SA, Farcasanu M, Abu-Ali G, *et al.*: *Lactobacillus*-deficient cervicovaginal bacterial communities are associated with increased HIV acquisition in young South African women. *Immunity* 2017, **46**:29-37.
165. van der Veer C, Bruisten SM, van der Helm JJ, de Vries HJ, van Houdt R: The cervicovaginal microbiota in women notified for *Chlamydia trachomatis* infection: a case-control study at the sexually transmitted Infection outpatient clinic in Amsterdam, the Netherlands. *Clin. Infect. Dis.* 2017, **64**:24-31.
166. Drell T, Lillsaar T, Tummeleht L, Simm J, Aaspollu A, *et al.*: Characterization of the vaginal micro- and mycobiome in asymptomatic reproductive-age Estonian women. *PLoS ONE* 2013, **8**:e54379.
167. Zhou X, Bent SJ, Schneider MG, Davis CC, Islam MR, *et al.*: Characterization of vaginal microbial communities in adult healthy women using cultivation-independent methods. *Microbiology* 2004, **150**:2565-2573.
168. Zhou X, Hansmann MA, Davis CC, Suzuki H, Brown CJ, *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.
169. Hickey RJ, Zhou X, Settles ML, Erb J, Malone K, *et al.*: Vaginal microbiota of adolescent girls prior to the onset of menarche resemble those of reproductive-age women. *mBio* 2015, **6**:1-14.
170. Jespers V, Menten J, Smet H, Poradosú S, Abdellati S, *et al.*: Quantification of bacterial species of the vaginal microbiome in different groups of women, using nucleic acid amplification tests. *BMC Microbiol.* 2012, **12**:83-83.
171. Hill JE, Goh SH, Money DM, Doyle M, Li A, *et al.*: Characterization of vaginal microflora of healthy, nonpregnant women by *chaperonin-60* sequence-based methods. *Am. J. Obstet. Gynecol.* 2005, **193**:682-692.
172. Ziklo N, Vidgen ME, Taing K, Huston WM, Timms P: Dysbiosis of the vaginal microbiota and higher vaginal kynurenine/tryptophan ratio reveals an association with *Chlamydia trachomatis* genital infections. *Front. Cell Infect. Microbiol.* 2018, **8**:1.
173. Srinivasan S, Liu C, Mitchell CM, Fiedler TL, Thomas KK, *et al.*: Temporal variability of human vaginal bacteria and relationship with bacterial vaginosis. *PLoS ONE* 2010, **5**:e10197-e10197.
174. Spear GT, Sikaroodi M, Zariffard MR, Landay AL, French AL, *et al.*: 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**:1131-1140.
175. Graspeuntner S, Bohlmann MK, Gillmann K, Speer R, Kuenzel S, *et al.*: Microbiota-based analysis reveals specific bacterial traits and a novel strategy for the diagnosis of infectious infertility. *PLoS ONE* 2018, **13**:e0191047.
176. Seo SS, Oh HY, Lee JK, Kong JS, Lee DO, *et al.*: Combined effect of diet and cervical microbiome on the risk of cervical intraepithelial neoplasia. *Clin. Nutr.* 2016, **35**:1434-1441.
177. Newton ER, Piper JM, Shain RN, Perdue ST, Peairs W: Predictors of the vaginal microflora. *Am. J. Obstet. Gynecol.* 2001, **184**:845-853; discussion 853-845.

178. Paxton LA, Sewankambo N, Gray R, Serwadda D, McNairn D*, et al.*: Asymptomatic non-ulcerative genital tract infections in a rural Ugandan population. *Sex. Transm. Infect.* 1998, **74**:421-425.
179. Lee JE, Lee S, Lee H, Song Y-M, Lee K*, et al.*: Association of the vaginal microbiota with human papillomavirus infection in a Korean twin cohort. *PLoS ONE* 2013, **8**:e63514-e63514.
180. Si J, You HJ, Yu J, Sung J, Ko G: *Prevotella* as a hub for vaginal microbiota under the influence of host genetics and their association with obesity. *Cell Host Microbe* 2017, **21**:97-105.
181. Eschenbach DA, Patton DL, Hooton TM, Meier AS, Stapleton A*, et al.*: Effects of vaginal intercourse with and without a condom on vaginal flora and vaginal epithelium. *J. Infect. Dis.* 2001, **183**:913-918.
182. Fashemi B, Delaney ML, Onderdonk AB, Fichorova RN: Effects of feminine hygiene products on the vaginal mucosal biome. *Microb. Ecol. Health Dis.* 2013, **24**.
183. Muhleisen AL, Herbst-Kralovetz MM: Menopause and the vaginal microbiome. *Maturitas* 2016, **91**:42-50.
184. Brotman RM, Shardell MD, Gajer P, Fadrosh D, Chang K*, et al.*: Association between the vaginal microbiota, menopause status, and signs of vulvovaginal atrophy. *Menopause* 2014, **21**:450-458.
185. Zhang R, Daroczy K, Xiao B, Yu L, Chen R*, et al.*: Qualitative and semiquantitative analysis of *Lactobacillus* species in the vaginas of healthy fertile and postmenopausal Chinese women. *J. Med. Microbiol.* 2012, **61**:729-739.
186. Hummelen R, Macklaim JM, Bisanz JE, Hammond J-A, McMillan A*, et al.*: Vaginal microbiome and epithelial gene array in post-menopausal women with moderate to severe dryness. *PLoS ONE* 2011, **6**:e26602-e26602.
187. Oh HY, Seo SS, Kong JS, Lee JK, Kim MK: Association between obesity and cervical microflora dominated by *Lactobacillus iners* in Korean women. *J. Clin. Microbiol.* 2015, **53**:3304-3309.
188. DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ*, et al.*: Temporal and spatial variation of the human microbiota during pregnancy. *Proc. Natl. Acad. Sci. USA* 2015, **112**:11060-11065.
189. Petricevic L, Domig KJ, Nierscher FJ, Sandhofer MJ, Fidesser M*, et al.*: Characterisation of the vaginal *Lactobacillus* microbiota associated with preterm delivery. *Sci. Rep.* 2014, **4**:5136-5136.
190. Romero R, Hassan SS, Gajer P, Tarca AL, Fadrosh DW*, et al.*: The vaginal microbiota of pregnant women who subsequently have spontaneous preterm labor and delivery and those with a normal delivery at term. *Microbiome* 2014, **2**:18-18.
191. Brooks JP, Edwards DJ, Blithe DL, Fettweis JM, Serrano MG*, et al.*: Effects of combined oral contraceptives, depot medroxyprogesterone acetate and the levonorgestrel-releasing intrauterine system on the vaginal microbiome. *Contraception* 2017, **95**:405-413.
192. Shoubnikova M, Hellberg D, Nilsson S, Mårdh PA: Contraceptive use in women with bacterial vaginosis. *55* 1997, **6**:355-358.
193. Eren AM, Zozaya M, Taylor CM, Dowd SE, Martin DH*, et al.*: Exploring the diversity of *Gardnerella vaginalis* in the genitourinary tract microbiota of monogamous couples through subtle nucleotide variation. *PLoS ONE* 2011, **6**:e26732.
194. Liu CM, Hungate BA, Tobian AA, Ravel J, Prodger JL*, et al.*: Penile microbiota and female partner bacterial vaginosis in Rakai, Uganda. *mBio* 2015, **6**:e00589.
195. Vodstrcil LA, Hocking JS, Law M, Walker S, Tabrizi SN*, et al.*: Hormonal contraception is associated with a reduced risk of bacterial vaginosis: a systematic review and meta-analysis. *PLoS ONE* 2013, **8**:e73055.
196. van de Wijgert JH, Verwijs MC, Turner AN, Morrison CS: Hormonal contraception decreases bacterial vaginosis but oral contraception may increase candidiasis: implications for HIV transmission. *AIDS* 2013, **27**:2141-2153.
197. Petrova MI, van den Broek M, Balzarini J, Vanderleyden J, Lebeer S: Vaginal microbiota and its role in HIV transmission and infection. *FEMS Microbiol. Rev.* 2013, **37**:762-792.
198. Nelson DB, Hanlon A, Nachamkin I, Haggerty C, Mastrogiannis DS*, et al.*: Early pregnancy changes in bacterial vaginosis-associated bacteria and preterm delivery. *Paediatr. Perinat. Epidemiol.* 2014, **28**:88-96.
199. Seo SS, Arokiyaraj S, Kim MK, Oh HY, Kwon M*, et al.*: High prevalence of *Leptotrichia amnionii*, *Atopobium vaginae*, *Sneathia sanguinegens*, and factor 1 microbes and association of spontaneous abortion among Korean women. *Biomed. Res. Int.* 2017, **2017**:5435089.

200. van Oostrum N, De Sutter P, Meys J, Verstraelen H: Risks associated with bacterial vaginosis in infertility patients: a systematic review and meta-analysis. *Hum. Reprod.* 2013, **28**:1809-1815.
201. Bediako PT, BeLue R, Hillemeier MM: A comparison of birth outcomes among Black, Hispanic, and Black Hispanic women. *J. Racial Ethn. Health Disparities* 2015, **2**:573-582.
202. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, *et al.*: Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc. Natl. Acad. Sci. USA* 2010, **107**:11971-11975.
203. Westerbeek EA, van den Berg A, Lafeber HN, Knol J, Fetter WP, *et al.*: The intestinal bacterial colonisation in preterm infants: a review of the literature. *Clin. Nutr.* 2006, **25**:361-368.
204. Young VB: The intestinal microbiota in health and disease. *Curr. Opin. Gastroenterol.* 2012, **28**:63-69.
205. Taylor BD, Darville T, Haggerty CL: Does bacterial vaginosis cause pelvic inflammatory disease? *Sex. Transm. Dis.* 2013, **40**:117-122.
206. Bhandari P, Prabha V: Evaluation of profertility effect of probiotic *Lactobacillus plantarum* 2621 in a murine model. *Indian. J. Med. Res.* 2015, **142**:79-84.
207. Barbonetti A, Cinque B, Vassallo MR, Mineo S, Francavilla S, *et al.*: Effect of vaginal probiotic lactobacilli on *in vitro*-induced sperm lipid peroxidation and its impact on sperm motility and viability. *Fertil. Steril.* 2011, **95**:2485-2488.
208. Haahr T, Jensen JS, Thomsen L, Duus L, Rygaard K, *et al.*: Abnormal vaginal microbiota may be associated with poor reproductive outcomes: a prospective study in IVF patients. *Hum. Reprod.* 2016, **31**:795-803.
209. Nardis C, Mosca L, Mastromarino P: Vaginal microbiota and viral sexually transmitted diseases. *Ann. Ig.* 2013, **25**:443-456.
210. King CC, Jamieson DJ, Wiener J, Cu-Uvin S, Klein RS, *et al.*: Bacterial vaginosis and the natural history of human papillomavirus. *Infect. Dis. Obstet. Gynecol.* 2011, **2011**:319460-319460.
211. Frank DN, Manigart O, Leroy V, Meda N, Valéa D, *et al.*: Altered vaginal microbiota are associated with perinatal mother-to-child transmission of HIV in African women from Burkina Faso. *J. Acquir. Immune Defic. Syndr.* 2012, **60**.
212. Nahui Palomino RA, Zicari S, Vanpouille C, Vitali B, Margolis L: Vaginal *Lactobacillus* inhibits HIV-1 replication in human tissues *ex vivo*. *Front. Microbiol.* 2017, **8**:906.
213. Mehta SD, Donovan B, Weber KM, Cohen M, Ravel J, *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**:e0116894-e0116894.
214. Kero K, Rautava J, Syrjanen K, Grenman S, Syrjanen S: Association of asymptomatic bacterial vaginosis with persistence of female genital human papillomavirus infection. *Eur. J. Clin. Microbiol. Infect. Dis.* 2017.
215. Gao W, Weng J, Gao Y, Chen X: Comparison of the vaginal microbiota diversity of women with and without human papillomavirus infection: a cross-sectional study. *BMC Infect. Dis.* 2013, **13**:271-271.
216. Kyrgiou M, Mitra A, Moscicki AB: Does the vaginal microbiota play a role in the development of cervical cancer? *Transl. Res.* 2016, **179**:168-182.
217. Champer M, Wong A, Champer J, Brito I, Messer P, *et al.*: The role of the vaginal microbiome in gynaecological cancer. *BJOG* 2017.
218. Mitra A, MacIntyre DA, Marchesi JR, Lee YS, Bennett PR, *et al.*: The vaginal microbiota, human papillomavirus infection and cervical intraepithelial neoplasia: what do we know and where are we going next? *Microbiome* 2016, **4**:58.
219. Piyathilake CJ, Ollberding NJ, Kumar R, Macaluso M, Alvarez RD, *et al.*: Cervical microbiota associated with higher grade cervical intraepithelial neoplasia in women infected with high-risk human papillomaviruses. *Cancer Prev. Res. (Phila)* 2016, **9**:357-366.
220. Gillet E, Meys JFA, Verstraelen H, Verhelst R, De Sutter P, *et al.*: Association between bacterial vaginosis and cervical intraepithelial neoplasia: systematic review and meta-analysis. *PLoS ONE* 2012, **7**:e45201-e45201.
221. Peters N, Van Leeuwen AM, Pieters WJ, Hollema H, Quint WG, *et al.*: Bacterial vaginosis is not important in the etiology of cervical neoplasia. *Sex. Transm. Dis.* 1995, **22**:296-302.
222. Verhoeven V, Renard N, Makar A, Van Royen P, Bogers JP, *et al.*: Probiotics enhance the clearance of human papillomavirus-related cervical lesions: a prospective controlled pilot study. *Eur. J. Cancer Prev.* 2013, **22**:46-51.

223. Al Kassaa I, Hober D, Hamze M, Chihib NE, Drider D: Antiviral potential of lactic acid bacteria and their bacteriocins. *Probiotics Antimicrob. Proteins* 2014, **6**:177-185.
224. Motevaseli E, Shirzad M, Akrami SM, Mousavi AS, Mirsalehian A*, et al.*: Normal and tumour cervical cells respond differently to vaginal lactobacilli, independent of pH and lactate. *J. Med. Microbiol.* 2013, **62**:1065-1072.
225. Anderson D, Politch JA, Pudney J: HIV infection and immune defense of the penis. *Am. J. Reprod. Immunol.* 2011, **65**:220-229.
226. Bonner K: Male circumcision as an HIV control Strategy: not a natural condom'. *Reprod. Health Matters* 2001, **9**:143-155.
227. Wasco MJ, Shah RB: Benign diseases and neoplasms of the penis. *Surg. Pathol. Clin.* 2009, **2**:161-197.
228. Pudney J, Anderson D: Innate and acquired immunity in the human penile urethra. *J. Reprod. Immunol.* 2011, **88**:219-227.
229. Price LB, Liu CM, Johnson KE, Aziz M, Lau MK*, et al.*: The effects of circumcision on the penis microbiome. *PLoS ONE* 2010, **5**:e8422-e8422.
230. O'Farrell N, Chung CK, Weiss HA: Foreskin length in uncircumcised men is associated with subpreputial wetness. *Int. J. STD AIDS* 2008, **19**:821-823.
231. O'Farrell N, Morison L, Moodley P, Pillay K, Vanmali T*, et al.*: Association between HIV and subpreputial penile wetness in uncircumcised men in South Africa. *J. Acquir. Immune Defic. Syndr.* 2006, **43**:69–77.
232. Schneider JA, Vadivelu S, Liao C, Kandukuri SR, Trikamji BV*, et al.*: Increased likelihood of bacterial pathogens in the coronal sulcus and urethra of uncircumcised men in a diverse group of HIV infected and uninfected patients in India. *J. Glob. Infect. Dis.* 2012, **4**:6-9.
233. Auvert B, Taljaard D, Lagarde E, Sobngwi-Tambekou J, Sitta R*, et al.*: Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: The ANRS 1265 Trial. *PLoS Med.* 2005, **2**:e298-e298.
234. Tobian AAR, Serwadda D, Quinn TC, Kigozi G, Gravitt PE*, et al.*: Male circumcision for the prevention of HSV-2 and HPV infections and syphilis. *N. Engl. J. Med.* 2009, **360**:1298-1309.
235. Vallely AJ, MacLaren D, David M, Toliman P, Kelly-Hanku A*, et al.*: Dorsal longitudinal foreskin cut is associated with reduced risk of HIV, syphilis and genital herpes in men: a cross-sectional study in Papua New Guinea. *J. Int. AIDS Soc.* 2017, **20**:1-11.
236. Castellsagué X, Bosch FX, Muñoz N, Meijer CJLM, Shah KV*, et al.*: Male circumcision, penile human papillomavirus infection, and cervical cancer in female partners. *N. Engl. J. Med.* 2002, **346**:1105-1112.
237. Bailey RC, Moses S, Parker CB, Agot K, Maclean I*, et al.*: Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. *Lancet* 2007, **369**:643-656.
238. Gray RH, Kigozi G, Serwadda D, Makumbi F, Watya S*, et al.*: Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial. *Lancet* 2007, **369**:657-666.
239. Hernandez BY, Wilkens LR, Zhu X, McDuffie K, Thompson P*, et al.*: Circumcision and human papillomavirus infection in men: a site-specific comparison. *J. Infect. Dis.* 2008, **197**:787-794.
240. Zhu YP, Jia ZW, Dai B, Ye DW, Kong YY*, et al.*: Relationship between circumcision and human papillomavirus infection: a systematic review and meta-analysis. *Asian. J. Androl.* 2017, **19**:125-131.
241. Larke N, Thomas SL, Dos Santos Silva I, Weiss HA: Male circumcision and human papillomavirus infection in men: a systematic review and meta-analysis. *J. Infect. Dis.* 2011, **204**:1375-1390.
242. Siegfried N, Muller M, Deeks JJ, Volmink J: Male circumcision for prevention of heterosexual acquisition of HIV in men. *Cochrane Database Syst. Rev.* 2009:CD003362.
243. Kim HH, Li PS, Goldstein M: Male circumcision: Africa and beyond? *Curr. Opin. Urol.* 2010, **20**:515-519.
244. Wilcken A, Keil T, Dick B: Traditional male circumcision in eastern and southern Africa: a systematic review of prevalence and complications. *Bull. World Health Organ.* 2010, **88**:907-914.
245. Connolly C, Simbayi LC, Shanmugam R, Nqeketo A: Male circumcision and its relationship to HIV infection in South Africa: results of a national survey in 2002. *S. Afr. Med. J.* 2008, **98**:789-794.
246. Sarvestani AS, Bufumbo L, Geiger JD, Sienko KH: Traditional male circumcision in Uganda: a qualitative focus group discussion analysis. *PLoS ONE* 2012, **7**:e45316.
247. Maffioli EM: Is traditional male circumcision effective as an HIV prevention strategy? Evidence from Lesotho. *PLoS ONE* 2017, **12**:e0177076.

248. Shaffer DN, Bautista CT, Sateren WB, Sawe FK, Kiplangat SC*, et al.*: The protective effect of circumcision on HIV incidence in rural low-risk men circumcised predominantly by traditional circumcisers in Kenya: two-year follow-up of the Kericho HIV Cohort Study. *J. Acquir. Immune Defic. Syndr.* 2007, **45**:371-379.
249. Thomas AG, Tran BR, Cranston M, Brown MC, Kumar R*, et al.*: Voluntary medical male circumcision: a cross-sectional study comparing circumcision self-report and physical examination findings in Lesotho. *PLoS ONE* 2011, **6**:e27561.
250. Gebremedhin S: Male circumcision and its association with HIV infection and sexually transmitted diseases: evidence from 18 demographic and health surveys in sub-Saharan Africa. *SAHARA J.* 2011, **8**:13-18.
251. Maughan-Brown B, Venkataramani AS, Nattrass N, Seekings J, Whiteside AW: A cut above the rest: traditional male circumcision and HIV risk among Xhosa men in Cape Town, South Africa. *J. Acquir. Immune Defic. Syndr.* 2011, **58**:499-505.
252. Peltzer K, Simbayi L, Banyini M, Kekana Q: HIV risk reduction intervention among medically circumcised young men in South Africa: a randomized controlled trial. *Int. J. Behav. Med.* 2012, **19**:336-341.
253. Peltzer K, Simbayi L, Banyini M, Kekana Q: HIV risk reduction intervention among traditionally circumcised young men in South Africa: a cluster randomized control trial. *J. Assoc. Nurses AIDS Care* 2011, **22**:397-406.
254. Liu CM, Hungate BA, Tobian AAR, Serwadda D, Ravel J*, et al.*: Male circumcision significantly reduces prevalence and load of genital anaerobic bacteria. *mBio* 2013, **4**:1-9.
255. Liu CM, Prodger JL, Tobian AA, Serwadda D, Galiwango RM*, et al.*: Genital anaerobic bacterial overgrowth and the PrePex male circumcision device, Rakai, Uganda. *J. Infect. Dis.* 2016, **214**:595-598.
256. Plummer EL, Vodstrcil LA, Danielewski JA, Murray GL, Fairley CK*, et al.*: Combined oral and topical antimicrobial therapy for male partners of women with bacterial vaginosis: acceptability, tolerability and impact on the genital microbiota of couples - a pilot study. *PLoS ONE* 2018, **13**:e0190199.
257. Nelson DE, Dong Q, Van der Pol B, Toh E, Fan B*, et al.*: Bacterial communities of the coronal sulcus and distal urethra of adolescent males. *PLoS ONE* 2012, **7**:e36298-e36298.
258. Grice EA, Segre JA: The skin microbiome. *Nat. Rev. Microbiol.* 2011, **9**:244-253.
259. Prodger JL, Gray RH, Shannon B, Shahabi K, Kong X*, et al.*: Chemokine levels in the penile coronal sulcus correlate with HIV-1 acquisition and are reduced by male crcumcision in Rakai, Uganda. *PLoS Pathog.* 2016, **12**:e1006025.
260. Gray RH, Kigozi G, Serwadda D, Nalugoda F, Watya S*, et al.*: The effects of male circumcision on female partners' genital tract symptoms and vaginal infections in a randomized trial in Rakai, Uganda. *Am. J. Obstet. Gynecol.* 2009, **200**:42.e41-47.
261. Mehta SD: Systematic review of randomized trials of treatment of male sexual partners for improved bacteria vaginosis outcomes in women. *Sex. Transm. Dis.* 2012, **39**:822-830.
262. Gray RH, Serwadda D, Kong X, Makumbi F, Kigozi G*, et al.*: Male circumcision decreases acquisition and increases clearance of high-risk human papillomavirus in HIV-negative men: a randomized trial in Rakai, Uganda. *J. Infect. Dis.* 2010, **201**:1455-1462.
263. Aldunate M, Tyssen D, Johnson A, Zakir T, Sonza S*, et al.*: Vaginal concentrations of lactic acid potently inactivate HIV. *J. Antimicrob. Chemother.* 2013, **68**:2015-2025.
264. Hyman RW, Fukushima M, Jiang H, Fung E, Rand L*, et al.*: Diversity of the vaginal microbiome correlates with preterm birth. *Reprod. Sci.* 2014, **21**:32-40.
265. Lamont RF, Sobel JD, Akins RA, Hassan SS, Chaiworapongsa T*, et al.*: The vaginal microbiome: new information about genital tract using molecular based techniques. *Brit. J. Obstet. Gynaec.* 2011, **118**:533-549.
266. Li J, McCormick J, Bocking A, Reid G: Importance of vaginal microbes in reproductive health. *Reprod. Sci.* 2012, **19**:235-242.
267. Watts DH, Fazzari M, Fazarri M, Minkoff H, Hillier SL*, et al.*: Effects of bacterial vaginosis and other genital infections on the natural history of human papillomavirus infection in HIV-1-infected and high-risk HIV-1-uninfected women. *J. Infect. Dis.* 2005, **191**:1129-1139.
268. de Sanjosé S, Diaz M, Castellsagué X, Clifford G, Bruni L*, et al.*: Worldwide prevalence and genotype distribution of cervical human papillomavirus DNA in women with normal cytology: a meta-analysis. *Lancet Infect. Dis.* 2007, **7**:453-459.
269. Quail MA, Smith M, Coupland P, Otto TD, Harris SR*, et al.*: A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 2012, **13**:341-341.

270. Mbulawa ZZA, Coetzee D, Marais DJ, Kamupira M, Zwane E*, et al.*: Genital human papillomavirus prevalence and human papillomavirus concordance in heterosexual couples are positively associated with human immunodeficiency virus coinfection. *J. Infect. Dis.* 2009, **199**:1514-1524.
271. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA*, et al.*: Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *PNAS* 2011, **15**:4516-4522.
272. Andrews S: FastQC: a quality control tool for high throughput sequence data. 2010. Accessed on 05 April 2018. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
273. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD*, et al.*: QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 2010, **7**:335-336.
274. Edgar RC: UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 2013, **10**:996-998.
275. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R: UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 2011, **27**:2194-2200.
276. Wang Q, Garrity GM, Tiedje JM, Cole JR: Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 2007, **73**:5261-5267.
277. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL*, et al.*: PyNAST: A flexible tool for aligning sequences to a template alignment. *Bioinformatics* 2010, **26**:266-267.
278. Philip D: Computer program review: VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* 2003, **14**:927-930.
279. Parks DH, Tyson GW, Hugenholtz P, Beiko RG: STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 2014, **30**:3123-3124.
280. Paulson JN, Talukder H, Pop M, Bravo HC: metagenomeSeq: statistical analysis for sparse high-throughput sequencing. 2016. Bioconductor package: 1.14.2.http://cbcb.umd.edu/software/metagenomeSeq.
281. Paulson JN, Stine OC, Bravo HC, Pop M: Robust methods for differential abundance analysis in marker gene surveys. *Nat. Methods* 2014, **10**:1200-1202.
282. Jensen DB, Ussery DW: Bayesian prediction of microbial oxygen requirement. *F1000Res* 2013, **2**:184.
283. Roesch LF, Silveira RC, Corso AL, Dobbler PT, Mai V*, et al.*: Diversity and composition of vaginal microbiota of pregnant women at risk for transmitting Group B *Streptococcus* treated with intrapartum penicillin. *PLoS ONE* 2017, **12**:e0169916.
284. Tishkoff SA, Williams SM: Genetic analysis of African populations: human evolution and complex disease. *Nat. Rev. Genet.* 2002, **3**:611-621.
285. Fettweis JM, Serrano MG, Girerd PH, Jefferson K, Buck GA: A new era of the vaginal microbiome: advances using next-generation sequencing. *Chem. Biodivers.* 2012, **9**:965-976.
286. Oakley BB, Fiedler TL, Marrazzo JM, Fredricks DN: Diversity of human vaginal bacterial communities and associations with clinically defined bacterial vaginosis. *Appl. Environ. Microbiol.* 2008, **74**:4898-4909.
287. Sobel JD: Bacterial vaginosis. *Annu. Rev. Med.* 2000, **51**:349-356.
288. Jarosik GP, Land CB, Duhon P, Chandler R, Mercer T: Acquisition of ion by *Gardnerella vaginalis*. *Infect. Immun.* 1998, **66**:5041-5047.
289. Roxby AC, Fredricks DN, Odem-Davis K, Ásbjörnsdóttir K, Masese L*, et al.*: Changes in vaginal microbiota and immune mediators in HIV-1-seronegative Kenyan women initiating depot medroxyprogesterone acetate. *J. Acquir. Immune Defic. Syndr.* 2016, **71**:359-366.
290. Zelezniak A, Andrejev S, Ponomarova O, Mende DR, Bork P*, et al.*: Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc. Natl. Acad. Sci. USA* 2015, **112**:6449-6454.
291. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D*, et al.*: Microbial co-occurrence relationships in the human microbiome. *PLoS Comput. Biol.* 2012, **8**:e1002606.
292. De Backer E, Verhelst R, Verstraelen H, Claeys G, Verschraegen G*, et al.*: Antibiotic susceptibility of *Atopobium vaginae*. *BMC Infect. Dis.* 2006, **6**:51.
293. Greene JF, Kuehl TJ, Allen SR: The Papanicolaou smear: inadequate screening test for bacterial vaginosis during pregnancy. *Am. J. Obstet. Gynecol.* 2000, **182**:1048-1049.
294. Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ: Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS ONE* 2012, **7**:e33865-e33865.

295. Fouhy F, Clooney AG, Stanton C, Claesson MJ, Cotter PD: 16S rRNA gene sequencing of mock microbial populations- impact of DNA extraction method, primer choice and sequencing platform. *BMC Microbiol.* 2016, **16**:123.
296. Siegwald L, Touzet H, Lemoine Y, Hot D, Audebert C*, et al.*: Assessment of common and emerging bioinformatics pipelines for targeted metagenomics. *PLoS ONE* 2017, **12**:e0169563.
297. Cozzuto L, Company C, Somavilla NA, Hecht J, Onywera DH*, et al.*: Benchmarking 16S rRNA gene sequencing and bioinformatics tools for identification of microbial abundances. In *NGS 16: Genome Annotation*; *Barcelona, Spain*: ISCB: 2016.
298. Yang B, Wang Y, Qian PY: Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics* 2016, **17**:135.
299. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C*, et al.*: Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 2013, **41**:e1.
300. Anahtar MN, Bowman BA, Kwon DS: Efficient nucleic acid extraction and 16S rRNA gene sequencing for bacterial community characterization. *J. Vis. Exp.* 2016.
301. Ghyselinck J, Pfeiffer S, Heylen K, Sessitsch A, De Vos P: The effect of primer choice and short read sequences on the outcome of 16S rRNA gene based diversity studies. *PLoS ONE* 2013, **8**:e71360-e71360.
302. Tremblay J, Singh K, Fern A, Kirton ES, He S*, et al.*: Primer and platform effects on 16S rRNA tag sequencing. *Front. Microbiol.* 2015, **6**:771.
303. Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J: Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. *PLoS ONE* 2014, **9**:e94249-e94249.
304. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE*, et al.*: Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* 2012, **30**:434-439.
305. Salipante SJ, Kawashima T, Rosenthal C, Hoogestraat DR, Cummings La*, et al.*: Performance comparison of Illumina and Ion Torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. *Appl. Environ. Microbiol.* 2014, **80**:7583-7591.
306. Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S*, et al.*: An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* 2014, **2**:6-6.
307. Illumina: 16S metagenomic sequencing library preparation. Preparing 16S ribosomal RNA gene amplicons for the Illumina MiSeq System. 2013. Accessed on 05 April 2018. Available from: https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf.
308. Takara Bio Inc.: *Ex Taq* DNA polymerase, hot-start version. 2018. Accessed on 08 October 2018. https://www.takarabio.com/products/pcr/high-yield-pcr/ex-taq-products/ex-taq-polymerase-hot-start.
309. Vázquez-baeza Y, Pirrung M, Gonzalez A, Knight R: EMPeror: a tool for visualizing high-throughput microbial community data. *Gigascience* 2013, **2**:16-16.
310. Peres-Neto PR, Jackson DA: How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia* 2001, **129**:169-178.
311. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J*, et al.*: Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 2012, **6**:1621-1624.
312. Bragg LM, Stone G, Butler MK, Hugenholtz P, Tyson GW: Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLoS Comput. Biol.* 2013, **9**:e1003031-e1003031.
313. Mizrahi-Man O, Davenport ER, Gilad Y: Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. *PLoS ONE* 2013, **8**:e53608-e53608.
314. Brooks JP, Edwards DJ, Harwich MD, Jr., Rivera MC, Fettweis JM*, et al.*: The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol.* 2015, **15**:66.
315. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI*, et al.*: Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* 2013, **10**:57-60.

316. Straight SW, Herman B, McCance DJ: The E5 oncoprotein of human papillomavirus type 16 inhibits the acidification of endosomes in human keratinocytes. *J. Virol.* 1995, **69**:3185-3192.
317. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D*, et al.*: Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 2013, **31**:814-821.
318. White JR, Nagarajan N, Pop M: Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput. Biol.* 2009, **5**:e1000352.
319. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L*, et al.*: Metagenomic biomarker discovery and explanation. *Genome Biol.* 2011, **12**:R60-R60.
320. Edgar RC: UNBIAS: an attempt to correct abundance bias in 16S sequencing, with limited success. 2017.
321. Edgar RC: UPARSE OTU analysis pipline. Misop tutorial - MiSeq 2x250 PE reads, mouse feces and mock community samples. Accessed on 05 April 2018. Available from: http://drive5.com/usearch/manual/upp_tut_misop_mock.html.
322. Bokulich NA, Rideout JR, Mercurio WG, Shiffer A, Wolfe B*, et al.*: mockrobiota: a public resource for microbiome bioinformatics benchmarking. *mSystems* 2016, **1**.
323. Narrowe AB, Albuthi-Lantz M, Smith EP, Bower KJ, Roane TM*, et al.*: Perturbation and restoration of the fathead minnow gut microbiome after low-level triclosan exposure. *Microbiome* 2015, **3**:6-6.
324. Morris EK, Caruso T, Buscot F, Fischer M, Hancock C*, et al.*: Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. *Ecol. Evol.* 2014, **4**:3514-3524.
325. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO*, et al.*: Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 2014, **12**:87-87.
326. Ling Z, Kong J, Liu F, Zhu H, Chen X*, et al.*: Molecular analysis of the diversity of vaginal microbiota associated with bacterial vaginosis. *BMC Genomics* 2010, **11**:488-488.
327. Forney LJ, Foster JA, W L: The vaginal flora of healthy women is not always dominated by *Lactobacillus* species. *J. Infect. Dis.* 2006, **194**:1468-1469.
328. Borgdorff H, Gautam R, Armstrong SD, Xia D, Ndayisaba GF*, et al.*: Cervicovaginal microbiome dysbiosis is associated with proteome changes related to alterations of the cervicovaginal mucosal barrier. *Mucosal Immunol.* 2015, **9**:621-633.
329. Brotman RM, He X, Gajer P, Fadrosh D, Sharma E*, et al.*: Association between cigarette smoking and the vaginal microbiota: a pilot study. *BMC Infect. Dis.* 2014, **14**:471.
330. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D*, et al.*: Microbial Co-occurrence Relationships in the Human Microbiome. *PLoS computational biology* 2012, **8**:e1002606.
331. Ebrahim S, Mndende XK, Kharsany AB, Mbulawa ZZ, Naranbhai V*, et al.*: High burden of human papillomavirus (HPV) infection among young women in KwaZulu-Natal, South Africa. *PLoS ONE* 2016, **11**:e0146603.
332. Lubelski J, Konings WN, Driessen AJ: Distribution and physiology of ABC-type transporters contributing to multidrug resistance in bacteria. *Microbiol. Mol. Biol. Rev.* 2007, **71**:463-476.
333. Soto SM: Role of efflux pumps in the antibiotic resistance of bacteria embedded in a biofilm. *Virulence* 2013, **4**:223-229.
334. Harwich MD, Jr., Serrano MG, Fettweis JM, Alves JM, Reimers MA*, et al.*: Genomic sequence analysis and characterization of *Sneathia amnii* sp. nov. *BMC Genomics* 2012, **13 Suppl 8**:S4.
335. Yeoman CJ, Yildirim S, Thomas SM, Durkin AS, Torralba M*, et al.*: Comparative genomics of *Gardnerella vaginalis* strains reveals substantial differences in metabolic and virulence potential. *PLoS ONE* 2010, **5**:e12411.
336. Harwich MDJ, Alves JM, Buck GA, Strauss III JF, Patterson JL*, et al.*: Drawing the line between commensal and pathogenic *Gardnerella vaginalis* through genome analysis and virulence studies. *BMC Genomics* 2010, **11**:375.
337. Cornejo OE, Hickey RJ, Suzuki H, Forney LJ: Focusing the diversity of *Gardnerella vaginalis* through the lens of ecotypes. *Evol. Appl.* 2018, **11**:312-324.
338. Nakao R, Senpuku H, Watanabe H: *Porphyromonas gingivalis* galE is involved in lipopolysaccharide O-antigen synthesis and biofilm formation. *Infect. Immun.* 2006, **74**:6145-6153.
339. Cuthbertson L, Kos V, Whitfield C: ABC transporters involved in export of cell surface glycoconjugates. *Microbiol. Mol. Biol. Rev.* 2010, **74**:341-362.

340. France MT, Mendes-Soares H, Forney LJ: Genomic comparisons of *Lactobacillus crispatus* and *Lactobacillus iners* reveal potential ecological drivers of community composition in the vagina. *Appl. Environ. Microbiol.* 2016, **82**:7063-7073.
341. Thurman AR, Kimble T, Herold B, Mesquita PM, Fichorova RN, *et al.*: Bacterial vaginosis and subclinical markers of genital tract inflammation and mucosal immunity. *AIDS Res. Hum. Retroviruses* 2015, **31**:1139-1152.
342. Chaban B, Hughes HV, Beeby M: The flagellum in bacterial pathogens: for motility and a whole lot more. *Semin. Cell Dev. Biol.* 2015, **46**:91-103.
343. Wood TK, Gonzalez Barrios AF, Herzberg M, Lee J: Motility influences biofilm architecture in Escherichia coli. *Appl. Microbiol. Biotechnol.* 2006, **72**:361-367.
344. O'Toole GA, Kolter R: Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. *Mol. Microbiol.* 1998, **30**:295-304.
345. Beamer MA, Austin MN, Avolia HA, Meyn LA, Bunge KE*, et al.*: Bacterial species colonizing the vagina of healthy women are not associated with race. *Anaerobe* 2017.
346. Flacher V, Bouschbacher M, Verronese E, Massacrier C, Sisirak V*, et al.*: Human Langerhans cells express a specific TLR profile and differentially respond to viruses and Gram-positive bacteria. *J. Immunol.* 2006, **177**:7959-7967.
347. Zhang J, Li G, Bafica A, Pantelic M, Zhang P*, et al.*: *Neisseria gonorrhoeae* enhances infection of dendritic cells by HIV type 1. *J. Immunol.* 2005, **174**:7995-8002.
348. Donoval BA, Landay AL, Moses S, Agot K, Ndinya-Achola JO*, et al.*: HIV-1 target cells in foreskins of African men with varying histories of sexually transmitted infections. *Am. J. Clin.* 2006, **125**:386-391.
349. Patterson BK, Landay A, Siegel JN, Flener Z, Pessis D*, et al.*: Susceptibility to human immunodeficiency virus-1 infection of human foreskin and cervical tissue grown in explant culture. *Am. J. Pathol.* 2002, **161**:867-873.
350. Ogawa Y, Kawamura T, Kimura T, Ito M, Blauvelt A*, et al.*: Gram-positive bacteria enhance HIV-1 susceptibility in Langerhans cells, but not in dendritic cells, via Toll-like receptor activation. *Blood* 2009, **113**:5157–5166.
351. McGowin CL, Ma L, Martin DH, Pyles RB: *Mycoplasma genitalium*-encoded MG309 activates NF-kappaB via Toll-like receptors 2 and 6 to elicit proinflammatory cytokine secretion from human genital epithelial cells. *Infect. Immun.* 2009, **77**:1175-1181.
352. de Jong MAWP, Geijtenbeek TBH: Human immunodeficiency virus-1 acquisition in genital mucosa: Langerhans cells as key-players. *J. Intern. Med.* 2009, **265**:18-28.
353. de Witte L, Nabatov A, Geijtenbeek TB: Distinct roles for DC-SIGN+-dendritic cells and Langerhans cells in HIV-1 transmission. *Trends Mol. Med.* 2008, **14**:12-19.
354. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M*, et al.*: Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 2009, **75**:7537-7541.
355. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD: Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* 2013, **79**:5112-5120. Accessed on 05 April 2018. Available from: https://www.mothur.org/wiki/MiSeq_SOP.
356. Good IJ: The population frequencies of species and the estimation of population parameters. *Biometrika* 1953, **40**:237–264.
357. McMurdie PJ, Holmes S: phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 2013, **8**:e61217-e61217.
358. Hugenholtz P, Goebel BM, Pace NR: Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* 1998, **180**:4765-4774.
359. Realini L, De Ridder K, Palomino J-C, Hirschel B, Portaels F: Microaerophilic conditions promote growth of *Mycobacterium genavense*. *J. Clin. Microbiol.* 1998, **36**:2565-2570.
360. Gottschick C, Deng ZL, Vital M, Masur C, Abels C*, et al.*: The urinary microbiota of men and women and its changes in women during bacterial vaginosis and antibiotic treatment. *Microbiome* 2017, **5**:99.
361. Allali I, Arnold JW, Roach J, Cadenas MB, Butz N*, et al.*: A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut microbiome. *BMC Microbiol.* 2017, **17**:194.
362. Mändar R, Punab M, Korrovits P, Turk S, Ausmees K*, et al.*: Seminal microbiome in men with and without prostatitis. *Int. J. Urol.* 2017, **24**:211-216.
363. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ*, et al.*: DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 2016, **13**:581-583.

364. Tobian AA, Kong X, Gravitt PE, Eaton KP, Kigozi G*, et al.*: Male circumcision and anatomic sites of penile high-risk human papillomavirus in Rakai, Uganda. *Int. J. Cancer* 2011, **129**:2970-2975.
365. Brooke JS: *Stenotrophomonas maltophilia*: an emerging global opportunistic pathogen. *Clin. Microbiol. Rev.* 2012, **25**:2-41.
366. Horz HP, Citron DM, Warren YA, Goldstein EJ, Conrads G: *Synergistes* group organisms of human origin. *J. Clin. Microbiol.* 2006, **44**:2914-2920.
367. Jumas-Bilak E, Carlier JP, Jean-Pierre H, Citron D, Bernard K*, et al.*: *Jonquetella anthropi* gen. nov., sp. nov., the first member of the candidate phylum '*Synergistetes*' isolated from man. *Int. J. Syst. Evol. Microbiol.* 2007, **57**:2743-2748.
368. Han XY, Andrade RA: *Brevundimonas diminuta* infections and its resistance to fluoroquinolones. *J. Antimicrob. Chemother.* 2005, **55**:853-859.
369. Doran TI: The role of Citrobacter in clinical disease of children - review. *Clin. Infect. Dis.* 1999, **28**:384-394.
370. Zenilman JM, Fresia A, Berger B, Mccormack WM: Bacterial vaginosis is not associated with circumcision status of the current male partner. *Sex. Transm. Inf.* 1999, **75**:347-348.
371. Holst E: Reservoir of four organisms associated with bacterial vaginosis suggests lack of sexual transmission. *J. Clin. Microbiol.* 1990, **28**:2035-2039.

Supplementary Data

I. Supplemental Tables

Supplemental Table 2.1. QIIME and UPARSE selected parameters for cervicovaginal microbiota analyses.

Step	QIIME/UPARSE	Script/Command Name	Parameters
Demultiplexing and quality-filtering	QIIME	split_libraries.py	-s, --min_qual_score (25); -b, --barcode_type (variable_length); -l, --min_seq_length (200); L, --max_seq_length (400); -d, --record_qual_scores; -j, --added_demultiplex_field (run_prefix); w, --qual_score_window (50); --z, reverse_primers (truncate_only)
Quality-filtering using UPARSE	UPARSE	usearch --fastq_filter	E, --fastq_maxee (0.5); L, --fastq_minlen 270
Dereplicating sequences	UPARSE	usearch --derep_fulllength	-
Abundance sorting and discarding singletons	UPARSE	usearch --sortbysize	--minsize (2)
UPARSE-OTU method for OTU clustering with the greedy clustering algorithm and representative sequence picking with abundance algorithm	UPARSE	usearch --cluster_otus	--otu_radius_pct (3.0)
Reference chimera checking and removal using uchime	UPARSE	usearch --uchime_ref	--mindiv (0.8); mindiffs (3); --db Greengenes, gg13_8
Mapping original quality-filtered reads back to OTUs	UPARSE	usearch --usearch_global	--id (0.97)
Assigning taxonomy to OTUs using RDP Classifier	QIIME	assign_taxonomy.py	-t Greengenes gg13_8
Sequence alignment	QIIME	align_seqs.py	-e, --min_length (75); -p, --minimum_percent_id (0.75)
Filtering sequence alignment	QIIME	filter_alignment.py	- g, --allowed_gap_frac (0.999999)
Inferring phylogeny	QIIME	make_phylogeny.py	-
Generating a dense/sparse representation of an OTU table (observation matrix)	QIIME	make_otu_table.py	-e, --exclude_otus_fp
Diversity analyses	QIIME	core_diversity_analyses.py	-e, --sampling_depth (5000); -c, --categories; -p, --parameter_fp (metrics: chao1, observed_species, Shannon, simpson, PD_whole_tree, unweighted_unifrac,weighted_unifrac,bray_curtis; single rarefaction depth: 5000)

Supplemental Table 3.1. QIIME and UPARSE selected parameters for cervicovaginal microbiota analyses for the Illumina dataset.

Step	QIIME/UPARSE	Script/Command Name	Parameters
Merging and quality-filtering	UPARSE	usearch –merge_fastq	-- E, --fastq_maxee (1.0); --stripleft (30), --trunclen (400), --min_overlap (32)
Quality-filtering using UPARSE	UPARSE	usearch --fastq_filter	E, --fastq_maxee (1.0); L, --fastq_minlen 400
Dereplicating sequences	UPARSE	usearch --derep_fulllength	-
Abundance sorting and discarding singletons	UPARSE	usearch --sortbysize	--minsize (2)
UPARSE-OTU method for OTU clustering with the greedy clustering algorithm and representative sequence picking with abundance algorithm	UPARSE	usearch --cluster_otus	--otu_radius_pct (3.0)
Reference chimera checking and removal using uchime	UPARSE	usearch --uchime_ref	--mindiv (0.8); mindiffs (3); --db Greengenes, gg13_8
Mapping original quality-filtered reads back to OTUs	UPARSE	usearch --usearch_global	--id (0.97)
Assigning taxonomy to OTUs using RDP Classifier	QIIME	assign_taxonomy.py	-t Greengenes gg13_8
Sequence alignment	QIIME	align_seqs.py	-e, --min_length (75); -p, --minimum_percent_id (0.75)
Filtering sequence alignment	QIIME	filter_alignment.py	- g, --allowed_gap_frac (0.999999)
Inferring phylogeny	QIIME	make_phylogeny.py	-
Generating a dense/sparse representation of an OTU table (observation matrix)	QIIME	make_otu_table.py	-e, --exclude_otus_fp
Diversity analyses	QIIME	core_diversity_analyses.py	-e, --sampling_depth (12,161); -c, --categories; -p, --parameter_fp (metrics: chao1, observed_species, Shannon, simpson, PD_whole_tree, unweighted_unifrac,weighted_unifrac,bray_curtis; single rarefaction depth: 12,161)

Supplemental Table 4.1. Composition and relative abundances of the genera in the even and staggered mock communities.

Genus	Species	16S rRNA Operon Count (operons per organism per µl)	
		HM-782D[#]	HM-783D[^]
Acinetobacter	*Acinetobacter baumannii*	100,000	10,000
Actinomyces	*Actinomyces odontolyticus*	100,000	1,000
Bacillus	*Bacillus cereus*	100,000	100,000
Bacteroides	*Bacteroides vulgatus*	100,000	1,000
Clostridium	*Clostridium beijerinckii*	100,000	100,000
Deinococcus	*Deinococcus radiodurans*	100,000	1,000
Enterococcus	*Enterococcus faecalis*	100,000	1,000
Escherichia	*Escherichia coli*	100,000	1,000,000
Helicobacter	*Helicobacter pylori*	100,000	100,000
Lactobacillus	*Lactobacillus gasseri*	100,000	10,000
Listeria	*Listeria monocytogenes*	100,000	10,000
Neisseria	*Neisseria meningitidis*	100,000	10,000
Propionibacterium	*Propionibacterium acne*	100,000	10,000
Pseudomonas	*Pseudomonas aeruginosa*	100,000	100,000
Rhodobacter	*Rhodobacter sphaeroides*	100,000	1,000,000
Staphylococcus	*Staphylococcus aureus*	100,000	100,000
Staphylococcus	*Staphylococcus epidermidis*	100,000	1,000,000
Streptococcus	*Streptococcus agalactiae*	100,000	100,000
Streptococcus	*Streptococcus mutans*	100,000	1,000,000
Streptococcus	*Streptococcus pneumoniae*	100,000	1,000

[#] Even mock community.

[^] Staggered mock community.

Supplemental Table 5.1. Top 40 most abundant families in penile microbiotas of heterosexually-active Black South African men.

Family	**All participants**	**HIV-negative men (N = 150)$^{\#}$**	**HIV-positive men (N = 88)***	**Oxygen requirement**
	% relative abundance	**% relative abundance (sd)**	**% relative abundance (sd)**	
Corynebacteriaceae	47.19	49.71 (28.61)	42.90 (26.76)	Facultative anaerobic
Prevotellaceae	6.56	5.58 (11.10)	8.22 (14.64)	Anaerobic
Unclassified *Clostridiales*	5.61	5.21 (11.22)	6.30 (10.52)	Unidentified
Porphyromonadaceae	4.94	4.47 (9.83)	5.73 (10.94)	Anaerobic
Staphylococcaceae	4.57	3.61 (5.09)	6.21 (9.50)	Facultative anaerobic
Bifidobacteriaceae	3.88	4.49 (10.98)	2.84 (7.66)	Anaerobic/Facultative anaerobic
Lactobacillaceae	3.81	3.92 (12.48)	3.64 (14.15)	Microaerophilic/Facultative anaerobic
Veillonellaceae	3.31	3.06 (4.68)	3.74 (4.99)	Anaerobic
Moraxellaceae	3.12	3.02 (8.26)	3.30 (9.35)	Aerobic
Flavobacteriaceae	2.55	2.58 (8.05)	2.51 (6.70)	Aerobic
Clostridiales Incertae Sedis XI	2.23	2.14 (2.67)	2.39 (2.90)	Anaerobic
Micrococcaceae	1.39	1.44 (4.09)	1.31 (3.70)	Aerobic
Leptotrichiaceae	1.13	0.97 (3.14)	1.42 (4.42)	Anaerobic/Facultative anaerobic
Unclassified Bacteria	0.85	0.91 (1.03)	0.74 (1.10)	Unidentified
Brevibacteriaceae	0.66	0.70 (1.69)	0.58 (1.00)	Aerobic
Intrasporangiaceae	0.60	0.82 (3.91)	0.23 (0.75)	Aerobic
Aerococcaceae	0.60	0.56 (0.73)	0.67 (1.09)	Facultative anaerobic
Unclassified *Actinomycetales*	0.57	0.65 (0.66)	0.44 (0.64)	Unidentified
Ruminococcaceae	0.55	0.57 (1.55)	0.51 (0.83)	Anaerobic/Facultative anaerobic
Rhodobacteraceae	0.55	0.46 (1.07)	0.70 (1.69)	Aerobic
Coriobacteriaceae	0.54	0.67 (1.92)	0.33 (0.52)	Anaerobic
Streptococcaceae	0.43	0.51 (2.75)	0.30 (0.60)	Facultative anaerobic
Dermabacteraceae	0.42	0.40 (0.51)	0.47 (0.52)	Facultative anaerobic
Xanthomonadaceae	0.41	0.37 (1.35)	0.47 (1.59)	Aerobic
Lachnospiraceae	0.36	0.33 (1.31)	0.41 (0.87)	Anaerobic
Microbacteriaceae	0.26	0.25 (0.57)	0.28 (0.54)	Microaerophilic/aerobic
Fusobacteriaceae	0.22	0.05 (0.33)	0.52 (2.48)	Anaerobic
Campylobacteraceae	0.18	0.17 (1.06)	0.21 (0.67)	Microaerophilic
Propionibacteriaceae	0.17	0.20 (0.48)	0.11 (0.39)	Microaerophilic/Facultative anaerobic
Comamonadaceae	0.15	0.12 (0.75)	0.21 (1.01)	Aerobic
Pseudomonadaceae	0.14	0.17 (1.32)	0.10 (0.24)	Aerobic
Unclassified *Bacteroidales*	0.11	0.18 (1.85)	0.01 (0.02)	Unidentified
Dietziaceae	0.11	0.14 (0.53)	0.06 (0.19)	Aerobic
Unclassified Proteobacteria	0.11	0.15 (1.67)	0.04 (0.24)	Unidentified
Neisseriaceae	0.11	0.09 (0.54)	0.15 (0.86)	Facultative anaerobic
Actinomycetaceae	0.10	0.09 (0.31)	0.11 (0.28)	Anaerobic/facultative anaerobic
Alcaligenaceae	0.09	0.07 (0.48)	0.13 (0.83)	Aerobic
Caulobacteraceae	0.09	0.07 (0.35)	0.12 (0.43)	Aerobic
Unclassified Bacteroidetes	0.08	0.06 (0.51)	0.12 (0.91)	Unidentified
Enterobacteriaceae	0.07	0.06 (0.23)	0.09 (0.32)	Facultative anaerobic

Abbreviations: sd – standard deviation.

$^{\#}$Circumcised men were 92.5% (135/146).

*Circumcised men were 97.6% (80/82).

Supplemental Table 5.2. Differentially abundant genera in men with *Corynebacterium*-dominated versus diverse penile microbiotas.

***Corynebacterium*-dominated penile microbiotas (CST-1)**	**Diverse penile microbiotas (CSTs 2-5)**
Corynebacterium, *Staphylococcus*, *Thauera*, *Enhydrobacter*, *Flavisolibacter*, *Brevibacterium*, *Blastococcus*, unclassified *Bacillales*, *Exiguobacterium*, *Eremococcus*, *Skermanella*, *Turicibacter*, unclassified *Lactobacillales*, *Gp*, *Planomicrobium*, *Leuconostoc*, unclassified *Dermabacteraceae*, *Microvirga*, *Facklamia*, *Propionibacterium*, unclassified *Planococcaeae*, *Anaerococcus*, unclassified Bacteria, *Neisseria*, unclassified *Oxalobacteraceae*, unclassified *Nocardioidaceae*, unclassified *Sphingomonadaceae*, *Anaerostipes*, *Delftia*, *Methylobacterium*, *Naxibacter*, *Zimmermannella*, *Nosocomiicoccus*, *Enterococcus*, *Dermabacter*, *Brachybacterium*, *Gemmiger*, *Clostridium sensu stricto*, *Bifidobacterium*, unclassified *Staphylococcaceae*, *Psychrobacter*, *Roseburia*, *Blautia*, *Jeotgalicoccus*, *Tessaracoccus*, *Salinicoccus*, unclassified *Neisseraceae*, *Sphingomonas*, *Ruminococcus*, *Paracoccus*, unclassified *Propionibacteriaceae*	*Prevotella*, unclassified *Clostridiales*, *Porphyromonas*, *Gardnerella*, *Negativicoccus*, *Sneathia*, *Dialister*, *Finegoldia*, unclassified Clostridia, *Saccharofermentans*, unclassified *Clostridiales Incertae Sedis XI*, *Atopobium*, unclassified *Fusobacteriacea*, *Campylobacter*, unclassified *Incertae Sedis XI*, *Micrococcus*, unclassified *Bacteroidales*, *Streptococcus*, *Slackia*, unclassified Bacteroidetes, *Pseudomonas*, unclassified *Porphyromonadaceae*, *Murdochiella*, unclassified *Micrococcaceae*, *Haemophilus*, *Hallella*, unclassified *Peptostreptococcaceae*, *Peptococcus*, *Parvimonas*, unclassified *Veillonellaceae*, *Sutterella*, unclassified *Coriobacteriaceae*, *Varibaculum*, unclassified *Prevotellaceae*, *Mobiluncus*, *Mycoplasma*, *Citrobacter*, *Aeromonas*, *Brucella*, unclassified Firmicutes, *Howardella*, *Pantoea*, *Pyramidobacter*, *Olsenella*, *Granulicatella*, unclassified *Aerococcaceae*, *Kocuria*, *Arcanobacterium*, unclassified *Brucellaceae*

Abbreviations: CST – community state type.

Lactobacillus-dominated penile microbiotas (CST-6) were excluded from these analyses because they were neither diverse nor dominated by *Corynebacterium*.

Only genera at logarithmic LDA scores >2.0 or <-2.0 ($p<0.05$) are shown.

Supplemental Table 5.3. Associations of male community state types (CSTs) with female sexual partners' CSTs.

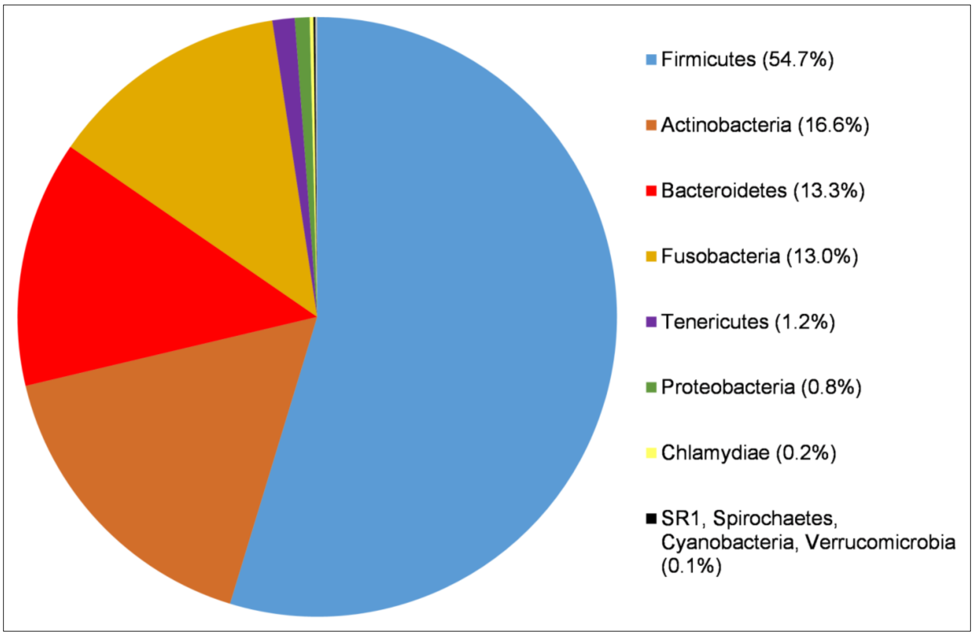
CST	*Lactobacillus*-dominated CVMs (% (n/N))	Non-*Lactobacillus*-dominated CVMs (% (n/N))	RR (95% CI) p-value[#]	*L. iners*-dominated CVMs (CST-3) (% (n/N))	Diverse CVMs (CST-8) (% (n/N))	RR (95% CI) p-value[#]
***Corynebacterium*-dominated (CST-1) vs. Non-*Corynebacterium*-dominated (CSTs 2-6)**						
Male CST-1	35.6 (21/59)	64.4 (38/59)	1.2 (0.9-1.7) 0.313	36.7 (18/49)	63.3 (31/49)	1.3 (0.9-1.9) 0.175
Male CSTs 2-6	26.5 (13/49)	73.5 (36/49)	Ref	23.9 (11/46)	76.1 (35/46)	Ref
***Corynebacterium*-dominated (CST-1) vs. Diverse (CSTs 2-5)**						
Male CST-1	35.6 (21/59)	64.4 (38/59)	1.4 (1.0-1.9) 0.082	36.7 (18/49)	63.3 (31/49)	1.5 (1.0-2.1) 0.054
Male CSTs 2-5	20.0 (9/45)	80.0 (36/45)	Ref	18.6 (8/43)	81.4 (35/43)	Ref
Highly diverse (CSTs 2-5) vs. Low-diversity (CSTs 1, 6)						
Male CSTs 2-5	20.0 (9/45)	80.0 (36/45)	0.5 (0.3-1.0) **0.030**	18.6 (8/43)	81.4 (35/43)	0.5 (0.3-1.0) **0.022**
Male CSTs 1, 6	39.7 (25/63)	60.3 (38/63)	Ref	40.4 (21/52)	59.6 (31/52)	Ref
***Lactobacillus*-dominated (CST-6) vs. Non-*Lactobacillus*-dominated (CSTs 1, 2-5)**						
Male CST-6	100.0 (4/4)	0.0 (0/0)	*	100.0 (3/3)	0.0 (0/0)	*
Male CSTs 1, 2-5	28.8 (30/104)	71.2 (74/104)		28.3 (26/92)	71.7 (66/92)	

Abbreviations: CST – community state type, CVMs – cervicovaginal penile microbiotas, RR – relative risk, CI – confidence interval.

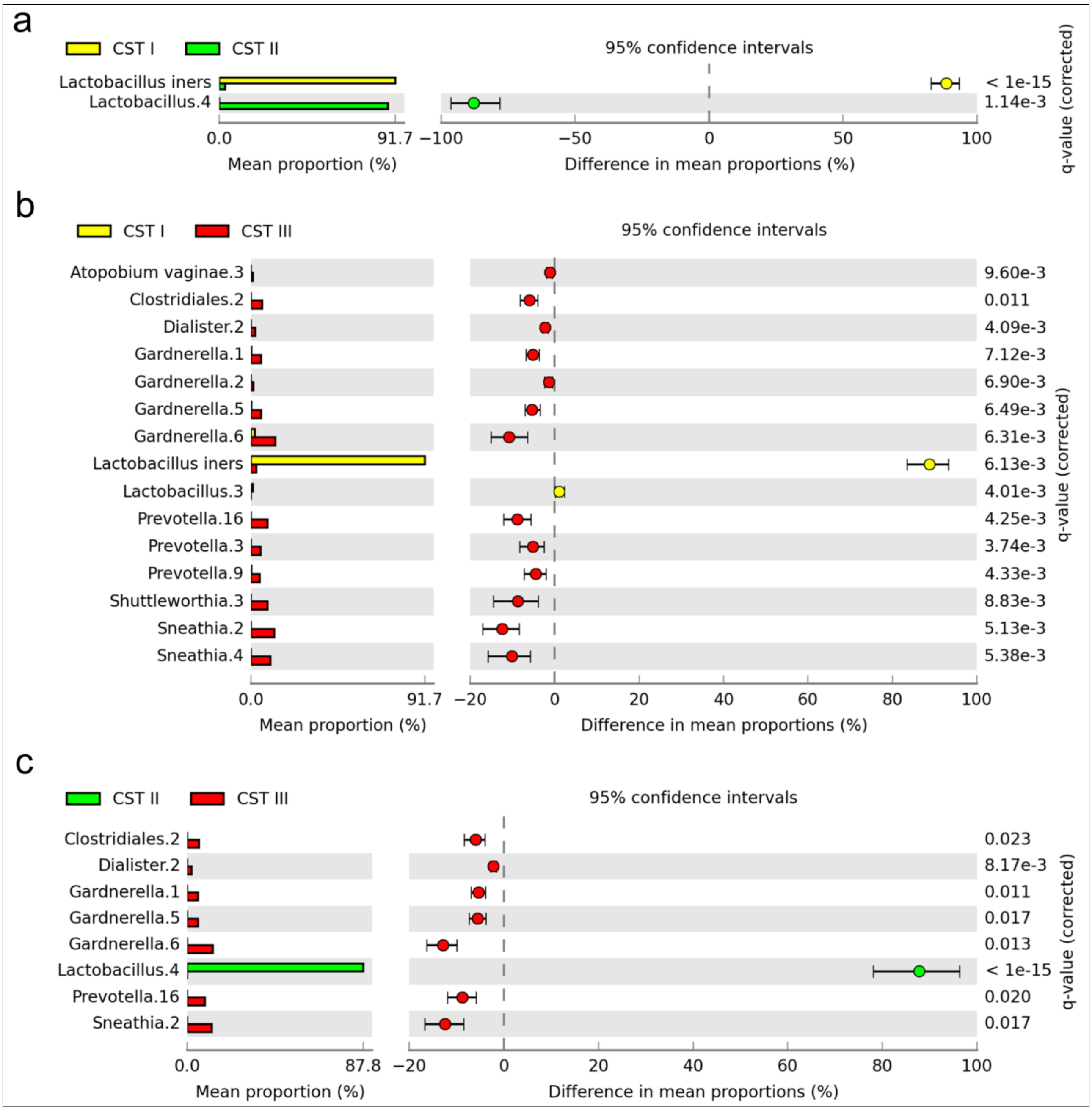
*Association of male CST-6 with female CSTs was not performed due to small number of men in CST-6.

[#]p-values were computed using Chi-square/Fisher's exact test. Significant p-values (<0.05) are shown in **bold**.

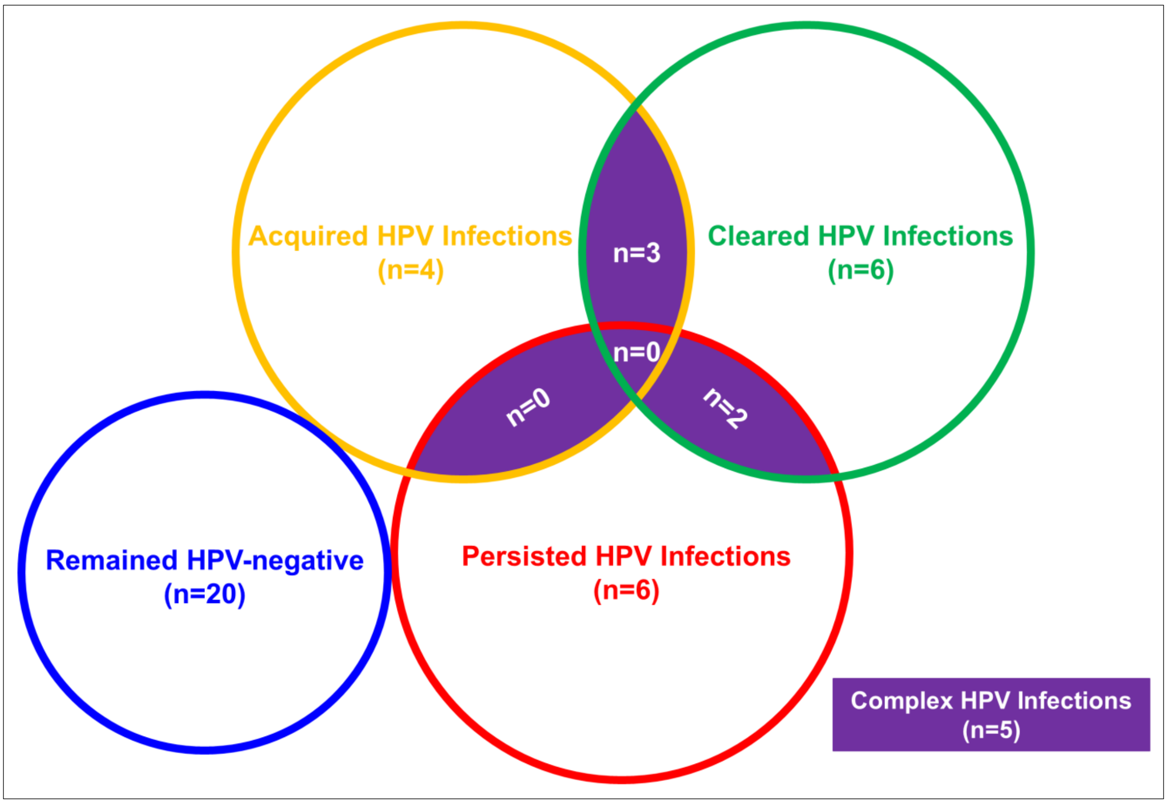
II. Supplemental Figures



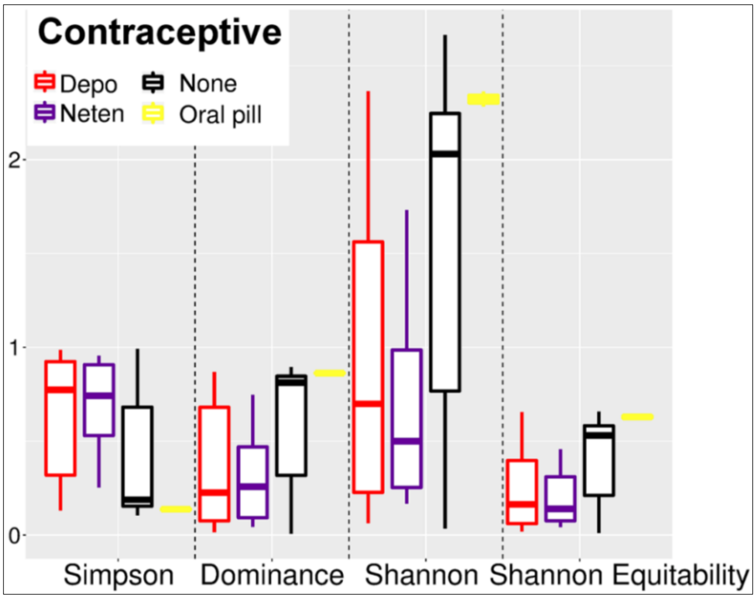
Supplemental Fig. 2.1. Proportions of the twelve phyla identified in the cervicovaginal microbiotas of the 62 women.



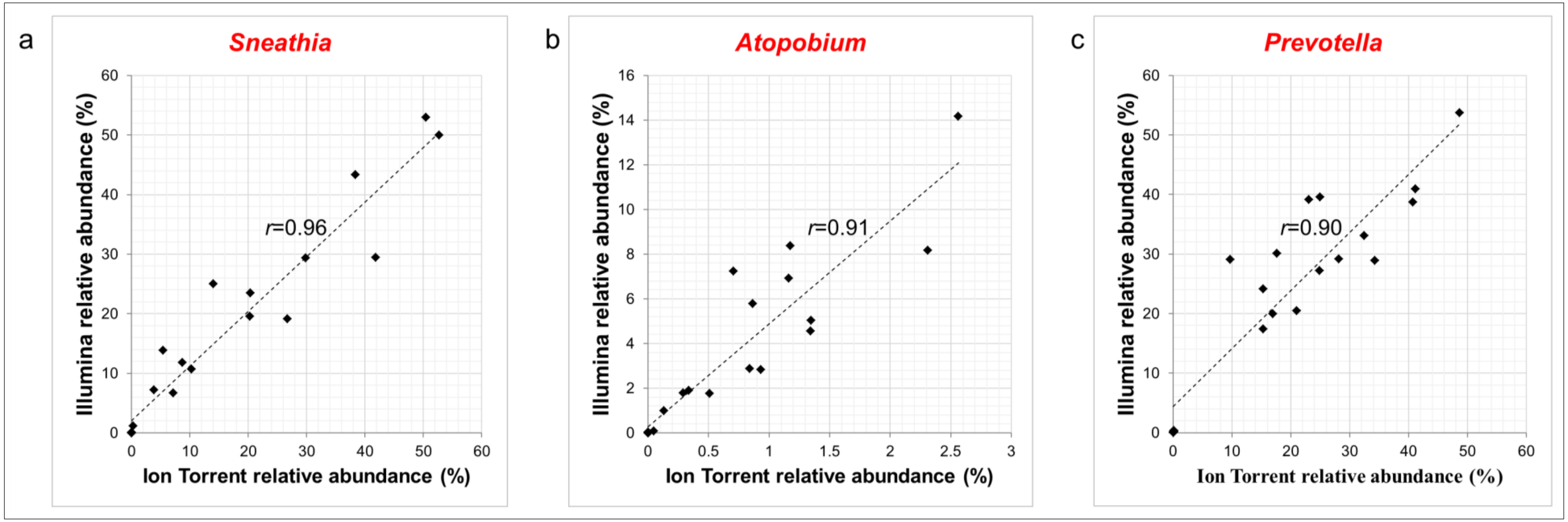
Supplemental Fig. 2.2. Extended error bar plots depicting differentially abundant bacterial taxa in each cervicovaginal community state type (CST). The p-values (adjusted by Benjamini Hochberg FDR correction to account for False Discovery Rates), effect size and confidence intervals (0.95, DP: bootstrap) as computed by White's non-parametric t-test (two-sided type) for each significant feature are indicated. Only bacterial taxa with greater than zero difference between their percentage proportions and ≤0.05 q-values and ≥1.0 effect size are shown by STAMP. a) CST I/CST II comparison, b) CST I/CST III comparison, and c) CST II/CST III comparison.



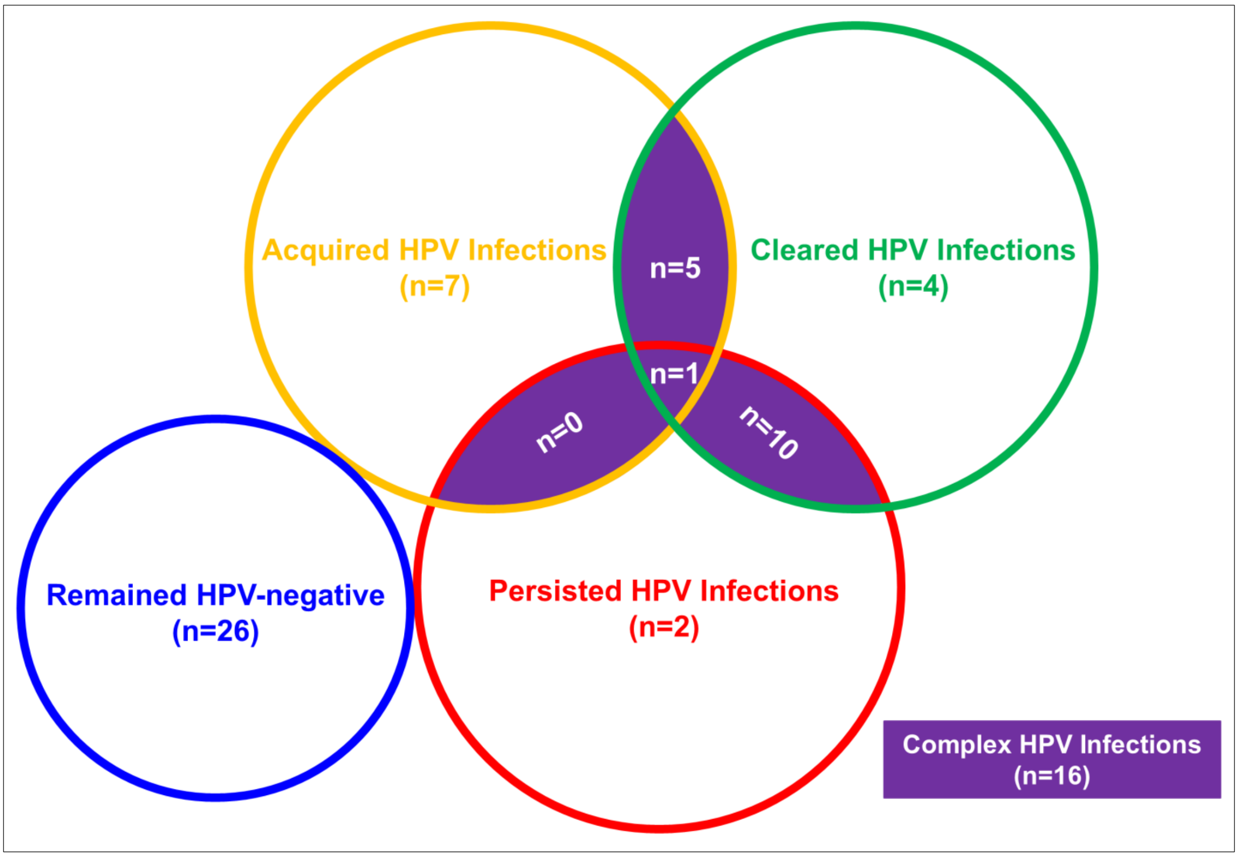
Supplemental Fig. 2.3. Distribution of the natural history of HPV infection at 6-month visit. The values in parentheses show the number of women in that particular HPV definition category.



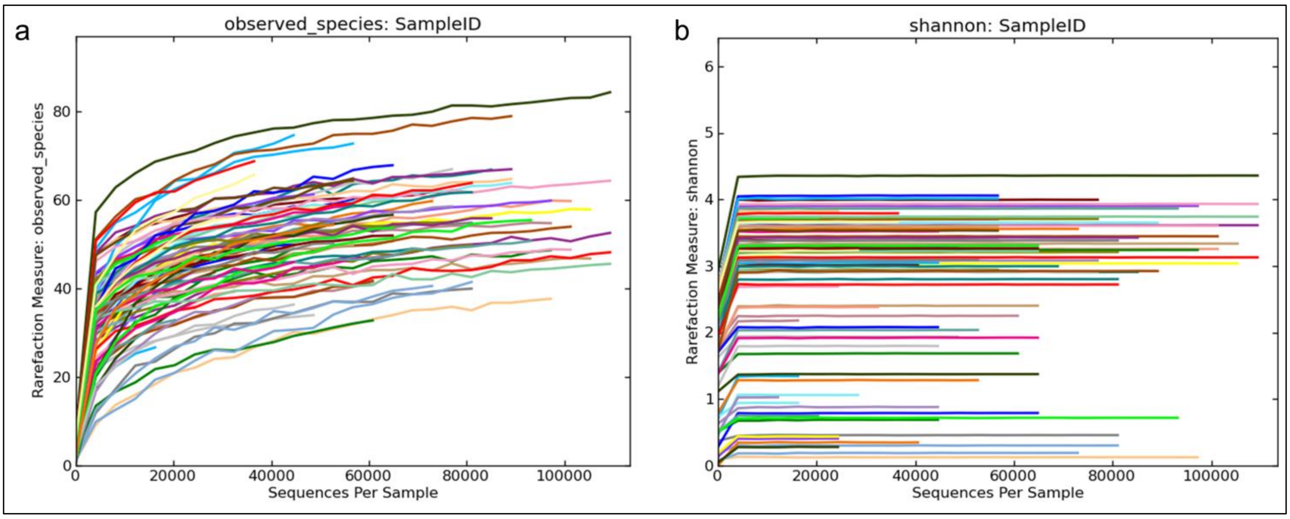
Supplemental Fig. 2.4. Alpha diversity grouped by hormonal contraception status and type of contraceptive. Each box plot is colour-coded according to the type of hormonal contraception. In each plot, the box ranges from the first to the third quartile, with the median represented by the line that divides the box into two. The whiskers extend to adjacent values within the lower and upper quartiles. Outliers are represented by the dots. Hormonal contraceptive users are categorised according to the type of contraceptive: oral (oral pills) or injectable (Depo: Depo-Provera and Neten: norethisterone enanthate) contraceptives.



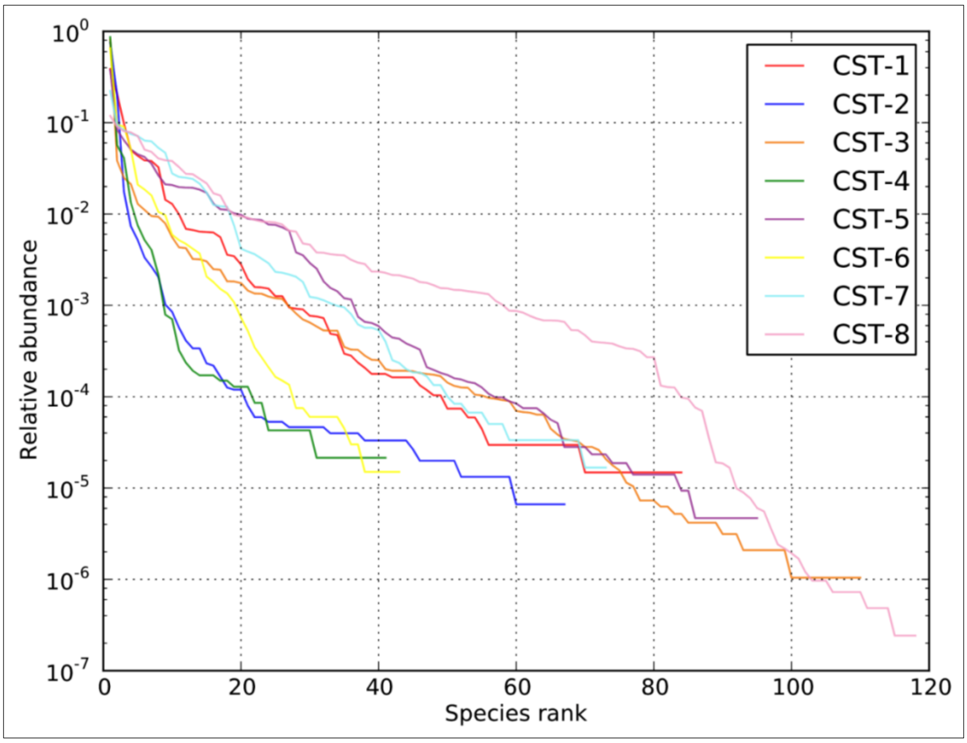
Supplemental Fig. 3.1. Comparison of the relative abundances of other genera with noticeable variation between Ion Torrent V4 and Illumina V3-V4 dataset. a) *Sneathia*, b) *Atopobium*, and c) *Prevotella*. The dashed line (trendline) is the line of identity (1=1). For each of these genera, the relative abundance was compared across the 19 samples in each dataset. Pearson's correlation (*r*) values are shown on the trendline. The relative abundance of each of the six highlighted genera from each method was compared across the 19 samples. Pearson's correlation (*r*) values are shown on the trendline.



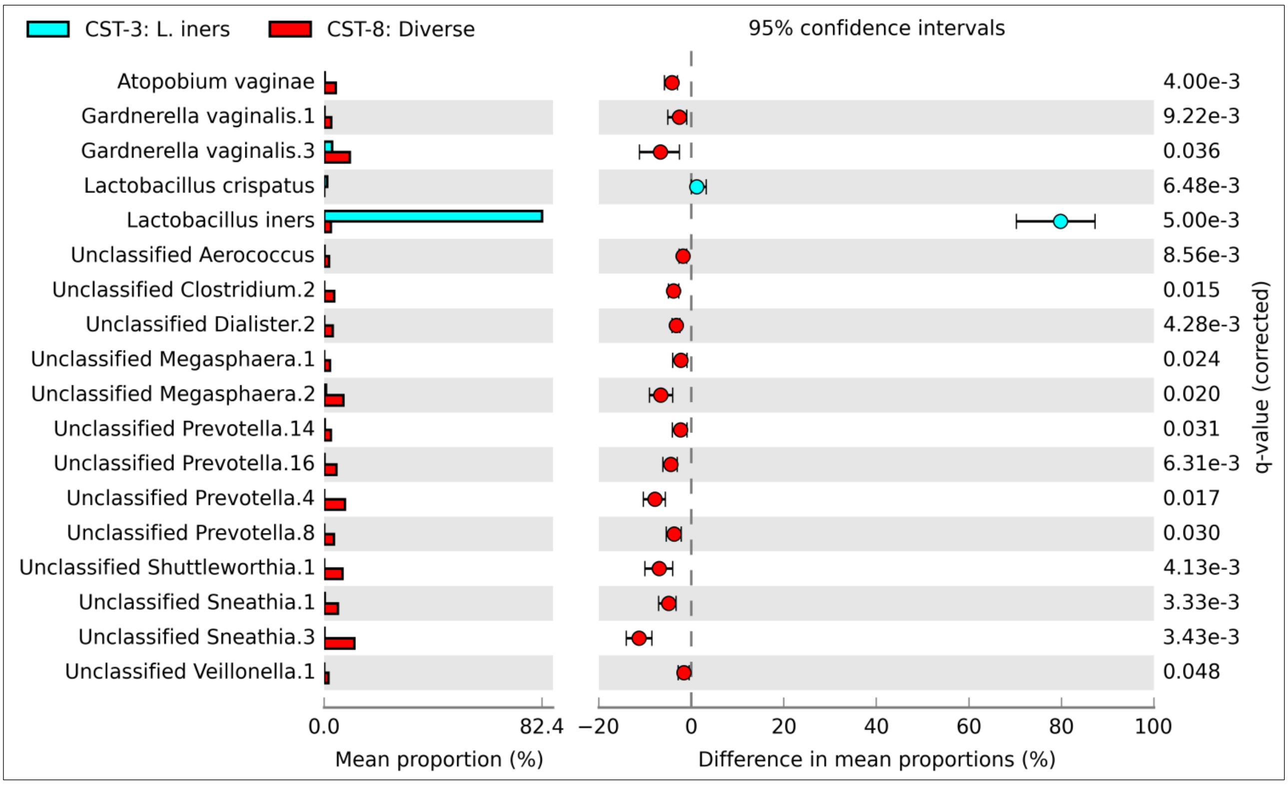
Supplemental Fig. 4.1. Distribution of the natural history of HPV infection at 6-month visit. The values in parentheses show the number of women in that particular HPV definition category.



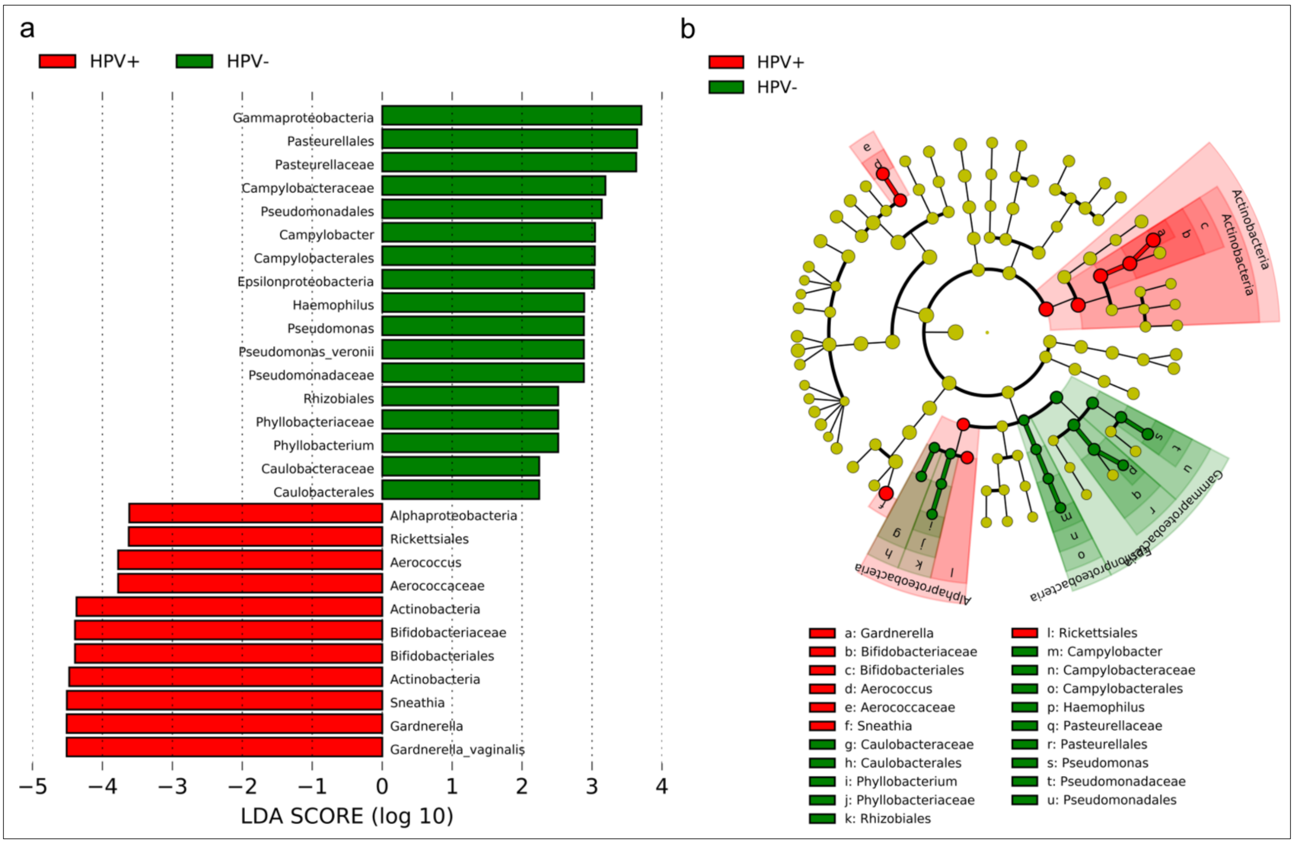
Supplemental Fig. 4.2. Rarefaction plots of the 87 cervicovaginal microbiotas. Rarefaction plot with a) observed_species and b) Shannon diversity estimators. The alpha diversity (community richness) was assessed at different sequencing depths. Each coloured-curve represents one of the eighty seven samples.



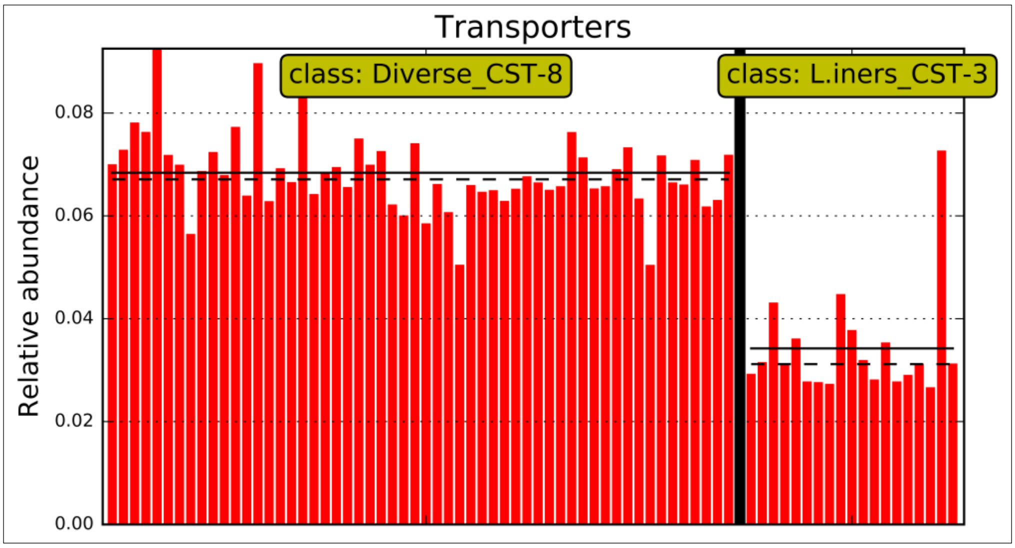
Supplemental Fig. 4.3. Rank abundance plots of bacterial taxa in the eight established community state types (CSTs). Rank abundance plots were used to estimate the bacterial diversity by measuring richness, dominance, and evenness. Richness is the distance a curve extends along the x-axis while dominance is the y-intercept. A higher value on the y-axis depicts increased dominance; hence, low diversity. Evenness is indicated by a low slope of the curve. Therefore, in a rank abundance plot, a high dominance, more richness and evenness in a bacterial community indicate a high ecological diversity.



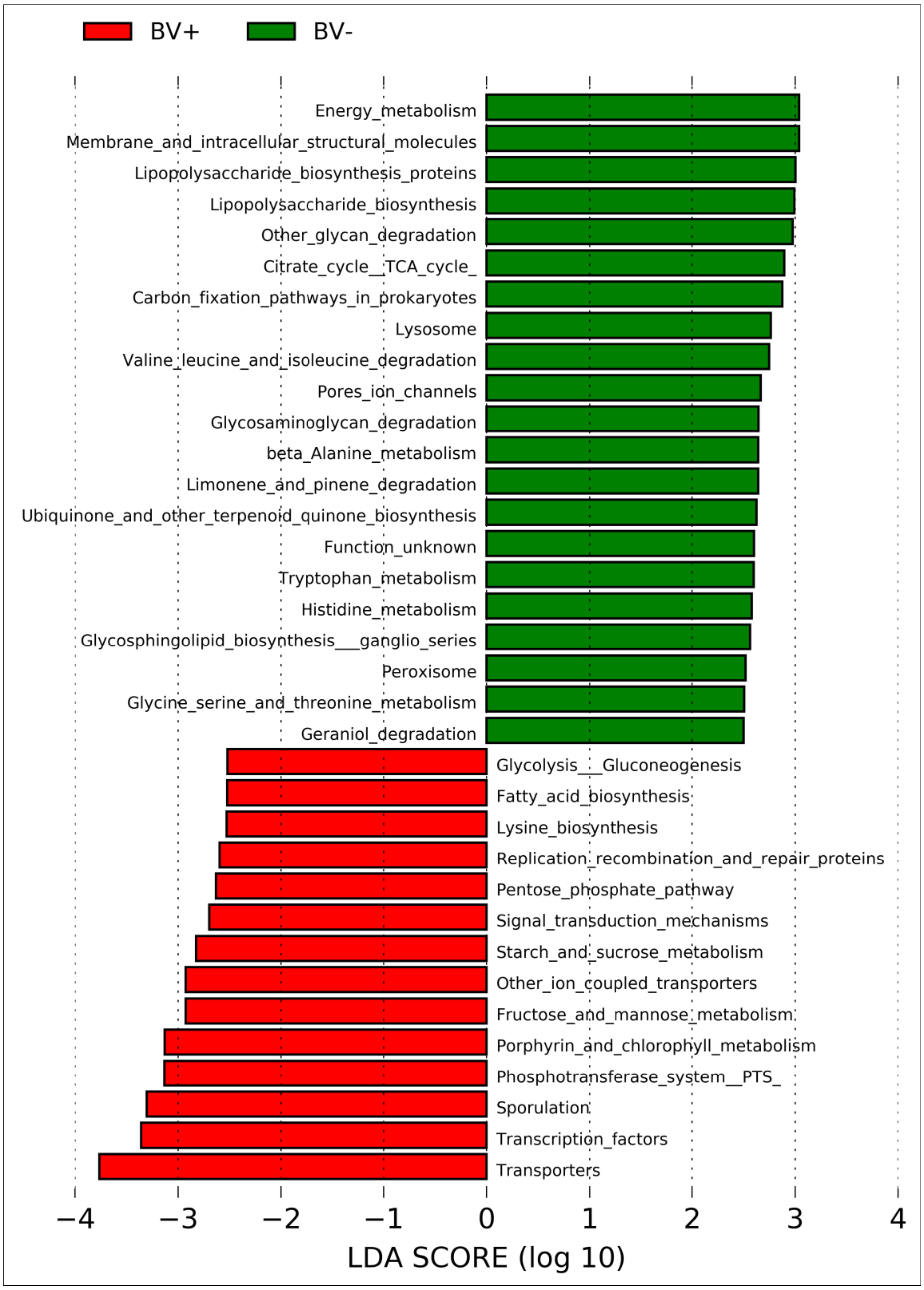
Supplemental Fig. 4.4. Extended error bar plots depicting enriched bacterial taxa in two community state types. Only features with greater than zero difference between their percentage proportions in CST-3 and CST-8 with an effect size ≥1.0 and q-value ≤0.05 are shown. The statistical comparison was carried out in STAMP. The p-values (adjusted by Benjamini Hochberg correction to account for False Discovery Rates), effect size and confidence intervals (0.95, DP: bootstrap) computed by White's non-parametric t-test (two-sided type) are indicated. The difference in mean proportions and 95% confidence intervals are shown on the bar plots.



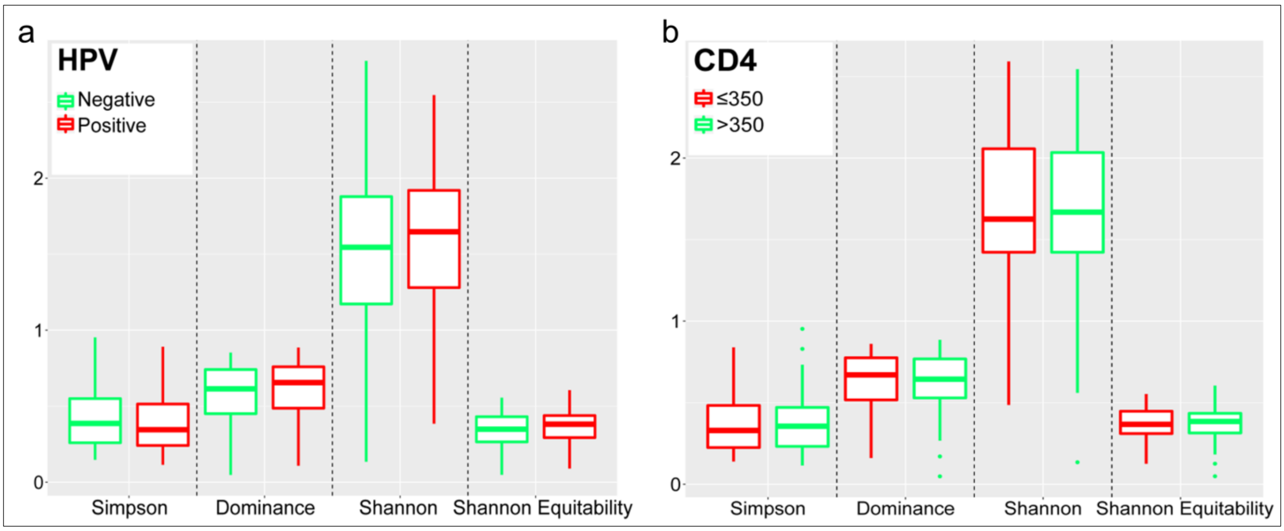
Supplemental Fig. 4.5. Potential biomarkers for HPV by LefSe. a) Histogram of differentially abundant taxa in cervicovaginal microbiotas of women with and without HPV infections, and b) A six-level cladogram with a taxonomic hierarchical structure. Each coloured solid represents a taxon and its diameter is proportional to the taxon’s relative abundance. Red and green solids represent statistically significant taxon ranks in HPV-positive and HPV-negative group, respectively. Only features with logarithmic LDA scores >2.0 or <-2.0 are shown.



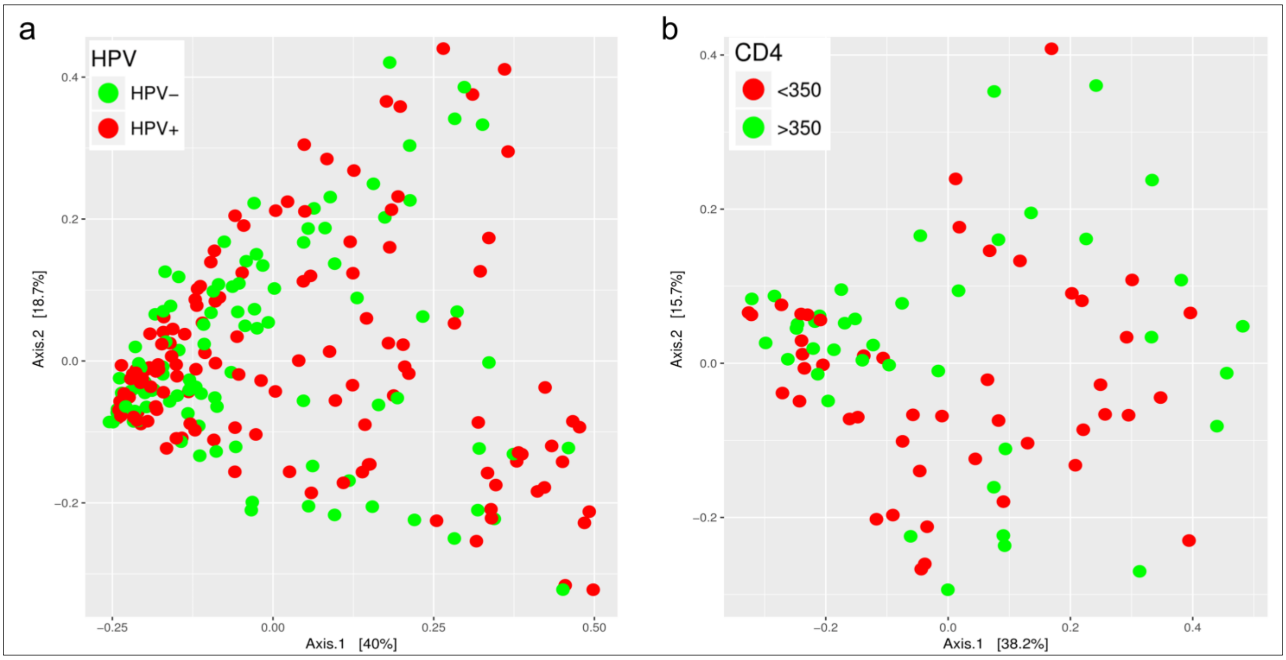
Supplemental Fig. 4.6. LefSe histograms of relative abundances of the KEGG module transporters in community state types (CSTs) dominated by *L. iners* (CST-3) and diverse bacteria (CST-8). Each bar plot represents the relative abundance of the predicted transporter in each sample. The mean and median relative abundances of these transporters are shown by solid and dashed lines, respectively.



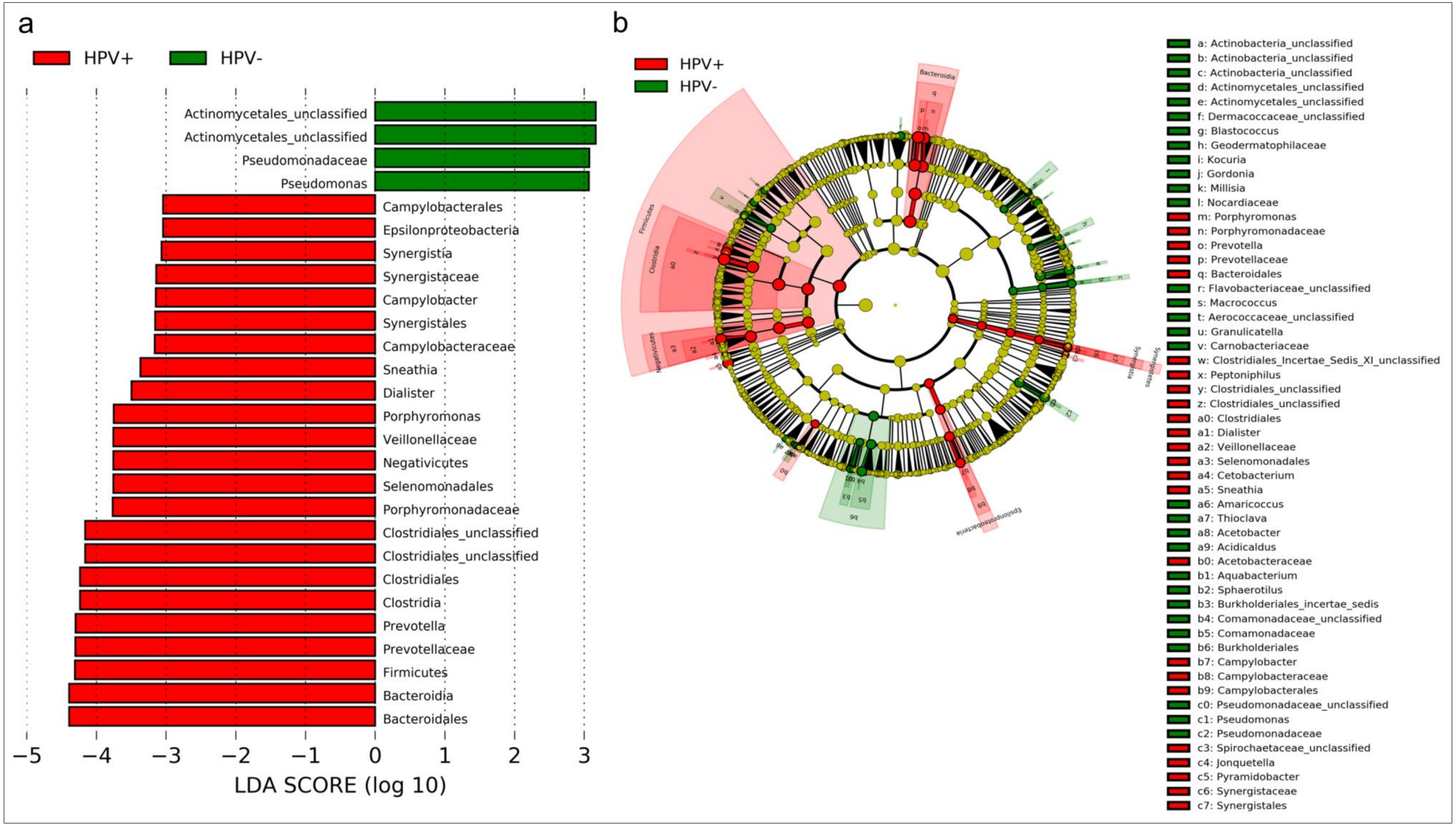
Supplemental Fig. 4.7. Histogram of the differentially enriched KEGG pathway and metabolic functional modules in women with and without bacterial vaginosis (BV). At LDA score >2.0, thirty three and 45 modules were significantly abundant in women with and without BV, respectively. Only modules with logarithmic LDA scores >2.5 or <-2.5 are shown.



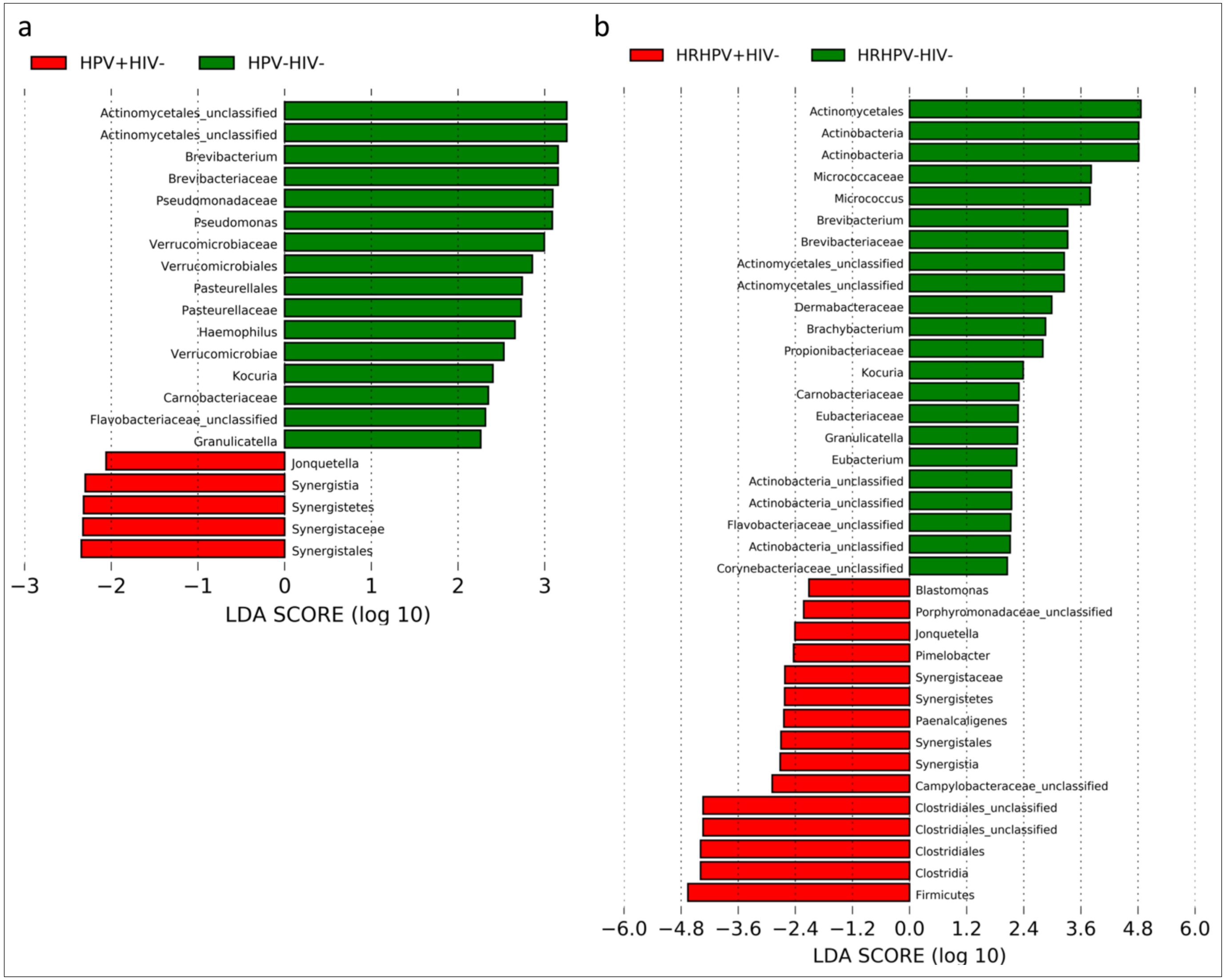
Supplemental Fig. 5.1. Alpha diversity measures of penile microbiotas. Comparison of alpha diversity measures of penile microbiotas grouped a) human papillomavirus (HPV) infection status, and b) CD4$^+$ T-cell count status. In the plot, the boxes range from the first to the third quartile, with the median represented by the horizontal line. The whiskers extend to the smallest and largest non-outliers and outliers are represented by dots.



Supplemental Fig. 5.2. Beta diversity of penile bacterial communities. Principal Coordinates Analysis (PCoA) plots of the weighted UniFrac distances of the penile microbiotas coloured according to a) human papillomavirus (HPV) infection status and b) CD4$^+$ T-cell count status. The two principal coordinate axes of variations (Axis.1: 40.0% and Axis.2: 18.7% (a) and Axis.1: 38.2% and Axis.2: 15.7% (b)) are shown. Each solid point is a bacterial community.



Supplemental Fig. 5.3. Potential biomarkers for HPV by LefSe. a) Histogram of differentially abundant taxa in penile microbiotas of men with and without HPV infections, and b) A six-level cladogram with a taxonomic hierarchical structure. Each coloured solid represents a taxon and its diameter is proportional to the taxon's relative abundance. Red and green solids represent statistically significant taxon ranks in HPV-positive and HPV-negative group, respectively. For the purposes of better visualization of the histogram, only features with logarithmic LDA scores >3.0 or <-3.0 are shown. For the cladogram, features are shown at logarithmic LDA scores >2.0 or <-2.0.



Supplemental Fig. 5.4. Potential biomarkers for HPV and high-risk HPV by LefSe. Histogram of differentially abundant taxa in penile microbiotas of HIV-negative men with and without a) HPV infections, and b) high-risk (HR)-HPV infections. Only features with logarithmic LDA scores >2.0 or <-2.0 are shown.