The association between the oral and vaginal microbiome of young South African women

By Rachel Esra
ESRRAC001

A dissertation submitted in the fulfilment of the requirements for the degree of
Master in Clinical Science and Immunology
Faculty of Health Science
February 2019

Supervisor: Dr. Heather Jaspan
Co-Supervisor: Christina Balle
Department of Pathology/Virology

University of Cape Town

Word Count:

University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

Plagiarism Declaration

I, Rachel Tal Esra, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work or any part thereof has been, is being or will be submitted for another degree in this or any other university.

I empower the university to reproduce, for the purposes of research, either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: 11 February 2019

Acknowledgements

I would like to acknowledge the substantial work by my co-supervisor Christina Balle in setting up and validation the DNA extraction, library preparation and bioinformatics pipeline used in this study. Christina, thank you for your guidance, incredible patience and all of the time you have put aside for me during this study.

To my supervisor Dr. Heather Jaspan thank you for your understanding and support that have allowed me to pursue my external work in addition to this study.

To my colleagues in the Jaspan lab – thank you for all the assistance and advice over the years. To Enock – thank you for all the help with lab work and experiment validation.

To my friends and family for your continued encouragement and interest in my research.

Finally, to my parents, thank you for your unwavering support throughout my studies.

Abstract

Bacterial vaginosis (BV) and periodontal disease (PD) are conditions characterised by reduction of healthy bacterial communities in the vaginal and oral microbiomes respectively. Both BV and PD are associated with an increased risk of preterm labour and negative birth outcomes, yet it is unknown whether PD and BV are independent risk factors or may be interrelated. Understanding the health risks associated with pregnancies in young women is critical for developing new preventative interventions and for informing guidelines. Current knowledge of what constitutes a healthy microbiome is largely based on North American studies and may not be applicable to the South African population. This study characterises the oral and vaginal microbiome of South African female adolescents and investigates the association between alterations in oral bacterial diversity and BV in young South African women. DNA was extracted from matched lateral vaginal wall, saliva and periodontal samples and V4 16S sequencing was performed using MiSeq technology. The composition of the core oral microbiome of South African female adolescents was found to be similar to descriptive studies published in other populations. We additionally report a description the vaginal microbiome that is in agreement with previous studies in the South African population. PD-associated bacterial species were enriched in the oral microbiome of women with clinically diagnosed BV and in those with *Lactobacillus iners* dominant vaginal community types (VCTs) compared to asymptomatic women and those with *L. crispatus* dominated VCTs respectively. While this data provides evidence in support of a relationship between oral and vaginal dysbiosis, it unclear in which compartment bacterial dysbiosis would originate, should the association holds true.

Table of Contents

Plagiarism Declaration ..... *i*

Acknowledgements ..... *ii*

Abstract ..... *iii*

List of Figures *6*

List of tables *7*

List of Abbreviations *8*

CHAPTER ONE: Literature review *9*

1.1 The human microbiome 9

1.1.1 Introduction 9
1.1.2 Culture-independent bacterial classification 10
1.1.3 Bioinformatic analysis 13

1.2 The oral microbiome 14

1.2.1 Defining a core microbiome 14
1.2.2 Ecological niches within the oral cavity 15
1.2.3 The oral microbiome in health and disease 17

1.3 The vaginal microbiome 19

1.3.1 Defining a core vaginal microbiome 19
1.3.2 Bacterial vaginosis (BV) 20
1.3.3 Redefining an optimal vaginal microbiome 21

1.4 PD, BV and pre-term birth (PTB) 22

1.4.1 BV and PTB 23
1.4.2 PD and PTB 25
1.4.3 Evidence for a relationship between PD and BV 28

1.5 Oral and vaginal microbiome of the South African population 29

CHAPTER TWO: Aims & Objectives *31*

CHAPTER THREE: Materials and Methods *32*

3.1 Sample Collection 32

3.1.1 Study Cohort 32
3.1.2 Ethical issues and risk assessment 33
3.1.3 Oral and vaginal sample collection 34

3.2 STI testing 36

3.3 16S rRNA gene V4 DNA extraction and amplification 37

3.3.1 Previously validated 16S rRNA gene V4 DNA extraction and validation protocol 37
3.3.2 Optimization of of 16S rRNA gene V4 region DNA extraction and validation protocol 41
3.3.3 Summary of findings and resultant protocol changes from 16S rRNA gene V4 DNA extraction optimization 51

3.4 Illumina 16S rRNA iTag library preparation 52

Deleted:

Deleted:

Deleted:

Deleted:

3.4.1 Previously validated Illumina 16S rRNA iTag library preparation protocol........52
3.4.2 Optimisation of Illumina 16S rRNA gene MiSeq library preparation protocol........56

3.5 MiSeq Sequencing........59

3.6 Bioinformatic pipeline........60

3.7 Statistical analysis........61

CHAPTER 4: Results........62

4.1 Library quality assessment........62
4.1.1 Rarefication and selection of samples based on read counts........62
4.1.2 Quality and compositional analysis of controls........62

4.3 Composition of the oral microbiota........68
4.3.1 Ecological niche and community type descriptions of the oral microbiota........68
4.3.2 Description of the core oral microbiota........73

4.4 Description of the vaginal microbiome........75

4.5 Comparison of the oral and vaginal microbiota........80

4.6 Association between oral and vaginal microbial dysbiosis........85

CHAPTER 5: Discussion........90

5.1 Study design and methodology........90

5.2 Development of a sample processing protocol for stored samples with low DNA yields........92

5.3 Characterisation of the oral microbiota........92

5.4 Characterisation of the vaginal microbial dysbiosis in South African female adolescents........95

5.5 The relationship between oral and vaginal microbial dysbiosis........98

5.6 Conclusions, relevance and further work........99

Reference List........103

Deleted:
Deleted:
Deleted:
Deleted: 9
Deleted:
Deleted:
Deleted:
Deleted:
Deleted:
Deleted:

List of Figures

Chapter One:

Figure 1.1: Differential composition of bacterial communities in different anatomical sites in the human body…9
Figure 1.2: Illumina® Miseq workflow……………………………………………………………………………………………………11
Figure 1.3: Phylogenetic visualization of 16S data in the CORE database…………………………………………………14
Figure 1.4: Comparison of composition of the salivary and periodontal microbiome………………………………15,16
Figure 1.5: Hypothesized causal pathways underlying the observation association between BV, PD and PTB……21

Chapter Two:

Figure 2.1: Instruction of saliva sample collection using the Salivette (Sarstedt) collection device……………………34
Figure 2.2: Demonstration of vaginal samples collection after the insertion of a gynaecological speculum……….35
Figure 2.3: Gel electrophoresis of amplified 16S V4 DNA extracted from stored oral samples using the Powersoil™ DNA Isolation Kit (MoBio)………41
Figure 2.4: Gel electrophoresis visualisation of PCR products of amplified 16S rRNA gene V4 product from fresh oral samples using the Powersoil™ DNA Isolation Kit (MoBio)…………………………………………………………………43
Figure 2.5: Gel electrophoresis visualisation of amplified 16S V4 DNA extracted from stored oral samples using the Powersoil™ DNA Isolation Kit (MoBio)……………………………………………………………………………………………45
Figure 2.6: Gel electrophoresis visualisation of amplified 16S V4 DNA extracted from stored oral samples using the Powersoil™ DNA Isolation Kit (MoBio) ……………………………………………………………………………………………46
Figure 2.7: Gel electrophoresis visualisation of amplified 16S V4 DNA extracted from stored oral samples using the Powersoil™ DNA Isolation Kit (MoBio)……………………………………………………………………………………………49
Figure 2.8: Workflow for PCR amplicon purification using the Agencourt AMPure XP PCR purification system
Figure 2.9: Size of denatured RNA fragments (bp) in initial DNA library measured by fluorescence [FU] using the Bioanalyzser 2100 (Agilent)………53
Figure 2.10: : Gel electrophoresis visualisation of PCR optimisation……………………………………………………………56

Chapter Three:

Figure 3.1: Schematic of sample selection after sequencing quality assessment and rarefication curves..…………62
Figure 2.2: Composition and relative abundance of sequencing controls…………………………………………………..…64
Figure 3.3: Analysis for technical bias introduced by DNA extraction methods and MiSeq sequencing run……….65
Figure 3.4: Ecological analysis of the oral microbiota…………………………………………………………………………….68
Figure 3.5: Bacterial composition of oral microbiota………………………………………………………………………………69
Figure 3.6: Taxonomical differences between oral community types and ecological niches. ……………………………71
Figure 3.7: Core analysis of the oral microbiota……………………………………………………………………………………73
Figure 3.8: Ecological diversity of vaginal microbiota by clinical BV status and vaginal community types………….76
Figure 3.9: Composition of vaginal bacterial community types…………………………………………………………………77
Figure 3.10: Differentially abundant bacterial taxa between vaginal community type. ………………………………78,79
Figure 3.11: Microbial communities in the vagina, saliva and periodontal space……………………………………………81
Figure 3.12: Microbial composition of oral and vaginal microbiota. ……………………………………………………………82
Figure 3.13: Ecological diversity of oral microbiota by clinical BV status and vaginal community type. ………..…87
Figure 3.14: Top classification features in random forest modelling using oral bacterial composition to predict clinical BV status and vaginal community type………………………………………………………………………………………89

List of tables

Chapter Two

Table 2.1: Reagent list for validation PCR..40

Table 2.2: Thermocycler protocol for validation PCR...40

Table 2.3: Concentration of dSDNA extracted from fresh oral and stored vaginal samples....................................43

Table 2.4: Concentration of dsDNA extracted from fresh oral and stored vaginal samples.....................................44

Table 2.5: Concentration of dsDNA extracted from fresh oral and stored vaginal samples.....................................46

Table 2.6: Concentration of dsDNA extracted from fresh oral and stored vaginal samples.....................................49

Table 2.7: Reagent list for PCR 2...54

Table 2.8: Thermocycler protocol for PCR 2..54

Table 2.9: Reagent list for PCR 2 optimisation...57

Chapter Three

Table 3.1: Cohort baseline characteristics...66

Table 3.2: Mean relative abundance of top 30 OTUs in oral community types..70

Table 3.3: Mean relative abundance of core oral taxa...74

Table 3.4: Mean relative abundance of top 30 most prevalent OTUs in all oral samples..83

Table 3.5: Mean relative abundance of top 30 most prevalent OTUs in all vaginal samples....................................84

Table 3.6: Distribution of clinical BV status and VCTs in OCTs...85

Table 3.7: Results of a negative binomial model detecting significant differences in the relative abundance of oral bacterial taxa between gparticipnats with and without and clinical BV diagnosis and differing vaginal community types..88

List of Abbreviations

BLAST	Basic Local Alignment Search Tool
BV	Bacterial Vaginosis
CDC	Centre for Disease Control
CHAMPS	Choices For Adolescent Prevention Methods for South Africa
CT	Community type
DNA	Deoxyribonucleic acid
DOH	Department of Health
DTHF	Desmond Tutu HIV Youth Foundation
GCP	Good Clinical Practice
HIV	Human Immunodeficiency Virus
HSV-2	Herpes Simplex Virus 2
LW	Vaginal lateral wall swab
NCBI	National Centre for Biotechnology Information
NGS	Next Generation Sequecing
NICD	National Institute of Communicable Disease
NMDS	Non-metric multidimensional scaling
OCT	Oral community type
OTU	Operational taxonomic unit
PCoA	Principal component analysis
PCR	Polymerase chain reaction
PD	Periodontal disease
PE	Periodontal swab
PROM	Premature rupture of membranes
PTB	Pre-term birth
rRNA	Ribosomal ribonucleic acid
SAL	Saliva sample
STI	Sexually transmitted infection
VCT	Vaginal community type

CHAPTER ONE: Literature review

1.1 The human microbiome

1.1.1 Introduction

The composition of the bacterial microbiome, the collection of bacterial microorganisms in the human body, has an important influence on human health (reviewed by Cho and Blaser 2012). Bacterial communities residing in different ecological niches within the human body play a large part in shaping immunity, development, physiology and nutrition. These distinct anatomical sites in the human body provide unique microhabitats supporting the growth of distinct microbial communities (Figure 1.1, Cho and Blaser 2012). It was originally understood that the human microbiome is acquired only after birth, but recent studies identifying low abundance microbial communities within the placenta have challenged the paradigm that the foetus exists in a sterile environment *in utero* (Aagaard et al. 2014). The infant microbiome is seeded from the maternal vagina and faeces (Mändar and Mikelsaar 1996) and mode of delivery (vaginal or via caesarean section) plays a role in the initial composition of the infant microbiome (Dominguez-bello et al. 2010; Chu et al. 2017).

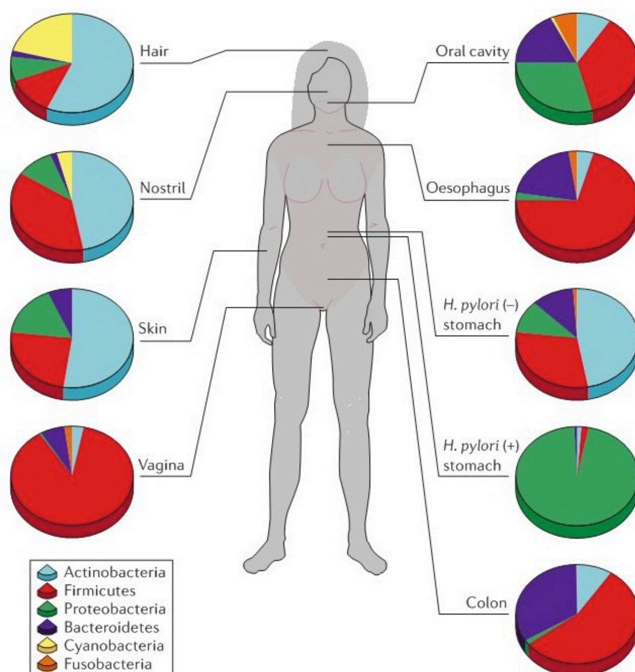


Figure 1.1: Differential composition of bacterial communities in different anatomical sites in the human body. Bacterial communities sampled from the same anatomical site from different individuals have been found to be more similar to each other than bacterial communities from different sites in the same individual (Sonnenburg and Fischbach, 2011). At the phylum level, bacterial communities have been found to be longitudinally stable within the same anatomical sites amongst different individuals. These stable bacterial communities have been described as "core" or "optimal" microbiomes. Core microbiomes have a mutualistic relationship with the human host and play a large role in maintaining host health. *Adapted from: Cho, I., & Blaser, M. J. (2012). The human microbiome: at the interface of health and disease. Nature Reviews Genetics, 13(4), 260.*

The composition of these microbial communities changes throughout our lifetime in response to hormonal fluctuations, host genetics and environmental exposures (The Human Microbiome Project Consortium 2012). The majority of microbiota play a mutualistic role within the human body, surviving off nutrients provided by the human host, while providing a first line of defence from opportunistic pathogens (Costello et al. 2012). Commensal bacteria maintain host health by competitively excluding invasive species from colonizing ecological niches and producing defence factors that hinder a pathogen's ability to multiply and/or produce virulence factors (Costello et al. 2012). Microbial dysbiosis, characterised by the reduction of mutualistic indigenous bacterial communities, results in the outgrowth of opportunistic pathogens causing chronic infections and potentially life threatening disease (Lloyd-Price et al. 2016). The 'core' microbiome refers to the most stable microbiota within an ecological niche that are shared by the majority of healthy individuals, whilst the 'variable' microbiome is determined by individual genetics, life-style and environmental exposures (Turnbaugh et al. 2007). There is a global interest in defining a 'healthy' or 'optimal' microbiome in order to better understand how deviations from stable microbial communities impact human health and disease.

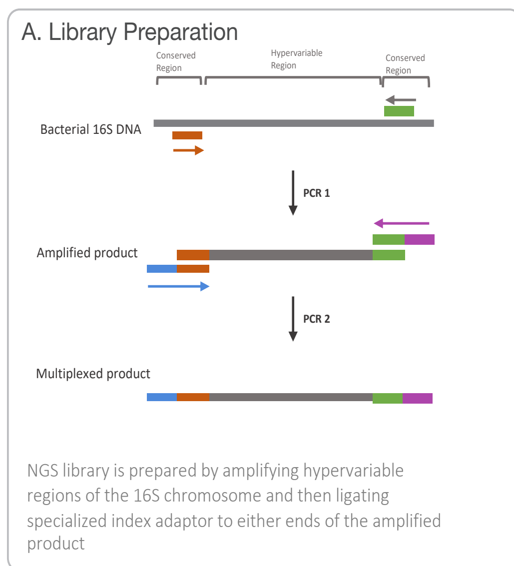
1.1.2 Culture-independent bacterial classification

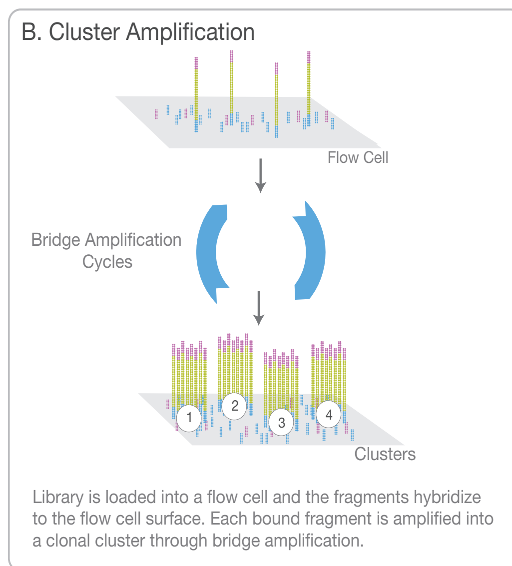
Historically, our knowledge of the composition of the human microbiome was limited to the <0.01% of bacterial species? that are amenable to laboratory cultivation (Amann et al. 1995). The advent of culture-independent approaches has resulted in a rapid deepening of our understanding of the composition and ecological function of the human microbiome. Advances in next-generation genetic sequencing of the 16S hypervariable region of the bacterial ribosomal gene has significantly expanded our knowledge of bacterial diversity (Clarrdige 2004). This region of the bacterial chromosome has been exploited in describing bacterial diversity due to its unique

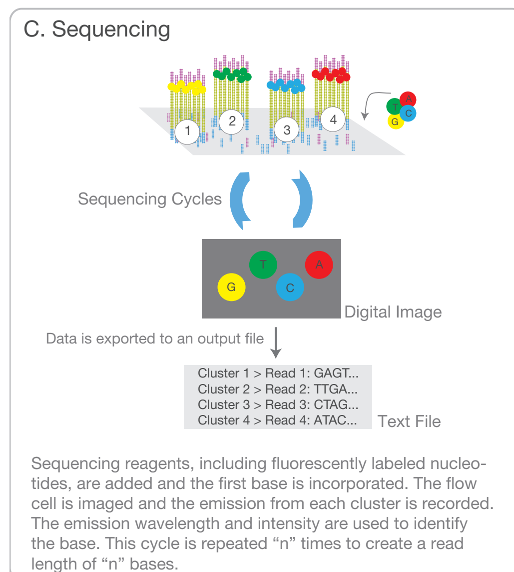
combination of evolutionary conserved regions alongside hypervariable regions - allowing for species level classification of bacterial phyla (Van de Peer et al. 2013). There are nine distinct hypervariable regions within the 16S region, numbered V_1 to V_9 (Chakravorty et al. 2008). Different variable regions have been shown to exhibit higher levels of variability for specific bacterial phylum and to provide varying amounts of species level resolution for selected bacterial families. V_1 demonstrates high level resolution for *Streptococcus* and *Staphylococcus* species, V_2 for *Mycobacterial* species and V_3 for *Haemophilus* species (Chakravorty et al. 2008). The $V_{4\text{-}6}$ region is conventionally used as representative barcode for the majority of bacterial phyla (Yang et al. 2016), although studies have found difficulty in identifying *Fusobacterium* species in this region (Kumar et al. 2011).

The development of high throughput next-generation sequencing (NGS) platforms has revolutionized genetics and genomic research, making high resolution, species level classification of non-culturable bacteria both possible and affordable. While a number of NGS platforms have been developed over the past ten years (Reviewed by Lin Liu et al. 2012), Illumina® sequencing-by-synthesis technology has emerged as an industry leader. In comparison to other NGS platforms, Illumina® technology delivers DNA sequencing data with the highest accuracy and lowest error rates per reads (Ross et al. 2013; Nakazato et al. 2013). The Illumina® MiSeq desktop sequencer, developed for research laboratories and clinical diagnostics, provides rapid and high quality 16S data for small scale studies (Liu et al. 2012). For the purpose of 16S sequencing, DNA libraries are prepared using polymerase chain reaction (PCR) to amplify the 16S bacterial chromosome using universal primers targeting the highly conserved regions alongside the hypervariable regions. Library preparation involves two rounds of PCRs; the first usually amplifies the desired region of 16S chromosome and the second to ligate specialized Illumina adapters containing index sequencing to the end of PCR amplicons (Figure 1.2). Indexed amplicons are pooled and loaded onto a flow cell where the amplicons hybridize to the flow cell's surface by the complementary chemistry of the ligated index barcodes. Each fragment is then amplified into distinct, clonal clusters through bridge amplification. Clonal clusters are then sequenced using proprietary sequencing-by-synthesis incorporating fluorescently labeled reversible-terminator

nucleotides. Four fluorescently labeled nucleotides, complementary to the cluster fragments, are incorporated into a DNA strand. With each labeled nucleotide incorporated, a fluorescent signal specific to the DNA base incorporated is emitted and recorded. The signals are recorded as fragment reads and filtered by the MiSeq sequencer to remove low quality reads. Reads from specific samples are identified by their unique index sequences. Raw sequence reads are then aligned to a reference genome using bioinformatics software (Figure 1.2).







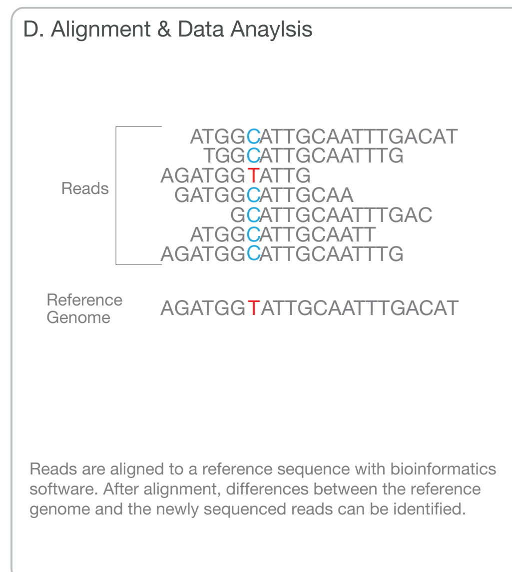


Figure 1.2: Illumina® Miseq workflow. (A) The bacterial 16S chromosome is amplified using PCR with primers targeting the evolutionarily conserved regions that flank hypervariable stretches. A second round of PCR ligates Ilumina® MiSeq adaptors containing unique barcodes to amplified products. **(B)** Pooled libraries are amplified on flow cells using bridge amplification to produce clonal clusters containing numerous replicas of amplified fragments. **(C)** Clonal clusters are sequenced using sequencing-by-synthesis technology before **(D)** raw read data is quality checked and aligned to reference genomes using bioinformatics software. *Adapted from: https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf, [accessed on 18 July 2018]*

1.1.3 Bioinformatic analysis

NGS methods generate large sequencing datasets that require robust bioinformatic tools for quality assessment and alignment of raw read data to curated reference genomes. Sequenced reads are compared to reference genomes in peer-reviewed and curated genomic databases, assigning bacterial identity and phylogeny to allow for rapid and sensitive taxonomic annotation. This is routinely done using the basic local alignment tool (BLAST; Edgar 2010) to identify and assign operational taxonomic units (OTUs) to sequence data against peer-reviewed genome databases, such as Greengenes or SILVA (Desantis et al. 2006; Pruesse et al. 2007), using the RDP classifier (Wang et al. 2007). Additional phylogenetically curated databases have been developed for the analysis of 16S rDNA data from specified anatomical niches in the human microbiome (Griffen et al. 2011). Common downstream ecological analysis of annotated taxonomic data includes assessing the bacterial diversity within a sample (alpha-diversity) and the relative bacterial diversity between samples (beta-diversity; Whittaker, Willis, and Field 2001). Alpha diversity is routinely calculated by the Shannon Diversity Index, a statistic representing the number of species and distribution of OTUs in a sample (Spellerberg & Fedor 2003). Beta-diversity measures [such as Bray Curtis (Curtis & Roger Bray 1957) or UniFrac (Lozupone et al. 2010)] are calculated based on the extent of overlap between defined parameters between two samples, for example phylogenetic distance or OTUs. Diversity measures are commonly visualised using multivariate statistics, such as principal component analysis (PCoA) and non-metric multidimensional scaling (NMDS; Faith, Minchin, and Belbin 1987; Gower 1966).

1.2 The oral microbiome

1.2.1 Defining a core microbiome

After the gut, the oral cavity has the most diverse bacterial composition compared to other sites in the human body, with more than 700 bacterial species identified (Dewhirst et al. 2010; Aas et al. 2005). In comparison to the vaginal and gut-microbiome, the oral microbiome most closely resembled the placental microbiome on a phylum level (Aagaard et al. 2014) and the placenta is believed to play a role in the development of prenatal tolerance to the maternal oral microbiome (Verma et al. 2018). Although the mechanisms determining the pioneer oral microbiome are not well understood, it is hypothesized that the placenta acts as an antigen collection site, seeded by maternal oral bacteria through a hematogenous route, and presented to the foetus to train immune tolerance (Zaura et al. 2014). Both, mode of delivery (vaginal or via caesarean section) and method of feeding (breast- or formula-fed) have shown to significantly impact the composition of the infant oral microbiome (Holgerson et al. 2013; Li et al. 2005)

Despite the number of physical and chemical exposures the oral cavity experiences on a daily basis, research suggest that a stable oral microbiome exists (Li et al. 2013; Zaura et al. 2009). NGS found 47% of species level OTUs to be shared between the oral microbiomes of three individuals (Zaura et al. 2009). The core oral microbiome contains six major phyla (Figure 1.3), representing 96% of oral bacteria found in the saliva of healthy individuals, namely: Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria, Spirochaetes and Fusobacteria (Dewhirst et al. 2010; Bik et al. 2010; Zaura et al. 2009). *Firmicutes* has been found to be the most prevalent bacteria, representing 37% of the oral microbiome, followed by *Bacteroidetes* (17.1%), *Proteobacteria* (17.1%), *Actinobacteria* (11.6%), *Spirochaetes* (7.9%) and *Fusobacteria* (5.2%) (Dewhirst et al. 2010). The extent to which host genetics determine the composition of the oral microbiome is not well understood. While one study found that the salivary microbiome did not vary across twelve locations worldwide (Nasidze et al. 2009), another study comparing saliva from Alaskan, German and African individuals found significantly higher bacterial diversity in the oral microbiomes of Africans compared to those from Northern countries (Li et al. 2014). Authors concluded that ethnicity was a stronger determinant of oral microbiome composition than age,

gender and dietary habits (Li et al. 2014). In a study of 2,343 healthy individuals, *Neisseria* and *Haemophilus* were found to dominate the salivary microbiome of Korean individuals, while *Prevotella* and *Veillonella* were most prevalent amongst Japanese individuals (Takeshita et al. 2016).

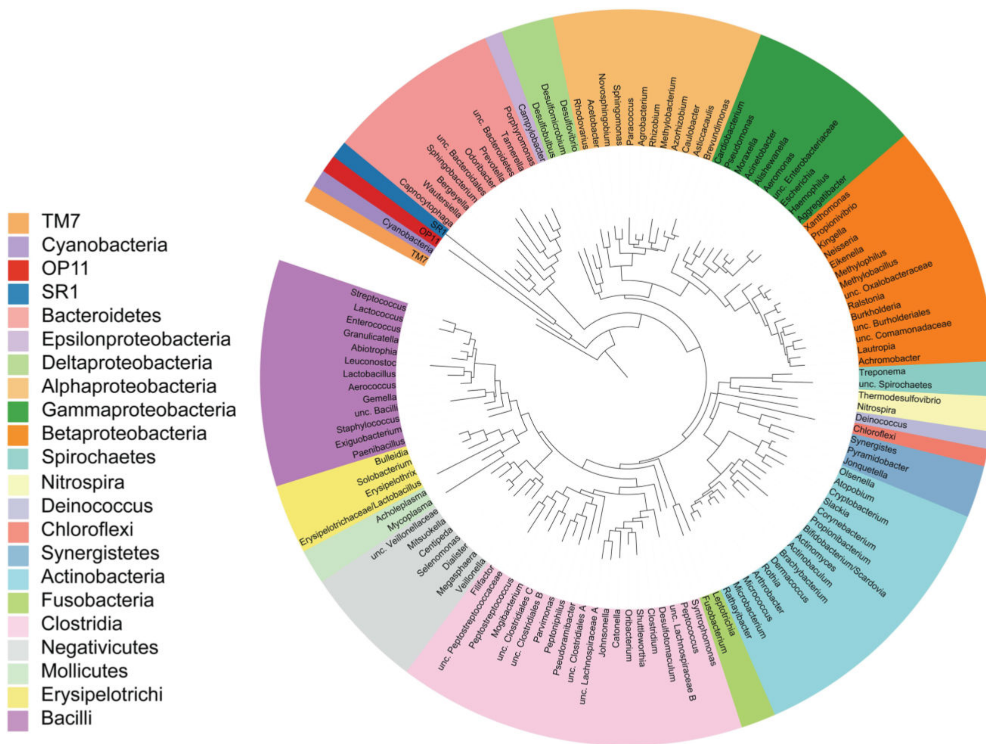
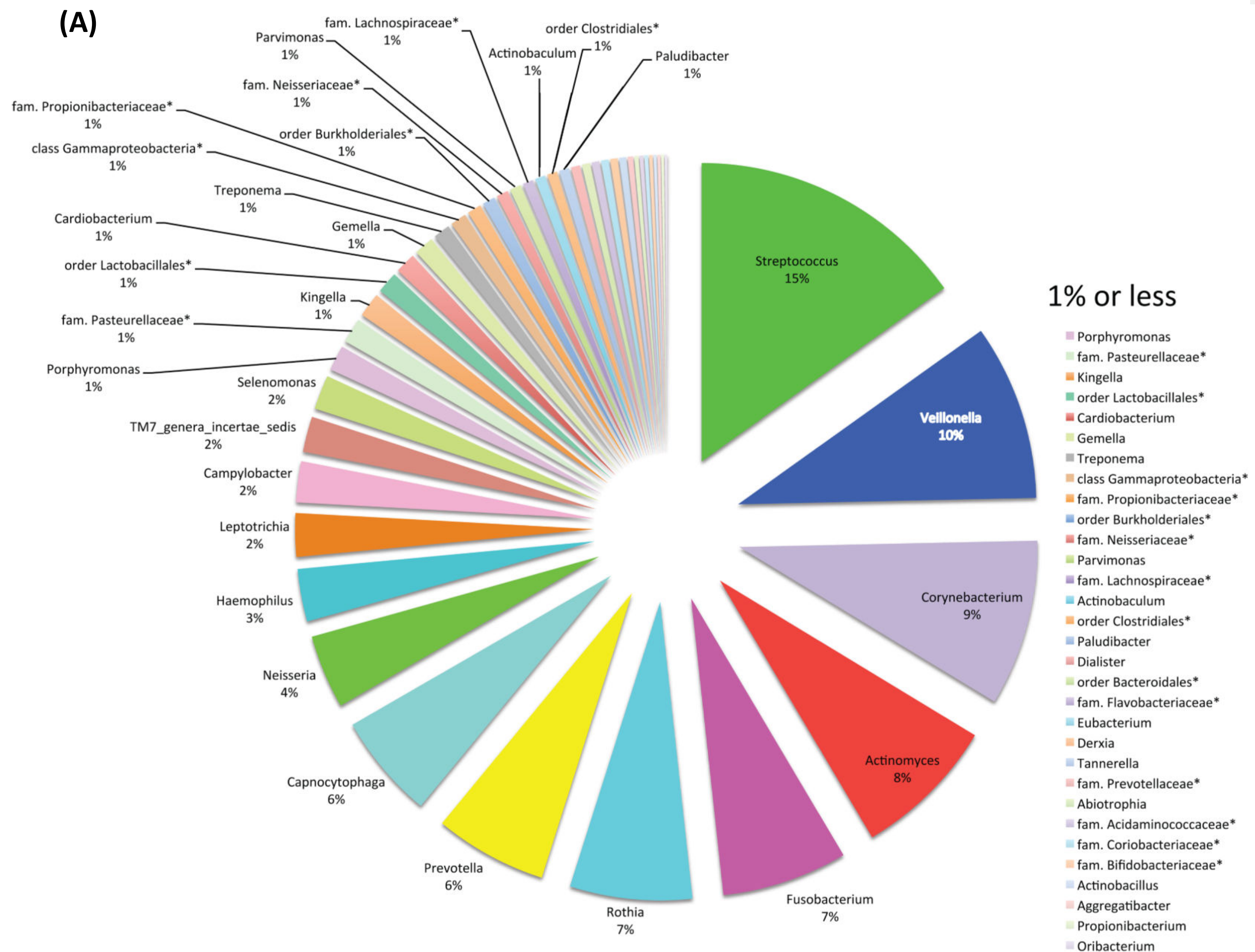


Figure 1.3: Phylogenetic visualization of 16S data in the CORE database. This database was created by using a list of bacteria gathered from a number of published studies surveying the oral microbiome and 16S sequence data from GenBank (Griffen et al. 2011). The oral microbiome was found to consist of predominately *Firmicutes, Proteobacteria, Bacteriodes, Actinobacteria, Spirochaetes, Fusobacteria* and *TM7*. Genera are colour-coded by phylum, except for the Firmicutes and Proteobacteria, which are shown at the level of class. Adapted from (Costalonga & Herzberg 2014)

1.2.2 Ecological niches within the oral cavity

The structure of the oral cavity leads to the formation of different ecological niches, causing colonization by distinct bacterial community types. Mucosal surfaces including the palate,

tongue, cheeks and tonsils house a variety of shedding bacteria that accumulate in saliva, whilst the enamelled covered surface of teeth or dentures favour the long-term formation of bacterial biofilms (Costalonga & Herzberg 2014). While Firmicutes and Bacteroidetes are most prevalent in all oral niches, Segata et al. (2012) found three distinct microbial communities when comparing microbiota sampled from seven different locations in the oral cavity. The microbial composition of the buccal mucosa, keratinized gingiva, and hard palate were dominated by Firmicutes followed in decreasing order of relative abundance by *Proteobacteria*, *Bacteroidetes* and either *Actinobacteria* or *Fusobacteria* (Segata et al. 2012). Saliva, tongue, tonsils, and throat communities contained fewer *Firmicutes* and relatively larger proportions of *Bacteroidetes, Fusobacteria, Actinobacteria* and *TM7* (Segata et al. 2012). Samples from the teeth, including both the sub- and supra-gingival plaque had the lowest proportion of *Firmicutes* and relatively larger proportion of *Actinobacteria* (Segata et al. 2012). These findings are in line with other studies, showing the oral mucosal surfaces to be largely dominated by *Firmicutes* and the dental plaque to host a more diverse range of *Firmicutes, Bacteriodetes, Actinobacteria, Proteobacteria* and *Fusobacteria* (He et al. 2015; Aas et al. 2005).



(B)

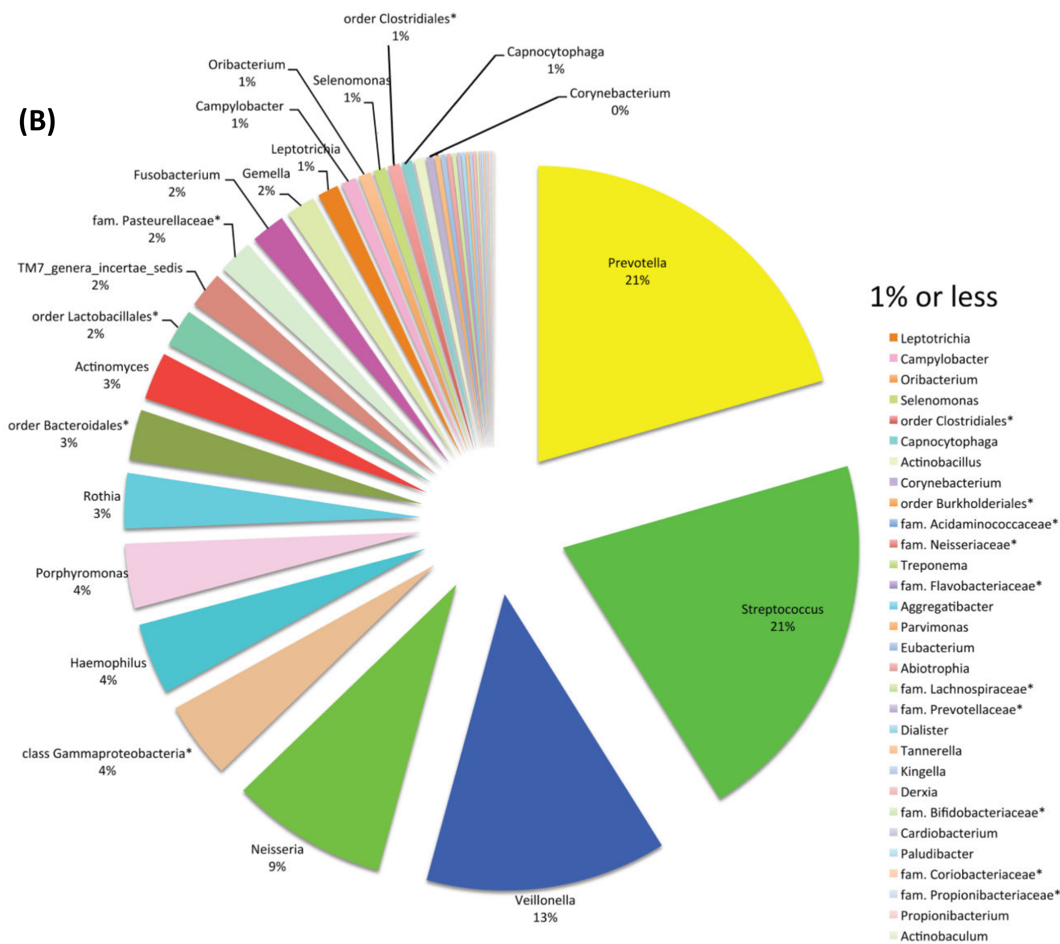


Figure 1.4: Comparison of composition of the salivary and periodontal microbiome. V6 sequence data was obtained with 454 pyrosequencing of (A) plaque (N=98) and (B) saliva (N= 71) and plaque samples collected from healthy adults who had not used antibiotics for at least three months (Keijser et al. 2008). The salivary microbiome was dominated by Firmicutes (genus *Streptococcus* and *Veillonella*) and Bacteroidetes (genus *Prevotella*), while the supragingival plaque was dominated by Firmicutes and Actinobacteria (genus *Corynebacterium* and *Actinomyces*) (Keijser et al. 2008). The asterisks (*) denote the best classification possible as adapted from a table in (Keijser et al. 2008). *Figure adapted from* (Costalonga & Herzberg 2014).

1.2.3 The oral microbiome in health and disease

The mouth is the primary gateway through which environmental pathogens gain access to the human body. Oral bacteria can gain entrance to the blood stream via the gingival crevice and can be the cause of infective endocarditis (Parahitiyawa et al. 2009) and abscesses of the brain (da

Silva et al. 2004) and liver (Schiff et al. 2003). In addition to being a reservoir for environmental pathogens, the oral microbiome has been identified as a marker of systemic disease. Altered oral microbiomes have been observed in individuals with diabetes (Long et al. 2017), bacteraemia (Roda et al. 2008), cancer (Michaud & Izard 2014), HIV (Kistler et al. 2018), autoimmune disease (Lerner et al. 2016) and systemic inflammation (Hayashi et al. 2010). Distinct oral microbiomes have additionally been observed for non-disease states, for example in pregnant and lactating women (Zarco et al. 2012), indicating that homeostatic alterations in the body may manifest in the oral microbiome. Commensal bacterial communities maintain oral health by occupying ecological niches in the oral cavity, preventing colonization by pathogenic bacterial species (Vollaard & Clasener 1994). Periodontal disease (PD) occurs when anaerobic bacteria colonize the space between the gingivae and teeth, resulting in an inflammatory periodontal pocket. PD is a polymicrobial condition that appears to be initiated by the reduction of commensal bacteria and the expansion of pathogenic bacterial communities, as opposed to the presence of a specific pathogen. PD is characterised by shifts away from the core oral microbiome and the expansion of keystone pathobiont bacterial species, such as *Porphyromonas gingivalis* and *Streptococcus mutans* (Costalonga & Herzberg 2014). Originally identified using culture-dependent methods, bacteria known as the ‘red complex’ (*P. gingivalis, Tannerella forsythia* and *Treponema denticolaI*) and ‘orange complex’ (*Fusobacterium*, *Prevotella*, and *Campylobacter* species) have been associated within progression to periodontal disease and oral inflammation (Hajishengallis & Lamont 2012; Socransky et al. 1998). Culture-independent studies have identified a number of new species belonging to the Bacteroidetes, Firmicutes, Proteobacteria, Spirochaetes and Synergistete*s* phyla as candidate pathogens associated with PD (Pérez-Chaparro et al. 2014).

Recent NGS has detected low levels of red complex bacteria in 95% of healthy individuals, indicating that they form part of the commensal oral bacterial community in the non-diseased population (Segata et al. 2012). Takeshita et al. (2012) identified two distinct community types in saliva, the first comprising of predominantly *P. histicola, P. melaninogenica, Veillonella parvula, V. atypica, S. salivarius,* and *S. parasanguinis* and the second predominantly *Neisseria flavescens, Haemophilus parainfluenzae, Porphyromonas pasteri, Gemella sanguinis, and Granulicatella*

adiacens (Takeshita & Yamashita 2012). High bacterial diversity and community composition similar to the first group was associated with poor oral health, including PD and tooth decay (Yamashita & Takeshita 2017). Like other microbial dysbiosis, More research is required to understand the predisposing factors that cause previously commensal bacteria to acquire virulence factors and become pathogenic.

1.3 The vaginal microbiome

1.3.1 Defining a core vaginal microbiome

Since the first description by Gustav Döderlein in the late 1900s, an optimal vaginal microbiome has been understood to be largely dominated by the *Lactobacillus* genus (Rogosa 1960). Lactobacilli are gram-positive, facultative anaerobes that ferment glycogen produced in vaginal secretions into D- and L- lactic acid (Witkin et al. 2014). Lactic acid production sustains a low vaginal pH (pH 3.5-4.5) that is protective against opportunistic pathogens (Boskey et al. 2014; Alakomi et al. 2000) and is believed to stimulate innate host immune factors (Mirmonsef et al. 2011). Lactobacilli provide the first line of defence against urogenital pathogens by competing for nutrients and host cell receptors at the epithelial cell surface (Zarate & Nader-Macias 2006). Further, Lactobacilli are believed to provide additional protection against invasive bacterial species through the production of bacteriocins and co-aggregation with pathogens, facilitating easy clearance of the organisms by host innate immune mechanisms (Boris et al. 1998; Spurbeck & Arvidson 2010). The vaginal microbiome is largely shaped by hormonal fluctuations during a woman's lifetime (Gajer et al. 2012). As oestrogen levels rise during puberty, glycogen is deposited in the vaginal epithelium supporting Lactobacilli growth and lactic acid fermentation. While influenced by a variety of external factors such as vaginal hygiene, sexually transmitted infections (STIs) and antibiotic use, lactobacilli dominate the vaginal microbiome of the majority of women of reproductive age (Gajer et al. 2012). Unlike other ecological niches within the human body, the human vaginal tract is believed to be dominated by one of four major Lactobacilli species – *L. iners, L. crispatus, L. jensenii* or *L. gasseri* (Ravel et al. 2011; Gajer et al. 2012).

1.3.2 Bacterial vaginosis (BV)

BV is the most prevalent urogenital disorder of reproductive age women with prevalence ranging from 25-50% in different populations (Kenyon et al. 2013). BV is a microbial dysbiosis, characterised by a reduction of Lactobacillus communities and an expansion of gram-negative and strictly anaerobic bacteria collectively known as BV associated bacteria (BVAB), including *Gardnerella, Atopobium, Mobiluncus* and *Prevotella* (Swidsinski et al. 2005; Fredricks et al. 2005). As a result of the reduction in lactic acid production and increase in proteolytic breakdown of amino acids into amines, symptomatic and asymptomatic BV presents as an increase in vaginal pH, release of proinflammatory cytokines and vaginal discharge and odour (Masson et al. 2015). While BV is not sexually transmitted, it may be sexually-associated or –enhanced, as increased numbers of sexual partners have been shown to be to be associated with an increased risk of BV (Fethers et al. 2008). The aetiology of BV is not well understood and vaginal douching, smoking, menstrual blood, previous HSV-2 infection and lack of condom use have all been identified as factors that may increase a woman's risk of BV (Fethers et al. 2008; Cherpes et al. 2008; Smart et al. 2004). While BV often resolves in the absence of any intervention, antibiotic treatment has been shown to restore vaginal microbial balance, however, the natural life cycle of BV is not well understood and BV re-occurs in 15-60% of women following treatment with antibiotics (Vodstrcil et al. 2017; Bradshaw et al. 2006).

Part of the difficulty in investigating the aetiology of BV is due to subjective diagnostic criteria, the Amsel criteria, used to identify BV in clinical settings. BV is diagnosed by the presence of at least three of the four Amsel criteria: (i) thin white or yellow homogenous vaginal wall discharge, (ii) microscopy confirmed presence of clue cells (squamous epithelial cells with adherent bacteria), (iii) vaginal pH > 4.5 and (iv) a fishy odour upon adding 10% potassium hydroxide to vaginal fluid on a glass slide (Amsel et al. 1983). In laboratory and research settings, BV is diagnosed by Nugent scoring (Nugent et al. 1991), which makes use of gram-staining a slide of vaginal fluid and scoring the proportion of gram-positive Lactobacilli, smaller gram variable curved and rod shaped bacteria. A combined score is calculate to diagnose BV positive (at or above seven), BV intermediate (between four to six) and BV negative (three and below) (Nugent

et al. 1991). Controversy exists around the reliability of Amsel criteria due to association of vaginal discharge and odour with a number of other urogenital disorders, as well as the lack of Amsel defined symptoms in up to 50% of BV cases identified by Nugent scoring (Moallaei & Namazi 2015). In addition to this, both Amsel criteria and Nugent scoring rely on subjective diagnostic criteria and can be influenced by the location and manner of vaginal sampling and the individual interpreting the gram stain (Guise et al. 2001). BV has been identified as an independent risk factor for the acquisition of STIs including HIV, pelvic inflammatory disease and adverse birth outcomes (Leitich et al. 2003; Wiesenfeld et al. 2003; Sweet 1995).

1.3.3 Redefining an optimal vaginal microbiome

Based largely on North American studies, a 'healthy' or 'normal' vaginal microbiome is defined as one containing predominantly *Lactobacillus* bacterial commensal species, having a low vaginal pH (3.5-4.5), and having no *Candida,* STI or BV present (Ma et al. 2012; Huang et al. 2015). In a study of 296 asymptomatic American women of reproductive-age, *Ravel et al.* (year) concluded that a "healthy" core vaginal microbiome does not exist, and identified five distinct community types (CTs) dominated by *L. crispatus* (26.25%), *L. gasseri* (6.3%), *L. iners* (34.1%), *L. jensenii* (5.3%) and non-lactobacilli bacteria (26%) (Ravel et al. 2011). Of the lactobacillus -dominated CTs, women with *L. crispatus*-dominant CTs had the lowest vaginal pH and those with predominantly *L. iners* CTs had the highest. This is in agreement with genomic studies describing protein families encoded by different lactobacillus species, indicating that different community compositions may confer differing levels of protection to the host (Mendes-soares et al. 2014; Sullivan et al. 2009). The vaginal microbiomes of the 26% of women not dominated by Lactobacilli were composed of strictly anaerobic bacteria, including *Prevotella, Dialister, Atopobium, Gardnerella, Megasphaera, Peptoniphilus, Sneathia, Eggerthella, Aerococcus, Finegoldia,* and *Mobiluncus* (Ravel et al. 2011). The authors additionally noted that while 80-90% of CTs of White and Asian women were dominated by lactobacilli, only ~60% of Hispanic and African-American women had CTs dominated predominately by lactobacilli (Ravel et al. 2011). These findings have been confirmed by a number of other studies (Fettweis et al. 2014; Zhou et al. 2011), indicating a distinct vaginal microbiome in African-American women characterised by high bacterial diversity and low

numbers of *Lactobacillus,* both with and without a clinical BV diagnosis. These studies have challenged the paradigm of using *Lactobacilli* as the marker of a “healthy’ vaginal microbiome and high bacterial diversity and presence of BVAB as the marker of a “unhealthy” vaginal microbiome (Ma et al. 2012). Our current understanding of the correct microbiome and molecular correlates for vaginal health are still incomplete and further research is needed into the extent to which host genetics play a role in determining vaginal microbial composition, whether there are functional redundancies by which non-lactobacillus bacterial communities may facilitate lactic acid fermentation to maintain a protective vaginal microbiota, and whether a non-lactobacillus dominated vaginal microbiome is actually a healthy state in some women.

1.4 PD, BV and pre-term birth (PTB)

Both PD and BV, conditions characterised by microbial dysbiosis, are associated with a two- to four-fold risk of PTB (Pretorius et al. 2007; Harper et al. 2012). PTB, defined as birth before 37 full weeks of gestation, is the leading cause of infant mortality globally (Liu et al. 2017). Up to 40% of PTB is thought to be the result of infection that results in either a foetal or maternal inflammatory response (Liu et al. 2017). Despite the strong epidemiological evidence for an association between BV, PD and PTB, it is unknown whether microbial dysbiosis in the oral and vaginal cavity are linked or act as two independent risk factors for PTB (Figure 1.5). Furthermore, the difficulty in diagnosing and treating microbial dysbiosis has yielded mixed results of clinical trials using PD and BV treatment to reduce adverse birth outcomes (Srinivasan et al. 2009).

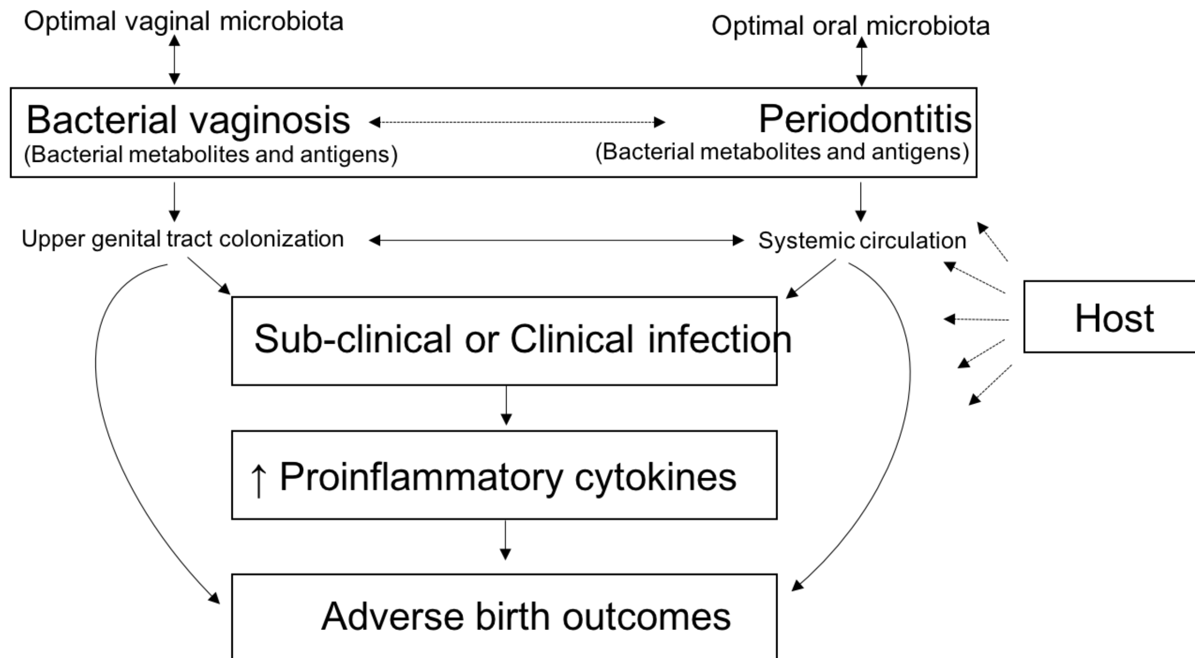


Figure 1.5: Hypothesized causal pathways underlying the observation association between BV, PD and PTB. During pregnancy, microbial dysbiosis in the oral and vaginal tract are thought to initiate a host and/or foetal immune response. Bacteria are hypothesized to access the uterine cavity by ascending through the genital tract or via systemic entry points, such as the oral cavity, triggering a clinical or subclinical inflammatory cascade that results in the release of proinflammatory cytokines and prostaglandins, inducing myometrial contractions and spontaneous pre-term birth. It is currently unknown to what extent this is a localized inflammatory response as the result of colonization of bacteria from distal sites to the uterine tissue, or a systemic inflammatory response triggered by bacterial dysbiosis at distal sites. Additionally, little is known about the extent to which host genetics modulate this inflammatory cascade. (Figure adapted from (Srinivasan et al. 2009))

1.4.1 BV and PTB

BV during pregnancy has been found to be an independent risk factor for PTB, increasing a woman's likelihood of delivering before term by two- to five-fold (Leitich et al. 2003). BV during pregnancy is also associated with an increased risk of low-birth weight, chorioamnionitis and premature rupture of amniotic membranes (PROM) (Leitich, Bodner-adler, and Brunbauer 2003). Studies investigating the size of the effect of BV on adverse birth outcomes have reported varying results, most likely due to the subjectivity in clinical BV diagnostic criteria. In a study including 221 women, expanding BV diagnosis to include atypical gram-positive bacteria and neutrophils on gram-stained vaginal smears was successful in predicting 75% of pre-term birth cases in comparison to 25% predicted using traditional Nugent scoring (Verstraelen et al. 2007). BVABs

produce proteolytic enzymes (such as sialidase, prolidase, elastase and mucinase) and prostaglandin precursors (such a phospholipase 2A) that have been shown to induce preterm birth in mice models (Bennett et al. 2000; Celik & Ayar 2002). Increased levels of sialidase activity in the vaginal tract are associated with both clinically diagnosed BV and PTB, with 70% of sialidase activity being attributed to the presence of BVAB species, such as *Prevotella, Bacteriodes* and *Gardnerella* (Briselden et al. 1992). In a study of pregnant Danish women, where clinically diagnosed BV was not associated with an increased risk of PTB, increased sialidase and/or prolidase activity in the genital tract in conjunction with a high vaginal pH was shown to significantly increase the risk of both PTB and low-birth weight (Cauci et al. 2005).

In addition to directly triggering a maternal or foetal immune response via accession through the genital tract and/or the production of inflammatory by-products in the uterine tissues, BV may also act as a marker of other urogenital infections associated with PTB risk (Srinivasan et al. 2009). It is hypothesized that increased activity of proteolytic enzymes may degrade the mucosal epithelial barrier of the vaginal tract, facilitating STI infections (Srinivasan et al. 2009). BV is associated with an increased risk of infection by *Neisseria gonorrhoeae, Chlamydia trachomatis* and HIV (Myer et al. 2005; Wiesenfeld et al. 2003) – STIs that all have been implicated in increasing the risk of PTB and other adverse birth outcomes (Dean et al. 2013). It has additionally been hypothesized that a lactobacillus-dominant vaginal microbiome may be protective against PTB, although trials using lactobacillus vaginal probiotics have been shown to be effective in reducing BV recurrence without any impact in reducing adverse birth outcomes (Othman et al. 2007).

South Africa currently follows a syndromic management approach to STI control where patients are treated based on the symptoms with which they present rather than deferring treatment until laboratory tests are available (Black et al. 2008). As per the Centre for Disease Control (CDC) recommendations clinically diagnosed BV is treated with antibiotics effective against anaerobic bacteria, such metronidazole or clindamycin (Centre for Disease Control 2015). Although there is no gold standard with regards to the type or dosage of antibiotics used, clindamycin has been

shown to be more effective at reducing BVAB populations with the caveat that it reduces protective lactobacilli populations at the same time (Lamont 2005). Studies investigating the use of BV testing and treatment to reduce adverse birth outcomes have reported mixed results, likely due to difference in BV diagnostic criteria and antibiotic treatment administered. A systematic review of 13 studies found antibiotic treatment to be effective in treating BV in pregnant women but to have no impact on the occurrence of PTB compared to a placebo (Brocklehurst et al. 2013). There was no difference in the effect observed when comparing type of antibiotic, route of delivery, antibiotic dosage or gestational time at time of treatment (Brocklehurst et al. 2013). Only two studies extended the exposure criteria to include women who had both clinically diagnosed BV (Nugent score greater than 7) and intermediate bacterial microbiota (Nugent score between 4-7)(Lamont et al. 2003; Ugwumadu et al. 2003). When combining the findings of these two trials, authors identified a significant reduction in the occurrence of preterm birth in pregnant women with abnormal vaginal microbiota (defined as Nugent score greater than 4) who received clindamycin treatment in comparison to those who did not (Brocklehurst et al. 2013). It is clear from our current knowledge that traditional clinical diagnostic criteria for BV are not sufficient in identifying microbial dysbiosis in the genital tract and that our current understanding of what constitutes a "healthy' vaginal microbiome may not be the case for certain generalisable to all women.

1.4.2 PD and PTB

The association between PD and PTB was first described by Offenbacher et al. (1996) in a case-control study where antepartum moderate-severe PD was found to increase the risk for spontaneous PTB two-fold (adjusted risk ratio (RR): 2.4, 95% CI)). Ten years later, a systematic review identified twenty-two observational studies (N participants= 7152) investigating the relationship between PD and negative birth outcomes. Of these studies, eighteen reported findings that PD was linked to increased risk of PTB, LBW babies (defined as a birth weight of less than 2500g) and PROM, while eight studies found no evidence of an association (Xiong et al. 2006). Meta-analyses of these studies indicated a positive association between PD and PTB and/or LBW (pooled odd ratio: 2.83; 95% CI: 1.95-4.10), but authors noted lower associations in

higher quality studies and recommended that the relationship be confirmed by larger randomised multi-centre trials (Vergnes & Sixou 2007). Of the seventeen studies in this metanalysis, only three assessed the microbial composition of the oral microbiota to diagnose PD. For those where microbial analysis was not done, a range of diagnostic criteria – including clinical attachment loss, bleeding on probing, plaque index, gingival index and dental mobility, were used to assign PD status (Vergnes & Sixou 2007). Clinical trials assessing the treatment of PD treatment during pregnancy to prevent adverse birth outcomes have reported mixed results (Iheozor-Ejiofor et al. 2017). A recent systematic review of 15 studies including 7161 participants noted that in addition to heterogeneity in PD diagnostic criteria, a number of different PD interventions were used, ranging from dental hygiene education to surgical and non-surgical therapies (Iheozor-Ejiofor et al. 2017). Authors identified a high risk of bias in all of the studies and concluded that there is a lack of quality evidence for the effectiveness of PD treatment to reduce PTB and other adverse birth outcomes (Iheozor-Ejiofor et al. 2017).

Although PD affects up to 50% of women in some populations, it is believed that hormonal shifts during pregnancy, specifically increases in progestogen levels, result in an increased incidence of PD in pregnant women (Figuero et al. 2013). The levels of PD-associated oral bacteria, including *P. gingivalis, T. forsythia, P. intermedia,* and *P. nigrescens,* were found to be significantly increased by at least two-fold in the oral cavity of women who had delivered pre-term compared to those who had not (Lin et al. 2007). The most commonly accepted hypothesis for the causal relationship between PD and PTB describes the hematogenous dissemination of inflammatory products released by red and orange complex bacteria in the oral cavity, resulting in a systemic inflammatory response that initiates uterine contractions and PTB (Pretorius et al. 2007). This has been supported by a number of studies identifying increased levels of inflammatory cytokines, such as IL1-β and $PGE_{2,}$ in the gingival crevicular fluid of women who deliver pre-term and/or LBW babies compared to those who delivered to term (Konopka et al. 2003; Offenbacher et al. 1998). In studies investigating the relationship between PD and cardiovascular disease, PD has been found to be associated with serological markers of systemic inflammation including increased levels of C-reactive proteins (CRP) and IL-6 (Amar et al. 2003; D'Aiuto et al. 2004). Few

studies have investigated the association between PD, systemic inflammation and adverse birth outcomes but one study found that pregnant women with both PD and increased systemic CRP were at a higher risk of developing preeclampsia compared to women without PD and both low and high levels of systemic CRP (Ruma et al. 2008). Based on identification of known oral pathogens such a Fusobacterium in the amniotic fluid of women who delivered pre-term and the absence thereof in term births (Hill 1998), it has been suggested that oral pathogens themselves may migrate to the uterine tract, causing decidual and chorioamniotic infections (Pretorius et al. 2007), although the exact mechanism through which this happens has yet to be elucidated

The mechanisms through which PD-associated bacteria might migrate to the genital tract and increase the risk of PTB through systemic infection is currently unclear, although invasive dental procedures and daily activities such as teeth brushing are thought to induce temporary bacteraemia that may create an opportunity for invasive bacteria to enter the blood stream (Toma et al. 2012; Forner et al. 2006). Additionally, both foetal and maternal immune responses seem to play a large role in determining the extent to which PD-associated oral bacteria impact adverse birth outcomes. In a study that found no significant differences in the levels of oral PD-associated bacteria between mothers who gave birth pre-term and at term, cord blood of preterm infants was found to have significantly higher rates of IgM seropositivity to bacteria belonging to the PD-associated red and orange cluster bacteria in comparison to infants born at term (Madianos et al. 2003). The same study found a lack of maternal IgG antibodies to red-cluster bacteria to be associated with 2.2-fold increased odds of PTB (Madianos et al. 2003). Subsequent studies have found foetal inflammation to be a modifier of the foetal immune response, with high cord blood IgM responses in conjunction with high levels of foetal inflammatory cytokines and CRPs to confer a greater risk of PTB compared to strong IgM responses alone (Boggess et al. 2005). While the protective effect of maternal IgG antibodies was observed in another study of African-American women (Dasanayake et al. 2003), a study of Hispanic women found no association between maternal IgM and PTB risk (Jarjoura et al. 2005), indicating that host genetics may play a role in mediating the immunological interaction between PD and PTB.

1.4.3 Evidence for a relationship between PD and BV

Although there is a large body of work investigating the independent association of BV and PD with PTB, the combined effect of microbial dysbiosis in the vaginal and oral microbiome has not been extensively studied. In the United States, PTB only occurs in around 12% of births (Martin et al. 2007), although the prevalence of BV and PD in pregnant women are thought to be approximately 15% and 50%, respectively (Vogt et al. 2012; Hillier et al. 1995). In a cohort study of 3569 non-pregnant women (80% of whom were African-American), the prevalence of BV and PD were 40% and 26%, respectively, and 12% of women had both BV and PD simultaneously, translating into 1.29-fold (95% CI: 1.13-1.47) greater risk of PD among women with BV. In a study of 180 non-pregnant women with a clinically-diagnosed BV prevalence of 27.2%, BV positive women were significantly more likely to have gingivitis (a reversible inflammatory condition of keratinized and non-keratinized gum tissues surrounding the teeth as a results of biofilm formation of PD_-associated bacteria) compared to BV negative women (Persson et al. 2009). Irrespective of clinically-diagnosed BV status, women with gingivitis were found to have significantly increased counts of BVAB, such as *P. bivia* and *P. disiens,* in their genital tracts (Persson et al. 2009). Similarly, women with both BV and gingivitis had higher counts of PD-associated bacteria in their genital tracts, including *P. gingivalis, T. forsythia* and Fusobacterium species (Persson et al. 2009). Cassini et al. (2013) directly investigated the synergistic relationship of PD and BV in pregnancy. In a cohort of 792 pregnant women, the prevalence of clinically-diagnosed BV and PD was 55% and 53%, respectively (Cassini et al. 2013). The prevalence of PD in women with a clinical BV diagnosis was 70% compared to a PD prevalence of 34% in women without BV (Cassini et al. 2013). This study found that neither BV alone, PD alone or both PD and BV together were associated with an increased risk of PTB or adverse birth outcomes (Cassini et al. 2013).

There is evidence for a hereditary component for PTB risk, with the risk of PTB found to be higher in women who were born preterm (Flint Porter et al. 1997), having a sibling born pre-term (Winkvist et al. 1998), or having a female family relative who experiences a preterm birth (Mercer

et al. 1999). A number of studies have identified genetic polymorphisms that are associated with an increased risk of PTB relating to the production of pro- and anti-inflammatory cytokines, including increased production of TNF-alpha and IL-6 and decreased production of IL-10, as well as innate defence factors such a toll-like receptors and metalloproteinases, (reviewed by Srinivasan et al. 2009). A number of alleles of inflammation-related genes have been associated with a predisposition towards PD and/or BV, (Srinivasan et al. 2009; Pretorius et al. 2007). There is additionally evidence for population-specific differences regarding genetically determined inflammatory responses, with reports indicating elevated IL-6 production to be associated with PTB in Caucasian women but not Africa-American women (Menon et al. 2008). In the same population of women with PTB cases, increased IL-6 production was observed in Caucasians and increased IL-1B was observed in African-Americans (Menon et al. 2007). Additionally, polymorphisms for maternal IL-12 and foetal IL-12B production were identified as PTB risk factors in a study assessing PTB risk factors in African -American women (Velez et al. 2009). This has led to the hypothesis that the association between PD, BV and PTB may be modulated by host genetics that may influence susceptibility to colonization by anaerobic bacteria, as well as mounting a maternal and/or foetal inflammatory immune reaction in response to microbial dysbiosis. While the mechanism through which PD initiates PTB via systemic inflammation or systemic infection is speculative, both maternal and foetal factors are likely involved.

1.5 Oral and vaginal microbiome of the South African population

Young women are disproportionately affected by HIV in Sub-Saharan Africa (Cowan & Pettifor 2009). Almost half of all new HIV infections occur in women under the age of 25, and HIV infection rates are up to eight times as high in women aged 15-24 compared to in men of the same age group (Cowan & Pettifor 2009). While high-risk sexual networks and sexual risk behaviours place adolescent women at high risk of HIV exposure, this alone cannot explain their increased risk of HIV acquisition (Katz & Low-Beer 2008; Pettifor et al. 2011). Evidence suggests that biological factors, such as a high prevalence of vaginal microbial dysbiosis, may contribute to the disproportionate rates of HIV acquisition amongst these adolescent females (Jaspan 2011). In addition to the risk of HIV acquisition, young women in South Africa are also at a high risk of

pregnancy, with 30% of 15-19 year olds reporting having ever been pregnant (Panday et al. 2009). Young women are also more at risk for adverse birth outcomes such as preterm delivery and small for gestational age infant (Khashan et al. 2010). Recent studies in young South African women have described vaginal microbiota characterised by low lactobacilli abundance, high bacterial diversity, a high prevalence of BV and a higher vaginal pH in comparison to their North American counterparts (Lennard et al. 2018; Anahtar et al. 2015). Anahtar *et al.* (2015) identified four distinct vaginal CTs in young HIV-negative South African women: three dominated by non-*L. iners* Lactobacilli, *L. iners* Lactobacilli, and *Gardnerella,* respectively, and one mixed CT. This study found bacterial diversity and the presence of specific bacteria including Fusobacterium, Aerococcus, Sneathia, Gemella, Mobiluncus, and Prevotella, to be associated with increased concentrations of pro-inflammatory cytokines in the genital tract (Anahtar et al. 2015). Previous work by our group in Cape Town, in the same study population that this study is based, identified three distinct vaginal CTs in these young South African women: two dominated by *L. iners,* or *L. crispatus,* respectively, and a third comprised of diverse BVAB, which was associated with high genital tract inflammation and persistent BV infection (Nugent score > 9; Lennard et al. 2018). In agreement with the previous study, the presence of certain BVABs, including Prevotella, Sneathia, Aerococcus, Fusobacterium and Gemella, were associated with genital tract inflammation (Lennard et al. 2018).

To my knowledge, there have been no studies characterizing the South African oral microbiome. Additionally, there are currently no studies investigating the epidemiological or causal interaction between PD and BV in South Africa women, despite the high prevalence of BV in this population. Young South African women are at a high risk for both, HIV infection and adverse birth outcomes, both of which are associated with shifts away from healthy oral and vaginal microbiomes, however current knowledge of what constitutes a healthy microbiome is largely based on North American studies. Based on the marked differences in the composition of the vaginal microbiome in young South African women and their north American counterparts, more research is required to better understand the interaction between microbial dysbiosis and preterm birth in this population.

CHAPTER TWO: Aims & Objectives

Aim: This study aims to characterise the oral and vaginal microbiome of South African female adolescents in order to investigate the association between bacterial vaginosis and periodontal disease.

Specific Objectives:

1. Characterisation of the oral and vaginal microbiome of South African adolescent females. Shifts away from healthy oral and vaginal microbial populations are associated with a number of adverse health outcomes that are prevalent in young South African women, however current knowledge of what constitutes a healthy microbiome are largely based on North American studies. To the best of our knowledge, no study has described the oral and vaginal microbiome of South African adolescent females using next-generation sequencing methods.

2. Investigate the association between alterations oral bacterial diversity and BV in young South African women

Previous data from our group suggests abnormal vaginal flora may be common in African adolescents, the group most at risk for HIV infection and teenage pregnancy. Dysbiosis in healthy oral and vaginal microbiomes is associated with negative reproductive health outcomes in women yet it is unknown whether PD and BV are independent risk factors or may be interrelated. Understanding the health risks associated with pregnancies in young women is critical for developing new preventative intervention and for informing guidelines.

CHAPTER THREE: Materials and Methods

3.1 Sample Collection

3.1.1 Study Cohort

For the purpose of this study, samples were obtained from the previously approved UChoose-A-Star (HREC REF 801/2014) sub-study, where participants were drawn from the pre-existing NIH funded parent study "Choices For Adolescent Prevention Methods for South Africa (CHAMPS): An Open-Label, Randomized Crossover Study to Evaluate the Acceptability and Feasibility of, and Adherence to, Contraceptive Choices" (R01AI094586; PI Prof Linda-Gail Bekker). The CHAMPs study investigated the acceptability and preference for contraceptive options in South African female adolescents, as proxy for HIV prevention methods. CHAMPS enrolled 150 healthy HIV-negative females age 15-19 years attending the Desmond Tutu HIV Foundation (DTHF) Youth Centre in Masiphumelele, Cape Town, South Africa. The UChoose-A-Star sub-study utilised the infrastructure of the CHAMPs study to investigate hormone-induced mucosal changes and HIV susceptibility in South African adolescents. For the purpose of this study, samples from 75 adolescent girls with (N = 35) and without (N = 40) bacterial vaginosis, HIV-negative, non-pregnant females aged 15 to 19 enrolled in the UChoose-A-Star study were included. The collection of all samples utilised in this study was previously approved by the University of Cape Town Human Research Ethics Council as part of the U-Choose-a-Star study protocol. Oral and vaginal samples are collected at the screening visit. No additional samples were collected for the purpose of this study. Additional eligibility criteria for the parent study included: contraceptive naïve or wish for a method change, no symptomatic STI within the prior 40 days, negative urine pregnancy test and no intent to be pregnant in the next 8 months, no contraindications to any study product, and willing to refrain from inserting any non-study vaginal products or objects into the vagina throughout the duration of study participation, willing to use condoms for anal and vaginal sex.

3.1.2 Ethical issues and risk assessment

All aspects of the parent study were conducted according to the International Conference on Harmonization (ICH) Good Clinical Practices (GCP), and with approval by all relevant Institutional Review Boards (IRB) in Cape Town and the USA. In addition, this study complied with the South African National Health Act (NHA, No. 61, 2003) and the South African Good Clinical Practice Guidelines (DOH 2006). Prior to enrolling participants in the study cohort, the parent study underwent ethical review and was approved by each academic institution and medical clinic where the proposed work was carried out. All samples used in this sub-study are deidentified.

For both the CHAMPS parent study and UChoose sub-study, consent was required from a parent or guardian, as well as assent from those <18 years. Although South African adolescents from age 12 years and older are allowed to consent to contraceptives and contraceptive advice independently in terms of the Children's Act (No. 38 of 2005), at present the South African National Health Act states that any health research on minors requires parental or guardian consent. The consent form(s) described the study in detail, including the purpose, study procedures, duration and frequency of visits, types of information and specimens to be collected, potential risks and benefits (if any) of study participation, and alternatives to study participation. Informed consent was obtained in the adolescent's preferred language. If participants were under the age of 18, consent from one parent was obtained in accordance with the South African Children's Act [No. 38 of 2005], Chapter 3, Part 20, and with US DHHS Human Subjects Research CFR 46.404. The consent forms were updated as necessary to reflect protocol revisions or new information. If an independent witness was necessary due to illiteracy, the witness was required to sign stating the process was adequately explained and the guardian fully understood. Original signed consent documents have been maintained in secure files. If the adolescent turn 18 years during the study, she signed the participant consent form.

A variety of mechanisms have been established to protect the confidentiality of medical records and data procured in this research. Databases are password protected, a unique study ID number was assigned at the beginning of the study and this number was used as the indirect identifier in databases and lab specimens. Patient privacy was maintained and this was communicated to the

study participants. All data was confidential, except STI and HIV results for referral clinicians only if necessary. Other results were coded, and kept under lock and key. Results of hormonal testing was not be made available to the adolescent unless they were of clinical consequence (for example if it became obvious she was not adherent to medication). Likewise, none of the microbiome analysis (except clinically relevant BV) or cervical immunology was discussed with the adolescent or their guardian. The consent and assent forms clearly stipulated that the results of the adolescents' STI and HIV test results would not be available for the legal guardian but that it was up to the adolescent to disclose results to a responsible adult, therefore results were only disclosed to the adolescent. Support was provided for such disclosures. The consent also clearly stated the conditions under which a breach of confidentiality is required according to the Children's Act (No. 38 of 2005) including knowledge of sexual abuse. The age of consensual sex is 16 years, and therefore underage sex is not an issue in this protocol (*Teddy Bear Clinic for Abused Children, and Prevention of Child Abuse and Neglect (RAPCAN) V Minister Of Justice And Constitutional Development* Case (Case Number 73300/10).

3.1.3 Oral and vaginal sample collection

Clinical samples for oral and vaginal microbiome analysis and STI testing were collected at the screening visit by clinicians at the Desmond Tutu HIV Youth Centre Clinic, Masiphumelele. Samples were then transported to the University of Cape Town, processed and placed into storage on the same day of collection. These samples include: a vulvo-vaginal swab for STI testing, a posterior fornix and lateral wall swab for bacterial vaginosis testing, a vaginal lateral wall swab for vaginal microbiome analysis, and periodontal fluid and a saliva specimen for oral microbiome analysis. With the exception of bacterial vaginosis (BV) and STI testing, all downstream sample processing was performed at the University of Cape Town.

3.1.3.1 Oral sample collection

Studies have identified distinct microbial communities existing in different environmental niches within the oral cavity (Zaura et al. 2009; Massimo Costalonga & Herzberg 2014). For the purpose of this study, two oral samples were collected in order to sample both the periodontal and

salivary bacterial composition. Participants were required to have refrained from eating or drinking anything other than water for at least 30 minutes prior to oral sample collection. Sterile toothpicks were used to scrape and collect supragingival plaque. Immediately after sampling, toothpicks were placed into an eppendorf tube with 500µl Tris-EDTA buffer (10 mM Tris HCl, 1 mM EDTA; pH 8.0) and cut. The eppendorf tube was stored upright at room temperature during transport. Saliva samples were collected using the Salivette®(Sarstedt) collection device (Figure 2.1). All handling of the swab was performed with sterile gloves. Participants were required to chew the swab from the Salivette for 60 seconds in order to stimulate saliva production. The swab was then spit back into the collection tube and stored upright at 4^{0}C during transport.

Figure 2.1: Instruction of saliva sample collection using the Salivette (Sarstedt) collection device. Edited from Salivette®: Hygienic saliva collection for diagnostics and monitoring, Sarstedt. Accessed online at: https://www.sarstedt.com/fileadmin/user_upload/99_Broschueren/Englisch/156_Salivette_GB_0813.pdf

3.1.3.2 Vaginal sample collection

Vaginal samples were collected at least 3-4 days after the end of a current menstrual cycle. Participants were instructed not to use spermicidal cream or gel for 2 days and to refrain from vaginal intercourse, douching, or inserting anything into the vagina for 1 day prior to collection of the sample. Vaginal samples were collected from different compartments of the vaginal tract as demonstrated in Figure 2.1 after the insertion of a gynaecological speculum. The vulvovaginal swab was collected by inserting a Dryswab™ Standard Tip Polyester (Medical Wire & Equipment,

MW102D) into the vagina and sampling the entire vulval area, the vaginal introitus and the lower vagina. A flocked swab (MDG diagnostics) was used to sample the right and left vaginal walls for bacterial vaginosis testing. The flocked swab was rolled onto a pH Fix 2.0-9.0 (Macherey-Nagel) testing strip and vaginal pH was recorded according to manufacturer's specifications. The same swab was then rolled onto a frosted glass slide and both the swab and glass slide were transported for further processing at room temperature. Lateral vaginal wall samples for vaginal microbiome analysis were collected using a Digene® Female Swab Specimen Collection Kit (Qiagen) according to the manufacturers specifications. All vaginal samples were transported at room temperature.

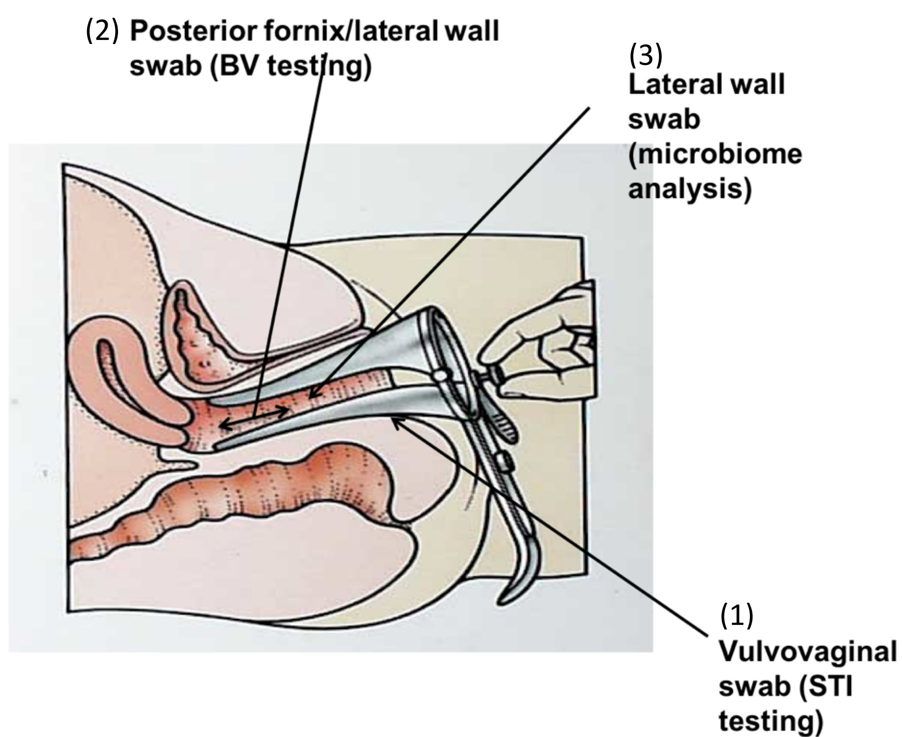


Figure 2.2: Demonstration of vaginal samples collection after the insertion of a gynaecological speculum. Three vaginal samples were collected: (1) a vulvo-vaginal swab for STI testing, (2) a posterior fornix and lateral wall swab for bacterial vaginosis testing and (3)a vaginal lateral wall swab for vaginal microbiome analysis,

3.2 STI testing

Vulvovaginal swabs and heat fixed posterior fornix/lateral wall slides were transported at -80° C to the National Institute of Communicable Disease (NICD) in Johannesburg, South Africa where polymerase chain reaction (PCR) testing for *N. gonorrhoea*, *C. trachomatis, T. vaginalis* and *M. genitalium* was performed as described by Lewis et al. (Lewis et al. 2012). Bacterial vaginosis status was determined by gram staining and Nugent scoring according to conventional protocols (Verstraelen & Verhelst 2009).

3.3 16S rRNA gene V4 DNA extraction and amplification

A previously optimised DNA extraction protocol had been developed for the extraction of DNA for sequencing of the hypervariable V4 16S DNA region from vaginal samples stored at -80°C (Lennard et al. 2018). Initially, this protocol and was used for DNA extraction from both stored oral and vaginal samples. During the course of this project, this protocol was optimised to amplify DNA extraction from oral samples found to have significantly lower DNA yields in comparison to the vaginal samples. This optimisation involved of the comparison of various oral sample collection and storage methods and the comparison of two commercial DNA extraction kits. For each optimization experiment the same overall DNA extraction and quantification protocol was followed. Briefly, stored samples were thawed and subjected to mechanical and chemical lysis prior to DNA extraction. Extracted 16S rRNA gene DNA was quantified using fluorescence, visualised and quality checked using PCR amplification and gel electrophoresis, before storage at - 20 ^{O}C. Section 3.3.1 describes the previously validated DNA extraction and quantification protocol and Section 3.3.2 describes various optimization experiments run to optimise the protocol for the extraction of sufficient bacterial DNA from oral samples for amplification of the V4 16S rRNA gene region.

3.3.1 Previously validated 16S rRNA gene V4 DNA extraction and validation protocol

3.3.1.1 Sample processing and storage

Samples were transported to the University of Cape Town on the same day of collection. Posterior fornix/lateral wall swabs, lateral wall swabs and vulvovaginal swabs were stored at -80^{0}C without further processing. Glass slides were heat fixed and stored at room temperature.

Periodontal samples were vortexed and the toothpicks discarded prior to storage at -80^0C. Salivette®(Sarstedt) tubes were centrifuged at room temperature for 5 minutes at 4000 rpm. The salivette swab was discarded and the saliva filtrate was transferred into a 2mL cryovial prior to storage at -80^0C.

3.3.1.2 Enzyme digestion

Bacterial DNA extraction methods can result in preferential sampling of gram negative bacteria, as DNA is not as easily released from gram positive bacteria with thick peptidoglycan cell walls (Yuan et al. 2012). Enzymatic digestion prior to DNA extraction ensures the rupture of gram positive cells walls, resulting in a more representative extraction of bacterial DNA present in the sample(Gill et al. 2016). A number of naturally occurring enzymes can be used to achieve this purpose: lysozyme hydrolyzes the β(1→4) linkages between *N*-acetylmuramic acid and *N*-acetyl-D-glucosamine residues in peptidoglycan, mutanolysin cleaves the β-N-acetylmuramyl-(1→4)-N-acetylglucosamine linkage of the bacterial cell wall peptidoglycan-polysaccharide (Yokogawa et al. 1974) and lysostaphin cleaves the polyglycine cross-links in peptidoglycan(Freire et al. 2010). Prior to DNA extraction, samples were removed from -80°C storage and thawed on ice. Thawed samples were briefly vortexed and 500µl of sample was removed for DNA extraction. An enzyme digestion mastermix containing; 50µl 1X Lysozyme from chicken egg white (Sigma-Aldrich, Cat. No. L6876), 6µl Mutanolysin from *Streptomyces globisporu* ATCC 21553 (Sigma Aldrich, Cat. No. M9901)and 3µl Lysostaphin from *Staphylococcus staphylolyticus* (Sigma Aldrich, Cat. No. L7386) was added to each sample and mixed by pipetting. Samples were incubated at 37°C for 60 minutes and vortexed every 20 minutes.

3.3.1.3 DNA extraction

DNA extraction from stored periodontal and saliva samples was initially performed using the Powersoil™ DNA Isolation Kit (MoBio) as was done in a previously validated protocol(Lennard et al. 2018). A negative extraction control was included where 500 µl of DNase free H_2O was used in the place of clinical samples. The manufacturer's instructions were followed with slight modifications as follows: All centrifugation steps were carried out using a microcentrifuge at 10 000g. After enzyme digestion as described in 3.3.1.2, 500µl of sample was added to the

provided PowerBead Tubes in addition to 60µL of Solution C1 and vortexed to mix. The PowerBead Tubes were then homogenized using the Thermo Savant FastPrep® 120 Cell Disrupter System (Thermo Savant) for 3 rounds of 30 seconds at 5.5 m/sec. Samples were then centrifuged for 30 seconds at room temperature. After 500µl of supernatant was transferred to a clean collection tube, 250µl of Solution C2 was added, samples were vortexed and incubated for 5 minutes at 4°C. Following a 1 minute centrifugation, 600µl of supernatant was transferred to a clean collection tube and 200µl of Solution C3 was added. Samples were then incubated and centrifuged as in the previous step and supernatant was transferred into a clean collection tube. Solution C4 was added to make up a final volume of 1750µl and 675µl of this mixture was loaded onto the Spin Filter and centrifuged for 1 minute. The filtrate was discarded and this step was repeated until all of the mixture from the previous step had been passed through the filter. The filter was then washed using 500µl of Solution C5 and centrifuged for 30 seconds. The filtrate was discarded and the Spin Filter was centrifuged for 1 minute to remove any residual C5. The Spin Filter was then transferred into a clean collection tube and 50µl of Solution C6 (heated to 60 ^{O} C) was added directly onto the filter membrane and centrifuged for 1 minute. This step was repeated using another 50µl of Solution C6 resulting in eluted DNA in a final volume of 100µl. Extracted DNA was stored at -20 ^{O} C for downstream processing.

3.3.1.4 DNA quantification

The dsDNA sample concentration was quantified using the Quant-iTTM High-Sensitivity DNA Assay Kit (Invitrogen) and Qubit 2.0 Fluorometer (Invitrogen) according to the manufacturer's instructions. As the reagents used in this kit are light sensitive, all steps of this protocol were carried out in the dark where possible. Assays were mixed in thin-wall, clear, 0.5 ml PCR tubes. The manufacturers protocol was followed. Briefly, the Qubit® dsDNA HS Reagent and Buffer were brought to room temperature and a working solution was created by diluting the Qubit® dsDNA HS Reagent 1:200 in Qubit® dsDNA HS Buffer. The required amount of working solution was prepared in plastic conical tubes and wrapped in tinfoil to protect from light. Standards were created using the supplied Qubit® dsDNA HS Standards and the working solution to make up 200µl standards with reagent concentrations of 0 ng/µl and 10 ng/µl. Extracted DNA (2µl) was

added to 198µl of working solution to make up 200 µl assays that were subsequently vortexed and incubated at room temperature for 2 minutes. For each DNA quantification, new standards were created and read using the Qubit® 2.0 Fluorometer to generate a double-stranded DNA standard curve used to calculate the DNA concentration of the subsequent sample assays. Sample assays were read using the Qubit® 2.0 Fluorometer and each assay was vortexed prior to reading. The Qubit® 2.0 Fluorometer displays DNA concentration in ng/ml and the following formula was used to calculate the concentration of the sample:

Sample concentration = Qubit value x $\frac{200}{x}$, where x = the number of microliters of sample added to the assay tube.

3.3.1.5 V4 16S PCR amplification and gel electrophoresis

Extracted DNA was removed from -20°C storage and thawed on ice. The V4 hypervariable region of the bacterial 16S rRNA genes was amplified using PCR according to the following protocol. A master mix was created containing: 1µl (5µM/µl) each of the universal forward and reverse primers 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3') and 12.5µl high fidelity DNA polymerase KAPA Hotstart Mix (Thermo Scientific™ Phusion™). Variable amounts of template DNA were added (maximum volume 10µl) to reach a final amount of ~20ng. Variable amounts of DNase free H_2O were then added to reach a final reaction volume of 25µL per sample (Table 2.1). PCR reaction were run in 0.5-mL PCR tubes using the GeneAmp® PCR System 9700 according to the following program: 1 cycle 95°C for 3 minutes, 35 cycles of; 95°C for 30 seconds, 55°C for 30 seconds and 72°C for 30 seconds, 1 cycle of 72°C for 5 minutes hold at 4°C (Table 2.2). All PCRs were run using a negative control where 10µL of DNase free H_2O was substituted for sample in the above described mastermix.

The size and quality of each amplified DNA sample was visualised using gel electrophoresis. 2µL of amplicon mixed with 0.5ul of 6X TriTrack DNA Loading Dye (Thermo Scientific) was run on a 1.2% agarose gel (Tris–Acetate–EDTA (TAE) buffer, 90 V, 60 min) with GelRed™ Nucleic Acid Gel Stain (Biotium) added to the molten agarose at a dilution of 1:10 000 according to manufacturer's instructions. Either the Gene Ruler 1Kb or O'Gene Ruler 100bp Plus DNA Ladders (both Thermo

Scientific) were included at 0.5µg/ml for DNA band size estimation. Successful amplification was determined by the presence of a clear band (~320bp).

Table 2.1: Reagent list for validation PCR

Reagent	Amount to add
Template DNA	Max 10µl (to ~20ng DNA)
Forward primer	1µl
Reverse primer	1µl
KAPA Hotstart Mix	12.5µl
PCR H_2O	0.5µl (up to final volume of 25µl)
Total volume:	25µl

Table 2.2: Thermocycler protocol for validation PCR

PCR phase	Duration/temperature
Initial denaturation	95°C for 3 minutes
Cycles (x35)	
Denaturation	95°C for 30 seconds
Annealing	55°C for 30 seconds
Elongation	72°C for 30 seconds
Final Elongation	72°C for 30 seconds
Storage	4°C

3.3.2 Optimization of of 16S rRNA gene V4 region DNA extraction and validation protocol

3.3.2.1 Optimization Experiment 1: DNA extraction from stored oral samples

We first investigated the effectiveness of a DNA extraction protocol optimised for DNA extraction from vaginal lateral wall samples stored at -80°C on periodontal and saliva samples stored under identical conditions. Saliva and periodontal samples were collected and processed as described in 3.1. DNA was extracted following the protocol described in 3.3.1.3 using the Powersoil™ DNA Isolation Kit (MoBio). Extracted DNA was quantified using the Quant-iTTM High-Sensitivity DNA Assay Kit (Invitrogen) and Qubit 2.0 Fluorometer (Invitrogen) as described in 3.3.1.4. Extracted DNA was then amplified using PCR and gel electrophoresis was performed as described in 3.3.1.5. DNA quantification revealed all samples to have below detectable concentrations of dsDNA (<0.05ng/µl). Gel electrophoresis (Experiment 1, Figure 2.3) revealed no DNA amplicons of the expected size (~450 bp) and a high concentration of small genomic bands (<100bp), most likely to be primer dimers. These results indicated that the previously validated protocol was not effective in extracting DNA from stored periodontal and saliva samples.

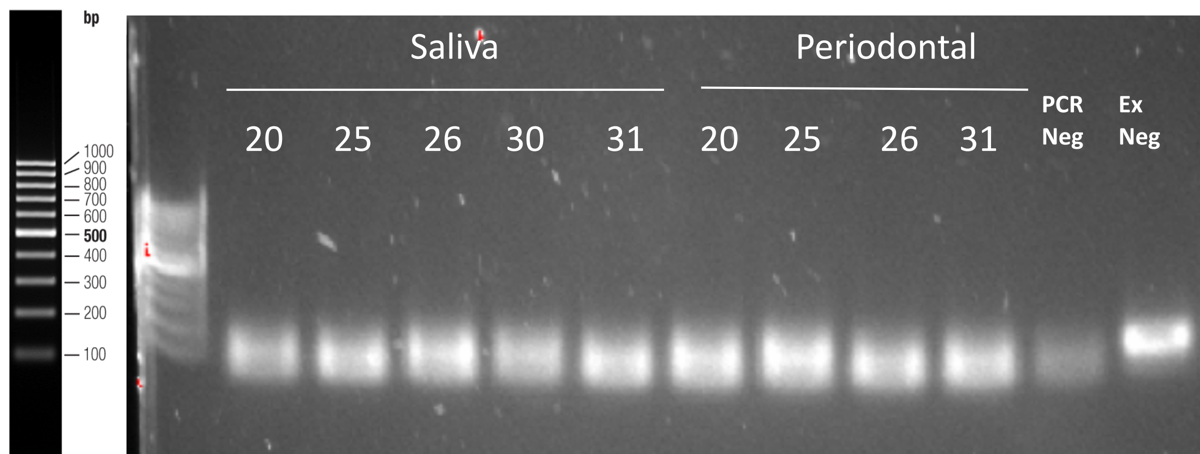


Figure 2.3: Gel electrophoresis of amplified 16S V4 DNA extracted from stored oral samples using the Powersoil™ DNA Isolation Kit (MoBio). Gel electrophoresis was carried out as described in 3.3.1.5. O'Gene Ruler 100bp Plus DNA Ladder (Thermo Scientific) was included at 0.5µg/ml in the left most lane for DNA band size estimation. Lanes 2-6 contain amplicon products after PCR amplification of extracted stored saliva samples. Lanes 7-10 contain amplicon products after PCR amplification of extracted stored periodontal samples. Lane 11 and 12 contain negative PCR and DNA extraction controls respectively.

3.3.2.2 Optimization Experiment 2: DNA extraction from fresh oral samples with additional processing prior to DNA extraction

We next investigated the effectiveness of a DNA extraction protocol optimised for stored vaginal lateral wall samples on fresh periodontal and saliva samples**.** Saliva gland secretions are thought to contain nucleases and proteins that may degrade DNA during extended sample storage and inhibit DNA amplification using PCR. Boiling saliva samples prior to DNA extraction and PCR amplification has been observed to result in increased yields of amplified DNA (Ochert et al. 1994). We therefore added the following to the processing of periodontal and saliva samples:

(1) Boiling of periodontal and saliva samples prior to extraction in order to deactivate potential inhibitors that may be interfering with DNA extraction and PCR reactions

(2) Concentration of salivary bacteria and removal of the majority salivary supernatant in order to eliminate potential inhibitors that may be interfering with DNA extraction and PCR reactions

Periodontal and saliva samples were collected as described in 1.2.1 from volunteers at the University of Cape Town. Saliva filtrate was transferred from the Salivette tube as described in 2.1.1. Saliva samples were then centrifuged for 15 minute centrifugation at 4000g at 4°C. Leaving 100µl to prevent excluding low-weight bacteria still present in the saliva, saliva supernatant was

removed and stored at -80^{0}C. The pellet was resuspended in the remaining 100µl of supernatant and 300µl Tris-EDTA (pH 8.0) buffer. The resuspended saliva pellets and periodontal samples were then boiled at 95°C from ten minutes. Samples were allowed to cool to room temperature prior to extraction. DNA was extracted following the protocol described in 3.3.1.3 using the Powersoil^{TM} DNA Isolation Kit (MoBio). In addition to previously described negative extraction control, a positive extraction control containing a stored vaginal lateral wall swab (collected and processed as described in 3.1) was included to assess the effectiveness of the previously validated DNA extraction protocol.

DNA was quantified using the Quant-iTTM High-Sensitivity DNA Assay Kit (Invitrogen) and Qubit 2.0 Fluorometer (Invitrogen) as described in 3.3.1.4. Extracted DNA was then amplified using PCR and gel electrophoresis was performed as described in 3.3.1.5. DNA quantification revealed 50% of the extracted oral samples to have DNA concentration lower than the detectable range (Table 3). The yield of DNA extracted from oral samples where DNA was detectable was lower than the DNA extracted from a stored lateral vaginal wall sample (Table 2.3). The DNA extraction negative control (Figure 2.4, Lane 13) showed only low weight molecular bands. The DNA extraction positive control (Figure 2.4, Lane 12) showed an amplicon of the expected size (~320 bp). Amplicons of the expected size were present for all but one of the oral samples (P3-Saliva, Figure 4, Lane 9). DNA quantification revealed majority of samples to have detectable concentrations of dsDNA (Table 2.3). This was confirmed by PCR amplification and gel electrophoresis revealing majority of the samples to have visible amplicon products around the expected size of 320 base pairs (Figure 2.4). DNA concentrations were calculated using fluorescence correlated with the intensity of the bands observed in gel electrophoresis. DNA extraction negative and positive controls demonstrated that there was no bacterial contamination during the DNA extraction and PCR steps. Additionally the protocol was confirmed to be effective in extracting V4 16S DNA from stored vaginal samples. Boiling oral samples and the removal of salivary supernatant directly after sample collection resulted in observable ~320bp bands amplified from DNA extracted from these samples. Faint amplicon bands (~320bp) were visible for samples where quantification using

fluorescence was not able to detect dsDNA, indicating V4 16S DNA amplification is possible from samples with very low DNA yields.

Table 2.3: Concentration of dSDNA extracted from fresh oral and stored vaginal samples

				Extracted DNA conc (ng/ µl)		
Sample ID	Pre-extraction processing	Sample type	DNA Extraction kit	Periodontal	Saliva	LWM
P1	Boiled	Fresh	Powersoil	0.104	0.066	-
P2	Boiled	Fresh	Powersoil	> 0.05	> 0.05	-
P3	Boiled	Fresh	Powersoil	> 0.05	> 0.05	-
P4	Boiled	Fresh	Powersoil	0.404	> 0.05	-
P5	Boiled	Fresh	Powersoil	0.214	0.142	-
P6	None	Fresh	Powersoil	-	-	0.812

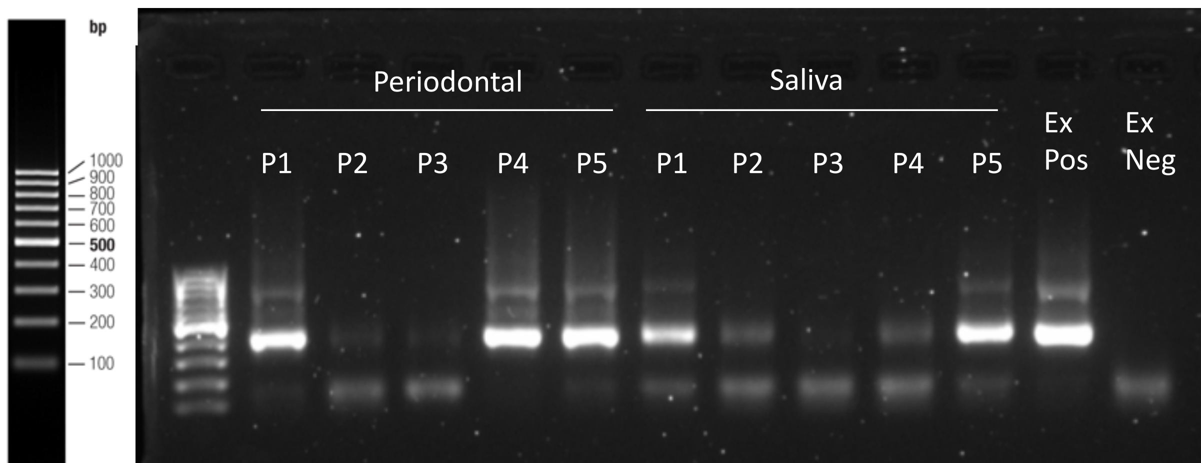


Figure 2.4: Gel electrophoresis visualisation of PCR products of amplified 16S rRNA gene V4 product from fresh oral samples using the Powersoil^{TM} DNA Isolation Kit (MoBio). Gel electrophoresis was carried out as described in 2.1.4. O'Gene Ruler 100bp Plus DNA Ladder (Thermo Scientific) was included at 0.5 µg/ml in the left most lane for DNA band size estimation. Lanes 2-6 contain amplicon products after PCR amplification of extracted periodontal samples. Lanes 7-11 contain amplicon products after PCR amplification of extracted periodontal samples. Lane 12 and 13 contain positive and negative DNA extraction controls respectively.

3.3.2.3 Optimization Experiment 3: DNA Extraction from stored oral samples with boiling prior to extraction

As a number of the samples had already been collected and stored at -80 ^{O}C, we next investigated post-thawing processing of stored oral samples prior to DNA extraction using DNA extraction protocol optimised for stored vaginal lateral wall samples**.** For this, periodontal and saliva

samples were removed from -80°C storage and thawed on ice. Prior to extraction, thawed samples were processed (by boiling and the removal of salivary supernatant) as described in 2.2.2. DNA was quantified using the Quant-iTTM High-Sensitivity DNA Assay Kit (Invitrogen) and Qubit 2.0 Fluorometer (Invitrogen) as described in 2.1.3. Extracted DNA was then amplified using PCR and gel electrophoresis was performed as described in 2.1.4. DNA quantification revealed 33% of the extracted oral samples to have DNA concentration lower than the detectable range (Table 2.4). Amplicons of the expected size were present for the majority of samples (Figure 2.5). Boiling oral samples and the removal of salivary supernatant directly after thawing samples from storage at -80°C resulted in observable ~320bp bands amplified from DNA extracted from these samples. This additional sample processing allowed for sufficient DNA extraction and amplification using the Powersoil^{TM} DNA Isolation Kit (MoBio) and procedures described in 2.1

Table 2.4: Concentration of dsDNA extracted from stored oral samples

				Extracted DNA conc (ng/ µl)	
Sample ID	Pre-extraction processing	Sample type	DNA Extraction kit	Periodontal	Saliva
UC-013	Boiled after thawing	Stored	Powersoil	0.981	0.043
UC-015	Boiled after thawing	Stored	Powersoil	0.760	1.23
UC-016	Boiled after thawing	Stored	Powersoil	0.690	0.142
UC-017	Boiled after thawing	Stored	Powersoil	> 0.05	0.997
UC-018	Boiled after thawing	Stored	Powersoil	> 0.05	> 0.05
UC-021	Boiled after thawing	Stored	Powersoil	> 0.05	1.02
UC-022	Boiled after thawing	Stored	Powersoil	0.234	0.098
UC-024	Boiled after thawing	Stored	Powersoil	0.187	0.654
UC-028	Boiled after thawing	Stored	Powersoil	0.344	0.056

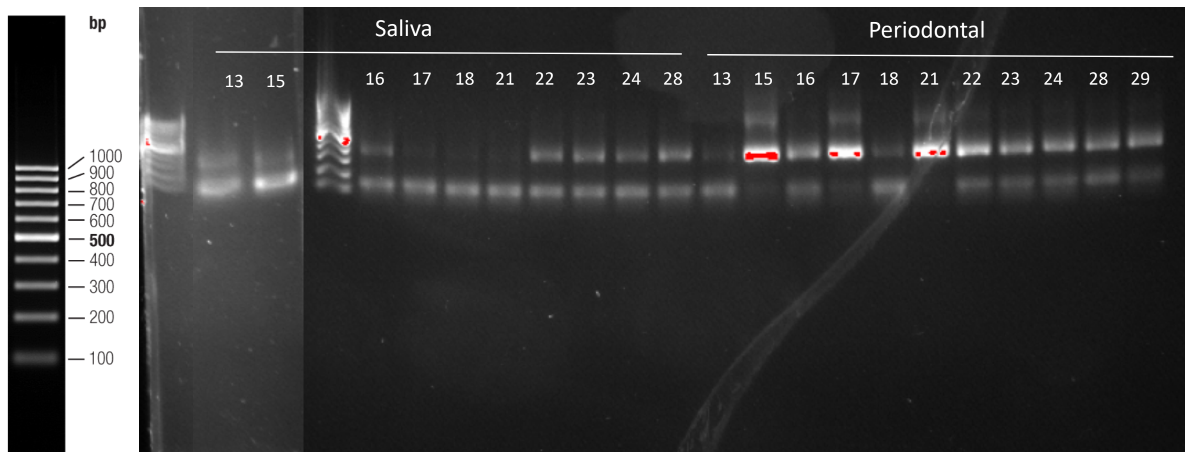


Figure 2.5: Gel electrophoresis visualisation of amplified 16S V4 DNA extracted from stored oral samples using the Powersoil™ DNA Isolation Kit (MoBio). Gel electrophoresis was carried out as described in 2.1.4. Lanes 2,3 and 5-12 contain amplicon products after PCR amplification of extracted saliva samples. Lanes 13-23 contain amplicon products after PCR amplification of extracted periodontal samples. Lane 1 and 4 contain O'Gene Ruler 100bp Plus DNA Ladder (Thermo Scientific) included at 0.5 µg/ml.

3.3.2.4 Optimization Experiment 4: DNA Extraction from stored samples with boiling prior to storage

We additionally investigated pre-storage processing of stored oral samples prior to DNA extraction using the DNA extraction protocol optimised for stored vaginal lateral wall sample. Upon receipt in the lab, periodontal samples were vortexed and were boiled at 95°C for 10 minutes prior to storage at -80°C. Salivette samples were centrifuged at room temperature for 5 minutes at 4000 rpm. The swab was discarded and the filtrate was transferred into a 2mL cryovial and boiled at 95°C for 10 minutes before a 15 minute centrifugation at 4000g (RCF) at 4°C. Leaving 100µl, supernatant was removed at stored at -80^{0}C. The pellet was resuspended in the remaining 100µl of supernatant and 300µl Tris (pH 8.0) buffer prior to storage at -80°C. After a minimum of 2 weeks storage at -80 ^{O}C, periodontal and saliva samples were thawed on ice and DNA was then extracted following the protocol described in 2.1 using the Powersoil™ DNA Isolation Kit (MoBio). DNA was quantified using the Quant-iTTM High-Sensitivity DNA Assay Kit (Invitrogen) and Qubit 2.0 Fluorometer (Invitrogen) as described in 2.1.3. Extracted DNA was then amplified using PCR and gel electrophoresis was performed as described in 2.1.4. DNA quantification revealed 65% of the extracted oral samples to have DNA concentration lower than the detectable range (Table 2.5). The DNA extraction negative control (Figure 2.6, Lane 14)

showed only low weight molecular bands and no detectable DNA was identified through fluorescence. Amplicons of the expected size (~320bp) were present for the majority of samples (Figure 2.6). Gel electrophoresis resulted distinctly higher yield amplicon products for periodontal samples in comparison to saliva samples (Figure 2.6). Pre-storage processing, including boiling oral samples and the removal of salivary supernatant prior to DNA extraction allowed for sufficient DNA extraction and amplification using the Powersoil^{TM} DNA Isolation Kit (MoBio) and procedures described in 2.1. DNA amplicon yields from periodontal samples are distinctly higher than those from saliva samples, most likely due to salivary PCR inhibitors.

Table 2.5: Concentration of dsDNA extracted from stored oral samples

				Extracted DNA conc (ng/ µl)	
Sample ID	Pre-extraction processing	Sample type	DNA Extraction kit	Periodontal	Saliva
UC-062	Boiled after thawing	Stored	Powersoil	<0.05	<0.05
UC-072	Boiled after thawing	Stored	Powersoil	0.462	1.52
UC-073	Boiled after thawing	Stored	Powersoil	<0.05	<0.05
UC-084	Boiled after thawing	Stored	Powersoil	0.269	<0.05
UC-089	Boiled after thawing	Stored	Powersoil	<0.05	<0.05
UC-090	Boiled after thawing	Stored	Powersoil	0.069	<0.05
UC-091	Boiled after thawing	Stored	Powersoil	0.069	<0.05

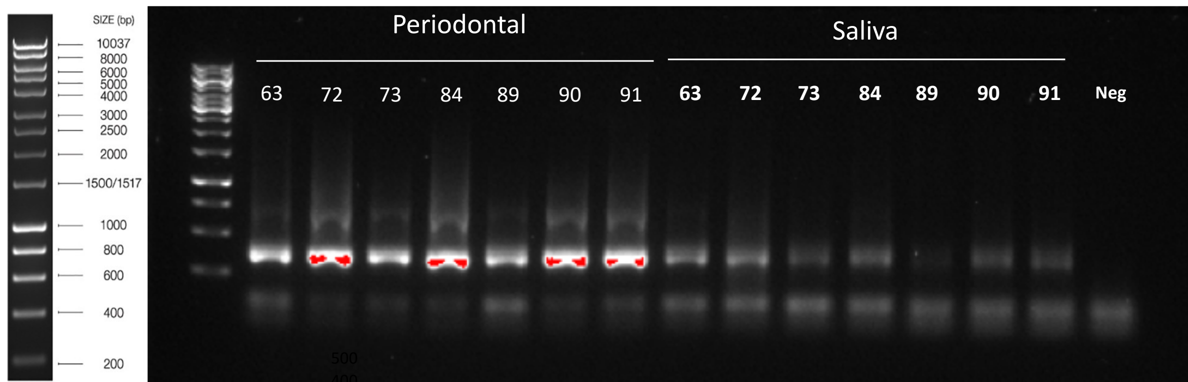


Figure 2.6: Gel electrophoresis visualisation of amplified 16S V4 DNA extracted from stored oral samples using the Powersoil^{TM} DNA Isolation Kit (MoBio). Gel electrophoresis was carried out as described in 2.1.4. O'Gene Ruler 100bp Plus DNA Ladder (Thermo Scientific) was included at 0.5 µg/mL in the left most lane for DNA band size estimation. Lanes 2-8 contain amplicon products after PCR amplification of extracted periodontal samples. Lanes 8-15 contain amplicon products after PCR amplification of extracted periodontal samples. Lane 16 contains the negative DNA extraction control.

3.3.2.5 Optimization Experiment 5: Comparison of two commercially available DNA extraction kits to extract 16S DNA from oral, vaginal and foreskin samples

While the Powersoil™ DNA Isolation Kit (MoBio) was shown to be sufficient for extracting 16S DNA from both oral and vaginal samples, we identified a second commercially available kit, the Quick-DNA™ Fungal/Bacterial Kit (Zymo Technologies), with a significantly shorter protocol time (~3 hours as opposed to the ~5 hours), and previously shown to be superior in DNA extraction for similar purposes (Claassen et al. 2013). This experiment aimed to compare performance of both kits in extracting bacterial DNA for the purpose of the previously described protocol. Stored saliva, lateral vaginal wall and foreskin swab samples were thawed on ice. Prior to storage, lateral vaginal wall and saliva samples had been processed and stored as described in 1.2.1. Foreskin swabs from a previous study (HREC ref: 566/2012) were included as additional low biomass samples. These were collected by swabbing the inner foreskin with a flocked swab (MDG diagnostics) and stored at -80 ^{O} C. All samples were split into duplicate aliquots and enzyme digestion was followed as described in 2.1. DNA was extracted from one set of duplicates using the Powersoil™ DNA Isolation Kit (MoBio) as described in 2.2.1. DNA was extracted from the other set of duplicates using the Quick-DNA™ Fungal/Bacterial Kit (Zymo Technologies) and the manufacturer's instructions were followed with slight modifications as follows: All centrifugations were carried out in a microcentrifuge. Briefly, 500µl of sample and 250µl of Bashing Bead™ Buffer was added to ZR Bashing Bead™ Lysis Tubes. Tubes were homogenised using TissueLyser LT (Qiagen) for 5 minutes at 50Hz. ZR Bashing Bead™ Lysis Tubes were then centrifuged at 10 000g for 1 minute. Up to 400µl supernatant was transferred to the Zymo-Spin™ III-F Filter in a Collection Tube and centrifuge at 8 000 g for 1 minute. The Zymo-Spin™ III-F Filter was the discarded and 1200µl of Genomic Lysis Buffer (with beta-mercaptoethanol added to a final dilution of 0.5%(v/v)) was added to the filtrate in the Collection Tube from the previous step. Up to 800µl of this mixture was transferred to Zymo-Spin™ IC Column in a Collection Tube and centrifuged at 10 000 g for 1 minute. The filtrate was discarded and this step was repeated before the Zymo-Spin™ IC Column was transferred to a new collection tube and 200µl of DNA Pre-Wash Buffer was added to the column. The column was centrifuged for at 10 000 g for 1 minute and the filtrate discarded. This step was then repeated using 500µl of g-DNA Wash Buffer. The Zymo-Spin™ IC Column was transferred to a clean 1 ml microcentrifuge tube and 50µl DNA Elution

Buffer was added directly to the column matrix and incubated for 1 minute. The tube was centrifuged at 10 000 g for 30 seconds to elute the DNA. This step was repeated using another 50µl DNA Elution Buffer resulting in eluted DNA in a final volume of 100µl. Extracted DNA was stored at -20°C for downstream processing. DNA was quantified using the Quant-iTTM High-Sensitivity DNA Assay Kit (Invitrogen) and Qubit 2.0 Fluorometer (Invitrogen) as described in 3.3.1.4 . Extracted DNA was then amplified using PCR and gel electrophoresis was performed as described in 3.3.1.5.

When comparing these two kits, DNA quantification of samples extracted using the Zymo kit revealed all samples to have detectable dsDNA concentrations (>0.05 ng/ml) (Table 2.6). Of the samples extracted using the Mobio kit, dsDNA was not detected in 33% of samples (Table 2.6). The DNA extraction negative control for both kits (Figure 2.7, Lanes 19 and 20) showed only low weight molecular bands and no detectable DNA was identified through fluorescence (Table 2.6). Amplicons of the expected size (~320bp) were present for all saliva samples (Figure 2.7, Lanes 1-6). For all saliva samples, the Zymo kit produced a brighter amplicon band in comparison to extraction done with the Mobio kit. Amplicons were present for 2 out of the 3 foreskin samples (Figure 2.7, Lanes 7-12) with no distinct difference in the brightness of amplicons bands between samples extracted using the difference kits observed. Amplicons of the expected size were present for all vaginal samples extracted from both kits, with slightly brighter bands observed for the Zymo extractions (Figure 2.7, Lanes 13-18). Gel electrophoresis resulted distinctly higher yield amplicon products for vaginal samples in comparison to saliva samples (Figure 2.7). DNA concentrations calculated using fluorescence did not correlate with the intensity of the bands observed in gel electrophoresis. In this experiment, dsDNA quantification did not give a reliable indication of the performance of the two extraction kits compared as high yield amplicon ~320 bp were observed for vaginal samples where dsDNA was recorded as below the detectable range. This may be a result of the light reactivity of the reagents used in the DNA quantification protocol resulting in inaccurate DNA readings. Based on the gel electrophoresis result and the corresponding DNA quantification using both Qubit and Nanodrop, for the majority of samples, the use of the Zymo kit resulted in higher yields of amplicons after V4 16S PCR amplification. Due

to this and the shorter protocol time required for the Zymo kit extractions, the Zymo kit was used for all further DNA extraction in this study.

Table 2.6: Concentration of dsDNA extracted from fresh oral and stored vaginal samples

		Extracted DNA conc (ng/ µl)			
		Qubit		Nanodrop	
Sample	Type	Zymo kit	MoBio kit	Zymo kit	MoBio kit
Saliva 1	Fresh	0.234	<0.05	8.9	1.6
Saliva 2	Fresh	0.145	0.263	11.1	1.3
Saliva 3	Fresh	0.131	0.098	13.1	0.9
Foreskin 1	Stored	2.93	3.44	11.2	4.9
Foreskin 2	Stored	1.05	0.761	16.4	3.4
Foreskin 3	Stored	0.731	0.748	17.5	3.8
Vaginal 1	Stored	0.190	<0.05	11.4	1.8
Vaginal 2	Stored	1.49	0.205	12.7	3.5
Vaginal 3	Stored	0.064	<0.05	14.3	2.3
Neg Control	-	<0.05	<0.05	4.0	1.2

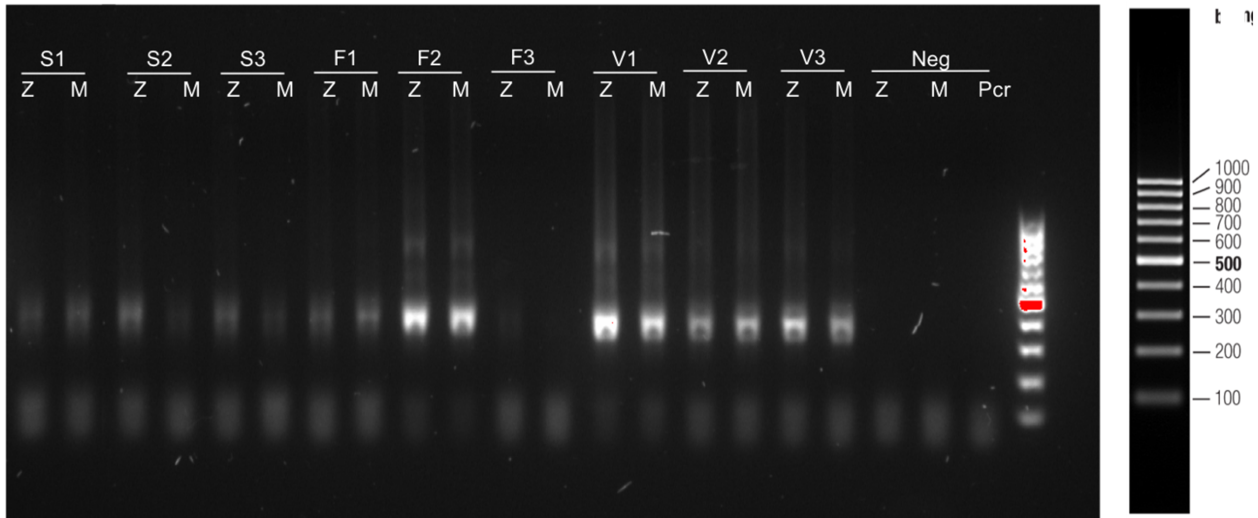


Figure 2.7: Gel electrophoresis visualisation of amplified 16S V4 DNA extracted from stored oral samples using the Powersoil^{TM} DNA Isolation Kit (MoBio). Gel electrophoresis was carried out as described in 2.1.4. O'Gene Ruler 100bp Plus DNA Ladder (Thermo Scientific) was included at 0.5 µg/ml in the right most lane for DNA band size estimation. Lanes 1-6 contain amplicon products after PCR amplification of extracted saliva samples. The gel contains amplicon products after PCR amplification for saliva (S, lanes 1-6), foreskin (F, lanes 7-12) and vaginal (V, lanes 13-18) extracted using the Mobio (M) and Zymo (Z) kits. Lanes 19,20 and 21 contain negative controls for the Mobio extraction, Zymo extraction and PCR.

3.3.3 Summary of findings and resultant protocol changes from 16S rRNA gene V4 DNA extraction optimization

The protocol described for enzyme digestion for the extraction of 16S microbial DNA (3.3.1.2), 16S DNA Quantification (3.3.1.4) and gel electrophoresis (3.3.1.5) remained unchanged. Based on the previously described optimisation experiments the following changes were made to the samples processing and DNA extraction protocols:

- Boiling and the removal of salivary supernatant from oral samples either prior or post -80°C storage increases the yield of bacterial DNA extracted from oral samples. This additional processing was included prior to -80° C storage for all periodontal and saliva samples. For all samples already in -80° C storage, this additional processing was included after samples were thawed prior to DNA extraction.
- The yield of DNA extracted from periodontal and saliva samples is lower than that of DNA extracted from lateral vaginal wall samples. Based on this, for the purpose of DNA library preparation for MiSeq sequencing, oral sample V4 16S rRNA gene amplification was planned to be performed in duplicate and pooled prior to subsequent protocol steps.
- V4 16S rRNA gene PCR amplification is possible from samples where quantification using fluorescence was not able to detect dsDNA. Based on this, the maximum volume (10uL) of all low yield samples (periodontal and saliva) were included during the V4 16S amplification step of DNA library preparation for MiSeq sequencing.
- In comparison to the MoBio, the Zymo kit resulted in improved amplification of V4 16S rRNA gene product. Due to this and the shorter protocol time required for the Zymo kit extractions, the Zymo kit was used for all further DNA extraction in this study.

3.4 Illumina 16S rRNA iTag library preparation

3.4.1 Previously validated Illumina 16S rRNA iTag library preparation protocol

As with the DNA extraction, a previously validated protocol developed for Illumina 16S rRNA iTag library preparation from V4 16S rRNA gene DNA extracted from vaginal lateral wall swabs was originally used to prepare DNA libraries containing matched oral and vaginal samples from this study (Lennard et al. 2018). DNA libraries were prepared in 96-well plates including 92-samples and 4 controls: PCR 1 negative control, PCR 2 negative control, extraction negative control and positive PCR control (detailed description to follow). Where possible, participant matched samples (lateral wall, periodontal and saliva) were included in the same library. Library preparation included the followings steps: An initial round of PCR to amplify the V4 hypervariable region of the 16S bacterial chromosome from extracted dsDNA, bead purification of amplicons from the first round of PCR, a second round of PCR to attach Illumina adaptors and sequencing barcodes, bead purification of amplicons from the second round of PCR, library pooling and gel purification of final pooled library.

3.4.1.1 PCR 1: Amplification of the V4 16S hypervariable region

The V4 hypervariable region of the bacterial 16S rRNA gene was amplified by PCR using the protocol described in 3.3.1.5. The following negative controls with the following substitutions for clinical samples were included: a negative PCR control with 10µl of DNase free H_2O and a negative extraction control with 10µl of an extraction control from a DNA extraction matching samples included in the library. For the purpose of a positive control for both the extraction efficiency, amplification and data processing pipeline, DNA was extracted from a microbial mock community of known composition (ATCC, HM-280) according to the protocol described in 3.3.1.3. HM-280 mock community C is an microbial cell mixture of even amounts of 22 different bacterial strains in phosphate buffered saline (PBS) and full details of the community composition can be found in the Appendix. Amplification of DNA from oral samples (periodontal and saliva) were performed in duplicate and pooled in the subsequent bead purification step. Amplicons were quantified using the Quant-iTTM High-Sensitivity DNA Assay Kit (Invitrogen) and Qubit 2.0 Fluorometer (Invitrogen) as described in 3.3.1.4. PCR amplicon plates were sealed and stored at -20°C for up to a week.

3.4.1.2 PCR 1 purification: Bead purification and visualization

The Agencourt AMPure XP PCR purification system (Beckman Coulter) were used to purify amplicons from the first round of PCR in order to removing excess primers, nucleotides, salts, and enzymes that may inhibit downstream amplification and sequencing processes. The manufacturer instructions were followed with the following changes: PCR products were thawed on ice and an aliquot of AMPure XP beads was brought to room temperature. PCR plates were centrifuged to collect condensation and the seals were carefully removed. Duplicate oral samples were pooled resulting in a single plate containing vaginal samples (25µl of amplicon) and pooled oral samples (50µl of amplicon). AMPure XP beads were vortexed until beads were evenly dispersed and an equivalent amount of beads was added to each amplicon sample (25µl for vaginal samples and 50µl for oral samples). Beads were mixed well by pipetting and incubated at room temperature for 10 minutes. The plate was then placed on the Magnetic Stand-96 (Invitrogen) for 2 minutes or until supernatant had cleared. The supernatant was discarded and 200µl of freshly prepared 80% ethanol was added to each sample well. The plate was incubated on the magnetic stand for 30 seconds before the supernatant was removed an discarded. This ethanol wash step was repeated before beads were air dried for 2 minutes. The plate was removed from the magnetic stand and 27.5µl of 10 mM Tris pH 8.5 was added to each well and mixed by pipetting until beads were fully resuspended. The plate was incubated at room temperature for 2 minutes before being placed on the magnetic stand. After 2 minutes or once the supernatant had cleared, 25µl of the supernatant was carefully transferred into a new PCR plate. Purified amplicons were stored at -20 ^{O} C for up to a week.

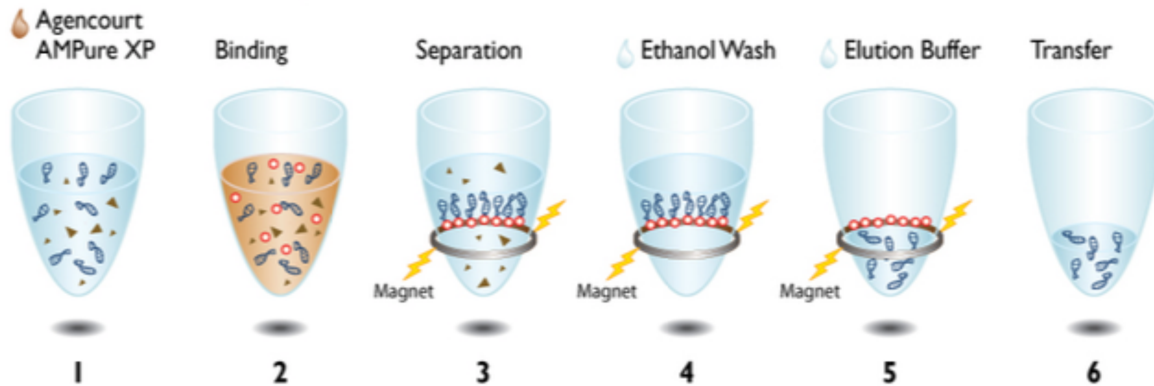


Figure 2.8: Workflow for PCR amplicon purification using the Agencourt AMPure XP PCR purification system. Edited from: Instructions for use: Agencourt AMPure XP PCR Purification, Beckman Coulter. Accessed online at: https://genome.med.harvard.edu/documents/sequencing/Agencourt_AMPure_Protocol.pdf

After bead purification, purified amplicons were quantified and visualised with gel electrophoresis as described in 3.3.1.4 and 3.3.1.5. Successful amplification was determined by the presence of a clear band (~320bp). For samples where no band was present, the first round of PCR and bead purification was repeated once using the maximum sample volume for the PCR reaction (10 µl). If no amplification was evident after repeated amplification and bead purification was, samples were excluded from the library. Purified amplicon plates were sealed and stored at -20 ^{O} C for up to a week.

3.4.1.3 PCR 2: Limited cycle multiplexing PCR

A second round of PCR was run in order to attach dual indices and Illumina sequencing adapters to the amplified V4 hypervariable region of the bacterial 16S rRNA gene using the Nextera XT Index Kit (llumina). Eight forward primers and twelve reverse primers were used in unique combinations to distinguish samples for MiSeq multiplexing. These primers were added to the amplicon products by a short cycle PCR described in Tables 2.7 and 2.8. A master mix was created containing: 5µL (2.5µM/µl) each of the unique forward and reverse primers (full combination supplied in appendix) and 25µl high fidelity DNA polymerase KAPA Hotstart Mix (Thermo Scientific™ Phusion™). Variable amounts of template amplicon from the initial round of PCR were added (max volume 5µl), to reach a 15-20ng of template amplicon in each reaction. Variable amounts of DNase free H_2O was then added to reach a final reaction volume of 50µL per sample (Table 2.7). PCR reactions were run using the GeneAmp® PCR System 9700 according to the

following program: 1 cycle 95°C for 3 minutes, 8 cycles of 95°C for 30 seconds, 55°C for 30 seconds and 72°C for 30 seconds, 1 cycle of 72°C for 5 minutes hold at 4°C (Table 2.8). A PCR 2 negative control was included where 5µl of DNase free H_2O was substituted for template amplicon in the above described mastermix.

Table 2.7: Reagent list for PCR 2

Reagent	Amount to add
Template DNA	Max 5µl (to ~15-20ng DNA)
Forward primer	5µl
Reverse primer	5µl
KAPA Hotstart Mix	20µl
PCR H_2O	10µl (up to final volume of 25µl)
Total volume:	50µl

Table 2.8: Thermocycler protocol for PCR 2

PCR phase	Duration/temperature
Initial denaturation	95°C for 3 minutes
Cycles (x8)	
Denaturation	95°C for 30 seconds
Annealing	55°C for 30 seconds
Elongation	72°C for 30 seconds
Final Elongation	72°C for 30 seconds
Storage	4°C

3.4.1.4 PCR 2: Bead purification

Amplicons from PCR 2 were purified using Agencourt AMPure XP beads (Beckman Coulter) as described in 3.4.1.2. After bead purification, purified amplicons were quantified and visualised with gel electrophoresis as described in 3.3.1.4 and 3.3.1.5. Successful amplification was determined by the presence of a clear band (~450bp). For samples where no band was present, the second PCR and bead purification was repeated once using the maximum samples volume (5µl). If no amplification was evident after the second PCR and purification were repeated, samples were excluded from the library. Purified amplicon plates were sealed and stored at -20 ^{O} C for up to a week.

3.4.1.5 Library normalizing and pooling

Amplicons were pooled in equimolar amounts. This was achieved by identifying the sample with the lowest amplicon concentration in the library and adding all remaining samples to the pooled aliquot at the same concentration. The concentration of the final pooled library was quantified using the Quant-iTTM High-Sensitivity DNA Assay Kit (Invitrogen) and Qubit 2.0 Fluorometer (Invitrogen) as described in 2.1.4 and stored at 4°C.

3.4.1.6 Pooled library purification (gel extraction)

The QIAquick PCR Purification Kit (Qiagen) was used to purify the final pooled library, in order to remove non-specific amplification and primer-dimers. Gel purification was run according to the manufacturers recommendations. Briefly, all centrifugations were carried out in a microcentrifuge at 13 000g. Buffer PB was added to the final pooled library at a 5:1 ratio. This mixture was added to a QIAquick column placed in a collection tube and centrifuged for one minute. The flow-through was discarded and the column was washed with 750µl before a one minute centrifugation. The QIAquick column was placed in a clean 1.5 ml microcentrifuge tube and DNA was eluted using 30µl Buffer EB (10 mM Tris·Cl, pH 8.5) and a one minute centrifugation. Loading dye was added to the flow through at a 5:1 ratio and mixed by pipetting. The DNA library was run on a 1% agarose gel (Tris–Acetate–EDTA (TAE) buffer, 60 V, 60 min). DNA was visualised using a handheld UV light and excised from the gel with a disposable scalpel. Gel fragments were weighed and placed in a collection tube with Buffer QG added at a 3:1 ratio (100mg ~100µl). The gel fragments were incubated at 50° C for 10 minutes or until the gel had completely dissolved. Isopropanol was added to the mixture at a ratio of 1:1 and mixed by pipetting. The sample was transferred to a QIAquick column in a collection tube and centrifuged for one minute. The flow through was discarded and the filter was washed using 500µl of Buffer QG and a one minute centrifugation. The flow through was again discarded and second wash was done as in the previous step using 750µl of Buffer PE after a 5 minute incubation at room temperature. The column was then dry centrifuged for one minute. After transferring the column into a clean collection tube, DNA was eluted in 30µl of elution buffer (10 mM Tris·Cl, pH 8.5) with a one minute centrifugation. The concentration of the purified library was determined using the Qubit 2.0 Fluorometer (Invitrogen) and the Quant-iTTM High-Sensitivity DNA Assay Kit (Invitrogen) according to the manufacturers protocol. Purified DNA libraries were stored at 4°C.

3.4.2 Optimisation of Illumina 16S rRNA gene MiSeq library preparation protocol

3.4.2.1 Initial library quality assessment

The Bioanalyzser 2100 (Agilent) was used to assess the quality of amplicons in the pooled library. Protocols were followed according to manufactures specifications. Briefly, the RNA 6000 Nano Assay (Agilent Technologies Inc., part No. G2941-90126) was used to prepare an RNA 6000 Nano

gel matrix according to manufacturer's specifications. Libraries were denatured at 70°C for 2 minutes and then centrifuged quickly before being placed on ice. Gel dye mix, RNA 6000 Nano Markers (15bp and 1500bp) and Ladder were added to the chip according to manufacturer's specifications. Libraries were added to the chip in 1µl aliquots and the chip was vortexed at 2400 rpm. The chip was loaded into the Bioanalyzer 2100 within 5 minutes of preparation and software analysis protocols were run according to manufacturer's specifications. Bioanalyzer results from the first library prepared revealed a peak correlating to an amplification product ~410bp, in addition to a peak correlating to the expected amplicon product (~450 bp) (Figure 2.9). This was an indication of non-specific amplification and the Section 3.3.2 describes the optimisation that took place to correct for this.

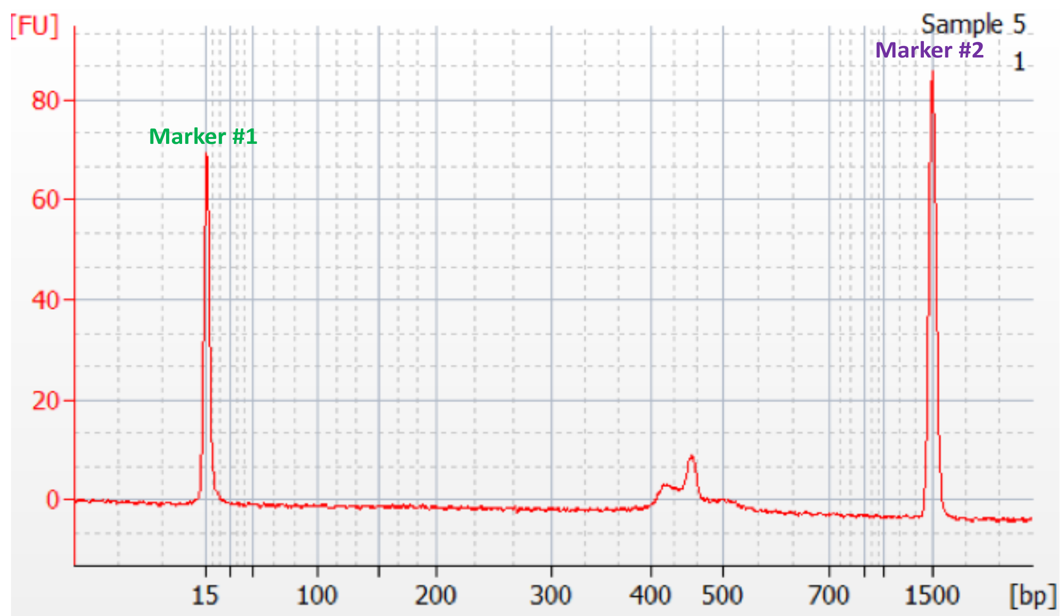


Figure 2.9: Size of denatured RNA fragments (bp) in initial DNA library measured by fluorescence [FU] using the Bioanalyzser 2100 (Agilent).

3.4.2.2 Optimisation Experiment 7: Optimisation of Limited cycle multiplexing PCR

In order to reduce non-specific product formation during the second round of PCR in the 16S rRNA MiSeq library preparation protocol, we assessed the impact following changes to the protocol described in 3.4.1:

- Increasing the primer concentration from 0.1 to 0.5µM
- Increasing cycle numbers from 8 to 15
- Varied concentration of template amplicon from the first round of PCR amplification

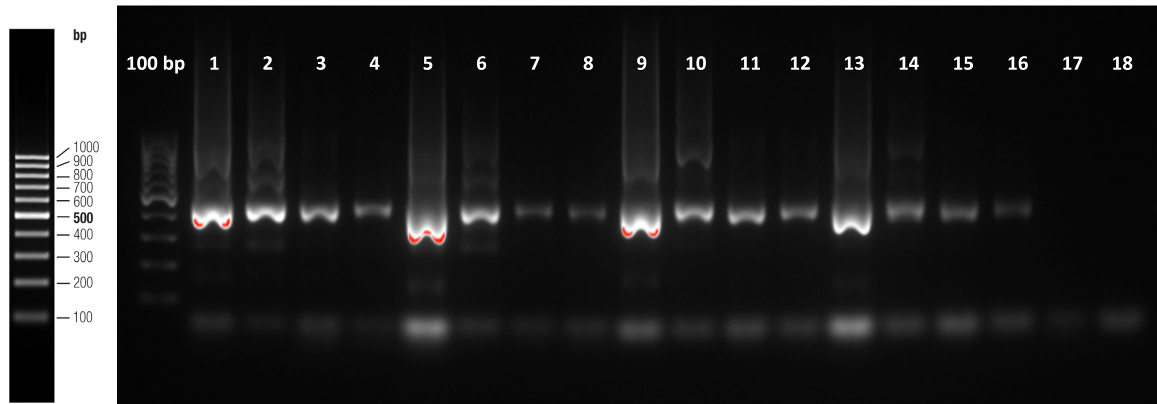
The reverse primers tested were chosen based on quality control results for previous libraries in which samples with primers S501 and S504 gave poor results. The forward primers were randomly chosen. The second round of PCR was run as described in 3.1.3 with the following alterations: 18 separate PCR reactions with different combinations of primer concentrations, cycle numbers and starting sample concentration according to Table 2.9.

Table 2.9: Reagent list for PCR 2 optimisation

Combination:	1	2	3	4	Control
Primer conc (uM)	0.1	0.5	0.1	0.5	0.1
No. of cycles	15	15	8	8	15
Amplicon conc (ng/ul)	13.7	13.7	6.9	6.9	-
Mastermix					
Template DNA (µl)*	10	10	5	5	-
Forward primer (µl)	5	5	5	5	5
Reverse primer (µl)	5	5	5	5	5
KAPA Hotstart Mix (µl)	20	20	20	20	20
PCR H_2O	-	-	-	-	10

Gel electrophoresis revealed that increasing the primer concentration from 0.1 µM to 0.5µM showed slight improvement in amplification specificity with fewer non-specific low molecular weight amplifications present in the reactions with higher primer concentrations (Figure 2.10). Increasing the number of PCR cycles from 8 to 15 resulted in a number of additional amplification products above the expected product size of 460 bp (Figure 2.10). Reducing the initial concentration of amplicons from the first PCR (template DNA for the 2nd PCR) also resulted in fewer non-specific amplifications (Figure 2.10).

Reaction	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Reverse primer	701	701	701	701	702	702	702	702	701	701	701	701	702	702	702	702	701	702
Forward primer	501	501	501	501	501	501	501	501	504	504	504	504	504	504	504	504	501	504
Primer conc (µM)	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.1
No. PCR Cycles	15	15	8	8	15	15	8	8	15	15	8	8	15	15	8	8	15	15
Amplicon conc (ng/µl)	13.7	13.7	13.7	13.7	6.9	6.9	6.9	6.9	13.7	13.7	13.7	13.7	6.9	6.9	6.9	6.9	H_2O	H_2O



...ation. Gel electrophoresis was carried out as ... Scientific) was included at 0.5 µg/mL in the left

...eriment, the 2nd PCR was run using 8 ... were diluted to ~3.0-6.0 ng/ul before the ... d to 0.2uM.

... pooled library was determined using the Qubit 2.0 Fluorometer ... library quantification kit (Roche) according to the manufacturers protocol. DNA libraries were diluted to a concentration of 4.0 nM using nuclease free H_2O. Diluted libraries were then denatured by adding 0.2N sodium hydroxide (NaOH) and a 5 minute incubation at room temperature. Denatured libraries were diluted to concentration of 1.3nM by adding 200nM Tris-HCl (pH 7) before an additional dilution to a concentration of 20pM with pre-chilled hybridization buffer. Finally, samples were heat denatured for 2 minutes at 96°C before MiSeq sequencing. For each run, a 20% % PhiX internal control spike-in was included at the same concentration of the DNA library. All reagents used were from the MiSeq Reagent Kits V3 (Illumina). Amplicon sequencing was performed on the Illumina MiSeq Desktop Sequencer (300-bp paired-end reads with V3 chemistry).

3.6 Bioinformatic pipeline

De-multiplexed, raw reads were pre-processed using usearch7 and modules included in the QIIME package (Quantitative Insights Into Microbial Ecology, http://qiime.org). The quality of raw reads was assessed using FastQC (Andrews 2010). Using USEARCH, 250 bp paired-end reads were merged and then quality filtered (merged reads were truncated to 300bp and reads with error scores larger than 0.1 discarded). Next, sequences were de-replicated (the process of identifying and sorting replicate sequences) whilst recording the level of replication for each sequence. De-replicated sequences were sorted by abundance (highest to lowest) and clustered de novo into operational taxonomic units (OTUs) at 97% similarity (Edgar 2018). Chimeric sequences were removed. Individual sequences were assigned to specific OTUs using a 97% similarity threshold. Taxonomic assignment was performed using the RDP classifier (against the Greengenes 13.8 database) at a confidence level of 0.5 (Wang et al. 2007). The representative sequence set was then aligned against the Greengenes 13.8 database using PyNAST (Caporaso *et al.* 2010). As previous publications have described the vaginal microbiome of girls from the same study population in depth, manual taxonomic annotation was not performed for OTUs during vaginal microbiome analysis, except for cases where no annotation was available after taxonomic assignment using the Greengenes database. For oral OTU's where species level annotation was not achieved using the previously described method, BLASTn searches were performed in the NCBI 16S ribosomal RNA sequence (bacteria and Archaea) database after excluding all uncultured bacteria. If more than one species mapped to an OTU, OTUs were annotated as follows: *Genus species A_species B_species C* for a maximum of three species. If more than three species mapped to an OUT or there was disagreement between the Greengenes and BLASTn annotation, the FASTA sequence was searched using BLASTn in the Human Oral Microbiome Database (eHOMD) and taxa previously identified from the human oral microbiome were selected. Finally, a phylogenetic tree was constructed using FastTree to relate OTUs from the multiple sequence alignment (Price et al. 2010). Based on rarefication analysis for optimal read count depth, samples with > 2000 reads were selected for downstream analyses.

3.7 Statistical analysis

All downstream statistical analysis was performed in R. Beta diversity non-metric dimensional scaling (NMDS) was performed using Bray-Curtis and alpha diversity principal component(PC) analysis was performed using Weighted-Unifrac. Alpha diversity boxplots were generated using Shannon's diversity and the Kruskal Wallis non-parametric variance test. Microbial community types were determined by fuzzy clustering with optimal k. For heat map and relative abundance plots, OTUs were merged at the lowest available taxonomic level using a custom script developed by *Lennard et al.* (Lennard et al. 2017). OTU tables were then standardized (i.e., transformed to relative abundance and multiplied by the median sample read depth) and filtered so that each OTU had at least 10 counts in at least 10% of samples. Heatmaps were constructed using Weighted-Unifrac as a distance metric and UPGMA unsupervised hierarchical clustering. Differential abundance testing was calculated using negative binomial models after filtering out OTUs where variance was lower than 30 and a predetermined level of significance (adjusted p-value <0,05). Fold differences > 1.25 and < 0.75 were considered clinically significant. Ecological diversity was calculated using the phyloseq (Mcmurdie & Holmes 2013) package, cluster (Rousseeuw et al. 2018) was used for community type clustering, vegan (Oksanen et al. 2018) for ordinations and redundancy analysis, NMF (Gaujoux 2014) for annotated heat maps and DeSeq (Anders & Huber 2010) for differential abundance testing.

CHAPTER 4: Results

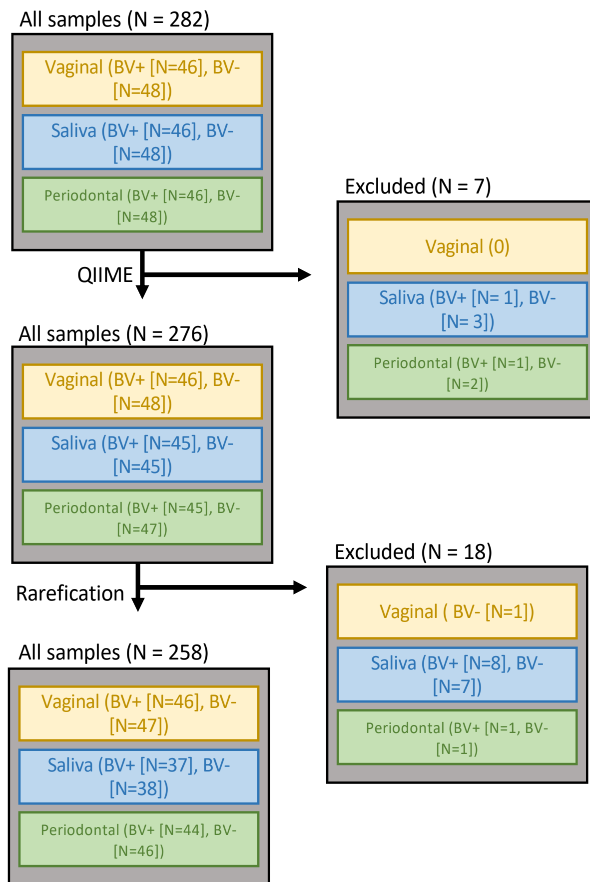
4.1 Library quality assessment

4.1.1 Rarefication and selection of samples based on read counts

Bacterial DNA was extracted from lateral vaginal wall (LW), saliva (SAL) and periodontal (PE) samples of 94 participants (N total samples = 282, Figure 3.1). The V4 region of the 16S rRNA gene in these samples were sequenced in six separate libraries keeping matched LW, SAL and PE samples for each participant in the same library where possible. Reads from the six libraries were combined for downstream analysis. After QIIME quality filtering, four SAL and two PE samples were excluded (N samples after filtering = 276, Figure 3.1). Rarefaction curves demonstrated that the mean OTU depth plateaued at around 2000 read counts and based on this, all samples with read counts below 2000 were excluded. Of the eighteen samples excluded with less than 2000 reads one originated from LW, 15 from SAL and two from PE samples (Figure 3.1). The final library used for downstream analysis contained 258 samples (93 LW, 75 SAL and 90 PE), 1 634 5216 reads (with a median of 41066 reads per sample) and 1904 unique OTUs.

4.1.2 Quality and compositional analysis of controls

For each library, a mock bacterial community consisting of an even mixture of 22 known bacterial species was included to assess variability across sequencing runs (Appendix, Table S1). The heatmap in Figure 3.2 A illustrates the standardised read counts of the top 30 OTUs amplified from the mock communities in each MiSeq run. With the exception of one sample, Mock-run1 where read counts were low overall, bacterial species known to be present in the mock community (Table 3.1) were present at much higher read counts in each sample in comparison to unexpected bacterial species (Figure 3.2A). Of the 22 bacterial species present in the mock community, only *Propionibacterium acnes* was not identified in any of the samples. Twelve OTUs were identified at the species level after searching representative sequences using BLASTn against the NCBI 16S ribosomal RNA (bacteria and Archaea) (Edgar 2018) database with identity scores >97%, including *Acinetobacter baumannii, Pseudomonas aeruginosa, Bifidobacterium adolescentis, Streptococcus mutans, Streptococcus agalactiae, Helicobacter pylori, Bacillus cereus, Rhodobacter sphaeroides, Porphyromonas gingivalis, Escherichia coli,*



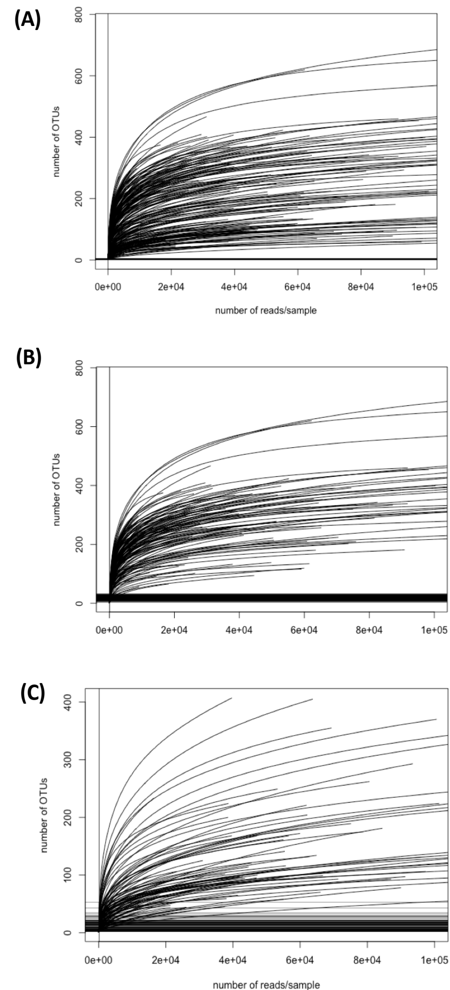


Figure 3.1: Schematic of sample selection after sequencing quality assessment and rarefication curves illustrating species richness (number of OTUs) with increasing number of reads per sample for (A) all samples and (B) and (C) vaginal and oral samples, respectively. Rarefaction was performed using the vegan (Oksanen et al. 2018) package in R.

Deinococcus radiodurans and *Bacteroides vulgatus*. Species level annotation was not available for the OTUs mapping to *Enterococcus*. Two OTUs mapped to the *Neisseria* genus, with one identifiable at *N. cinerea* and species level resolution not possible for the other. There was only one OTU of *Staphylococcus* despite two *Staphylococcus* species being present in the mock community. We were thus unable to distinguish between *Staphylococcus aureus* and *S. epidermis* and this OTU was therefore annotated as *S. aureus_epidermis_cluster* for the purpose of this study. Additionally, as observed in other studies (Petricevic et al. 2012; Kiss et al. 2007; Vasquez et al. 2002), a single OTU mapped to both *Lactobacillus gasseri* and *L. jonsonnii*. Due to the fact that *L. gasseri* was known to be present in the mock community this OTU is referred to as *L. gasseri* for the purpose of this study. The same logic was applied in downstream analysis where

some sequences mapped to more than one unique species and we used previous publications to determine the most likely identity. Putative sequences mapping to both *Lactobacillus crispatus* and *L. acidophilus* were referred to as *L. crispatus* due to previous literature noting the difficulty in differentiating these two species and identifying *L. crispatus* as a keystone vaginal species in the South African population (Lennard et al. 2018; Balle et al. 2018). Similarly the OTU identified as both *A. odontolyticus* and *A. meyeri* was annotated as *A. odontolyticus,* a keystone oral bacterium*.* The OTU mapping to the *Clostridium* genus was identified as *C. butyricum.* Due to the documented difficulty in distinguishing between *C. butyricum* and *C. beijerinckii,* the species known to be present in the mock community, this OTU was annotated as *C. beijerinckii_butyricum_cluster. Listeria selegrii* was identified instead of the expected *L. monocytegens* by both BLAST and Greengenes, likely the result of a sequencing error. In each sample, a number of bacteria known to be present at high proportions in the oral (*Delftia acidovorans)* and vaginal (*Lactobacillus iners, Prevotella amnii, Megasphaera, BVAB-1, Sneathia sanguinegens* and *Gardnerella vaginalis*) microbiotas respectively, were present at low read counts, possibly a result of spill over from oral and vaginal samples. In all samples, high read counts of *Lysinibacillus* were present. We were unable to determine the source from which this OTU originated, but may have been an environmental contaminant. As it was present across all mock community samples and was not picked up in downstream analysis or as one of the top 30 OTUs in our analysis of negative controls (Figure 3.2 B), it likely the result of contamination specific to the mock communities introduced during handling of the mock community DNA. Similar to previous studies using the same even mock community (Fouhy et al. 2016), DNA extraction and MiSeq sequencing resulted in non-uniform amplification of different bacterial phyla with *Enterrococus, N.cinerea, P. aeruginosa* and *E. coli* overrepresented across all runs (Figure 3.2 A) despite a large variability in read count across mock communities (Appendix, Table S2). Negative controls, containing PCR grade nuclease free water in the place of genomic material were included for DNA extraction and/or PCR amplification in each MiSeq run. PCR controls for DNA extraction from sequencing run 1 were excluded in the QIIME quality filtering step, likely due to very low read counts. Although a range of bacteria were present in the negative controls (Figure 3.2B), the read counts for all controls (with the exception of PCR-2 Run 6) were well below

the 2000 read cut off point and there was no strong evidence of contamination.

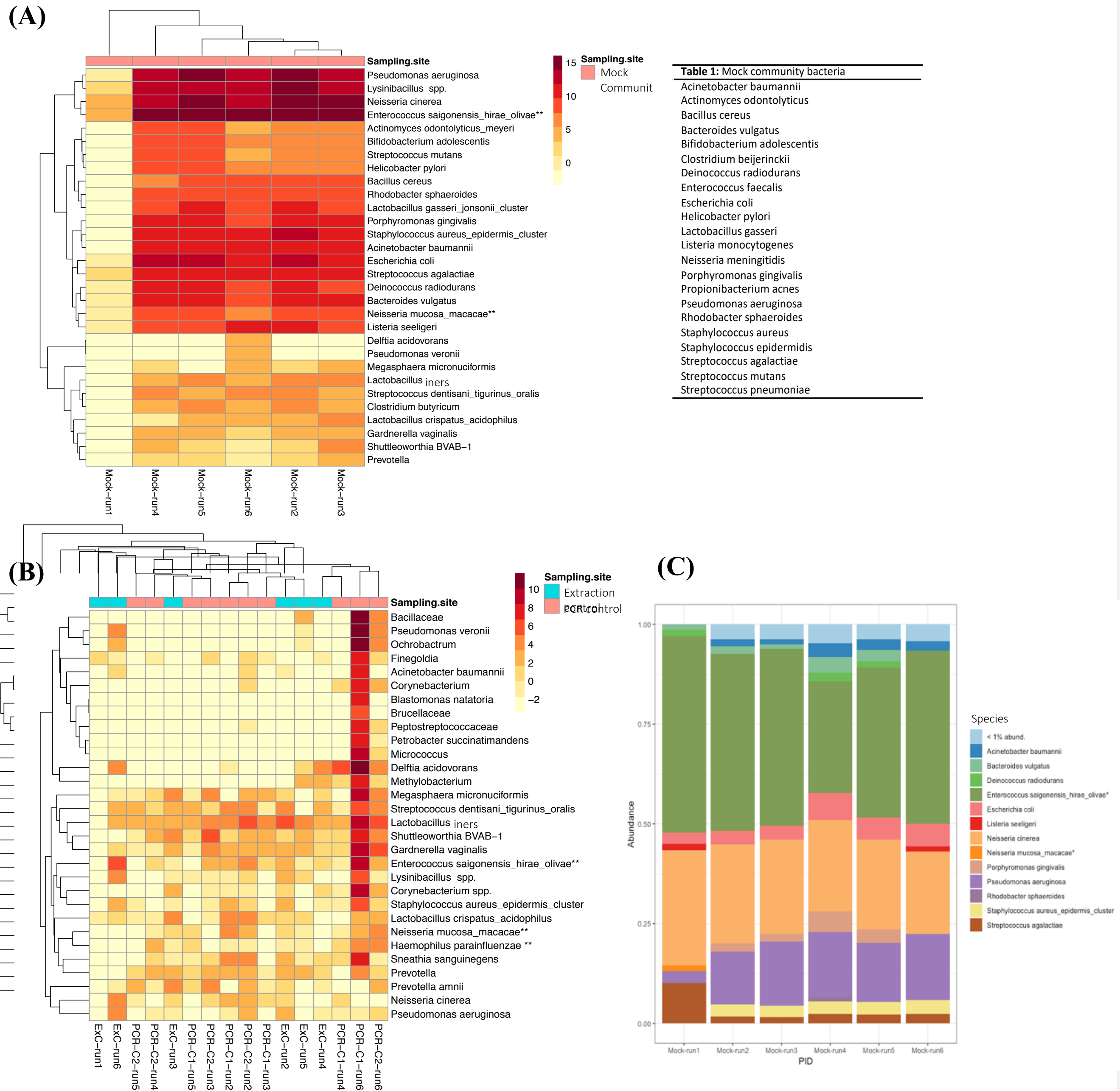


Table 1: Mock community bacteria
Acinetobacter baumannii
Actinomyces odontolyticus
Bacillus cereus
Bacteroides vulgatus
Bifidobacterium adolescentis
Clostridium beijerinckii
Deinococcus radiodurans
Enterococcus faecalis
Escherichia coli
Helicobacter pylori
Lactobacillus gasseri
Listeria monocytogenes
Neisseria meningitidis
Porphyromonas gingivalis
Propionibacterium acnes
Pseudomonas aeruginosa
Rhodobacter sphaeroides
Staphylococcus aureus
Staphylococcus epidermidis
Streptococcus agalactiae
Streptococcus mutans
Streptococcus pneumoniae

Figure 3.2: Composition and relative abundance of sequencing controls. Relative standardised read counts of the top 30 bacterial phyla and genera present in controls clustered by unweighted pair group method with arithmetic mean (UPGMA) hierarchical clustering using Weighted-Unifrac distances for **(A)** mock communities (N=6) and **(B)** negative controls (N=16). Colour scale indicates the relative read count of bacterial taxa present in each experimental control. **(C)** Relative abundance of bacterial taxa present in mock communities agglomerated at the lowest taxonomic level possible. The x-axis denotes sample type and the y-axis denotes % relative abundance of bacterial taxa contributing to the total makeup of the microbiota (100%). * Indicate species annotation with lower than 97% identity using BLASTn to search the eHOMD database.

4.1.3 Technical bias analysis

Unsupervised hierarchical clustering of all samples illustrated that there was no clustering according to DNA extraction kit type (MoBio or Zymo) or MiSeq Run (Figure 3.3 A). In addition to this, NMDS analysis of all samples indicated that there were no distinct pattern in the beta diversity of samples introduced by either extraction kit type or MiSeq run (Figure 3.3 B). This analysis provides no evidence of technical bias introduced by DNA extraction or MiSeq sequencing and based on that, all samples were analysed together regardless of extraction kit or MiSeq run.

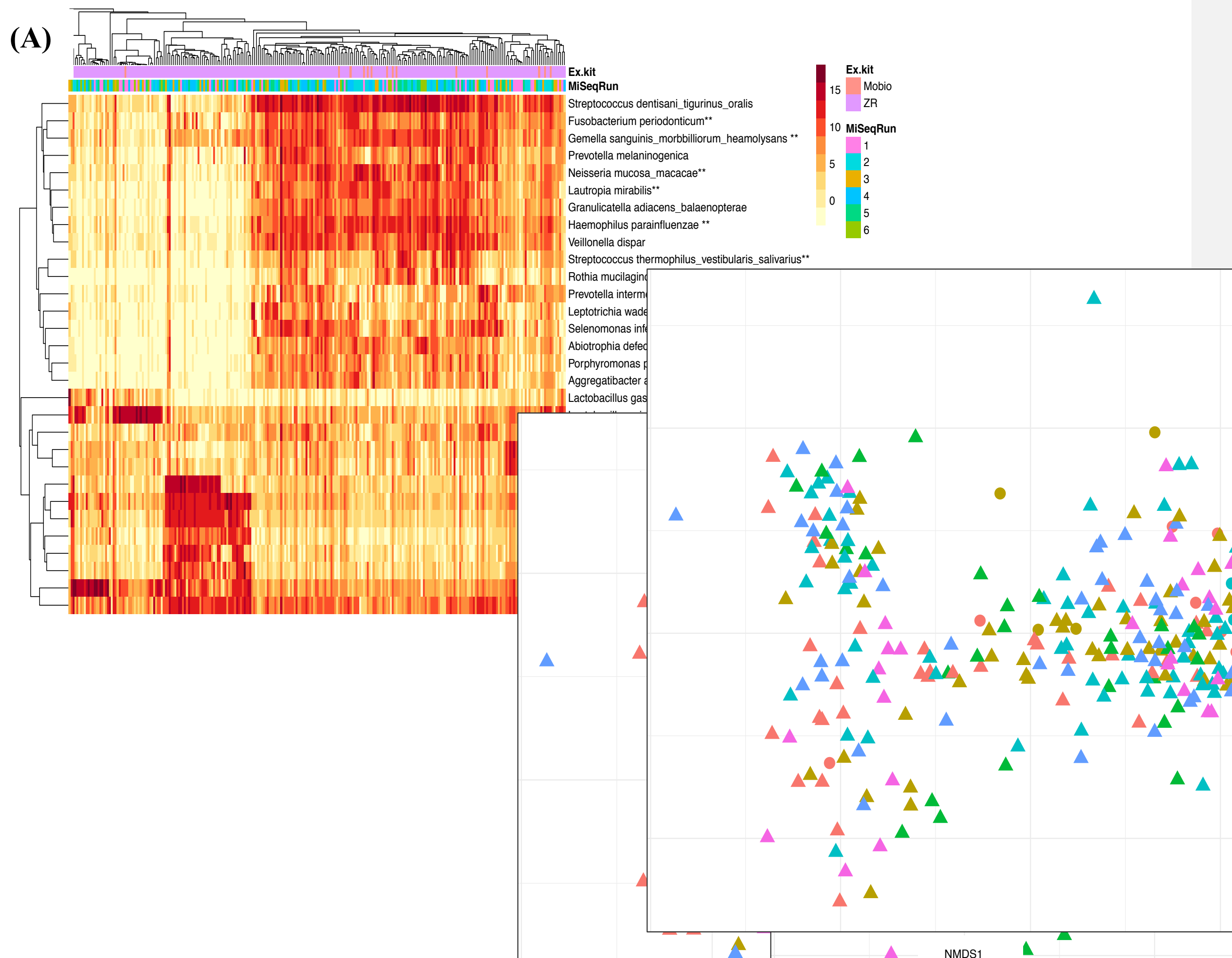


Figure 3.3: Analysis for technical bias introduced b
Relative standardised read counts of the 30 top mo
= 258) clustered by UPGMA hierarchical clustering
by log2() for readability) of bacterial taxa in each sa
(B) using NMDS and Bray Curtis as a distance metric. The shape of the data points indicate the DNA extraction kit used and the colour indicates the MiSeq run in which samples were included. * Indicate species annotation with lower than 97% identity using BLASTn to search the eHOMD database.

4.2 Study cohort characteristics

Table 3.1: Cohort baseline characteristics

	Clinical BV Diagnosis	
	Negative (N=46)	Positive (N=47)
Mean Age (Std. Deviation)	17 (1,4)	17 (3,0)
Mean BMI (Std. Deviation)	26 (4,6)	26 (5,4)
Previous Pregnancy	11% (N=5)	15% (N=7)
Any bacterial STI	48% (N=22)	53% (N=25)
Chlamydia	35% (N=16)	36% (N=17)
Gonorrhoea	11% (N=5)	21% (N=10)
Trichomonas vaginalis	7% (N=3)	7% (N=3)
HSV-2*	35% (N=16)	34% (N=16)
Mycoplasma genitalium	4% (N=2)	4% (N=2)
Vaginal samples included	46	47
Saliva samples included	38	37
Periodontal samples included	46	44

After quality filtering as described previously, a total of 258 samples were included with at least one LW, SAL or PE samples originating from one of 93 participants. The mean age of participants in this cohort was 17 years with a higher distribution of ages in BV positive participants. Age, BMI, number of previous pregnancies and laboratory diagnosis of at least one bacterial STI were similar amongst BV positive and BV negative participants. In accordance with previous studies in this population (Balle et al. 2018; Lennard et al. 2017; Masson et al. 2015), the prevalence of bacteria STIs was relatively high compared to reports from the general population (Johnson & Geffen 2016). Chlamydia was the most prevalent STI in this cohort at 34%, followed by Herpes Simplex Virus 2 (HSV-2) at 34%, gonorrhoea at 15%, trichomonas at 6% and mycoplasma at 4%. While the prevalence of chlamydia, *Trichomonas vaginalis*, HSV-2 and *Mycoplasma genitalium* was similar in both experimental groups, the prevalence of gonorrhoea in girls with a clinical diagnosis of BV (21%) was double that of BV negative participants (10%). After quality filtering there were no large differences in the number of LW, SAL and PE samples included for each experimental group.

4.3 Composition of the oral microbiota

4.3.1 Ecological niche and community type descriptions of the oral microbiota

Fuzzy clustering of the oral samples with optimal k (described in method section) resulted in two distinctive oral community types (OCT 1 and OCT 2) that did not correlate to sampling site (Figure 3.4 B) or by hierarchical clustering (Figure 3.5 A). The median alpha diversity measured by Shannon's diversity Index was significantly ($p<0.001$) higher in OCT 1 [Median (variance) = 3.65 (0.29)] compared to OCT 2 [Median (variance) = 2.78 (0.27)] (Figure 3.4 B). Both OCTs were dominated by Firmicutes with a relative abundance of 50% and 59% in OCT 1 and OCT 2, respectively. OCT 1 had a more diverse makeup at a genus level comprising of *Streptococcus* (15%), *Prevotella* (13%), *Lactobacillus* (9%), *Selenomonas* (5%), *Staphylococcus* (5%), *Haemophilus* (5%) and a number other genera with less than 5% abundance (Figure 3.5 B, Table 3.2) compared to OCT 2 comprised predominately of *Streptococcus* (40%), followed by *Haemophilus* (10%), *Neisseria* (7%), *Lautrophia* (5%) and a number of low abundant genera (Figure 3.5 B, Table 3.2). *S. dentisani_tigurinus_oralis* was the most prevalent species in both OCTs, comprising 15.6% and 39% of the reads in OCT 1 and OCT 2, respectively (Figure 3.6 D, Table 3.2). Negative binomial modelling performed using Deseq2 (Paulson et al. 2013) identified a number of OTUs with significantly different relative abundances between OCTs, with only *Streptococcus dentisani_tigurinus_oralis* enriched in OCT 2 compared to OCT 1 with a clinically relevant fold difference of 0.25 (Appendix, Table S3). These results were supported by random forest modelling using all oral samples to predict OCT (Sensitivity (Sn) = 0.84, Specificity (Sp) = 0.78, N= 165 with 85 and 80 samples per class), with *S. dentasani_tigurinus_oralis* identified as the most important bacteria distinguishing OCT 1 from OCT 2, followed by *Granulicatella adiacens_balaenopterae*, *Gemella* and *Haemophilus.* The full results of the random forest model are summarised in Figure 3.6 C and Appendix Table S5. Unsupervised hierarchical clustering of all oral samples (N=119) did not reveal any site specific differences in the relative abundance of the top 30 most prevalent OTUs between the salivary and periodontal microbiota (Figure 3.5 A). Negative binomial modelling identified significant differences in the less prevalent taxa between the SAL and PE samples as summarised in Figure 3.6 B and Appendix Table S4. Firmicutes, including *Oribacterium* spp. (O. *parvum, O. asaccharolyticum* and *O. sinus*) and *Streptococcus*

anginosus were enriched in the salivary microbiota, in addition to *Prevotella nanceiensis* and Actinobacteria including *Actinomyces graevenitzii* and *Rothia mucilaginosa.* The Fusobacterium, *Leptotrichia* (species level annotation not available) was present at a significantly higher proportion in the periodontal space. The results of a random forest model (Sn = 0.77, Sp = 0.92, N= 165 with 90 and 75 samples per class) identified *R. mucilaginosa*, *O. sinus*, *S. thermophilus_vestibularis_salivarius, A. graevenitzii*, *P. nanceiensis* and *O. parvum* as the most important species for accurately predicting the separation of the periodontal and salivary microbiota. The full results of the model and the most important classifying features are summarised in Figure 6D and Appendix Table S6.

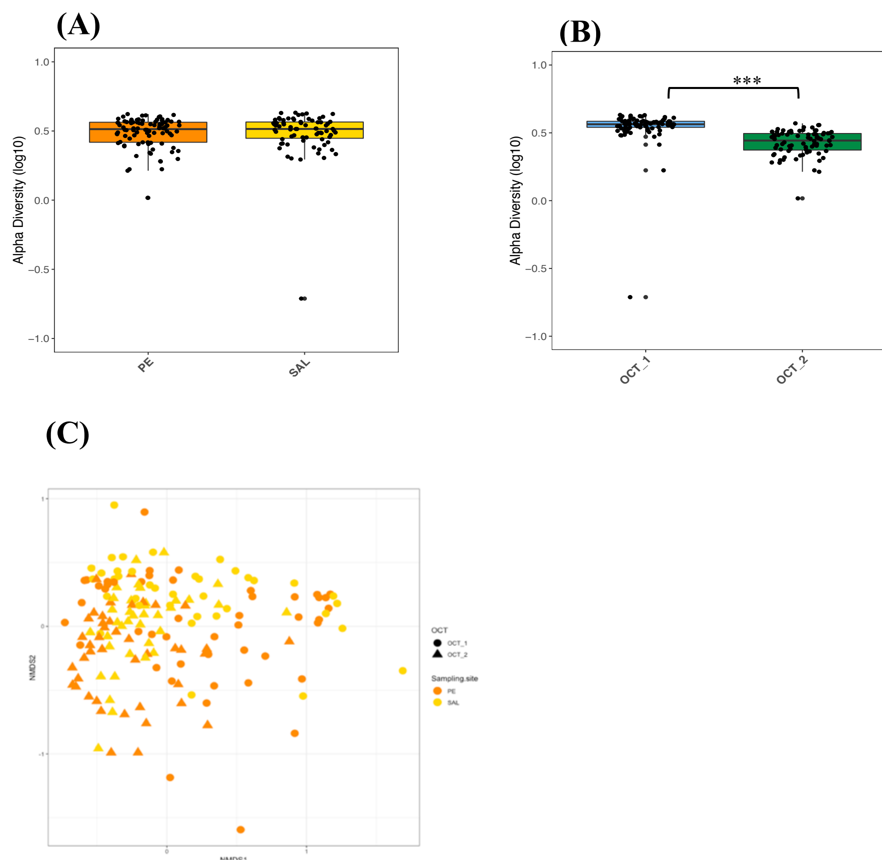


Figure 3.4: Ecological analysis of the oral microbiota (A) Comparison of alpha diversity (measured by Shannon's Index) of oral samples by (A) sampling site (SAL = saliva and PE = periodontal) and (B) oral community types (OCT 1 and OCT 2). Bars represent 95% confidence intervals. **(C)** NMDS analysis of beta diversity using Bray-Curtis as a distance metric of microbial communities in the mouth (saliva and periodontal samples, N = 165). The colours keys indicates the sampling site from which samples originated with yellow denoting saliva (SAL) and orange periodontal (PE) and the shape denotes the oral community type (OCT) derived from fuzzy clustering with an optimal k. *** denotes p <0,005.

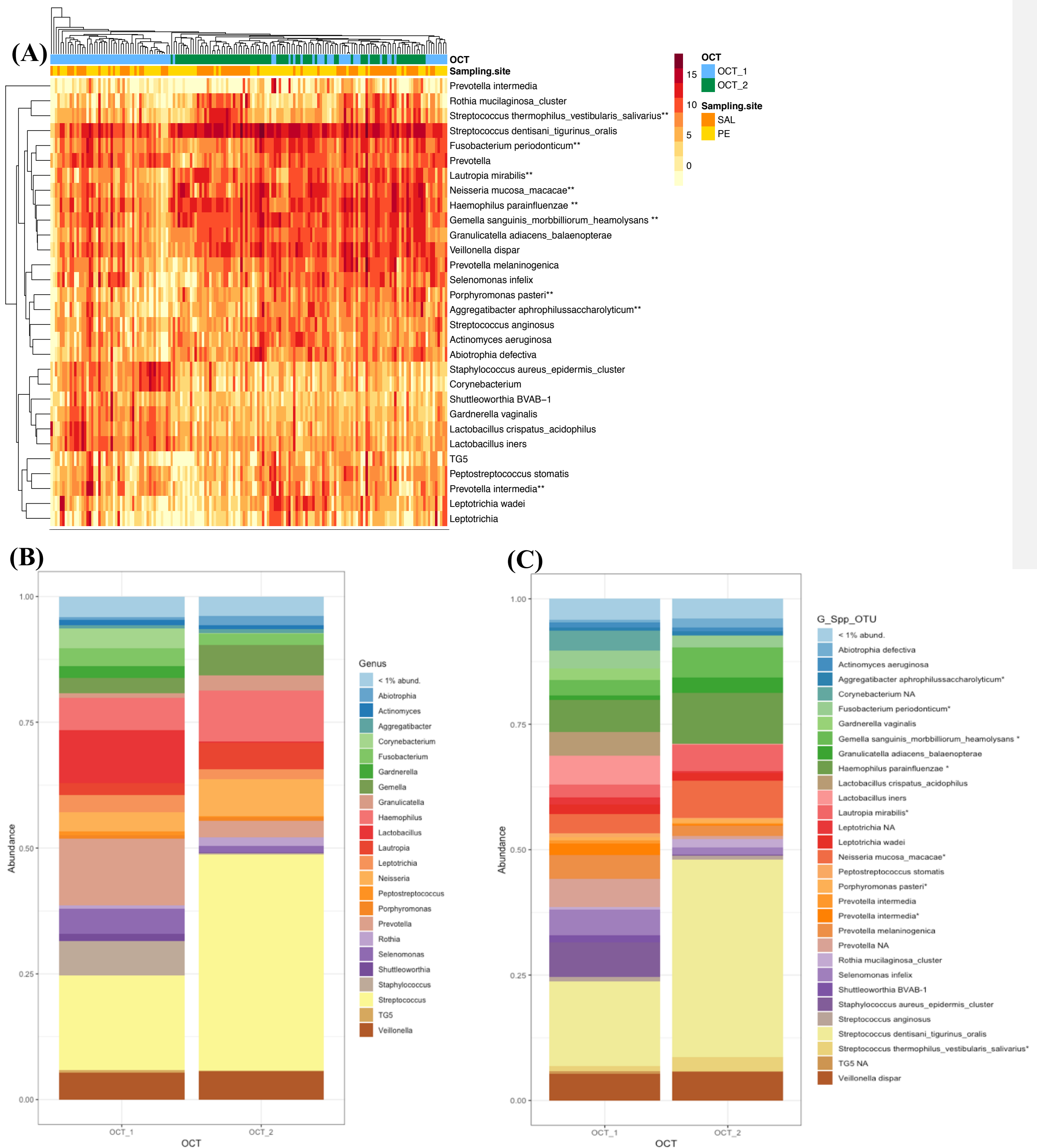


Figure 3.5: Bacterial composition of oral microbiota (A) Relative standardised read counts of the 30 top most abundant bacterial taxa in oral samples (N=165)) clustered by UPGMA hierarchical clustering with colour scale indicating the relative read count of different bacterial taxa in each sample and colour key denoting the sample site (orange for saliva (SAL), yellow for periodontal (PE)) and oral community type (blue for OCT 1 and green for OCT 2). The composition of oral community types were also visualised with the relative abundance of bacterial taxa present in mock communities agglomerated at the on a **(B)** genus and **(C)** species level. The x-axis denotes sample type and the y-axis denotes % relative abundance of bacterial taxa contributing to the total makeup of the microbiota (100%). * Species annotation with lower than 97% identity using BLASTn to search the eHOMD database.

Table 3.2: Mean relative abundance of top 30 OTUs in oral community types**

Phylum	Genus	Species	Mean relative abundance	
			OCT 1	OCT 2
Actinobacteria	*Actinomyces*	*oris_naeslundii _viscosus**	2,3%	2,4%
	Corynebacterium		3,5%	0%
	Gardnerella	*vaginalis*	2,0%	0%
	Rothia	*mucilaginosa_cluster*	0%	1,7%
Bacteroidetes	*Capnocytophaga*	*leadbetteri*	0%	0,5%
	Capnocytophaga		0%	0,5%
	Porphyromonas	*pasteri*	3,6%	1,3%
	Prevotella		6,6%	0%
	Prevotella	*melaninogenica*	4,2%	2,0%
	Prevotella	*intermedia*	2,7%	0,7%
	Prevotella		0%	1,6%
Firmicutes	*Abiotrophia*	*defective*	0%	2,0%
	Anaerococcus		1,6%	0%
	Catonella	*morbi*	0%	0,2%
	Finegoldia		1,8%	0%
	Lactobacillus	*iners*	4,8%	0%
	Lactobacillus	*crispatus*	4,4%	0%
	Megasphaera	*micronuciformis*	1,3%	0%
	Granulicatella	*adiacens_balaenopterae*	1,1%	2,9%
	Not annotated		0%	0,3%
	Peptostreptococcus	*stomatis*	0,9%	0,3%
	Selenomonas	*infelix*	5,3%	1,7%
	BVAB-1		1,2%	0%
	Staphylococcus		4,8%	0%
	Streptococcus	*dentisani_tigurinus_oralis*	15,6%	39,4%
	Streptococcus	*anginosus*	0%	0,9%
	Veillonella	*dispar*	4,6%	5,2%
			0%	0,1%
	Gemella	*sanguinis_morbbilliorum_heamolysans*	2,6%	5,7%
Fusobacteria	*Fusobacterium*	*periodonticum*	3,7%	2,4%
	Leptotrichia	*wadei*	4,3%	2,6%
	Sneathia		0,9%	0%
Proteobacteria	*Aggregatibacter*	*aphrophilussaccharolyticum*	1,4%	1,3%
	Campylobacter	*showae_rectus*	2,3%	0,8%
	Kingella	*denitrificans*	0%	0,8%
	Haemophilus	*parainfluenzae*	5,5%	9,4%
	Haemophilus	*influenzae*	0%	0,7%
	Lautropia	*mirabilis*	2,2%	4,9%
	Neisseria	*mucosa_macacae*	3,3%	7,0%
Spirochaetes	*Treponema*		0,9%	0,3%
SR1			0,6%	0%
Synergistetes	*TG5*		0%	0,2%

*Taxa merged at lowest annotation level possible.

* Species annotation with lower than 97% identity using BLASTn to search the eHOMD database

Figure 3.6: Taxonomical differences between oral community types and ecological niches. Oral samples (N=165) were used to build a negative binomial model to identify bacterial present at relatively different proportions between **(A)** oral community types defined using fuzzy clustering with an optimal k and **(B)** oral sampling site (salivary (SAL) versus periodontal (PE)). Dashed lines indicate a fold change > 1.25 or <0.75. Relative fold differences have been displayed as log2 (fold change) for illustrative purposes and the true fold differences are summarised in Appendix, Table S3. All OTUs included had a significantly different relative abundance (adjusted p-value < 0,05). The same sample set was used to train a random forest model to predict the most important taxa differentiating **(C)** oral community types (OCT1 vs OCT2) and **(D)** oral sampling site (SAL vs PE), with the mean decrease in the Gini coefficient for the top 30 most important taxa displayed in **(B)**. For **(A)** and **(C)**, OCT 1 and PE samples were used as reference levels for (A)+(C) and (B)+(D) respectively. * Indicate species annotation with lower than 97% identity using BLASTn to search the eHOMD database.

4.3.2 Description of the core oral microbiota

Ecological community analysis (Figure 3.4 A and 3.4 C) and unsupervised hierarchical clustering of oral samples (Figure 3.5 A) did not reveal any sampling site specific separation between SAL and PE samples. Based on this, samples from both sampling sites were analysed together for descriptions of a core oral microbiota. Of the 775 OTUs present in at least 10% of all oral samples at a relative abundance higher than 0.01%, 560 were ubiquitous across oral sampling sites and 101 and 114 were present in only PE and SAL samples, respectively. For the description of the core microbiota, OTUs present in at least 50% of both SAL and PE samples with a relative abundance larger than 0.01% were included. Figures 3.7 illustrates how the prevalence of core oral microbiota changes when using different relative abundance detection thresholds and Table 3.3 summarizes the relative abundance of core oral taxa on a genus and species level. The oral core microbiota was dominated by Firmicutes (41%) followed by Proteobacteria (26%), Bacteriodes (18%), Fusobacteria (9%) and Actinobacteria (6%) (Table 3.3). On a genus and species level, the core oral microbiota was comprised of *Streptococcus* [*S. dentisani_tigurinus_oralis* (29%)*, S. anginosus* (1%)], *Haemophilus* [*H. parainfluenzae* (8%), *H. influenzae* (1%)], *Prevotella* [*P. melaninogenica* (3%), *P. intermedia* (2%), *P. oris* (1%)], *LactobacilLu*s [*L. iners* (3%), *L. crispatus* (3%)], *Neisseria* [*N. mucosa_macacae* (5%)] and *Veillonella* [*V. dispar* (5%)] (Table 3.3)

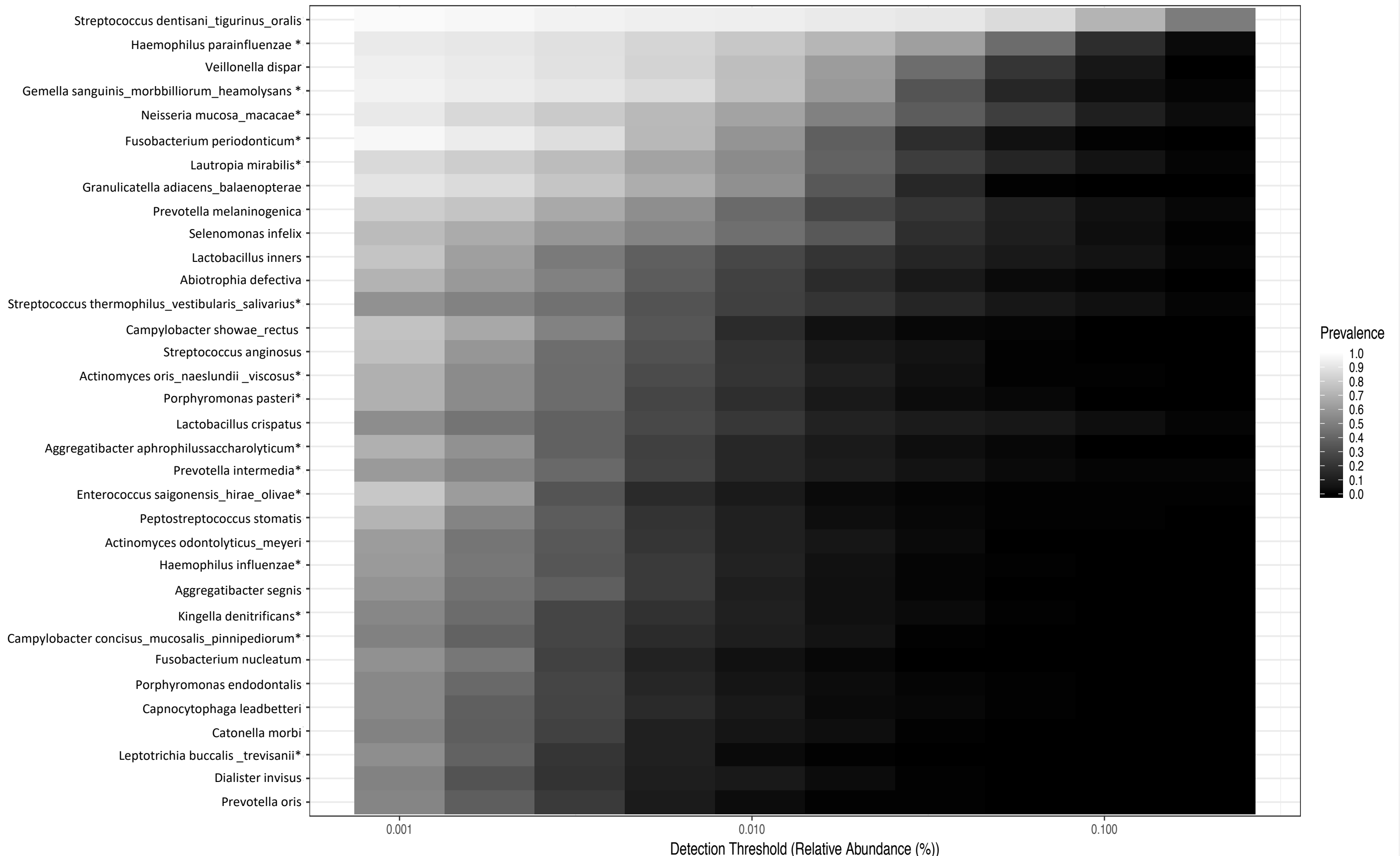


Figure 3.7: Core analysis of the oral microbiota Relative prevalence of core OTUs at different cut offs of minimal relative OTU prevalence. All core analysis were done on combined periodontal and saliva samples (N= 165) including OTUs with at least 10 counts present in at least 50% of samples. * Indicate species annotation with lower than 97% identity using BLASTn to search the eHOMD database**.**

Table 3.3: Mean relative abundance of core oral taxa**

OTU	Phylum	Genus	Species	Mean Relative Abundance
OTU_32	Actinobacteria	*Actinomyces*	*oris_naeslundii _viscosus**	2%
OTU_24		*Corynebacterium*		2%
OTU_6		*Gardnerella*		1%
OTU_39		*Rothia*	*mucilaginosa_cluster*	1%
OTU_29	Bacteroidetes	*Porphyromonas*	*pasteri**	1%
OTU_18		*Prevotella*	*intermedia**	2%
OTU_14		*Prevotella*	*melaninogenica*	3%
OTU_100		*Prevotella*	*oris*	1%
OTU_19	Firmicutes	*Abiotrophia*	*defective*	1%
OTU_67		*Catonella*	*morbi*	0%
OTU_1		*Lactobacillus*	*iners*	3%
OTU_3		*Lactobacillus*	*crispatus*	3%
OTU_713		*Granulicatella*	*adiacens_balaenopterae*	2%
OTU_26		*Peptostreptococcus*	*stomatis*	1%
OTU_28		*Selenomonas*	*infelix*	4%
OTU_4		*BVAB-1*		1%
OTU_16		*Staphylococcus*		4%
OTU_58		*Streptococcus*	*anginosus*	1%
OTU_2		*Streptococcus*	*dentisani_tigurinus_oralis*	29%
OTU_17		*Veillonella*	*Dispar*	5%
OTU_8		*Gemella*	*sanguinis_morbbilliorum_heamolysans **	4%
OTU_31	Fusobacteria	*Fusobacterium*	*periodonticum**	3%
OTU_38		*Leptotrichia*	*wadei*	3%
OTU_30	Proteobacteria	*Aggregatibacter*	*aphrophilussaccharolyticum**	1%
OTU_41		*Campylobacter*	*showae_rectus*	2%
OTU_40		*Kingella*	*denitrificans**	1%
OTU_5		*Haemophilus*	*parainfluenzae **	8%
OTU_934		*Haemophilus*	*influenzae**	1%
OTU_12		*Lautropia*	*mirabilis**	4%
OTU_9		*Neisseria*	*mucosa_macacae**	5%

* Species annotation with lower than 97% identity using BLASTn to search the eHOMD database
** Taxa merged at lowest annotation level possible.

4.4 Description of the vaginal microbiome

In accordance with previously published research (Ravel et al. 2011; Lennard et al. 2018), participants with a clinical diagnosis of BV determined by Nugent scoring presented with a diverse vaginal microbiota comprised of a high relative abundance of BV-associated bacteria including *Gardnerella, Prevotella,* BVAB-1*, Aerococcus, Megasphaera, Sneathia* and *Atopobium* (Figure 3.9). In comparison to participants diagnosed with BV, BV negative participants presented with a vaginal microbiota dominated by *Lactobacillus* spp. (Figure 3.9). In agreement with previous publications on the vaginal microbiota of participants from the same population as this study (Balle et al. 2018; Lennard et al. 2017), three distinct vaginal community types (termed VCT 1, VCT 2 and VCT 3 for distinction from oral community types) were delineated using fuzzy clustering with an optimal k (Figure 3.9). Of the 93 samples that were retained after quality assessment, 51% (N = 48) belonged to community type 2 (VCT 2), distinguished by a diverse vaginal microbiota dominated by *BVAB-1* (21%), *Prevotella* (19%), *Gardnerella* (19%), *Megasphaera* (11%), *Sneathia* (9%) and *Lactobacillus* (9%) (Figure 3.9 C). The remainder of the participants were split between VCT 1 (25%, N= 23) and VCT 3 (22%, N=22) and displayed low diversity community types dominated by *Lactobacillus* (93,7% and 92% respectively), with VCT 1 consisting of majority *L. crispatus* (87.3%) and VCT 3 *L. iners* (80%) (Figure 3.9 C). Shannon's diversity Index was significantly higher in VCT 2 [Median (variance) = 2.14 (0.27)] compared to both VCT 1 [Median (variance) = 0.49 (0.47)] and VCT 3 [Median (variance) = 0.61 (0.34)] (Figure 3.8 A). NMDS analyses of beta diversity revealed VCT 1 and VCT 3 to be similar in terms of community membership while both distinct from VCT 2 (Figure 3.8 B). Linear regression revealed VCT 2 to have significantly higher Shannon's diversity ($p<0.005$) in comparison to both VCT 1 and VCT 3, between which there were no significant differences. The majority of the participants with a VCT 2 vaginal microbiota had a clinical diagnosis of BV at the time samples were collected while 0% and 4% of participants with to VCT 1 and VCT 3 microbiotas respectively, were diagnosed as BV negative through Nugent scoring. We used negative binomial modelling to determine the quantitative relative difference between bacterial taxa in vaginal communities of participants with and without clinical BV diagnoses and different VCTs. Participants with a clinical BV diagnosis had significantly increased proportions of Actinobacteria (*Gardnerella, Atopobium, Mobiluncus*

and *OTU_104* (Not annotated), Bacteroidetes (*Prevotella* and *OTU_181* (Not annotated), Fusobacteria (*Sneathia*) and TM7 in comparison to participants without clinical BV (Figure 3.10 A, Appendix, Table S7). The Firmicutes; *Aerococcus, Clostridium, Dialister, Gemella, Megasphaera, Parvimonas,* BVAB-1 and *Veillonella* were enriched in the vaginal microbiota of BV positive participants, while a number *Lactobacillus* spp. (*L. iners, L. reuteri, L. jensenii* and *L. crispatus*) were enriched in participants with no clinical vaginal microbial dysbiosis. Similar differences were observed when comparing VCT 2, the VCT associated with a clinical BV diagnosis with VCT 1 and VCT 3 (Figure 3.10 B-C, Table S7). Compared to VCT 1 and VCT 3, VCT 2 was enriched in Actinobacteria (*Gardnerella, Atopobium, Corynebacterium* and *OTU 104* (Not annotated), Bacteroidetes (*Prevotella* spp. and *OTU_181* (Not annotated), Fusobacteria (*Sneathia*) and a number of Firmicutes (*Aerococcus, Clostridium, Dialister, Gemella, Megasphaera, Parvimonas* and BVAB-1) (Figure 3.10 B and C). In comparison to VCT 2, both VCT 1 and VCT 3 were enriched in a number of *Lactobacillus* spp., *Corynebacterium*, *Clostridium* and *Mogibacterium*, with VCT 1 having higher proportions of *Acinetobacter* and VCT 3 having higher proportions of *Porphymonas, 1-68* and *Staphylococcus*. Additionally, this analysis illustrated the relative differences between the *Lactobacillus* dominant community types, with VCT 1 having a relatively larger presence of *L. crispatus* and VCT 3 of *L. iners* (Figure 3.10 D).

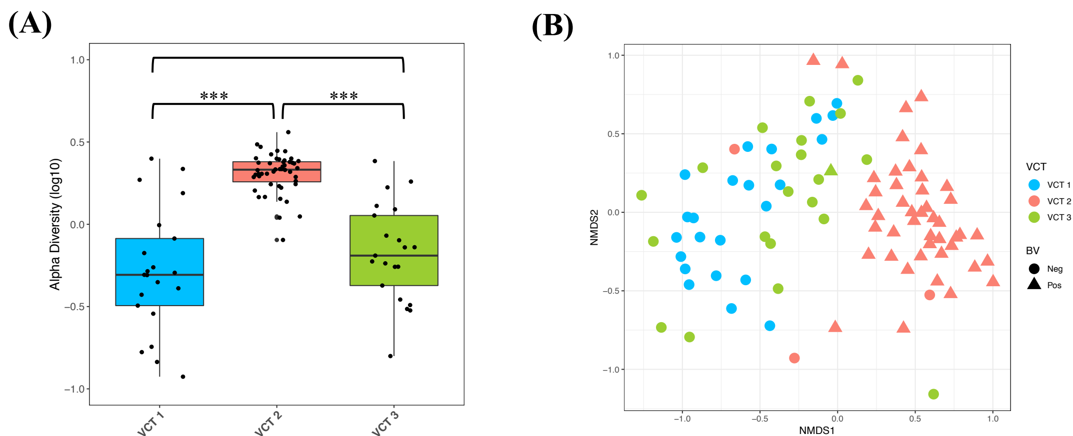


Figure 3.8: Ecological diversity of vaginal microbiota by clinical BV status and vaginal community types. Comparison of ecological diversity of vaginal samples (N=93) by alpha diversity measured by Shannon's Index **(A)** and NMDS analysis of beta diversity using Bray-Curtis as a distance metric **(B)** of bacterial community diversity in the vagina. Colour denotes vaginal community type (VCT 1 = blue, VCT 2 = pink, VCT 3 = green) determined by fuzzy clustering using optimal k and shape denotes clinical bacterial vaginosis diagnosis (round = negative, triangle = positive). *** denotes p <0,001.

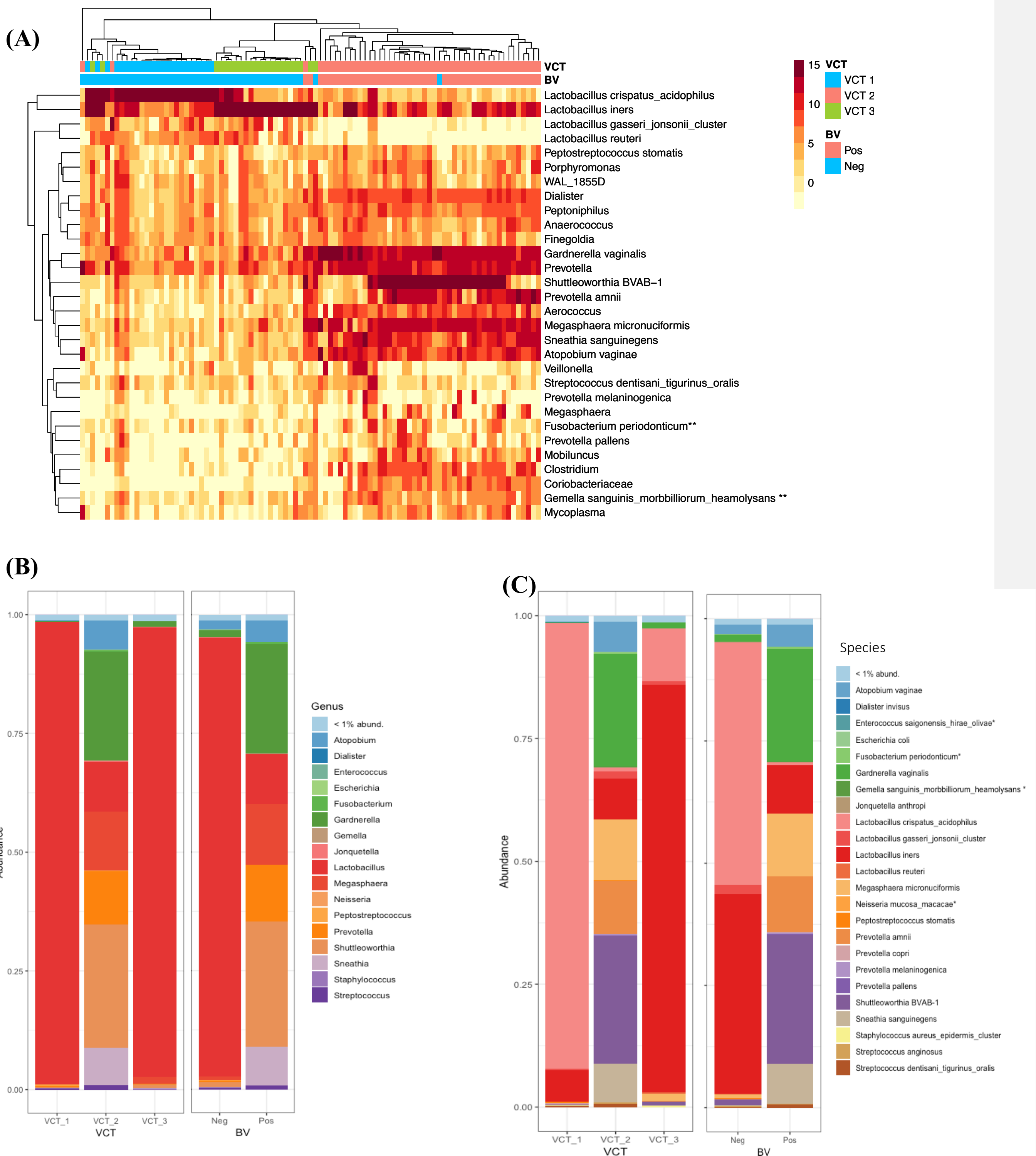


Figure 3.9: Composition of vaginal bacterial community types (A) Relative standardised read counts of the 30 top most abundant bacterial taxa in vaginal samples (N=95) clustered by UPGMA hierarchical clustering with colour scale indicating the relative read count of different bacterial taxa in each sample and colour key denoting vaginal community type (VCT) and clinical BV status. The composition of vaginal community types and differences in vaginal microbiota of participants diagnosed with BV compared to those without were additionally visualised with the relative abundance of bacterial taxa agglomerated at the on a **(B)** genus and **(C)** species level. The x-axis denotes sample type and the y-axis denotes % relative abundance of bacterial taxa contributing to the total makeup of the microbiota (100%).

(A)

log2FoldChange

5
0
−5

Not Annotated
Atopobium vaginae
Gardnerella vaginalis
Mobiluncus mulieris
Not Annotated
Prevotella amnii
Prevotella buccalis
Prevotella colorans
Prevotella pallens
Prevotella timonensis
Aerococcus christensenii.
BVAB−1
Clostridium
Clostridium perfringens
Dialister micraerophilus
Dialister propionicifaciens
Gemella asaccharolytica
Lactobacillus crispatus_acidophilus
Lactobacillus iners
Lactobacillus jensenii
Lactobacillus reuteri
Megasphaera
Megasphaera (
Megasphaera micronuciformis
Mogibacterium
Parvimo s micra
Shuttleworthia BVAB−1
Veillonella montpellierensis
Sneathia sanguinegens
Not Annotated

phylum
Actinobacteria
Bacteroidetes
Firmicutes
Fusobacteria
TM7

(B)

log2FoldChange

5
0
−5

Not Annotated
Atopobium vaginae
Corynebacterium variabile
Gardnerella vaginalis
Not Annotated
Prevotella amnii
Prevotella buccalis
Prevotella colorans
Aerococcus christensenii.
BVAB−1
Clostridium butyricum
Clostridium perfringens
Dialister micraerophilus
Dialister propionicifaciens
Gemella asaccharolytica
Lactobacillus crispatus_acidophilus
Lactobacillus jensenii
Lactobacillus reuteri
Megasphaera
Megasphaera micronuciformis
Mogibacterium
Parvimo s micra
ph2
Shuttleworthia BVAB−1
Sneathia sanguinegens
Acinetobacter baumannii

phylum
Actinobacteria
Bacteroidetes
Firmicutes
Fusobacteria
Proteobacteria

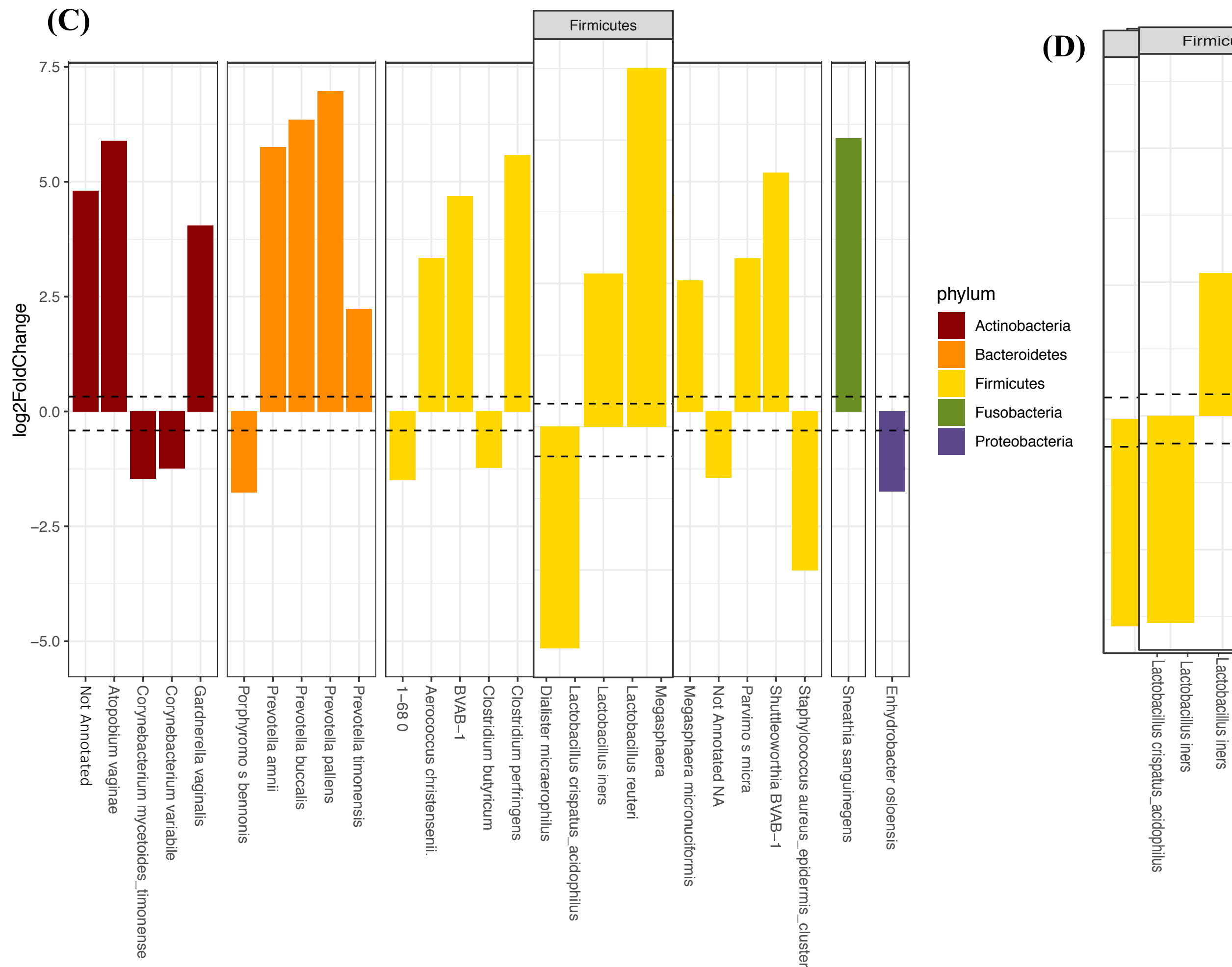


Figure 3.10: Differentially abundant vaginal bacterial taxa between vaginal community type. Vaginal samples (N=93) were used to build a negative binomial model to identify bacteria present at relatively different proportions (p adjusted <0.05) in the vaginal microbiota between **(A)** participants with and without a clinical diagnosis of BV (N=93). We additionally identified differentially abundant taxa between **(B)** VCT 2 and VCT 1 (N =71), **(C)** VCT 2 and **(D)** VCT 3 (N=70) and VCT 1 and VCT 3 (N=45). Relative fold differences have been displayed as log2(fold change) for illustrative purposes and the true fold differences are summarised in Appendix, Table S7. Dashed lines indicate a fold change > 1.25 or <0.75.

4.5 Comparison of the oral and vaginal microbiota

Bacterial communities in the mouth and vagina of South African adolescent females were distinct in terms of both alpha and beta diversity (Figure 3.11 A and 8B). The bacterial composition of the oral microbiota was found to be more diverse than that of the vaginal microbiota with 866 unique OTUs identified from oral samples compared to 364 OTUs identified from vaginal samples. When combining SAL and PE samples, median alpha diversity as measured by Shannon's Index was significantly higher than that of vaginal samples (p<0,001, Figure 3.11 A). Tables 3.4 and 3.5 summaries the top most abundant OTUs present across all oral and vaginal samples respectively. Unsupervised hierarchical clustering of standardised read counts resulted in distinct oral and vaginal clusters (Figure 3.11 C). Figure 3.11 C illustrates the top 30 most abundant OTUs present across all libraries and standardised read counts of these OTUs across all samples. The majority of oral samples clustered into one distinct group characterised by high read counts of bacterial genera known to be present in the oral microbiota including *Streptococcus, Veillonella, Neisseria, Haemophilus, Porphyromonas, Actinomyces* and *Gemella* (as described in section 4.3, Figure 3.11 C, 3.11 A and 3.11 B). The vaginal samples separated into two distinct clusters with one group having higher read counts of *Lactobacillus* spp. and the other enriched with BV-associated bacteria including BVAB-1*, Gardnerella, Megasphaera, Atopobium, Prevotella* and *Sneathia* (Figure 3.11 C, 3.11 A and 3.11 B). Within the lactobacilli dominant group there was distinct separation of two groups, one dominated by *L. iners* and the other by *L. crispatus*.

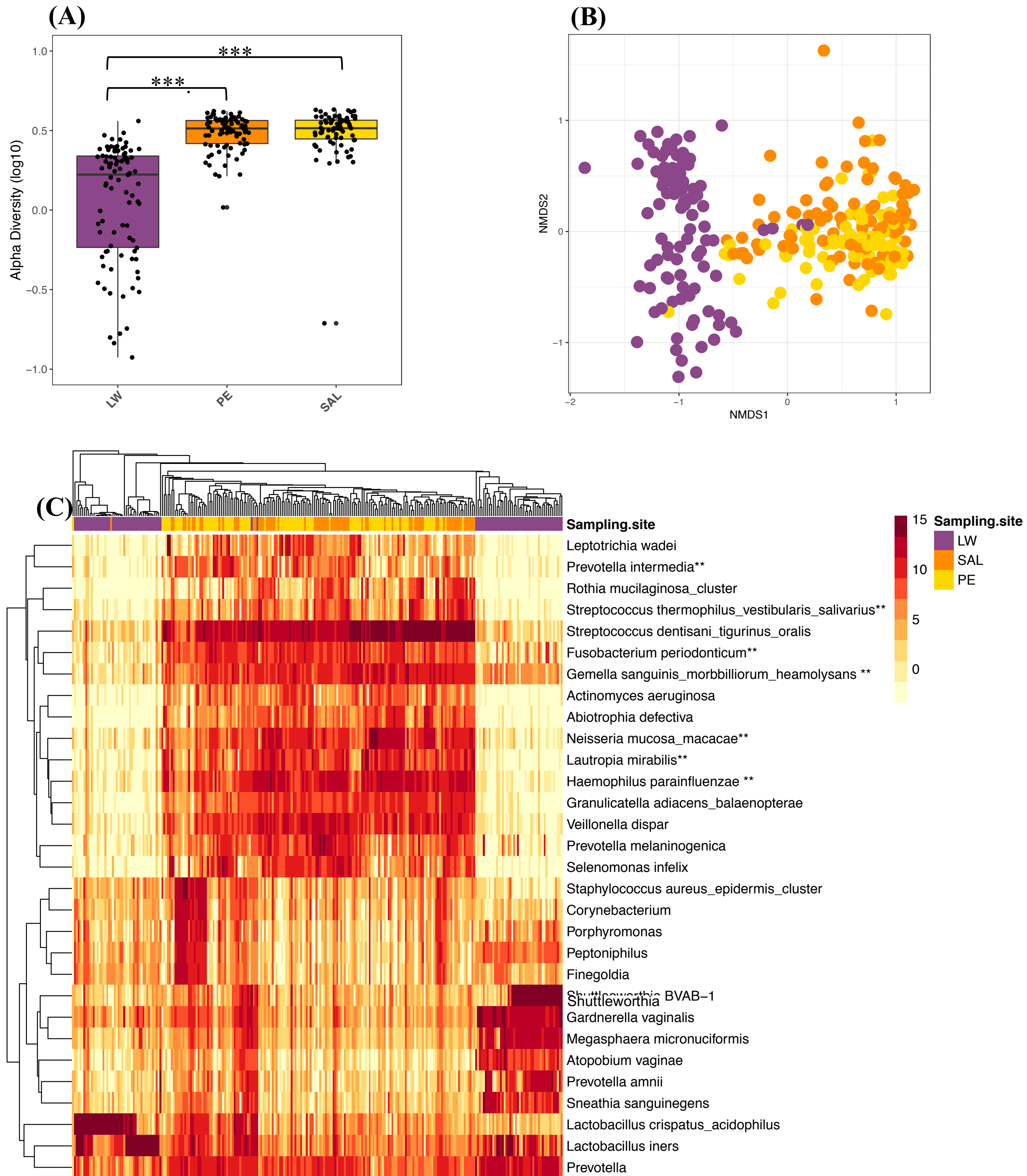


Figure 3.11: Microbial communities in the vagina, saliva and periodontal space (N= 258). Comparison of alpha diversity of oral and vaginal samples by Shannon's diversity **index (A)** and NMDS analysis of beta diversity using Bray-Curtis as a distance metric **(B).** Standardised mean read counts of oral and vaginal samples were grouped using UPGMA hierarchical clustering **(C)**. Taxa were merged at the lowest taxonomic level available and only OTUs with more than 10 counts present in at least 10% of samples were selected. Colour scale indicates the relative read count of bacterial taxa present in each sample. The colours key indicates sampling site from which samples originated with purple denoting vagina (LW), yellow saliva (SAL) and orange periodontal (PE). * Indicate species annotation with lower than 97% identity using BLASTn to search the eHOMD database.

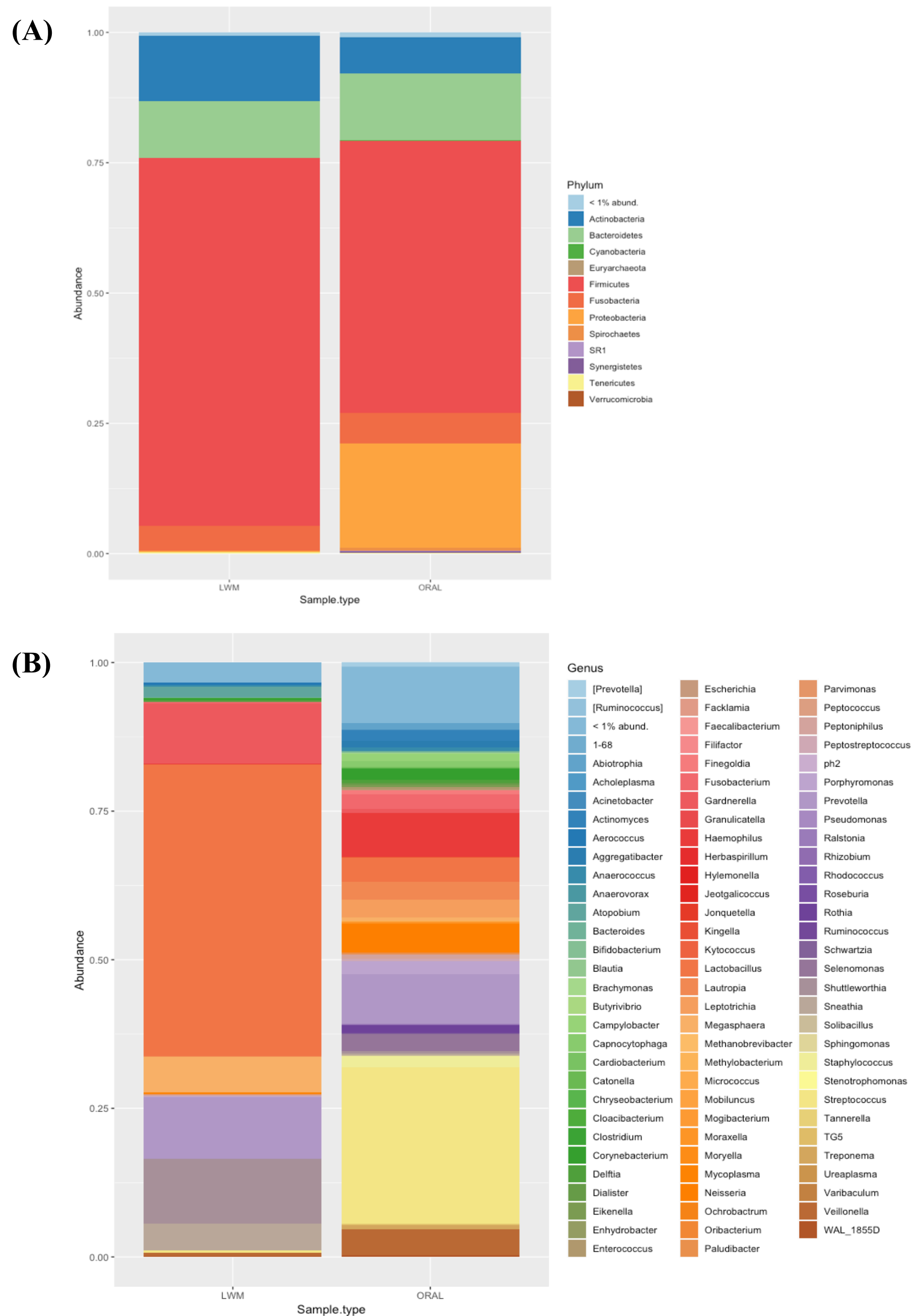


Figure 3.12: Microbial composition of oral and vaginal microbiota. Comparison of the relative abundance of sample counts merged at the **(A)** phylum and **(B)** genera level including only OTUs where at least 10% of samples have more than 10 counts of that OTU. The x-axis denotes sample type and the y-axis denotes % relative abundance of bacterial taxa contributing to the total makeup of the microbiota (100%).

Table 3.4: Mean relative abundance of top 30 most prevalent OTUs in all oral samples**

OTU	Phylum	Genus	Species	Mean Relative abundance
OTU_32	Actinobacteria	*Actinomyces*	*oris_naeslundii _viscosus**	2%
OTU_24		*Corynebacterium*		2%
OTU_6		*Gardnerella*	*Vaginalis*	1%
OTU_39		*Rothia*	*mucilaginosa_cluster*	1%
OTU_29	Bacteroidetes	*Porphyromonas*	*pasteri**	3%
OTU_18		*Prevotella*	*intermedia**	2%
OTU_14		*Prevotella*	*Melaninogenica*	3%
OTU_10		*Prevotella*		4%
OTU_19	Firmicutes	*Abiotrophia*	*defective*	1%
OTU_1		*Lactobacillus*	*Iners*	3%
OTU_3		*Lactobacillus*	*Crispatus*	2%
OTU_7		*Megasphaera*	*Micronuciformis*	1%
OTU_713		*Granulicatella*	*adiacens_balaenopterae*	2%
OTU_26		*Peptostreptococcus*	*Stomatis*	1%
OTU_28		*Selenomonas*	*Infelix*	4%
OTU_4		*BVAB-1*		1%
OTU_16		*Staphylococcus*		3%
OTU_58		*Streptococcus*	*Anginosus*	1%
OTU_2		*Streptococcus*	*dentisani_tigurinus_oralis*	27%
OTU_17		*Veillonella*	*Dispar*	5%
OTU_8		*Gemella*	*sanguinis_morbbilliorum_heamolysans **	4%
OTU_31	Fusobacteria	*Fusobacterium*	*periodonticum**	3%
OTU_38		*Leptotrichia*	*Wadei*	4%
OTU_11		*Sneathia*		1%
OTU_30	Proteobacteria	*Aggregatibacter*	*aphrophilussaccharolyticum**	1%
OTU_41		*Campylobacter*	*showae_rectus*	2%
OTU_5		*Haemophilus*	*parainfluenzae **	7%
OTU_12		*Lautropia*	*mirabilis**	4%
OTU_9		*Neisseria*	*mucosa_macacae**	5%
OTU_143	Spirochaetes	*Treponema*		1%

* Species annotation with lower than 97% identity using BLASTn to search the eHOMD database
** Taxa merged at lowest annotation level possible

Table 3.5: Mean relative abundance of top 30 most prevalent OTUs in all vaginal samples**

OTU	Phylum	Genus	Species	Mean relative abundance
OTU_22	Actinobacteria	*Atopobium*	*vaginae*	2,0%
OTU_24		*Corynebacterium*		0,1%
OTU_6		*Gardnerella*	*vaginalis*	10,3%
OTU_106		*Mobiluncus*		0,2%
OTU_104	Not annotated			0,2%
OTU_54	Bacteroidetes	*Porphyromonas*		0,3%
OTU_14		*Prevotella*	*melaninogenica*	0,1%
OTU_10		*Prevotella*		10,1%
OTU_121		*Prevotella*	*pallens*	0,3%
OTU_48	Firmicutes	*Aerococcus*		0,6%
OTU_71		*Anaerococcus*		0,3%
OTU_56		*Clostridium*		0,4%
OTU_49		*Dialister*		0,6%
OTU_27		*Finegoldia*		0,2%
OTU_1		*Lactobacillus*	*iners*	23,9%
OTU_3		*Lactobacillus*	*crispatus*	25,7%
OTU_115		*Lactobacillus*	*reuteri*	0,2%
OTU_7		*Megasphaera*	*micronuciformis*	6,0%
OTU_8		*Gemella*	*sanguinis_morbbilliorum_heamolysans* *	0,1%
OTU_1680		*Parvimonas*		0,3%
OTU_53		*Peptoniphilus*		0,4%
OTU_26		*Peptostreptococcus*	*stomatis*	0,1%
OTU_4		*BVAB-1*		11,1%
OTU_16		*Staphylococcus*		0,1%
OTU_2		*Streptococcus*	*dentisani_tigurinus_oralis*	0,4%
OTU_34		*Veillonella*		0,6%
OTU_65		*WAL_1855D*		0,2%
OTU_31	Fusobacteria	*Fusobacterium*	*periodonticum**	0,3%
OTU_11		*Sneathia*		4,5%
OTU_77	Tenericutes	*Mycoplasma*		0,3%

* Species annotation with lower than 97% identity using BLASTn to search the eHOMD database

** Taxa merged at lowest annotation level possible

4.6 Association between oral and vaginal microbial dysbiosis

To our knowledge, no study has characterised the oral microbiota of women with and without clinical BV or investigated how different vaginal bacterial community types may relate to oral community types. In order to gain a greater understanding of the relationship between the oral and vaginal microbiota, we compared the oral microbiota of female adolescents with both different VCTs and clinical BV. The prevalence of clinical BV was higher in participants with OCT 1 (55%) compared to OCT 2 (43%) (Table 3.6). This was confirmed by a higher proportion of girls in OCT 1 the diverse BV-associated VCT 2 (60%) compared to 46% of those in OCT 2 (Table 3.6). Considering alpha diversity measured by Shannon's Index, NMDS analysis of the beta diversity of oral taxa (Figure 3.13 A-C) and linear regression of Shannon diversity, we did not find evidence that bacterial community diversity in the oral cavity is related to clinical BV status or VCT. Additionally, unsupervised hierarchical clustering of oral samples did not reveal any patterns in the microbial composition of the oral microbiota relating to clinical BV diagnosis or VCT (Figure 3.13 D).

Table 3.6: Distribution of clinical BV status and VCTs in OCTs

		OCT		
		OCT 1 (N = 85)	**OCT 2 (N= 80)**	p-value*
BV prevalence		55% (N = 47)	43% (N = 34)	0.1004
VCT distribution	VCT 1	22% (N = 19)	24% (N = 19)	0.8313
	VCT 2	60% (N = 51)	46% (N = 37)	0.0768
	VCT 3	18% (N = 16)	30% (N = 24)	0.0941

* z-score two-sample test for proportions

The results of negative binomial modelling are summarised in Table 3.7, listing taxa that were found to be differentially abundant (adjusted p<0.05) in the oral microbiota when comparing the cohort with respect to clinical BV status and VCT. *Bacteroides heparinolyticus*, a saccharolytic *Bacteroides* species commonly isolated from human periodontitis lesions (Bailey et al. 2019; Ashimoto et al. 1995), was found to be significantly enriched in the oral microbiota of participants with a clinical diagnosis of BV compared to those without (Table 3.7). A number of other PD associated bacteria including *Prevotella* spp. and *Oribacterium parvum* as well as *Butyrivibrio hungatei* were enriched in the oral microbiota of participants with BV compared to

participants with no clinically diagnosed vaginal microbial dysbiosis (Table 3.7). A reduction in *Bacteriodetes pyogenes* was observed in oral microbiota of BV positive participants compared to BV negative participants. When comparing the composition of the oral microbiota of this cohort with respect to VCT, there was a lower relative proportion of *SR1* in the oral microbiota of participants with *L. crispatus* dominated VCT 1 microbiotas compared to *L. iners* dominated VCT 2. A number of other bacterial taxa were found to be significantly different in this analysis but the fold-change was negligible (Table 3.7). Negative binomial modelling did not identify significant differences in the relative proportion of other red complex bacteria (*Tannerella* and *Treponema*) or PD associated bacteria (*Fusobacterium, Eikenella, Peptostreptococcus* and *Campylobacter*) in oral microbiota of participants with and without a clinical diagnosis of BV or different VCTs. Random forest modelling using all oral microbiota samples (N = 165) to predict clinical BV status (Sn = 0.63, Sp = 0.57) identified *Delftia* as being an important bacteria in differentiating between the oral microbiota of clinically BV negative and BV positive participants (Figure3.14 A) although negative binomial modelling did not find a significant difference in the oral presence of *Delftia* between participants with and without clinical BV (fold change = 0.32, p=1.00). Random forest models were additionally constructed to assess the ability of the composition of the oral microbiota to predict vaginal community between VCT 2 and VCT 1 (Sn = 1.00, Sp = 0.00), VCT 2 and VCT 3 (Sn = 1.00, Sp = 0.00) and VCT 1 and VCT 3 (Sn = 0.52, Sp =0.51) and the important bacteria distinguishing these models are summarised in Figure 3.14 A- 3.14 D.

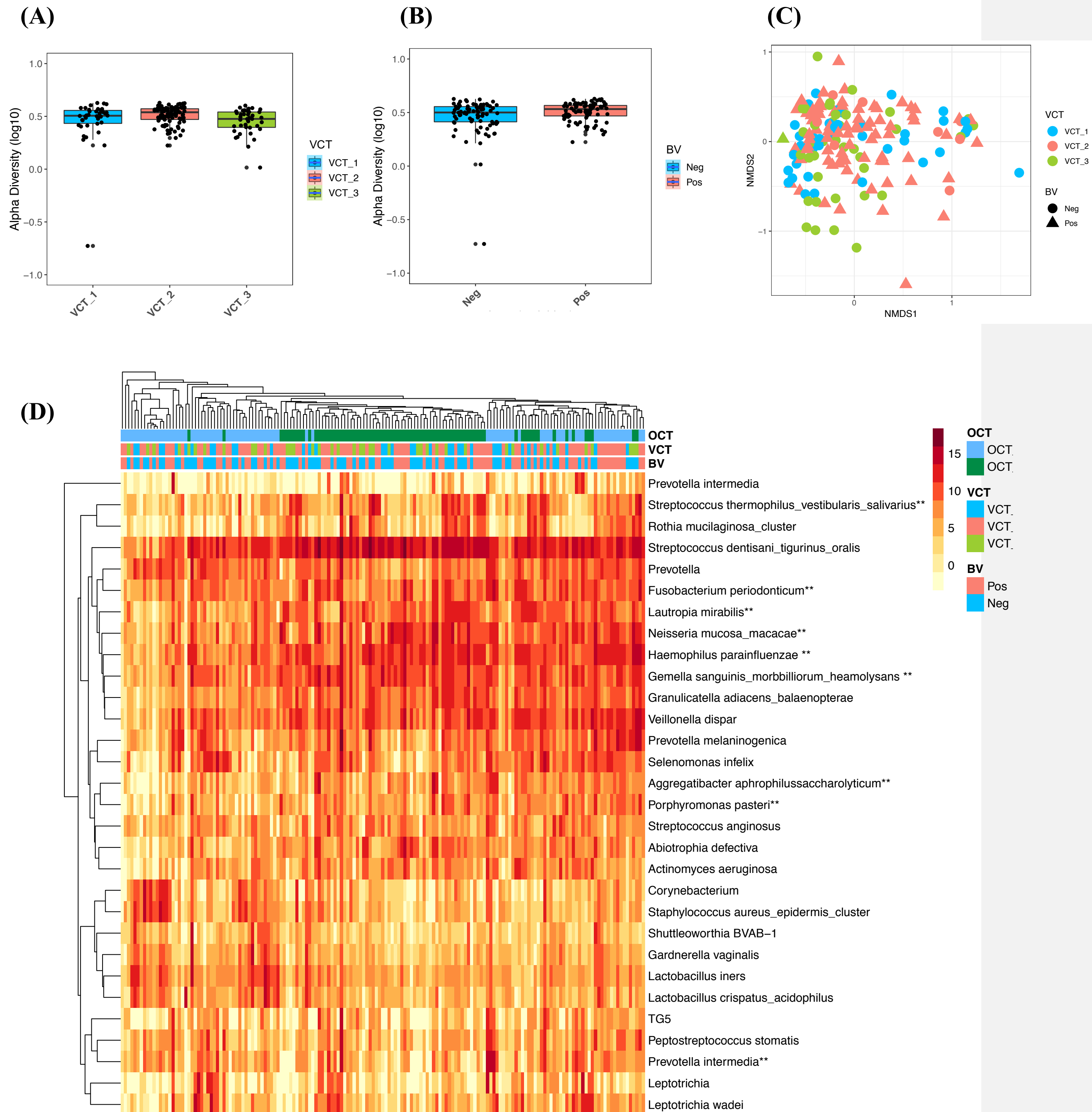


Figure 3.13: Ecological diversity of oral microbiota by clinical BV status and vaginal community type. (A) Shannon's diversity and **(B)** NMDS analysis of beta diversity using Bray-Curtis as a distance metric of oral microbial communities with colour denoting vaginal community type (VCT 1 = blue, VCT 2 = pink, VCT 3 = green) denoted by fuzzy clustering of vaginal bacterial composition using optimal k and shape denoting clinical BV diagnosis by Nugent scoring (round = negative, triangle = positive).

Table 3.7: Results of a negative binomial model detecting significant differences in the relative abundance of oral bacterial taxa between participants with and without and clinical BV diagnosis and differing vaginal community types

				Relative abundance fold change (adjusted p-value)		
				Clinical BV	Vaginal Community Type	
Phylum	Phylum	Genus	Species	Pos vs. Neg	VCT 2 vs. VCT 1	VCT 2 vs. VCT 3
OTU_151	Bacteroidetes	*Prevotella*	*shahii_loescheii*	2.06 (0.009)*		
OTU_221			*albensis*	1.29 (0.0326)		
OTU_307			*enoeca*	1.52 (0.0188)*		
OTU_405		*Alloprevotella*	*rava*	1.17 (0.0003)		
OTU_588		*Porphyromonas*	*circumdentaria*	0.97 (0.013)		
OTU_279			*gingivicanis_circumdentaria_pasteri*	1.09 (0.0037)		
OTU_424		*Bacteroides*	*zoogleoformans*	0.88 (<0.0001)	1.09 (<0.0001)	1.05 (<0.0001)
OTU_562			*pyogenes*	0.75 (0.0469)*		
OTU_241			*heparinolyticus*	3.29 (0.0014)*		
OTU_135	Firmicutes	*Butyrivibrio*	*hungatei*	3.07 (0.0159)*		
OTU_470		*Centipeda*	*periodontii*	0.97 (<0.0001)	1.07 (<0.0001)	1.07 (<0.0001)
OTU_492		*Lachnospiraceae*	*bacterium*	0.95 (<0.0001)		
OTU_529		*Oribacterium*	*parvum*	1.3 (0.0025)*		
OTU_602		*Peptostreptococcaceae*	*saphenum_sulci_infirmum*	0.9 (<0.0001)		
OTU_420	Proteobacteria	*Desulfovibrio*	*desulfuricans*	0.94 (<0.0001)		
OTU_1229	SR1				0.67 (<0.0001)*	
OTU_447	Synergistetes	*Pyramidobacter*	*piscolens*	0.83 (0.0009)		

* p-value > 0,05

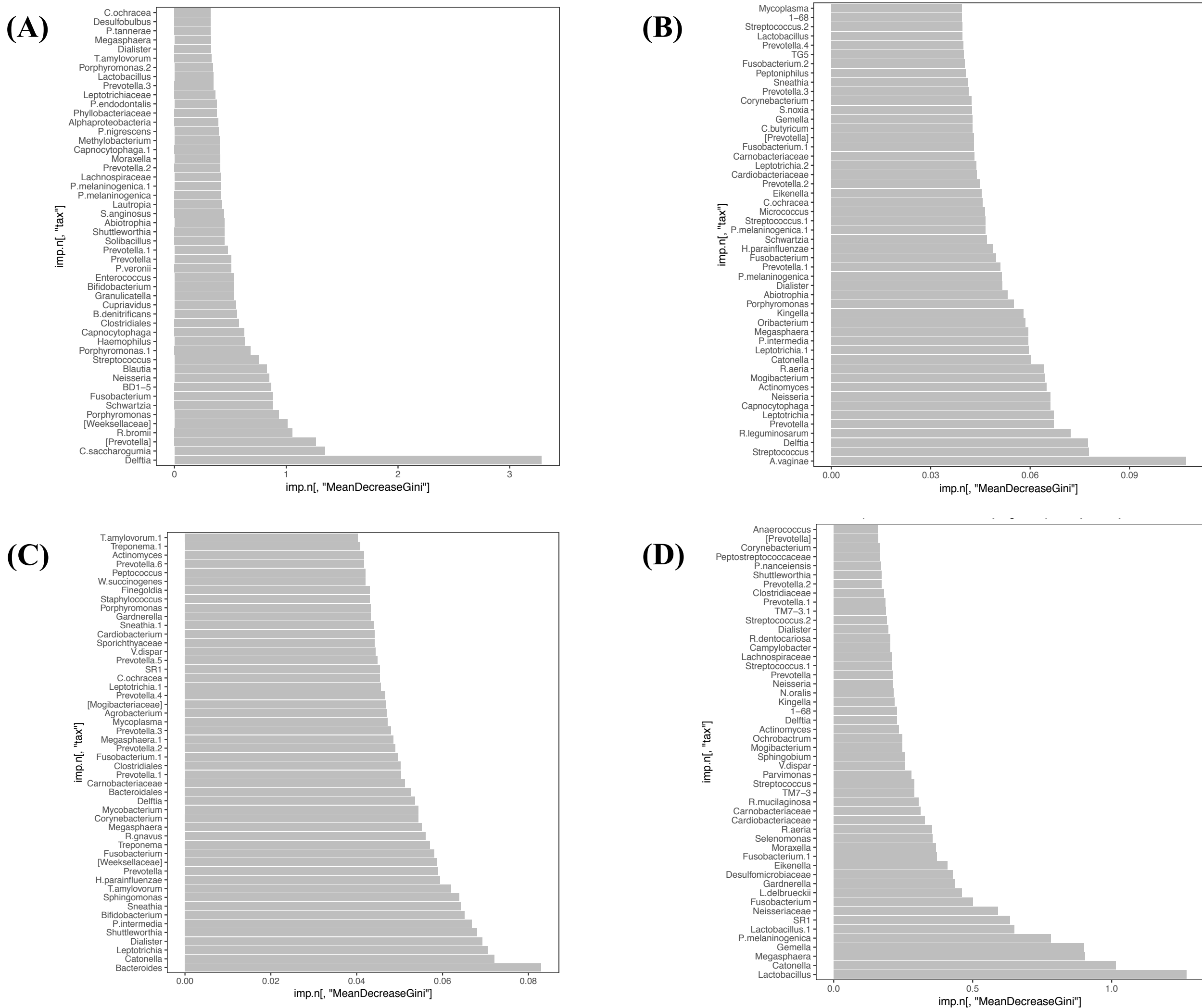


Figure 3.14: Top classification features in random forest modelling using oral bacterial composition to predict clinical BV status and vaginal community type. The oral samples were used to train a random forest model to predict the most important taxa to classify clinical BV diagnosis (N=258) **(A)** and vaginal community types (VCT) with VCT 2 vs. VCT 1 (N=60) **(B),** VCT 2 vs. VCT 3 (N=61) **(C)** and VCT 1 vs. VCT 3 (N=45) **(D)**. The mean decrease in the Gini coefficient for the top 30 most important taxa are displayed.

CHAPTER 5: Discussion

5.1 Study design and methodology

To our knowledge, this study presents the first data on the characterization of the South African oral microbiota. Another strength of our study is the size of the cohort included and the use of MiSeq next generation sequencing for characterization of the oral microbiota. Our cohort consisted of 94 South African women between the ages of 15 and 19 years enrolled in the UChoose-A-Star clinical trial at the Desmond Tutu HIV Youth Centre in Masiphumelele, Cape Town. The size of our cohort is relatively larger than the majority of previously published studies describing the oral microbiota in other populations (Aas et al. 2005; Dalwai et al. 2007; Zaura et al. 2009; Dewhirst et al. 2010; Bik et al. 2010; Xie et al. 2010; Belda-Ferre et al. 2011; Pushalkar et al. 2012; Schmidt et al. 2014; Fourie et al. 2016; Al-hebshi et al. 2017). We utilised high throughput Illumina Miseq technology which has been shown, in comparison to targeted PCR approaches and other sequencing platforms, to provide DNA sequencing data with high accuracy and low error rates per reads (Ross et al. 2013; Nakazato et al. 2013). Using this technology, we were able to characterise a larger proportion of the existing taxa in both the oral and vaginal microbiota compared to the aforementioned studies. As this was a retrospective study utilising samples from a previous study focused on reproductive health, we were unable to collect data on important baseline factors including dental hygiene, smoking habits and antibiotic usage that have been shown to impact the composition of the oral microbiota, a limitation that must be considered when interpreting these results. Most importantly, we could not assess for the presence of dental caries or gingivitis/ periodontitis and thus could not determine the specific factors that may be driving the separation we observed between different oral community types. We were able to asses a number of other baseline characteristics known to be important for vaginal colonization (summarised in Results section 2) and did not find any significant difference between our comparator groups of those with and without a clinical diagnosis of BV.

A common methodological challenge of 16S rRNA gene sequencing is achieving consistent and representative amplification of bacterial taxa present in biological samples. Differing DNA extraction methods and sequencing platforms have been shown to affect the representation

of specific bacterial taxa. Additionally, the amplification of different variable regions of the 16S bacterial ribosomal gene have been shown to exhibit higher levels of variability for specific bacterial phyla and to provide varying amounts of species level resolution for selected bacterial families (Graspeuntner *et al.* 2018). In order to control for bias introduced by batch effects of differing MiSeq runs, we included all matched samples (vaginal lateral wall (LW), saliva (SAL) and periodontal (PE)) from the same participants within the same MiSeq run where possible. Our final sequence library was pooled from six MiSeq runs using samples extracted from two different extraction kits. Hierarchical clustering and ecological diversity analysis revealed no evidence for technical bias relating to DNA extraction methods or MiSeq runs. Negative controls included for both DNA extraction and PCR amplification steps, gave evidence for minimal environmental contamination. The inclusion of an even mock community of known bacterial composition demonstrated that although bacterial taxa were amplified consistently across MiSeq runs, amplification was not evenly representative for different taxa. Of all 22 bacterial taxa known to be present in the mock community, *Propionibacterium acnes* was not identified in any of the runs and *Enterococcus, Neisseria cinerea, Pseudomonas aeruginosa* and *Escherichia coli* were consistently overrepresented across all runs. The majority of studies describing the oral microbiome have amplified the V1 and V2 hypervariable regions of the of the bacterial ribosomal gene, likely due to these regions demonstrating high level resolution for keystone oral bacteria including *Streptococcus* and *Staphylococcus* species (Chakravorty et al. 2008). For the purpose of this study we amplified the V4 region to facilitate a comparison of sequences from the oral and vaginal samples for which our group had already developed a V4 sequencing protocol. While the V4-6 region is conventionally used as representative barcode for the majority of bacterial phyla (Yang et al. 2016), some studies have found difficulty in identifying *Fusobacterium* species in this region (Kumar et al. 2011). Results from our amplification of the mock communities demonstrated that we were able to differentiate between the two *Streptococcus* species present (*S. mutans* and *S. agalactiae*) but were unable to make a distinction between *Staphylococcus aureus* and *S. epidermis*, a limitation that must be considered when comparing the results of this study to other descriptions of the oral microbiota.

Variability in sequencing depth is an additional difficulty encountered when analysing MiSeq sequencing data as the number of reads generated per sample can vary by orders of

magnitude within a single sampling run. For the purpose of this study we combined several MiSeq runs including both high DNA yield vaginal and low DNA-yield oral samples in our comparison. While historically studies have adjusted for high variability in read counts by sub-sampling using read count rarefaction, this method results in a loss of valuable data and is inappropriate for determining differential abundance in sequencing libraries with high read count variability (Mcmurdie & Holmes 2014). To account for this variability in differential abundance testing we standardised our OTU tables by transforming each OTU count to relative abundance multiplied by median sample read and used a negative binomial model (utilised in the DeSeq package (Anders & Huber 2010)) to estimate differences in the relative abundance of bacterial taxa between comparison groups.

5.2 Development of a sample processing protocol for stored samples with low DNA yields

As detailed in Chapter 1, the study utilised a bacterial DNA extraction and 16S rRNA gene amplification protocol that had been previously developed for the preparation of bacterial 16S V4 gene sequencing libraries using Illumina Miseq. Our initial attempts to extract DNA from stored saliva and periodontal samples did not result in sufficient DNA yields for downstream PCR amplification. This likely due to the overall low abundance of bacterial product present in these samples, in addition to the presence of salivary gland secretions thought to contain nucleases and proteins that may degrade DNA during extended sample storage and inhibit DNA amplification using PCR. After adding additional steps to our sample processing protocol, including boiling of samples (prior or post storage at -80°C) and removal of salivary supernatant to denature and remove potential inhibitors that may be interfering with DNA extractions and PCR reactions, for most samples collected we were able to extract sufficient DNA yields for PCR and downstream MiSeq DNA library preparation. The development of this protocol allowed for the extraction of 16S rRNA gene DNA from stored low-yield samples.

5.3 Characterisation of the oral microbiota

This study presents a description of the oral microbiota composition of 94 South African women aged 15-19, originating from salivary and periodontal samples. After filtering for OTUs that had at least 10 read counts in at least 10% samples, we identified 866 unique OTUs mapping to 213 genera and 193 identifiable species. This is higher than previous description

of oral bacterial diversity in other large scale population studies (Takeshita et al. 2016; Sarkar, Stoneking, and Nandineni 2017; Li et al. 2014a).

Despite the number of physical and chemical exposures the oral cavity experiences on a daily basis, research indicates that a stable core oral microbiota exists (Li et al. 2013; Zaura et al. 2009). In line with what has been previously published (Dewhirst et al. 2010; Bik et al. 2010; Zaura et al. 2009) we found the core oral microbiota to be to be dominated by Firmicutes and Proteobacteria followed by Actinobacteria, Fusobacteria and to a small extent Spirochaetes. As expected, *Streptococcus* was the most prevalent bacterial genus, followed by *Prevotella*, *Haemophilus*, *Lactobacillus*, *Neisseria*, *Veillonella, Leptotrichia, Lautropia, Selenomonas, Fusobacterium, Staphylococcus and Porphyromonas.* A number of other bacterial genera species previously identified in the oral microbiota were identified at low levels including *Actinomyces, Corynebacterium, Campylobacter, Aggregatibacter, Abiotrophia, Gardnerella, Rothia, Megasphaera, Treponema, Shuttleworthia, Peptostreptococcus* and *Sneathia*.

The structure of the oral cavity leads to the formation of different ecological niches, causing colonization by distinct bacterial community types. Mucosal surfaces including the palate, tongue, cheeks and tonsils house a variety of shedding bacteria that accumulate in saliva, whilst the enamelled covered surface of teeth or dentures favour the long-term formation of bacterial biofilms (Costalonga & Herzberg 2014). Of the 775 OTUs present in all oral samples, 560 were ubiquitous across saliva and periodontal samples while 101 and 114 were unique to saliva and periodontal, respectively. In this study, we did not identify differences between the salivary and periodontal microbiota in terms of ecological diversity or the composition of the most abundant bacterial taxa. Differential abundance testing of less prevalent bacterial species revealed higher proportions of Firmicutes (*Oribacterium* spp. and *Streptococcus anginosus*), *Prevotella nanceiensis* and Actinobacteria (*Actinomyces graevenitzii* and *Rothia mucilaginosa*) in the saliva compared to the periodontal space. In contrast, periodontal samples were enriched in a single Fusobacteria OTU mapping to the *Leptotrichia* genus. This is in agreement with previous studies reporting the oral mucosal surfaces to be largely dominated by Firmicutes and the dental plaque to host a more diverse range of Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria and Fusobacteria (He et al. 2015; Aas et al.

2005). In a study comparing nine ecological niches in the oral environment, samples from the teeth, including both the sub- and supra-gingival plaque, had the lowest proportion of Firmicutes and relatively larger proportion of Actinobacteria (Segata et al. 2012).

The extent to which host genetics determine the composition of the oral microbiota is not well understood. While one study found that the salivary microbiota did not vary across individuals from twelve locations worldwide (Nasidze et al. 2009), another study comparing saliva from Alaskan, German and African (originating from Democratic republic of Congo (DRC), Sierra Leone and Uganda) individuals found significantly higher bacterial diversity in the oral microbiotas of Africans compared to those from Northern countries (Li et al. 2014b). This study, which utilised the Genome Sequencer FLX platform, identified 62 genera in the Alaskan cohort, 58 genera in the German cohort and 100 genera in the African cohort. In a study of 2,343 healthy individuals, *Neisseria* and *Haemophilus* were found to dominate the salivary microbiota of Korean individuals, while *Prevotella* and *Veillonella* were most prevalent amongst Japanese individuals (Takeshita et al. 2016). In our study, a single OTU comprised 27% of the oral microbiota mapping to a cluster of *Streptococcus* species including S. *dentisani, S. oralis* and *S. tigurinus.* All three *Streptococcus* species have been identified in the oral microbiome. While *S. tigurinus* has been associated with infective endocarditis (Zbinden et al. 2012; Zbinden et al. 2015; Zbinden et al. 2014), S. *dentisani* and *S. oralis* are associated with oral health and are both candidates for landmark oral species (López-López et al. 2017). *S. oralis* is the most commonly identified in the oral microbiome, reported as one of the early colonisers of the oral microbiome and involved in maintaining oral biofilm integrity allowing for further bacterial colonisation of species associated with good oral health (including *S. parasanguinis*, *A. defectiva*, *S. mitis*, and *S. sanguinis*) (Peterson et al. 2013; Corby et al. 2005; Paddick et al. 2003). *S. dentisani* has been reported to confer dual probiotic action in the oral bacterial ecosystem by inhibiting the growth of major oral pathogens through the production of bacteriocins and buffering acidic oral pH (the main cause of dental carries) through the production of ammonia in the presence of arginine (López-López et al. 2017).

A study in 161 Italian participants identified 3 "salivary types' characterised by increased relative abundance of *Prevotella*, *Streptococcus/Gemella* and *Fusobacterium/Neisseria* with the *Prevotella* and *Streptococcus/Gemella* salivary composition associated with poor oral

health (Filippis et al. 2014). In another cohort of 200 healthy Japanese individuals, three salivary types were defined as being *Prevotella/Veillonella*-dominant, *Streptococcus*-dominant and *Neisseria*, *Haemophilus*, or *Aggregatibacter/Porphyromonas*-dominant with the first two salivary types associated with a higher percentage of periodontal pockets. We identified two distinct oral community types (OCTs) that did not correlate with sampling site. Oral community type one (OCT 1) was characterised by a diverse make up of oral bacteria while OCT 2 was enriched in the *Streptococcus dentisani_tigurinus_oralis* cluster previously described as the most abundant OTU across all oral samples. A limitation of this study was our inability to establish oral health (defined by periodontitis diagnosis) and oral health practices at the time sample collection. Due to this we are unable to determine whether a *Streptococcus* dominated oral microbiota in the South African population is associated with poor oral health as previously described or if the dominant *Streptococcus* species is indeed *S. dentisani* or *S. oralis* conferring protection against invasive oral bacterial species. Targeted qPCR would assist in resolving the species level classification and should be a consideration for future work.

While one study has attributed differences in the oral microbiota of African hunter-gatherers from Uganda compared to agricultural groups in DRC and Sierra Leone to differences in lifestyle and diet (Nasidze et al. 2009), another study found no differences in composition of oral microbiota of omnivores, ovo-lacto-vegetarians and vegan individuals (Filippis et al. 2014). Altered oral microbiota have been identified as a marker of systemic disease (Long et al. 2017; Roda et al. 2008; Michaud & Izard 2014; Kistler et al. 2018; Lerner et al. 2016; Hayashi et al. 2010). Distinct oral microbiota have additionally been observed for non-disease states, for example in pregnant and lactating women (Zarco et al. 2012), indicating that homeostatic alterations in the body may manifest in the oral microbiota. There it is important to understand what constitutes a 'healthy' or 'optimal' oral bacterial make up in order to recognize deviations from the state predictive of systemic disease. More research is needed into the association between oral and systemic health and the composition of the oral microbiota in the South African population.

5.4 Characterisation of the vaginal microbial dysbiosis in South African female adolescents

The current understanding of the composition of common vaginal community types (VCTs) is based on studies in North-American women where five or six vaginal community types have been described – four of which are dominated by a single lactobacilli species (*L. crispatus*, *L. iners*, *L. gasseri* or *L. jensenii*) and one to two characterised by a diverse vaginal microbiota comprised of strictly anaerobic BV associated bacteria ((Ravel et al. 2011; Kroon et al. 2018). North American studies describing the vaginal microbiome of asymptomatic women have shown Lactobacilli dominant VCTs to be more common in White and Asian women in while asymptomatic Hispanic and African-American women sometimes present with high bacterial diversity and low numbers of *Lactobacillus*(Fettweis et al. 2014; Zhou et al. 2011; Ravel et al. 2011).

This has led to a need for a change in the definition of what was previously considered a 'healthy' or 'optimal' vaginal microbiota – that being one dominated by *Lactobacillus* bacterial commensal species, having a low vaginal pH (3.5-4.5), and having no *Candida,* STI or BV present. These studies originate from populations in which overall prevalence of BV is relatively low compared to the South African population in which the vaginal microbiota of women of reproductive age is characterised by low lactobacilli abundance, high bacterial diversity, a high prevalence of BV and a higher vaginal pH in comparison to their North American counterparts (Lennard et al. 2018; Anahtar et al. 2015). In this study, participants with clinically diagnosed BV (Nugent score >7) presented with a vaginal microbiota comprised of several BV associated bacteria in comparison to their BV-negative counterparts with lactobacilli dominant vaginal microbiotas. We additionally identified three distinct VCTs by fuzzy clustering. VCT 2, characterised by a diverse composition of BV-associated bacteria, was highly correlated with a clinical diagnosis of BV and in accordance with previous studies in the same population of young South African women (Lennard et al. 2017), participants without diagnosed BV clustered into two distinct VCTs dominated by either *L. crispatus* (VCT 1) or *L. iners* (VCT 3). These results were similar to a study in a different South African population where participants clustered in to four VCTs, two dominated by lactobacilli (*L. crispatus* and *L. iners*) and two BVAB VCTs differentiated by a dominance of *Gardnerella* and a consistent presence of *Prevotella* (Anahtar et al. 2015). Our study, in addition to the previously described studies in the South African populations, did not observe the separation into four distinct lactobacilli dominated VCTs as reported by Ravel et al. 2012. As in previous studies in the

South African population, *L. gasseri* and *L. jensenii* are only identified at low proportions in the South African vaginal microbiome, likely due to a combination of population specific variable including host genetics and lifestyle factors.

Our results, in addition to the previously described studies, challenge the paradigm of associating a *Lactobacillus* dominant vaginal microbiome with functional characteristics of a 'heathy' vaginal microbiome and high bacterial diversity and presence of BV-associated bacteria as the marker of an "unhealthy" vaginal microbiota characterised by high pH and inflammation on a clinical and sub-clinical level (Ma et al. 2012). Particularly the role of *L. iners* as a protective vaginal bacteria, with this *Lactobacillus* species identified in both optimal and dysbiotic VCTs (Petrova et al. 2017; Borgdorff et al. 2016). Studies have reported an *L. crispatus* dominated VCT to be protective against BV, while *L. iners* dominance has been associated with increased vaginal pH (Ravel et al. 2011) and marker of a transition to BV characterised by the expansion of BV-associated bacteria (Van De Wijgert et al. 2014; Verstraelen et al. 2009). The underlying reason for the high prevalence of BV in African-American women in comparison to their Caucasian counterparts is currently unknown (Peipert et al. 2008). Studies characterising non-dysbiotic vaginal microbiomes have reported a higher proportion of *L. iners* dominated VCTs in African-American women compared Caucasian women who were dominated by *L. crispatus* (Srinivasan et al. 2012). In addition to this, some asymptomatic African-American and black South African women present with diverse, non-lactobacilli dominant VCTs (Ravel et al. 2011; Van De Wijgert et al. 2014). These studies indicate that women of African descent are less likely to harbour *L. crispatus* vaginal bacteria and more likely to be colonized by *L. iners*. Based on the potential contribution of *L. iners* to BV pathology, this may be a driver for the high BV prevalence observed in women of African descent although the low rates of BV in some African countries leaves more to be determined (Anahtar et al. 2015; Kenyon et al. 2013). Further research is needed into the extent to which host genetics play a role in determining vaginal microbial composition, whether there are functional redundancies by which non-lactobacillus bacterial communities may facilitate lactic acid fermentation to maintain a protective vaginal microbiota, and to better define the determinants of a healthy vaginal microenvironment.

5.5 The relationship between oral and vaginal microbial dysbiosis

Both PD and BV, conditions characterised by microbial dysbiosis, are associated with a two- to four-fold risk of PTB (Pretorius et al. 2007; Harper et al. 2012). Although ecological evidence of the co-occurrence of BV, PD and PTB supports an association between bacterial dysbiosis of the oral and vaginal microbiotas, few etiological studies have investigated a possible biological relationship between the oral and vaginal microbiotas. Furthermore, the difficulty in diagnosing and treating microbial dysbiosis has yielded mixed results of clinical trials of PD and/or BV treatment to reduce adverse birth outcomes (Srinivasan et al. 2009). One study identified increased proportions PD-associated oral bacteria including *P. gingivalis, T. forsythia, P. intermedia,* and *P. nigrescens* in oral cavity of women who had delivered preterm compared to those who did not (Lin et al. 2007). To our knowledge, no study has characterised the oral microbiota of women with and without clinical BV or investigated how different bacterial VCTs may relate to OCTs. In order to gain a greater understanding of the relationship between the oral and vaginal microbiota, we compared the oral microbiota of female adolescents with both different VCTs and with or without clinical BV. In this study, *Bacteroides heparinolyticus*, a saccharolytic *Bacteroides* species commonly isolated from human periodontitis lesions (Bailey et al. 2019; Ashimoto et al. 1995), was found to be significantly enriched in the oral microbiota of participants with a clinical diagnosis of BV compared to those without. A number of other PD associated bacteria (*Prevotella* spp., *Oribacterium parvum* and *Butyrivibrio hungatei*) were enriched in the oral microbiota of participants with BV compared to participants with no clinically diagnosed vaginal microbial dysbiosis. When considering VCTs of participants without diagnosed BV, those with an *L. iners* dominated VCT 3 had significantly increased proportions of oral SR1 (Curtis et al. 2011; Campbell et al. 2013), a bacterial phylum elevated in PD-states, in comparison to those with an *L. crispatus* dominated vaginal microbiota. This data supports the hypothesis that vaginal microbial dysbiosis in linked to the expansion of sub-optimal lactobacillus species in the oral cavity, especially when considering the questionable role of *L. iners* as a dominant vaginal bacteria. The functional significance of this relationship is yet to be determined. The most commonly accepted hypothesis for the causal relationship between PD and PTB describes the hematogenous dissemination of inflammatory products released by red and orange complex bacteria in the oral cavity, resulting in a systemic inflammatory response that initiates uterine contractions and PTB (Pretorius et al. 2007). This has been supported by a number of studies

identifying an association between PD and increased markers of inflammation in the oral compartment (Konopka et al. 2003; Offenbacher et al. 1998) and systemically (Amar et al. 2003; D'Aiuto et al. 2004). Few studies have investigated the association between PD, systemic inflammation and adverse birth outcomes but one study found that pregnant women with both PD and increased systemic CRP were at a higher risk of developing preeclampsia compared to women without PD and both low and high levels of systemic CRP (Ruma et al. 2008). Based on identification of known oral pathogens such a Fusobacterium in the amniotic fluid of women who delivered pre-term and the absence thereof in term births (Hill 1998), it has been suggested that oral pathogens themselves may migrate to the uterine tract, causing decidual and chorioamniotic infections (Pretorius et al. 2007), although the exact mechanism through which this happens has yet to be elucidated. If in fact there is an association between oral and vaginal bacterial dysbiosis, the direction of the relationship with regards to which compartment the dysbiosis originates (as summarised in Chapter Two, Figure 2.5) is still unclear.

While these results strengthen the argument for a link between bacterial dysbiosis in the genital and oral tract, there are still a number of factors in the etiology of PD and BV that need to be elucidated. Studies identifying hereditary risk for PTB preterm (Flint Porter et al. 1997; Winkvist et al. 1998) and alleles of inflammation-related genes associated with a predisposition towards PD and/or BV has led to the hypothesis that the association between PD, BV and PTB may be modulated by host genetics that may influence susceptibility to colonization by anaerobic bacteria. Alternatively, maternal and/or foetal inflammatory immune reaction in response to microbial dysbiosis may be genetically determined. A better understanding of how differences in host genetics determine the composition of both VCTs and OCTs and how this in turn impacts inflammation on a localized and systemic level, is required to disentangle the relationship between BV, PD and PTB.

5.6 Conclusions, relevance and further work

To our knowledge, this study presents the first data on the characterization of the South African oral microbiota. We present findings in support of previous studies from other non-African populations that the oral microbiota is one of the most diverse in human body. We

additionally reported a description the vaginal microbiome that is in agreement with previous studies in the South African population and in contrast with data originating in North American populations. Due to the retrospective nature of this study we were unable to establish key baseline characteristics that may impact oral health. Additionally we were unable to distinguish between key bacterial *Streptococcus* species comprising majority of the oral microbiota. These limitations impacted our ability to determine the functional and clinical relevance of the distinct oral community types we observed. In future, more advanced molecular typing to achieve a greater level of taxonomic resolution will greatly improve our understanding of what constitutes the optimal composition of the South African oral microbiome. Hormonal fluctuations, both endogenous and exogenous, have been shown to influence the composition of the oral and vaginal microbiome (Gajer et al. 2012, Prasana et al. 2018). Additionally, both hormonal shifts during pregnancy and exposure to oral contraceptives have been linked to adverse periodontal health (Figuero et al. 2013, Prasana et al. 2018, Prachi et al. 2019). Due to the cross sectional nature of this study, we were unable to investigate temporal shifts in the oral and vaginal microbiome relating to endogenous hormonal fluctuations and exposure to exogenous hormonal contraceptives. A longitudinal analysis of these compartments considering exposure to exogenous hormonal contraceptive is critical to better understand causality in the in the relationship between PD, BV and PTB and has been identified as potential for future work.

This study additionally investigated the association between bacterial dysbiosis of the oral and vaginal tracts by comparing the oral bacterial composition of women with and without vaginal bacterial dysbiosis (measured by clinical BV criteria and vaginal community typing). We did not identify an association between vaginal microbial dysbiosis and an increased presence of “red-complex” of bacteria traditionally associated with PD, we report increased proportions of additional PD-associated bacterial species in the oral microbiome of women with clinically diagnosed BV compared to asymptomatic women. Additionally we observed the oral microbiome of women with *L. iners* dominant VCTs to be enriched with PD-associated bacterial species, an interesting finding considering recent reports of the expansion of *L. iners* populations as a marker of transitioning from an optimal to dysbiotic vaginal microbiota. While this data provides evidence in support of a relationship between oral and vaginal dysbiosis, it unclear in which compartment bacterial dysbiosis would originate, should the

association holds true. Assessing the relationship between PD and BV could lead to potential screening and intervention programs to effectively identify and treat risk factors during pregnancy and reduce the amount of preventable maternal and infant deaths in Sub Saharan Africa.

Reference List

Aaard, K. et al., 2014. The Placenta Harbors a Unique Microbiome. *Sci Transl Med*, 6.
Aas, J.A. et al., 2005. Defining the Normal Bacterial Flora of the Oral Cavity. *Journal of Clinical Microbiology*, 43(11)
Al-hebshi, N.N. et al., 2017. Inflammatory bacteriome featuring Fusobacterium nucleatum and Pseudomonas
Aagaard, K. et al., 2014. The Placenta Harbors a Unique Microbiome. *Sci Transl Med*, 6.
Aas, J.A. et al., 2005. Defining the Normal Bacterial Flora of the Oral Cavity. *Journal of Clinical Microbiology*, 43(11), pp.5721–5732.
Al-hebshi, N.N. et al., 2017. Inflammatory bacteriome featuring Fusobacterium nucleatum and Pseudomonas aeruginosa identified in association with oral squamous cell carcinoma. *Scientific Reports*, (May), pp.1–10. Available at: http://dx.doi.org/10.1038/s41598-017-02079-3.
Alakomi, H.L. et al., 2000. Lactic Acid Permeabilizes Gram-Negative Bacteria by Disrupting the Outer Membrane. *Appl*, 66(5), pp.2001–2005.
Amann, R.I., Ludwig, W. & Schleifer, K.-H., 1995. Phylogenetic Identification and In Situ Detection of Individual Microbial Cells without Cultivation. *Microbiological reviews*, 59(1), pp.143–169.
Amar, S. et al., 2003. Periodontal Disease Is Associated With Brachial Artery Endothelial Dysfunction and Systemic Inflammation. , pp.1–6.
Amsel, R. et al., 1983. Nonspecific vaginitis: diagnostic criteria and microbial and epidemiologic associations. *The American journal of medicine*, 74(1), pp.14–22.
Anahtar, M.N. et al., 2015. Cervicovaginal Bacteria Are a Major Modulator of Host Inflammatory Responses in the Female Genital Tract. *Immunity*, 42(5), pp.965–976. Available at: http://dx.doi.org/10.1016/j.immuni.2015.04.019.
Anders, S. & Huber, W., 2010. Differential expression analysis for sequence count data. *Genome Biology*, 11.
Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data.
Ashimoto, A., Slots, M.J. & J., F., 1995. Molecular genetic detection of Bacteroides heparinolyticus in adult periodontitis. *Oral microbiology and immunology*, 10(5), pp.284–287.
Bailey, G.D. et al., 2019. Bacteroides heparinolyticus : Deoxyribonucleic Acid Relatedness of Strains from the Oral Cavity and Oral-Associated Disease Conditions of Horses , Cats , and Humans. , pp.42–44.
Balle, C. et al., 2018. Endocervical and vaginal microbiota in South African adolescents with asymptomatic Chlamydia trachomatis infection. *Scientific Reports*.
Belda-Ferre, P. et al., 2011. The oral metagenome in health and disease. *ISME Journal*, (June).
Bennett, W.A. et al., 2000. Intrauterine endotoxin infusion in rat pregnancy induces preterm delivery and increases placental prostaglandin F2α metabolite levels. *American Journal of Obstetrics and Gynecology*, 182(6), pp.1496–1501.
Bik, E.M. et al., 2010. Bacterial diversity in the oral cavity of ten healthy individuals Elisabeth. *ISME Journal*, 4(8), pp.962–974.
Black, V. et al., 2008. The detection of urethritis pathogens among patients with the male urethritis syndrome, genital ulcer syndrome and HIV voluntary counselling and testing

clients: should South Africa's syndromic management approach be revised? *Sexually transmitted infections*, 84(4), pp.254–258.
Boggess, K.A. et al., 2005. Fetal immune response to oral pathogens and risk of preterm birth. *American journal of obstetrics and gynecology*, 193, pp.1121–6.
Borgdorff, H. et al., 2016. Unique Insights in the Cervicovaginal Lactobacillus iners and L. crispatus Proteomes and Their Associations with Microbiota Dysbiosis. *PLoS ONE*, 11(3).
Boris, S. et al., 1998. Adherence of Human Vaginal Lactobacilli to Vaginal Epithelial Cells and Interaction with Uropathogens. *Infection and Immunity*, 66(5), pp.1985–1989.
Boskey, E.R. et al., 2014. Origins of vaginal acidity: high D/L lactate ratio is consistent with bacteria being the primary source. *Human Reproduction Volume*, 16(9), pp.1809–1813.
Bradshaw, C.S. et al., 2006. High Recurrence Rates of Bacterial Vaginosis over the Course of 12 Months after Oral Metronidazole Therapy and Factors Associated with Recurrence. *JID*, 193.
Briselden, A.N.N.M. et al., 1992. Sialidases (Neuraminidases) in Bacterial Vaginosis and Bacterial Vaginosis-Associated Microflora. *Journal of clinical microbiology*, 30(3), pp.663–666.
Brocklehurst, P. et al., 2013. Antibiotics for treating bacterial vaginosis in pregnancy. *Cochrane Database of Systematic Reviews*, (1).
Campbell, J.H. et al., 2013. UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *PNAS*, pp.1–6.
Caporaso, J.G. et al., 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7(5).
Cassini, M. a. et al., 2013. Periodontal bacteria in the genital tract: Are they related to adverse pregnancy outcome? *International Journal of Immunopathology and Pharmacology*, 26(4), pp.931–939.
Cauci, S. et al., 2005. Combination of vaginal pH with vaginal sialidase and prolidase activities for prediction of low birth weight and preterm birth.
Celik, H. & Ayar, A., 2002. Effects of erythromycin on pregnancy duration and birth weight in lipopolysaccharide-induced preterm labor in pregnant rats. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 103, pp.22–25.
Centre for Disease Control, 2015. *Sexually Transmitted Diseases Treatment Guidelines , 2015*,
Chakravorty, S. et al., 2008. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiology Methods*, 69(2), pp.330–339.
Cherpes, T.L. et al., 2008. 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. *Sexually Transmistted Diseases*, 35(1), pp.78–83.
Cho, I. & Blaser, M.J., 2012. The human microbiome at the interface of health and disease. *Nature Genetics*, 13.
Chu, D.M. et al., 2017. Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nature Medicine*, 23(3), p.314.
Claassen, S. et al., 2013. A comparison of the efficiency of five different commercial DNA extraction kits for extraction of DNA from faecal samples. *Journal of Microbiological Methods*, 94(2), pp.103–110. Available at:

https://doi.org/10.1016/j.mimet.2013.05.008.
Corby, P.M. et al., 2005. Microbial Risk Indicators of Early Childhood Caries. *Journal of Clinical Microbiology*, 43(11), pp.5753–5759.
Costalonga, M. & Herzberg, M.C., 2014. The oral microbiome and the immunobiology of periodontal disease and caries. *Immunology Letters*, 162, pp.22–38.
Costalonga, M. & Herzberg, M.C., 2014. The oral microbiome and the immunobiology of periodontal disease and caries. *Immunology Letters*, 162(2), pp.22–38. Available at: http://dx.doi.org/10.1016/j.imlet.2014.08.017.
Costello, E.K. et al., 2012. The Application of Ecological Theory. *Science*, 336, pp.1255–1263.
Cowan, F. & Pettifor, A., 2009. HIV in adolescents in sub-Saharan Africa. *Current Opinion in HIV and AIDS*, 4(4), pp.288–293. Available at: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=01222929-200907000-00011.
Curtis, J.T. & Roger Bray, J., 1957. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, 27(4), pp.325–349.
Curtis, M.A., Zenobia, C. & Darveau, R.P., 2011. The relationship of the oral micrpbiota to periodontal health and disease. *Cell Host and Microbe*, 10(4), pp.302–306.
D'Aiuto, F. et al., 2004. Periodontitis and Systemic Inflammation_ Control of the Local Infection is Associated with a Reduction in Serum Inflammatory Markers. *Journal of Dental Research*.
Dalwai, F., Spratt, D.A. & Pratten, J., 2007. Use of Quantitative PCR and Culture Methods To Characterize Ecological Flux in Bacterial Biofilms. *Journal of Clinical Microbiology*, 45(9), pp.3072–3076.
Dasanayake, A.P. et al., 2003. Preterm low birth weight and periodontal disease among African Americans. *Dental Clinics of North America*, 47, pp.115–125.
Dean, S. V et al., 2013. Born Too Soon: Care before and between pregnancy to prevent preterm births : from evidence to action. *Reproductive Health*, 10(Suppl 1), pp.1–16.
Desantis, T.Z. et al., 2006. Greengenes , a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Applied and Environmental Microbiology*, 72(7), pp.5069–5072.
Dewhirst, F.E. et al., 2010. The Human Oral Microbiome. *Journal of Bacteriology*, 192(19), pp.5002–5017.
Dominguez-bello, M.G. et al., 2010. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *PNAS*, 107(26), pp.11971–11975.
Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19), pp.2460–2461.
Edgar, R.C., 2018. Search and clustering orders of magnitude faster than BLAST. , 26(19), pp.2460–2461.
Faith, D.P., Minchin, P.R. & Belbin, L., 1987. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio*, 69, pp.57–68.
Fethers, K.A. et al., 2008. Sexual Risk Factors and Bacterial Vaginosis : A Systematic Review and Meta-Analysis. *Clinical Infectious Diseases*, 47.
Fettweis, J.M. et al., 2014. Differences in vaginal microbiome in African American women versus women of European ancestry. *Microbiology*, 160, pp.2272–2282.
Figuero, E. et al., 2013. Effect of pregnancy on gingival inflammation in systemically healthy women: a systematic review. *Journal of Clinical Periodontology*, 40, pp.457–473.

Filippis, F. De et al., 2014. The Same Microbiota and a Potentially Discriminant Metabolome in the Saliva of Omnivore , Ovo-Lacto- Vegetarian and Vegan Individuals. , 9(11).
Flint Porter, T. et al., 1997. The Risk of Preterm Birth Across Generations. *Obstetrics and gynecology*, 90(1).
Forner, L. et al., 2006. Incidence of bacteremia after chewing , tooth brushing and scaling in individuals with periodontal inflammation. , pp.401–407.
Fouhy, F. et al., 2016. 16S rRNA gene sequencing of mock microbial populations- impact of DNA extraction method , primer choice and sequencing platform. *BMC Microbiology*, 16(123), pp.1–13. Available at: http://dx.doi.org/10.1186/s12866-016-0738-z.
Fourie, N.H. et al., 2016. The microbiome of the oral mucosa in irritable bowel syndrome. *Gut Microbes*, 7(4), pp.286–301. Available at: http://dx.doi.org/10.1080/19490976.2016.1162363.
Fredricks, D.N., Fiedler, T.L. & Marrazzo, J.M., 2005. Molecular Identification of Bacteria Associated with Bacterial Vaginosis. *New England Journal of Medicine*, 353, pp.1899–1911.
Freire, C. De et al., 2010. Lysostaphin: A Staphylococcal Bacteriolysin with Potential Clinical Applications. *Pharmaceuticals*, 3, pp.1139–1161.
Gajer, P. et al., 2012. Temporal Dynamics of the Human Vaginal Microbiota. *Sci Transl Med*, 4(132).
Gaujoux, R., 2014. Generating heatmaps for Non negative Matrix Factorization. *R Foundation for Statistical Computing, Vienna, Austria*.
Gill, C. et al., 2016. Evaluation of Lysis Methods for the Extraction of Bacterial DNA for Analysis of the Vaginal Microbiota. *PLoS ONE*, 11(9), pp.1–16.
Gower, A.J.C., 1966. Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis. *Biometrika*, 53(3), pp.325–338.
Graspeuntner, S. et al., 2018. Selection of validated hypervariable regions is crucial in 16S-based microbiota studies of the female genital tract. *Scientific Reports*, 8(9678), pp.4–10.
Griffen, A.L. et al., 2011. CORE : A Phylogenetically-Curated 16S rDNA Database of the Core Oral Microbiome. *PLoS One*, 6(4), pp.1–10.
Guise, J. et al., 2001. Screening for Bacterial Vaginosis in Pregnancy. *American Journal of Preventative Medicine*, 20(3S).
Hajishengallis, G. & Lamont, R.J., 2012. Beyond the red complex and into more complexity: the polymicrobial synergy and dysbiosis (PSD) model of periodontal disease etiology. *Molecular Oral Microbiology*, 27(6), pp.409–419.
Harper, L.M. et al., 2012. The interaction effect of bacterial vaginosis and periodontal disease on the risk of preterm delivery. *Am J Perinatol*, 29(5), pp.347–52.
Hayashi, C. et al., 2010. Pathogen-induced inflammation at sites distant from oral infection: bacterial persistence and induction of cell-specific innate immune inflammatory pathways. *Molecular Oral Microbiology*, 25(5), pp.305–316.
He, J. et al., 2015. The oral microbiome diversity and its relation to human diseases. *Folia Microbiol*, 60, pp.69–80.
Hill, G., 1998. Preterm Birth: Associations With Genital and Possibly Oral Microflora. *Annals of Peridontology*, 3(1), pp.222–232.
Hillier, S.L. et al., 1995. Association between bacterial vaginosis and preterm delivery of a low- birth-weight infant. *New England Journal of Medicine*, 333(26), pp.1737–1742.
Holgerson, P.L. et al., 2013. Oral Microbial Profile Discriminates Breast-fed From Formula-

fed Infants. *Hepatology and Nutrition*, 56(2), pp.127–136.
Huang, B. et al., 2015. The Changing Landscape of the Vaginal Microbiome. *Clin Lab Med*, 34(4), pp.747–761.
Iheozor-Ejiofor, Z. et al., 2017. Treating periodontal disease for preventing adverse birth outcomes in pregnant women (Review). , (6).
Jarjoura, K. et al., 2005. Markers of periodontal infection and preterm birth. *American journal of obstetrics and gynecology*, 192, pp.513–9.
Jaspan, H.B., 2011. The Wrong Place at the Wrong Time: Geographic Disparities in Young People's HIV Risk. *Journal of Adolescent Health*, 49(3).
Jill E. Clarrdige III, 2004. Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. *Clinical Microbiology Reviews*, 17(4), pp.840–862.
Johnson, L.F. & Geffen, N., 2016. A Comparison of Two Mathematical Modeling Frameworks for Evaluating Sexually Transmitted Infection Epidemiology. *Sexually transmitted diseases*, 43(3), pp.139–46. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26859800.
Katz, I. & Low-Beer, D., 2008. Why Has HIV Stabilized in South Africa, Yet Not Declined Further? Age and Sexual Behavior Patterns Among Youth. *Sexually Transmitted Diseases*, 35(10), pp.837–842.
Keijser, B.J.F. et al., 2008. Pyrosequencing Analysis of the Oral Microflora of Healthy Adults. *Journal of Dental Research*, 87(11), pp.1016–1020.
Kenyon, C., Colebunders, R. & Crucitti, T., 2013. The global epidemiology of bacterial vaginosis: a systematic review. *The American Journal of Obstetrics & Gynecology*, pp.505–523. Available at: http://dx.doi.org/10.1016/j.ajog.2013.05.006.
Khashan, A.S., Baker, P.N. & Kenny, L.C., 2010. Preterm birth and reduced birthweight in first and second teenage pregnancies: a register-based cohort study. *BMC Pregnancy & Childbirth*, 10(36).
Kiss, H. et al., 2007. Vaginal Lactobacillus microbiota of healthy women in the late first trimester of pregnancy. *BJOG*, 114, pp.1402–1407.
Kistler, J.O. et al., 2018. The oral microbiome in human immunodeficiency virus (HIV)-positive individuals. *Journal of Medical Microbiology (2015),*, 64(2015), pp.1094–1101.
Konopka, T. et al., 2003. The secretion of prostaglandin E2 and interleukin 1-beta in women with periodontal diseases and preterm low-birth-weight. *ulletin du Groupement International pour la Recherche Scientifique en Stomatologie et Odontologie*, 45(1), pp.18–28.
Kroon, S.J., Ravel, J. & Huston, W.M., 2018. Cervicovaginal microbiota, women's health, and reproductive outcomes. *Fertility and Sterility*, 110(3), pp.327–336. Available at: https://doi.org/10.1016/j.fertnstert.2018.06.036.
Kumar, P.S. et al., 2011. Target Region Selection Is a Critical Determinant of Community Fingerprints Generated by 16S Pyrosequencing. *PLos ONE*, 6(6), pp.1–8.
Lamont, R.F., 2005. Can antibiotics prevent preterm birth — the pro and con debate. *BJOG: An International Journal of Obstetrics and Gynaecology*, 112(Supplement 1), pp.67–73.
Lamont, R.F. et al., 2003. The efficacy of vaginal clindamycin for the treatment of abnormal genital tract flora in pregnancy. , 11(4), pp.181–189.
Leitich, H., Bodner-adler, B. & Brunbauer, M., 2003. Bacterial vaginosis as a risk factor for preterm delivery: A meta-analysis. *American journal of obstetrics and gynecology*, 189(1), pp.139–147.

Lennard, K. et al., 2017. Microbial composition predicts genital tract inflammation and persistent bacterial vaginosis in adolescent South African women. *American Society for Microbiology*, (October).

Lennard, K. et al., 2018. Microbial Composition Predicts Genital Tract Inflammation and Persistent Bacterial Vaginosis in South African Adolescent Females. *Infection and Immunity*, 86(1), pp.1–18.

Lerner, A. et al., 2016. Dysbiosis May Trigger Autoimmune Diseases via Inappropriate Post-Translational Modification of Host Proteins. *Frontiers in Microbilogy*, 7, pp.1–6.

Lewis, D. a et al., 2012. Urethritis/cervicitis pathogen prevalence and associated risk factors among asymptomatic HIV-infected patients in South Africa. *Sexually transmitted diseases*, 39(7), pp.531–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22706215.

Li, J. et al., 2014. Comparative analysis of the human saliva microbiome from different climate zones : Alaska , Germany , and Africa. , pp.1–13.

Li, K., Bihan, M. & Methe, B.A., 2013. Analyses of the Stability and Core Taxonomic Memberships of the Human Microbiome. *Plos One*, 8(5).

Li, Y. et al., 2005. Mode of Delivery and Other Maternal Factors Influence the Acquisition of Streptococcus mutans in Infants. *Journal of Dental Research*, 84(9).

Liu, L. et al., 2012. Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology*.

Liu, L. et al., 2017. Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. *The Lancet*, 388, pp.3027–3035. Available at: http://dx.doi.org/10.1016/S0140-6736(16)31593-8.

Lloyd-Price, J., Abu-Ali, G. & Huttenhower, C., 2016. The healthy human microbiome. *Genome Medicine*, 8(51), pp.1–11. Available at: http://dx.doi.org/10.1186/s13073-016-0307-y.

Long, J. et al., 2017. Association of oral microbiome with Type 2 diabetes risk. *Journal of Periodontal Research*, 52(3), pp.636–643.

López-López, A. et al., 2017. Health-Associated Niche Inhabitants as Oral Probiotics : The Case of Streptococcus dentisani. *Frontiers in Microbiology*, 8(March), pp.1–12.

Lozupone, C. et al., 2010. UniFrac : an effective distance metric for microbial community comparison. *The ISME Journal*, 5(2), pp.169–172. Available at: http://dx.doi.org/10.1038/ismej.2010.133.

Ma, B., Forney, L.J. & Jacques Ravel, 2012. The vaginal microbiome: rethinking health and diseases. *Annu. Rev. Microbiol.*, 66, pp.371–389.

Madianos, P.N. et al., 2003. Maternal Periodontitis and Prematurity, Part II: Maternal Infection and Fetal Exposure. *Annals of Peridontology*, 6, pp.175–182.

Mändar, R. & Marika Mikelsaar, 1996. Transmission of Mother’s Microflora to the Newborn. *Neonatology*, 69(1), pp.30–365.

Martin, J.A. et al., 2007. Births: Final Data for 2005. *National Vital Statistics Reports*, 56(6).

Masson, L. et al., 2015. Genital Inflammation and the Risk of HIV Acquisition in Women. *Clinical Infectious Diseases*, 61(2), pp.260–269. Available at: http://cid.oxfordjournals.org/lookup/doi/10.1093/cid/civ298.

Mcmurdie, P.J. & Holmes, S., 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, 8(4).

Mcmurdie, P.J. & Holmes, S., 2014. Waste Not , Want Not : Why Rarefying Microbiome Data

Is Inadmissible. , 10(4).
Mendes-soares, H. et al., 2014. Comparative functional genomics of Lactobacillus spp . reveals possible mechanisms for specialization of vaginal lactobacilli to their environment. *Journal of Bacteriology*.
Menon, R. et al., 2008. Amniotic fluid interleukin-6 increase is an indicator of spontaneous preterm birth in white but not black Americans. *American journal of obstetrics and gynecology*, 198(77).
Menon, R., Williams, S.M. & Fortunato, S.J., 2007. Amniotic Fluid Interleukin-1B and Interleukin-8 Concentrations: Racial Disparity in Preterm Birth. *Reproductive Sciences*, 14(3).
Mercer, B.M. et al., 1999. The Preterm Prediction Study: Effect of gestational age and cause of preterm birth on subsequent obstetric outcome. *American journal of obstetrics and gynecology*, 181(5), pp.1216–1221.
Michaud, D.S. & Izard, J., 2014. Microbiota, Oral Microbiome, and Pancreatic Cancer Dominique. *Cancer Journal*, 20(3), pp.203–206.
Mirmonsef, P. et al., 2011. The Effects of Commensal Bacteria on Innate Immune Responses in the Female Genital Tract. *American Journal of Reproductive Immunology*, 65, pp.190–195.
Moallaei, H. & Namazi, M.J., 2015. Evaluation and Comparison between Amsel ' s Criteria and Nugent ' s Score Methods in Diagnosis of Bacterial Vaginos is in Non-pregnant Women Complete Evaluation and Comparison between Amsel ' s Criteria and Nugent ' s Score Methods in Diagnosis of Bacterial Vaginos is in Non-pregnant Women. , (January).
Myer, L. et al., 2005. Bacterial vaginosis and susceptibility to HIV infection in South African women: a nested case-control study. *The Journal of infectious diseases*, 192(8), pp.1372–1380.
Nakazato, T., Ohta, T. & Bono, H., 2013. Experimental Design-Based Functional Mining and Characterization of High-Throughput Sequencing Data in the Sequence Read Archive. *PLoS ONE*, 8(10).
Nasidze, I. et al., 2009. Global diversity in the human salivary microbiome. *Genome Research*, pp.636–643.
Nugent, R.P., Krohn, M.A. & Hillier, S.L., 1991. Reliability of Diagnosing Bacterial Vaginosis Is Improved by Standardized Method of Gram Stain Interpretation a. *Journal of clinical microbiology*, 29(2), pp.297–301.
Ochert, A.S. et al., 1994. Inhibitory effect of salivary fluids on PCR : potency and removal. *Genome Research*, pp.6–11.
Offenbacher, S. et al., 1998. Potential Pathogenic Mechanisms of Periodontitis-Associated Pregnancy Complications. *Journal of Periodontology*, 3(1).
Oksanen, A.J. et al., 2018. Package "vegan ." *R Foundation for Statistical Computing, Vienna, Austria.*
Othman, M., Alfirevic, Z. & Neilson, J., 2007. Probiotics for preventing preterm labour. *Chochrane Database for Sytematic Reviews*, (1).
Paddick, J.S. et al., 2003. Effect of the Environment on Genotypic Diversity of Actinomyces naeslundii and Streptococcus oralis in the Oral Biofilm. *Applied and Environmental Microbiology*, 69(11), pp.6475–6480.
Panday, S. et al., 2009. Teenage pregnancy in South Africa: with a specific focus on school-going learners. *South African National Department of Edication*.

Parahitiyawa, N.B. et al., 2009. Microbiology of Odontogenic Bacteremia: beyond Endocarditis. *Clinical Microbiology Reviews*, 22(1), pp.46–64.
Paulson, J.N. et al., 2013. Differential abundance analysis for microbial marker-gene surveys. *Nature Methods*, 10(12).
Van de Peer, Y., Chapelle, S. & de Wachter, R., 2013. A quantitative map of nucleotide substitution rates in bacterial rRNA. *Nucleic Acids Research*, 24(13), pp.3381–3391.
Peipert, J.F. et al., 2008. Bacterial Vaginosis, Race, and Sexually Transmitted Infections: Does Race Modify the Association ? *Sexually Transmistted Diseases*, 35(4), pp.363–367.
Pérez-Chaparro, P.J. et al., 2014. Newly Identified Pathogens Associated with Periodontitis: A Systematic Review. *J Dent Res*, 93(9), pp.846–858.
Persson, R. et al., 2009. The vaginal microflora in relation to gingivitis. *BMC infectious diseases*, 9, p.6.
Peterson, S.N. et al., 2013. The Dental Plaque Microbiome in Health and Disease. *PLoS ONE*, 8(3), pp.e58487–e58487.
Petricevic, L. et al., 2012. European Journal of Obstetrics & Gynecology and Reproductive Biology Characterisation of the oral , vaginal and rectal Lactobacillus flora in healthy pregnant and postmenopausal women. *European Journal of Obstetrics and Gynecology*, 160(1), pp.93–99. Available at: http://dx.doi.org/10.1016/j.ejogrb.2011.10.002.
Petrova, M. et al., 2017. Lactobacillus iners: Friend or Foe?: Trends in Microbiology. *Trends in Microbiology*, 25(3).
Pettifor, A.E. et al., 2011. A Tale of Two Countries: Rethinking Sexual Risk for HIV Among Young People in South Africa and the United States. *Journal of Adolescent Health*, 49(3), p.237–243.e1. Available at: http://dx.doi.org/10.1016/j.jadohealth.2010.10.002.
Prasanna, J.S., Karunakar, P., Sravya, M.N., Madhavi, B. and Manasa, A., 2018. Detrimental consequences of women life cycle on the oral cavity. *Journal of Oral Research and Review*, *10*(1), p.39.
Prachi, S., Jitender, S., Rahul, C., Jitendra, K., Priyanka, M., & Disha, S. (2019). Impact of oral contraceptives on periodontal health. *African health sciences*, *19*(1), 1795-1800.
Pretorius, C., Jagatt, A. & Lamont, R.F., 2007. The relationship between periodontal disease, bacterial vaginosis and preterm birth. *Journal of Perinatal Medicine*, 35, pp.93–99.
Price, M.N., Dehal, P.S. & Arkin, A.P., 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE*, 5(3).
Pruesse, E. et al., 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, 35(21), pp.7188–7196.
Pushalkar, S. et al., 2012. Comparison of oral microbiota in tumor and non-tumor tissues of patients with oral squamous cell carcinoma. *BMC Microbiology*, 12(144).
Ravel, J. et al., 2011. Vaginal microbiome of reproductive-age women. *PNAS*, 108(Suppl. 1).
Roda, R.P. et al., 2008. Bacteremia originating in the oral cavity . A review. *Medicina Oral Patologia Oral y Cirugia Bucal*, 13(6), pp.355–362.
Rogosa, M., 1960. Species Differentiation of Human Vaginal Lactobacilli. *J. gen. Microbiol.*
Ross, M.G. et al., 2013. Characterizing and measuring bias in sequence data. *Genome Biology*, 14(R51).
Rousseeuw, P. et al., 2018. Finding Groups in Data: Cluster Analysis Extended.
Ruma, M. et al., 2008. Maternal periodontal disease, systemic inflammation, and risk for preeclampsia. *American journal of obstetrics and gynecology*, 198(4).
Sarkar, A., Stoneking, M. & Nandineni, M.R., 2017. Unraveling the human salivary

microbiome diversity in Indian populations. *Plos One*, 12(9), pp.1–17.
Schiff, E. et al., 2003. Multiple Liver Abscesses After Dental Treatment. *Journal of Clinical Gastroenterology*, 36(4), pp.369–371.
Schmidt, B.L. et al., 2014. Changes in Abundance of Oral Microbiota Associated with Oral Cancer. *PLoS ONE*, 9(6).
Segata, N. et al., 2012. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biology*, 13(6), p.R42. Available at: http://genomebiology.com/2012/13/6/R42.
da Silva, R.M. et al., 2004. Characterization of Streptococcus constellatus strains recovered from a brain abscess and periodontal pockets in an immunocompromised patient. *Journal of Periodontology*, 75, pp.1720–1723.
Smart, S., Singal, A. & Mindel, A., 2004. Social and sexual risk factors for bacterial vaginosis. *Sex Transm Infect*, 80, pp.58–62.
Socransky, S.S. et al., 1998. Microbial complexes in subgingival plaque. *Journal of Clinical Periodontology*.
Spellerberg, I.F. & Fedor, P.J., 2003. Richness, Species Diversity and the ’ Shannon-Wiener ’ Index A. *Global Ecology and Biogeography*, 12(3), pp.177–179.
Spurbeck, R.R. & Arvidson, C.G., 2010. Lactobacillus jensenii Surface-Associated Proteins Inhibit Neisseria gonorrhoeae Adherence to Epithelial Cells. *Infection and Immunity*, 78(7), pp.3103–3111.
Srinivasan, S. et al., 2012. Bacterial communities in women with bacterial vaginosis: High resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. *PLoS ONE*, 7(6).
Srinivasan, U. et al., 2009. Vaginal and Oral Microbes, Host Genotype and Preterm Birth. *Medical Hypotheses*, 6(73), pp.963–975.
Sullivan, O.O. et al., 2009. Comparative genomics of lactic acid bacteria reveals a niche-specific gene set. *BMC Microbiology*, 9(50).
Sweet, R.L., 1995. Role of bacterial vaginosis in pelvic inflammatory disease. *Clin Infect Dis*, 20(Suppl 2).
Swidsinski, A. et al., 2005. Adherent Biofilms in Bacterial Vaginosis. *Obstetrics and gynecology*, 106(5), pp.1013–1023.
Takeshita, T. et al., 2016. Bacterial diversity in saliva and oral health-related conditions : the Hisayama Study. *Nature Publishing Group*, (November 2015), pp.1–11. Available at: http://dx.doi.org/10.1038/srep22164.
Takeshita, T. & Yamashita, Y., 2012. Evaluation of salivary microbiota from an ecological perspective. *Journal of Oral Biosciences*, 54(3), pp.128–131. Available at: http://dx.doi.org/10.1016/j.job.2012.04.003.
The Human Micrbiome Project Consortium, 2012. Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402), pp.207–214. Available at: http://dx.doi.org/10.1038/nature11234.
Toma, I. et al., 2012. Periodontal health status and bacteraemia from daily oral activities : systematic review / meta - analysis. , pp.213–228.
Ugwumadu, A. et al., 2003. Effect of Early Oral Clindamycin on Late Miscarriage and Preterm Delivery in Asymptomatic Women With Abnormal Vaginal Flora and Bacterial Vaginosis A Randomized, Controlled Trial. *Lancet*, 361, pp.983–988.
Vasquez, A. et al., 2002. Vaginal Lactobacillus Flora of Healthy Swedish Women. *Journal of Clinical Microbiology*, 40(8), pp.2746–2749.

Velez, D.R. et al., 2009. Spontaneous preterm birth in African Americans is associated with infection and inflammatory response gene variants. *Am J Obstetrics and Gynecology*, 200(2).

Vergnes, J. & Sixou, M., 2007. Preterm low birth weight and maternal periodontal status: a meta-analysis. *American journal of obstetrics and gynecology*, 196(135).

Verma, D. et al., 2018. Insights into the human oral microbiome. *Archives of Microbiology*, 200(4), pp.525–540. Available at: http://dx.doi.org/10.1007/s00203-018-1505-3.

Verstraelen, H. et al., 2009. 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 Microbiology*, 9(116).

Verstraelen, H. et al., 2007. Modified classification of Gram-stained vaginal smears to predict spontaneous preterm birth : a prospective cohort. *American journal of obstetrics and gynecology*, 196, p.528.e1-528.e6.

Verstraelen, H. & Verhelst, R., 2009. Bacterial vaginosis: An update on diagnosis and treatment. *Expert Review of Anti-infective Therapy*, 7(9).

Vodstrcil, L.A. et al., 2017. The influence of sexual activity on the vaginal microbiota and Gardnerella vaginalis clade diversity in young women. *PLoS ONE*, 12(2), pp.1–15.

Vogt, M. et al., 2012. Factors associated with the prevalence of periodontal disease in low-risk pregnant women. *Reproductive Health*, 9(1), p.3. Available at: http://www.reproductive-health-journal.com/content/9/1/3.

Vollaard, E.J. & Clasener, H.A.L., 1994. Colonization Resistance. *Antimicrobial Agents and Chemotherapy*, 38(3), pp.409–414.

Wang, Q. et al., 2007. Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and*, 73(16), pp.5261–5267.

Whittaker, R.J., Willis, K.J. & Field, R., 2001. Scale and Species Richness : Towards a General, Hierarchical Theory of Species Diversity. *Journal of Biogeography*, 28(4), pp.453–470.

Wiesenfeld, H.C. et al., 2003. Bacterial Vaginosis Is a Strong Predictor of Neisseria gonorrhoeae and Chlamydia trachomatis Infection. *HIV/AIDS*, 36, pp.663–668.

Van De Wijgert, J.H.H.M. et al., 2014. The Vaginal Microbiota : What Have We Learned after a Decade of Molecular Characterization ? *PLOS one*, 9(8).

Winkvist, A., Morgen, I. & Hgberg, U., 1998. Familial patterns in birth characteristics: impact on individual and population risks. *International Journal of Epidemiology*, 27, pp.248–254.

Witkin, S.S. et al., 2014. Influence of Vaginal Bacteria and D - and L -Lactic Acid Isomers on Vaginal Extracellular Matrix Metalloproteinase Inducer : Implications. *mBio*, 4(4), pp.1–8.

Xie, G. et al., 2010. Community and gene composition of a human dental plaque microbiota obtained by metagenomic sequencing. *Molecular Oral Microbiology*, 25(6), pp.391–405.

Xiong, X. et al., 2006. Periodontal disease and adverse pregnancy outcomes: a systematic review. *BJOG: An International Journal of Obstetrics and Gynaecology*, pp.135–143.

Yamashita, Y. & Takeshita, T., 2017. The oral microbiome and human health. *Journal of Oral Science*, 59(2), pp.201–206.

Yang, B., Wang, Y. & Qian, P., 2016. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics*, pp.1–8. Available at: http://dx.doi.org/10.1186/s12859-016-0992-y.

Yokogawa, K. et al., 1974. Mutanolysin, Bacteriolytic Agent for Cariogenic Streptococci: Partial Purification and Properties. *Antimicrobial Agents and Chemotherapy*, 6(2), pp.156–165.
Yuan, S. et al., 2012. Evaluation of Methods for the Extraction and Purification of DNA from the Human Microbiome. *PLoS ONE*, 7(3).
Zarate, G. & Nader-Macias, M.E., 2006. Influence of probiotic vaginal lactobacilli on in vitro adhesion of urogenital pathogens to vaginal epithelial cells. *Letters in Apllied Microbiology*, 43, pp.174–180.
Zarco, M.F., Vess, T.J. & Ginsburg, G.S., 2012. The oral microbiome in health and disease and the potential impact on personalized dental medicine. *Oral Diseases*, 18, pp.109–120.
Zaura, E. et al., 2014. Acquiring and maintaining a normal oral microbiome : current perspective. *Frontier ins Cellular and Infection Microbiology*, 4(June), pp.1–8.
Zaura, E. et al., 2009. Defining the healthy “core microbiome” of oral microbial communities. *BMC Microbiology*, 9(1471–2180 (Electronic)), p.259.
Zbinden, A. et al., 2014. Frequent detection of Streptococcus tigurinus in the human oral microbial flora by a specific 16S rRNA gene real-time TaqMan PCR. *BMC Microbiology*, 14(231), pp.1–7.
Zbinden, A. et al., 2012. Streptococcus tigurinus , a Novel Member of the Streptococcus mitis Group , Causes Invasive Infections. , 50(9), pp.2969–2973.
Zbinden, A., Bostanci, N. & Belibasakis, G.N., 2015. The novel species Streptococcus tigurinus and its association with oral infection. *Virulence*, 6(3), pp.177–182.
Zhou, X. et al., 2011. The Vaginal Bacterial Communities of Japanese Women Resemble Those of Women in Other Racial Groups. *FEMS Immunol Med Microbiol*, 58(2), pp.1–19.
aeruginosa identified in association with oral squamous cell carcinoma. *Scientific Reports*, (2)6
Alakomi, H.L. et al., 2000. Lactic Acid Permeabilizes Gram-Negative Bacteria by Disrupting the Outer Membrane. *Applied and Environmental Microbiology*, 66(5)
Amann, R.I., Ludwig, W. & Schleifer, K.-H., 1995. Phylogenetic Identification and In Situ Detection of Individual Microbial Cells without Cultivation. *Microbiological reviews*, 59(1)
Amar, S. et al., 2003. Periodontal Disease Is Associated With Brachial Artery Endothelial Dysfunction and Systemic Inflammation.
Amsel, R. et al., 1983. Nonspecific vaginitis: diagnostic criteria and microbial and epidemiologic associations. *The American journal of medicine*, 74(1)
Anahtar, M.N. et al., 2015. Cervicovaginal Bacteria Are a Major Modulator of Host Inflammatory Responses in the Female Genital Tract. *Immunity*, 42(5)
Anders, S. & Huber, W., 2010. Differential expression analysis for sequence count data. *Genome Biology*, 11.
Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data.
Ashimoto, A., Slots, M.J. & J., F., 1995. Molecular genetic detection of Bacteroides heparinolyticus in adult periodontitis. *Oral microbiology and immunology*, 10(5)
Bailey, G.D. et al., 2019. Bacteroides heparinolyticus : Deoxyribonucleic Acid Relatedness of Strains from the Oral Cavity and Oral-Associated Disease Conditions of Horses , Cats , and Humans.
Balle, C. et al., 2018. Endocervical and vaginal microbiota in South African adolescents with asymptomatic Chlamydia trachomatis infection. *Scientific Reports*.

Belda-Ferre, P. et al., 2011. The oral metagenome in health and disease. *ISME Journal*,
Bennett, W.A. et al., 2000. Intrauterine endotoxin infusion in rat pregnancy induces preterm delivery and increases placental prostaglandin F2α metabolite levels. *American Journal of Obstetrics and Gynecology*, 182(6)
Bik, E.M. et al., 2010. Bacterial diversity in the oral cavity of ten healthy individuals Elisabeth. *ISME Journal*, 4(8)
Black, V. et al., 2008. The detection of urethritis pathogens among patients with the male urethritis syndrome, genital ulcer syndrome and HIV voluntary counselling and testing clients: should South Africa's syndromic management approach be revised? *Sexually transmitted infections*, 84(4)
Boggess, K.A. et al., 2005. Fetal immune response to oral pathogens and risk of preterm birth. *American journal of obstetrics and gynecology*, 193, pp.1121–6.
Borgdorff, H. et al., 2016. Unique Insights in the Cervicovaginal Lactobacillus iners and L. crispatus Proteomes and Their Associations with Microbiota Dysbiosis. *PLoS ONE*, 11(3).
Boris, S. et al., 1998. Adherence of Human Vaginal Lactobacilli to Vaginal Epithelial Cells and Interaction with Uropathogens. *Infection and Immunity*, 66(5.
Boskey, E.R. et al., 2014. Origins of vaginal acidity: high D/L lactate ratio is consistent with bacteria being the primary source. *Human Reproduction Volume*, 16(9)
Bradshaw, C.S. et al., 2006. High Recurrence Rates of Bacterial Vaginosis over the Course of 12 Months after Oral Metronidazole Therapy and Factors Associated with Recurrence. *JID*, 193.
Briselden, A.N.N.M. et al., 1992. Sialidases (Neuraminidases) in Bacterial Vaginosis and Bacterial Vaginosis-Associated Microflora. *Journal of clinical microbiology*, 30(3)
Brocklehurst, P. et al., 2013. Antibiotics for treating bacterial vaginosis in pregnancy. *Cochrane Database of Systematic Reviews*, (1).
Campbell, J.H. et al., 2013. UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *PNAS*
Caporaso, J.G. et al., 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7(5).
Cassini, M. a. et al., 2013. Periodontal bacteria in the genital tract: Are they related to adverse pregnancy outcome? *International Journal of Immunopathology and Pharmacology*, 26(4)
Cauci, S. et al., 2005. Combination of vaginal pH with vaginal sialidase and prolidase activities for prediction of low birth weight and preterm birth.
Celik, H. & Ayar, A., 2002. Effects of erythromycin on pregnancy duration and birth weight in lipopolysaccharide-induced preterm labor in pregnant rats. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 103
Centre for Disease Control, 2015. *Sexually Transmitted Diseases Treatment Guidelines , 2015*,
Chakravorty, S. et al., 2008. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiology Methods*, 69(2), pp.330–339.
Cherpes, T.L. et al., 2008. 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. *Sexually Transmistted Diseases*, 35(1)
Cho, I. & Blaser, M.J., 2012. The human microbiome at the interface of health and disease.

Nature Genetics, 13.
Chu, D.M. et al., 2017. Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nature Medicine*, 23(3)
Claassen, S. et al., 2013. A comparison of the efficiency of five different commercial DNA extraction kits for extraction of DNA from faecal samples. *Journal of Microbiological Methods*, 94(2)
Corby, P.M. et al., 2005. Microbial Risk Indicators of Early Childhood Caries. *Journal of Clinical Microbiology*, 43(11)
Costalonga, M. & Herzberg, M.C., 2014. The oral microbiome and the immunobiology of periodontal disease and caries. *Immunology Letters*, 162
Costalonga, M. & Herzberg, M.C., 2014. The oral microbiome and the immunobiology of periodontal disease and caries. *Immunology Letters*, 162(2)
Costello, E.K. et al., 2012. The Application of Ecological Theory. *Science*, 336, pp.1255–1263.
Cowan, F. & Pettifor, A., 2009. HIV in adolescents in sub-Saharan Africa. *Current Opinion in HIV and AIDS*, 4(4)
Curtis, J.T. & Roger Bray, J., 1957. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, 27(4)
Curtis, M.A., Zenobia, C. & Darveau, R.P., 2011. The relationship of the oral micrpbiota to periodontal health and disease. *Cell Host and Microbe*, 10(4)
D'Aiuto, F. et al., 2004. Periodontitis and Systemic Inflammation_ Control of the Local Infection is Associated with a Reduction in Serum Inflammatory Markers. *Journal of Dental Research*.
Dalwai, F., Spratt, D.A. & Pratten, J., 2007. Use of Quantitative PCR and Culture Methods To Characterise Ecological Flux in Bacterial Biofilms. *Journal of Clinical Microbiology*, 45(9)
Dasanayake, A.P. et al., 2003. Preterm low birth weight and periodontal disease among African Americans. *Dental Clinics of North America*, 47
Dean, S. V et al., 2013. Born Too Soon: Care before and between pregnancy to prevent preterm births : from evidence to action. *Reproductive Health*, 10(Suppl 1)
Desantis, T.Z. et al., 2006. Greengenes , a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Applied and Environmental Microbiology*, 72(7)
Dewhirst, F.E. et al., 2010. The Human Oral Microbiome. *Journal of Bacteriology*, 192(19),
Dominguez-bello, M.G. et al., 2010. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *PNAS*, 107(26)
Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19)
Edgar, R.C., 2018. Search and clustering orders of magnitude faster than BLAST. , 26(19),
Faith, D.P., Minchin, P.R. & Belbin, L., 1987. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio*, 69
Fethers, K.A. et al., 2008. Sexual Risk Factors and Bacterial Vaginosis : A Systematic Review and Meta-Analysis. *Clinical Infectious Diseases*, 47.
Fettweis, J.M. et al., 2014. Differences in vaginal microbiome in African American women versus women of European ancestry. *Microbiology*, 160, pp.2272–2282.
Figuero, E. et al., 2013. Effect of pregnancy on gingival inflammation in systemically healthy women: a systematic review. *Journal of Clinical Periodontology*, 40
Filippis, F. De et al., 2014. The Same Microbiota and a Potentially Discriminant Metabolome in the Saliva of Omnivore , Ovo-Lacto- Vegetarian and Vegan Individuals. , 9(11).

Flint Porter, T. et al., 1997. The Risk of Preterm Birth Across Generations. *Obstetrics and gynecology*, 90(1).
Forner, L. et al., 2006. Incidence of bacteremia after chewing , tooth brushing and scaling in individuals with periodontal inflammation.
Fouhy, F. et al., 2016. 16S rRNA gene sequencing of mock microbial populations- impact of DNA extraction method , primer choice and sequencing platform. *BMC Microbiology*, 16(123)
Fourie, N.H. et al., 2016. The microbiome of the oral mucosa in irritable bowel syndrome. *Gut Microbes*, 7(4)
Fredricks, D.N., Fiedler, T.L. & Marrazzo, J.M., 2005. Molecular Identification of Bacteria Associated with Bacterial Vaginosis. *New England Journal of Medicine*, 353
Freire, C. De et al., 2010. Lysostaphin: A Staphylococcal Bacteriolysin with Potential Clinical Applications. *Pharmaceuticals*, 3
Gajer, P. et al., 2012. Temporal Dynamics of the Human Vaginal Microbiota. *Sci Transl Med*, 4(132).
Gaujoux, R., 2014. Generating heatmaps for Non negative Matrix Factorization. *R Foundation for Statistical Computing, Vienna, Austria*.
Gill, C. et al., 2016. Evaluation of Lysis Methods for the Extraction of Bacterial DNA for Analysis of the Vaginal Microbiota. *PLoS ONE*, 11(9)
Gower, A.J.C., 1966. Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis. *Biometrika*, 53(3)
Graspeuntner, S. et al., 2018. Selection of validated hypervariable regions is crucial in 16S-based microbiota studies of the female genital tract. *Scientific Reports*, 8(9678)
Griffen, A.L. et al., 2011. CORE : A Phylogenetically-Curated 16S rDNA Database of the Core Oral Microbiome. *PLoS One*, 6(4)
Guise, J. et al., 2001. Screening for Bacterial Vaginosis in Pregnancy. *American Journal of Preventative Medicine*, 20(3S).
Hajishengallis, G. & Lamont, R.J., 2012. Beyond the red complex and into more complexity: the polymicrobial synergy and dysbiosis (PSD) model of periodontal disease etiology. *Molecular Oral Microbiology*, 27(6)
Harper, L.M. et al., 2012. The interaction effect of bacterial vaginosis and periodontal disease on the risk of preterm delivery. *Am J Perinatol*, 29(5)
Hayashi, C. et al., 2010. Pathogen-induced inflammation at sites distant from oral infection: bacterial persistence and induction of cell-specific innate immune inflammatory pathways. *Molecular Oral Microbiology*, 25(5)
He, J. et al., 2015. The oral microbiome diversity and its relation to human diseases. *Folia Microbiol*, 60
Hill, G., 1998. Preterm Birth: Associations With Genital and Possibly Oral Microflora. *Annals of Peridontology*, 3(1)
Hillier, S.L. et al., 1995. Association between bacterial vaginosis and preterm delivery of a low- birth-weight infant. *New England Journal of Medicine*, 333(26)
Holgerson, P.L. et al., 2013. Oral Microbial Profile Discriminates Breast-fed From Formula-fed Infants. *Hepatology and Nutrition*, 56(2)
Huang, B. et al., 2015. The Changing Landscape of the Vaginal Microbiome. *Clin Lab Med*, 34(4)
Iheozor-Ejiofor, Z. et al., 2017. Treating periodontal disease for preventing adverse birth outcomes in pregnant women

Jarjoura, K. et al., 2005. Markers of periodontal infection and preterm birth. *American journal of obstetrics and gynecology*, 192
Jaspan, H.B., 2011. The Wrong Place at the Wrong Time: Geographic Disparities in Young People's HIV Risk. *Journal of Adolescent Health*, 49(3).
Jill E. Clarrdige III, 2004. Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. *Clinical Microbiology Reviews*, 17(4)
Johnson, L.F. & Geffen, N., 2016. A Comparison of Two Mathematical Modeling Frameworks for Evaluating Sexually Transmitted Infection Epidemiology. *Sexually transmitted diseases*, 43(3)
Katz, I. & Low-Beer, D., 2008. Why Has HIV Stabilized in South Africa, Yet Not Declined Further? Age and Sexual Behavior Patterns Among Youth. *Sexually Transmitted Diseases*, 35(10)
Keijser, B.J.F. et al., 2008. Pyrosequencing Analysis of the Oral Microflora of Healthy Adults. *Journal of Dental Research*, 87(11)
Kenyon, C., Colebunders, R. & Crucitti, T., 2013. The global epidemiology of bacterial vaginosis: a systematic review. *The American Journal of Obstetrics & Gynecology*, pp.505–523. Available at: http://dx.doi.org/10.1016/j.ajog.2013.05.006.
Khashan, A.S., Baker, P.N. & Kenny, L.C., 2010. Preterm birth and reduced birthweight in first and second teenage pregnancies: a register-based cohort study. *BMC Pregnancy & Childbirth*, 10(36).
Kistler, J.O. et al., 2018. The oral microbiome in human immunodeficiency virus (HIV)-positive individuals. *Journal of Medical Microbiology (2015),*, 64(2015)
Konopka, T. et al., 2003. The secretion of prostaglandin E2 and interleukin 1-beta in women with periodontal diseases and preterm low-birth-weight. *Recherche Scientifique en Stomatologie et Odontologie*, 45(1)
Kroon, S.J., Ravel, J. & Huston, W.M., 2018. Cervicovaginal microbiota, women's health, and reproductive outcomes. *Fertility and Sterility*, 110(3)
Kumar, P.S. et al., 2011. Target Region Selection Is a Critical Determinant of Community Fingerprints Generated by 16S Pyrosequencing. *PLos ONE*, 6(6)
Lamont, R.F., 2005. Can antibiotics prevent preterm birth — the pro and con debate. *BJOG: An International Journal of Obstetrics and Gynaecology*, 112(Supplement 1),
Lamont, R.F. et al., 2003. The efficacy of vaginal clindamycin for the treatment of abnormal genital tract flora in pregnancy. , 11(4)
Leitich, H., Bodner-adler, B. & Brunbauer, M., 2003. Bacterial vaginosis as a risk factor for preterm delivery: A meta-analysis. *American journal of obstetrics and gynecology*, 189(1)
Lennard, K. et al., 2017. Microbial composition predicts genital tract inflammation and persistent bacterial vaginosis in adolescent South African women. *American Society for Microbiology*,
Lennard, K. et al., 2018. Microbial Composition Predicts Genital Tract Inflammation and Persistent Bacterial Vaginosis in South African Adolescent Females. *Infection and Immunity*, 86(1)
Lerner, A. et al., 2016. Dysbiosis May Trigger Autoimmune Diseases via Inappropriate Post-Translational Modification of Host Proteins. *Frontiers in Microbilogy*, 7, pp.1–6.
Lewis, D. a et al., 2012. Urethritis/cervicitis pathogen prevalence and associated risk factors among asymptomatic HIV-infected patients in South Africa. *Sexually transmitted*

diseases, 39(7)

Li, J. et al., 2014. Comparative analysis of the human saliva microbiome from different climate zones : Alaska , Germany , and Africa.

Li, K., Bihan, M. & Methe, B.A., 2013. Analyses of the Stability and Core Taxonomic Memberships of the Human Microbiome. *Plos One*, 8(5).

Li, Y. et al., 2005. Mode of Delivery and Other Maternal Factors Influence the Acquisition of Streptococcus mutans in Infants. *Journal of Dental Research*, 84(9).

Liu, L. et al., 2012. Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology*.

Liu, L. et al., 2017. Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. *The Lancet*, 388,

Lloyd-Price, J., Abu-Ali, G. & Huttenhower, C., 2016. The healthy human microbiome. *Genome Medicine*, 8(51)

Long, J. et al., 2017. Association of oral microbiome with Type 2 diabetes risk. *Journal of Periodontal Research*, 52(3),

López-López, A. et al., 2017. Health-Associated Niche Inhabitants as Oral Probiotics : The Case of Streptococcus dentisani. *Frontiers in Microbiology*, 8(March), pp.1–12.

Lozupone, C. et al., 2010. UniFrac : an effective distance metric for microbial community comparison. *The ISME Journal*, 5(2)

Ma, B., Forney, L.J. & Jacques Ravel, 2012. The vaginal microbiome: rethinking health and diseases. *Annu. Rev. Microbiol.*, 66

Madianos, P.N. et al., 2003. Maternal Periodontitis and Prematurity, Part II: Maternal Infection and Fetal Exposure. *Annals of Peridontology*, 6, pp.175–182.

Mändar, R. & Marika Mikelsaar, 1996. Transmission of Mother's Microflora to the Newborn. *Neonatology*, 69(1)

Martin, J.A. et al., 2007. Births: Final Data for 2005. *National Vital Statistics Reports*, 56(6).

Masson, L. et al., 2015. Genital Inflammation and the Risk of HIV Acquisition in Women. *Clinical Infectious Diseases*, 61(2)

Mcmurdie, P.J. & Holmes, S., 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, 8(4).

Mcmurdie, P.J. & Holmes, S., 2014. Waste Not , Want Not : Why Rarefying Microbiome Data Is Inadmissible. , 10(4).

Mendes-soares, H. et al., 2014. Comparative functional genomics of Lactobacillus spp . reveals possible mechanisms for specialization of vaginal lactobacilli to their environment. *Journal of Bacteriology*.

Menon, R. et al., 2008. Amniotic fluid interleukin-6 increase is an indicator of spontaneous preterm birth in white but not black Americans. *American journal of obstetrics and gynecology*, 198(77).

Menon, R., Williams, S.M. & Fortunato, S.J., 2007. Amniotic Fluid Interleukin-1B and Interleukin-8 Concentrations: Racial Disparity in Preterm Birth. *Reproductive Sciences*, 14(3).

Mercer, B.M. et al., 1999. The Preterm Prediction Study: Effect of gestational age and cause of preterm birth on subsequent obstetric outcome. *American journal of obstetrics and gynecology*, 181(5

Michaud, D.S. & Izard, J., 2014. Microbiota, Oral Microbiome, and Pancreatic Cancer Dominique. *Cancer Journal*, 20(3)

Mirmonsef, P. et al., 2011. The Effects of Commensal Bacteria on Innate Immune Responses in the Female Genital Tract. *American Journal of Reproductive Immunology*, 65
Moallaei, H. & Namazi, M.J., 2015. Evaluation and Comparison between Amsel ' s Criteria and Nugent ' s Score Methods in Diagnosis of Bacterial Vaginos is in Non-pregnant Women
Myer, L. et al., 2005. Bacterial vaginosis and susceptibility to HIV infection in South African women: a nested case-control study. *The Journal of infectious diseases*, 192(8),
Nakazato, T., Ohta, T. & Bono, H., 2013. Experimental Design-Based Functional Mining and Characterization of High-Throughput Sequencing Data in the Sequence Read Archive. *PLoS ONE*, 8(10).
Nasidze, I. et al., 2009. Global diversity in the human salivary microbiome. *Genome Research*
Nugent, R.P., Krohn, M.A. & Hillier, S.L., 1991. Reliability of Diagnosing Bacterial Vaginosis Is Improved by Standardized Method of Gram Stain Interpretation a. *Journal of clinical microbiology*, 29(2)
Ochert, A.S. et al., 1994. Inhibitory effect of salivary fluids on PCR : potency and removal. *Genome Research*.
Offenbacher, S. et al., 1998. Potential Pathogenic Mechanisms of Periodontitis-Associated Pregnancy Complications. *Journal of Periodontology*, 3(1).
Oksanen, A.J. et al., 2018. Package "vegan ." *R Foundation for Statistical Computing, Vienna, Austria.*
Othman, M., Alfirevic, Z. & Neilson, J., 2007. Probiotics for preventing preterm labour. *Chochrane Database for Sytematic Reviews*, (1).
Paddick, J.S. et al., 2003. Effect of the Environment on Genotypic Diversity of Actinomyces naeslundii and Streptococcus oralis in the Oral Biofilm. *Applied and Environmental Microbiology*, 69(11), pp.6475–6480.
Panday, S. et al., 2009. Teenage pregnancy in South Africa: with a specific focus on school-going learners. *South African National Department of Edication*.
Parahitiyawa, N.B. et al., 2009. Microbiology of Odontogenic Bacteremia: beyond Endocarditis. *Clinical Microbiology Reviews*, 22(1), pp.46–64.
Paulson, J.N. et al., 2013. Differential abundance analysis for microbial marker-gene surveys. *Nature Methods*, 10(12).
Van de Peer, Y., Chapelle, S. & de Wachter, R., 2013. A quantitative map of nucleotide substitution rates in bacterial rRNA. *Nucleic Acids Research*, 24(13), pp.3381–3391.
Peipert, J.F. et al., 2008. Bacterial Vaginosis, Race, and Sexually Transmitted Infections: Does Race Modify the Association ? *Sexually Transmistted Diseases*, 35(4), pp.363–367.
Pérez-Chaparro, P.J. et al., 2014. Newly Identified Pathogens Associated with Periodontitis: A Systematic Review. *J Dent Res*, 93(9), pp.846–858.
Persson, R. et al., 2009. The vaginal microflora in relation to gingivitis. *BMC infectious diseases*, 9, p.6.
Peterson, S.N. et al., 2013. The Dental Plaque Microbiome in Health and Disease. *PLoS ONE*, 8(3), pp.e58487–e58487.
Petrova, M. et al., 2017. Lactobacillus iners: Friend or Foe?: Trends in Microbiology. *Trends in Microbiology*, 25(3).
Pettifor, A.E. et al., 2011. A Tale of Two Countries: Rethinking Sexual Risk for HIV Among Young People in South Africa and the United States. *Journal of Adolescent Health*, 49(3), p.237–243.e1. Available at: http://dx.doi.org/10.1016/j.jadohealth.2010.10.002.
Pretorius, C., Jagatt, A. & Lamont, R.F., 2007. The relationship between periodontal disease,

bacterial vaginosis and preterm birth. *Journal of Perinatal Medicine*, 35, pp.93–99.
Price, M.N., Dehal, P.S. & Arkin, A.P., 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE*, 5(3).
Pruesse, E. et al., 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, 35(21), pp.7188–7196.
Pushalkar, S. et al., 2012. Comparison of oral microbiota in tumor and non-tumor tissues of patients with oral squamous cell carcinoma. *BMC Microbiology*, 12(144).
Ravel, J. et al., 2011. Vaginal microbiome of reproductive-age women. *PNAS*, 108(Suppl. 1).
Roda, R.P. et al., 2008. Bacteremia originating in the oral cavity . A review. *Medicina Oral Patologia Oral y Cirugia Bucal*, 13(6), pp.355–362.
Rogosa, M., 1960. Species Differentiation of Human Vaginal Lactobacilli. *J. gen. Microbiol.*
Ross, M.G. et al., 2013. Characterizing and measuring bias in sequence data. *Genome Biology*, 14(R51).
Rousseeuw, P. et al., 2018. Finding Groups in Data: Cluster Analysis Extended.
Ruma, M. et al., 2008. Maternal periodontal disease, systemic inflammation, and risk for preeclampsia. *American journal of obstetrics and gynecology*, 198(4).
Sarkar, A., Stoneking, M. & Nandineni, M.R., 2017. Unraveling the human salivary microbiome diversity in Indian populations. *Plos One*, 12(9), pp.1–17.
Schiff, E. et al., 2003. Multiple Liver Abscesses After Dental Treatment. *Journal of Clinical Gastroenterology*, 36(4), pp.369–371.
Schmidt, B.L. et al., 2014. Changes in Abundance of Oral Microbiota Associated with Oral Cancer. *PLoS ONE*, 9(6).
Segata, N. et al., 2012. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biology*, 13(6), p.R42. Available at: http://genomebiology.com/2012/13/6/R42.
da Silva, R.M. et al., 2004. Characterization of Streptococcus constellatus strains recovered from a brain abscess and periodontal pockets in an immunocompromised patient. *Journal of Periodontology*, 75, pp.1720–1723.
Smart, S., Singal, A. & Mindel, A., 2004. Social and sexual risk factors for bacterial vaginosis. *Sex Transm Infect*, 80, pp.58–62.
Socransky, S.S. et al., 1998. Microbial complexes in subgingival plaque. *Journal of Clinical Periodontology*.
Spellerberg, I.F. & Fedor, P.J., 2003. Richness, Species Diversity and the ' Shannon-Wiener ' Index A. *Global Ecology and Biogeography*, 12(3), pp.177–179.
Spurbeck, R.R. & Arvidson, C.G., 2010. Lactobacillus jensenii Surface-Associated Proteins Inhibit Neisseria gonorrhoeae Adherence to Epithelial Cells. *Infection and Immunity*, 78(7), pp.3103–3111.
Srinivasan, S. et al., 2012. Bacterial communities in women with bacterial vaginosis: High resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. *PLoS ONE*, 7(6).
Srinivasan, U. et al., 2009. Vaginal and Oral Microbes, Host Genotype and Preterm Birth. *Medical Hypotheses*, 6(73), pp.963–975.
Sullivan, O.O. et al., 2009. Comparative genomics of lactic acid bacteria reveals a niche-specific gene set. *BMC Microbiology*, 9(50).
Sweet, R.L., 1995. Role of bacterial vaginosis in pelvic inflammatory disease. *Clin Infect Dis*, 20(Suppl 2).

Swidsinski, A. et al., 2005. Adherent Biofilms in Bacterial Vaginosis. *Obstetrics and gynecology*, 106(5), pp.1013–1023.
Takeshita, T. et al., 2016. Bacterial diversity in saliva and oral health-related conditions : the Hisayama Study. *Nature Publishing Group*, (November 2015), pp.1–11. Available at: http://dx.doi.org/10.1038/srep22164.
Takeshita, T. & Yamashita, Y., 2012. Evaluation of salivary microbiota from an ecological perspective. *Journal of Oral Biosciences*, 54(3), pp.128–131. Available at: http://dx.doi.org/10.1016/j.job.2012.04.003.
The Human Micrbiome Project Consortium, 2012. Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402), pp.207–214. Available at: http://dx.doi.org/10.1038/nature11234.
Toma, I. et al., 2012. Periodontal health status and bacteraemia from daily oral activities : systematic review / meta - analysis. , pp.213–228.
Ugwumadu, A. et al., 2003. Effect of Early Oral Clindamycin on Late Miscarriage and Preterm Delivery in Asymptomatic Women With Abnormal Vaginal Flora and Bacterial Vaginosis A Randomized, Controlled Trial. *Lancet*, 361, pp.983–988.
Velez, D.R. et al., 2009. Spontaneous preterm birth in African Americans is associated with infection and inflammatory response gene variants. *Am J Obstetrics and Gynecology*, 200(2).
Vergnes, J. & Sixou, M., 2007. Preterm low birth weight and maternal periodontal status: a meta-analysis. *American journal of obstetrics and gynecology*, 196(135).
Verma, D. et al., 2018. Insights into the human oral microbiome. *Archives of Microbiology*, 200(4), pp.525–540. Available at: http://dx.doi.org/10.1007/s00203-018-1505-3.
Verstraelen, H. et al., 2009. 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 Microbiology*, 9(116).
Verstraelen, H. et al., 2007. Modified classification of Gram-stained vaginal smears to predict spontaneous preterm birth : a prospective cohort. *American journal of obstetrics and gynecology*, 196, p.528.e1-528.e6.
Verstraelen, H. & Verhelst, R., 2009. Bacterial vaginosis: An update on diagnosis and treatment. *Expert Review of Anti-infective Therapy*, 7(9).
Vodstrcil, L.A. et al., 2017. The influence of sexual activity on the vaginal microbiota and Gardnerella vaginalis clade diversity in young women. *PLoS ONE*, 12(2), pp.1–15.
Vogt, M. et al., 2012. Factors associated with the prevalence of periodontal disease in low-risk pregnant women. *Reproductive Health*, 9(1), p.3. Available at: http://www.reproductive-health-journal.com/content/9/1/3.
Vollaard, E.J. & Clasener, H.A.L., 1994. Colonization Resistance. *Antimicrobial Agents and Chemotherapy*, 38(3), pp.409–414.
Wang, Q. et al., 2007. Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and*, 73(16), pp.5261–5267.
Whittaker, R.J., Willis, K.J. & Field, R., 2001. Scale and Species Richness : Towards a General, Hierarchical Theory of Species Diversity. *Journal of Biogeography*, 28(4), pp.453–470.
Wiesenfeld, H.C. et al., 2003. Bacterial Vaginosis Is a Strong Predictor of Neisseria gonorrhoeae and Chlamydia trachomatis Infection. *HIV/AIDS*, 36, pp.663–668.
Van De Wijgert, J.H.H.M. et al., 2014. The Vaginal Microbiota : What Have We Learned after a Decade of Molecular Characterization ? *PLOS one*, 9(8).

Winkvist, A., Morgen, I. & Hgberg, U., 1998. Familial patterns in birth characteristics: impact on individual and population risks. *International Journal of Epidemiology*, 27, pp.248–254.

Witkin, S.S. et al., 2014. Influence of Vaginal Bacteria and D - and L -Lactic Acid Isomers on Vaginal Extracellular Matrix Metalloproteinase Inducer : Implications. *mBio*, 4(4), pp.1–8.

Xie, G. et al., 2010. Community and gene composition of a human dental plaque microbiota obtained by metagenomic sequencing. *Molecular Oral Microbiology*, 25(6), pp.391–405.

Xiong, X. et al., 2006. Periodontal disease and adverse pregnancy outcomes: a systematic review. *BJOG: An International Journal of Obstetrics and Gynaecology*, pp.135–143.

Yamashita, Y. & Takeshita, T., 2017. The oral microbiome and human health. *Journal of Oral Science*, 59(2), pp.201–206.

Yang, B., Wang, Y. & Qian, P., 2016. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics*, pp.1–8. Available at: http://dx.doi.org/10.1186/s12859-016-0992-y.

Yokogawa, K. et al., 1974. Mutanolysin, Bacteriolytic Agent for Cariogenic Streptococci: Partial Purification and Properties. *Antimicrobial Agents and Chemotherapy*, 6(2), pp.156–165.

Yuan, S. et al., 2012. Evaluation of Methods for the Extraction and Purification of DNA from the Human Microbiome. *PLoS ONE*, 7(3).

Zarate, G. & Nader-Macias, M.E., 2006. Influence of probiotic vaginal lactobacilli on in vitro adhesion of urogenital pathogens to vaginal epithelial cells. *Letters in Apllied Microbiology*, 43, pp.174–180.

Zarco, M.F., Vess, T.J. & Ginsburg, G.S., 2012. The oral microbiome in health and disease and the potential impact on personalized dental medicine. *Oral Diseases*, 18, pp.109–120.

Zaura, E. et al., 2014. Acquiring and maintaining a normal oral microbiome : current perspective. *Frontier ins Cellular and Infection Microbiology*, 4(June), pp.1–8.

Zaura, E. et al., 2009. Defining the healthy “core microbiome” of oral microbial communities. *BMC Microbiology*, 9(1471–2180 (Electronic)), p.259.

Zbinden, A. et al., 2014. Frequent detection of Streptococcus tigurinus in the human oral microbial flora by a specific 16S rRNA gene real-time TaqMan PCR. *BMC Microbiology*, 14(231), pp.1–7.

Zbinden, A. et al., 2012. Streptococcus tigurinus , a Novel Member of the Streptococcus mitis Group , Causes Invasive Infections. , 50(9), pp.2969–2973.

Zbinden, A., Bostanci, N. & Belibasakis, G.N., 2015. The novel species Streptococcus tigurinus and its association with oral infection. *Virulence*, 6(3), pp.177–182.

Zhou, X. et al., 2011. The Vaginal Bacterial Communities of Japanese Women Resemble Those of Women in Other Racial Groups. *FEMS Immunol Med Microbiol*, 58(2), pp.1–19.