Characterising interactions between bacteria in human vaginal microbiomes

Emily Fotopoulou

MSc by Research

University of York

Biology

July 2017

ABSTRACT

The human body is colonised by an immense number of microbial organisms inhabiting various tissues and body sites and although most microbiomes are beneficial for the host, environmental disturbances can lead to negative clinical consequences. Microenvironment disruption has been linked with various disorders in the vaginal tissue including Bacterial Vaginosis, HIV and other Sexually Transmitted Infections. Microbiome studies have proven a useful tool in characterising microorganisms associated with health and disease in humans. Amplicon data can provide information on the relationship between bacterial community composition and ecosystem function. This study aimed to identify correlations between members of the vaginal microbiomes from different individuals with gynaecological disorders, to gain insight into the microbial interactions that affect community assembly. Although positive and negative correlations between bacterial taxa may give us insight to bacterial relationships, they can be enhanced by exploring the metabolic properties of these taxa. A pipeline was designed here to allow cultivation-free, bioinformatics analysis on existing amplicon data from vaginal microbiome studies. QIIME (Quantitative Insights Into Microbial Ecology) and other purpose-written Python scripts were designed to complete taxonomy assignment, diversity and clustering analysis, as well as to assess the statistical significance of the correlations from the interactions observed. Analysis suggests strong correlations between various anaerobes, linked with dysbiosis in bacterial communities. A novel correlation between *Dialister* and *Prevotella* genera is presented, which can be reinforced by the presence of metabolic links. Succinate is a shared metabolite, that is a product of fermentation in *Prevotella* and a substrate for *Dialister* in propionate production. The findings identify links between the human microbiome and pathogenicity, thus providing insight into vaginal microbiome structure and composition, particularly so in the gynaecological syndrome of bacterial vaginosis. In conclusion, microbiome analyses studies show the prospect of new approaches to diagnosis and therapy.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I wish to express my gratitude to my supervisor Professor James Moir for his constant support and feedback, as well as to Timothy Rozday for his computational expertise.

AUTHOR’S DECLARATION

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

1. INTRODUCTION

1.1 Importance of Microbiomes

The human body is colonised by an immense number of microbial organisms cohabiting in various tissues and body sites throughout surfaces of the body. Human microbiomes play a crucial role in health and disease, whilst also being linked to nutrition, metabolism and immunity protection [1]. The number and scope of human microbiome studies has been greatly expanded over the past decade due to the technical advances in sequencing technologies [2]. Additionally, metagenomic tools have promoted microbiome study analyses, by offering insight into metabolic functionality. Computational tools are now available which can analyse microbiome community composition and functionality as a complete system, thus avoiding potential cultivation bias [3].

Due to advances in sequencing and accompanying metagenomics technologies scientists have been able to understand the importance of microbiome composition underlying function in multiple different tissues and organs throughout the body [4],[5]. Gut, oral, skin and vaginal microbiomes are colonised by distinct microbial communities, that offer a mutually beneficial system for both host and resident microbes. Microbiomes are mainly composed of bacteria, though they can also contain viruses, protozoa and fungi that play key roles in digestion and immunity defence. Gut microbiomes are composed of very diverse communities with approximately 800 microbe species [6]. Skin is the largest organ in the human body, and like the gut or the vagina, is colonised by various beneficial microorganisms comprising stable structures based on microbiome interactions [7]. D'Argenio et al. 2015 report that key organisms in the gut flora such as *Firmicutes*, *Bacteroidetes* and *Actinobacteria* assist in polysaccharide digestion [8]. This leads to production of various vitamins (such as vitamin B) which play a role in immune system development and defence against infections [9]. On the other hand, the microbes in turn flourish within their human hosts, benefiting from nutrients and an advantageous growth environment.

Microbiomes of asymptomatic healthy individuals, represent dynamic, structured bacterial communities forming symbiotic relationships and regulating various metabolic functions [10]. Symbiosis is a term describing on going relationships between organisms. More specifically mutualism refers to mutually beneficial symbiotic relationships, and in this instance is commonly related to metabolic properties [11]. Mutualistic symbiotic metabolic interactions between members of the gut microbiome as well as between the microbiome and host have been studied extensively [12]. For example, Neish et al. 2000 revealed that nonvirulent *Salmonella* strains inhibit inflammatory cytokine production in intestinal epithelia cells, via the IkB pathway, blocking further nuclear translocation of the NF-kB dimer[13], as a result, illustrating the mutualistic symbiosis between host and *Salmonella*. Xu et al. 2013 discuss bacteria-bacteria symbiosis in their study on Gram-negative anaerobe *Bacteroides thetaiotaomicron* [14]. They reveal that bacteria *Bacteroides thetaiotaomicron*, *Clostridium perfringens*, *Bifidobacterium longum*, and *Escherichia coli* have the ability to utilise various polysaccharides depending on environmental availability, thus efficiently sharing environmental resources and creating a "metabolic milieu of the intestinal ecosystem" [14].

Equally, vaginal microbiomes reveal various symbiotic bacteria-bacteria metabolic relationships due to the compositional stability of the microbiome. In 1997 Pybus et al. proposed a mutualistic symbiotic metabolic relationship between *Gardnerella vaginalis* and *Prevotella bivia* [15]. They revealed that *Gardnerella vaginalis* and *Prevotella bivia* cycle ammonia and amino acids, with *P. bivia* utilising amino acids for growth and producing ammonia whilst *G. vaginalis* utilised ammonia for growth and produced amino acids [15]. *Neisseria gonorrhoeae*, a pathogenic bacterium responsible for vaginal gonorrhoea infections, is also known to display syntrophic interactions. *N. gonorrhoeae* contains a conserved genomic island, the *prp* gene cluster, which enables propionic acid utilisation as a carbon source, especially under stress conditions [16]. Propionic acid is produced by various anaerobic bacteria as an end product of fermentation. This proves useful to vaginal microbiome communities as multiple bacteria (eg. *Corynebacterium*) utilised propionic acid to generate pyruvate, a key carbon source [17] [18].

Humans experience multiple microbiome composition variation phenomena throughout their life span, from infancy to puberty, adulthood and finally to less diverse elderly microbiota [19]–[23]. However, some compositional changes driven by environmental stress (eg. pregnancy, psychological stress [24]) have been implicated with increased susceptibility to disease or infections [25]–[27]. Microbiome fluctuations have been linked to vaginosis [28], obesity [29], bowel disease [30], and even behavioural habits [31]. As Falony et al. 2015 discussed in their study, metabolites in the gut microbiome, such as trimethylamine, can turn harmful and promote atherosclerosis, although strong correlation to causation was not proven [32]. Additionally, Gosmann et al. suggests in their 2017 study, a correlation between vaginal microbiome composition and human immunodeficiency virus (HIV) susceptibility [27]. Their analysis reports links between decreased *L. crispatus* and increased anaerobes (such as *Prevotella*, *Sneathia* and others) with elevated activated genital CD4$^+$ T cells [27]. Vaginal communities with increased activated CD4$^+$ T cells were additionally associated with increased HIV susceptibility, due to their ability in expressing HIV co-receptors and thus allowing enhanced viral replication [33]. Their data were used for the purpose of this research, to approach complete microbiome analyses via a new proposed pipeline. (The study will be referred in the upcoming chapters by its Sequence Read Archive (SRA) code HIV).

1.2 Dysbiosis in Vaginal Microbiomes

The focus of this study was to identify community links between vaginal microbiomes under various dysbiotic conditions based on microbiome composition. Vaginal flora are very beneficial to their host by providing the first line of defence against infections and colonisation of pathogenic organisms. As Cribby et al. 2008 mention in their study, more than 50 unique microbial species inhabit human vaginas [6]. Asymptomatic female vaginal tracts usually consist of aerobic and anaerobic bacterial communities. The majority of healthy vaginal microbiomes consist of bacterial communities with predominantly one of four *Lactobacillus* species [28]. *L. iners*, *L. crispatus*, *L. gasseri*, *L. jenesenii* are the four most commonly present *Lactobacillus* species, involved in key processes maintaining a balanced microbiome environment [34]. *Lactobacilli* maintain a low pH (4-4.5) by producing lactic acid, an antimicrobial compound in its own right, thus creating an unsuitable environment for various pathogenic organisms and preventing colonisation [28], [35]–[37]. Vaginal *Lactobacilli* are also involved in hydrogen peroxide production, generating an additional barrier to pathogenic colonisation [38]. However, multiple healthy asymptomatic women, present low *Lactobacillus* abundance microbiome communities. Interestingly, the microbiome will balance lactic-acid production by replacing *Lactobacilli* with other lactic acid–producing bacteria such as *Atopobium vaginae*, *Megasphaera*, and *Leptotrichia* [39]. In conclusion, female health is significantly regulated by the vaginal microbiome. Reduction in *Lactobacillus* abundance has been linked to various vaginal syndromes, most commonly Bacterial Vaginosis (BV) and HIV [40].

Females are commonly diagnosed with vaginal inflammation syndromes (non-specific vaginitis), however treatment for vaginitis can prove challenging as various conditions can be responsible for its cause [41]. Most common infections are caused by yeast or bacteria colonisation, due to hormonal changes, medical prescriptions or even sexually transmitted diseases [41]. Bacterial vaginosis (BV), Chlamydia, Genital herpes and Gonorrhoea are some of the most common causes of non-specific vaginitis [42]. All disorders include unique symptoms however the most common are discharge, odour and irritation. Unfortunately, the symptom similarities between disorders causes complications with diagnosis, as women assume yeast infections (which are commonly self-treated). The misconception is common in BV [43] and Gonorrhoea patients [44], thus increasing the percentage of falsely medicated cases. Self-medicating or antibiotic over-prescription can lead to increased antibiotic resistant bacterial strains, as observed in the latest years with *Neisseria gonorrhoeae* (the bacteria responsible for bacterial gonorrhoea) [45].

Disturbance of the vaginal microbiota's ecosystem can result in moderate infections or severe vaginal conditions. Bacterial vaginosis (BV) is a common clinical syndrome resulting from disruption of the environmental equilibria in the vaginal microbiome [46]. BV is an "alert state" of the vaginal microbiome usually characterised by the loss of *lactobacilli* and an increase in anaerobes and Gram-negative bacteria [47]. *Gardnerella spp* have also been reportedly high in BV patients [48]. BV state is characterised by a number of typical symptoms, including topical irritation, increased pH and thin grey-white vaginal discharge, however the reasons leading to this environmental disruption are not fully understood [49]. Despite the common occurrence, even though common, BV is strongly linked with gynaecologic implications such as an increased chance to acquire STI infections and HIV [40], [50]. Saxena et al. 2012 discuss that the association between BV and HIV susceptibility could be a result of mucosal permeability defects, driven by BV [51]. BV has also been linked with pregnancy miscarriages as well as an increased rate of premature labour [52], [53]. BV has severe implications to female health, accordingly it is crucial to investigate BV microbiomes to better understand the relationships between bacteria, host and the causation of bacteria imbalance.

While all vaginitis conditions have major implications on microbiome composition and patient health, the correlation of microbiome instability to HIV susceptibility is concerning [54]. As Zhou et al. 2007 mention in their study, more than 90% of HIV infections originate from heterosexual intercourse with a 2-4-fold increase ratio if females were BV carriers [55]. HIV is a lentivirus, which over time can be responsible for Acquired Immunodeficiency Syndrome (AIDS), a severe immune failure disorder. Due to hindered immune system, women infected with HIV present microbiomes with higher species richness [27]. Spear et al. 2010 results demonstrated that *Lactobacillus iners* were significantly less present (1.3-fold difference) in HIV-positive women, compared to HIV-negative patients [56]. Additionally, Hummelen et al. 2010 focused on sequencing Tanzanian women's microbiomes and revealed a strong increase of *Clostridiales* order level taxonomies (with *Prevotella bivia* dominating most patients), in co-infected HIV and BV individuals [57]. Equally, Ravel et al. 2011 report *Prevotella* as the most abundant genus taxonomy, in low *Lactobacillus* samples

[28]. *Prevotella* appears to be significantly associated with decreased presence of *Lactobacilli*, either as a result of HIV or BV microbiome imbalance. Interestingly, *Prevotella* is a pathogen responsible for aspiration pneumonia, lung abscess and other respiratory tract infections, due to its ability in invading epithelial cells and thus triggering inflammatory response [58]. Consequently, there is a large need to study vaginal microbiomes and understand the interactions of their members to identify the causes of the changes that can turn its environment toxic to its host.

1.2.1 Definition of dysbiosis

As discussed, disturbance of microbiome communities due to external triggers can result in major health implications. However, fluctuations in community structures do not always lead to increased vulnerability. All humans undergo multiple microbiome community fluctuations during their life span causing temporary microbiome instability. Vaginal microbiomes consisting of atypical microbial communities can be characterised as dysbiotic. Dysbiotic is an ambiguous term, as various studies provide different definitions. It is commonly expressed as the presence of microbial imbalance. Tamboli et al. 2004 define dysbiosis as an imbalance of "healthy" vs "harmful" bacteria in the intestinal microbiome [59]. Dysbiosis is commonly related to damaging, susceptible or diseased microbiomes, [60]–[62] and not focused on microbiome composition and structure. However, for the purpose of this study the term "dysbiotic" will signify the lack of a "common" asymptomatic vaginal microbiome community. In particular, a dysbiotic vaginal sample will describe a patient's lack of, or irregularly low abundance of, *Lactobacilli* or the presence of uncommon bacterial community structures. Dysbiosis will not be associated with patient's medical status or condition. Therefore, asymptomatic healthy patients with atypical vaginal microbiomes, as observed in Ma et al. 2012 study - where healthy vaginal microbiomes not dominated by *Lactobacilli* were detected, will be described as dysbiotic [36]. In conclusion for the purpose of this study dysbiosis will not be synonymous to symptomatic, vulnerability or diseased but simply represent atypical vaginal microbiome communities.

1.3 Metagenomics

Human microbiome studies have been proven to be a very useful tool in identifying and characterising microorganisms associated with health and disease in humans. Microbiome studies via metagenomic analysis can provide information on the relationship between bacterial community composition and the ecosystem function in human tissues colonised by microorganisms. Prior to high throughput sequencing technologies, microbiome studies required individually cultured community members in order to investigate associations and community structure [63]. Culturing methodologies are optimised for a small number of well characterised organisms, thus atypical bacteria would cause complications in analysis. Additionally, culture-dependent techniques can create bias; depending on culturing conditions, the easily cultured bacteria will be overrepresented. Consequently, culture based microbial analysis present limitations.[64], [65]. Fortunately, high-throughput sequencing provides an efficient approach to investigate members of microbial communities by analysing DNA samples directly from the source [66].

Metagenomics is a term used to define both research techniques (commonly associated with amplicon data and computational analyses), and a research field studying genomic material from uncultured microbial populations [67]. Metagenomics allows investigation of microbiomes aiming to gain insight on microbial behaviours and environmental interactions at a genomic level [68]. A number of studies have investigated vaginal microbial communities via metagenomic analysis [16], [40], [47], [69], providing a large numbers of amplicon data to investigate bacterial interactions. Typically, metagenomic analysis of microbiome bacterial communities is implemented on multiple samples of varied composition. This would allow comparison between microbiomes and distinction between microbial communities in response to function. For example, study [70], [71] collected samples from various areas of the human digestive tract, comparing microbiome composition between gut and faecal microbiomes.

Metagenomics sequencing techniques can produce hundreds of thousands of reads, depending on the size and sample properties of the experiment. The large numbers of reads assist with accuracy and provide better assessment of microbiome composition while avoiding cultural bias [72]. The sequencing reads can be utilised to perform diversity and correlation analysis to understand microbiome community composition thus gain a better understanding of function. Various computational tools such as Quantitative Insights Into Microbial Ecology (QIIME), have been developed allowing sample assignment, operational taxonomic unit (OTU) picking and taxonomic identification [73]. Using these tools on data sets originating from multiple published studies focused on various dysbiotic microbiomes could allow insight into potential links between dysbiotic vaginal environments.

16S ribosomal RNA data has been one of the preferred methods for sequencing studies, when characterising members of a microbiome since the late 1980's, and particularly since the invention of PCR [74]. 16S rRNA data have proven beneficial as they contain conserved regions, as well as variable regions helping with taxonomy assignment [75]. 16S rRNA are part of ribosomes, which are ubiquitous and have had conserved structural and functional properties over the course of evolution, thus allowing direct organism identification and assessment of phylogenetic relatedness [76]. However, multiple studies have criticised 16S rRNA ability to identify taxonomies to species level. Multiple studies such as Becker et al. 2004 have attempted new approaches to 16S rRNA analysis to overcome species level phylogenic disadvantages [77]. Ribosomal RNA identification is most commonly used in bacterial metagenomics analysis due to the well-established reference databases assigning taxonomy to 16S sequences [64], [78], [79].

Cultivation-free analysis with bioinformatics permits fast and deep understanding of microbiome composition and structure as well as the effect of environmental changes. As mentioned above, vaginal flora undergo multiple disturbance events from pregnancy to medication prescriptions during women's life span [36]. It has been suggested that stability of the microbiome, if exposed to environmental stress is dependent on microbiome composition [80]. It is therefore interesting to investigate whether short term conservation changes of the microbiome, have an effect on its ability to fight infection. For that reason, various dysbiotic vaginal microbiomes will be analysed. Data originating from existing microbial diversity studies [27], [81]–[84] may allow insight into any potential links between composition and dysbiosis thus gaining a clearer understanding of medical disorders and their associated bacterial interactions.

1.4 Aims

Focusing on the impact of dysbiosis on vaginal microbiomes, five studies were selected to carry out a computational analysis. This would draw out information on potential links between dysbiotic vaginal microbiomes and the interactions between members of those microbiomes. Aiming to identify associations between specific members of the microbiome, a new approach to microbiome analysis is suggested. The study hypothesises strong correlations between specific bacteria constituting the microbiome driven by complementary metabolic traits. These will be tested via various bioinformatics software and tools. Bioinformatic tools such as Quantitative Insights Into Microbial Ecology (QIIME) and Clustered Image Maps (CIM)miner were utilised to perform OTU and taxonomy assignment, thus identifying microbiome composition; as well as diversity and clustering analyses, thus investigating microbiome interactions. Various python programs were developed to carry out additional clustering assessments (illustrated in heatmaps and dendrograms) as well as statistical tests to confirm the significance of inter-species correlations. Various previous studies have focused on correlations between microbiome composition and environmental state, whereas this project focuses on identifying strong correlations between specific organisms dependant on microbiome community structure. In conclusion, this study aims to provide insight to potential associations between specific bacteria members of a microbiome and express the likelihood of metabolic relationships being the driving force of these correlations.

2. METHODS AND METHOD DEVELOPMENT

The recent “hype” in microbiome studies, paired with the advanced feasibility of genome-wide sequencing has resulted in a diverse database of sequenced microbiomes. This provides the materials to assess human microbiome communities by computational methods [85]. Bioinformatics allow rapid assimilation of a vast number of sequences thus permitting investigation of a microbiome’s diversity, structure, composition and even ecosystem in a reduced-resource and high-throughput manner.

One of the aims of this project was the design of a new pipeline to manage 16S rRNA data for metagenomic analysis. The lack of a universal amplicon pipeline steered the optimisation of the methodology to carry out whole microbiome community assessment. The design of this methodology aspired to investigate interpersonal microbiome variation, bacteria and environmental interaction analyses, on any given amplicon dataset. Most currently available methodologies specialise on certain statistical or correlation tests rather than profiling complete microbiomes and their interactions [35], [76], [86]–[88].

Here a pipeline was developed to perform data acquisition and reformatting on 16S rRNA human vaginal microbiome samples collected from former microbiome studies. Aiming to study vaginal community structures under different pathology states, the pipeline contains Operational Taxonomic Unit (OTU) and taxonomy assignments, allowing profiling of the microbiome. Diversity analyses in the pipeline reveal microbiome community structures and correlations within its members as well as in-between the ecosystems. Additionally, executing statistical tests such as Shapiro and Wilks, Spearman’s rank correlation coefficient and computing principal component analysis (PCA) provides a statistical significance on the observed associations. Clustering analysis, dendrograms and heatmaps are more tools added to the pipeline, aiding in visualising interactions within and in-between the microbiomes.

Most of the pipeline steps were implemented through Python programming (see Appendices). However key tools such as Quantitative Insights Into Microbial Ecology (QIIME) and CIM miner online tools were used for data filtering, microbiome classification, heatmap generations and investigation of key correlation links.

Further development of the pipeline dedicated on metabolic interactions and associations within the microbiome would allow a complete illustration of a microbiome’s contribution to health and disease at an intrapersonal level. Completion of this pipeline could provide an improved and faster way for medical diagnostics using bioinformatics as an assessment tool.

2.1 Data selection and acquisition

Metagenomic analysis on microbiome studies is a very common approach with multiple applications in research. For this study, amplicon data produced through high throughput sequencing methods (454 or Illumina), were selected to investigate human vaginal microbiomes under various dysbiotic conditions. 16S rRNA data have proven very useful in microbiome assessment due to their high success in profiling complex microbiome communities [89]. 16S rRNA allows microbiome profiling, particularly for low abundance species, with deep sequencing depth [90]. As mentioned previously, healthy vaginal microbiomes vary in diversity however most are dominated by key bacteria such as the *Lactobacillus* species. As this research focused on dysbiotic vaginal microbiomes, it was essential that the data included an accurate representation of low abundance species. Therefore, 16S rRNA sequences were the amplicon data chosen for this study in order to observe variations between medical syndromes driven by microbiome interactions.

Sequences were collected from previous metagenomic studies containing vaginal samples; either tissue or swab samples. Samples were accessed from the Sequence Read Archives (SRA) database, a function of NCBI's (National Centre for Biotechnology Information) database. Each sequence file contains sequences from a single tissue sample. Tissue samples do not necessarily represent an individual in a study. Typically patients were required to provide multiple samples during the course of a study, to review intrapersonal variations over time [91]. Unfortunately, the true "identity" of the samples was not always available, due to annotation issues and will for the purpose of this study be perceived as independent samples.

It is important to mention that the final dataset assembled for this study (see Table 1) went through a series of quantitative and qualitative edits. The details of data acquisition and sorting are discussed to present possible means of approaching computational setbacks and challenges faced. The initial search for sequences was performed through utilising NCBI's database through the search box, typing "vaginal microbiome". This search (01/02/2016-18/02/2016) listed 6928 experiments with approximately 41000 SRR (SRA Run Brower) accession codes, each containing sequence and technique details of a single run. Aiming to create a diverse sample size database containing various 16S rRNA human vaginal samples, all 41000 SRR's were included in the initial dataset. The attempted download of 41000 SRR was performed through the "SRA Toolkit" via "prefetch" command (suitable for Windows operating systems) in command prompt. NCBI offers SRA Toolkit, a collection of tools and libraries available for Sequence Read Archive data analysis.

```
>> prefetch [options] <path/SRA file | path/kart file> [<path/file> ...]
```

"Prefetch" calls the SRA accession number corresponding to a single study run and generates a .txt file containing the sequences of that SRR file (see Appendix 1). Following this step, "fastq-dump" was employed to reformat the "SRR*.txt" file to a more computationally friendly .fastq file. "fastq-dump" is an additional SRA Toolkit utility which was again implemented in command prompt (command listed below). Fastq files are commonly used text file formats, containing sequence reads and qualitative information of the sequence reads, thus assisting with sequence display for bioinformatics analyses.

```
>> fastq-dump [options] <path/file> [<path/file> ...]
```

However, the immense size of the database appeared to decrease the speed of the download. As this process proved too computationally heavy for a Windows operating system, Ubuntu Linux operating system was installed. Linux provides a safer, more efficient system when programming, with great advantages in memory management allowing faster processing in comparison to Windows. "fastq-dump" command works differently in a Linux operating system. Unlike in Windows, "fastq-dump" in Linux does not need a preceding step in order to download SRR sequences. It downloads and directly converts and stores the files into a .fastq or .fasta file.

```
>> fastq-dump -X 5 -Z SRR390728
```
(see more in Appendix 2)

During the initial database selection step, non-human samples were detected; as the database consisted of every study listed under "vaginal microbiome". Removal of the non-human vaginal samples resulted in a new dataset of 7963 SRR's. Download for the second dataset was performed through the fastq-dump command in Linux terminal. However, a number of accession codes were flagged thus preventing their download. Upon examination of the returned SRA codes (550 SRA false files), some codes corresponded to blank SRR sequence files, thus were deleted; whereas others had to be downloaded individually via "prefetch". The precise reasoning behind troubleshooting error-flagged files via prefetch are not fully understood, however it is speculated that file format issues caused download blocking.

Upon completion of all downloads, data reformatting followed. Samples downloaded through fastq-dump, thus stored in a .fastq file format, were then converted into .fasta files via seqtk command.

```
>> seqtk seq -a input.fq > output.fa
```

The files downloaded via prefetch were stored as .sra files, thus had a step preceding seqtk. The .sra files were initially formatted into .fastq files (via fastq-dump), followed by seqtk conversion (into .fasta files). Fasta files are very similar to fastq files; however, fasta files lack quality data for each sequence run. Fasta files permit straightforward data analysis and sequence visualisation, due to their smaller file size. Sequence manipulation is also possible in fastq files, however fastq increased file size hinders processing power and thus sequence visualisation in a text editor. For the first version of the pipeline proposed, converting fastq files into fasta proved obsolete, as the fasta files remained too large to allow any visualisation advantages. Additionally, the QIIME tools and scripts employed to approach microbiome studies exhibited compatibility with fastq files.

Manipulation of the second selected study dataset, proved its impractical size. The second study dataset contained 20 human vaginal microbiome studies with collectively 7963 SRA samples. The database might have permitted extensive in depth sampling, but would also present an ambiguous study with massive time restrictions. For that reason, the selection of studies was further filtered through a number of additional criteria. The studies selected had to contain human vaginal amplicon data (16S rRNA), sequenced through high throughput sequencing techniques; such as Illumina and 454. Moreover, the number of SRA samples was taken into consideration as broader diverse sampling studies were preferred (>50 SRAs) for the purpose of this research. At that stage, the focus was turned on the presence of primers for each study with the hopes of identifying enough data with matching primers to allow analysis of multiple studies under the same pipeline. It soon became apparent that most studies did not contain identical primer reads, so these criteria were excluded for study sorting. Implementing these criteria reduced the dataset to 18 studies (with 3119 SRA samples in total). However, the number of studies remained excessive; thus an additional criteria based on vaginal microbiome condition, was considered for study selection. Studies focused on dysbiotic or diseased microbiomes were favoured; thus creating the finalised dataset; including 7 studies (1927 SRR's) on HIV, candiditis, herpes and Bacterial Vaginosis vaginal microenvironments (refer to Table1).

Study Description	Designated Abbreviations	SRA Project Accession number	Amplicon	Sequencing Method	Number of runs	Human Vaginal Samples
Certain species of vaginal bacteria can increase a woman's susceptibility to HIV	HIV	ERP017263	Yes	Illumina	168	Yes
Diverse vaginal microbiomes in reproductive-age women with vulvovaginal candidiasis	CANDIDIASIS	ERP003902	Yes	Illumina	223	Yes
Complementary seminovaginal microbiome in couples	SV	ERP009682	Yes	Illumina	69	Yes
Characterization of the Vaginal Microbiota among Sexual Risk Behavior Groups of Women with Bacterial Vaginosis	BV	SRP045868	Yes	454	112	Yes
Distinct effects of the cervico-vaginal microbiota and herpes simplex type 2 infection on female genital tract immunology	HSV2	SRP071021	Yes	Illumina	51	Yes
Vaginal microbiome of reproductive-age women *	PRJNA329618	SRP090242	Yes	Illumina	366	Yes
Endometrial cancer microbiome *	PRJNA295859	SRP064295	Yes	Illumina	238	Yes

Table 1: Summary table of selected studies utilised for data analysis. A table summarising the 7 studies composing the finalised dataset with a total of 1927 sequence files. All studies listed were applied to the pipeline presented, in order to investigate vaginal microbiomes under various microbiome disorders. *Studies were not included in the final analysis due to sequencing file errors (see Appendix 3 for details.) For Experiment accession codes refer to Appendix 4)

Once the finalised dataset of 7 studies was established (Table 1), aiming to maximise time efficiency when handling 16S rRNA data, a new method of downloading sequences was established. EMBL-EBI (European Molecular Biology Laboratory – European Bioinformatics Institute) offers a direct and simplified approach to downloading amplicon data[1]. Accessing a study's sequence files can be achieved by sourcing EBI's ENA (The European Nucleotide Archive) web based function and quoting an SRA Project Accession number or Experiment Accession Number (attained from NCBI's database) in the ”Text Search” box. ENA offers bulk download of all files enclosed within any study. Launch of this application can only be accomplished through Java software, consequently installation might be required. ENA loads a new window with details of the downloading files, as well as offers the choice of selecting a specific download directory for the files to be saved in. Downloading time differs depending on the number of runs contained within each study, however ENA includes a download status bar permitting live monitoring. In comparison to the methodology presented previously, it is safe to state that EBI proposes the simplest method, with limited steps, to approach 16S rRNA sequence download.

ENA's download generates numerous fastq files originating from an individual study. Once all fastq files were acquired, conversion to fasta files was essential for the following step to commence. QIIME's split_libraries.py command performs data de-multiplexing, a crucial step for microbiome amplicon analysis, which requires a fasta (.fna) file (details of this script will be discussed in detail in the following sections). Unlike the approach followed previously, fasta reformatting was accomplished through a QIIME command. Convert_fastaqual_fastq.py script assists in generating two files per fastq file; a fasta file containing all sequence runs and their IDs in a text file format (stored as .fna file) and a qual file containing quality scores for each sequence run.

```
>> Convert_fastaqual_fastq.py (see Appendix 5)
```

In conclusion it is important to state that the finalised version of the pipeline proposed here for microbiome research, includes several different techniques for data acquisition and reformatting.

2.2 QIIME toolkit

QIIME is a powerful open–source pipeline that allows several amplicon microbiome data analyses from taxonomy assignment to statistics and diversity analyses. It offers a number of python scripts running in Unix shells with multiple modifiable parameters to match any study's focus or requirements. For the purpose of this study de-multiplexing, OTU and taxonomy assignment, were employed through QIIME scrips followed by investigation of microbiome composition variation and evaluation of alpha and beta diversity in microbiomes.

[1] http://www.ebi.ac.uk/ena

Installation of QIIME proved complex as Windows operating systems require installation of a VirtualBox (VB). The Oracle VirtualBox version 5.0.26 was downloaded, running a virtual Ubuntu based system, containing QIIME 1.9.1 with pre-installed dependencies and scripts. All QIIME and python scripts were developed in IPython Jupyter Notebook; an interactive shell tool supporting data visualisation and providing access to GUI toolkits. Installation of IPython Jupyter Notebook is very suitably bundled within Anaconda, a package and environment manager that installs Python (for this study python 2.7 was installed) and other analytical scientific packages [2].

Due to QIIME's computationally expensive scripts and large 16S rRNA file sizes, a standard desktop computer faced memory and size restrictions, thus hindering command completion. Struggling to complete such tasks in a timely manner, the University of York's computer cluster (YARCC - York Advanced Research Computing Cluster) was installed and a personal server filesystem was set to run most QIIME scripts. Although a computationally challenging task, as python libraries and QIIME scripts had to be installed individually on a personal file system, this proved beneficial when working with large files. This caused a significant processing speed increase, thus reducing duration of each task to less than half the time previously observed in the VB.

It is important to emphasise that two distinct QIIME pipelines were employed to design an optimal methodology approaching microbial community analysis (see Figure 1 a,b). Even though both pipelines performed the same tests and investigations, they vary in stages and scopes, proving the importance of exploring different methodologies at various stages during a research project. In this section the steps followed to design and employ an amplicon analysis on 16S rRNA vaginal microbiomes sequences via QIIME will be described in detail.

The following diagram illustrates all the main stages, followed to create two QIIME pipelines which will be discussed in detail in this section (Figure 1).

[2] https://docs.continuum.io/.

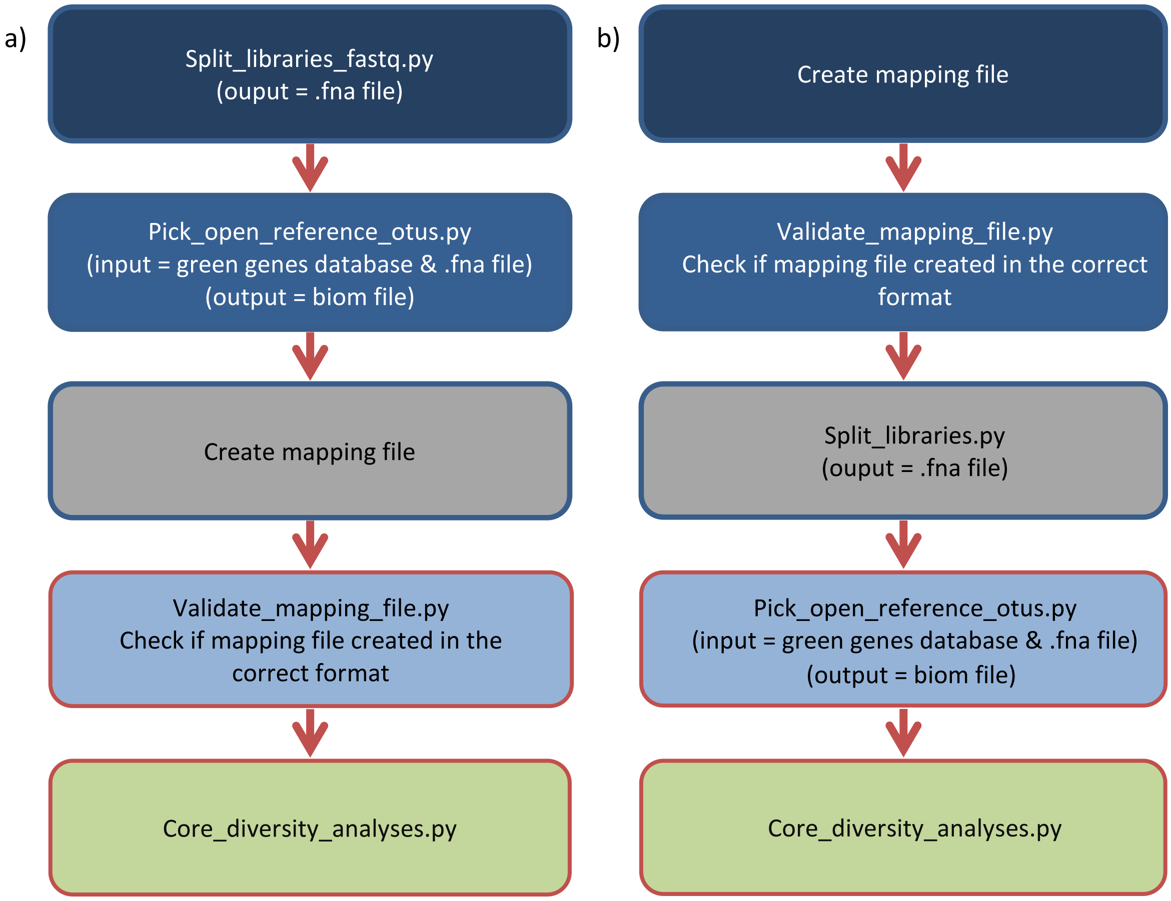

a)
Split_libraries_fastq.py
(ouput = .fna file)
Pick_open_reference_otus.py
(input = green genes database & .fna file)
(output = biom file)
Create mapping file
Validate_mapping_file.py
Check if mapping file created in the correct format
Core_diversity_analyses.py
b)
Create mapping file
Validate_mapping_file.py
Check if mapping file created in the correct format
Split_libraries.py
(ouput = .fna file)
Pick_open_reference_otus.py
(input = green genes database & .fna file)
(output = biom file)
Core_diversity_analyses.py

Figure 1: QIIME pipelines followed for microbiome analyses. Figure 1 displays a step to step illustration of two the pipelines employed through QIIME, to investigate diversity within and in-between human vaginal microbiomes. Figure 1a demonstrates a diagram of the initial pipeline, pipeline one, designed to carry out diversity analyses in QIIME. Figure 1b represents the improved and optimised version of the QIIME pipeline, pipeline two, applied on 16S rRNA data.

2.2.1 Reformatting and de-multiplexing SRA data

Regardless of a researcher's aimed end product, the first step in most QIIME pipelines is consistent. De-multiplexing of 16S rRNA sequence reads is the primary stage of any analysis in QIIME as it converts the raw data into a functioning format for QIIME to use [92]. Quality filtering and reformatting are crucial steps that assure successful results through QIIME.

Therefore, the first task was to assign multiplexed 16S rRNA reads to groups based on their nucleotide barcode. Split_libraries.py is a QIIME command which performs quality filtering based on the quality features of each sequence by removing poor or ambiguous reads. Split_libraries.py additionally executes quality control by introducing thresholds on sequence lengths, end-trimmings and on minimum quality scores. De-multiplexing, reformatting and concatenating millions of sequence reads from an individual study, results in a computationally heavy process. Completion of split_libraries.py script was possible through a virtual box, even for an abundantly sampled study (tested on study SRP062720 with a total of 511 sample runs; run for approximately 48hours). However due to the long duration of the process, a cluster computer system was the preferred method. Subsequently, for all studies selected, split_libraries.py script was exclusively performed in YARCC reducing the time of the run (20-40minutes depending on SRA file sizes).

To perform split_libraries.py a mapping file along with the fasta file names of a single study, were required as input files (see Appendix 6a and 6b for full scripts). The mapping file contained information used for the sequence groupings, in order to execute effective de-multiplexing. The mapping file assists with assignment of unique barcodes, allowing parallel analysis and facilitates arrangement into sample groups (refer to Appendix 7 for format of mapping files). split_libraries.py –m and –f arguments instruct the input mapping and 16S rRNA sequence fna files, respectively. All output files can be directed to a directory or folder by –o argument.

```
>>      python      /<absolute      path>      /split_libraries.py      -m
mapping_tableHIV_corrected.txt                                           -f
ERR1679496_1_barcoded_linkedPrimer.fna,ERR1679497_1_barcoded_linkedPrimer
.fna,ERR1679498_1_barcoded_linkedPrimer.fna,………..   -o   /<absolute   path>
/<output folder>/
```

QIIME's explicit requirements on the format of a mapping file are not to be overlooked. Any compatibility errors with the mapping file will cause major malfunctions with de-multiplexing. For that reason, QIIME provides validate_mapping_file.py; a script that will ensure the file's contents and format. If any errors are detected, QIIME will create a log file stating the faults. Once corrected, the mapping file can be applied to the split_libraries.py script.

Unfortunately, all 7 studies selected did not include barcodes in their sample sequences. Thus a python script was composed; generating unique randomised barcodes 12base pair long, to be assigned and added on all sequence runs of a study (see Appendix 8 for python script). Upon testing, the need for Linker Primers became apparent as split_libraries.py requires them for assortment during de-multiplexing. A randomised Linker Primer sequence was designed (ATGCTGCCTCCCGTAGGAGT) and added to both fasta and mapping files, to be employed in the split_libraries.py. QIIME script. This study proposed de-multiplexing of all sequence runs depending on nucleotide barcodes and sample IDs. As both barcodes and Linker Primer sequences were only designed to assist de-multiplexing and had no influence in biological organism identification, their features were not significant for the purpose of this study. Therefore, Linker Primers remained in the pipeline exclusively for formatting purposes, as QIIME requires Linker Primers to complete de-multiplexing. Subsequently all SRA files were modified so that each 16S rRNA sequence run contained a unique barcode followed by a consistent Linker Primer (ATGCTGCCTCCCGTAGGAGT) and finally the sample sequence (Figure 2). Modified sequences were saved as a fasta file, with the `_barcoded_linkedPrimer.fna` extension ID to be later applied in split libraries script.

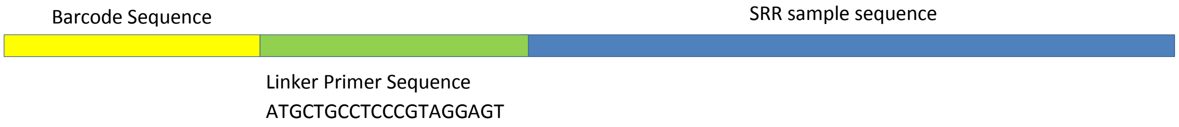


Figure 2: Remodified SRA sequence runs. Figure 2 represents the final format of all fasta sequences before employed for the QIIME scripts. A python script modified all SRR sequences to contain unique barcodes (12 base pairs) and identical Linker Primer sequences (ATGCTGCCTCCCGTAGGAGT). This format that would allow successful de-multiplexing though command split_libraries.py in QIIME.

Completion of split_libraries generates an output of three new files; histograms.txt, seqs.fna and split_library_log.txt. Files histograms.txt and split_library_log.txt contain information and the specifics of the split command, whereas seqs.fna is substantially large fasta file containing concatenated and reformatted sequences from a single study. Seqs.fna contained high quality reads assigned to unique barcodes and clustered accordingly. Success of this step results in completed de-multiplexed 16S rRNA data.

During the course of this research QIIME released a new update (07/11/2016 – QIIME 2), during which a number of optional parameters for the QIIME commands turned compulsory. Split_libraries_fastq.py script allows sequence de-multiplexing without the need of a mapping file containing details for the groupings. If no barcodes were passed through this command (optional parameter: --barcode_type 'not-barcoded') the split would depend on the sample IDs passed (see Appendix 6b). The script would again output a seq.fna file, concatenating all sequence reads of one study. Each sequence run contained a unique sample ID and an origin SRA accession number, thus de-multiplexing the original 16S rRNA sequences. This approach was utilised for the design of the first pipeline, pipeline one (Figure 1a), to study microbiome community interactions. However, the methodology described previously via QIIME's split_libraries.py, allows extra specification and more thorough control of sequence assessment into biological groups. Subsequently, the optimised pipeline two is a superior methodology to approach 16S rRNA de-multiplexing and is therefore recommended for future 16S rRNA microbiome studies.

The script bellow illustrates the split_libraries_fastq.py QIIME command utilised in pipeline one, to perform de-multiplexing based on sample IDs with a "dummy" mapping file. All QIIME scripts run in ipython, included ! generating a bash subprocess shell internally. –i argument instructs the input fastq files and -o generates an output directory. Optional parameters such as --sample_ids define an alias ID for all sample sequences and --barcode_type express the presence/ or absence of barcodes sequences within the sequence files. Additionally --phred_offset parameter, controls substitution errors.

```
>> !split_libraries_fastq.py -i
SRR1823471.fastq,SRR1823472.fastq,SRR1823473.fastq,…  --sample_ids
SRR1,SRR2,SRR3,… -o /<absolute path>/split_libraries_fastq_output --
barcode_type 'not-barcoded' --phred_offset 33
```

2.2.2 OTU picking and Taxonomy assessment

The first aim of our analysis was assigning taxonomies to the 16S rRNA samples. This was achieved by using the pick_open_reference_otus.py QIIME command. QIIME's default algorithm for OTU assessment is uclust OTU clustering tool. Pick_open_reference_otus.py is a complex script consisting of four stages resulting in OTU assessment.

The script commences with close-reference OTU picking, where sample sequences get clustered against a reference sequence database. For the purpose of this research, GreenGenes' 2010 database was downloaded as a reference database. Although more recently updated databases are available, the present study is focused on vaginal microbiota, consisting of thoroughly characterised and established organisms. Therefore, the GreenGenes reference dataset can provide sufficient data coverage.

Closed reference OTU picking generates two key files; containing the identified sequences and the unmatched sequence reads. The unmatched sequences are further progressed through de novo clustering. De novo OTU picking performs sequence clustering by matching sequences against each other with no additional reference dataset. Due to its computationally heavy methodology, only the sequences that failed to be assigned get assessed through it. De novo OTU picking forms sequence clusters where each cluster centroid is subsequently used as a “new reference sequence”. Stage 3 follows up with closed reference OTU picking, where unidentified reads are clustered against the “new reference sequences” (created through de novo picking). Any remaining unidentified sequences advance to the final stage of OTU picking through an additional de novo OTU picking process. The small number of remaining unidentified sequences makes de novo OTU assessment computationally feasible. Once every OTU assignment stage has been completed, pick_open_reference_otus.py produces a OTU mapping file containing all successful OTU samples. Finally, the OTU mapping file is later used to apply taxonomy assessment on all representative sequences of the OTU table.

QIIME offers the option of running any individual OTU picking command, included in the pick_open_reference_otus.py script, as separate commands. However, for the purpose of this study pick_open_reference_otus.py was the most appropriate command. Due to the research’s high volume of amplicon data the most time efficient test was preferred. Although de novo picking might have provided a more in depth taxonomy assignment, it would prove too computationally demanding for the present scope due the sheer number of sequences. Vaginal microbiomes consist of well characterised organisms thus an extensively in depth assessment would not have produced vastly improved results. Thus pick_open_reference_otus.py script was the appropriately chosen script for OTU assignment.

Pick_open_reference_otus.py command requires as input files; a seq.fna concatenated file (created previously via split_libraries.py command) as well as a reference sequence database and a parameters file (consult complete command on Appendix 9a). As mentioned previously, pick_open_reference_otus.py script combines multiple QIIME commands; thus a parameters file defining the preferences for each command is essential. For the purpose of this study, the parameters file contained feature details on OTU picking, taxonomy assignment and filtering quality control (see Appendix 9b). A reference sequence database was essential for the script’s completion, as during the initial stage of the analysis (closed reference OTU picking), reads get clustered against a reference database. GreenGenes 2010 database was selected for this purpose (gg_97_otus_6oct2010_aligned.fasta).

```
>> !pick_open_reference_otus.py -f -i split_librariesSV/seqs.fna -r
current_Bacteria_aligned.fa -o otusSV/ -p params.txt --
suppress_align_and_tree
```

To conclude, the key output files of pick_open_reference_otus.py script are an otu_table_mc_w_tax.biom and a rep_set_tax_assignment.txt. otu_table_mc_w_tax.biom contained information on a study's OTU assessment as well as their abundance values per sample (see Appendix 10). otu_table_mc_w_tax.biom file represented the classic format of a OTU table. On the other hand, pick_open_reference_otus.py additionally created a blast_assigned_taxonomy folder containing two files; one of which the rep_set_tax_assignment.txt file. The text file contains taxonomy assessment of the OTU samples (provided in the biom file) as well as the details of the blast search (see Appendix 11). Otu_table_mc_w_tax.biom and rep_set_tax_assignment.txt are fundamental files for the subsequent statistical stages of the suggested pipeline. These files would be further modified to allow statistical tests investigating distribution and correlation within microbiomes.

Pipeline one focused on assigning taxonomies, thus pick_open_reference_otus.py was the first script to be applied. Upon completion, the core_diversity_analyses.py script followed, where a number of diversity tests remained incomplete. The importance of de-multiplexing was recognised, as this step was identified as the cause of errors in the diversity tests. The de-multiplexing step was modified for the second pipeline and core diversity analyses were completed. High throughput sequencing methods generate multiple reads per run causing complications in tests due to the collective volume of sequences. Split_libraries_fastq.py as mentioned previously, allows libraries to be split according to their individual barcodes and thus multiple sequencing runs can be processed simultaneously and the results can be clustered according to their sample IDs.

2.2.3 Core Diversity analysis

Upon completing OTU and taxonomy assessment the next aim for the pipeline designed was to investigate diversity within and between microbiomes and their members. QIIME offers python core_diversity_analyses.py script for such analyses. It is a script consisting of an extensive workflow of assessing rarefaction, beta diversity, alpha diversity and microbiome composition.

Core_diversity_analyses.py requires inputs of a .biom file (generated during pick_open_reference_otus.py script) followed by the same mapping file created when splitting libraries as well as specifying a sampling depth (-e) (see Appendix 12a for full script). The –e value had to be no larger than the number of sequences present in the smallest sample (information enclosed in the output folder of pick_open_reference_otus.py).

For the purpose of this analysis the parameter --nonphylogenetic_diversity was passed as non-phylogenetic alpha (chao1 and observed_otus) and beta (bray_curtis) diversity calculations were preferred. Bray Curtis was the favoured test as it displays the presence or absence of dissimilarity between different sites, something really useful for comparative metagenomics study. Chao1 and observed_otus are non-parametric tests allowing accurate testing with minimal bias in large data sets [93]. They permit estimation of a communities' species richness thus assisting in investigating microbiome correlations.

```
>> python /<absolute path>/core_diversity_analyses.py -i /<absolute path>/otu_table_mc2_w_tax.biom -o /<absolute path for output folder>/ -m /<absolute path>/mapping_tableBV_corrected.txt -e 7000 --nonphylogenetic_diversity
```

Core_diversity_analyses.py script outputs a number of files assisting with visualising patterns in data but not necessarily resolving queries. Nevertheless, core_diversity_analyses.py results can assess beta diversity through a Principal Component Analysis (PCA) 3D graph, by visualising sample clustering to identify variation between the microbiomes from different individual donors (as presented herein by Figures 13-17, 25, 26); alpha diversity and species taxonomic richness via rarefaction plots and bar plots.

Interestingly, core_diversity_analyses.py analysis did not differ between pipeline one and pipeline two. Unlike with previously discussed QIIME scripts, where the parameters and format of the commands varied between the two designed pipelines, all parameters for core_diversity_analyses.py script remained identical for pipeline one and two (Appendix 12b).

2.3 BIOM file reformatting

As discussed previously, a BIOM file (or OTU table) is essential to any microbiome studies. The original format generated through pick_open_reference_otus.py QIIME's script consists of an OTU table of a complete study with all sample sequence identifiers (column data), all OTU identifiers (row data), as well as metadata for each present OTU (counts of each OTU per sample). Although an extremely useful file, the .biom format proves problematic to employ. Therefore, reformatting is essential for further analysis.

The first step was converting the .biom file into an easily handled text file. The biom convert command applied in a Linux bash shell converts a .biom file into a tab-delaminated file format allowing data visualisation and shift between sparse and dense file formats. CSV (comma separated values) was the format chosen for the biom files storing tabular data applied through the following command in terminal.

```
>> biom convert -i table.biom -o table.from_biom.csv
```

Below follows the python script designed to carry out the OTU table file reformatting. A python library was loaded in an ipython shell permitting .csv file processing. Pandas is an open source python library allowing easy data structure visualisation and processing. In this case pandas allows visualisation of a .csv file as a tab- delimited file. The first obstacle to be observed was the presence of OTU identifiers instead of more informative taxonomies, in the biom file. For that reason, the rep_set_tax_assignment text file matching OTUs to taxonomies, was utilised. The file contained the results of a blast search during pick open reference OTUs; thus listing the assigned OTU IDs with their corresponding taxonomies and unique blast e-value qualities. However, this illustrated an additional issue. Due to the similarity score parameter during the pick OTUs step (0.97), QIIME had excessively assigned OTUs. This created multiple samples under the taxonomies thus creating duplicates of the same entries. To overcome the multiple taxonomy entries, a script was created, grouping and summarising all identical taxonomy entries to a single data submission. 16S rRNA data have been extensively discussed for their difficulty with species level sequence identification, resulting from their stable gene structure and functionality not easily disturbed by time [94]. Therefore, for the purpose of this study, only the successfully genus ranked taxonomies were used. The output of the following script was an excel file containing unique genus level taxonomies as well their abundance data per sample (see Appendix 13 for complete format).

Modifying BIOM file script:

```
# loading files and python libraries
import pandas as pd
import numpy as np
file1 = pd.read_table("otu_table_mc2_w_taxHIV.from_biom.csv", header=1,
sep="\t") #important not to define index cause np tables do not work
otherwise
file2 = pd.read_excel("HIV_rep_set_tax_assignment.xlsx")

#compair both files to test if OTUs from file1 match OTUs from file2
OTUs1 = file1["#OTU ID"].values.tolist()
OTUs2= file2["OTUs"].values.tolist()

def cmp(OTUs1,OTUs2):
    for item in OTUs1:
        if item in OTUs2:
            item=item
            print ("found ") + (item)
        else:
            print ("not found") + (item)
print cmp(OTUs1, OTUs2)

#create 2D lists of the OTUs with the correlating taxonomies from the
assingment file
otus2_taxa2_LIST = list(file2[["OTUs", "Taxa"]].values)

#create dictionary with corresponding OTUs and Taxomomies
otu2_tax2_dict={}
for i, item in enumerate(otus2_taxa2_LIST):
    otus = item[0]
    taxa = item[1]
    otu2_tax2_dict[str(otus)] = taxa

# create the list of the taxomonies by including only the ones that have
genus (taxalevel=6)
data_list = list(file1.values)
taxa_level = 6
genus_data_dict = {}

#loading the responding data to the 2D list
for n, data_row in enumerate(data_list):
    otu = data_row[0]
    taxa = otu2_tax2_dict[otu] #replace
    if taxa[0] == "k":
        genus = "".join(taxa.split(';')[0:taxa_level])
        if genus[-1] == '_':
            #genus = "".join(taxa.split(';')[0:taxa_level-1])
             genus = "excluded"
        #now to match the abundance data with according taxonomies
        if genus in genus_data_dict:
            genus_data_dict[genus].append(data_row)
        else:
            genus_data_dict[genus] = [data_row]

# Condence the .biom table reformed above "genus_data_dict" into a numpy
table
final_table = {}

for name, row_list in genus_data_dict.items():
    numpy_table = np.array(row_list)
    final_table[name] = numpy_table[:,1:].sum(axis=0)

#Create New Reformed Biom file in a dataframe
SRR_samples = file1.columns[1:]
new_taxonomies = final_table.keys()
abundances = final_table.values()
new_biomfile = pd.DataFrame(abundances, index=new_taxonomies,
columns=SRR_samples)

#write biom file to excel
writer = pd.ExcelWriter("Final_HIV_biom.xlsx")
new_biomfile.to_excel(writer, sheet_name="Sheet1")
writer.save()
```

2.4 Spearman's Rank Correlation Coefficient and Statistics

Studying correlations between bacteria composing the microbiomes under various medical conditions (BV, gonorrhoea, STIs), would offer insights on potential links between microenvironments and the disorders. Hence correlation models were added to the pipeline.

The first correlation model attempted was Pearson correlation coefficient (r) which was run through a script from the scipy library module pearsonr, programmed in ipython notebook (Appendix 14). The python program utilises the previously modified OTU table containing all sample abundances and genus level taxonomic identities, to perform the correlation analysis. Prior to the Pearson correlation model, taxonomies with an abundance sum of less than 60 were excluded, as they represented insignificant rare taxa. Although not essential, it allowed clearer visualisation of correlation links, especially as most vaginal microbiomes are dominated by *lactobacilli* with extraordinary high abundance scores (e.g. study HIV listed an abundance sum of 7541501 *lactobacilli* out of a 1019104 total study abundance).

Python library scipy lists module scipy.stats.pearsonr(), calculating Pearson correlation coefficient (r) and the statistical p-values of the correlation test. The module applied on the taxonomy abundance data created a single three dimensional data structure containing both correlation and p-values. The file was stored as a numpy array to assist with data handling. Finally, the numpy array was separated into two asymmetric tables one containing the Pearson correlation coefficient data and the other enlisting the p values of the statistical test. Both tables were saved in a tab-delimited format. The Pearson correlation table consisted of asymmetrical values ranging from -1 to 1, as well as their corresponding taxonomies. Likewise, the p-values followed the same file structure covering p-values from 0 to 1.

However, upon additional testing it became apparent that the abundance data were not normally distributed (due to the large amount of zero reads). Thus Pearson correlation proved inadequate to estimate correlation within the microbiomes as it should only be used with normally distributed data. To assure that the data were not normally distributed, a Shapiro – Wilk normality test was performed. Once again, a python programme was written to perform the normality test with scipy python library's assistance (module scipy.stats.shapiro() – see Appendix 15). The test returned a value of 0.039 (anything less than 0.055 reveals non normally distributed data) indicating that the data were not normally distributed, thus a new correlation test had to be implemented.

Spearman’s rank correlation coefficient (ρ) offers non-parametric correlation coefficient testing. Spearman’s model measures the degree of similarity between two ranked variables and estimates the correlation significance between them. Unlike Pearson’s correlation model where linear relationships are tested, Spearman’s coefficient reviews monotonic relationships. An “absolute” correlation between taxa would be indicated if each variable is a monotone function of the other variable and therefore resulting in ρ values of +1 or -1. Just like with the Pearson python script discussed previously, Spearman test was carried out in an ipython shell. The output consisted again of a three dimensional numpy array, containing Spearman’s correlation coefficient along their p-values. Equally the array was split into two asymmetrical tables; one containing the ρ values between every taxa and the other consisting of the corresponding p values. The p-values represented the probability of obtaining a correlation relationship against the probability of the event occurrence. However, significance of the Spearman correlations was determined through on the ρ values for each study, gaining 95 % confidence that a correlation is true. Details of the estimated correlation significance threshold will be discussed in detail in section 2.4. The python script listed below carries out Spearman’s rank correlation coefficient test.

Spearmans’ ranked correlation coefficient analysis script:

```
#because data are not normalised i need to do a different correlation test: Spearman's rank correlation coefficient
n = biom.shape[0]

#creates a numpy table full of zeros (x,z,y) which will then be filled with the data bellow
output_table = np.zeros([n,n,2])

#.values changes Dataframe into numpy array
otutable_data = biom.values
for row1 in range(n):
    for row2 in range(row1,n):
        row = otutable_data[row1,1:]
        col = otutable_data[row2,1:]
        output_table[row1,row2,:] = scipy.stats.spearmanr(row, col)
np.save("SpearmanTableERP003902.npy", output_table)

#see numpy 2D table only with the spearman values
numpy3D = np.load("SpearmanTableERP003902.npy")
Spearm2D_spearm = pd.DataFrame(numpy3D[:,:,0])
print pearson2D_pears[:10]

#name columns and rows
b = output_sum_taxa.keys()
Spearm2D_spearm.columns = b
Spearm2D_spearm.index = b
print Spearm2D_spearm

#write spearman 2D file to excel
writer = pd.ExcelWriter("Spearman_values.xlsx")
Spearm2D_spearm.to_excel(writer, sheet_name="Sheet1")
writer.save()

#see numpy 2D table only with the spearman p-values
numpy3D = np.load("SpearmanTableERP003902.npy")
Spearm2D_Pval = pd.DataFrame(numpy3D[:,:,1])
print pearson2D_pears[:10]
```

```
#name columns and rows
b = output_sum_taxa.keys()
Spearm2D_Pval.columns = b
Spearm2D_Pval.index = b
print Spearm2D_Pval

#write pearson 2D file to excel
writer = pd.ExcelWriter("Spearman_P_values.xlsx")
Spearm2D_Pval.to_excel(writer, sheet_name="Sheet1")
writer.save()
```

Due to the large sample size of data, Bonferroni correction was applied on the p-values files. Bonferroni correction adjusted the p values to scale to the study sample size. P value was corrected by the number of genus-genus pairs. The correction was performed by dividing all p-values (α), with the total number of correlations observed (e.g. for study HIV, 38 taxa * 38 taxa -1 = 1443, thus α/1443) (Appendix 16). Hence, lowering the threshold at which a p value was considered significant (original p = 0.05). This correction ensured reduced chances of acquiring false positive reads (type I errors) in statistical analyses investigating correlation relationships. The threshold of spearman rank correlation coefficient (ρ) value representing 95 % confidence that a correlation is significant, was calculated. The corresponding rho value was determined from the Bonferroni-adjusted p-value using a statistical table relating rho and p in Excel. Briefly, the t statistic was calculated using the relationship t = ρ * SQRT[(n-2)/(1- ρ^2)] (where n is the number of samples in the study). The p-value was calculated for values of t using the Excel function TDIST. This correction ensured reduced chances of acquiring false positive reads (type I errors) in statistical analyses investigating correlation relationships.

Aiming to identify intrapersonal bacterial interactions focus was redirected to the Spearman correlation data, where high positive and negative correlations were selected. The selected data pairs exceeded the threshold for rho-value representing > 95 % confidence. This value of rho varied between studies, as it depends on sample size and number of genus-genus pairs. Once certified, the strongly correlated taxonomies were visualised through the design of linear graphs (Appendix 17, 18) via python programming (script not shown). To visualise the correlation between two taxa, the abundance data of two species were plotted against each other, where a linear association (best fit line) illustrated the intensity and the type of correlation (positive or negative correlation).

2.5 Clustering and Principal Component analyses

Clustering analysis was conducted, as it could show patterns of similarity and coexistence. Clustering analysis would create groupings of samples according to a distance similarity matrix, a density threshold and statistical distributions. A cluster can be a changeable term depending on parameters and algorithm types selected to run the analysis. A large assortment of algorithms was available for clustering, varying from Hierarchical and K-means clustering to statistical models like Principal Component Analysis.

For the purpose of this study multiple clustering algorithms were attempted. However hierarchical clustering was the preferred algorithm, as it did not require prior knowledge of the number of clusters or the values of centroid centres (the centres of each cluster group). In contrast, a successful K-means analysis can only be achieved through knowing or guessing these parameters.

Hierarchical clustering uses distance connectivity models to create grouping, unlike K-means which uses centroid models representing each cluster by a single mean vector. In other words, hierarchical clustering is based on the theory that samples in closer proximity will be more closely related than samples with a larger distance between them. An additional beneficial feature of hierarchical clustering is the utilisation of dendrograms to assist with clustering visualisation. Dendrograms represent the hierarchy of sample groupings into clusters, which combine with other sample groups at certain distances. The axis in a dendrogram displayed the distance between members of the same or different clusters. More specifically in this study the distance similarity was calculated through a Euclidean distance model converting distance into a metric space. Clustering analyses additionally require linkage criteria algorithms, to order and estimate the distance of each sample within one cluster. Average linkage clustering or otherwise named: Unweighted Pair Group Method with Arithmetic Mean (UPGMA) method was used for this study's data due to the larger data coverage.

The first attempt in performing clustering analysis was carried through CIMminer (Cluster Image Maps) open source tool[3]. CIMminer offers one or two matrix clustering analysis. As the abundance data files (.biom file/ OTU table) were two dimensional; with a format of taxonomy rows against sample columns, one matrix approach was the appropriate technique. CIMminer requires a .txt file input format, thus the abundance data were converted and uploaded onto their server, followed by customisation of clustering parameters. Clustering was performed on both taxonomy (rows) and samples (columns) values, through Euclidean distance method and average linkage clustering algorithm. CIMminer results were emailed to an address provided, containing 6 files. The key output file was an html index file enclosing a clustering heatmap with two corresponding dendrograms for both bacteria and sample data (Appendix 19). Although the resulting heatmap offered clear visualisation of patterns and relationships between both samples and bacteria, the application tool needed refining as lacks room for customisation. For that reason, analysis progressed through other applied bioinformatics tools approaching clustering (Cluster 3.0- Java TreeView).

The next open source software employed was Cluster 3.0 which likewise offers a number of different clustering methods[4]. Cluster 3.0 was downloaded on a windows drive and the abundance data file was again uploaded as a text file. Hierarchical clustering was chosen for both "genes" (representing row-wise means) and "arrays" (column-wise means) [95]. Finally, Euclidean distance and average linkage clustering models were chosen once again. Cluster 3.0 generates a .cdt output file, only compatible and legible through the assistance of a second software; Java TreeView. Java TreeView is an open source program allowing visualisation and interactive analysis of the data created by Cluster 3.0. The output generates a file consisting of a heatmap with the corresponding dendrograms for both column and row values. Unfortunately, although an initial test ran successfully, Cluster 3.0 ran into a major program and server malfunction with the application failing to resume. Despite all troubleshooting efforts, the underlying issue could not be traced. Remaining unexplained, the approach had to be removed from the methodology proposed here.

[3] https://discover.nci.nih.gov/cimminer/

[4] http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm

2.5.1 Implementing python scripts to generate Heatmaps

As presented thus far, Hierarchical Clustering analyses can be effectively visualised through heatmaps. A heat map allows representation of numerical data (i.e. a data table) by using colours and a spectrum to signify values. It also allows qualitative and quantitative control on the clustering data represented. The clustering matrix in a heatmap would display the similarity or dissimilarity of values via generation of a distance matrix.

The applied bioinformatics tools discussed previously, lacked parameter and illustration modification, thus a python script was programmed to perform clustering analysis that would allow control on parameters and visual marks. The code was designed with the help of ploty python library and was divided into 3 main sections: 1) creating a dendrogram for the bacteria values (rows), a 2) generating a dendrogram for the sample values (columns), and 3) constructing a heatmap in a three dimensional matrix. The full python script created to carry Hierarchical clustering via Euclidean distance and average linkage algorithms is enclosed bellow:

```
Clustering and Heatmap analysis script:

import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import scipy
import pylab
import scipy.cluster.hierarchy as sch
import matplotlib

file1 = pd.read_excel("Log_Abundances_BV.xlsx")

# Override the default linewidth.
matplotlib.rcParams['lines.linewidth'] = 10
data = np.transpose(np.array(file1))

# Compute and plot first dendrogram. X=SAMPLES
fig = pylab.figure(figsize=(80,80))
ax1 = fig.add_axes([0.3,0.71,0.6,0.2])
X = sch.linkage(data, "average")
Z1 = sch.dendrogram(X, orientation='top')
ax1.set_xticks([])
ax1.set_yticks([])

# Plot colorbar
cbar = fig.colorbar(im, ticks=[-1, 0.5, 2.5, 5], orientation='top',
shrink=0.3, aspect=10)
cbar.ax.set_xticklabels(['Low', 'Medium', 'High'], fontsize=50)  #
horizontal colorbar

# Compute and plot second dendrogram. Y=BACTERIA
data2 = np.array(file1)
ax2 = fig.add_axes([0.09,0.1,0.2,0.6])
Y = sch.linkage(data2, "complete")
Z2 = sch.dendrogram(Y, orientation='left')
ax2.set_xticks([])
ax2.set_yticks([])

#clusterind data
clustered_data = np.zeros(data2.shape)
for i, j in enumerate(Z1["leaves"]):
    clustered_data[:,i] = data2[:,j]
clustered_data2 = np.zeros(data2.shape)

for i, j in enumerate(Z2["leaves"]):
    clustered_data2[i,:] = clustered_data[j,:]

#heatmap
# Plot distance matrix.
axmatrix = fig.add_axes([0.3,0.1,0.6,0.6])
im = axmatrix.matshow(clustered_data2, aspect='auto', origin='lower',
cmap=pylab.cm.YlGnBu)
axmatrix.set_xticks([])
axmatrix.set_yticks([])

#LABELS SAMPLES
samples = Z1["leaves"]
names = file1.columns
idx1 = []
for i in samples:
    for n,SRR in enumerate(names):
        if i==n:
            i=SRR
            idx1.append(i)

axmatrix.set_xticks(range(112))
axmatrix.set_xticklabels(idx1, minor=False)
axmatrix.xaxis.set_label_position('bottom')
axmatrix.xaxis.tick_bottom()
pylab.xticks(rotation="vertical", fontsize=20)

#LABELS BACTERIA
bacteria = Z2["leaves"]
taxa = file1.index
idx2 = []
for i in bacteria:
    for n,name in enumerate(taxa):
        if i==n:
            names=name.split("_")
            i=names[-1]
            idx2.append(i)
axmatrix.set_yticks(range(109))
axmatrix.set_yticklabels(idx2, minor=False)
axmatrix.yaxis.set_label_position('right')
axmatrix.yaxis.tick_right()
pylab.yticks(rotation="horizontal", fontsize=30)

fig.show()
fig.savefig('dendrogram_BV_FIX.png')
```

2.5.2 Implementing python scripts to investigate Principal Component Analysis

An additional way to study clustering of samples is though Principal Component Analysis. Principal Component Analysis is a statistical model that transforms multidimensional data into two or three dimensions to allow visualisation. In other words, converts correlated non-linear data into linear linked data called principal components. This results in a new output file where the number of principal components reflects the number of original samples.

For this study, Principal Component analysis model was carried out through two vastly different approaches. The first application was achieved through QIIME's core_diversity_analyses.py script, in the form of a 3D graph displaying sample groupings (discussed in section 2.2.3). A python script was additionally created to carry PCA to allow further interactive clustering investigation. Sklearn's decomposition python library was utilised to allow two dimensional and three dimensional PCA testing (see Appendix 20 for full analysis and graphs). Unlike the Principal Component analysis created through QIIME, python allowed parameters and features modification as well as focus on single principal components. Unlike with QIIME's pipeline, where total sample clustering was observed; the script emphasised the number of clusters created within one variable (samples or bacteria). To reduce time and size restrictions for the computational runs, only three principal components were calculated due to the lack of significant variation within the data processed for this study. By observing the biodiversity distribution, it was evident that most of the total variance of a study, could be explained by the first Principal Component. Once PC1 was isolated and displayed as a histogram, clustering of either samples (column values) or bacteria (row values) could be visualised very clearly (Appendix 20).

3. EXPLORING COMPOSITION DIVERSITY IN HUMAN VAGINAL MICROBIOMES

3.1 Investigating composition of dysbiotic vaginal microbiomes

The aim of this study was to examine diversity of the bacteria in vaginal microbiomes. Diversity analyses assist in interpreting microbiome community structure and function [96]. Publically available datasets were used, which included samples from healthy individuals and individuals with various diseases or clinical syndromes which might be associated with dysbiosis. In this chapter, the microbial diversity amongst individuals within these five selected studies were analysed using various methods. Microbiome studies have proven very useful in establishing health and disease within individuals. As Turnbaugh *et al.* 2016 discuss in their study on gut microbiome associations with obesity, important questions on human health and disease can be addressed through microbiome diversity analyses [97]. Therefore, investigation of diversity within and in-between various microbiome studies is essential to examine human and ecosystem disorders; as human vaginal disorders might be linked to a breakdown of normal microbial community structure (and function).

The first attempt was to visualise α microbial diversity within samples as well as β – diversity among samples of a single study. α diversity represents the number of unique species within an individual, whereas β – diversity illustrates the differences in species composition between individuals. α - diversity was estimated via taxon based methods and more specifically the generation of taxonomy bar plots. Taxonomy bar plots were created, as mentioned previously, through QIIME's core_diversity_analyses.py script on studies; HIV, HSV2, BV, CANDIDIASIS, SV which contain samples from healthy individuals and individuals with conditions such as HIV, BV, Herpes and candidiasis (Figures 3-12). Study HIV contained healthy atypical vaginal samples; HSV2 sampled HIV, BV, HSV-1, HSV-2 and yeast infected females; BV consisted of healthy and BV infected samples; CANDIDIASIS enlisted BV and vulvovaginal candidiasis patients; and finally SV sampled healthy vaginal microbiomes.

The taxonomy bar plots, created by QIIME, represent the collective bacterial richness of an individual study as well as illustrating interpersonal variation between patients. The bar charts display all assigned taxonomies within each sample of a single study. The x-axis lists the sample ID's and the y-axis displays different coloured bars representing individual taxonomies identified through QIIME's pick_open_reference_otus.py command. Taxonomies are presented in the format assigned via QIIME, which have been consistently displayed at the Genus level for all five studies. Additional bar charts at either Family or Order taxonomic levels are also included for each study to signify the total level of variation within each study. Family or Order level bar taxonomies would allow investigation of the level of differentiation at various levels of taxonomy. The length of each coloured bar signifies the relative abundance of a specific taxon. Bacterial abundance could lead to information about microbiome and community structure through correlation studies, which is discussed in the following chapters 4 and 5.

Bar charts are very informative illustrations when working with microbiome data, providing visual representation of diversity within and between samples. Bar charts not only allow insight into patients' intrapersonal bacterial variation and total bacterial richness, but also display common and rare taxa. Exhibiting microbiome composition with categorical quantitative data allows straightforward comparisons between bacteria, patients and studies. Bacterial diversity comparison between studies, via taxonomic bar charts, is possible as bar taxonomies overcome irregularities caused by dissimilarities in rRNA sampling or sequencing. For this project, comparison of the variation between studies was a beneficial feature offered by the bar charts, which could assist identification of potential links between various dysbiotic vaginal environments. Taxonomic bar charts permit visualisation of total diversity within a microenvironment or a complete study, thus enabling comparisons between all five selected studies. Bar charts display a detailed representation of each taxon contained within each sample. However, due to the large amounts of sample data, it is possible to overlook certain less dominant taxonomic interactions within a study. It is important to state that even though this approach is not as accurate when focusing on intrapersonal variation and composition, universal diversity patterns leading to assumptions on community relationships could be identified.

The suggestive relationships were further investigated and tested through various means and statistical models. The analyses supported the presence of certain community structures in dysbiotic vaginal microbiomes. These will be discussed in detail in the following sections (chapters 4 and 5). Due to metadata sequence annotation issues and lack of sample metadata descriptions, it was not possible to assign relationships between sample and any particular medical syndromes which were associated with the donor of that sample. However, the analysis presented here allows comparisons between samples and studies where clear associations between members of the microbiomes and the microbiomes themselves can be traced. This pipeline offers the prospect of application on datasets with sample descriptions to identify correlations between medical syndromes by comparing various microbiomes.

Through observing the bar charts, it can be concluded that some studies demonstrate more diverse communities with patients containing diverse and varying bacterial communities, whereas others are mainly composed by numerous monoclonal samples. By comparing total taxa composition diversity within studies, studies HIV and CANDIDIASIS display the highest levels of diversity within the selected studies used for this project (Figure 3, 4). Study HIV had fewer samples than CANDIDIASIS, yet consisted of more patients with higher bacteria diversity and fewer monoclonal microbiomes. Figure 3 depicts considerable organismal diversity, with most samples containing multiple high abundance taxa and fewer samples consisting of exclusively or most abundantly of green bars (in Figure 3 depicting *lactobacilli*). On the other hand, study CANDIDIASIS catalogues more unique taxa (as seen in Figure 4b) and submits greater sampling depth, however contains a greater number of monoclonal samples. Most samples in Figure 4a carry an abundance of *Lactobacilli* illustrated with pale peach coloured bars. Although study CANDIDIASIS identifies additional unique taxa (in comparison to HIV), most atypical genera remain in relatively low abundances, therefore not contributing to total community variation. In other words, study HIV exhibits higher intrapersonal variation (α diversity), whereas study CANDIDIASIS represents higher β diversity levels (variation between individuals).

a)

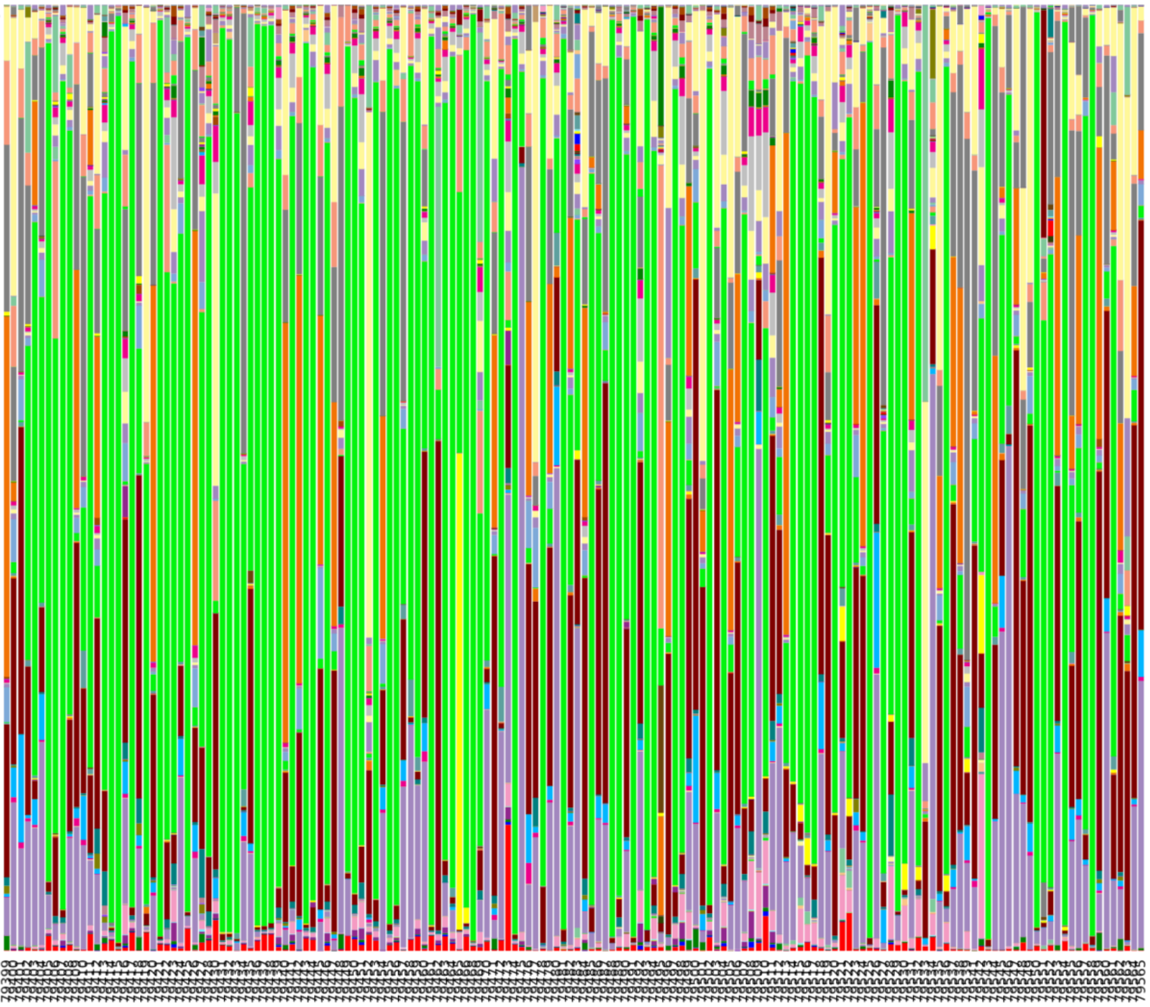


Figure 3: Study HIV genus level taxonomy bar chart. Figure 3a the bar chart displays all assigned taxonomies from each sample. The x-axis lists sample IDs with various colour bars representing individual taxonomies assigned via QIIME scripts. The length of each coloured bar signifies the relative abundance of a specific taxon. The green bars represent Lactobacillus as the most abundant organism in the complete study. Figure 3b lists the taxonomy IDs corresponding to the bar chart of Figure 3a.

b)

- 🟥 No blast hit;Other;Other;Other;Other;Other
- 🟦 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinomycetaceae;g__Actinomyces
- 🟧 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinomycetaceae;g__Arcanobacterium
- 🟩 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinomycetaceae;g__Mobiluncus
- 🟪 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinomycetaceae;g__Varibaculum
- 🟨 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Brevibacteriaceae;g__Brevibacterium
- 🟦 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Corynebacteriaceae;g__
- 🟪 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium
- 🟩 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Intrasporangiaceae;g__
- 🟫 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Intrasporangiaceae;g__Serinicoccus
- ⬜ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Microbacteriaceae;g__Pseudoclavibacter
- 🟩 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micrococcaceae;g__Arthrobacter
- 🟧 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micrococcaceae;g__Micrococcus
- 🟦 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Propionibacteriaceae;g__
- 🟧 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Propionibacteriaceae;g__Propionibacterium
- 🟩 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium
- 🟪 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Bifidobacteriales;f__Bifidobacteriaceae;g__Gardnerella
- 🟨 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__;g__
- ⬜ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__Coriobacteriaceae;g__
- 🟪 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__Coriobacteriaceae;g__Adlercreutzia
- 🟦 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__Coriobacteriaceae;g__Atopobium
- 🟫 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__Coriobacteriaceae;g__Collinsella
- 🟩 k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__;g__
- 🟪 k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides
- 🟫 k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__
- 🟩 k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Porphyromonas
- 🟥 k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella
- 🟩 k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__
- 🟧 k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Haloanella
- 🟥 k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Wautersiella
- 🟦 k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__;g__Exiguobacterium
- 🟧 k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__;g__Gemella
- 🟩 k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Anaerobacillus
- 🟪 k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__
- 🟨 k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus
- 🟦 k__Bacteria;p__Firmicutes;c__Bacilli;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Haloplasma
- 🟪 k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae;g__
- 🟩 k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae;g__Aerococcus
- 🟫 k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae;g__Granulicatella
- ⬜ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Enterococcaceae;g__Enterococcus
- 🟩 k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus
- 🟧 k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus
- 🟦 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__
- 🟧 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__
- 🟩 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__Clostridium
- 🟪 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XI. Incertae Sedis;g__
- 🟨 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XI. Incertae Sedis;g__Anaerococcus
- ⬜ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XI. Incertae Sedis;g__Finegoldia
- 🟪 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XI. Incertae Sedis;g__Peptoniphilus
- 🟦 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XIII. Incertae Sedis;g__Anaerovorax
- 🟫 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XIII. Incertae Sedis;g__Mogibacterium
- 🟩 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Eubacteriaceae;g__Eubacterium
- 🟪 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__
- 🟫 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Blautia
- 🟩 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Butyrivibrio
- 🟥 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Clostridium
- 🟩 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Coprococcus
- 🟧 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Moryella
- 🟥 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Roseburia
- 🟦 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Ruminococcus
- 🟧 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Shuttleworthia
- 🟩 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptococcaceae;g__Peptococcus
- 🟪 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;g__Filifactor
- 🟨 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;g__Peptostreptococcus
- 🟦 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__
- 🟪 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Eubacterium
- 🟩 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Faecalibacterium
- 🟫 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Oscillospira
- ⬜ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__
- 🟩 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Anaeroglobus
- 🟧 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister
- 🟦 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Megasphaera
- 🟧 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Mitsuokella
- 🟩 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Veillonella
- 🟪 k__Bacteria;p__Fusobacteria;c__Fusobacteria (class);o__Fusobacteriales;f__Fusobacteriaceae;g__Fusobacterium
- 🟨 k__Bacteria;p__Fusobacteria;c__Fusobacteria (class);o__Fusobacteriales;f__Fusobacteriaceae;g__Sneathia
- ⬜ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulobacteraceae;g__Brevundimonas
- 🟪 k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Haematobacter
- 🟦 k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Oligella
- 🟫 k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Sutterella
- 🟩 k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Neisseria
- 🟪 k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__Desulfobacteraceae;g__
- 🟫 k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae;g__Campylobacter
- 🟩 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__Succinivibrionaceae;g__Succinivibrio
- 🟥 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__
- 🟩 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Cronobacter
- 🟧 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Morganella
- 🟥 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Providencia
- 🟦 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Pseudomonadaceae;g__Pseudomonas
- 🟧 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__
- 🟩 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Haemophilus
- 🟪 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__
- 🟨 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Acinetobacter
- 🟦 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Psychrobacter
- 🟪 k__Bacteria;p__Synergistetes;c__Synergistia;o__Synergistales;f__Dethiosulfovibrionaceae;g__Jonquetella
- 🟩 k__Bacteria;p__Synergistetes;c__Synergistia;o__Synergistales;f__Dethiosulfovibrionaceae;g__Pyramidobacter
- 🟫 k__Bacteria;p__TM7;c__TM7-3;o__CW040;f__;g__
- ⬜ k__Bacteria;p__TM7;c__TM7-3;o__I025;f__;g__
- 🟩 k__Bacteria;p__Tenericutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__
- 🟧 k__Bacteria;p__Tenericutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Bulleidia
- 🟦 k__Bacteria;p__Tenericutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Catenibacterium
- 🟧 k__Bacteria;p__Tenericutes;c__Mollicutes;o__Acholeplasmatales;f__Acholeplasmataceae;g__Candidatus Phytoplasma
- 🟩 k__Bacteria;p__Tenericutes;c__Mollicutes;o__Mycoplasmatales;f__Mycoplasmataceae;g__Mycoplasma
- 🟪 k__Bacteria;p__Tenericutes;c__Mollicutes;o__Mycoplasmatales;f__Mycoplasmataceae;g__Ureaplasma
- 🟨 k__Bacteria;p__Tenericutes;c__Mollicutes;o__RF39;f__;g__

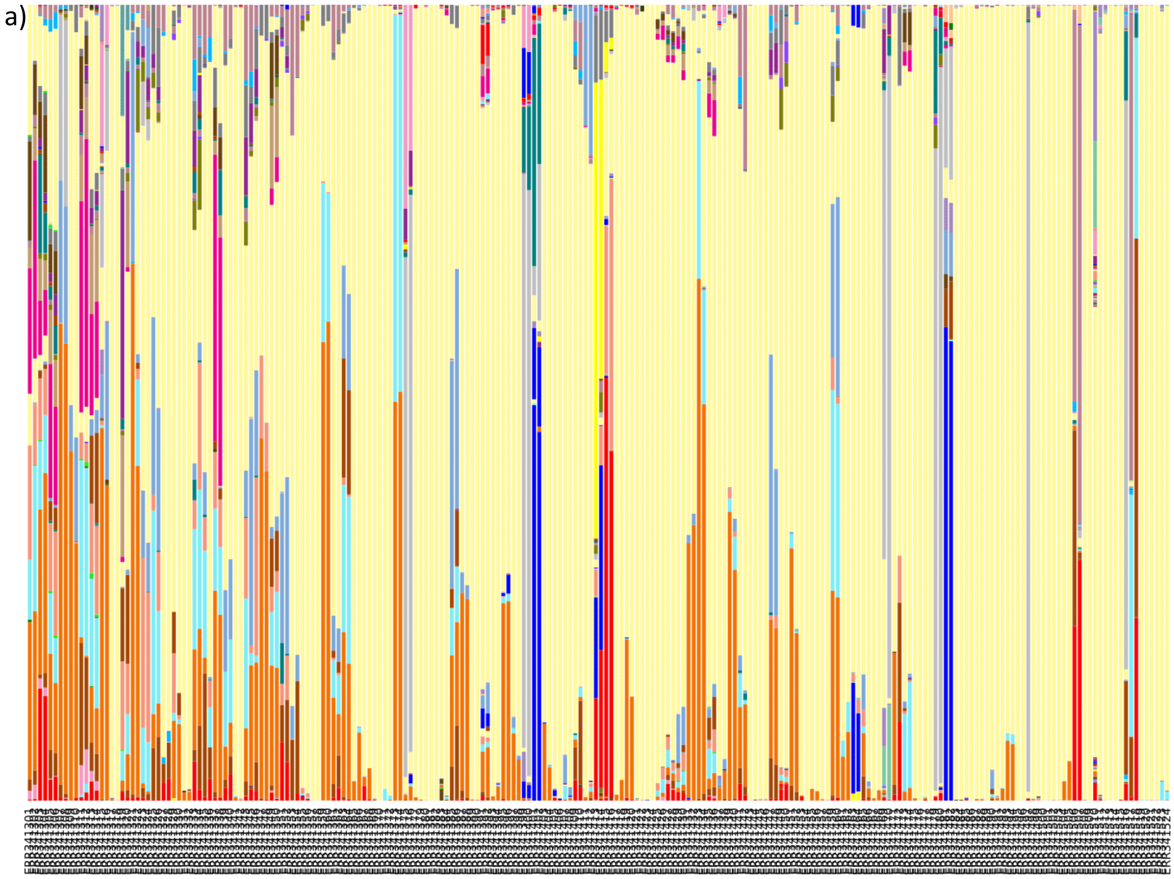
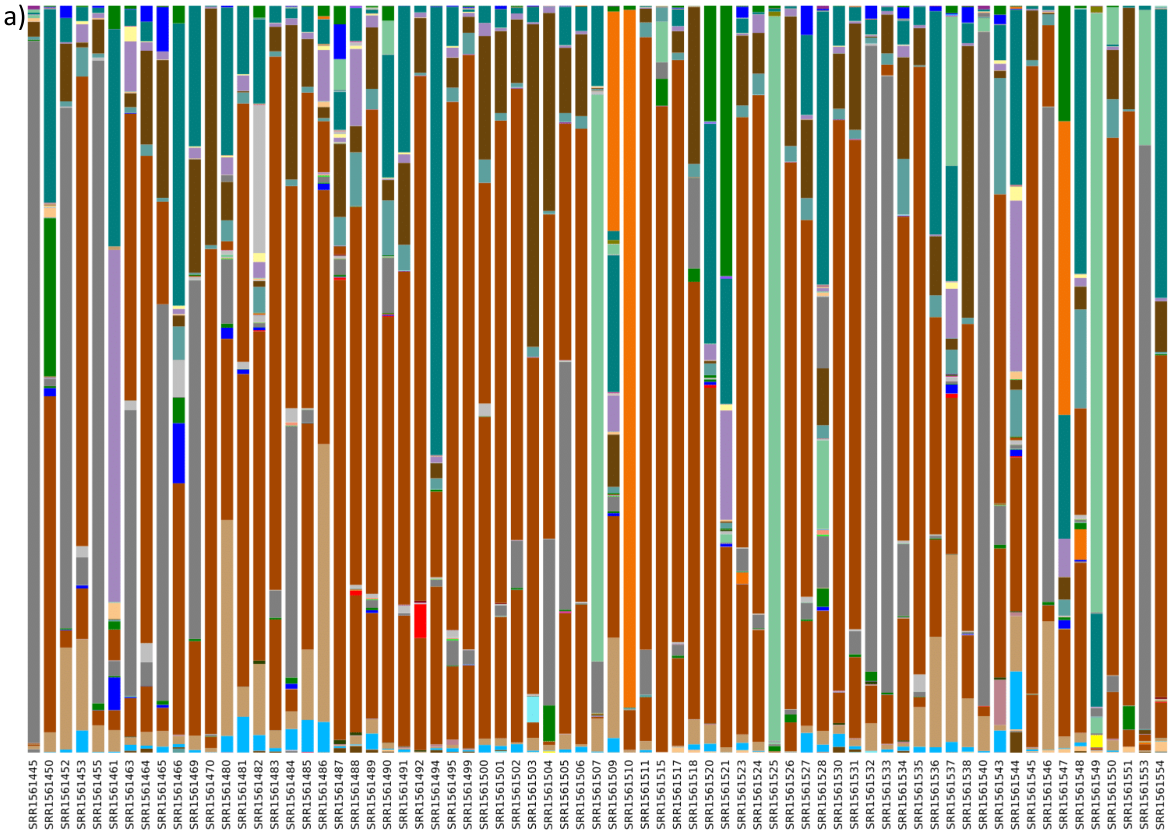


Figure 4: Study CANDIDIASIS genus level taxonomy bar chart. Figure 4a illustrates a bar chart representing all assigned taxonomies for the samples. The x-axis displays the sample IDs with colour bars representing the assigned taxonomies. The length of each bar displays the relative abundance of a taxon within a single sample. The salmon colour bars represent the most abundant bacteria in the study, Lactobacillus. Figure 4b lists the taxonomic identities of each coloured bar in Figure 4a.

b)
No blast hit;Other;Other;Other;Other;Other
k__Bacteria;p__Acidobacteria;c__Acidobacteria (class);o__Acidobacteriales;f__;g__
k__Bacteria;p__Acidobacteria;c__Chloracidobacteria;o__;f__;g__
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__;f__;g__
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__ACK-M1;g__
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinomycetaceae;g__Actinomyces
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinomycetaceae;g__Arcanobacterium
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinomycetaceae;g__Mobiluncus
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Corynebacteriaceae;g__
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Dermabacteraceae;g__Brachybacterium
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Dermabacteraceae;g__Dermabacter
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Dietziaceae;g__Dietzia
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Intrasporangiaceae;g__
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Intrasporangiaceae;g__Janibacter
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Intrasporangiaceae;g__Serinicoccus
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Microbacteriaceae;g__Agromyces
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Microbacteriaceae;g__Leucobacter
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Microbacteriaceae;g__Microbacterium
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micrococcaceae;g__Citricoccus
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micrococcaceae;g__Kocuria
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micrococcaceae;g__Micrococcus
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micrococcaceae;g__Nesterenkonia
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micrococcaceae;g__Rothia
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nocardiaceae;g__Rhodococcus
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nocardioidaceae;g__Nocardioides
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Propionibacteriaceae;g__Propionibacterium
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Streptomycetaceae;g__Streptomyces
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Bifidobacteriales;f__Bifidobacteriaceae;g__
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Bifidobacteriales;f__Bifidobacteriaceae;g__Alloscardovia
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Bifidobacteriales;f__Bifidobacteriaceae;g__Gardnerella
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__;g__
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__Coriobacteriaceae;g__
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__Coriobacteriaceae;g__Adlercreutzia
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__Coriobacteriaceae;g__Atopobium
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Solirubrobacterales;f__;g__
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Solirubrobacterales;f__Patulibacteraceae;g__
k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Solirubrobacterales;f__Solirubrobacteraceae;g__
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__;g__
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Porphyromonas
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes
k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__;g__Candidatus Sulcia
k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__
k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__;g__
k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Flexibacteraceae;g__Cytophaga
k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Sphingobacteriaceae;g__
k__Bacteria;p__Chlamydiae;c__Chlamydiae (class);o__Chlamydiales;f__Chlamydiaceae;g__Chlamydia
k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__;f__;g__
k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__H39;f__;g__
k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__S0208;f__;g__
k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__WCHB1-50;f__;g__
k__Bacteria;p__Chloroflexi;c__Ktedonobacteria;o__;f__;g__
k__Bacteria;p__Cyanobacteria;c__;o__;f__;g__
k__Bacteria;p__Cyanobacteria;c__;o__Oscillatoriales;f__;g__
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__;g__Gemella
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Anoxybacillus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Bacillus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Geobacillus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Oceanobacillus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Paenibacillaceae;g__Brevibacillus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Paenibacillaceae;g__Paenibacillus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__Paenisporosarcina
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__Sporosarcina
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__;g__
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae;g__Aerococcus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae;g__Facklamia
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae;g__
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae;g__Granulicatella
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Enterococcaceae;g__Enterococcus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Enterococcaceae;g__Vagococcus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__Clostridium
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XI. Incertae Sedis;g__
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XI. Incertae Sedis;g__Anaerococcus
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XI. Incertae Sedis;g__Finegoldia
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XI. Incertae Sedis;g__Peptoniphilus
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Blautia
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Clostridium
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Coprococcus
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Lachnospira
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Moryella
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Roseburia
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptococcaceae;g__Peptococcus
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;g__Peptostreptococcus
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Clostridium
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Eubacterium
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Faecalibacterium
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Phascolarctobacterium
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Thermosinus
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Veillonella
k__Bacteria;p__Firmicutes;c__Clostridia;o__Desulfitobacterales;f__Desulfitobacteraceae;g__
k__Bacteria;p__Firmicutes;c__Clostridia;o__SHA-98;f__;g__
k__Bacteria;p__Fusobacteria;c__Fusobacteria (class);o__Fusobacteriales;f__Fusobacteriaceae;g__Fusobacterium
k__Bacteria;p__Fusobacteria;c__Fusobacteria (class);o__Fusobacteriales;f__Fusobacteriaceae;g__Sneathia
k__Bacteria;p__Gemmatimonadetes;c__Gemmatimonadetes (class);o__;f__;g__
k__Bacteria;p__NKB19;c__;o__;f__;g__
k__Bacteria;p__Nitrospirae;c__Nitrospira (class);o__Nitrospirales;f__FW;g__GOUTA7
k__Bacteria;p__OP8;c__OP8;o__;f__;g__
k__Bacteria;p__Planctomycetes;c__Phycisphaerae;o__Phycisphaerales;f__;g__
k__Bacteria;p__Planctomycetes;c__Planctomycea;o__Gemmatales;f__Gemmataceae;g__
k__Bacteria;p__Planctomycetes;c__Planctomycea;o__Gemmatales;f__Isosphaeraceae;g__
k__Bacteria;p__Planctomycetes;c__Planctomycea;o__Pirellulales;f__;g__Rhodopirellula
k__Bacteria;p__Planctomycetes;c__agg27;o__OM190;f__;g__
k__Bacteria;p__Planctomycetes;c__vadinHA49;o__;f__;g__
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulobacteraceae;g__
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulobacteraceae;g__Brevundimonas
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulobacteraceae;g__Caulobacter
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulobacteraceae;g__Phenylobacterium
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae;g__
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae;g__Bradyrhizobium
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae;g__Rhodopseudomonas
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Methylobacteriaceae;g__Methylobacterium
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Methylocystaceae;g__
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Phyllobacteriaceae;g__
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Rhizobiaceae;g__Shinella
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Rhodobacteraceae;g__
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Rhodobacteraceae;g__Pannonibacter
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Rhodobiaceae;g__
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Rhodobiaceae;g__Rhodobium
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Xanthobacteraceae;g__Azorhizobium
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Amaricoccus
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Paracoccus
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Rhodospirillaceae;g__
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__;g__
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingomonas
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingopyxis
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__;g__
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__;g__Aquabacterium
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__;g__Tepidimonas
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Achromobacter
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Alcaligenes
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Azohydromonas
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Sutterella
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Limnobacter
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Ralstonia
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Acidovorax
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Brachymonas
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Hydrogenophaga
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Hylemonella
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Limnohabitans
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Polaromonas
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Rhodoferax
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Massilia
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Hydrogenophilales;f__Hydrogenophilaceae;g__
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Vogesella
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Rhodocyclales;f__;g__
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Rhodocyclales;f__Rhodocyclaceae;g__
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Rhodocyclales;f__Rhodocyclaceae;g__Methyloversatilis
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Rhodocyclales;f__Rhodocyclaceae;g__Zoogloea
k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Bdellovibrionales;f__Bacteriovoracaceae;g__
k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfobacterales;f__;g__
k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfovibrionales;f__Desulfovibrionaceae;g__
k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfuromonadales;f__Geobacteraceae;g__Geobacter
k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__Desulfobacteraceae;g__
k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__Syntrophaceae;g__
k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__Syntrophaceae;g__Syntrophus
k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__Syntrophobacteraceae;g__Syntrophobacter
k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae;g__Campylobacter
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__Aeromonadaceae;g__Aeromonas
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Buchnera
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Enterobacter
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Klebsiella
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Plesiomonas
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Providencia
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Raoultella
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__Coxiellaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__Legionellaceae;g__Legionella
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Pseudomonadaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Pseudomonadaceae;g__Pseudomonas
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Chelonobacter
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Haemophilus
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Acinetobacter
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Alkanindiges
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiotrichales;f__Piscirickettsiaceae;g__Methylophaga
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiotrichales;f__Thiotrichaceae;g__Beggiatoa
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Vibrionaceae;g__Photobacterium
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Vibrionaceae;g__Vibrio
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Lysobacter
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Pseudoxanthomonas
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Stenotrophomonas
k__Bacteria;p__Spirochaetes;c__SP_WWE1;o__;f__;g__
k__Bacteria;p__TM7;c__TM7-1;o__;f__;g__
k__Bacteria;p__Tenericutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Bulleidia
k__Bacteria;p__Tenericutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__L7A_E11
k__Bacteria;p__Tenericutes;c__Mollicutes;o__Mycoplasmatales;f__Mycoplasmataceae;g__Ureaplasma
k__Bacteria;p__Tenericutes;c__Mollicutes;o__RF39;f__;g__
k__Bacteria;p__Thermi;c__Deinococci;o__Deinococcales;f__Deinococcaceae;g__Deinococcus
k__Bacteria;p__Verrucomicrobia;c__Spartobacteria;o__;f__;g__Chthoniobacter
k__Bacteria;p__WS3;c__PRR-12;o__;f__;g__

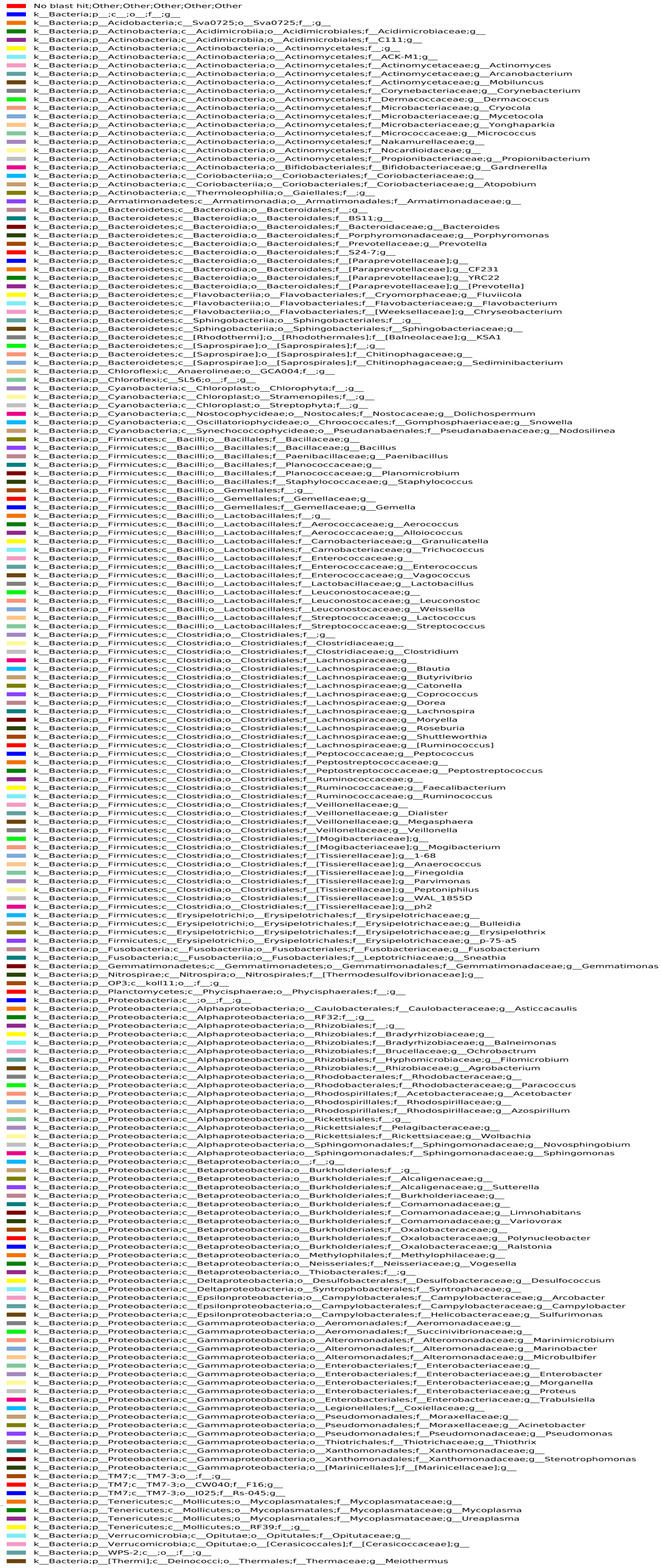
Studies BV, HSV2 and SV do not represent as highly diverse systems as the two previously mentioned studies. Though characterised by an immense number of unique taxa, their abundances remain substantially low with the majority of samples appearing monoclonal while dominated by *Lactobacilli*. In more detail, though study BV implemented thorough sampling depth (similar to study HIV), it is apparent that sampling does not affect total study diversity, due to the low taxon richness present in most samples (Figure 5). Though fewer monoclonal patients are present, bacteria appear relatively balanced within samples with most patients carrying similar abundance taxa. In other words, community structures appear more stable with most samples consisting of 8 dominant bacteria (*Lactobacilli, Shuttleworthia, Prevotella, Megasphaera, Sneathia, Parvimonas, Atopobium* and *Dialister*).

Study SV has multiple diverse samples, however half of the samples documented in the bar chart originated from male seminal samples, as study SV examined the influence of intercourse on vaginal microbiomes [82]. The first 23 samples (ERR769967-ERR769989) listed in the chart enclosing purple coloured bars represent seminal samples (Figure 7a). The taxa chart generated through QIIME shows that most male microbiomes are dominated by *Flavobacterium*, *Lactobacillus* and *Acinetobacter* (not *Corynebacterium* as reported in Mandar et al. 2015 published results [82]). Seminal microbiome samples are noticeably more diverse compared to vaginal samples, whilst most vaginal samples are dominated exclusively by *Lactobacilli* (illustrated by pale pink bars seen in Figure 7a) and supplementary bacteria are only present in very low abundances. In other words, male seminal microbiomes have a higher α- and β –diversity, compared to vaginal microbiomes with the majority of the vaginal communities appearing homogenous. As the present project was solely focused on vaginal microbiome composition, study SV was considered to present low β-diversity levels, even though total bacteria richness is excessively high (due to the immense number of unique taxonomies assigned illustrated in Figure 7b). Finally study HSV2 represents similar levels of diversity at both intrapersonal and total study diversity levels (Figure 6). Bacterial richness as displayed in Figure 6b remains low, followed by low abundances in most rare taxa (Figure 6a). The majority of the samples appear monoclonal and colonised by *Lactobacilli* as illustrated by the purple bars in Figure 6. Only a handful of study HSV2 samples demonstrate intrapersonal diversity, with a number of abundantly represent bacteria, of which most common is *Gardnerella*.



<u>Figure 5: Study BV genus level taxonomy bar chart.</u> Figure 5a illustrates a bar chart representing all taxonomies assigned in each sample from the study. The x-axis lists the sample IDs and the colour bars display the assigned taxonomies within each sample. The bar length represents the relative abundance of each taxon from a single sample. Brown bars illustrate *Shuttleworthia* as the most abundant organism. Figure 5b lists the taxonomic identities of each coloured bar.

b)


No blast hit;Other;Other;Other;Other;Other
k__Bacteria;p__;c__;o__;f__;g__
k__Bacteria;p__Acidobacteria;c__Sva0725;o__Sva0725;f__;g__
k__Bacteria;p__Actinobacteria;c__Acidimicrobiia;o__Acidimicrobiales;f__Acidimicrobiaceae;g__
k__Bacteria;p__Actinobacteria;c__Acidimicrobiia;o__Acidimicrobiales;f__C111;g__
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__;g__
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__ACK-M1;g__
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Actinomyces
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Arcanobacterium
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Mobiluncus
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Dermacoccaceae;g__Dermacoccus
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Microbacteriaceae;g__Cryocola
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Microbacteriaceae;g__Mycetocola
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Microbacteriaceae;g__Yonghaparkia
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Micrococcaceae;g__Micrococcus
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Nakamurellaceae;g__
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Nocardioidaceae;g__
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Propionibacteriaceae;g__Propionibacterium
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Gardnerella
k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__
k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__Atopobium
k__Bacteria;p__Actinobacteria;c__Thermoleophilia;o__Gaiellales;f__;g__
k__Bacteria;p__Armatimonadetes;c__Armatimonadia;o__Armatimonadales;f__Armatimonadaceae;g__
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__;g__
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__BS11;g__
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Porphyromonas
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__S24-7;g__
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__[Paraprevotellaceae];g__
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__[Paraprevotellaceae];g__CF231
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__[Paraprevotellaceae];g__YRC22
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__[Paraprevotellaceae];g__[Prevotella]
k__Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__Cryomorphaceae;g__Fluviicola
k__Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__Flavobacteriaceae;g__Flavobacterium
k__Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__[Weeksellaceae];g__Chryseobacterium
k__Bacteria;p__Bacteroidetes;c__Sphingobacteriia;o__Sphingobacteriales;f__;g__
k__Bacteria;p__Bacteroidetes;c__Sphingobacteriia;o__Sphingobacteriales;f__Sphingobacteriaceae;g__
k__Bacteria;p__Bacteroidetes;c__[Rhodothermi];o__[Rhodothermales];f__[Balneolaceae];g__KSA1
k__Bacteria;p__Bacteroidetes;c__[Saprospirae];o__[Saprospirales];f__;g__
k__Bacteria;p__Bacteroidetes;c__[Saprospirae];o__[Saprospirales];f__Chitinophagaceae;g__
k__Bacteria;p__Bacteroidetes;c__[Saprospirae];o__[Saprospirales];f__Chitinophagaceae;g__Sediminibacterium
k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__GCA004;f__;g__
k__Bacteria;p__Chloroflexi;c__SL56;o__;f__;g__
k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Chlorophyta;f__;g__
k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Stramenopiles;f__;g__
k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta;f__;g__
k__Bacteria;p__Cyanobacteria;c__Nostocophycideae;o__Nostocales;f__Nostocaceae;g__Dolichospermum
k__Bacteria;p__Cyanobacteria;c__Oscillatoriophycideae;o__Chroococcales;f__Gomphosphaeriaceae;g__Snowella
k__Bacteria;p__Cyanobacteria;c__Synechococcophycideae;o__Pseudanabaenales;f__Pseudanabaenaceae;g__Nodosilinea
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Bacillus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Paenibacillaceae;g__Paenibacillus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__Planomicrobium
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Gemellales;f__;g__
k__Bacteria;p__Firmicutes;c__Bacilli;o__Gemellales;f__Gemellaceae;g__
k__Bacteria;p__Firmicutes;c__Bacilli;o__Gemellales;f__Gemellaceae;g__Gemella
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__;g__
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae;g__Aerococcus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae;g__Alloiococcus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae;g__Granulicatella
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae;g__Trichococcus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Enterococcaceae;g__
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Enterococcaceae;g__Enterococcus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Enterococcaceae;g__Vagococcus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Leuconostocaceae;g__
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Leuconostocaceae;g__Leuconostoc
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Leuconostocaceae;g__Weissella
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Lactococcus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__Clostridium
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Blautia
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Butyrivibrio
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Catonella
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Coprococcus
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Dorea
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Lachnospira
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Moryella
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Roseburia
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Shuttleworthia
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__[Ruminococcus]
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptococcaceae;g__Peptococcus
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;g__
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;g__Peptostreptococcus
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Faecalibacterium
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcus
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Megasphaera
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Veillonella
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Mogibacteriaceae];g__
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Mogibacteriaceae];g__Mogibacterium
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__1-68
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Anaerococcus
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Finegoldia
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Parvimonas
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Peptoniphilus
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__WAL_1855D
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__ph2
k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__
k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Bulleidia
k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Erysipelothrix
k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__p-75-a5
k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Fusobacteriaceae;g__Fusobacterium
k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Leptotrichiaceae;g__Sneathia
k__Bacteria;p__Gemmatimonadetes;c__Gemmatimonadetes;o__Gemmatimonadales;f__Gemmatimonadaceae;g__Gemmatimonas
k__Bacteria;p__Nitrospirae;c__Nitrospira;o__Nitrospirales;f__[Thermodesulfovibrionaceae];g__
k__Bacteria;p__OP3;c__koll11;o__;f__;g__
k__Bacteria;p__Planctomycetes;c__Phycisphaerae;o__Phycisphaerales;f__;g__
k__Bacteria;p__Proteobacteria;c__;o__;f__;g__
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulobacteraceae;g__Asticcacaulis
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__RF32;f__;g__
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__;g__
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae;g__
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae;g__Balneimonas
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Brucellaceae;g__Ochrobactrum
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Hyphomicrobiaceae;g__Filomicrobium
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Rhizobiaceae;g__Agrobacterium
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Paracoccus
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Acetobacteraceae;g__Acetobacter
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Rhodospirillaceae;g__
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Rhodospirillaceae;g__Azospirillum
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__;g__
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__Pelagibacteraceae;g__
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__Rickettsiaceae;g__Wolbachia
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Novosphingobium
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingomonas
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__;f__;g__
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__;g__
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Sutterella
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Limnohabitans
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Variovorax
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Polynucleobacter
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Ralstonia
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Methylophilales;f__Methylophilaceae;g__
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Vogesella
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Thiobacterales;f__;g__
k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfobacterales;f__Desulfobacteraceae;g__Desulfococcus
k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__Syntrophaceae;g__
k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae;g__Arcobacter
k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae;g__Campylobacter
k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Helicobacteraceae;g__Sulfurimonas
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__Aeromonadaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__Succinivibrionaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Alteromonadaceae;g__Marinimicrobium
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Alteromonadaceae;g__Marinobacter
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Alteromonadaceae;g__Microbulbifer
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Enterobacter
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Morganella
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Proteus
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Trabulsiella
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__Coxiellaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Acinetobacter
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiotrichales;f__Thiotrichaceae;g__Thiothrix
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Stenotrophomonas
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__[Marinicellales];f__[Marinicellaceae];g__
k__Bacteria;p__TM7;c__TM7-3;o__;f__;g__
k__Bacteria;p__TM7;c__TM7-3;o__CW040;f__F16;g__
k__Bacteria;p__TM7;c__TM7-3;o__I025;f__Rs-045;g__
k__Bacteria;p__Tenericutes;c__Mollicutes;o__Mycoplasmatales;f__Mycoplasmataceae;g__
k__Bacteria;p__Tenericutes;c__Mollicutes;o__Mycoplasmatales;f__Mycoplasmataceae;g__Mycoplasma
k__Bacteria;p__Tenericutes;c__Mollicutes;o__Mycoplasmatales;f__Mycoplasmataceae;g__Ureaplasma
k__Bacteria;p__Tenericutes;c__Mollicutes;o__RF39;f__;g__
k__Bacteria;p__Verrucomicrobia;c__Opitutae;o__Opitutales;f__Opitutaceae;g__
k__Bacteria;p__Verrucomicrobia;c__Opitutae;o__[Cerasicoccales];f__[Cerasicoccaceae];g__
k__Bacteria;p__WPS-2;c__;o__;f__;g__
k__Bacteria;p__[Thermi];c__Deinococci;o__Thermales;f__Thermaceae;g__Meiothermus

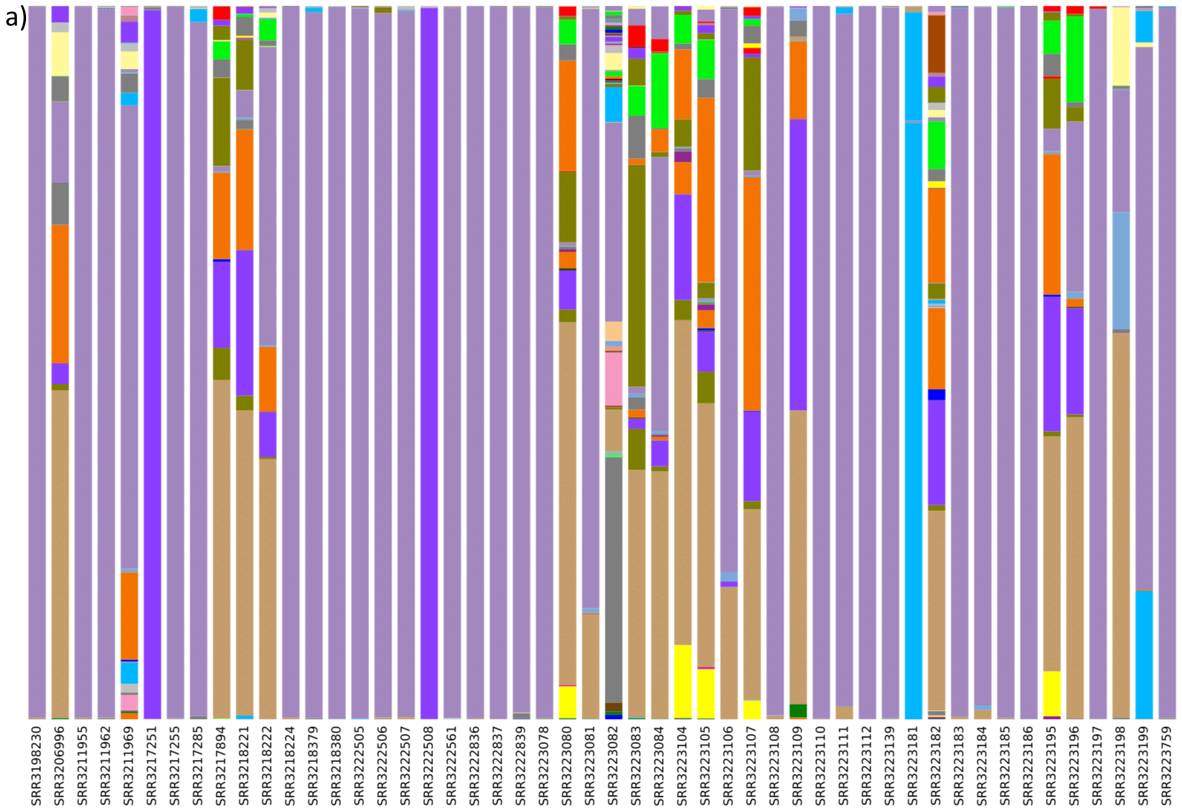


Figure 6: Study HSV2 genus level taxonomy bar chart. Figure 6a depicts a bar chart created via QIIME commands and represents all taxonomies identified within each sample of the study. The x-axis lists sample IDs and bars represent the taxa within each sample. The length of each coloured bar demonstrates the relative abundance of each taxon within each sample. Purple coloured bars represent Lactobacillus, which is the most abundant organism and displays multiple homogenous samples. Figure 6b lists the taxonomic identities of each coloured bar from Figure 6a.

b)

- ■ No blast hit;Other;Other;Other;Other;Other
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__;g__
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Actinobaculum
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Actinomyces
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Arcanobacterium
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Mobiluncus
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Trueperella
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Varibaculum
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinosynnemataceae;g__
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Brevibacteriaceae;g__Brevibacterium
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Dermabacteraceae;g__Dermabacter
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Jonesiaceae;g__
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Microbacteriaceae;g__Pseudoclavibacter
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Micrococcaceae;g__
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Micrococcaceae;g__Rothia
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Propionibacteriaceae;g__
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Propionibacteriaceae;g__Propionibacterium
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Propionibacteriaceae;g__Propionimicrobium
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium
- ■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Gardnerella
- ■ k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__
- ■ k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__Atopobium
- ■ k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__Collinsella
- ■ k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__Eggerthella
- ■ k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__Slackia
- ■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__;g__
- ■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides
- ■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Parabacteroides
- ■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Porphyromonas
- ■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella
- ■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__S24-7;g__
- ■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__[Paraprevotellaceae];g__YRC22
- ■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__[Paraprevotellaceae];g__[Prevotella]
- ■ k__Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__[Weeksellaceae];g__Wautersiella
- ■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__
- ■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus
- ■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Gemellales;f__Gemellaceae;g__
- ■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Gemellales;f__Gemellaceae;g__Gemella
- ■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__;g__
- ■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae;g__
- ■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae;g__Aerococcus
- ■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae;g__Facklamia
- ■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae;g__Granulicatella
- ■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus
- ■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Pediococcus
- ■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Leuconostocaceae;g__
- ■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__
- ■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__Clostridium
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Eubacteriaceae;g__Pseudoramibacter_Eubacterium
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Blautia
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Coprococcus
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Dorea
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Lachnospira
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Moryella
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Roseburia
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Shuttleworthia
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__[Ruminococcus]
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptococcaceae;g__Peptococcus
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;g__Peptostreptococcus
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Faecalibacterium
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcus
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Megasphaera
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Phascolarctobacterium
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Veillonella
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Mogibacteriaceae];g__
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__1-68
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Anaerococcus
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Finegoldia
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__GW-34
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Gallicola
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Helcococcus
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Parvimonas
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__Peptoniphilus
- ■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae];g__WAL_1855D
- ■ k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__
- ■ k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Bulleidia
- ■ k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__[Eubacterium]
- ■ k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Fusobacteriaceae;g__Fusobacterium
- ■ k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Leptotrichiaceae;g__Sneathia
- ■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Oligella
- ■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Sutterella
- ■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__
- ■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Herbaspirillum
- ■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__
- ■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Neisseria
- ■ k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae;g__Campylobacter
- ■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Shewanellaceae;g__Shewanella
- ■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__
- ■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Serratia
- ■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Halomonadaceae;g__Halomonas
- ■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Aggregatibacter
- ■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Haemophilus
- ■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas
- ■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__
- ■ k__Bacteria;p__TM7;c__TM7-3;o__I025;f__Rs-045;g__
- ■ k__Bacteria;p__Tenericutes;c__Mollicutes;o__Mycoplasmatales;f__Mycoplasmataceae;g__Mycoplasma
- ■ k__Bacteria;p__Tenericutes;c__Mollicutes;o__Mycoplasmatales;f__Mycoplasmataceae;g__Ureaplasma

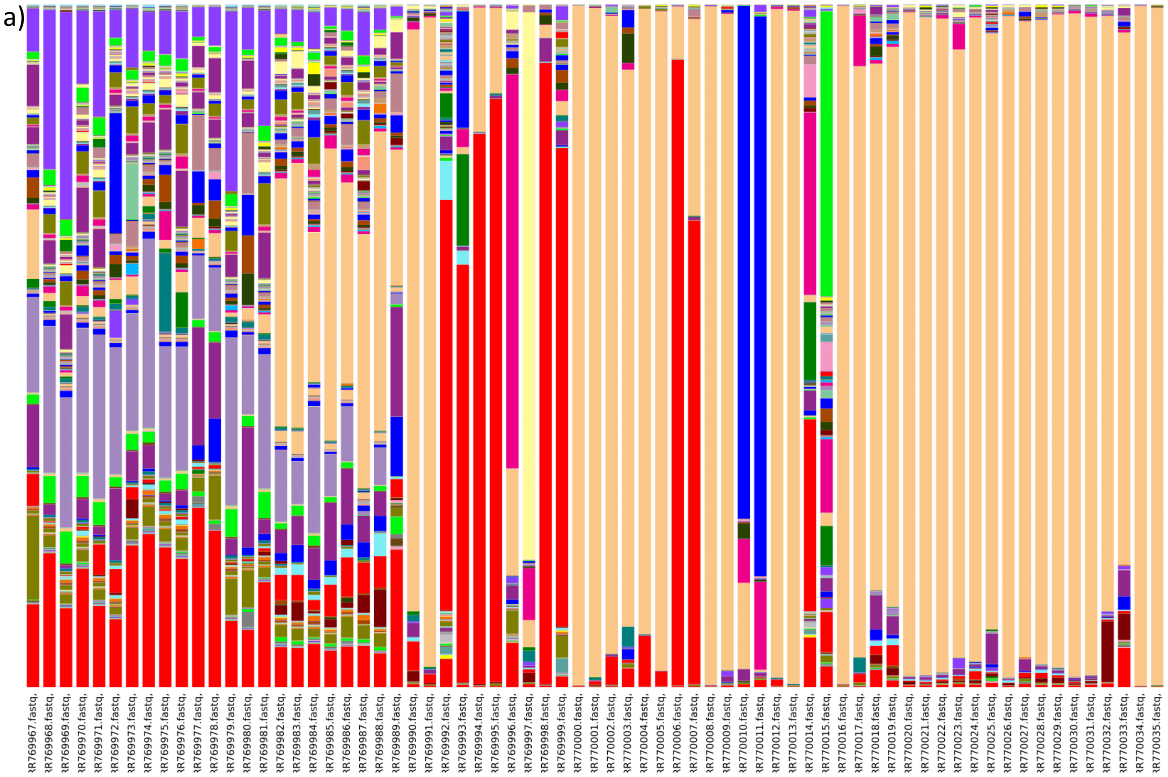


Figure 7: Study SV genus level taxonomic bar chart. Figure 7a depicts QIIME's genus level bar chart presenting all taxonomies identified within each sample of the study. The x-axis displays the sample IDs. Colour bars represent the taxonomy characterised for each sample. The length of the colour bars demonstrates the relative abundance of each taxon present. ERR769967-ERR769989 samples containing purple bars represent seminal specimens whereas the remaining samples originate from vaginal samples. Salmon colour bars depict Lactobacillus, which is the most abundant organism and multiple near-monoclonal samples are present. Figure 7b lists the taxonomic identities of each coloured bar in Figure 7a.

■ No blast hit;Other;Other;Other;Other;Other
■ k__Bacteria;p__;c__;o__;f__;g__
■ k__Bacteria;p__ABY1_OD1;c__;o__;f__;g__
■ k__Bacteria;p__AD3;c__ABS-6;o__;f__;g__
■ k__Bacteria;p__AD3;c__JG37-AG-4;o__;f__;g__
■ k__Bacteria;p__Acidobacteria;c__;o__;f__;g__
■ k__Bacteria;p__Acidobacteria;c__;o__;f__Koribacteraceae;g__
■ k__Bacteria;p__Acidobacteria;c__;o__;f__Koribacteraceae;g__Candidatus Koribacter
■ k__Bacteria;p__Acidobacteria;c__Acidobacteria (class);o__Acidobacteriales;f__;g__
■ k__Bacteria;p__Acidobacteria;c__Acidobacteria (class);o__Acidobacteriales;f__Acidobacteriaceae;g__
■ k__Bacteria;p__Acidobacteria;c__Acidobacteria-5;o__;f__;g__
■ k__Bacteria;p__Acidobacteria;c__Chloracidobacteria;o__;f__;g__
■ k__Bacteria;p__Acidobacteria;c__Holophagae;o__Holophagales;f__Holophagaceae;g__
■ k__Bacteria;p__Acidobacteria;c__Holophagae;o__Holophagales;f__Holophagaceae;g__Geothrix
■ k__Bacteria;p__Acidobacteria;c__MVS-40;o__;f__;g__
■ k__Bacteria;p__Acidobacteria;c__OS-K;o__;f__;g__
■ k__Bacteria;p__Acidobacteria;c__PAUC37f;o__;f__;g__
■ k__Bacteria;p__Acidobacteria;c__RB25;o__;f__;g__
■ k__Bacteria;p__Acidobacteria;c__Solibacteres;o__Solibacterales;f__Solibacteraceae;g__Candidatus Solibacter
■ k__Bacteria;p__Acidobacteria;c__Sva0725;o__;f__;g__
■ k__Bacteria;p__Acidobacteria;c__iii1-8;o__;f__;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__;f__;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__0319-7L14;f__;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Acidimicrobiales;f__;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Acidimicrobiales;f__AKIW874;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Acidimicrobiales;f__CL500-29;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Acidimicrobiales;f__EB1017;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Acidimicrobiales;f__Iamiaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Acidimicrobiales;f__Iamiaceae;g__Iamia
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Acidimicrobiales;f__JdFBGBact;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Acidimicrobiales;f__Microthrixaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Acidimicrobiales;f__SC3-41;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Acidimicrobiales;f__wb1_P06;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__ACK-M1;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinomycetaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinomycetaceae;g__Actinobaculum
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinomycetaceae;g__Actinomyces
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinomycetaceae;g__Arcanobacterium
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinomycetaceae;g__Mobiluncus
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinomycetaceae;g__Varibaculum
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinopolysporaceae;g__Actinopolyspora
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinospicaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinosynnemataceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Bogoriellaceae;g__Georgenia
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Brevibacteriaceae;g__Brevibacterium
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Catenulisporaceae;g__Catenulispora
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Cellulomonadaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Cellulomonadaceae;g__Cellulomonas
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Cellulomonadaceae;g__Demequina
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Corynebacteriaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Cryptosporangiaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Cryptosporangiaceae;g__Cryptosporangium
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Dermabacteraceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Dermabacteraceae;g__Brachybacterium
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Dermabacteraceae;g__Dermabacter
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Dermacoccaceae;g__Dermacoccus
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Dermacoccaceae;g__Kytococcus
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Dietziaceae;g__Dietzia
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Frankiaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Geodermatophilaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Geodermatophilaceae;g__Geodermatophilus
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Glycomycetaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Gordoniaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Gordoniaceae;g__Gordonia
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Intrasporangiaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Intrasporangiaceae;g__Janibacter
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Intrasporangiaceae;g__Phycicoccus
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Intrasporangiaceae;g__Serinicoccus
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Intrasporangiaceae;g__Terracoccus
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Kineosporiaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Kineosporiaceae;g__Kineococcus
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Microbacteriaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Microbacteriaceae;g__Agrococcus
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Microbacteriaceae;g__Agromyces
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Microbacteriaceae;g__Candidatus Rhodoluna
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Microbacteriaceae;g__Cryocola
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Microbacteriaceae;g__Curtobacterium
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Microbacteriaceae;g__Leucobacter
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Microbacteriaceae;g__Microbacterium
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Microbacteriaceae;g__Microcella
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Microbacteriaceae;g__Pseudoclavibacter
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Microbacteriaceae;g__Rathayibacter
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Microbacteriaceae;g__Subtercola
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micrococcaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micrococcaceae;g__Arthrobacter
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micrococcaceae;g__Citricoccus
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micrococcaceae;g__Kocuria
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micrococcaceae;g__Micrococcus
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micrococcaceae;g__Nesterenkonia
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micrococcaceae;g__Renibacterium
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micrococcaceae;g__Rothia
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micrococcaceae;g__Zhihengliuella
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micromonosporaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micromonosporaceae;g__Actinoplanes
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micromonosporaceae;g__Catellatospora
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micromonosporaceae;g__Micromonospora
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micromonosporaceae;g__Virgisporangium
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Mycobacteriaceae;g__Mycobacterium
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nakamurellaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nocardiaceae;g__Nocardia
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nocardiaceae;g__Rhodococcus
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nocardioidaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nocardioidaceae;g__Actinopolymorpha
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nocardioidaceae;g__Aeromicrobium
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nocardioidaceae;g__Friedmanniella
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nocardioidaceae;g__Kribbella
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nocardioidaceae;g__Marmoricola
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nocardioidaceae;g__Nocardioides
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nocardioidaceae;g__Pimelobacter
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nocardioidaceae;g__Propionicimonas

■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nocardiopsaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nocardiopsaceae;g__Nocardiopsis
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nocardiopsaceae;g__Thermobifida
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Promicromonosporaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Promicromonosporaceae;g__Cellulosimicrobium
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Promicromonosporaceae;g__Myceligenerans
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Promicromonosporaceae;g__Xylanimicrobium
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Propionibacteriaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Propionibacteriaceae;g__Aestuariimicrobium
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Propionibacteriaceae;g__Luteococcus
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Propionibacteriaceae;g__Propionibacterium
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Propionibacteriaceae;g__Tessaracoccus
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Pseudonocardiaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Pseudonocardiaceae;g__Actinoalloteichus
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Pseudonocardiaceae;g__Actinomycetospora
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Pseudonocardiaceae;g__Amycolatopsis
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Pseudonocardiaceae;g__Pseudonocardia
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Pseudonocardiaceae;g__Saccharomonospora
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Pseudonocardiaceae;g__Saccharopolyspora
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Sanguibacteraceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Sanguibacteraceae;g__Sanguibacter
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Streptomycetaceae;g__Streptacidiphilus
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Streptomycetaceae;g__Streptomyces
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Streptosporangiaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Streptosporangiaceae;g__Microbispora
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Streptosporangiaceae;g__Nonomuraea
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Thermomonosporaceae;g__Actinomadura
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Tsukamurellaceae;g__Tsukamurella
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Williamsiaceae;g__Williamsia
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Yaniellaceae;g__Yaniella
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Bifidobacteriales;f__Bifidobacteriaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Bifidobacteriales;f__Bifidobacteriaceae;g__Alloscardovia
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Bifidobacteriales;f__Bifidobacteriaceae;g__Gardnerella
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Bifidobacteriales;f__Bifidobacteriaceae;g__Parascardovia
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Bifidobacteriales;f__Bifidobacteriaceae;g__Scardovia
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__Coriobacteriaceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__Coriobacteriaceae;g__Adlercreutzia
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__Coriobacteriaceae;g__Atopobium
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__Coriobacteriaceae;g__Collinsella
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__Coriobacteriaceae;g__Eggerthella
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__Coriobacteriaceae;g__Olsenella
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__Coriobacteriaceae;g__Slackia
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__MC47;f__;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Rubrobacterales;f__Rubrobacteraceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Rubrobacterales;f__Rubrobacteraceae;g__Rubrobacter
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Solirubrobacterales;f__;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Solirubrobacterales;f__Patulibacteraceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Solirubrobacterales;f__Patulibacteraceae;g__Patulibacter
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Solirubrobacterales;f__Solirubrobacteraceae;g__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Solirubrobacterales;f__Solirubrobacteraceae;g__Solirubrobacter
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__koll13;f__;g__
■ k__Bacteria;p__BRC1;c__;o__;f__;g__
■ k__Bacteria;p__Bacteroidetes;c__;o__;f__;g__
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__;g__
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Marinilabiaceae;g__
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Dysgonomonas
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Odoribacter
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Paludibacter
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Parabacteroides
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Porphyromonas
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Tannerella
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__Alistipes
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__;f__;g__
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__;g__
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Cryomorphaceae;g__
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Cryomorphaceae;g__Fluviicola
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Aequorivita
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Capnocytophaga
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Chryseobacterium
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Elizabethkingia
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Flavobacterium
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Gillisia
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Haloanella
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Leeuwenhoekiella
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Maribacter
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Myroides
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Olleya
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Riemerella
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Robiginitalea
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Salegentibacter
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Salinimicrobium
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Sejongia
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Ulvibacter
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae;g__Wautersiella
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__;g__
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__;g__Chitinophaga
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__;g__Flavisolibacter
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__;g__Niastella
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__;g__Segetibacter
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Cyclobacteriaceae;g__
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Cyclobacteriaceae;g__Algoriphagus
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Cyclobacteriaceae;g__Aquiflexum
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Cyclobacteriaceae;g__Cyclobacterium
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Flammeovirgaceae;g__
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Flammeovirgaceae;g__Roseivirga
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Flexibacteraceae;g__
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Flexibacteraceae;g__Cytophaga
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Flexibacteraceae;g__Dyadobacter
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Flexibacteraceae;g__Emticicia
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Flexibacteraceae;g__Flectobacillus
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Flexibacteraceae;g__Hymenobacter
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Flexibacteraceae;g__Pontibacter
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Flexibacteraceae;g__Runella
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Flexibacteraceae;g__Spirosoma
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Saprospiraceae;g__
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Saprospiraceae;g__Candidatus Aquirestis
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Saprospiraceae;g__Haliscomenobacter
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Sphingobacteriaceae;g__
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Sphingobacteriaceae;g__Olivibacter
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Sphingobacteriaceae;g__Pedobacter
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Sphingobacteriaceae;g__Sphingobacterium
■ k__Bacteria;p__CCM11b;c__;o__;f__;g__
■ k__Bacteria;p__Caldithrix_KSB1;c__Caldithrixae;o__Caldithrixales;f__;g__
■ k__Bacteria;p__Caldithrix_KSB1;c__c5LKS36;o__;f__;g__
■ k__Bacteria;p__Chlamydiae;c__Chlamydiae (class);o__Chlamydiales;f__;g__
■ k__Bacteria;p__Chlamydiae;c__Chlamydiae (class);o__Chlamydiales;f__Chlamydiaceae;g__
■ k__Bacteria;p__Chlamydiae;c__Chlamydiae (class);o__Chlamydiales;f__Chlamydiaceae;g__Chlamydia
■ k__Bacteria;p__Chlamydiae;c__Chlamydiae (class);o__Chlamydiales;f__Parachlamydiaceae;g__
■ k__Bacteria;p__Chlamydiae;c__Chlamydiae (class);o__Chlamydiales;f__Parachlamydiaceae;g__Candidatus Protochlamydia
■ k__Bacteria;p__Chlamydiae;c__Chlamydiae (class);o__Chlamydiales;f__Rhabdochlamydiaceae;g__Candidatus Rhabdochlamydia
■ k__Bacteria;p__Chlamydiae;c__Chlamydiae (class);o__Chlamydiales;f__Simkaniaceae;g__
■ k__Bacteria;p__Chlamydiae;c__Chlamydiae (class);o__Chlamydiales;f__Waddliaceae;g__Waddlia
■ k__Bacteria;p__Chlorobi;c__;o__;f__;g__
■ k__Bacteria;p__Chlorobi;c__BSV19;o__;f__;g__
■ k__Bacteria;p__Chlorobi;c__OPB56;o__;f__;g__

■ k__Bacteria;p__Chlorobi;c__SJA-28;o__;f__;g__
■ k__Bacteria;p__Chlorobi;c__ZB1;o__;f__;g__
■ k__Bacteria;p__Chloroflexi;c__;o__;f__;g__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__;f__;g__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__A4b;f__;g__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__Anaerolineales;f__Anaerolinaceae;g__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__Anaerolineales;f__Anaerolinaceae;g__Anaerolinea
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__Anaerolineales;f__Anaerolinaceae;g__Levilinea
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__Anaerolineales;f__Anaerolinaceae;g__T78
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__Anaerolineales;f__Anaerolinaceae;g__WCHB1-05
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__Caldilineales;f__Caldilineaceae;g__Caldilinea
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__GCA004;f__;g__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__H39;f__;g__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__OPB11;f__;g__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__S0208;f__;g__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__SHA-20;f__;g__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__SJA-101;f__SHA-31;g__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__WCHB1-50;f__;g__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__envOPS12;f__;g__
■ k__Bacteria;p__Chloroflexi;c__Bljii12;o__;f__;g__
■ k__Bacteria;p__Chloroflexi;c__Chloroflexi (class);o__Roseiflexales;f__;g__
■ k__Bacteria;p__Chloroflexi;c__Chloroflexi (class);o__Roseiflexales;f__Kouleothrixaceae;g__
■ k__Bacteria;p__Chloroflexi;c__Chloroflexi (class);o__Roseiflexales;f__Kouleothrixaceae;g__Kouleothrix
■ k__Bacteria;p__Chloroflexi;c__Chloroflexi-4;o__;f__;g__
■ k__Bacteria;p__Chloroflexi;c__Ktedonobacteria;o__;f__;g__
■ k__Bacteria;p__Chloroflexi;c__SOGA31;o__;f__;g__
■ k__Bacteria;p__Chloroflexi;c__TK17;o__;f__;g__
■ k__Bacteria;p__Chloroflexi;c__Thermobacula;o__Thermobaculales;f__Thermobaculaceae;g__Thermobaculum
■ k__Bacteria;p__Chloroflexi;c__Thermomicrobia;o__HN1-15;f__;g__
■ k__Bacteria;p__Chloroflexi;c__Thermomicrobia;o__Thermomicrobiales;f__;g__
■ k__Bacteria;p__Cyanobacteria;c__;o__;f__;g__
■ k__Bacteria;p__Cyanobacteria;c__;o__Chroococcales;f__;g__
■ k__Bacteria;p__Cyanobacteria;c__;o__Chroococcales;f__;g__Cyanobium
■ k__Bacteria;p__Cyanobacteria;c__;o__Chroococcales;f__;g__Synechococcus
■ k__Bacteria;p__Cyanobacteria;c__;o__Nostocales;f__Nostocaceae;g__
■ k__Bacteria;p__Cyanobacteria;c__;o__Nostocales;f__Nostocaceae;g__Anabaena
■ k__Bacteria;p__Cyanobacteria;c__;o__Nostocales;f__Nostocaceae;g__Anabaenopsis
■ k__Bacteria;p__Cyanobacteria;c__;o__Nostocales;f__Nostocaceae;g__Cylindrospermopsis
■ k__Bacteria;p__Cyanobacteria;c__;o__Nostocales;f__Nostocaceae;g__Dolichospermum
■ k__Bacteria;p__Cyanobacteria;c__;o__Nostocales;f__Rivulariaceae;g__Calothrix
■ k__Bacteria;p__Cyanobacteria;c__;o__Nostocales;f__Scytonemataceae;g__Scytonema
■ k__Bacteria;p__Cyanobacteria;c__;o__Oscillatoriales;f__;g__
■ k__Bacteria;p__Cyanobacteria;c__;o__Oscillatoriales;f__;g__Leptolyngbya
■ k__Bacteria;p__Cyanobacteria;c__;o__Oscillatoriales;f__;g__Limnothrix
■ k__Bacteria;p__Cyanobacteria;c__;o__Oscillatoriales;f__;g__Microcoleus
■ k__Bacteria;p__Cyanobacteria;c__;o__Oscillatoriales;f__;g__Oscillatoria
■ k__Bacteria;p__Cyanobacteria;c__;o__Oscillatoriales;f__;g__Phormidium
■ k__Bacteria;p__Cyanobacteria;c__;o__Oscillatoriales;f__;g__Planktothricoides
■ k__Bacteria;p__Cyanobacteria;c__;o__Oscillatoriales;f__;g__Planktothrix
■ k__Bacteria;p__Cyanobacteria;c__;o__Oscillatoriales;f__;g__Pseudanabaena
■ k__Bacteria;p__Cyanobacteria;c__;o__Pleurocapsales;f__;g__
■ k__Bacteria;p__Cyanobacteria;c__;o__Pleurocapsales;f__;g__Cyanobacterium
■ k__Bacteria;p__Cyanobacteria;c__;o__Pleurocapsales;f__;g__Microcystis
■ k__Bacteria;p__Cyanobacteria;c__;o__Pleurocapsales;f__;g__Pleurocapsa
■ k__Bacteria;p__Cyanobacteria;c__;o__Pleurocapsales;f__;g__Snowella
■ k__Bacteria;p__Cyanobacteria;c__;o__Pleurocapsales;f__;g__Spirulina
■ k__Bacteria;p__Cyanobacteria;c__S15B-MN24;o__;f__;g__
■ k__Bacteria;p__Cyanobacteria;c__SM1D11;o__;f__;g__
■ k__Bacteria;p__Cyanobacteria;c__YS2;o__;f__;g__
■ k__Bacteria;p__Cyanobacteria;c__mle1-12;o__;f__;g__
■ k__Bacteria;p__Deferribacteres;c__Deferribacteres (class);o__Deferribacterales;f__Deferribacteraceae;g__Flexistipes
■ k__Bacteria;p__Elusimicrobia;c__Elusimicrobia (class);o__;f__;g__
■ k__Bacteria;p__Elusimicrobia;c__Elusimicrobia (class);o__Elusimicrobiales;f__;g__
■ k__Bacteria;p__Elusimicrobia;c__Elusimicrobia (class);o__Elusimicrobiales;f__Elusimicrobiaceae;g__
■ k__Bacteria;p__Elusimicrobia;c__Endomicrobia;o__;f__;g__
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__;f__;g__Thermicanus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__;g__
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__;g__Exiguobacterium
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__;g__Gemella
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Alicyclobacillaceae;g__Alicyclobacillus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Anoxybacillus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Bacillus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Geobacillus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Lentibacillus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Oceanobacillus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Sediminibacillus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Thalassobacillus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Virgibacillus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Paenibacillaceae;g__
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Paenibacillaceae;g__Ammoniphilus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Paenibacillaceae;g__Brevibacillus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Paenibacillaceae;g__Paenibacillus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Paenibacillaceae;g__Thermobacillus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__Bacillus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__Kurthia
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__Lysinibacillus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__Marinibacillus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__Paenisporosarcina
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__Planococcus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__Planomicrobium
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__Solibacillus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__Sporosarcina
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__Ureibacillus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__Viridibacillus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Jeotgalicoccus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Salinicoccus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Thermoactinomycetaceae;g__
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Turicibacter
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__;g__
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae;g__
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae;g__Aerococcus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae;g__Facklamia
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae;g__
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae;g__Alkalibacterium
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae;g__Alloiococcus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae;g__Carnobacterium
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae;g__Granulicatella
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae;g__Trichococcus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Enterococcaceae;g__
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Enterococcaceae;g__Enterococcus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Enterococcaceae;g__Tetragenococcus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Enterococcaceae;g__Vagococcus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Pediococcus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Leuconostocaceae;g__Leuconostoc
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Leuconostocaceae;g__Weissella
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Lactococcus
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Catabacteriaceae;g__
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__Caloramator
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__Clostridium
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__Sarcina
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XI. Incertae Sedis;g__
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XI. Incertae Sedis;g__Anaerococcus
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XI. Incertae Sedis;g__Finegoldia
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XI. Incertae Sedis;g__Helcococcus

■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XI. Incertae Sedis;g__Peptoniphilus
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XI. Incertae Sedis;g__Sedimentibacter
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XI. Incertae Sedis;g__Sporanaerobacter
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XII. Incertae Sedis;g__
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XII. Incertae Sedis;g__Fusibacter
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XIII. Incertae Sedis;g__
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XIII. Incertae Sedis;g__Anaerovorax
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XIII. Incertae Sedis;g__Eubacterium
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XIII. Incertae Sedis;g__Mogibacterium
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Dehalobacteriaceae;g__Dehalobacterium
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Eubacteriaceae;g__
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Eubacteriaceae;g__Acetobacterium
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Eubacteriaceae;g__Anaerofustis
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Eubacteriaceae;g__Pseudoramibacter
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Blautia
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Butyrivibrio
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Catonella
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Clostridium
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Coprococcus
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Dorea
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Eubacterium
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Johnsonella
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Lachnobacterium
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Lachnospira
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Moryella
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Oribacterium
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Pseudobutyrivibrio
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Roseburia
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Ruminococcus
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Shuttleworthia
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptococcaceae;g__
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptococcaceae;g__Desulfosporosinus
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptococcaceae;g__Desulfotomaculum
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptococcaceae;g__Pelotomaculum
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;g__Peptostreptococcus
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Anaerofilum
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Anaerotruncus
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Clostridium
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ethanoligenens
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Eubacterium
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Faecalibacterium
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Oscillospira
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcus
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Subdoligranulum
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Sulfobacillaceae;g__Thermaerobacter
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Syntrophomonadaceae;g__Syntrophomonas
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Thermoanaerobacterales Family III. Incertae Sedis;g__Thermoanaerobacterium
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Anaeroglobus
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Desulfosporomusa
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Megamonas
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Megasphaera
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Phascolarctobacterium
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Selenomonas
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Thermosinus
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Veillonella
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Desulfitobacterales;f__Desulfitobacteraceae;g__Desulfitobacter
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__MBA08;f__;g__
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Natranaerobiales;f__Anaerobrancaceae;g__KF-Gitt2-16
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Natranaerobiales;f__Natranaerobiaceae;g__Natranaerobius
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Natranaerobiales;f__YAB3B13;g__
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__OPB54;f__;g__
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Thermoanaerobacterales;f__Thermoanaerobacterales Family III. Incertae Sedis;g__Thermovenabulum
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Thermoanaerobacterales;f__Thermodesulfobiaceae;g__Coprothermobacter
■ k__Bacteria;p__Fusobacteria;c__Fusobacteria (class);o__Fusobacteriales;f__Fusobacteriaceae;g__Cetobacterium
■ k__Bacteria;p__Fusobacteria;c__Fusobacteria (class);o__Fusobacteriales;f__Fusobacteriaceae;g__Fusobacterium
■ k__Bacteria;p__Fusobacteria;c__Fusobacteria (class);o__Fusobacteriales;f__Fusobacteriaceae;g__J2-29
■ k__Bacteria;p__Fusobacteria;c__Fusobacteria (class);o__Fusobacteriales;f__Fusobacteriaceae;g__Leptotrichia
■ k__Bacteria;p__Fusobacteria;c__Fusobacteria (class);o__Fusobacteriales;f__Fusobacteriaceae;g__Sebaldella
■ k__Bacteria;p__Fusobacteria;c__Fusobacteria (class);o__Fusobacteriales;f__Fusobacteriaceae;g__Sneathia
■ k__Bacteria;p__Fusobacteria;c__Fusobacteria (class);o__Fusobacteriales;f__Fusobacteriaceae;g__Streptobacillus
■ k__Bacteria;p__Fusobacteria;c__Fusobacteria (class);o__Fusobacteriales;f__Fusobacteriaceae;g__u114
■ k__Bacteria;p__GN02;c__;o__;f__;g__
■ k__Bacteria;p__GN02;c__GKS2-174;o__;f__;g__
■ k__Bacteria;p__Gemmatimonadetes;c__Gemmatimonadetes (class);o__;f__;g__
■ k__Bacteria;p__Gemmatimonadetes;c__Gemmatimonadetes (class);o__Gemmatimonadales;f__;g__
■ k__Bacteria;p__Gemmatimonadetes;c__Gemmatimonadetes (class);o__Gemmatimonadales;f__Gemmatimonadaceae;g__
■ k__Bacteria;p__Gemmatimonadetes;c__Gemmatimonadetes (class);o__Gemmatimonadales;f__Gemmatimonadaceae;g__Gemmatimonas
■ k__Bacteria;p__Lentisphaerae;c__Lentisphaerae (class);o__Victivallales;f__;g__
■ k__Bacteria;p__Lentisphaerae;c__Lentisphaerae (class);o__Victivallales;f__Victivallaceae;g__
■ k__Bacteria;p__Lentisphaerae;c__Lentisphaerae (class);o__Z20;f__;g__
■ k__Bacteria;p__NC10;c__;o__;f__;g__
■ k__Bacteria;p__NKB19;c__;o__;f__;g__
■ k__Bacteria;p__Nitrospirae;c__Nitrospira (class);o__Nitrospirales;f__FW;g__4-29
■ k__Bacteria;p__Nitrospirae;c__Nitrospira (class);o__Nitrospirales;f__Nitrospiraceae;g__JG37-AG-70
■ k__Bacteria;p__Nitrospirae;c__Nitrospira (class);o__Nitrospirales;f__Nitrospiraceae;g__Nitrospira
■ k__Bacteria;p__Nitrospirae;c__Nitrospira (class);o__Nitrospirales;f__Thermodesulfovibrionaceae;g__HB118
■ k__Bacteria;p__OP10;c__CL500-48;o__;f__;g__
■ k__Bacteria;p__OP10;c__S1a-1H;o__;f__;g__
■ k__Bacteria;p__OP11;c__OP11-2;o__KD3-145;f__;g__
■ k__Bacteria;p__OP3;c__;o__;f__;g__
■ k__Bacteria;p__OP8;c__OP8;o__;f__;g__
■ k__Bacteria;p__Planctomycetes;c__Koll-18;o__;f__;g__
■ k__Bacteria;p__Planctomycetes;c__PW285;o__;f__;g__
■ k__Bacteria;p__Planctomycetes;c__Phycisphaerae;o__;f__;g__
■ k__Bacteria;p__Planctomycetes;c__Planctomycea;o__Gemmatales;f__Gemmataceae;g__
■ k__Bacteria;p__Planctomycetes;c__Planctomycea;o__Pirellulales;f__;g__
■ k__Bacteria;p__Planctomycetes;c__Planctomycea;o__Planctomycetales;f__Planctomycetaceae;g__Planctomyces
■ k__Bacteria;p__Planctomycetes;c__vadinHA49;o__;f__;g__
■ k__Bacteria;p__Proteobacteria;c__;o__;f__;g__
■ k__Bacteria;p__Proteobacteria;c__;o__;f__;g__Magnetococcus
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulobacteraceae;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulobacteraceae;g__Asticcacaulis
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulobacteraceae;g__Brevundimonas
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulobacteraceae;g__Caulobacter
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulobacteraceae;g__Mycoplana
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulobacteraceae;g__Phenylobacterium
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Aurantimonadaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Aurantimonadaceae;g__Aurantimonas
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bartonellaceae;g__Bartonella
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Beijerinckiaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Beijerinckiaceae;g__Beijerinckia
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae;g__Afipia
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae;g__Balneimonas
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae;g__Bosea
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae;g__Bradyrhizobium
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Brucellaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Brucellaceae;g__Ochrobactrum
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Brucellaceae;g__Pseudochrobactrum
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Hyphomicrobiaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Hyphomicrobiaceae;g__Devosia
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Hyphomicrobiaceae;g__Hyphomicrobium
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Hyphomicrobiaceae;g__Pedomicrobium
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Hyphomicrobiaceae;g__Rhodoplanes
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Methylobacteriaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Methylobacteriaceae;g__Methylobacterium
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Methylocystaceae;g__

■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Methylocystaceae;g__Methylocystis
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Methylocystaceae;g__Methylopila
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Methylocystaceae;g__Methylosinus
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Methylocystaceae;g__Pleomorphomonas
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Phyllobacteriaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Phyllobacteriaceae;g__Aquamicrobium
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Phyllobacteriaceae;g__Defluvibacter
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Phyllobacteriaceae;g__Mesorhizobium
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Phyllobacteriaceae;g__Nitratireductor
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Rhizobiaceae;g__Agrobacterium
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Rhizobiaceae;g__Candidatus Liberibacter
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Rhizobiaceae;g__Kaistia
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Rhizobiaceae;g__Mycoplana
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Rhizobiaceae;g__Rhizobium
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Rhodobiaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Xanthobacteraceae;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Xanthobacteraceae;g__Ancylobacter
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Xanthobacteraceae;g__Labrys
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Xanthobacteraceae;g__Xanthobacter
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Hyphomonadaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Hyphomonadaceae;g__Hyphomonas
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Amaricoccus
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Loktanella
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Paracoccus
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Rhodobaca
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Rhodobacter
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Rhodovulum
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Roseivivax
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Rubellimicrobium
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Sulfitobacter
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Thioclava
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Acetobacteraceae;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Acetobacteraceae;g__Roseococcus
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Acetobacteraceae;g__Roseomonas
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Rhodospirillaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Rhodospirillaceae;g__Azospirillum
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Rhodospirillaceae;g__Inquilinus
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Rhodospirillaceae;g__Magnetospirillum
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Rhodospirillaceae;g__Novispirillum
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Rhodospirillaceae;g__Phaeospirillum
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Rhodospirillaceae;g__Skermanella
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__;g__Caedibacter
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__;g__Candidatus Odyssella
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__;g__Candidatus Pelagibacter
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__Rickettsiaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__Rickettsiaceae;g__Rickettsia
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Erythrobacteraceae;g__Erythrobacter
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Erythrobacteraceae;g__Erythromicrobium
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Erythrobacteraceae;g__Lutibacterium
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Blastomonas
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Kaistobacter
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Novosphingobium
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingobium
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingomonas
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Sphingopyxis
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae;g__Zymomonas
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__;g__Aquabacterium
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__;g__Ideonella
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__;g__Leptothrix
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__;g__Methylibium
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__;g__Paucibacter
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__;g__Roseateles
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__;g__Rubrivivax
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__;g__Sphaerotilus
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__;g__Tepidimonas
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__;g__Thiomonas
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Achromobacter
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Alcaligenes
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Azohydromonas
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Oligella
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Pigmentiphaga
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Sutterella
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Burkholderia
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Candidatus Glomeribacter
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Chitinimonas
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Lautropia
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Limnobacter
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae;g__Polynucleobacter
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Acidovorax
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Alicycliphilus
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Aquamonas
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Brachymonas
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Caldimonas
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Comamonas
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Diaphorobacter
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Giesbergeria
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Hydrogenophaga
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Hylemonella
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Lampropedia
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Limnohabitans
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Polaromonas
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Ramlibacter
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Rhodoferax
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae;g__Schlegelella
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Collimonas
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Janthinobacterium
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Massilia
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Oxalobacter
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Gallionellales;f__Gallionellaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Gallionellales;f__Gallionellaceae;g__Gallionella
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Hydrogenophilales;f__Hydrogenophilaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Hydrogenophilales;f__Hydrogenophilaceae;g__Hydrogenophilus
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Hydrogenophilales;f__Hydrogenophilaceae;g__Thiobacillus
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Methylophilales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Methylophilales;f__Methylophilaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Methylophilales;f__Methylophilaceae;g__Methylophilus
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Methylophilales;f__Methylophilaceae;g__Methylotenera
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Methylophilales;f__Methylophilaceae;g__Methylovorus
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Aquitalea
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Chromobacterium
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Conchiformibius
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Deefgea
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Eikenella
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Kingella
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Microvirgula
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Neisseria
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Vitreoscilla
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae;g__Vogesella
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Nitrosomonadales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Nitrosomonadales;f__Nitrosomonadaceae;g__

■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Nitrosomonadales;f__Nitrosomonadaceae;g__Nitrosomonas
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Rhodocyclales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Rhodocyclales;f__Rhodocyclaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Rhodocyclales;f__Rhodocyclaceae;g__Dechloromonas
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Rhodocyclales;f__Rhodocyclaceae;g__Methyloversatilis
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Rhodocyclales;f__Rhodocyclaceae;g__Rhodocyclus
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Rhodocyclales;f__Rhodocyclaceae;g__Thauera
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Rhodocyclales;f__Rhodocyclaceae;g__Zoogloea
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__;f__Syntrophorhabdaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Bdellovibrionales;f__Bacteriovoracaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Bdellovibrionales;f__Bdellovibrionaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Bdellovibrionales;f__Bdellovibrionaceae;g__Bdellovibrio
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__CTD005-82B-02;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfobacterales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfobacterales;f__Desulfobulbaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfobacterales;f__Desulfobulbaceae;g__Desulfobulbus
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfobacterales;f__Desulfobulbaceae;g__Desulfocapsa
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfobacterales;f__Desulfobulbaceae;g__Desulfotalea
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfovibrionales;f__Desulfohalobiaceae;g__Desulfonatronovibrio
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfovibrionales;f__Desulfohalobiaceae;g__Desulfonauticus
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfovibrionales;f__Desulfomicrobiaceae;g__Desulfomicrobium
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfovibrionales;f__Desulfonatronumaceae;g__Desulfonatronum
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfovibrionales;f__Desulfovibrionaceae;g__Bilophila
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfovibrionales;f__Desulfovibrionaceae;g__Desulfovibrio
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfuromonadales;f__Desulfuromonadaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfuromonadales;f__Desulfuromonadaceae;g__Desulfuromonas
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfuromonadales;f__Geobacteraceae;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfuromonadales;f__Geobacteraceae;g__Geobacter
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Entotheonellales;f__Entotheonellaceae;g__Candidatus Entotheonella
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__MIZ46;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__Cystobacteraceae;g__Cystobacter
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__Haliangiaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__Myxococcaceae;g__Anaeromyxobacter
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__Myxococcaceae;g__Corallococcus
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__Nannocystaceae;g__Plesiocystis
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__Polyangiaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__Polyangiaceae;g__Chondromyces
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__NB1-j;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__NB1-j;f__JTB38;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__NB1-j;f__MND4;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__NB1-j;f__NB1-i;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__Desulfobacteraceae;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__Desulfobacteraceae;g__Desulfobacterium
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__Syntrophaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__Syntrophaceae;g__Desulfobacca
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__Syntrophaceae;g__Desulfomonile
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__Syntrophaceae;g__Syntrophus
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__Syntrophobacteraceae;g__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__Syntrophobacteraceae;g__Desulfacinum
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__Syntrophobacteraceae;g__Syntrophobacter
■ k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__;g__Sulfurovum
■ k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae;g__Arcobacter
■ k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae;g__Campylobacter
■ k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae;g__Sulfurospirillum
■ k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Helicobacteraceae;g__Helicobacter
■ k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Helicobacteraceae;g__Sulfuricurvum
■ k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Helicobacteraceae;g__Sulfurimonas
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__;f__;g__Kangiella
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Acidithiobacillales;f__Acidithiobacillaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Acidithiobacillales;f__Acidithiobacillaceae;g__Acidithiobacillus
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__Aeromonadaceae;g__Aeromonas
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__Aeromonadaceae;g__Oceanisphaera
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__Aeromonadaceae;g__Tolumonas
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__Succinivibrionaceae;g__Anaerobiospirillum
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__Succinivibrionaceae;g__Succinivibrio
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__;g__Alkalimonas
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Alteromonadaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Alteromonadaceae;g__Alishewanella
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Alteromonadaceae;g__Glaciecola
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Chromatiaceae;g__Rheinheimera
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Colwelliaceae;g__Colwellia
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Ferrimonadaceae;g__Ferrimonas
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Idiomarinaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Idiomarinaceae;g__Pseudidiomarina
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Pseudoalteromonadaceae;g__Pseudoalteromonas
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Psychromonadaceae;g__Psychromonas
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Shewanellaceae;g__Shewanella
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Cardiobacteriales;f__Cardiobacteriaceae;g__Cardiobacterium
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Cardiobacteriales;f__Cardiobacteriaceae;g__Dichelobacter
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Chromatiales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Chromatiales;f__Chromatiaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Chromatiales;f__Chromatiaceae;g__Allochromatium
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Chromatiales;f__Chromatiaceae;g__Chromatium
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Chromatiales;f__Chromatiaceae;g__Nitrosococcus
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Chromatiales;f__Chromatiaceae;g__Thiocapsa
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Chromatiales;f__Chromatiaceae;g__Thiococcus
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Chromatiales;f__Ectothiorhodospiraceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Chromatiales;f__Ectothiorhodospiraceae;g__Ectothiorhodospira
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Chromatiales;f__Halothiobacillaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Chromatiales;f__Halothiobacillaceae;g__Halothiobacillus
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Chromatiales;f__Sinobacteraceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Chromatiales;f__Sinobacteraceae;g__Steroidobacter
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Arsenophonus
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Averyella
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Brenneria
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Buchnera
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Cedecea
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Citrobacter
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Enterobacter
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Erwinia
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Escherichia
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Klebsiella
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Leclercia
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Pantoea
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Plesiomonas
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Proteus
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Providencia
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Rahnella
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Raoultella
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Salmonella
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Serratia
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Sodalis
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Trabulsiella
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Xenorhabdus
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Yersinia
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__Coxiellaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__Coxiellaceae;g__Aquicella
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__Coxiellaceae;g__Rickettsiella
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__Legionellaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__Legionellaceae;g__Legionella
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__Legionellaceae;g__Tatlockia
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Methylococcales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Methylococcales;f__Crenotrichaceae;g__Crenothrix
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Methylococcales;f__Methylococcaceae;g__

■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Methylococcales;f__Methylococcaceae;g__Methylomicrobium
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Methylococcales;f__Methylococcaceae;g__Methylomonas
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__;g__Kangiella
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Alcanivoracaceae;g__Alcanivorax
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Alteromonadaceae;g__Marinobacter
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Alteromonadaceae;g__Microbulbifer
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Halomonadaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Halomonadaceae;g__Candidatus Portiera
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Halomonadaceae;g__Chromohalobacter
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Halomonadaceae;g__Haererehalobacter
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Halomonadaceae;g__Halomonas
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Halomonadaceae;g__Kushneria
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Oceanospirillaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Oceanospirillaceae;g__Marinobacterium
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Oceanospirillaceae;g__Marinomonas
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Pseudomonadaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Pseudomonadaceae;g__Cellvibrio
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Pseudomonadaceae;g__Pseudomonas
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Actinobacillus
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Aggregatibacter
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Chelonobacter
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Gallibacterium
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Haemophilus
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Mannheimia
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Nicoletella
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Pasteurella
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Acinetobacter
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Alkanindiges
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Moraxella
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Psychrobacter
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiotrichales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiotrichales;f__Piscirickettsiaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiotrichales;f__Piscirickettsiaceae;g__Methylophaga
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiotrichales;f__Thiotrichaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiotrichales;f__Thiotrichaceae;g__Thiothrix
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Vibrionaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Vibrionaceae;g__Aliivibrio
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Vibrionaceae;g__Enterovibrio
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Vibrionaceae;g__Photobacterium
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Vibrionaceae;g__Vibrio
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__;g__Solimonas
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Sinobacteraceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Sinobacteraceae;g__Hydrocarboniphaga
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Sinobacteraceae;g__Nevskia
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Dokdonella
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Ignatzschineria
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Luteimonas
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Lysobacter
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Pseudoxanthomonas
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Rhodanobacter
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Stenotrophomonas
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae;g__Thermomonas
■ k__Bacteria;p__Proteobacteria;c__Zetaproteobacteria;o__Mariprofundales;f__Mariprofundaceae;g__Mariprofundus
■ k__Bacteria;p__SC3;c__;o__;f__;g__
■ k__Bacteria;p__SC4;c__KD3-113;o__;f__;g__
■ k__Bacteria;p__SM2F11;c__;o__;f__;g__
■ k__Bacteria;p__SPAM;c__;o__;f__;g__
■ k__Bacteria;p__Spirochaetes;c__;o__;f__;g__
■ k__Bacteria;p__Spirochaetes;c__GN05;o__;f__;g__
■ k__Bacteria;p__Spirochaetes;c__Leptospirae;o__Leptospirales;f__Leptospiraceae;g__
■ k__Bacteria;p__Spirochaetes;c__Leptospirae;o__Leptospirales;f__Leptospiraceae;g__SJA-88
■ k__Bacteria;p__Spirochaetes;c__Spirochaetes (class);o__Spirochaetales;f__Spirochaetaceae;g__SA-8
■ k__Bacteria;p__Spirochaetes;c__Spirochaetes (class);o__Spirochaetales;f__Spirochaetaceae;g__Spirochaeta
■ k__Bacteria;p__Spirochaetes;c__Spirochaetes (class);o__Spirochaetales;f__Spirochaetaceae;g__TM3
■ k__Bacteria;p__Spirochaetes;c__Spirochaetes (class);o__Spirochaetales;f__Spirochaetaceae;g__Treponema
■ k__Bacteria;p__Synergistetes;c__Synergistia;o__Synergistales;f__Aminiphilaceae;g__Aminiphilus
■ k__Bacteria;p__Synergistetes;c__Synergistia;o__Synergistales;f__Anaerobaculaceae;g__
■ k__Bacteria;p__TM6;c__;o__;f__;g__
■ k__Bacteria;p__TM7;c__;o__;f__;g__
■ k__Bacteria;p__TM7;c__MJK10;o__;f__;g__
■ k__Bacteria;p__TM7;c__TM7-1;o__;f__;g__
■ k__Bacteria;p__TM7;c__TM7-3;o__CW040;f__;g__
■ k__Bacteria;p__TM7;c__TM7-3;o__EW055;f__;g__
■ k__Bacteria;p__TM7;c__TM7-3;o__I025;f__;g__
■ k__Bacteria;p__Tenericutes;c__;o__;f__vadinHA31;g__
■ k__Bacteria;p__Tenericutes;c__;o__;f__vadinHA31;g__RFN20
■ k__Bacteria;p__Tenericutes;c__ML615J-28;o__;f__;g__
■ k__Bacteria;p__Tenericutes;c__Mollicutes;o__Anaeroplasmatales;f__Anaeroplasmataceae;g__Anaeroplasma
■ k__Bacteria;p__Tenericutes;c__Mollicutes;o__RF39;f__;g__
■ k__Bacteria;p__Thermi;c__Deinococci;o__Deinococcales;f__Deinococcaceae;g__Deinococcus
■ k__Bacteria;p__Thermi;c__Deinococci;o__Thermales;f__Thermaceae;g__Meiothermus
■ k__Bacteria;p__Thermi;c__Deinococci;o__Thermales;f__Thermaceae;g__Thermus
■ k__Bacteria;p__Thermodesulfobacteria;c__Thermodesulfobacteria (class);o__Thermodesulfobacteriales;f__Thermodesulfobacteriaceae;g__Thermodesulfatator
■ k__Bacteria;p__Verrucomicrobia;c__;o__;f__;g__
■ k__Bacteria;p__Verrucomicrobia;c__;o__Methylacidiphilales;f__;g__
■ k__Bacteria;p__Verrucomicrobia;c__Opitutae;o__;f__Opitutaceae;g__
■ k__Bacteria;p__Verrucomicrobia;c__Opitutae;o__;f__Opitutaceae;g__Opitutus
■ k__Bacteria;p__Verrucomicrobia;c__Opitutae;o__Puniceicoccales;f__Puniceicoccaceae;g__
■ k__Bacteria;p__Verrucomicrobia;c__Spartobacteria;o__;f__;g__Chthoniobacter
■ k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__;g__
■ k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Verrucomicrobia subdivision 3;g__
■ k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Verrucomicrobiaceae;g__
■ k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Verrucomicrobiaceae;g__Akkermansia
■ k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Verrucomicrobiaceae;g__Luteolibacter
■ k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Verrucomicrobiaceae;g__Prosthecobacter
■ k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Verrucomicrobiaceae;g__Verrucomicrobium
■ k__Bacteria;p__WPS-2;c__;o__;f__;g__
■ k__Bacteria;p__WS3;c__;o__;f__;g__
■ k__Bacteria;p__WS3;c__GN04;o__;f__;g__
■ k__Bacteria;p__WS3;c__PAUC34f;o__;f__;g__
■ k__Bacteria;p__WS3;c__PRR-12;o__;f__;g__
■ k__Bacteria;p__WS3;c__PRR-12;o__LCP-67;f__;g__
■ k__Bacteria;p__WS3;c__PRR-12;o__PBS-III-9;f__;g__
■ k__Bacteria;p__WS3;c__PRR-12;o__PRR-10;f__;g__
■ k__Bacteria;p__ZB2;c__;o__;f__;g__
■ k__Bacteria;p__ZB3;c__;o__;f__;g__

3.2 Providing supplementary taxonomic bar charts to test study variation

While total diversity of a study can be inferred through genus level taxonomy bar charts, it is important to test diversity at multiple taxonomic levels, to investigate different levels of divergence between microbial communities. For that reason, as mentioned previously, all bar charts are represented on two distinct taxonomic levels as observed in Figures 8-12. One of QIIME's advantageous features when estimating diversity, is the creation of taxonomy summary bar plots at all taxonomic levels (phylum, class, order, family, genus and species). QIIME's default parameters for taxa charts generation do not include analysis on species levels, thus avoiding errors in taxonomic assignment. Though parameters can be altered to allow species level analysis, for the purpose of this study this was not considered appropriate. While most bacteria inhabiting human vaginal microbiomes are well-established organisms, investigating microbiome composition through 16S rRNA data could prove inaccurate at species level. Small sequence changes (due to random evolutionary events) could not be distinguished between species in 16S rRNA reads, thus creating false groupings and not illustrating an accurate description of microbiome's variation.

Reviewing taxonomy plots at multiple taxonomic levels additionally provides qualitative control of the bacterial communities reported by Genus level bar taxonomies, as well as providing clearer information on the predominantly abundant organisms. Figure 8 unquestionably supports all previous diversity statements for study HIV. Both α and β diversities are significantly high, with large numbers of individual families assigned (Figure 8b), proportionately significant relative abundances for most families and low representation of monoclonal patients (Figure 8a). Additionally, it is crucial to point out that relative abundance between Figure 3 and Figure 8 are similar, therefore verifying accuracy of the reported abundant bacterial communities. This proves QIIME's appropriate approach when assigning OTUs and taxonomies.

Equally to study HIV, study CANDIDIASIS taxonomy plots at order taxonomic level (Figure 9), display identical diversity patterns to those illustrated at Genus level (Figure 4). Study CANDIDIASIS QIIME taxonomy charts were presented at both genus and Order level, as the diversity plots between Genus and Family levels illustrated similar levels of diversity. Unable to observe consistent bacterial patterns at Family level taxonomy charts of the study, Order level taxonomies were utilised for bar chart analysis. Examining Order level taxonomic composition through bar charts allowed clearer visualisation of the most abundant members of the microbial communities avoiding misperception by detailed intrapersonal diversity data. Therefore, it is concluded that study CANDIDIASIS consists a number of highly diverse samples with high relative abundances. However, by comparing the family taxa plots of study HIV and the order taxa plots from study CANDIDIASIS, it becomes apparent that study CANDIDIASIS does not contain as high of an α-diversity as study HIV (Figures 8 and 9). Study CANDIDIASIS contains fewer unique organisms with lower relative abundances. Thus the overall increased study diversity represented by the great numbers of unique individuals (Figure 9b) could be a result of the more in-depth sampling, with study HIV containing only 168 samples whereas study CANDIDIASIS containing 224 vaginal samples.

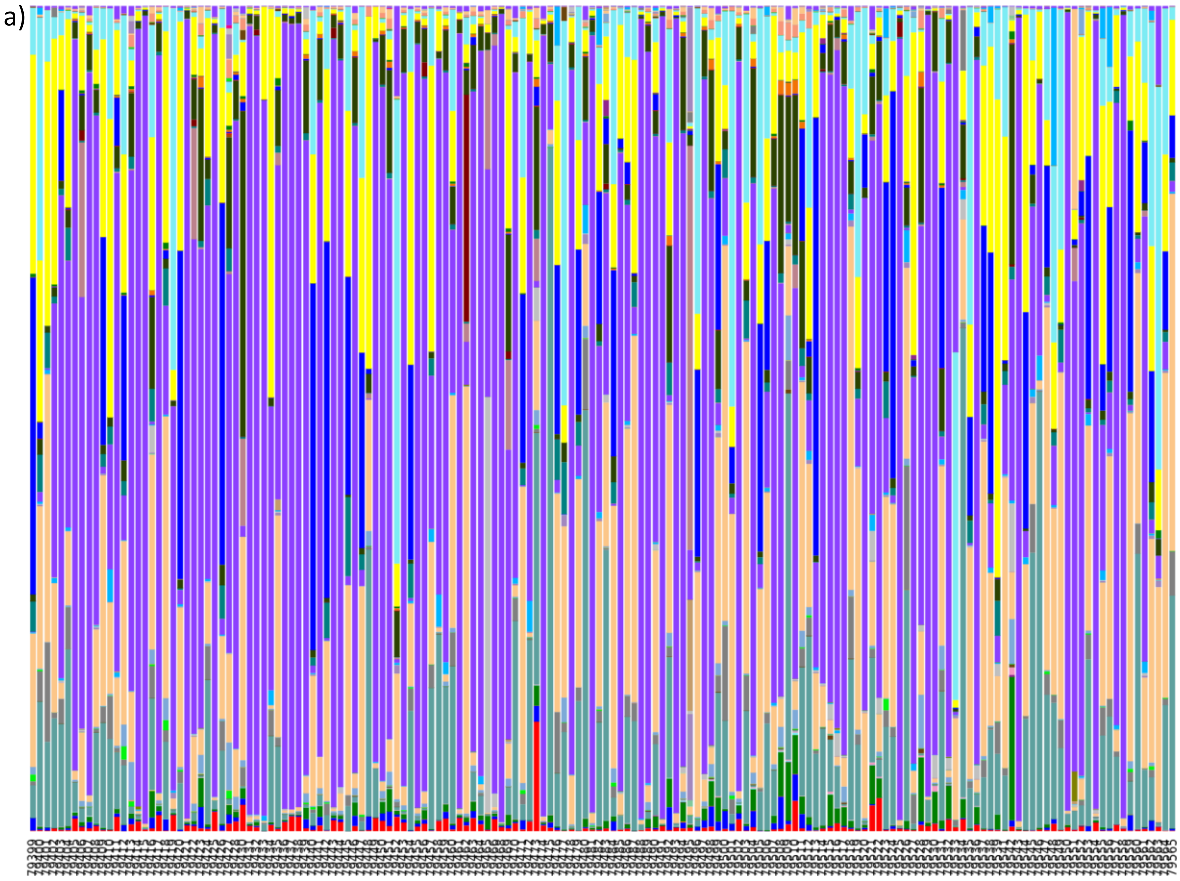


Figure 8: Study HIV family level taxonomy bar chart. Figure 8a displays individualistic families assigned within each study. The x-axis of the chart lists the sample IDs. Coloured bars represent the taxonomic families composing the samples. The size of the bars indicates the relative abundance of an organism within a single sample. *Lactobacillaceae*, *Bififobacteriaceae*, *Veillonellaceae, Prevotellaceae, Lachnospiraceae, Fusobacteria* are the most frequently present bars thus illustrating the most abundant organisms. Figure 8b lists the family identities of the coloured bars observed in Figure 8a.

b) ■ No blast hit;Other;Other;Other;Other

■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinomycetaceae

■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Brevibacteriaceae

■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Corynebacteriaceae

■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Intrasporangiaceae

■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Microbacteriaceae

■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micrococcaceae

■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Propionibacteriaceae

■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Bifidobacteriales;f__Bifidobacteriaceae

■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__

■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__Coriobacteriaceae

■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__

■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae

■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae

■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae

■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae

■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__

■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae

■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae

■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Erysipelotrichales;f__Erysipelotrichaceae

■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae

■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae

■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Enterococcaceae

■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae

■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae

■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__

■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae

■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XI. Incertae Sedis

■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XIII. Incertae Sedis

■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Eubacteriaceae

■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae

■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptococcaceae

■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae

■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae

■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae

■ k__Bacteria;p__Fusobacteria;c__Fusobacteria (class);o__Fusobacteriales;f__Fusobacteriaceae

■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulobacteraceae

■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae

■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae

■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae

■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__Desulfobacteraceae

■ k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae

■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__Succinivibrionaceae

■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae

■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Pseudomonadaceae

■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae

■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae

■ k__Bacteria;p__Synergistetes;c__Synergistia;o__Synergistales;f__Dethiosulfovibrionaceae

■ k__Bacteria;p__TM7;c__TM7-3;o__CW040;f__

■ k__Bacteria;p__TM7;c__TM7-3;o__I025;f__

■ k__Bacteria;p__Tenericutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae

■ k__Bacteria;p__Tenericutes;c__Mollicutes;o__Acholeplasmatales;f__Acholeplasmataceae

■ k__Bacteria;p__Tenericutes;c__Mollicutes;o__Mycoplasmatales;f__Mycoplasmataceae

■ k__Bacteria;p__Tenericutes;c__Mollicutes;o__RF39;f__

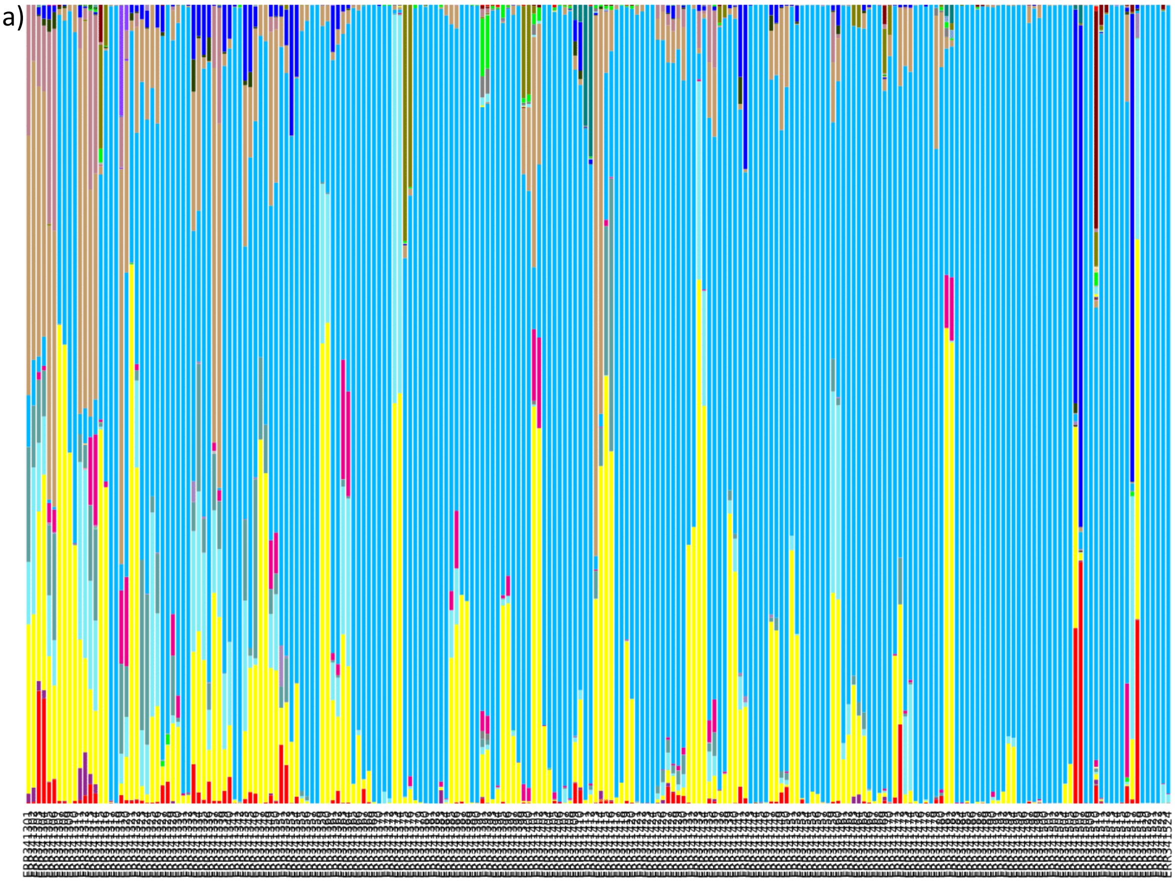


Figure 9: Study CANDIDIASIS order level taxonomy bar chart. Figure 9a illustrates a bar chart consisting of order level taxonomies identified in the samples. The sample IDs are displayed on the x-axis of the chart. Coloured bars represent individual organisms and the length of the bars depicts the relative abundancy of an organism within one sample. Bififobacteriales, Coriobacteriales, Lactobacillales, Bacillales, Clostridiales, Gemmatales and finally a non-identified blast search result appear as the most abundant organisms in the complete study. Figure 9b lists the order taxa identities of the coloured bars observed in Figure 9a.

b)

- 🟥 No blast hit;Other;Other;Other
- 🟦 k__Bacteria;p__Acidobacteria;c__Acidobacteria (class);o__Acidobacteriales
- 🟧 k__Bacteria;p__Acidobacteria;c__Chloracidobacteria;o__
- 🟩 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__
- 🟪 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales
- 🟨 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Bifidobacteriales
- 🟦 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales
- 🟪 k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Solirubrobacterales
- 🟩 k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales
- 🟫 k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales
- ⬜ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales
- 🟩 k__Bacteria;p__Chlamydiae;c__Chlamydiae (class);o__Chlamydiales
- 🟧 k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__
- 🟦 k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__H39
- 🟧 k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__S0208
- 🟩 k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__WCHB1-50
- 🟪 k__Bacteria;p__Chloroflexi;c__Ktedonobacteria;o__
- 🟨 k__Bacteria;p__Cyanobacteria;c__;o__
- ⬜ k__Bacteria;p__Cyanobacteria;c__;o__Oscillatoriales
- 🟪 k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales
- 🟦 k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales
- 🟫 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales
- 🟩 k__Bacteria;p__Firmicutes;c__Clostridia;o__Desulfitobacterales
- 🟪 k__Bacteria;p__Firmicutes;c__Clostridia;o__SHA-98
- 🟫 k__Bacteria;p__Fusobacteria;c__Fusobacteria (class);o__Fusobacteriales
- 🟩 k__Bacteria;p__Gemmatimonadetes;c__Gemmatimonadetes (class);o__
- 🟥 k__Bacteria;p__NKB19;c__;o__
- 🟩 k__Bacteria;p__Nitrospirae;c__Nitrospira (class);o__Nitrospirales
- 🟧 k__Bacteria;p__OP8;c__OP8;o__
- 🟥 k__Bacteria;p__Planctomycetes;c__Phycisphaerae;o__Phycisphaerales
- 🟦 k__Bacteria;p__Planctomycetes;c__Planctomycea;o__Gemmatales
- 🟧 k__Bacteria;p__Planctomycetes;c__Planctomycea;o__Pirellulales
- 🟩 k__Bacteria;p__Planctomycetes;c__agg27;o__OM190
- 🟪 k__Bacteria;p__Planctomycetes;c__vadinHA49;o__
- 🟨 k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales
- 🟦 k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales
- 🟪 k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales
- 🟩 k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales
- 🟫 k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales
- ⬜ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales
- 🟩 k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales
- 🟧 k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Hydrogenophilales
- 🟦 k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales
- 🟧 k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Rhodocyclales
- 🟩 k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Bdellovibrionales
- 🟪 k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfobacterales
- 🟨 k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfovibrionales
- ⬜ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfuromonadales
- 🟪 k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales
- 🟦 k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales
- 🟫 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales
- 🟩 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales
- 🟪 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales
- 🟫 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales
- 🟩 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales
- 🟥 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales
- 🟩 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiotrichales
- 🟧 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales
- 🟥 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales
- 🟦 k__Bacteria;p__Spirochaetes;c__SP_WWE1;o__
- 🟧 k__Bacteria;p__TM7;c__TM7-1;o__
- 🟩 k__Bacteria;p__Tenericutes;c__Erysipelotrichi;o__Erysipelotrichales
- 🟪 k__Bacteria;p__Tenericutes;c__Mollicutes;o__Mycoplasmatales
- 🟨 k__Bacteria;p__Tenericutes;c__Mollicutes;o__RF39
- 🟦 k__Bacteria;p__Thermi;c__Deinococci;o__Deinococcales
- 🟪 k__Bacteria;p__Verrucomicrobia;c__Spartobacteria;o__
- 🟩 k__Bacteria;p__WS3;c__PRR-12;o__

Study BV exemplifies an average vaginal microbiome diversity study. Both genus and order level taxonomy bar charts display high numbers of individual taxa with low relative abundance values (Figures 5 and 10). Despite the large numbers of identified taxa (Figure 10b), the order level taxonomy chart illustrates that study BV represents relatively lower β – diversity compared to the previously examined studies, as most samples are characterised by five highly abundant dominant organisms (orders: *Clostridiales, Lactobacillales, Fusobacteriales, Bacteroidales, Coriobacteriales*) (Figure 10a). The lack of monoclonal samples is noticeable in both taxonomy bar charts, thus representing a moderately increased α - diversity. Interestingly, study BV displays uncommon interpersonal variation patterns. Although numerous low taxon richness samples are enlisted they are not dominantly colonised by *Lactobacilli*, as portrayed in most healthy vaginal microbiomes, but are instead dominated by *Shuttleworthia* (k__Bacteria p__Firmicutes c__Clostridia o__Clostridiales f__Lachnospiraceae g__Shuttleworthia). As the study focused on characterising the impact of various sexually active groups on BV affected microbiomes [83], the samples contained various dysbiotic samples which is apparent from the lack of *Lactobacilli*. Females with dysbiotic vaginal microenvironments do not necessarily represent diseased individuals but rather samples of microbiomes with a “different than usual” bacterial community composition.

On the other hand, studies SV and HSV2 represent the lowest spectrum of diversity and bacteria richness in the present dataset of selected studies. By observing the family level taxonomy bar charts total study diversity can be seen, with both studies representing only three dominant organisms. Study HSV2 consists of mainly *Lactobacillaceae*, *Bififobacteriaceae* and *Coriobacteriaceae* (Figure 11), whereas study SV lists *Lactobacillaceae*, *Bififobacteriaceae* and *Veillonellaceae* as the three most abundant bacterial families (Figure 12). Therefore, both studies reveal low β – diversities, even though focused on different vaginal environments. Study HSV2 represents a higher α – diversity with fewer monoclonal samples, than the one observed in study SV (Figure 11). Although most HSV2 samples were colonised by fewer organisms, less monoclonal samples exist, increasing the levels of intrapersonal bacterial diversity. Instead, the majority of study SV samples consisted of a collection of monoclonal samples (excluding the male samples as previously mentioned). Female vaginal samples were dominated by *Lactobacillaceae* and only approximately 6 patients; out of 69 patients in total (12 samples – due to multiple sampling collections during the course of the study) were dominated by *Bifidobacteriaceae* or *Veillonellaceae* (samples SRR769992-9, SRR770006-7, SRR770014-15) (Figure 12). Vaginal samples in study SV contradicted expectations, since impact of sexual intercourse was expected to result in increased variation within the vaginal samples. Although fluctuations were reported in vaginal microbiome composition by Mandar et al. 2015, change remained at bacteria specific and intrapersonal levels, thus not affecting overall low level of diversity and coinciding with the present results.

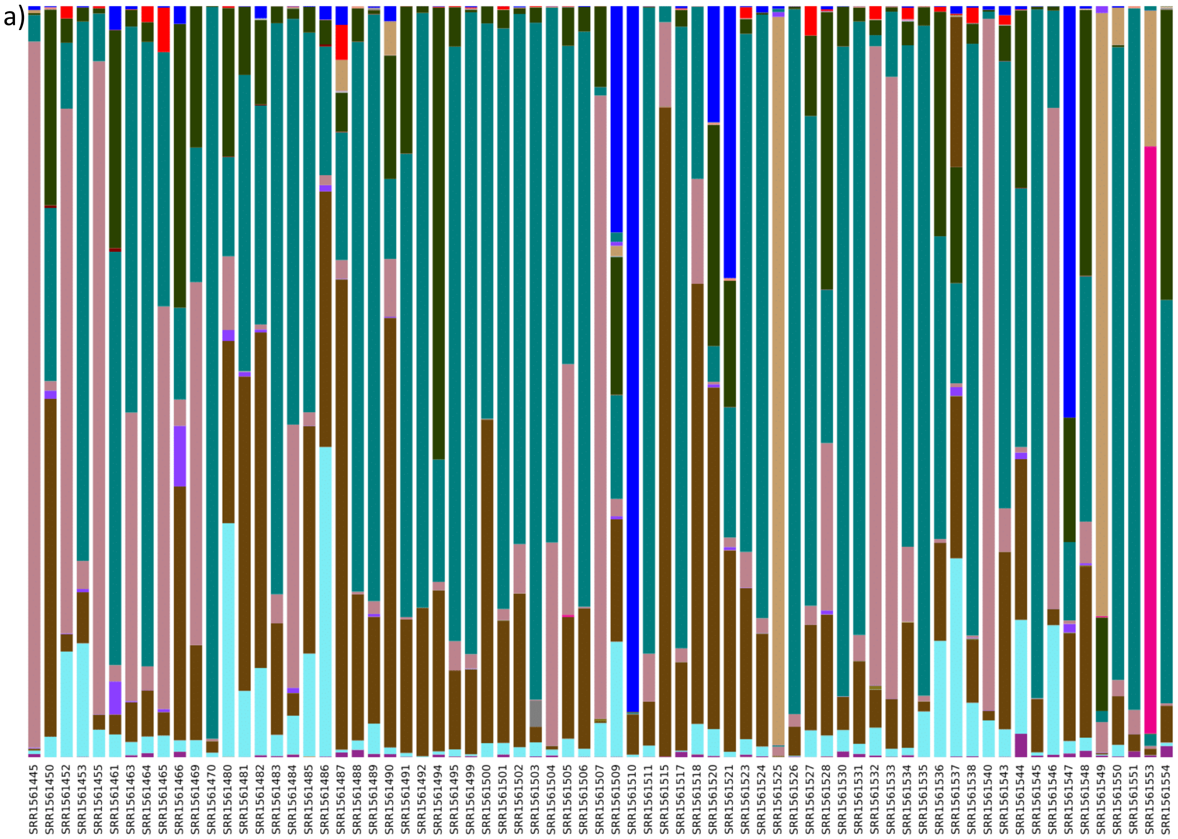


Figure 10: Study BV order level taxonomy bar chart. Figure 10a displays a taxonomic bar chart with order level taxonomies assigned from samples in the study. The sample IDs are listed on the x-axis. Coloured bars illustrate the different taxonomies contained within each sample and bar length represents abundance. *Clostridiales, Lactobacillales, Fusobacteriales, Bacteroidales, Coriobacteriales* appear as the most abundant bacteria within the study. Figure 10b lists the identities of the order level taxonomies displayed in Figure 10a.

b) 🟥 No blast hit;Other;Other;Other
🟦 k__Bacteria;p__;c__;o__
🟧 k__Bacteria;p__Acidobacteria;c__Sva0725;o__Sva0725
🟩 k__Bacteria;p__Actinobacteria;c__Acidimicrobiia;o__Acidimicrobiales
🟪 k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales
🟨 k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales
🟦 k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales
🟥 k__Bacteria;p__Actinobacteria;c__Thermoleophilia;o__Gaiellales
🟩 k__Bacteria;p__Armatimonadetes;c__Armatimonadia;o__Armatimonadales
🟫 k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales
⬛ k__Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales
🟩 k__Bacteria;p__Bacteroidetes;c__Sphingobacteriia;o__Sphingobacteriales
🟧 k__Bacteria;p__Bacteroidetes;c__[Rhodothermi];o__[Rhodothermales]
🟦 k__Bacteria;p__Bacteroidetes;c__[Saprospirae];o__[Saprospirales]
🟧 k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__GCA004
🟩 k__Bacteria;p__Chloroflexi;c__SL56;o__
🟪 k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Chlorophyta
🟨 k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Stramenopiles
⬜ k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta
🟥 k__Bacteria;p__Cyanobacteria;c__Nostocophycideae;o__Nostocales
🟦 k__Bacteria;p__Cyanobacteria;c__Oscillatoriophycideae;o__Chroococcales
🟫 k__Bacteria;p__Cyanobacteria;c__Synechococcophycideae;o__Pseudanabaenales
🟩 k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales
🟪 k__Bacteria;p__Firmicutes;c__Bacilli;o__Gemellales
🟫 k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales
🟩 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales
🟥 k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales
🟩 k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales
🟫 k__Bacteria;p__Gemmatimonadetes;c__Gemmatimonadetes;o__Gemmatimonadales
🟥 k__Bacteria;p__Nitrospirae;c__Nitrospira;o__Nitrospirales
🟦 k__Bacteria;p__OP3;c__koll11;o__
🟧 k__Bacteria;p__Planctomycetes;c__Phycisphaerae;o__Phycisphaerales
🟩 k__Bacteria;p__Proteobacteria;c__;o__
🟪 k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales
🟨 k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__RF32
🟦 k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales
🟥 k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales
🟩 k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales
🟫 k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales
⬛ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales
🟩 k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__
🟧 k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales
🟦 k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Methylophilales
🟧 k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales
🟩 k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Thiobacterales
🟪 k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfobacterales
🟨 k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales
⬜ k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales
🟥 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales
🟦 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales
🟫 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales
🟩 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales
🟪 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales
🟫 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiotrichales
🟩 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales
🟥 k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__[Marinicellales]
🟩 k__Bacteria;p__TM7;c__TM7-3;o__
🟫 k__Bacteria;p__TM7;c__TM7-3;o__CW040
🟥 k__Bacteria;p__TM7;c__TM7-3;o__I025
🟦 k__Bacteria;p__Tenericutes;c__Mollicutes;o__Mycoplasmatales
🟧 k__Bacteria;p__Tenericutes;c__Mollicutes;o__RF39
🟩 k__Bacteria;p__Verrucomicrobia;c__Opitutae;o__Opitutales
🟪 k__Bacteria;p__Verrucomicrobia;c__Opitutae;o__[Cerasicoccales]
🟨 k__Bacteria;p__WPS-2;c__;o__
🟦 k__Bacteria;p__[Thermi];c__Deinococci;o__Thermales

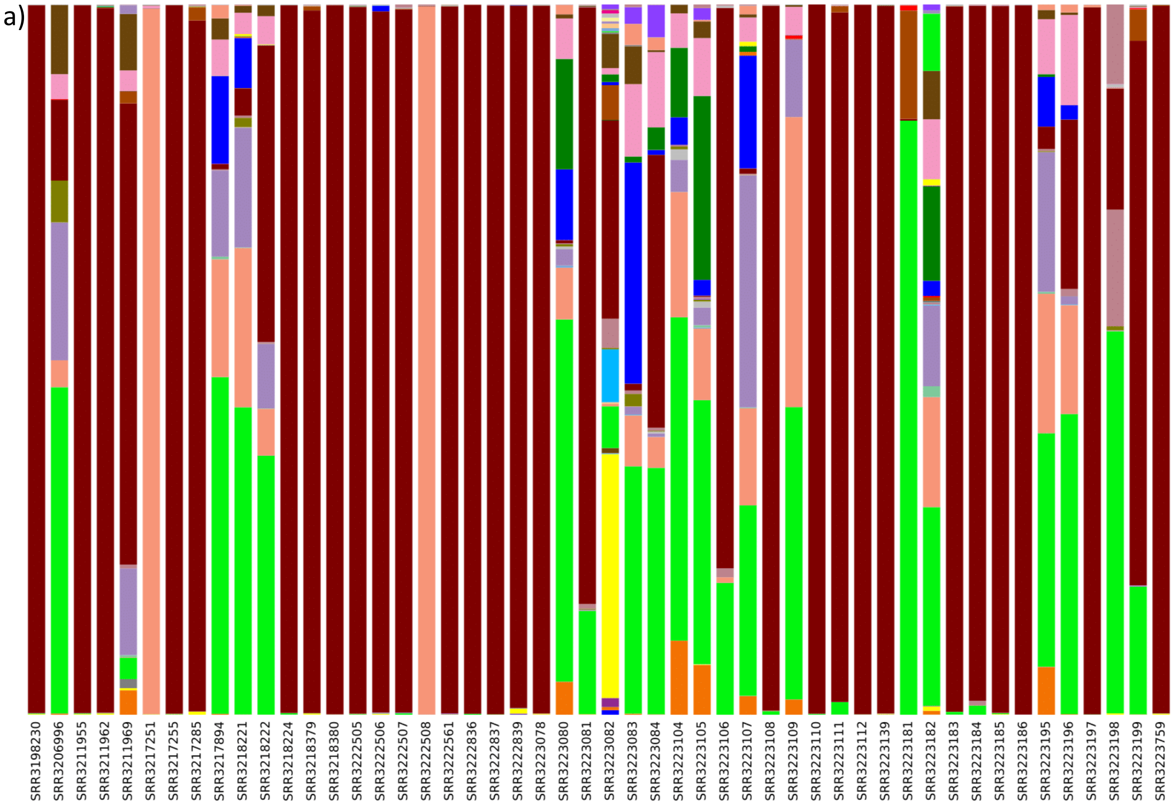


Figure 11: Study HSV2 family level taxonomy bar chart. Figure 11a demonstrates microbiome diversity within samples from the study. Sample IDs are listed on the x-axis. Colour bars signify the assigned family level taxonomies. The length of the bars suggests that *Lactobacillaceae, Bififobacteriaceae* and *Coriobacteriaceae* are the most abundant taxa in the complete study. Figure 11b displays the IDs at family level of the bars presented on Figure 11a.

b) 🟥 No blast hit;Other;Other;Other;Other
🟦 k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__
🟧 k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae
🟩 k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinosynnemataceae
🟪 k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Brevibacteriaceae
🟨 k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Dermabacteraceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Jonesiaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Microbacteriaceae
🟫 k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Micrococcaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Propionibacteriaceae
🟩 k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae
■ k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__S24-7
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__[Paraprevotellaceae]
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteriia;o__Flavobacteriales;f__[Weeksellaceae]
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Gemellales;f__Gemellaceae
🟪 k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Leuconostocaceae
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae
🟥 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__
🟦 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae
🟧 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Eubacteriaceae
🟩 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae
🟪 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptococcaceae
🟨 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Mogibacteriaceae]
🟫 k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae]
■ k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae
🟩 k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Fusobacteriaceae
■ k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Leptotrichiaceae
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae
■ k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Shewanellaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Halomonadaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae
🟪 k__Bacteria;p__TM7;c__TM7-3;o__I025;f__Rs-045
■ k__Bacteria;p__Tenericutes;c__Mollicutes;o__Mycoplasmatales;f__Mycoplasmataceae

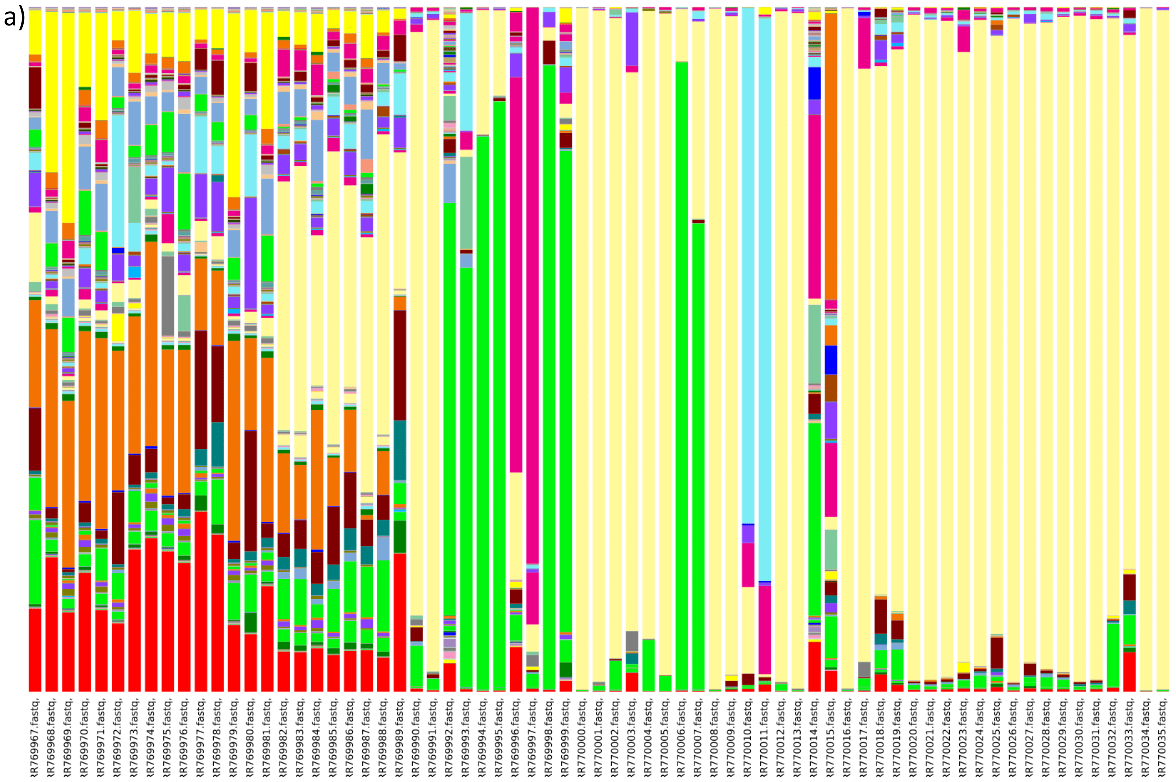


Figure 12: Study SV family level taxonomy bar chart. Figure 12a demonstrates family level taxonomy variation at both interpersonal and study levels. X-axis displays the sample IDs. Coloured bars represent the family level taxonomies present within each sample. Bar length illustrates taxon abundance. Samples ERR769967-ERR769989 are seminal samples and they display significantly more diverse communities. Seminal samples illustrate *Flavobacterium*, *Lactobacili* and *Acinetobacter* as the most dominant organisms. Vaginal samples suggest *Lactobacillaceae*, *Bififobacteriaceae* and *Veillonellaceae* as the most dominant organisms. Figure 12b displays the identities of the family level taxonomies illustrated as coloured bars in Figure 12a.

b)
■ No blast hit;Other;Other;Other;Other
■ k__Bacteria;p__;c__;o__;f__
■ k__Bacteria;p__ABY1_OD1;c__;o__;f__
■ k__Bacteria;p__AD3;c__ABS-6;o__;f__
■ k__Bacteria;p__AD3;c__JG37-AG-4;o__;f__
■ k__Bacteria;p__Acidobacteria;c__;o__;f__
■ k__Bacteria;p__Acidobacteria;c__;o__;f__Koribacteraceae
■ k__Bacteria;p__Acidobacteria;c__Acidobacteria (class);o__Acidobacteriales;f__
■ k__Bacteria;p__Acidobacteria;c__Acidobacteria (class);o__Acidobacteriales;f__Acidobacteriaceae
■ k__Bacteria;p__Acidobacteria;c__Acidobacteria-5;o__;f__
■ k__Bacteria;p__Acidobacteria;c__Chloracidobacteria;o__;f__
■ k__Bacteria;p__Acidobacteria;c__Holophagae;o__Holophagales;f__Holophagaceae
■ k__Bacteria;p__Acidobacteria;c__MVS-40;o__;f__
■ k__Bacteria;p__Acidobacteria;c__OS-K;o__;f__
■ k__Bacteria;p__Acidobacteria;c__PAUC37f;o__;f__
■ k__Bacteria;p__Acidobacteria;c__RB25;o__;f__
■ k__Bacteria;p__Acidobacteria;c__Solibacteres;o__Solibacterales;f__Solibacteraceae
■ k__Bacteria;p__Acidobacteria;c__Sva0725;o__;f__
■ k__Bacteria;p__Acidobacteria;c__iii1-8;o__;f__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__;f__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__0319-7L14;f__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Acidimicrobiales;f__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Acidimicrobiales;f__AKIW874
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Acidimicrobiales;f__CL500-29
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Acidimicrobiales;f__EB1017
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Acidimicrobiales;f__Iamiaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Acidimicrobiales;f__JdFBGBact
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Acidimicrobiales;f__Microthrixaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Acidimicrobiales;f__SC3-41
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Acidimicrobiales;f__wb1_P06
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__ACK-M1
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinomycetaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinopolysporaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinospicaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Actinosynnemataceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Bogoriellaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Brevibacteriaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Catenulisporaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Cellulomonadaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Corynebacteriaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Cryptosporangiaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Dermabacteraceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Dermacoccaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Dietziaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Frankiaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Geodermatophilaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Glycomycetaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Gordoniaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Intrasporangiaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Kineosporiaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Microbacteriaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micrococcaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Micromonosporaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Mycobacteriaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nakamurellaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nocardiaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nocardioidaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Nocardiopsaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Promicromonosporaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Propionibacteriaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Pseudonocardiaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Sanguibacteraceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Streptomycetaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Streptosporangiaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Thermomonosporaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Tsukamurellaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Williamsiaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Actinomycetales;f__Yaniellaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Bifidobacteriales;f__Bifidobacteriaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Coriobacteriales;f__Coriobacteriaceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__MC47;f__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Rubrobacterales;f__Rubrobacteraceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Solirubrobacterales;f__
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Solirubrobacterales;f__Patulibacteraceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Solirubrobacterales;f__Solirubrobacteraceae
■ k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__koll13;f__
■ k__Bacteria;p__BRC1;c__;o__;f__
■ k__Bacteria;p__Bacteroidetes;c__;o__;f__
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Marinilabiaceae
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae
■ k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__;f__
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Cryomorphaceae
■ k__Bacteria;p__Bacteroidetes;c__Flavobacteria;o__Flavobacteriales;f__Flavobacteriaceae
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Cyclobacteriaceae
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Flammeovirgaceae
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Flexibacteraceae
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Saprospiraceae
■ k__Bacteria;p__Bacteroidetes;c__Sphingobacteria;o__Sphingobacteriales;f__Sphingobacteriaceae
■ k__Bacteria;p__CCM11b;c__;o__;f__
■ k__Bacteria;p__Caldithrix_KSB1;c__Caldithrixae;o__Caldithrixales;f__
■ k__Bacteria;p__Caldithrix_KSB1;c__c5LKS36;o__;f__
■ k__Bacteria;p__Chlamydiae;c__Chlamydiae (class);o__Chlamydiales;f__
■ k__Bacteria;p__Chlamydiae;c__Chlamydiae (class);o__Chlamydiales;f__Chlamydiaceae
■ k__Bacteria;p__Chlamydiae;c__Chlamydiae (class);o__Chlamydiales;f__Parachlamydiaceae
■ k__Bacteria;p__Chlamydiae;c__Chlamydiae (class);o__Chlamydiales;f__Rhabdochlamydiaceae
■ k__Bacteria;p__Chlamydiae;c__Chlamydiae (class);o__Chlamydiales;f__Simkaniaceae
■ k__Bacteria;p__Chlamydiae;c__Chlamydiae (class);o__Chlamydiales;f__Waddliaceae
■ k__Bacteria;p__Chlorobi;c__;o__;f__
■ k__Bacteria;p__Chlorobi;c__BSV19;o__;f__
■ k__Bacteria;p__Chlorobi;c__OPB56;o__;f__
■ k__Bacteria;p__Chlorobi;c__SJA-28;o__;f__
■ k__Bacteria;p__Chlorobi;c__ZB1;o__;f__
■ k__Bacteria;p__Chloroflexi;c__;o__;f__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__;f__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__A4b;f__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__Anaerolineales;f__Anaerolinaceae
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__Caldilineales;f__Caldilineaceae
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__GCA004;f__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__H39;f__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__OPB11;f__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__S0208;f__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__SHA-20;f__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__SJA-101;f__SHA-31
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__WCHB1-50;f__
■ k__Bacteria;p__Chloroflexi;c__Anaerolineae;o__envOPS12;f__
■ k__Bacteria;p__Chloroflexi;c__Bljii12;o__;f__
■ k__Bacteria;p__Chloroflexi;c__Chloroflexi (class);o__Roseiflexales;f__
■ k__Bacteria;p__Chloroflexi;c__Chloroflexi (class);o__Roseiflexales;f__Kouleothrixaceae
■ k__Bacteria;p__Chloroflexi;c__Chloroflexi-4;o__;f__
■ k__Bacteria;p__Chloroflexi;c__Ktedonobacteria;o__;f__
■ k__Bacteria;p__Chloroflexi;c__SOGA31;o__;f__
■ k__Bacteria;p__Chloroflexi;c__TK17;o__;f__
■ k__Bacteria;p__Chloroflexi;c__Thermobacula;o__Thermobaculales;f__Thermobaculaceae
■ k__Bacteria;p__Chloroflexi;c__Thermomicrobia;o__HN1-15;f__
■ k__Bacteria;p__Chloroflexi;c__Thermomicrobia;o__Thermomicrobiales;f__
■ k__Bacteria;p__Cyanobacteria;c__;o__;f__
■ k__Bacteria;p__Cyanobacteria;c__;o__Chroococcales;f__
■ k__Bacteria;p__Cyanobacteria;c__;o__Nostocales;f__Nostocaceae
■ k__Bacteria;p__Cyanobacteria;c__;o__Nostocales;f__Rivulariaceae
■ k__Bacteria;p__Cyanobacteria;c__;o__Nostocales;f__Scytonemataceae
■ k__Bacteria;p__Cyanobacteria;c__;o__Oscillatoriales;f__
■ k__Bacteria;p__Cyanobacteria;c__;o__Pleurocapsales;f__
■ k__Bacteria;p__Cyanobacteria;c__S15B-MN24;o__;f__
■ k__Bacteria;p__Cyanobacteria;c__SM1D11;o__;f__
■ k__Bacteria;p__Cyanobacteria;c__YS2;o__;f__
■ k__Bacteria;p__Cyanobacteria;c__mle1-12;o__;f__
■ k__Bacteria;p__Deferribacteres;c__Deferribacteres (class);o__Deferribacterales;f__Deferribacteraceae
■ k__Bacteria;p__Elusimicrobia;c__Elusimicrobia (class);o__;f__
■ k__Bacteria;p__Elusimicrobia;c__Elusimicrobia (class);o__Elusimicrobiales;f__
■ k__Bacteria;p__Elusimicrobia;c__Elusimicrobia (class);o__Elusimicrobiales;f__Elusimicrobiaceae
■ k__Bacteria;p__Elusimicrobia;c__Endomicrobia;o__;f__
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__;f__
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Alicyclobacillaceae
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Paenibacillaceae
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Thermoactinomycetaceae
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Erysipelotrichales;f__Erysipelotrichaceae
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Aerococcaceae
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Enterococcaceae
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Leuconostocaceae
■ k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Catabacteriaceae
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XI. Incertae Sedis
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XII. Incertae Sedis
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XIII. Incertae Sedis
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Dehalobacteriaceae
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Eubacteriaceae
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptococcaceae
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Sulfobacillaceae
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Syntrophomonadaceae
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Thermoanaerobacterales Family III. Incertae Sedis
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Desulfitobacterales;f__Desulfitobacteraceae
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__MBA08;f__
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Natranaerobiales;f__Anaerobrancaceae
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Natranaerobiales;f__Natranaerobiaceae
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Natranaerobiales;f__YAB3B13
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__OPB54;f__

■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Thermoanaerobacterales;f__Thermoanaerobacterales Family III. Incertae Sedis
■ k__Bacteria;p__Firmicutes;c__Clostridia;o__Thermoanaerobacterales;f__Thermodesulfobiaceae
■ k__Bacteria;p__Fusobacteria;c__Fusobacteria (class);o__Fusobacteriales;f__Fusobacteriaceae
■ k__Bacteria;p__GN02;c__;o__;f__
■ k__Bacteria;p__GN02;c__GKS2-174;o__;f__
■ k__Bacteria;p__Gemmatimonadetes;c__Gemmatimonadetes (class);o__;f__
■ k__Bacteria;p__Gemmatimonadetes;c__Gemmatimonadetes (class);o__Gemmatimonadales;f__
■ k__Bacteria;p__Gemmatimonadetes;c__Gemmatimonadetes (class);o__Gemmatimonadales;f__Gemmatimonadaceae
■ k__Bacteria;p__Lentisphaerae;c__Lentisphaerae (class);o__Victivallales;f__
■ k__Bacteria;p__Lentisphaerae;c__Lentisphaerae (class);o__Victivallales;f__Victivallaceae
■ k__Bacteria;p__Lentisphaerae;c__Lentisphaerae (class);o__Z20;f__
■ k__Bacteria;p__NC10;c__;o__;f__
■ k__Bacteria;p__NKB19;c__;o__;f__
■ k__Bacteria;p__Nitrospirae;c__Nitrospira (class);o__Nitrospirales;f__FW
■ k__Bacteria;p__Nitrospirae;c__Nitrospira (class);o__Nitrospirales;f__Nitrospiraceae
■ k__Bacteria;p__Nitrospirae;c__Nitrospira (class);o__Nitrospirales;f__Thermodesulfovibrionaceae
■ k__Bacteria;p__OP10;c__CL500-48;o__;f__
■ k__Bacteria;p__OP10;c__S1a-1H;o__;f__
■ k__Bacteria;p__OP11;c__OP11-2;o__KD3-145;f__
■ k__Bacteria;p__OP3;c__;o__;f__
■ k__Bacteria;p__OP8;c__OP8;o__;f__
■ k__Bacteria;p__Planctomycetes;c__Koll-18;o__;f__
■ k__Bacteria;p__Planctomycetes;c__PW285;o__;f__
■ k__Bacteria;p__Planctomycetes;c__Phycisphaerae;o__;f__
■ k__Bacteria;p__Planctomycetes;c__Planctomycea;o__Gemmatales;f__Gemmataceae
■ k__Bacteria;p__Planctomycetes;c__Planctomycea;o__Pirellulales;f__
■ k__Bacteria;p__Planctomycetes;c__Planctomycea;o__Planctomycetales;f__Planctomycetaceae
■ k__Bacteria;p__Planctomycetes;c__vadinHA49;o__;f__
■ k__Bacteria;p__Proteobacteria;c__;o__;f__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__;f__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulobacteraceae
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Aurantimonadaceae
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bartonellaceae
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Beijerinckiaceae
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Bradyrhizobiaceae
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Brucellaceae
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Hyphomicrobiaceae
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Methylobacteriaceae
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Methylocystaceae
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Phyllobacteriaceae
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Rhizobiaceae
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Rhodobiaceae
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Xanthobacteraceae
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Hyphomonadaceae
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Acetobacteraceae
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodospirillales;f__Rhodospirillaceae
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rickettsiales;f__Rickettsiaceae
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Erythrobacteraceae
■ k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphingomonadaceae
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__;f__
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Burkholderiaceae
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Comamonadaceae
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Gallionellales;f__Gallionellaceae
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Hydrogenophilales;f__Hydrogenophilaceae
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Methylophilales;f__
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Methylophilales;f__Methylophilaceae
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Neisseriales;f__Neisseriaceae
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Nitrosomonadales;f__
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Nitrosomonadales;f__Nitrosomonadaceae
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Rhodocyclales;f__
■ k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Rhodocyclales;f__Rhodocyclaceae
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__;f__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__;f__Syntrophorhabdaceae
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Bdellovibrionales;f__Bacteriovoracaceae
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Bdellovibrionales;f__Bdellovibrionaceae
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__CTD005-82B-02;f__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfobacterales;f__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfobacterales;f__Desulfobulbaceae
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfovibrionales;f__Desulfohalobiaceae
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfovibrionales;f__Desulfomicrobiaceae
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfovibrionales;f__Desulfonatronumaceae
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfovibrionales;f__Desulfovibrionaceae
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfuromonadales;f__Desulfuromonadaceae
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfuromonadales;f__Geobacteraceae
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Entotheonellales;f__Entotheonellaceae
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__MIZ46;f__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__Cystobacteraceae
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__Haliangiaceae
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__Myxococcaceae
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__Nannocystaceae
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Myxococcales;f__Polyangiaceae
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__NB1-j;f__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__NB1-j;f__JTB38
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__NB1-j;f__MND4
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__NB1-j;f__NB1-i
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__Desulfobacteraceae
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__Syntrophaceae
■ k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Syntrophobacterales;f__Syntrophobacteraceae
■ k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__;f__
■ k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__
■ k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae
■ k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Helicobacteraceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__;f__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Acidithiobacillales;f__Acidithiobacillaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__Aeromonadaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__Succinivibrionaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Alteromonadaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Chromatiaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Colwelliaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Ferrimonadaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Idiomarinaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Pseudoalteromonadaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Psychromonadaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Alteromonadales;f__Shewanellaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Cardiobacteriales;f__Cardiobacteriaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Chromatiales;f__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Chromatiales;f__Chromatiaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Chromatiales;f__Ectothiorhodospiraceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Chromatiales;f__Halothiobacillaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Chromatiales;f__Sinobacteraceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__Coxiellaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Legionellales;f__Legionellaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Methylococcales;f__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Methylococcales;f__Crenotrichaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Methylococcales;f__Methylococcaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Alcanivoracaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Alteromonadaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Halomonadaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Oceanospirillaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Oceanospirillales;f__Pseudomonadaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiotrichales;f__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiotrichales;f__Piscirickettsiaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Thiotrichales;f__Thiotrichaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Vibrionaceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Sinobacteraceae
■ k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae
■ k__Bacteria;p__Proteobacteria;c__Zetaproteobacteria;o__Mariprofundales;f__Mariprofundaceae
■ k__Bacteria;p__SC3;c__;o__;f__
■ k__Bacteria;p__SC4;c__KD3-113;o__;f__
■ k__Bacteria;p__SM2F11;c__;o__;f__
■ k__Bacteria;p__SPAM;c__;o__;f__
■ k__Bacteria;p__Spirochaetes;c__;o__;f__
■ k__Bacteria;p__Spirochaetes;c__GN05;o__;f__
■ k__Bacteria;p__Spirochaetes;c__Leptospirae;o__Leptospirales;f__Leptospiraceae
■ k__Bacteria;p__Spirochaetes;c__Spirochaetes (class);o__Spirochaetales;f__Spirochaetaceae
■ k__Bacteria;p__Synergistetes;c__Synergistia;o__Synergistales;f__Aminiphilaceae
■ k__Bacteria;p__Synergistetes;c__Synergistia;o__Synergistales;f__Anaerobaculaceae
■ k__Bacteria;p__TM6;c__;o__;f__
■ k__Bacteria;p__TM7;c__;o__;f__
■ k__Bacteria;p__TM7;c__MJK10;o__;f__
■ k__Bacteria;p__TM7;c__TM7-1;o__;f__
■ k__Bacteria;p__TM7;c__TM7-3;o__CW040;f__
■ k__Bacteria;p__TM7;c__TM7-3;o__EW055;f__
■ k__Bacteria;p__TM7;c__TM7-3;o__I025;f__
■ k__Bacteria;p__Tenericutes;c__;o__;f__vadinHA31
■ k__Bacteria;p__Tenericutes;c__ML615J-28;o__;f__
■ k__Bacteria;p__Tenericutes;c__Mollicutes;o__Anaeroplasmatales;f__Anaeroplasmataceae
■ k__Bacteria;p__Tenericutes;c__Mollicutes;o__RF39;f__
■ k__Bacteria;p__Thermi;c__Deinococci;o__Deinococcales;f__Deinococcaceae
■ k__Bacteria;p__Thermi;c__Deinococci;o__Thermales;f__Thermaceae
■ k__Bacteria;p__Thermodesulfobacteria;c__Thermodesulfobacteria (class);o__Thermodesulfobacteriales;f__Thermodesulfobacteriaceae
■ k__Bacteria;p__Verrucomicrobia;c__;o__;f__
■ k__Bacteria;p__Verrucomicrobia;c__;o__Methylacidiphilales;f__
■ k__Bacteria;p__Verrucomicrobia;c__Opitutae;o__;f__Opitutaceae
■ k__Bacteria;p__Verrucomicrobia;c__Opitutae;o__Puniceicoccales;f__Puniceicoccaceae
■ k__Bacteria;p__Verrucomicrobia;c__Spartobacteria;o__;f__
■ k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__
■ k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Verrucomicrobia subdivision 3
■ k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Verrucomicrobiaceae
■ k__Bacteria;p__WPS-2;c__;o__;f__
■ k__Bacteria;p__WS3;c__;o__;f__
■ k__Bacteria;p__WS3;c__GN04;o__;f__
■ k__Bacteria;p__WS3;c__PAUC34f;o__;f__
■ k__Bacteria;p__WS3;c__PRR-12;o__;f__
■ k__Bacteria;p__WS3;c__PRR-12;o__LCP-67;f__
■ k__Bacteria;p__WS3;c__PRR-12;o__PBS-III-9;f__
■ k__Bacteria;p__WS3;c__PRR-12;o__PRR-10;f__
■ k__Bacteria;p__ZB2;c__;o__;f__
■ k__Bacteria;p__ZB3;c__;o__;f__

3.3 Criticising limitations of taxonomic bar charts generated in QIIME

Diversity analyses allow investigation of microbiome's composition as well as permit insight into microbiome variation between healthy and diseased states. Although advantageous for comparisons of variations between studies, diversity analyses have limitations for examining microbiome interactions. It is crucial to state that due to the lack of metadata information, on the sequence sample reads provided in most studies, the intrapersonal variation presented cannot be linked to medical condition. Thus only speculations of the microbiome state based on microbiome composition were possible. For that reason, additional analyses need to be developed to further support the evidence illustrated by the diversity analyses.

To ensure accuracy of the bar charts and quantify the results observed, the average abundance data were further examined. Although bar charts provide a straight-forward method for visualising the most abundant taxa, it is only accurately representative for intrapersonal variation and not total study bacterial abundance. A homogenous sample would include one long mono-colour bar representing the most abundant organism in that particular sample. However, the length of that bar would not be proportional to the absolute abundances of other samples, only the relative abundance. In other words, it is challenging to quantify abundance of an organism depending on bar length and almost unfeasible to calculate collective organism colonisation abundance of a complete study. For that reason, Table 2 was created from the relative abundance values originating from .biom OTU files. Table 2 represents lists of the 10 most abundant bacteria present in each study followed by percentage total abundance in all SRR samples of the study.

Table 2 displays different dominant microbial communities than the ones previously observed on the bar charts. Therefore, demonstrating the problem with bar taxonomies being exclusively used to study composition and variation in-between different studies. Due to study HIV having a relatively high diversity, genus taxonomy charts proved challenging in identifying the most dominant organism with accuracy. However, family level taxa bar charts pointed six commonly present bacteria (*Lactobacillaceae*, *Bififobacteriaceae*, *Veillonellaceae, Prevotellaceae, Lachnospiraceae, Fusobacteria*) (Figure 8). Although useful information, it is not possible to detect an order of dominance by observing the bar charts. Once focused on the bacterial abundances within a study, *Lactobacillus, Prevotella, Gardnerella, Shuttleworthia, Sneathia* and *Dialister* were quantifiably classified as the top 6 most abundant organisms, supportive of the results observed in the bar charts (Table 2).

Both HIV and CANDIDIASIS studies represent a diverse system with high numbers of identified taxonomies making it difficult to predict the most abundant organisms through bar length assessment. The order level taxonomy for study CANDIDIASIS (Figure 9), suggested 7 most abundant taxa (*Bififobacteriales*, *Coriobacteriales, Lactobacillales, Bacillales, Clostridiales, Gemmatales* and finally *Nitrospirales*). Examining the abundance values within the study, Table 2 summarises *Lactobacillus, Gardnerella, Streptococcus, Atopobium, Nitrospirales (family level taxonomy), Prevotella, Gemella* as the first 7 most dominant bacterial genera. Although the data from the abundance table (Table 2) and the order level bar chart (Figure 9) appear different, they are in fact the same with Table 2 listing the correct taxonomies. This is due to the difficulty in interpreting abundances through taxonomic bar charts.

	ERP017263		SRP045868		ERP003902		ERP009682		SRP071021	
Rank Order	**Bacteria**	**Percentage Abundances**	**Bacteria**	**Percentage Abundance**	**Bacteria**	**Percentage Abundance**	**Bacteria**	**Percentage Abundance**	**Bacteria**	**Percentage Abundance**
1	*Lactobacillus*	40.32%	*Shuttleworthia*	30.03%	*Lactobacillus*	79.78%	*Lactobacillus*	54.64%	*Lactobacillus*	61.60%
2	***Prevotella***	18.24%	***Prevotella***	19.12%	*Gardnerella*	6.19%	*Gardnerella*	16.17%	*Gardnerella*	14.80%
3	*Gardnerella*	11.15%	*Lactobacillus*	13.11%	*Streptococcus*	4.18%	*Veillonella*	5.41%	*Atopobium*	8.18%
4	*Shuttleworthia*	8.81%	*Sneathia*	11.62%	*Atopobium*	2.85%	*Flavobacterium*	3.74%	***Prevotella***	3.15%
5	*Sneathia*	8.57%	*Megasphaera*	8.76%	*Nitrospirales (genus ID not characterised)*	1.25%	*Streptococcus*	3.72%	*Clostridium*	2.07%
6	***Dialister***	2.25%	*Atopobium*	4.71%	***Prevotella***	1.01%	*Acinetobacter*	2.24%	*Shuttleworthia*	2.02%
7	*Atopobium*	1.55%	*Parvimonas*	3.21%	*Gemella*	0.81%	***Prevotella***	2.24%	*Bifidobacterium*	1.51%
8	*Finegoldia*	0.82%	***Dialister***	2.68%	*Peptostreptococcus*	0.69%	*Pseudomonas*	1.13%	*Megasphaera*	1.34%
9	*Streptococcus*	0.81%	*Streptococcus*	1.36%	***Dialister***	0.64%	*Corynebacterium*	0.99%	*Mobiluncus*	0.84%
10	*Anaerococcus*	0.74%	*Mycoplasma*	1.12%	*Sneathia*	0.54%	***Dialister***	0.65%	***Dialister***	0.75%

Table 2: Total percentage bacteria abundances per study. Table 2 lists the top 10 most abundant bacteria present within each study. The percentages were calculated form the corresponding OTU tables for each study. The table illustrates multiple shared bacteria between the studies (such as *Lactobacillus*, *Dialister*, *Prevotella, Gardnerella* and *Atopobium*). *Lactobacilli* are not as dominantly abundant as expected, representing only a 3-fold abundance increase to the next most abundant organism

QIIME performs taxonomic assignment by blasting the sequences provided against a database. If no results in the database match the minimum length, e-value and percentage requirements QIIME catalogues them under “No blast hit;Other;Other;Other**”.** Such data points are not sufficient to affect the overall results of the analysis as all remaining taxonomies consist of high quality matches representing expected organisms in the vaginal community. The non-identified blast matches were removed from the abundance data to avoid over representation of organisms. Consequently, QIIME proves a useful tool for taxonomy assessment in 16S rRNA data, even though encountering issues with Blast analysis.

Studies BV, SV and HSV2 as previously mentioned represent lower levels of diversity. The order level taxonomy chart for study BV assisted in the identification of five highly abundant organisms (*Clostridiales, Lactobacillales, Fusobacteriales, Bacteroidales, Coriobacteriales*) (Figure 10). Interestingly, coinciding not only with the abundance percentages (Table 2), but also with the results presented by Muzny’s study [83]. They reported *Lactobacillus, Lachnospiraceae*, *Prevotella*, and *Sneathia* as the four most dominant taxa, corresponding with the genus level *Shuttleworthia, Prevotella, Lactobacillus, Sneathia* observed in the taxonomic bar charts created by QIIME (Figure 5). Thus certifying the accuracy and suitability of the methodology performed.

Additionally, study SV once exclusively focused on vaginal samples displayed low diversity throughout with most samples appearing homogenous. A family level taxonomy chart revealed *Flavobacterium*, *Lactobacillus* and *Acinetobacter* as the most common organisms (Figure 12). However, by analysing the total abundance data of the study, *Lactobacillus, Gardnerella, Veillonella, Flavobacterium* and *Streptococcus* surfaced as highly present bacteria (Table 2). *Lactobacillus* and *Gardnerella* were the most commonly dominant organisms, followed by the atypical bacteria *Veillonella* and equally abundant *Flavobacterium* and *Streptococcus*. Unlike all previous analysed studies, study SV does not entirely match the results presented by Mandar et al. 2015. Male samples partially contradict the results presented in bar chart analysis created in this project, whilst Mandar et al. 2015 predict *Corynebacterium* as one of the three most abundant bacteria and not *Acinetobacter* as reported here. This could be a result of green genes human vaginal database used as a reference for the blast and taxonomy assessment analysis performed through QIIME. Although this exposes flaws with the pipeline followed for this study, seminal samples were not relevant to the focus of this project. Moreover, the degree of divergence between results was not substantial, thus could be excluded from the analysis. Nevertheless, the results from the vaginal samples observed in the bar charts were consistent with the abundance table (Table 2) and the reported results from Mandar et al. 2015. In conclusion, vaginal samples from study SV were colonised by *Lactobacillus, Gardnerella, Veillonella, Flavobacterium* and *Streptococcus* with the most dominant organisms being *Lactobacillus* and *Gardnerella* (Figure 12 and Table 2).

Finally, study HSV2 consisted of mostly homogenous patients dominated by *Lactobacilli*, with few more diverse samples observed in Figure 11. The family level taxonomy chart illustrated three typically abundant taxa: *Lactobacillaceae, Bififobacteriaceae* and *Coriobacteriaceae*. The results followed the same community patterns and were further supported by the percentage abundance data, observed in Table 2.

Lastly, Table 2 not only listed the most abundant bacteria present in each study, but interestingly revealed five shared dominant organisms (within the first ten most abundant bacteria within each study) between most of the five selected studies. *Lactobacillus*, as expected, was shared between all five studies, however despite expectations was not consistently exceedingly dominant, with a 5-fold average increase dominance over the rest of the organisms composing the microbiomes. Interestingly, *Dialister* and *Prevotella* were the next consistently abundant bacteria present in all five studies. *Gardnerella*, and *Atopobium* were equally present in most studies with high percentage abundances. Numerous other key organisms were shared between studies, like *Streptococcus* with relatively high abundances. However, abundance data do not reveal information on interactions between members of the microbiome, and thus we undertook clustering analysis to explore this further.

4. INVESTIGATING MICROBIOME COMMUNITY STRUCTURES VIA CLUSTERING ANALYSIS

Clustering analyses were implemented, with the aim to investigate interactions between vaginal microbiomes and their members, in order to characterise any potential links between bacteria belonging to the vaginal microbiome. Clustering analysis allows visualisation of interactions between members of a microbiome by establishing proximity between data. Clustering as discussed previously, can be achieved through various algorithms, which differ in defining the concept of a "cluster". Hierarchical clustering was utilised as the most advantageous algorithm, for the purpose of this analysis. Hierarchical clustering defines sample similarity through dendrograms, where the proximity of each leaf demonstrates the degree of similarity or dissimilarity. Hierarchical clustering is commonly presented alongside a heatmap, thus representing the complete data matrix of a study. Another common way of visualising data clustering is through Principal Component Analysis (PCA). PCA is a statistical model allowing visualisation of multi-dimensional data into the "principal components", demonstrating patterns of variance between them. Both methods effectively display interaction patterns as well as variance and similarities between microbiomes and their members. However, results are more reliable if both approaches are utilised in parallel, as the two technique provide complementary information.

4.1 Principal component analysis 3D plot designed through QIIME

Principal component analysis permits visualisation of a complete study's data matrix as well as the similarities in microbiome composition between samples of the study. The first approach in utilising PCA, was through QIIME's core_diversity_analyses.py script. One of the commands in QIIME's core_diversity_analyses.py script, make_emperor.py creates a 3D PCA graph consisting of the first three principal components displaying variation between samples. In QIIME's PCA plots, close proximity clustering/ grouping between samples signifies sample similarity dependant on bacterial composition and abundance. QIIME assessed principal components by estimating Bray-Curtis dissimilarity index of a study's complete abundance data matrix. As mentioned previously, Bray-Curtis assesses dissimilarity by generating a distance matrix with values ranging from 0 to 1. Zero values signify samples with equivalent bacteria abundances and 1 values represent samples that do not share any microbiome abundance similarities [98].

3D PCA plots were generated for all selected studies (HIV, HSV2, BV, CANDIDIASIS, SV) in order to investigate microbiome differences, as well as identify potential patterns between patients, giving us insight to microbiome conditions. Figures 13 – 17 display QIIME's 3D PCA plots for all studies, with the x-axis representing the first principal component (PC1) spanning the highest amount of bacterial compositional variation; the y- axis demonstrating the second principal component (PC2) illustrating the second highest bacterial compositional variance; and finally the z-axis signifying the third principal component (PC3) classifying the third highest percentage variance in bacterial composition. It is crucial to criticise that most of the variation will be driven by the high abundance in *Lactobacilli*, thus raising questions on clustering efficiency as the majority of human vaginal microbiomes are typically dominated by *Lactobacilli*. Through PCA the data matrices are reformatted in a 3 dimensional plane with commonly 2 dimensions covering the majority of the bacterial composition variation between samples, thus allowing easier pattern visualisation. Therefore, universally, if samples appear clustered in close proximity, samples are composed with matching bacteria abundances. Throughout this chapter various clusters will be discussed, however unlike clusters presented through k-means analysis, PCA illustrates subjective groupings. Consequently, all clusters presented in Figures 13-17 are ambiguous and only selected to illustrate variation and resemblances within samples of the studies.

Figure 13 displays a 3D PCA plot for study HIV with x, y and z axis representing the first 3 principal components. The axis percentages illustrate the percentage of bacterial composition variation covered by each corresponding principal component. PCA components 1, 2 and 3 explain about 57% of the total variance in the complete study, thus ensuring significance of the grouping patterns observed. Evidently, principal component one and two explain most variation, illustrated by the high variation percentages (PC1 = 28.69, PC2 = 21.65). As most variation is explained through the first two principal components, most clustering patterns can be visualised at a two dimensional plane. The PCA plot illustrates 3 distinct clusters (clusters A, B and C in Figure 13) with the majority of samples belonging to cluster A. Lacking metadata descriptions, it is not possible to link clusters with microbiome medical conditions. However, it is evident that samples fall in three classifications depending on their composition.

3D PCA plots offer advantages on visualising microbiome community structures and identifying similarities within samples of a specific study. In this case, Figure 13 illustrates three major groupings, which are based on bacterial composition. It is speculated that cluster formation is driven from samples composition dominant in *Lactobacillus, Prevotella and Gardnerella*, the three most abundant genera in this study. Cluster A represents the largest cluster, consisting of the most samples, thus *Lactobacilli* could be the driving source of this cluster. An alternative hypothesis is that the samples are clustered according to having several shared species, rather than each cluster being driven by a single taxon. Distinguishing the true causation behind QIIME's PCA sample clustering, proved challenging and time consuming as each sample point had to be investigated for its bacterial composition. Instead hierarchical clustering along with heatmap charts, offered a simpler approach to identifying similarities in sample composition within a study. Hierarchical clustering will be discussed in detail in section 4.2. Though, PCA is not self-sufficient to support conclusive evidence on microbiome interactions and structure, composition similarities were reported between patients, thus suggesting possible bacteria interactions and shared microbiome structures. Succeeding, focus should shift to investigating correlations within the selected studies for this project.

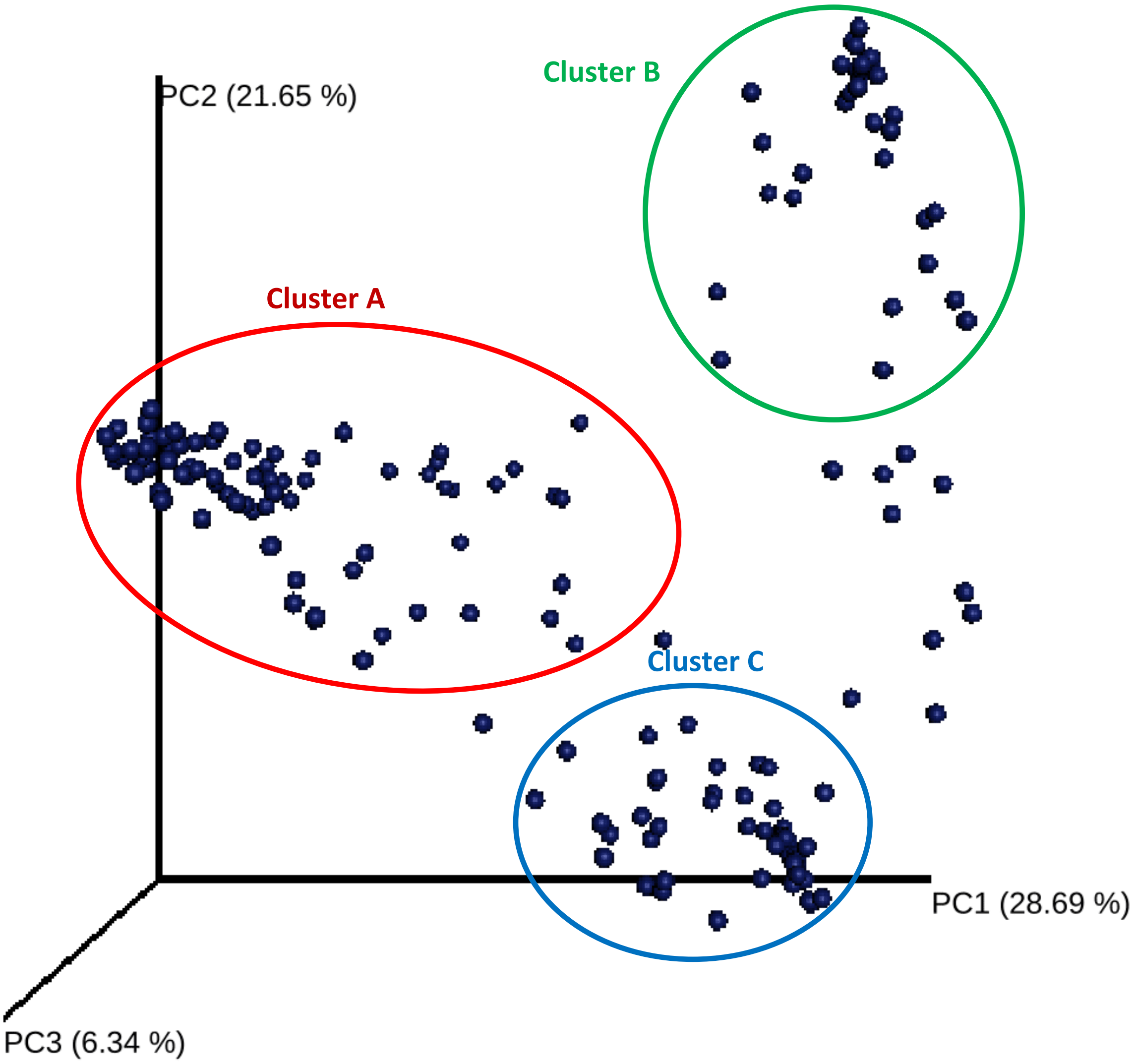


Figure 13: 3D Principal Component Plot for study HIV. Samples from study HIV plotted against the first three principal components displaying the highest amount of variance between sample bacterial compositions. Three suggestive sample clusters are formed and illustrated by the red, blue and green circles. Cluster A demonstrates the most populated cluster.

Figure 14 represents QIIME's designed PCA plot for study CANDIDIASIS. Unlike study HIV the principal component axes portray low percentages of explained sample composition variance, with only 10% of the total study variation being covered by the 3D PCA plot. The second (PC2) and third (PC3) components consist of almost equally low variation percentages, thus clustering patterns can only be visualised in a three dimensional plane. Although well-defined clusters are not easy to distinguish, the total abundance table (Table 2) raises composition factors, which could support the PCA patterns. Almost 80% of the total bacterial abundance in all present study samples, consists of *Lactobacilli*, with the remainder organisms present in very low abundances. Thus Figure 14 displays numerous small clusters with various sub-clusters within them, making patterns hard to detect. Low PC percentages illustrate the multidimensionality of the data matrix, due to the high in-depth sampling and high study bacteria richness, thus each principal component would explain a small percentage of the total study diversity. The majority of sample variation would be expressed by the abundance of *Lactobacilli* as they dominate most samples within the study.

Interestingly, it is worth restating that study CANDIDIASIS was analysed through the first pipeline, discussed in the "QIIME toolkit" section 2.2. Consequently, it could be argued that the lack of pattern observed in the PCA plot could be a result of the difference in the de-multiplexing approach. Although both pipelines designed for this project were effective and accurate for microbiome assessment, de-multiplexing is a complex factor and a step which could affect generation of the abundance matrix and in consequence PCA plot. In conclusion the PCA for study CANDIDIASIS illustrates an uncharacteristic arrangement of microbiome structure compared to the other studies, and supports the results presented on both taxonomy bar charts (Figures 4 and 9) which were indicative of high sample variation.

QIIME's script enabled PCA analysis for study BV, presented in Figure 15. As mentioned previously, the PCA plot illustrates three axes representing the first three principal components covering the percentage of bacteria abundance variation in the study. The total amount of study variation explained through PCA is more than 50%, ensuring confidence in the clustering patterns. Principal component one and two axes cover most of the sample variation within the study, with PC3 displaying a significantly lower percentage of the variation (6.58 %). This is additionally supported by the ability to distinctly visualise the complete PCA array in two dimensions. Interestingly, study BV depicts similar clustering and patterns to Figure 13 for study HIV. Three evident clusters can be observed with cluster A, consisting of most samples. As discussed previously, the most abundant species could be driving the majority of the compositional similarities between samples. Table 2 enlists *Shuttleworthia, Prevotella and Lactobacillus* as the three most abundant organisms. Therefore, can be proposed that the highly dominant *Shuttleworthia* is driving the establishment of cluster A. However, the identity of clusters B and C in Figure 15 could be a result of a combination of shared taxa between samples. Therefore, although composition and organism variance differs, clustering patterns appear similar between studies HIV and BV, possibly indicating shared stable microbiome community structures.

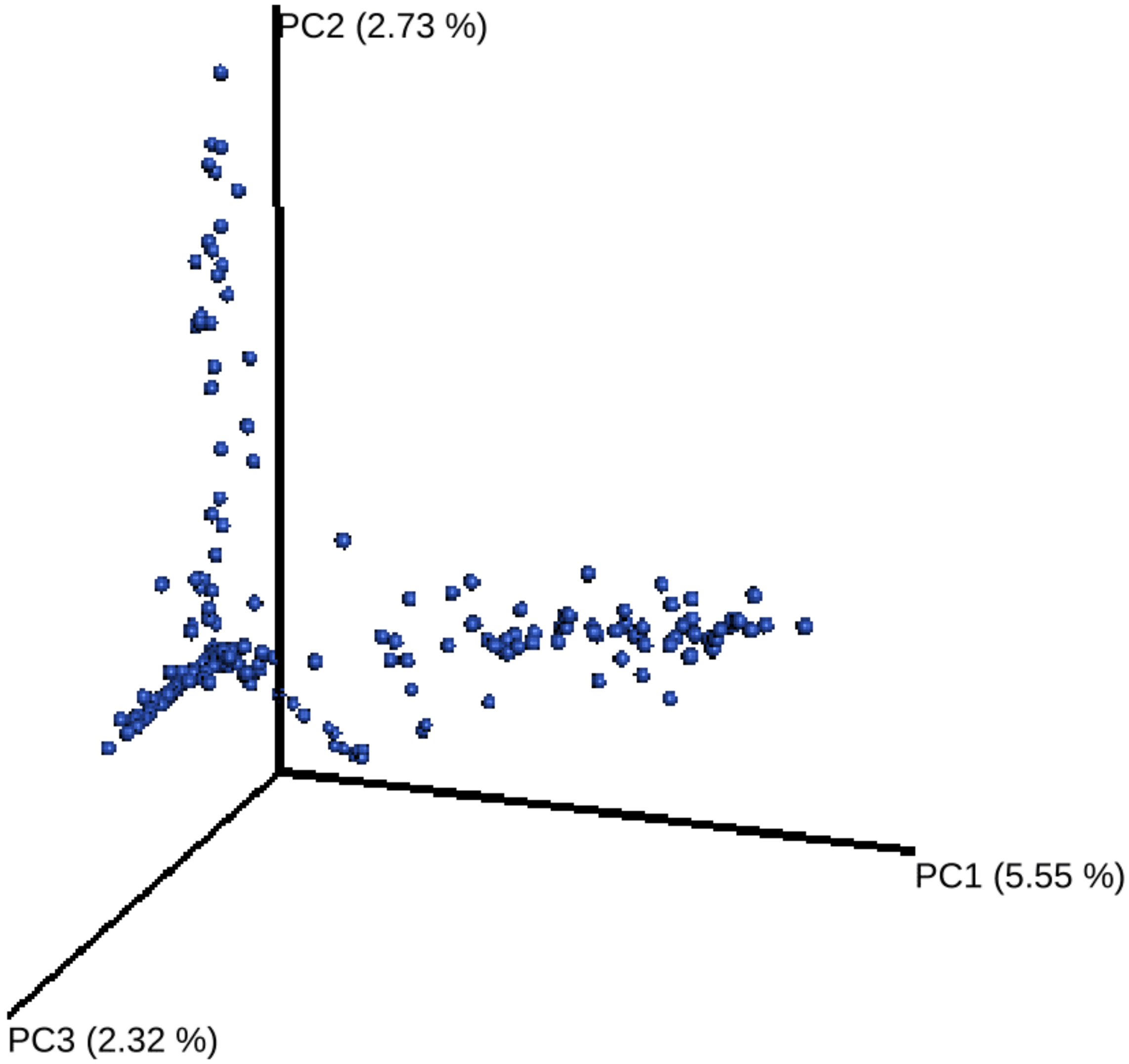


Figure 14: 3D Principal Component Plot for study CANDIDIASIS. Blue data points represent all samples collected for study CANDIDIASIS plotted against the first three principal components. Each principal component explains the highest amount of variance between sample bacterial composition. Figure 14 illustrates relatively low variance percentages and a lack of sample clustering.

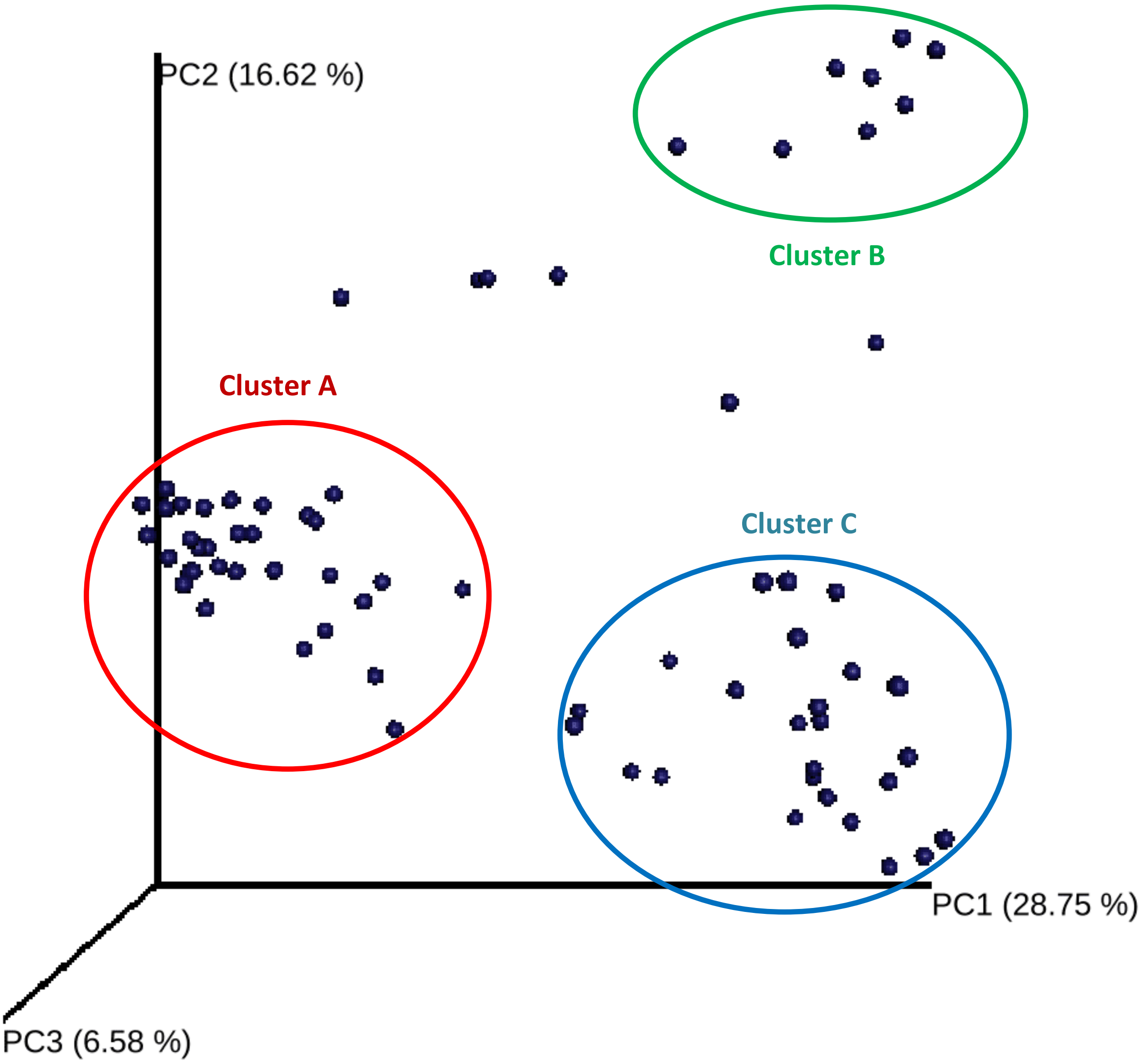


Figure 15: Study BV 3D Principal Component Graph. The principal component plot consists of three axis representing the first three principal components, explaining the most bacterial composition variance within samples. Data points of patients from the study cluster according to composition similarities with three distinct clusters forming which are highlighted by the red, green and blue circles

The PCA based on study's HSV2 data is illustrated in Figure 16. Once again the three axes represent the first three principal components along the percentages responsible for the sample variance. Study HSV2 exemplifies a relatively small study with only 51 samples (whereas an average of 143 samples is present in most selected studies). Thus due to lack of greater sampling depth PCA reveals a single dominant cluster. The detectable cluster is most likely driven by the 62% abundant *Lactobacilli* reported in Table 2. As seen previously on Figure 14 (PCA for study CANDIDIASIS), the PC percentages for study HSV2 are relatively low, representing only 19% of the total sample abundance variation of the study. Unlike study CANDIDIASIS, the de-multiplexing of study HSV2 was achieved through the second and final pipeline designed for this project. Consequently, both HSV2 and CANDIDIASIS studies consist of "loosely" structured microbiomes with atypical compositions causing multidimensionality to their PCA matrix. Thus each individual principal component will explain a small percentage of the total sample abundance variance of a complete study. The lack of further sample clustering in study CANDIDIASIS further challenges the drawing of conclusions on microbiome structure and organism interactions.

Finally, study SV displays very dissimilar PCA patterns, to all previously discussed studies (Figure 17). Figure 17 typically encloses the first three principal components, including the percentages covering the amount of sample abundance variation within the study. However, study SV is the only study which displays significant variance in the third principal component (PC3 = 14.9%), thus all PC axes have corresponding sample data points. PC1 as expected, represents the majority of the variance with 27%, whereas PC2 and PC3 have comparably high values with 17% and 15% correspondingly. Study SV displays one of the highest amounts of PCA coverage with about 60% sample bacteria abundance variance explained by the first three principal components. Interestingly unlike HIV, CANDIDIASIS, BV, HSV2 studies, more than four clusters are distinguishable from the PCA plot, even though the overall bacterial structure presented on Table 2 remains equivalent to the other studies. Cluster A demonstrates the most populated cluster, thus is likely to be determined by *Lactobacilli* consisting 55% of the total study. Once more, although the 3D PCA plot suggests multiple interesting patterns, there is no conclusive evidence of consistent bacterial structures or microbiome organism interactions. Seminal samples have contributed to increased clustering, due to their high bacterial diversity. Numerous speculations on microbiome structure providing stability to a microbiome have been brought to the surface. Further investigation on the causation of sample clustering was essential, therefore, the subsequent step focused on hierarchical clustering. Hierarchical clustering along with heatmap charts would allow easy visualisation of the composition similarities that drive the generation of the groupings/ clusters between samples of a study, thus suggesting association between organisms as well as microbiome structure.

Additionally, a python script was programmed to carry out PCA analysis to focus on different aspects of Principal Component analysis to ensure no additional patterns could be traced. In more detail python programming allowed plotting of specific principal components combinations, giving a more powerful analytical tool (Appendix 20). Results proved parallel to QIIME's plots thus were excluded from the analysis.

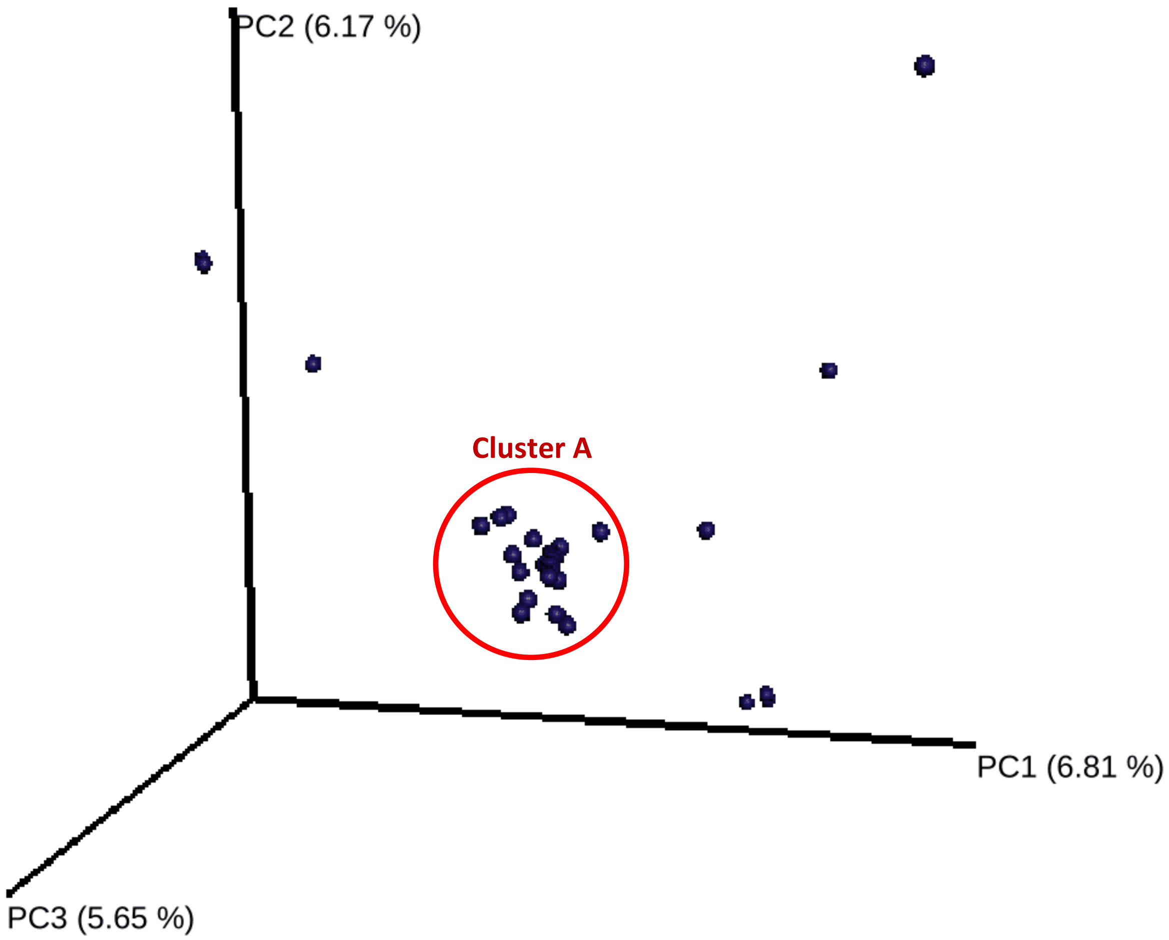


Figure 16: Principal Component 3D plot for study HSV2. Principal Component analysis plot illustrates low percentages of explained variance in all three principal component axis as well as a single sample cluster (cluster A) indicating sample composition similarities between patients.

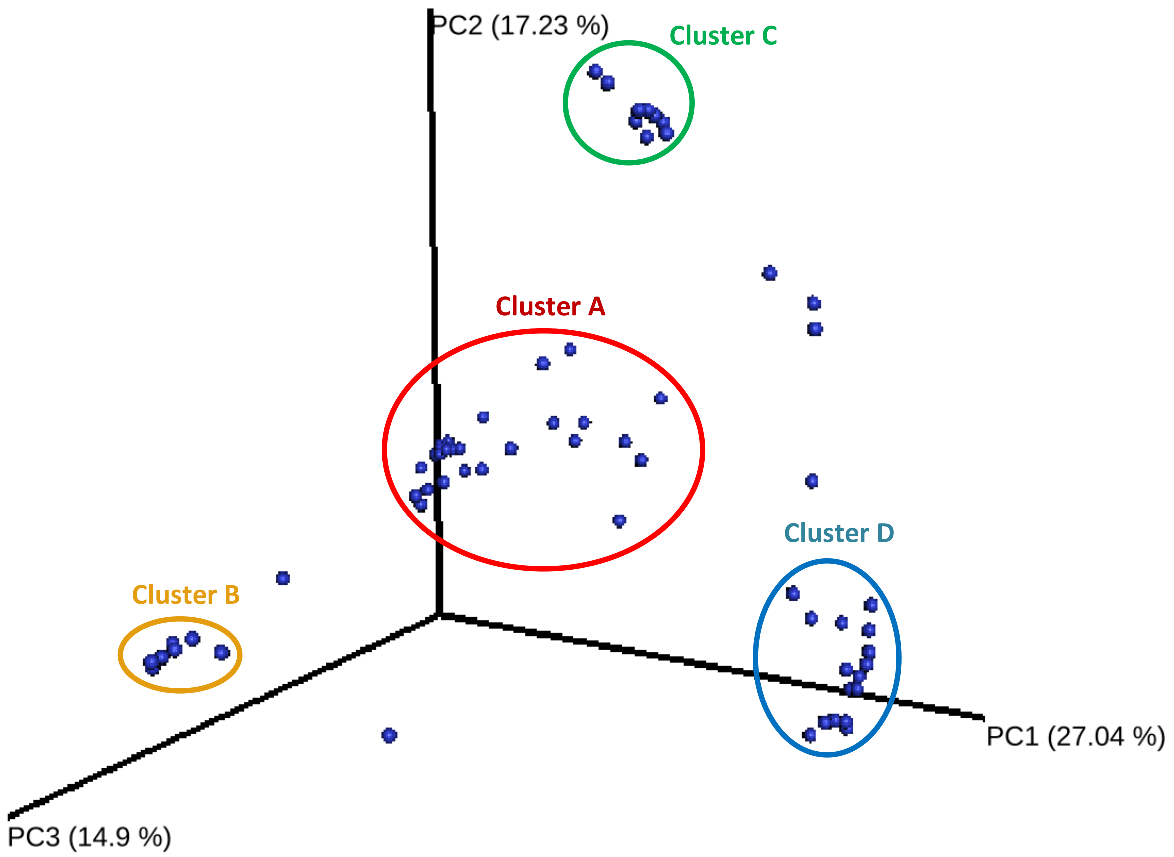


Figure 17: 3D Principal component graph for study SV. Sample data plotted in a 3D graph against the first three principal components represents the highest amounts of sample bacterial composition. The analysis suggests four sample clusters. Each distinct cluster demonstrates shared bacterial composition similarities, which are illustrated by the red, orange, green and blue circles; with Cluster A being the most prevalent.

4.2 Hierarchical clustering displayed in the form of heatmaps

The python script discussed in chapter 2.5.1, was designed to carry out hierarchical clustering along with heatmap analysis on HIV, CANDIDIASIS, BV, HSV2 and SV studies. Heatmaps allow clear visualisation of clustering, as well as fast assessment of potential bacterial interactions both at intrapersonal and at collective study level. Unlike previous methods, heatmaps evaluate similarity patterns between samples' bacterial composition, determine organism interactions based on relative abundance, display the most abundant bacteria within each sample and in a complete study with no ambiguity, as well as providing a visual representation of microbial community structure.

The first stages of the script generated dendrograms for both sample and bacteria data for each study. Hierarchical clustering dendrograms represent similarities between sample or bacteria values in the form of clusters, as well as illustrating the order of clustering. These dendrograms have a similar structure to phylogenetic trees, but with no phylogeny assessment. Dendrograms' leaves clustered in close proximity illustrate high levels of compositional or abundance similarity, with Branch length exhibiting the degree of dissimilarity and order of clustering. Both x and y axes of the heatmap display dendrograms respectively for sample and bacteria values. The following stages of the python program, carried out abundance data clustering, thus ensuring global scaling of the data matrix. The relative abundance data matrix was normalised and logged for the purpose of the heatmap. This ensured easy visualisation of the complete data matrix, as well as of the suggestive patterns generated from analysis. Global scaling assisted with clear, non-arbitrary visualisation and quantification of sample specific bacterium abundance composition, thus gaining insight on interpersonal sample composition by detecting the most abundant taxa within each sample. The relative abundance matrix was clustered and scaled from a Euclidian distance similarity based algorithm, converting "similarity" to a quantifiable variable. In conclusion, the script generated heatmap charts for all selected studies (Figure 18- 22) with sample and bacteria dendrograms on both axes and a similarity relative abundance data matrix. The rows of each heatmap display the taxonomically assigned 16S rRNA sequences (recorded at genus level), whereas the columns represent the samples from which the sequences originate. A colour bar quantified the measures of relative abundance data, within a given study, while displayed in a similarity distance matrix.

Figure 18, illustrates the heatmap generated for study HIV. The top x-axis displays the dendrogram created for the sample data, with the bottom x-axis listing the corresponding sample IDs. The sample dendrogram expresses clustering between samples depending on microbiome composition. On the other hand, the left y- axis along the heatmap illustrate the bacteria dendrogram reporting clustering between all bacteria present within study HIV. Bacterial clustering is based upon bacterial abundance within samples with no correlation to phylogenetic assessment. The genus level taxonomic identities corresponding to the y-axis bacteria dendrogram are listed on the right side of the heatmap. The relative abundance data matrix is scaled and displayed as a colour range from white to blue, with the darkest blue illustrating the highest relative abundance values. The relative abundance data were normalised and logged to ensure accuracy and assist visualisation of interaction patterns.

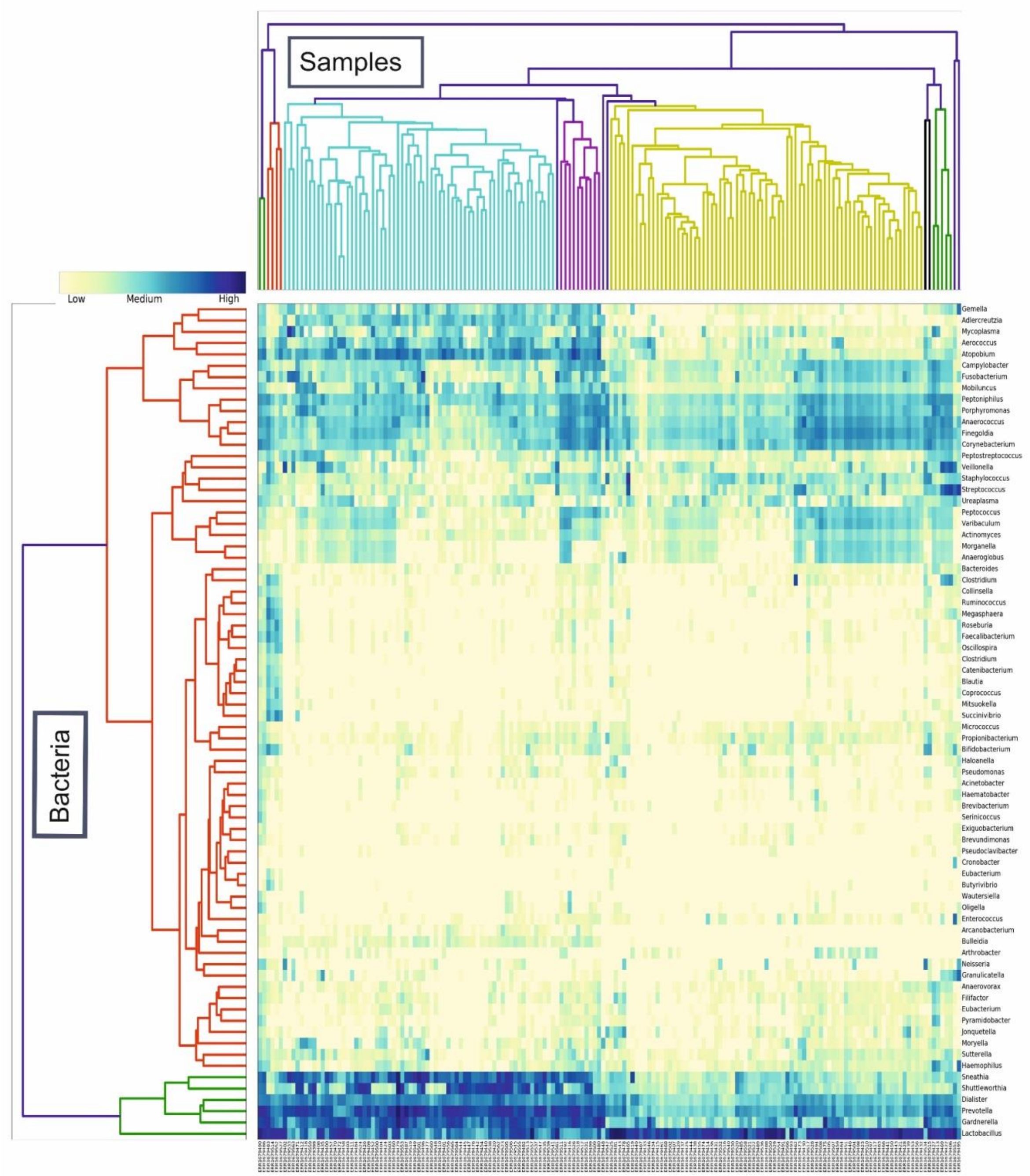


Figure 18: Heatmap representing vaginal microbiome profiles of study HIV. The heatmap includes hierarchical clustering for sample and bacterial data on the top x-axis and the left y-axis respectively. The bottom x-axis as well as the right y-axis contain the corresponding sample or bacterial IDs for each leaf of the dendrograms. The data matrix illustrated in the heatmap presents the logged abundances of the bacteria taxa found in the study. The values are displayed as colour scales with dark blue representing the highest values and white representing zero abundance. Hierarchical clustering suggests three major sample clusters as well as three bacterial clusters dependant on microbiome composition.

Observing study's HIV sample dendrogram, five clusters are visible, two of which comprise most samples. On the other hand, bacteria dendrograms present numerous smaller clusters consisted of multiple sub-clusters, thus difficult to distinguish exact clustering. The relative abundance data matrix reveals the most abundant taxa, as well as bacteria interactions driven by abundance similarities. Bacteria with consistently high abundances shared in multiple samples cluster closely (Figure 18), representing potential organism interactions. This suggests stability in the microbiome structure. Interestingly, the samples dendrogram on the x-axis, partially matches the clustering patterns observed in the PCA plot (Figure 13). The PCA displayed three sample clusters based on bacteria composition, with cluster A dominantly represented by *Lactobacilli*. Equally the two major sample clusters visualised on the x-axis of the heatmap are driven by the relative abundance values of *Lactobacilli*. Though the sample dendrogram displays five clusters, two clusters consist of few samples; unlike the three highly populated clusters observed in the PCA plot (Figure 13). Therefore, both PCA and dendrogram sample clustering present equivalent patterns, ensuring accuracy of the results. Focusing on bacteria patterns, a number of interactions appear to be linked with *Lactobacillus* abundancy. All sample clusters experience a significant increase of *Gardnerella, Prevotella, Dialister, Shuttleworthia, Sneathia* abundance, when associated with a decrease in *Lactobacilli* (Figure 18). Thus, interactions between these organisms can be speculated. Interestingly, all closely clustered organisms are listed in the top six most abundant bacteria in Table 2. Additionally, multiple clustering patterns are visible in a number of sample clusters between the first 11 listed bacteria (Gemella-Corynebacterium) (Figure 18). In conclusion, study HIV represents a number of clear sample clusters signifying similarities in microbiome composition, whereas bacterial clustering is influenced by multiple parameters thus resulting in less structured clusters, yet an apparent link between *Lactobacilli* abundance and bacteria clustering is present. However, all major clustering patterns observed are only informally suggestive of microbial interactions, thus further testing with correlation models was followed.

Figure 19 displays the heatmap from study CANDIDIASIS, illustrating diversity patterns as well as relationships between microbiome composition. Like in study HIV the heatmap consists of x-axis displaying samples dendrogram along the sample IDs, whereas the y-axis demonstrates bacterial dendrogram along the corresponding taxonomic IDs. Each row represents a specific taxon identified from the 16S rRNA sequences, whereas each column corresponds to a specific sample of the study. Supporting the clustering observed in the PCA analysis (Figure 14), the sampling dendrogram does not display clear clustering. Instead multiple small clusters with various sub-clusters are visualised. Therefore, this confirms the high β-diversity of study CANDIDIASIS, following previous claims based on the taxonomic bar charts (Figures 4 and 9). Although the study displays high bacteria richness (shown by bacteria dendrograms) the low relative abundances cause decreased total variation. *Lactobacilli* dominate the majority of the samples in study CANDIDIASIS, coinciding with the 80% relative abundance data presented on Table 2. The bacteria dendrograms display two major clusters with *Dialister, Prevotella, Atopobium and Gardnerella* appear strongly linked. Interestingly, *Dialister*, *Prevotella* and *Gardnerella* are three bacteria which appeared correlated in study HIV. However, unlike in study HIV where the link was driven by the presence or absence of *Lactobacilli*, study CANDIDIASIS does not display such link. Although the three organisms appear interconnected irrespectively of the *Lactobacilli* dominance, their relative abundances are equivalently low. While clustering is suggestive of strong correlations, the links need to be validated through correlation statistical models.

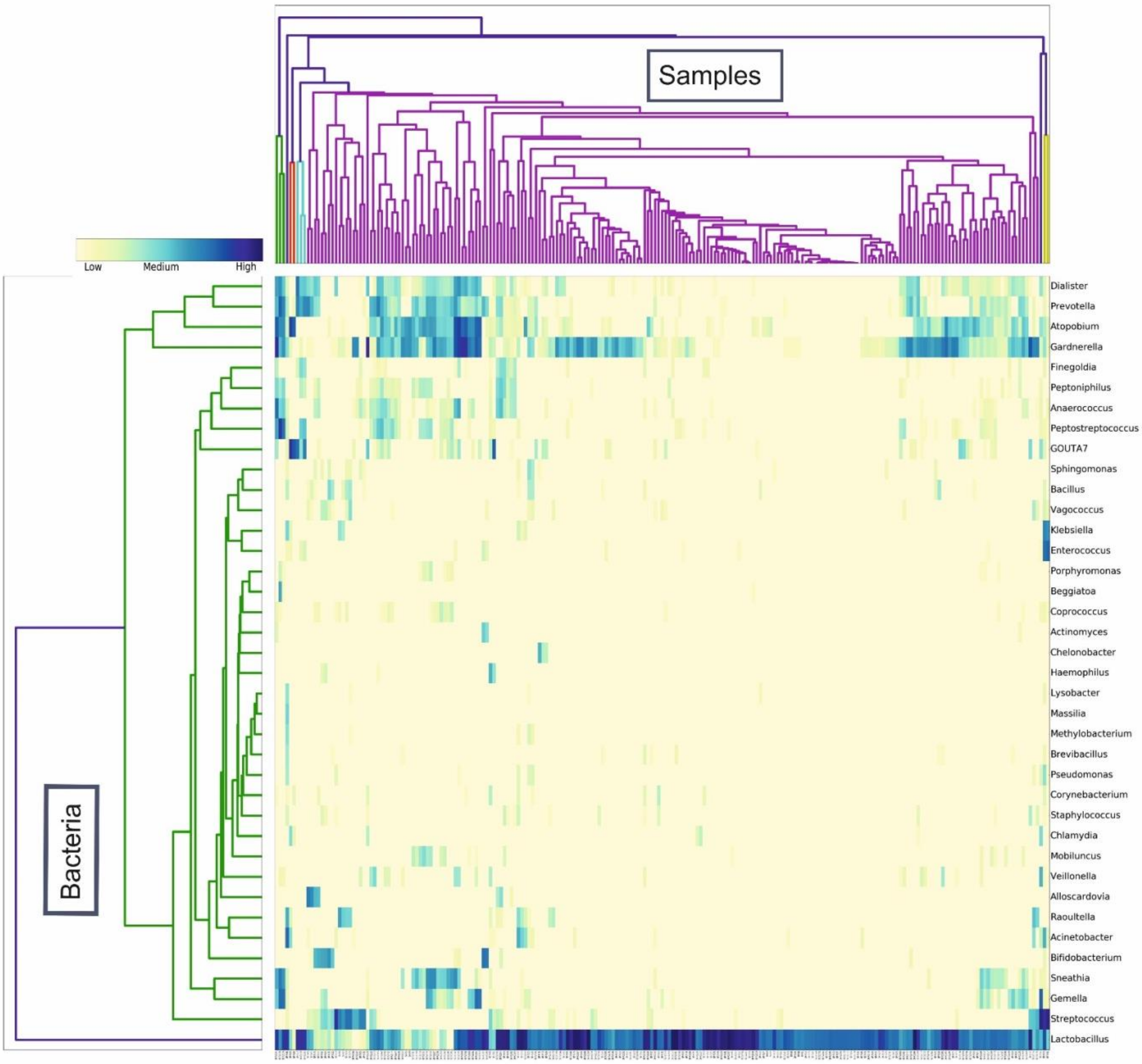


Figure 19: Heatmap of vaginal microbiomes present in study CANDIDIASIS. Hierarchical clustering for sample and bacterial data are displayed on the top x-axis and the left y-axis respectively. The sample and bacteria IDs corresponding to the dendrograms are listed at the bottom x-axis and right y-axis respectively. The data matrix representing the logged abundances of the bacteria taxa identified in study CANDIDIASIS are displayed as colour scales. Dark blue represents high abundancy whereas white represents zero abundance. Hierarchical clustering suggests various small sample clusters as well as two major bacterial clusters dependant on microbiome bacterial composition.

Respectively, Figure 20 illustrates the heatmap generated for study BV. Identical to both previously discussed heatmaps, Figure 20 assists visualisation of diversity and clustering patterns though the presentation of sample and bacteria dendrograms along a scaled and logged data matrix of relative abundance. Each dendrogram was created through hierarchical clustering, permitting visualisation of similarity patterns through clustering. The x-axis sample dendrogram illustrates one dominant cluster composed by the majority of the study's samples, followed by numerous smaller clusters. Clustering patterns are partially supportive of the clustering observed on the PCA plot in Figure 15. PCA as previously discussed, illustrated three well-defined clusters with cluster A consisting of most samples, thus possibly driven by *Shuttleworthia* - the most dominant bacteria present in study BV (30% total abundancy as reported in Table 2). Although one dominant cluster is visible on the sample dendrogram, clusters B and C from Figure 15 were probably representing a combination of the smaller clustering groups observed in Figure's 20 sample dendrogram. In contrast, bacteria dendrograms displayed on the y-axis, represent three clusters driven by each taxon's relative abundance. The red cluster lists the eight most abundant bacteria (*Shuttleworthia*, *Lactobacillus, Prevotella*, *Megasphaera, Sneathia, Parvimonas,* Atopobium and *Dialister*) identifying various microbiome community structures. Additionally, bacteria *Peptoniphilus, Gemella, Mycoplasma, Aerococcus and Clostridium* appear linked even though they present relatively low abundance values. Study BV is the third study supporting a community link between *Dialister* and *Prevotella* even in the absence of *Lactobacillus* dominance. Correlation analysis is essential to test the validity of a possible link between *Dialister* and *Prevotella*.

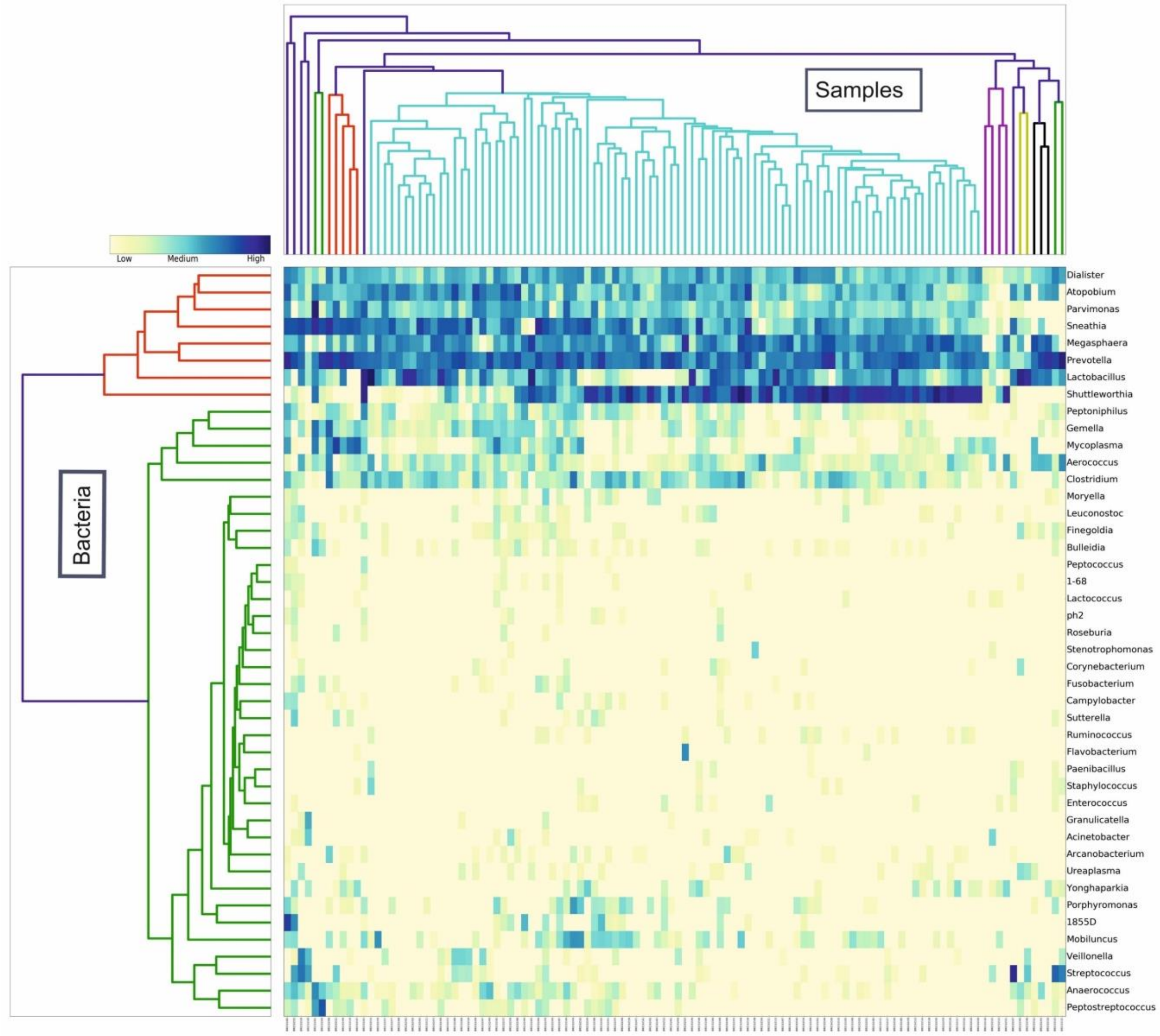
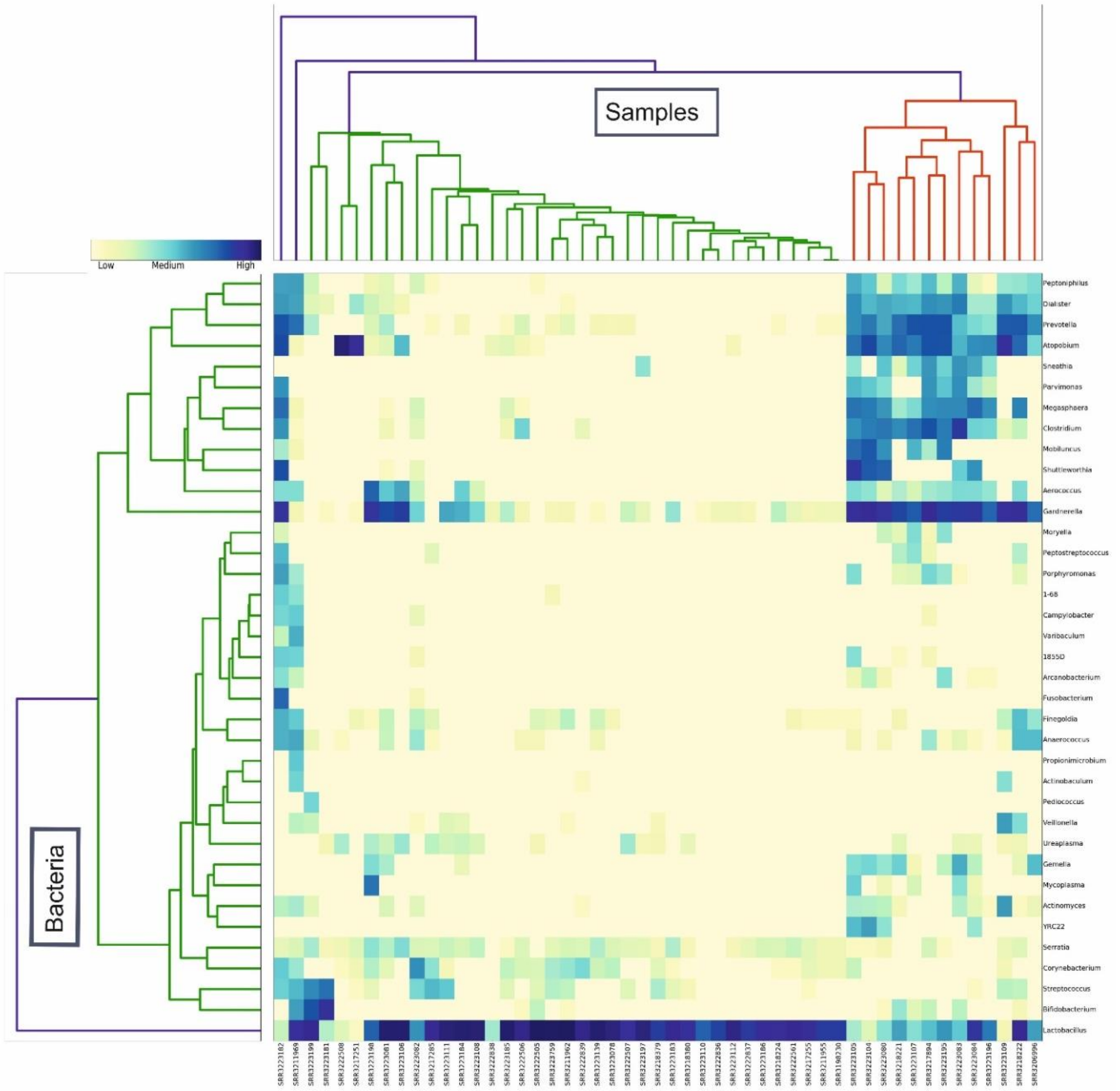


Figure 20: Heatmap of patients' vaginal microbiomes present in study BV. Hierarchical clustering for sample data is illustrated as a dendrogram on the top x-axis. Hierarchical dendrogram on the left y-axis demonstrates clustering for the bacterial data. The corresponding sample and bacteria IDs of the dendrograms are listed at the bottom x-axis and the right y-axis respectively. The logged bacterial abundances data matrix is displayed as a colour scale. Dark blue represents high values whereas white represents zero abundance values. The dendrograms illustrate one dominant sample cluster as well as three major bacterial clusters dependant on microbiome bacterial composition.

Study's HSV2 data generated hierarchical clustering along a heatmap presented on Figure 21 allowing visualisation of clustering patterns and thus portray potential microbiome community structures. Alike studies HIV, CANDIDIASIS, BV hierarchical clustering is displayed in the form of dendrograms for both sample and bacteria values. Due to the multidimensionality of the data arrays with various representative conditions (microbiome type, sample, abundance, bacteria) a heatmap permits straightforward visualisation of the data matrix and thus diversity and clustering patterns. The x-axis of the heatmap display the dendrogram generated for sample clustering, along the sample IDs at the bottom of the heatmap. Sample dendrogram illustrates two visible clusters, however samples SRR3223167 and SRR3211969 do not fall into any clustering due to their distinctive composition. Shannon B. 2017 et al. does not list metadata descriptions in the NCBI database, however patients were only sampled once throughout the course of their study with no follow up visits [84]. Therefore, outlier samples SRR3223167 and SRR3211969 represent two distinct patients. The green cluster presented is significantly populated, however the samples composing it lack β-diversity due to the low average abundances. On the other hand, the red cluster consists of fewer samples with higher abundances and more complex microbiome communities. Bacteria dendrograms display two clusters with *Lactobacillus* as an outlier, signifying the excessive level of dominance (62% of total study abundance). The second bacteria cluster does not represent bacterial interactions due to low bacteria relative abundance. However, the first bacterial cluster represents strong clustering between 13 diverse samples with high relative abundances of *Peptoniphilus, Dialister, Prevotella, Atopobium, Sneathia, Parvimonas, Megasphaera, Clostridium, Mobiluncus, Shuttleworthia, Aerococcus and Gardanella*. Interestingly not all bacteria coincide with the most highly abundant organisms listed on Table 2. Therefore, verifying that microbiome structure is not exclusively driven by abundancy. Multiple organisms reported in previous studies similarly appear related in study HSV2, such as *Dialister*, *Prevotella, Atopobium, Gardanella* and others, thus confirming the need for further investigation.



<u>Figure 21:</u> <u>Heatmap representing patients' vaginal microbiome profiles from study HSV2</u>. The heatmap exhibits hierarchical clustering in the form of dendrograms for sample and bacterial data on the top x-axis and the left y-axis respectively. The bottom x-axis and the right y-axis list the corresponding sample or bacterial IDs for each leaf of the dendrograms. The logged abundance data matrix illustrated is displayed as a colour scale. Dark blue represents the highest abundance whereas white represents zero abundance. Hierarchical clustering suggests two major sample clusters as well as two distinct bacterial clusters dependant on microbiome composition.

Finally, study SV equally displays the same analysis as discussed for all other heatmaps (Figure 22). The x-axis sample dendrogram illustrates two very distinct clusters (red and green) with two sample outliers each. The green cluster consists of the majority of samples and though taxon richness is high, unique taxonomies are present in very low abundances. By examining the taxonomies presented on the right it is evident that the x-axis green cluster illustrates female vaginal samples. Female vaginal samples are evidently less diverse, in comparison to the adjacent red cluster depicting male samples. Male samples are characterised by atypically low *Lactobacilli* and demonstrate elevated bacterial diversity. For the purpose of this analysis, study SV will be perceived here as a low diversity study, as most of the β – diversity visualised on the heatmap is driven by high levels of male sample diversity.

Bacteria dendrograms presented on the left y-axis, display an extraordinary number of taxonomies. Once again a result of seminal samples. Bacteria clustering appears puzzling with few defined clusters, thus proving difficult to distinguish between female and male clusters. The two sexes share organisms but not community structures with divergence in abundance values of certain taxa. Focusing on vaginal samples, a small bacterial cluster of 17 organisms (red cluster-bacterial clustering y-axis) illustrates association between some of the most abundant organisms including *Lactobacillus, Dialister, Prevotella, Gardanella, Streptococcus* and others. Interestingly, all female clustered bacteria are dominantly present in male samples. Study SV examined the effect of sexual intercourse on vaginal microbiomes. Mandar et al. 2015 report microbiome changes at intrapersonal level, post sexual intercourse, which defends the presence of shared bacteria between female and male samples. As mentioned before, the relationships reported for the shared organisms do not appear dependant on *Lactobacillus* abundance.

In conclusion, although study's SV heatmap appears complex, most taxonomic variation is driven by seminal samples, though more female samples were collected. Additionally, the heatmap displays strong links between all commonly present organisms discussed in the previous studies (e.g. *Dialister, Prevotella*), which are clustered in close proximity within the red y-axis cluster on the bacterial dendrogram. Strongly associated bacteria from vaginal samples (illustrated through bacterial dendrogram clustering) are not necessarily listed in the ten most abundant taxa (Table 2). This is due to male seminal organisms' impact on the total percentage abundance, due to their increased bacterial diversity. Sample dendrogram clustering did not offer further insight to PCA clustering patterns (Figure 17) as multiple factors affecting the basis of the clustering were present for this study (sex, sample type, microbiome, abundances). Finally, the heatmaps designed through the bespoke python programming allowed illustration of both bacteria abundance and bacteria composition thus proposing bacterial relationships for the various microbiomes within each study. Though presenting multiple patterns and potential community links, establishment of microbiome interactions is not possible without correlation analysis, needed to quantify and ensure confidence in the relationships.

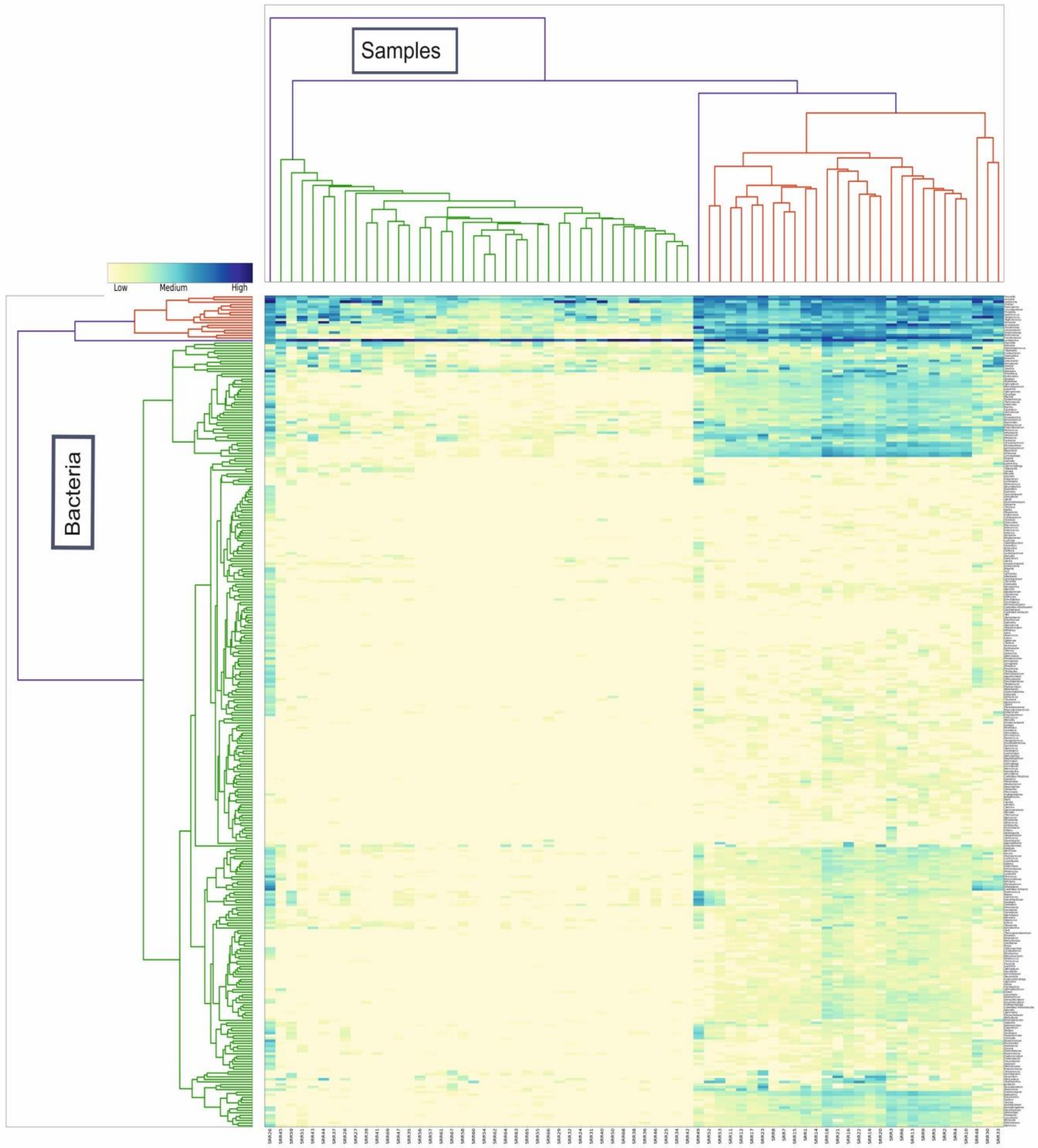


Figure 22: Heatmap of vaginal microbiomes present in study SV. Hierarchical clustering for sample and bacterial data are displayed as dendrograms on the top x-axis and the left y-axis respectively. The corresponding sample and bacteria IDs are listed at the bottom x-axis and the right y-axis respectively. The data matrix of the logged bacteria abundances are illustrated as colour scales. Dark blue represents the highest values, whereas white represents zero abundance. Hierarchical clustering suggests various small bacterial clusters; with one visibly distinct cluster (red cluster on left y-axis), as well as two major sample clusters dependant on microbiome bacterial composition.

5. CORRELATION STATISTICS ANALYSIS

The previous chapter covered the results of clustering analyses which provided insight to potential microbiome interactions within the vaginal microbiome. Interestingly, multiple patterns were shared between the selected studies. Although the heatmaps suggested associations between vaginal bacteria indicating potential microbiome structures, the links have to be further tested through correlation statistics analysis. To quantify any correlations between pairs of bacterial genera in the vaginal microbiome Spearman Rank Correlation Coefficient (ρ) model was utilised.

Spearman's rank correlation coefficient permits correlation analysis on non-normally distributed data by ranking the abundance of bacteria within a data matrix. The relative abundance data from all selected studies were not normally distributed, an effect that was due to the large number of zero values (or values close to zero) for many species in many samples, thus skewing the influence of the abundances towards the most abundant bacteria. For the purpose of this study correlation analysis assesses correlation between two data variables; in this case bacteria. Spearman's correlation generates asymmetrical tables (Figure 23) with values ranging from -1 to 1, with 1 representing a perfect positive correlation, -1 a negative correlation and finally 0 the presence of no correlation between two bacteria. Correlation models require p-values certifying statistical significance of the correlation links. However due to ample bacteria taxon richness within each study, Bonferroni correction on p-values and a Spearman correlation cut off value were estimated to avoid ambiguity in the significance of the correlations. The Spearman correlation cut off value delivered 95% confidence that a correlation is not a result of random associations driven by high abundance, but rather a true correlation.

Heatmaps identified various associations between bacterial community members depending on abundance composition. Spearman's correlation characterises the nature of the observed associations, by investigating for linear correlations. The first captivating association was drawn between *Dialister* and *Prevotella*, two bacteria dominantly present in all five selected studies (Table 3). As discussed previously, *Dialister* and *Prevotella* were consistently presented within the top ten most abundant organisms of the microbiomes (Table 2). Interestingly within all studies, *Prevotella* dominates *Dialister* in abundance, irrespective of the microbiome's conditional state. The heatmap from study HIV (Figure 18) reveals that both *Prevotella* and *Dialister* experience a visible increase within the samples consisting of lower *Lactobacilli*. Samples consisting of high *Lactobacilli* abundances present an abundance decrease in *Dialister* and *Prevotella*, however the two remain correlated. Therefore, again confirming that an association between *Dialister* and *Prevotella* is not dependant on *Lactobacilli* abundance. Finally, all studies (HIV, CANDIDIASIS, BV, HSV2, SV) consistently demonstrate strong clustering (Figures 18-22) between the two organisms even when relative abundances are not proportional. Therefore, all findings suggest strong association between *Prevotella* and *Dialister* as well as indicate strong influence on community structure and stability.

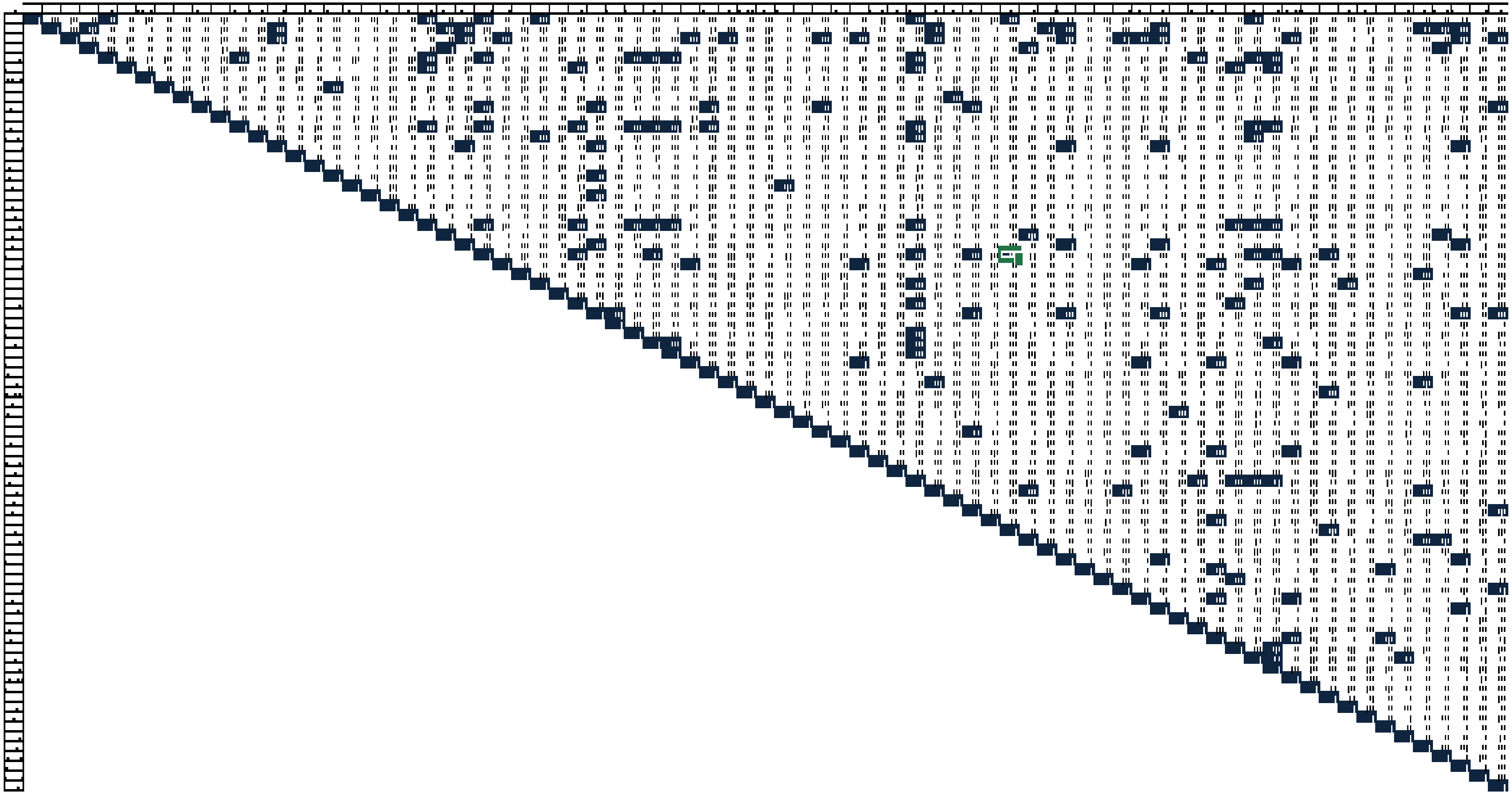


Figure 23: Spearman Rank Correlation Coefficient assymentric table for *study HSV2.* Figure 23 displays the assymentrical table enclosing Spearman correlation values generated for study HSV2. The highlighted blue cells illustrate correlation values over 0.6, signifying the presentage of overall positive correlations within a study.

	Dialister – Prevotella	95% Spearman correlation confidence threshold
HIV	0.81	0.5
BV	0.44	0.41
CANDIDIASIS	0.68	0.27
SV	0.91	0.55
HSV2	0.75	0.57

Table 3: Spearman Ranked Correlation Coefficient analysis between *Dialister* and *Prevotella* for all selected studies. The values generated through a python script running Spearman's correlation are listed on Table3. The significance threshold is additionally displayed for all studies, representing 95% certainty of a correction being true and not a result of random association by chance. Thus validating significance of the correlations.

Spearman's rank correlation coefficient was carried out for all possible bacteria-bacteria associations within each study, however only key relationships will be discussed. (For full Spearman correlation tables of each study refer to Appendix electronic files). Table 3 lists all Spearman correlation values generated between *Dialister* and *Prevotella* for all five selected studies. To validate correlation results, the 95% spearman correlation coefficient cut off values were included in Table 3. As expected, correlation values were consistently and significantly high, with the exception of study BV which illustrated moderately positive, yet highly probable correlation between the two organisms. Study BV focused on identifying microbiome composition differences between female homosexual couples, female heterosexual couples and finally heterosexual BV female carriers couples [83]. Due to the presence of 44 dysbiotic vaginal microbiomes infected with BV, the overall study composition is shifted, thus altering *Dialister* and *Prevotella* abundances. However, the 95% certainty threshold for study BV was 0.41 thus still representing a very favourable correlation. Lower positive correlations between *Dialister* and *Prevotella* could be a result of more unstable microbiome communities, due to high levels of bacteria diversity driven by BV. On the other hand, studies HIV, CANDIDIASIS, HSV2, SV demonstrate substantially high positive correlations with studies HIV and SV almost representing a perfect correlation. All Spearman correlation values exceed the 95% significance threshold thus supporting that these correlations are genuine. In other words, it is extremely probable that *Dialister* and *Prevotella* are correlated in vaginal microbiomes irrespectively on microbiome state or microbiome composition.

Graphical representation of the correlations would allow straight-forward visualisation of the type of correlation and strength of the correlations. Scatter plots were generated for all studies sporting a *Dialister* and *Prevotella* link (Figure 24). The green best fit line represents the correlation link between the two variables (*Dialister* and *Prevotella*) and illustrates the strength (angle of the line) and type (positive, negative or no correlation) of correlation. Each subplot within Figure 24 represents a single study. Coinciding with the previous results, all studies represent strong positive correlations, with the exception of study BV, which illustrates less sharp angles of the linear positive correlation between the two bacteria. The type of the correlation (i.e. positive or negative) is described through the angle of the best fitted line, whilst the strength of the correlation is dependent on the data proximity to the best fit line.

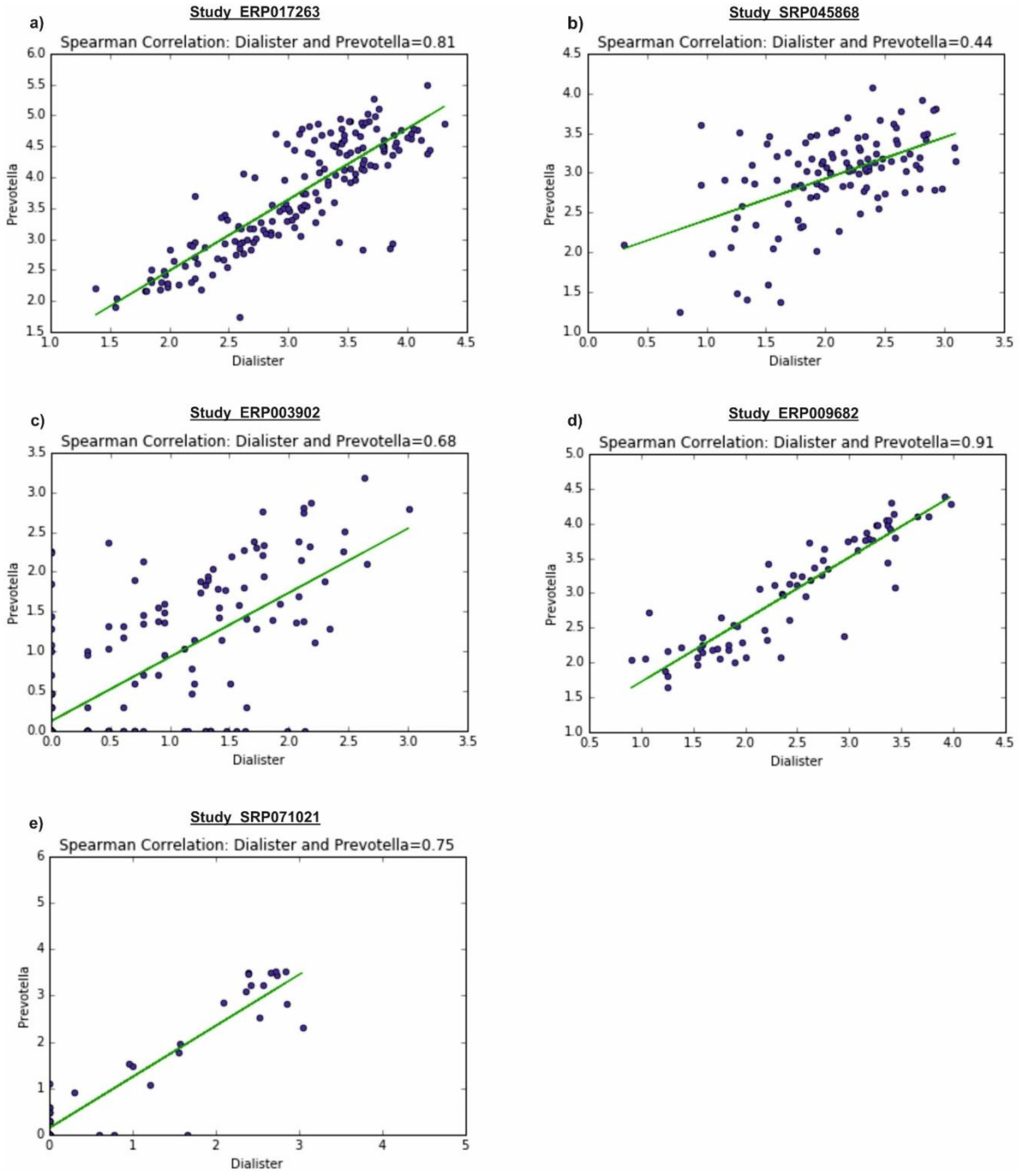


Figure 24: Scatter plots of *Dialister* and *Prevotella* correlations for all five selected studies. Figure 24 illustrates scatter plots for all five studies (HIV, BV, CANDIDIASIS, SV, HSV2). The blue data points represent the relative logged abundance data for *Dialister* (x-axis) against the relative logged abundances of *Prevotella* (y-axis). The green best fit line illustrates the correlation link between the two variables (*Dialister* and *Prevotella*). Each subplot lists the Spearman Rank Correlation Coefficient value for each study. All subplots suggest strong positive correlations between *Dialister* and *Prevotella* with high correlation values and steep best fit green lines.

Figures 24a,c,d,e illustrate strong, positive correlations between *Dialister* and *Prevotella* for studies HIV, CANDIDIASIS, HSV2, SV. Study BV displays a weaker positive correlation demonstrated by the lower best fit angle and the ρ value of 0.44 (spearman correlation 95% certainty threshold 0.41). All correlations significantly exceeded the 95% Spearman correlation cut off value, thus ensuring confidence in the nature of the correlations and signifying the low probabilities of the correlation being a result of random association due to high abundances. In conclusion, all results strongly support a relationship between *Dialister* and *Prevotella* bacteria and therefore suggesting a key role in microbiome structure stability. To this point, the correlation detected between *Dialister* and *Prevotella* is a novel correlation which has yet to be investigated. Although the true nature of the correlation between *Dialister* and *Prevotella* is not known, speculations on its influence on microbiome community structure can be made. Perhaps the degree of microbiome susceptibility to disease or infection (observed in BV and HIV infected microbiomes), is influenced by the association between *Dialister* and *Prevotella* [88], [99], [100]. However, the true character of the association could be investigated through metabolic patterns and interactions provided in KEGG followed by culture experiments.

Dialister and *Prevotella* were not the only dominant associations observed in the heatmap analyses. *Gardnerella and Atopobium*, were two additional bacteria clustered in close proximity with *Dialister* and *Prevotella*. As previously mentioned, *Gardnerella and Atopobium*, *Dialister* and *Prevotella* are shared bacteria within most of the five studies and are listed within the top ten most abundant organisms of the microbiomes within each study. Table 4 contains the Spearman correlation values for all possible associations between *Gardnerella, Atopobium*, *Dialister* and *Prevotella* tested within each study. The new 95% significance threshold implemented through the Bonferroni correction is included on Table 3.

	HIV	BV	CANDIDIASIS	SV	HSV2
Atopobium - Gardnerella	0.79	0.20	0.47	0.22	0.59
Dialister - Gardnerella	0.68	-0.04	0.50	0.37	0.66
Prevotella - Gardnerella	0.68	-0.09	0.35	0.38	0.60
Atopobium - Prevotella	0.78	0.03	0.53	0.50	0.58
Atopobium - Dialister	0.72	0.02	0.58	0.39	0.71
Streptococcus - Gardnerella	0.04	0.16	-0.07	0.33	-0.14
Streptococcus - Atopobium	-0.10	0.18	-0.11	0.21	-0.11
Streptococcus - Prevotella	-0.02	0.08	0.05	0.53	-0.01
Streptococcus - Dialister	0.02	0.03	0.09	0.50	0.06
Sneathia - Gardnerella	0.59	0.04	0.32	0.36	0.51
Sneathia - Atopobium	0.75	0.01	0.47	0.46	0.42
Sneathia - Prevotella	0.81	0.12	0.46	0.50	0.36
Sneathia - Dialister	0.66	0.22	0.40	0.53	0.48

Table 4: Spearman Rank Correlation Coefficient analysis between key bacteria of the vaginal microbiome. Table 4 displays the results of Spearman's Correlation analysis between *Dialister*, *Prevotella, Gardnerella*, *Atopobium, Streptococcus* and *Sneathia* within all five studies. Table 4 lists the correlation values with the green values meeting the 95% spearman correlation cut off value for each study as illustrated in Table 3.

Unlike the consistently high correlation values observed between *Dialister* and *Prevotella*, the values presented on Table 4 for correlations between bacteria *Gardnerella, Atopobium*, *Dialister* and *Prevotella*, appear study dependent. Study HIV and HSV2 are the only two studies with consistently high Spearman correlation values for all bacteria pairs. The 95% Spearman correlation significance threshold for study HIV is 0.5, therefore all correlations are significantly likely to be true correlations. Equally study HSV2 exhibits a 0.57 95% cut off value, which is significantly lower than all correlations between bacterial pairs, thus confirming high likelihood of the correlations. On the other hand, study BV exhibits exceptionally low correlation values representing almost no correlation between the bacteria, and a 95% correlation coefficient cut off value of 0.41. Therefore, all the correlations could be suggestive of false associations due to random chance. Studies CANDIDIASIS and SV represent moderate to low Spearman correlations for most bacteria groupings. However, study SV has a relatively high 95% Spearman correlation significance cut off value of 0.55, thus suggesting that any suggested correlations between *Gardnerella, Atopobium*, *Dialister* and *Prevotella* could be due to random chance. Instead study CANDIDIASIS displays highly possible correlations between the anaerobes with a very low 95% Spearman threshold 0.27. Consequently, the correlations between *Gardnerella, Atopobium*, *Dialister* and *Prevotella* in study CANDIDIASIS are highly significant even though not as strong in comparison to studies HIV and HSV2.

Interestingly, the exceptionally low correlation values between the four dominant bacteria of study BV, provides further support for the positive correlation discussed previously between *Dialister* and *Prevotella*. *Gardnerella* is not included in the ten most abundant bacteria for study BV (Table 2), thus the low Spearman correlation values could be driven by lower relative abundance hindering the correlation due to other more dominant associations. However, this is not the case with *Atopobium,* as it consists of nearly 5% of total study bacteria abundancy. Nevertheless, *Atopobium* shows no correlation with *Gardnerella*, *Dialister* or *Prevotella*, with correlation values close to zero. As mentioned previously the Spearman correlations don't meet the threshold, thus suggest increased chances that a correlation between these key members of most vaginal microbiomes are due to chance and not true correlations. The results for study BV coincide with Muzny et al. 2013 findings as dysbiotic females influenced by BV experienced a decrease in both *Atopobium* and *Prevotella* [83]. It is possible, that the unbalance of the dysbiotic vaginal environments affected the overall structure of the communities within the microbiome, which could explain the low yet positive correlation observed between *Dialister* and *Prevotella.*

Equally for study SV the Spearman correlation coefficient values are relatively low (Table 4). Study SV aimed to identify the effect of sexual intercourse on vaginal microbiomes. For that reason, both vaginal and seminal samples were collected [82]. Seminal samples were significantly more diverse as seen on Figure 22, however most highly abundant bacteria were shared with females, whereas unique exclusively male bacteria remained in very low abundances. Therefore, the overall total abundance table (Table 2) was not skewed by male bacteria. *Dialister*, *Prevotella* and *Gardnerella* remained dominantly shared within both male and female samples. Even though the correlation between *Dialister* and *Prevotella* was an almost perfect correlation (0.91), no other correlations were significantly high between *Dialister*, *Prevotella, Gardnerella* or *Atopobium* (correlations did not meet the 95% Spearman correlation cut off value 0.55). It is not possible to argue that the lack of correlations derives from seminal bacteria diversity abundancy, as percentages were low. Table 2 enlists *Gardnerella* as the second most abundant bacteria (16%) in the study following *Lactobacillus*. If correlations were dependant on relative abundancy, *Gardnerella* should have displayed strong positive correlations between these dominant members of the vaginal microbiome (as seen on heatmap in Figure 22). However, the correlation results on Table 4 contradict this theory, thus ensuring correlations to be a result of true microbiome community associations between the bacteria. In other words, study SV does not support evidence of strong positive or negative correlations between *Dialister*, *Prevotella, Gardnerella* or *Atopobium* but displays an almost perfect, reliable correlation between *Dialister* and *Prevotella.*

Study CANDIDIASIS represents a few moderately high positive correlations (Table 4). *Dialister* and *Atopobium* represent the highest Spearman correlation with a value of 0.58 which significantly exceeds the 95% correlation threshold 0.27. Both *Dialister* and *Atopobium* are listed in Table 2 within the top ten most abundant bacteria. *Atopobium* demonstrates a 4-fold abundance over *Dialister*, the low positive correlation could be a result of few metabolic interactions between the two organisms. *Atopobium* appears to depict similar patterns with *Prevotella* with Spearman value of 0.53. Equally *Dialister* and *Gardnerella* illustrate low positive correlations with Spearman value of 0.50 and higher almost 10-fold abundance difference between the two bacteria. The significance of the correlations suggests purposeful correlations and not ones driven by bacteria abundance.

Finally studies HSV2 and HIV illustrate strong high positive correlations between all commonly shared bacteria (*Dialister*, *Prevotella, Gardnerella* and *Atopobium*). Both studies enlist all four bacteria in Table 2 displaying their top ten highest abundance bacteria. In study HSV2 *Dialister*, *Prevotella, Gardnerella* and *Atopobium* correlations remain above 0.55, with the highest one between *Atopobium* and *Dialister* with a Spearman correlation value of 0.71. Thus all higher than the estimated 95% Spearman correlation threshold verifying confidence in correlations. In conclusion, study HSV2 suggests microbiome stability and community structure based on the associations between bacteria *Dialister*, *Prevotella, Gardnerella* and *Atopobium*. Similarly study HIV represents even stronger positive correlations between *Dialister*, *Prevotella, Gardnerella* and *Atopobium*. In this case Spearman values exceed 0.67 values illustrating stable correlations within communities of the microbiomes. Once again the correlations exceed the correlation confidence threshold, validating the statistical likelihood of the correlations. Strong, non-random correlations between organisms are not driven by bacteria abundance. Studies HIV and HSV2 included dysbiotic female samples in their datasets, resulting to overall diverse microbiome composition. Despite the fluctuating microbiome composition usually reported in dysbiotic vaginal microbiomes, correlation between *Dialister*, *Prevotella, Gardnerella* and *Atopobium* remained prominent. Thus suggesting that *Dialister*, *Prevotella, Gardnerella* and *Atopobium* could be playing a key role in microbiome stability and assist organisation of structured microbiome communities. In conclusion, strong correlations between bacteria illustrate microbiome community associations linked to community structure.

Although these correlations appear significant and are supported by both clustering and statistical analyses, it can be argued that correlations were biased due to subjective selection of abundant species. Figure 23 depicts the asymmetrical table generated for study HSV2 consisting the Spearman correlation coefficient data. The blue highlighted cells represent values above 0.6, thus the majority of study's HSV2 Spearman correlation values are low, most of which do not meet the 95% correlation threshold (0.57). Consequently, the high correlations enlisted are not false positives or exclusive representations of high abundances, but true representations of community association and structure.

Lastly, the excessive dominance of *Lactobacilli* could be responsible for driving the strong correlations. In other words, strong positive correlations could be originating from negative correlations between the anaerobes (*Dialister*, *Prevotella, Gardnerella* and *Atopobium)* and *Lactobacilli,* due to their excessive abundances. Studies HIV, CANDIDIASIS, HSV2, SV are dominated by *Lactobacilli*. To avoid ambiguity of the results, correlations between *Lactobacilli* and the key anaerobes were examined (Table 5). As expected, the majority of values illustrate negative correlations driven by the *Lactobacillus* dominance, however the values remain at low levels close to zero signifying the lack of significant correlation, with no correlations meeting the 95% confidence correlation threshold. Additionally, no study dependent patterns are detected, demonstrating the lack of bacteria abundance influence on bacterial relationships. On the contrary, study BV presents *Shuttleworthia* as the most dominant bacteria (30% of total abundance – Table 2) and not *Lactobacillus*. *Shuttleworthia* are anaerobic, Gram-positive bacilli characterised in human oral microbiomes [101]. Muzny et al. 2013 show association between *Shuttleworthia* and BV infected patients within increased microbiome diversity [83]. The correlations observed between *Shuttleworthia* and the key anaerobes are representative of the patterns seen with *Lactobacilli*. In other words *Shuttleworthia* abundancy does not drive the correlations between the anaerobes, as no correlation relationships between *Shuttleworthia* and *Dialister*, *Prevotella, Gardnerella* and *Atopobium* meet the 95% Spearman correlation threshold. Therefore, once again the correlations observed between *Dialister* and *Prevotella* in study BV are not a result of random association due to high abundances.

	HIV	BV	CANDIDIASIS	SV	HSV2
Lactobacillus - Gardnerella	-0.28	0.23	-0.15	-0.07	-0.36
Lactobacillus - Atopobium	-0.40	0.22	-0.17	-0.20	-0.55
Lactobacillus - Prevotella	-0.47	-0.19	-0.19	-0.39	-0.32
Lactobacillus - Dialister	-0.38	-0.05	-0.16	-0.35	-0.54
Shuttleworthia - Gardnerella	0.40	-0.15	No Shuttleworthia	0.12	0.50
Shuttleworthia - Atopobium	0.50	-0.23	No Shuttleworthia	0.42	0.31
Shuttleworthia - Prevotella	0.57	-0.02	No Shuttleworthia	0.43	0.45
Shuttleworthia - Dialister	0.41	0.12	No Shuttleworthia	0.42	0.45

Table 5: Spearman Rank Correlation Coefficient analysis between *Lactobacilli and Shuttleworthia* with the key bacteria *Dialister, Prevotella, Gardnerella* and *Atopobium.* Table 5 displays the correlation values to test the nature of the association between organisms *Dialister*, *Prevotella, Gardnerella, Atopobium* and Shuttleworthia. Green correlation values record correlations that met the correlation cut off threshold, thus ensuring certainty in the correlation. Samples from study CANDIDIASIS did not include Shuttleworthia, thus no correlations between *Shuttleworthia* and the other key anaerobes could be reported.

Further investigation is needed to test the nature of correlations between *Lactobacillus, Prevotella* and *Atopobium* bacteria. In conclusion, although the *Lactobacillus* dominance is prominent and affects microbiome associations, most positive correlations are not driven by *Lactobacillus* dominance with most of the impact emerging from structured microbiome communities. Most studies revealed apparent correlations between key anaerobic bacteria (*Dialister*, *Prevotella, Gardnerella* and *Atopobium)*, however not universally shared patterns. However, it is possible to conclude that *Dialister* and *Prevotella* correlation is a consistently observed relationship among all five studies, suggesting key involvement in metabolism and microbiome structure.

6. DISCUSSION

Although vaginal microbiomes have been extensively studied, their association to health and disease is commonly focused on one disorder, with few studies investigating similarities between medical syndromes [1], [2], [13]. Although the analysis proposed here was not able to report links between the microbiomes and medical syndromes associated with the sample donors; access to metadata information would allow further insight into the disorders. Suggesting similarities between vaginal medical disorders such as bacterial vaginosis (BV), HIV and even gonorrhoea, could prove very useful not only in understanding the metabolic structure of the disease but may also propose new approaches to diagnostics and treatment. Mimee et al. 2016 review the possible strategies to approach host treatment through microbiome manipulation as well as the challenges faced when developing "microbiome-based therapeutics" [102]. They discuss three possible therapies; through addition of natural or engineered bacteria; exclusion of harmful bacteria and finally "modulatory therapies" administrating non-living agents or prebiotics to manipulate microbiome communities [102]. Probiotics have been discussed in multiple studies for their suitability in treating the dysbiotic microbiome environments which are the root cause of certain disorders [6], [10]. For example, probiotics have been shown to benefit Inflammatory Bowel Disease by preventing pathogenic bacteria growth and improving immunity by increasing intestinal barrier function and regulating the host's immune response [103].

This study focused on characterising links between bacteria and microbiome ecosystem in various dysbiotic vaginal samples. A pipeline was designed, utilising publicly available data from five individual studies; [27], [81]–[84]. Each selected study focused on different dysbiotic vaginal environments. This study aimed to identify unique characteristics within each microbiome by examining and comparing microbiome community structures. The compositional diversity between the microbiomes, as well as interactions between the bacteria comprising the microbiomes were studied via various bioinformatic tools. Dysbiotic vaginal microbiomes are expected to be linked to a breakdown of microbial community composition and function.

The results suggest multiple strong correlations between specific organisms of the microbiomes associated with dysbiotic microbiome ecosystems. In other words, correlations between bacteria appeared stronger in some dysbiotic samples but not necessarily in others. However as mentioned before, no links could be drawn between specific microbiome structures and existing medical conditions due to the lack of metadata information. Interestingly, the analysis here suggests a novel correlation between *Dialister* and *Prevotella* genera which appears consistently strong and significant between all five selected studies.

6.1 Reviewing studies and analysis outcomes

Study HIV (Gosmann et al. 2017) investigated HIV susceptibility in healthy, asymptomatic South African women with atypically variant microbiomes. They report samples with high diversity bacterial communities and individuals with lower than average *Lactobacillus* abundance [27]. The bar taxonomies designed here (Figure 3,8), display matching patterns with the results as presented by Gosmann et al.(2017), illustrating a highly diverse system, with strong intrapersonal bacteria variance within individuals. Although the majority of the samples contain *Lactobacillus* they do not dominate the ecosystem. In fact, Table 2 confirms the low abundance of *Lactobacilli*, with *Lactobacilli* representing only 40% of total relative abundance with only a 2-fold increase compared to the second most common bacteria (*Prevotella*). Gosmann et al. 2017 analysis equally supports high abundances of *Prevotella* and classify *Prevotella*, *Gardnerella*, *Shuttleworthia*, *Sneathia* as key anaerobes responsible for increased inflammation and thus increased HIV acquisition. The pipeline followed here investigated correlation relationships between all bacteria comprising the microbiomes. Clustering analyses (Figure 13,18) confirmed stable shared community structures in patients. Supporting the hypotheses that specific bacteria-bacteria correlations would appear more prevalent in some dysbiotic microbiomes, Tables 3,4 and 5 list Spearman correlation values between key genera *Lactobacillus*, *Prevotella*, *Gardnerella*, *Atopobium* and *Dialister*. Although the Gosmann et al. 2017 study does not report any strong bacteria-bacteria associations, the heatmap presented here shows correlations between *Prevotella*, *Gardnerella*, *Atopobium* and *Dialister* taxa; which were further confirmed by Spearman Rank Correlation coefficient test. Spearman correlation values exhibit the highest positive correlation between *Dialister* and *Prevotell*a (ρ=0.81, 95% confidence threshold = 0.5). As previously mentioned, the significance of the Spearman correlation test was calculated, representing 95% confidence in the observed bacteria associations. Interestingly, the lack of strong negative correlations between *Lactobacilli* and the key bacteria (*Prevotella*, *Gardnerella*, *Atopobium* and *Dialister*), suggests a true association between them, which is not driven by the dominance of *Lactobacilli*. The strong correlations suggest probable metabolic relationships between the bacteria. Furthermore, the correlations between taxa *Prevotella*, *Gardnerella*, *Atopobium* and *Dialister* (Table 3,4), suggest stable community structures in dysbiotic microbiome communities associated with HIV susceptibility.

Liu et al. 2013 characterised composition and diversity between women with vulvovaginal candidiasis (VVC), bacterial vaginosis (BV) and finally women infected with both vulvovaginal candidiasis (VVC) and bacterial vaginosis (BV). Their data were utilised for the purpose of this research and accessed through their SRA project accession code ERP003902. Liu et al. 2013 report high diversity and intrapersonal variation within VVC patients [81]. Due to the lack of patient metadata information, the analyses followed here could not distinguish patterns specific to disorders. However, the taxonomy bar charts equally illustrated highly diverse communities as well as high species richness within individual samples (Figures 4,9). Reviewing the relative abundance data from all patient samples (Table 2), *Lactobacillus, Gardnerella, Streptococcus* and *Atopobium* were reported as the most abundant taxa, with *Lactobacilli* dominating the majority of samples. Moreover, almost 80% of the study's bacterial abundance is explained by *Lactobacilli.*

Liu et al. 2013 associated microbiome composition to medical syndromes and correspondingly identified BV patients consisting of higher *Gardnerella*, *Atopobium*, *Dialister*, *Sneathia*, *Mobiluncus*, and *Prevotella* with lower than typical *Lactobacilli*; BV and VVC infected patients illustrating microbiome patterns of both BV and "normal" microbiomes (*Lactobacilli* dominance followed by increased levels of *Prevotella, Gardnerella* and *Atopobium*); and finally VVC infected patients displaying high abundancy of *Lactobacilli* (lower levels than in normal vaginal microbiomes) and multiple microbiome community profiles. Unlike their reported compositional distinction between medical disorders, the results created by the suggested pipeline here, do not report patient sample clustering (Figures 14, 19). This could be a result of sample ID de-multiplexing, instead of the barcode and linker primer sequence methodology carried out in studies HIV, BV and HSV2.

Confirming the compositional results reported by Liu et al. 2013 hierarchical clustering (Figure 19) suggested links between *Gardnerella, Atopobium, Prevotella* and *Dialister*. The links were further confirmed as significant bacterial correlations through Spearman correlation statistical analysis. The results exposed a previously unreported strong positive correlation between *Prevotella* and *Dialister* bacteria. Even though *Lactobacilli* abundancy would be expected to drive most of the microbiome community patterns, both the heatmap (Figure 19) and Spearman correlation analyses (Table 5) verify no significant impact from *Lactobacilli* abundancy. The lack of sample patient information limited the pipeline and the possible outcomes for this study, however it was able to reveal strong correlations between key bacteria responsible for dysbiotic environments, suggesting stable microbiome community structures probably associated with metabolic function.

Study BV focused on identifying microbiome composition patterns between various "sexual risk behaviour groups". Muzny et al. 2013 support the theory that BV is a sexually transmitted disorder, which is more prevalent in females who have sexual intercourse with females. For that reason, they studied BV infected females who have sexual intercourse with men, women and finally both men and women [83]. Against their hypothesis, they report more diverse communities between women that have sex with men, with their microbiomes consisting high abundances of *Atopobium*, *Parvimonas* and *Prevotella*, all key bacteria in charactering BV. Additionally, they report exceptionally high levels of *Lachnospiraceae* abundance. *Lachnospiraceae* is a family level taxon, reported by Muzny et al. 2013 to be highly specific to BV infections [83].

Correspondingly the analysis suggested here reports nearly identical relative abundance percentages to the results of Muzny et al. 2013, across all samples [83]. *Shuttleworthia* (originating from *Lachnospiraceae* family taxa), *Prevotella, Lactobacillus, Sneathia, Megasphaera, Atopobium, Parvimonas, Dialister* are reported as the most abundant genera in descending order in both studies (Table 2). Study BV did not include metadata details of the patients in NBCI database, thus the analysis was not able to distinguish compositional patterns between sexual groups. However, the taxonomic bar charts generated (Figures 5,10), illustrate similar patterns of high intrapersonal variation between some patients, whereas others are almost entirely dominated by *Shuttleworthia*.

The results of Muzny et al. 2013 illustrate clear sample grouping between individuals depending on microbiome composition based on sexual behaviour [83]. Equally, both PCA and heatmap Figures (Figures 15, 20) demonstrate clear sample clustering. However, the heatmap's sample dendrograms illustrates smaller groups consisting of multiple clustering pairs. This could be explained through the microbiome composition similarities between the patients, as all females were infected with BV. Although each sexual group was characterised by specific microbiome profiles, BV is characterised by key organisms which were shared between all samples. Therefore, when analysing bacterial abundances of the complete study sample, patterns would not be easy to distinguish. The driving influence of the PCA clusters cannot be easily determined, however it has been proposed that cluster A (the most populated cluster) could be generated due to *Shuttleworthia* dominance.

In their analyses Muzny et al. 2013 did not carry out correlation tests [83]. Spearman's rank correlation coefficient implemented here reveals significant positive correlations between *Dialister* and *Prevotella* taxa, but not between other key BV specific anaerobic bacteria, as expected. *Prevotella, Atopobium, Gardnerella* and *Dialister*, demonstrate low, insignificant correlations, suggesting no association between the bacteria. Study BV does not support links between *Prevotella, Atopobium, Gardnerella* and *Dialister* despite their high abundances and clustering observed in the heatmap graph. Unlike previously discussed studies HIV and CANDIDIASIS, no strong microbiome communities or structures can be confirmed. However, *Shuttleworthia, Prevotella, Lactobacillus, Sneathia, Megasphaera, Atopobium, Parvimonas, Dialister* are once again confirmed as key organisms associated with BV infection. Additionally, study BV supports bacterial associations between *Dialister* and *Prevotella* enhancing the hypothesis of a metabolic link between them.

Study HSV2 aimed to identify links between cervicovaginal microbiomes, genital immunology and HSV-2 infection in African, Caribbean and Black (ACB) women. Shannon et al. 2017 sampled patients with diagnosed BV, HSV-1, HSV-2, intermediate vaginal flora, papillomavirus, and yeast infections [84]. For the purpose of their analysis, samples were grouped dependant on "community state types" with one group representing samples dominated by *L. crispatus* (CST-I), another dominated by *L. gasseri* (CST-II), a third group dominated by *L. iners* (CST-III) and finally a fourth group representing low *Lactobacilli* abundances with increased diversity and abundances of anaerobes (CST-IV). They report the highest level of diversity within patients consisting of low *Lactobacilli* and high anaerobes. Additionally, sample group three (CTS-III) and four (CST-IV) included BV infected patients thus increasing the overall bacteria richness. On the other hand, sample groups one (CST-I) and two (CST-II) were characterised as healthy vaginal microbiomes. CST-III and CST-IV were associated with genital inflammation and proinflammatory cytokines. Although Shannon et al. 2017 identified synergy between BV and HSV-2, they were not able to report links between the sample groups (CTSs) and HSV-2 infection [84].

Following the same patterns, the taxonomic analyses performed here (Figures 16,21), demonstrated high *Lactobacillus* abundancy with low total diversity including a number of samples which appeared monoclonal; whereas others illustrated higher intrapersonal variation. This could be explained by the presence of diseased or infected patients discussed by B. Shannon. Therefore, it can be hypothesised that more diverse samples reflect infected individuals grouped in CST-III or CST-IV, while monoclonal samples dominated by *Lactobacilli* represent healthy individuals grouped in CST-I or CST-II. However due to the lack of metadata information the medical diagnosis or CST condition of the samples was not available, thus prohibiting confirmation of the hypothesis. Interestingly, the dendrograms presented on the generated heatmap (Figure 21) illustrate clear clustering between patients' dependant on *Lactobacillus* presence as well as anaerobic bacteria diversity. The red sample cluster depicts low *Lactobacilli* abundance, followed by diverse anaerobic communities. Therefore, it can be suggested that dysbiotic microenvironments are a result of BV, HSV-1, HSV-2, intermediate vaginal flora, papillomavirus, or yeast infections, whereas the green sample cluster illustrates healthy individuals.

Overall high *Lactobacilli* abundance is supported by Table 2, where a 4-fold dominance of *Lactobacilli* against *Gardnerella* is evident. PCA analysis in Figure 16 additionally supports this dominance by demonstrating a single strong cluster driven by high *Lactobacillus* abundance samples. The pipeline suggested here performed Spearman correlation analysis, focusing on the primarily dominant anaerobes *Prevotella, Atopobium, Gardnerella* and *Dialister.* The results suggested significant strong positive correlations between all anaerobes (Tables 3,4). *Dialister* and *Prevotella* displayed the greatest positive correlation relationship between the anaerobes. To ensure that the correlations were not driven by the communities' lack of *Lactobacilli*, Table 5 lists the Spearman correlation values of the anaerobes against *Lactobacilli*. All correlations were insignificant between the key anaerobes and *Lactobacilli* thus implying the associations are due to interactions between the anaerobes, rather than due to the absence of *Lactobacillus*, possibly linked to metabolic functionality. Even though the analysis could not report links between causation of dysbiosis, study HSV2 supports dysbiotic microbiome community structures with low *Lactobacilli* abundance and correlations between *Dialister* and *Prevotella* as well as with other key anaerobes potentially reinforced by metabolic associations.

Finally study SV aimed to identify the impact of sexual intercourse on vaginal microbiota. Consequently, complementary seminovaginal microbiomes were studied prior and post sexual intercourse [82]. Mändar et al. 2015 reveal seminal microbiomes with increased diversity communities and low bacterial abundances and no predominant bacteria; in comparison to nearly homogenous vaginal communities dominated by *Lactobacilli* or *Gardnerella vaginalis* (dominant in half of female samples associated with Leukocytospermia). Additionally, Mändar et al. 2015 identified four men dominated by *Prevotella* and *Porphyromonas* and suggest a possible association with inflammation in the upper genital tract. Their results presented shared organisms between seminal and vaginal microbiomes (such as *Lactobacillus, Veillonella, Streptococcus, Porphyromonas and Atopobium*). Investigating microbiome composition, they identified male patients with high *Porphyromonas* abundances, others with high proportions of *Prevotella* sp. followed by high presence of *Porphyromonas*. On the other hand, most vaginal microbiota were consistent of *Lactobacillus iners* and *Lactobacillus crispatus,* as well as *Lactobacillus jensenii* and *Lactobacillus gasseri*, with some patients revealing *Gardnerella vaginalis* as the most dominant species, and other females listing *Streptococcus, Enterobacteriaceae, Veillonella, Pseudomonas, Atopobium and* other aerobic communities. After intercourse, Mändar et al. 2015 reveal a significant decrease of *Lactobacillus crispatus* relative abundance in females driven by seminal microbiomes. Thus they conclude significant concordance between seminal and vaginal samples with regards to *Gardnerella vaginalis* predominance (in vaginal microbiomes) and association to inflammation in male genital tracts.

The bioinformatics approach followed here utilising data from study SV, aimed to investigate interactions between and within atypical (dysbiotic) vaginal microbiomes. For that reason, seminal samples were not suitable for this analysis. However, the data uploaded in NCBI's database did not include sample identify information at the stage of data acquisition, thus distinguishing between male and female samples was not possible. For that reason, both male and female samples were included in the analysis even though the focus remained on female microbiota.

When analysing the bar taxonomies generated, male samples are easily distinguished in 23 samples (ERR769967-ERR769989). Male samples consisted of drastically lower *Lactobacilli* abundances with the microbiome communities not compensating with additional lactic acid producing bacteria, such as *Atopobium, Corynebacterium, Anaerococcus, Peptoniphilus, Prevotella, Gardnerella, Sneathia*, as observed in asymptomatic atypical female vaginal microbiomes [104]. Male samples illustrate significantly higher bacterial diversity with no single dominating taxa. Equally female samples follow the same patterns as discussed in Mändar et al. 2015 results, with most samples representing *Lactobacilli* dominance and fewer samples revealing *Gardnerella* predominance. Table 2 demonstrates a 3-fold universal dominance of *Lactobacilli* over *Gardnerella*. Although male samples exhibited high genera richness, the heatmap displayed in Figure 22 illustrated low relative abundances. Therefore, the overall study's bacteria abundances presented on Table 2 will not be affected by male samples and the dominating relationships between bacteria will be driven by vaginal microbiota.

It is difficult to suggest a driving force of the PCA clustering (Figure 17) as samples are clustered according to sample sex identity (male/ female) and microbiome composition. However, cluster A will be driven by the high abundance of *Lactobacilli*, due to the large number of samples comprising it. Interestingly the PCA is suggesting association and composition similarities between seminal and vaginal samples as more than two clear clusters exist not exclusively dependant on samples gender identity. On the contrary, the sample dendrogram represented on Figure's 22 heatmap, illustrates clustering between male and female samples, with the green cluster representing female samples and the red cluster displaying male samples. Once again, the heatmap graph supports evidence of shared organisms between seminal and vaginal samples as well as suggesting association between the sample communities. Bacteria such as *Prevotella, Gardnerella*, *Dialister, Veillonella*, *Flavobacterium,* and *Corynebacterium* are shared between female and male couples (following equal patterns as reported from Mändar et al. 2015). Additional Spearman correlation analysis not examined in the SV study, suggested an almost perfect correlation between *Dialister* and *Prevotella*, but no correlation between other key anaerobes. Correlations would be driven by high bacterial abundances as present in vaginal microbiomes samples. Therefore, the lack of correlation between key anaerobes is a result of low anaerobe colonisation in healthy females. Although *Prevotella, Gardnerella* and *Dialister* are dominant in male microbiota, the low male bacterial abundances would not impact bacterial correlations.

High levels of *Prevotella* and *Gardnerella* in certain female and male samples, identified in both analyses, confirm association between bacteria and dysbiotic environments or genital inflammation. Once again, the lack of *Lactobacillus* dependency of the correlation between *Dialister* and *Prevotella* (Table 5) proves the strength of the correlation and is consistent with a metabolic link. Interestingly, even though study SV mainly sampled healthy vaginal microbiomes dominated by *Lactobacilli*, key bacteria such as *Dialister* and *Prevotella* still indicate strong community associations, suggesting community structures.

To investigate universal microbiome patterns between all studies, PCA clustering was performed on studies HIV, BV and HSV2 (studies CANDIDIASIS and SV were excluded due to de-multiplexing compatibility issues). Figure 25 demonstrates a 3D PCA plot for studies HIV, BV and HSV2, where each colour represents a different study (green colour samples represent patients from study HIV, black samples depict study's HSV2 samples and finally blue samples originate from study BV). Figure 25 does not illustrate distinct sample clustering, even though studies were shown to share microbiome community patterns. Instead samples appear clustered by study. Despite PCA's short-coming on comparing between studies, the analysis suggested here reports strong patterns of shared bacteria, (as seen from hierarchical clustering, heatmaps and Spearman correlation tests), particularly with respect to the *Dialister* and *Prevotella* link. The lack of clustering between studies, could be due to sample preparation, sequencing techniques and number of reads. Although bacteria *Prevotella, Atopobium, Gardnerella* and *Dialister* were universally abundant in most samples of the studies, correlations between them were only proven in studies HIV, CANDIDIASIS and HSV2. However, *Dialister* and *Prevotella* associations were strongly shared between all five studies (Table 3). All studies with the exception of BV demonstrated significant dominance of *Lactobacilli* irrespective of the microbiome's condition focus for each study. Therefore, it can be argued that any potential clustering observed in Figure 25 would be driven by the overpowering abundance of *Lactobacilli*.

To test this hypothesis *Lactobacilli* were removed from all studies individually and PCA was performed again. Figure 26 illustrates a 3D PCA plot of the same studies excluding *Lactobacilli* from its samples. The hypothesis is confirmed as all studies cluster independently from each other, with each cluster representing a single study containing exclusively its original samples and no other samples originating from other studies. Hence, any clustering suggested on Figure 25 would be driven by *Lactobacilli* with no other microbiome community similarities presented within patients. The lack of sample clustering does not affect the significance of the bacteria-bacteria correlations reported, but is instead suggesting the significance of sampling methodologies. In other words, the clustering observed in Figures 25 and 26 are exclusively dependent on sequencing and sampling methodologies rather than inter-personal variation between individuals. Therefore, the impact of sequencing techniques, sample preparation and number of sequence reads will affect the nature of the samples, creating additional variation between the sequences of a specific study. PCA clustering represents sequencing methodology similarities between samples of the same study. In conclusion the absence of clustering is driven by the difference in sample "type" and does not reflect microbiome structure similarities between dysbiotic environments.

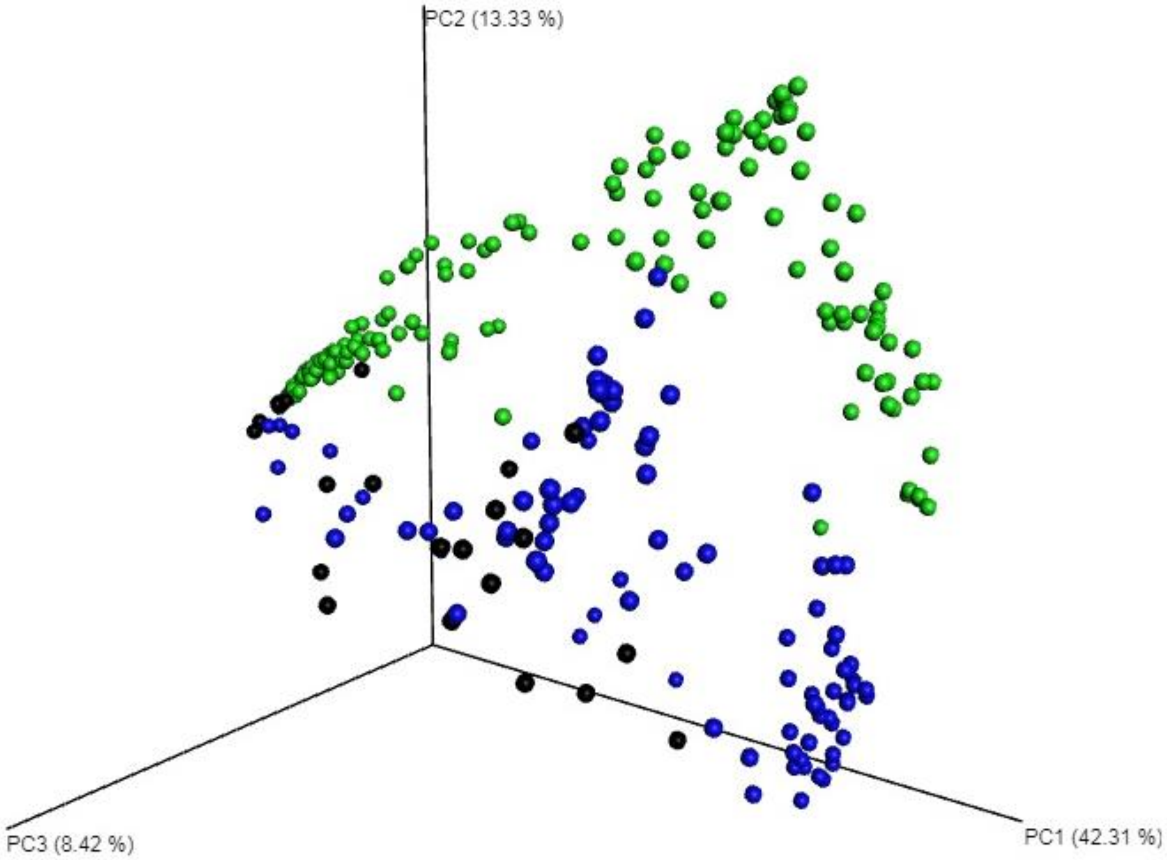


Figure 25: Principal Component Analysis 3D graph for studies HIV, BV and HSV2. Sample data from all three studies plotted against the first three principal components representing the highest levels of variance explained by each principal component. Each study is presented by a different colour; with data from study HSV2 illustrated in black, data from study HIV in green and data from study BV in blue. The 3D plot suggests no clear sample clustering.

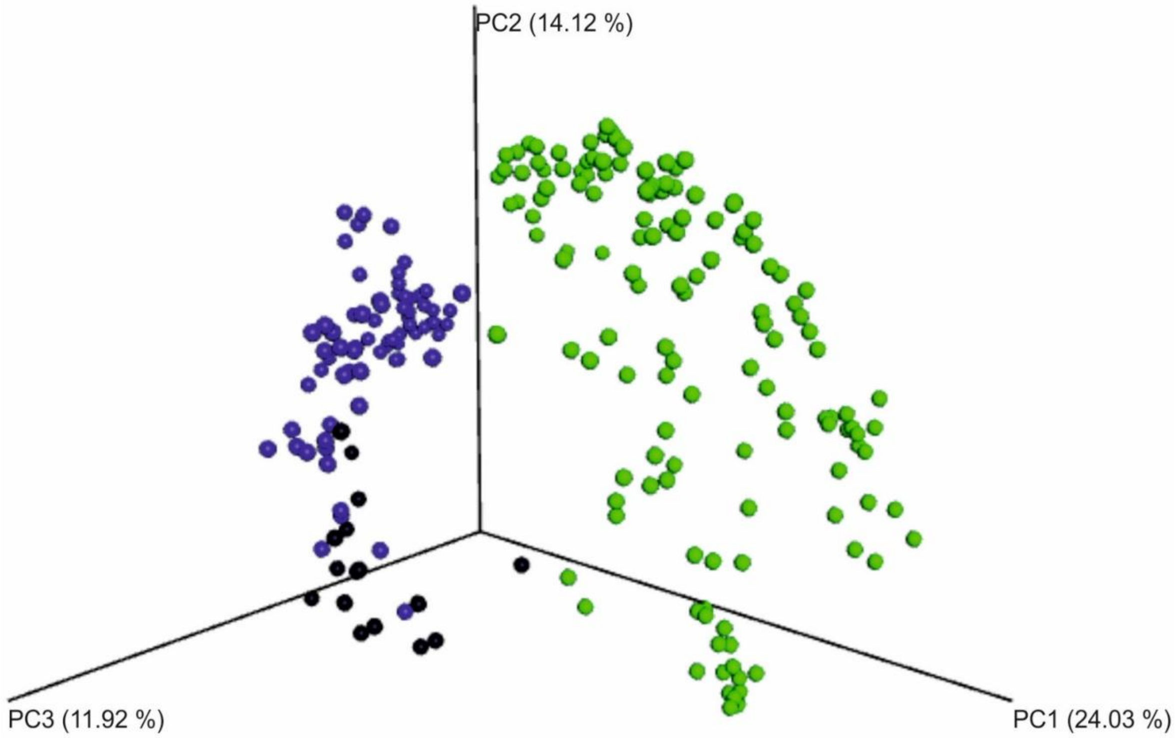


Figure 26: Principal Component Analysis 3D graph for studies HIV, BV and HSV2 excluding *Lactobacilli* taxonomies. Sample data excluding *Lactobacilli* taxonomies from all three studies plotted against the first three principal components, representing the highest levels of variance explained by each principal component. Study HIV is illustrated by green data points, study HSV2 is represented by black data points and finally study BV is depicted by blue data points. The removal of *Lactobacilli* taxonomies confirms the lack of sample clustering, as the only suggested clusters illustrated are study dependent

6.2 Supporting evidence from literature

The five studies selected for this research (HIV, CANDIDIASIS, BV, HSV2, SV) report evidence of microbiome structures, with composition suggesting some direct bacteria-bacteria interactions. Relationships between members of the microbiome suggest structure within vaginal microbiomes. Even though dysbiotic communities fluctuate in composition and are therefore less structured, compared to atypical, asymptomatic vaginal microbiomes; some microbiome community structures appear consistent in all environments. An example of this is the shared correlation observed in all five studies between *Dialister* and *Prevotella.* Dysbiotic environments illustrate less established microbiome structures with microbiome composition varying between stages of dysbiosis. Therefore, some correlations between bacteria appear more convincing within certain studies, even though present in most studies.

Dysbiosis in vaginal microbiomes is not characterised by a single microbiome composition or structure but instead describes a range of microbiome states from typical to “unhealthy”. Independent of the samples’ microbiome states, all five studies demonstrated strong positive interactions between *Dialister* and *Prevotella* bacteria. All studies listed both organisms within the top ten most abundant organisms with *Prevotella* always dominating *Dialister* in abundancy. This evidence suggests a stable relationship between the bacteria regardless of the microbiome state. Thus, structured communities can be driven by *Dialister* and *Prevotella* bacterial interactions, with individuals carrying both organisms consisting of more stable microbiomes. Although none of the five selected studies reported the *Dialister* and *Prevotella* correlation in their published papers, when analysed by different tests the correlation is abundantly present. All results presented here propose the possibility of a metabolic functionality link between *Dialister* and *Prevotella*, which needs to be further investigated.

A different study by Srinivasan et al. 2012 focused on BV infected microbiomes and does not report a correlation between *Dialister and Prevotella,* although their analysis on Figure 4 of their published paper illustrates strong positive correlations between the bacteria [88]. The correlation was possibly overlooked due to their focus on higher positive correlations, even though their supplementary table lists positive correlations between multiple *Dialister and Prevotella* species with Pearson values ranging from 0.2 to 0.7 (P <0.05) (Srinivasan et al. 2012 – Supplementary Table S7). In line with the results presented here, they speculate polyamine (such as putrescine, cadaverine, and trimethylamine) metabolic correlations, due to “amine odour” of samples [88]. “Amine odour” reported from a Whiff tests has been linked to BV infected individuals, representing increased species richness, increased anaerobe abundances (such as *Atopobium vaginae*, *Veillonellaceae*, *Prevotella* spp., BVAB1 and *Dialister micraerophilus*) and decreased *Lactobacilli* [88]. Therefore, polyamine metabolic correlations coincide with *Dialister and Prevotella* correlations, both characterising dysbiotic microbiomes. Additionally, Nelson et al. 2015 report high *Dialister* and *Prevotella* abundances in the presence of low *Lactobacilli,* however they do not report correlation between the genera [99]. C. J. Yeoman state high abundances in both *Dialister* and *Prevotella*, as well as demonstrating correlations with polyamine presence (responsible for vaginal odour), thus suggesting key contribution to BV state. They conclude that BV symptoms could be a result of individual metabolic processes originating from *Dialister* spp., *Gardnerella* spp., *Mobiluncus* spp., or other bacteria [100]. All the above-mentioned associations suggest a correlation between *Dialister* and *Prevotella* to dysbiosis in vaginal microbiomes.

The analyses presented here, reported a unique reoccurring significant correlation between vaginal bacteria *Dialister and Prevotella*. *Dialister* was isolated and studied in faecal microbiomes, where it was shown that *Dialister succinatiphilus sp. nov*. growth utilises succinate and produces propionate and acetate as end metabolic by-products [87]. On the other hand, *Prevotella intermedia* and *Prevotella nigrescens* produce succinate as an end-product of glucose metabolism [105]. Therefore, it can be speculated that the strong correlations observed in all selected studies for this project, are not only a result of stable microbiome communities but also represent a possible metabolic link between decarboxylation of succinate to propionate. In other words, it is probable that *Prevotella* produces succinate as an end-product of its glucose pathway, which is then utilised by *Dialister* to enhance anaerobic growth and generate propionate. Thus, suggesting a linear metabolic pathway illustrated in Figure 27.

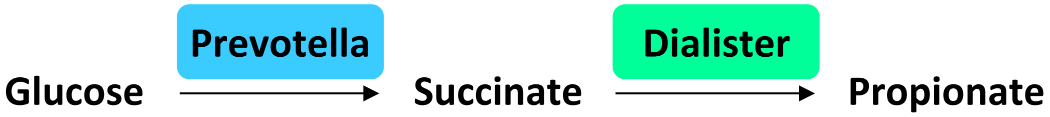


Figure 27: Diagram of the suggested metabolic correlation between *Dialister* and *Prevotella*. Glucose decarboxylation via *Prevotella* produces succinate, which is then taken up by *Dialister* leading to an end product of propionate.

A potential link between *Dialister* and *Prevotella* through glucose metabolism could explain the association to dysbiosis. Propionate is a key metabolite utilised by various vaginal microorganisms which have been associated with BV infections [106]. In fact L. V. Hill utilised high levels of propionate to characterise BV infected individuals [107]. Therefore, high levels of propionate and succinate would define dysbiotic microbiota communities and more specifically the presence of BV infections. This hypothesis could explain the consistently strong correlations between *Dialister* and *Prevotella*, expressed through this analysis, as all studies included more diverse dysbiotic females. Consequently, it can be suggested that a *Dialister* and *Prevotella* link, if driving propionate metabolism, could prove an association between health, dysbiosis and disease in female genital tracts.

Additionally, other key anaerobes such as *Atopobium* and *Gardnerella* also appeared linked to acquisition and microbiome interactions through this analysis. Various studies have linked *Prevotella, Atopobium* and *Gardnerella* with dysbiotic vaginal microbiota, usually characterised by the reduced *Lactobacilli* abundance [86]. High levels of *Prevotella*, *Gardnerella and Atopobium* have also been correlated with various vaginal inflammation disorders such as HSV-2 infection [84], BV [69], [81], [83] upper genital tract inflammation [82] and HIV susceptibility [27]. For example, increased abundances of *Atopobium* and *Gardnerella* followed by lowered levels of *Lactobacilli* were found in vaginal microbiomes infected by BV, in a study focused on vaginal immunity [108]. The analysis suggested here illustrates similar patterns, where studies consisting of more diverse samples express higher levels of correlations between these key anaerobes. This could be suggestive of a correlation between microbiome composition and susceptibility to infection. Verhelst et al. 2004 report a strong correlation between *Gardnerella vaginalis* and *Atopobium vaginae* in BV infected patients [109], however do not suggest any potential metabolic links between them. Studies HIV, CANDIDIASIS and HSV2 analysed here, equally represent high positive correlations between the two bacteria with Spearman correlation values ranging from 0.5 to 0.8 (95% confidence in correlation < 0.57).

Although *Prevotella*, *Gardnerella* and *Atopobium* are present in healthy, typical vaginal flora [110], high abundances followed by lower than normal levels of *Lactobacilli* appear linked to dysbiosis. Thus, it can be hypothesised that those key anaerobes influence microbiome structure in healthy microbiomes, which if triggered by environmental changes (such as yeast infection, or pregnancy etc.) will lead to an increase in their numbers turning into dysbiosis. The consistency of increased abundances of the anaerobes in multiple vaginosis states suggests a microbiome structure even though environments are dysbiotic. To our knowledge, although no metabolic associations have been suggested in literature the stability of the associations proves their importance for further investigation.

6.3 Restrictions with bioinformatics in microbiome studies

Fast sequencing techniques followed by the collection of large data pools has led to the popular use of microbiome studies accompanied by a massive increase of interest in the –omics fields. Although such studies have proven useful in establishing relationships between an organism and its microbiome, as well as studying links between health and disease, questions are being raised on their scientific implication and approach. Due to the "informal" approach that computational analyses offer, most bioinformatics research starts with a lack of a clear hypothesis [111]. Instead the study turns into a search for results causing the "fishing for significance" phenomenon [112]. "Fishing for significance" is a term that A. L. Boulesteix uses to define the phenomenon of over optimised research results. Due to the rapid progression in bioinformatics tools it is now easy to submit data in online programs to scout for links [113].

Hanno Teeling et al. 2012 discuss the challenges faced when analysing microbiome samples through bioinformatic tools [114]. Data submission generated various challenges in this research due to lack of metadata information as well as annotation imbalances. Teeling et al. 2012 comment on the issue caused by the lack of a universal standardised annotation model implemented on publically available databases [114]. Most accessible online tools permit fast analyses using standardised parameters. Therefore, online tools need to be used with caution. Verifying the tool's parameters and algorithm to be compatible with the data type imported as well as assessing their suitability for the type of analyses carried out can prevent misuse of the programs. Unfortunately, the benefits of open access tools can be abused; due to lack of precaution, as the generated results get presented as research findings with no further supporting evidence or testing.

This analysis aimed to identify compositional links related to microbiome structure in various dysbiotic vaginal microbiomes. A novel correlation between two common vaginal bacteria *Dialister* and *Prevotella* is proposed here. In addition, the study suggested links between specific anaerobic bacteria, such as *Prevotella*, *Gardnerella* and *Atopobium*, and dysbiotic vaginal communities. The reoccurring correlation between them proves of great importance as a metabolic link is theorised. It is proposed that *Dialister* and *Prevotella* are associated through a linear metabolic pathway (illustrated in Figure 27), utilising succinate and producing propionate as an end product. As mentioned before, propionate is a crucial metabolite utilised by various bacteria and is linked to BV infections. Thus, it can be hypothesised that the *Dialister* and *Prevotella* correlation is associated with dysbiotic and more specifically BV vaginal environments. However, these are simple speculations and more analyses need to be run to test this hypothesis. Online tools such as Kyoto Encyclopedia of Genes and Genomes (Kegg) could provide additional information on the potential metabolic links. This would then portray any metabolic association and therefore give means to create metabolic maps. Most importantly, cultural experiments should follow up to demonstrate the metabolic interaction. A possible example of this would be a propionate accumulation study via gas chromatography in mixed population of *Dialister* and *Prevotella* compared to monocultures of each. If over time there is an accumulation of propionate in the mixed population media, in comparison to little or no propionate presence in the monocultures, this could be an indication of a *Dialister* and *Prevotella* metabolic relationship. As a negative control a third non-interacting partner should also be selected from the pool of available vaginal microbiota and also be grown in a mixed population with *Dialister* and then *Prevotella* alone. Such an experiment would help us understand highly functional metabolic pathways vital for the vaginal microbiome and how they are utilised by the organism inhabiting the microbiome as well as human health. Metagenomics applied in microbiome studies could prove very beneficial to diagnostics as well as improve personalized treatments by carefully studying the environmental and ecosystem changes. Above all a potential metabolic link between various dysbiotic communities could provide a whole new approach to diagnosis and treatment of vaginitis.

APPENDICES

Appendix 1:

Python script running in ipython notebook via the Oracle VM Virtual Box, calling all SRA accession codes of a study to be applied to the prefetch command.

```
file1 = open ("E:/Database_files/NCBI/all_sras - Copy.txt")

import subprocess
import time
import sys

count = 0

for line in file1:
    total_count = len(line.split(" "))
    accesion = line[0:10]
    for acc_num in line.split(" "):
        count +=1
        print acc_num + ": " + str(count) + "/" + str(total_count)
+ " (" + time.strftime("%H:%M:%S") + ")"
        sys.stdout.flush()
        subprocess.call(["prefetch", acc_num])
```

Appendix 2:

Fastq-dump command for Linux operating systems. Fastq-dump in a linux shell did not require prior sra download of the sequence files within a study. The command connects to NCBI and downloads each SRR sequence file accessed through the SRR accession code, retrieved via the python script listed in Appendix 1. Parameters X and Z are optional and were added in this study to print the first five spots (-X 5) of the file on the screen(-Z) to ensure success.

```
fastq-dump -X 5 -Z SRR1804553
```

Appendix 3:

Studies PRJNA329618 - Vaginal microbiome of reproductive-age women and PRJNA295859 - Endometrial cancer microbiome were not included in the final analysis as the sequence files for both studies included sequence duplicates and annotation errors. More specifically, the SRR sequence files for both studies did not follow the typical fasta format of unique sequence IDs followed by the sequence:

```
>sequence_name_1
CAGTAACAGACCAGAGAGCCGCCTTCGCCACCGGTGTTCTTCCATATATCTACGCATTTCACCGCT
ACGGCATT
>sequence_name_2
TCTAATTGATTACCGTCAAACAAAGGTCAGTTACTACCCCTGTCCTTCTTCACCAACAACAGAGCT
TTACGAGCT
```

Instead the SRR files contained duplicates of the sequence IDs followed by incomplete sequences and distorted characters (not readable in Linux or Windows operating systems). An example from study PRJNA295859 is illustrated in Appendix 3a.

```
a)
'>SRR2533924\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
```

```
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
```

```
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
```

```
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00
\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x
00\x00\x00\x00\x00\x00GTAGCGTGCAGGATGACGGCCCTATGGGTTGTAAACTGCTTTTA
TGTGGGGATAAAGTGCGTGACGTGTCATGCATTGCAGGTACCACATGAATAAGGACCGGCTAATTC
CGTGCCAGCAGCCGCGGTAATACGGAAGGTTCGGGCGTTATCCGGATTTATTGGGTTTAAAGGGAG
CGTAGGCTGTCTATTAAGCGTGTTGTGAAATTTAACGGCTCAACCGGTGGCTTGCAGCGCGAACTG
GTCGCATTGACTATGGA'
```

A python script was programmed to rewrite all .fna files (Appendix 3b), however the size of the studies caused major time restrictions. Therefore, studies PRJNA329618 and PRJNA295859 were not included in the final analysis.

b) fix fna sequences:

```
import glob

import re

from subprocess import call

study1path = glob.glob('/scratch/tef504/SRP064295/SRR25339*')


nameslist = []

for n in study1path:

    nameslist.append(n[26:36])


seq_name = "bad"

for filepath in study1path:

    infile = open(filepath, 'r')

    outfile = open(filepath.replace('out', 'FIXED'), 'w')


    for line in infile:

        if re.match('^>[A-Z]{3}[0-9]+.*', line):

            if line[1:11] in nameslist:

                seq_name = line


        else:
            if re.match("^[ATGC]+\r\n?|\n$", line) and not
seq_name == "bad":

                sequence = line.replace("\n", "")

                cmd = "echo " + "'" + seq_name + sequence + "'" +
" >> " + (filepath.replace('out', 'FIXED'))

                seq_name = "bad"

                call(cmd, shell=True)
```

Appendix 4: List of Study Accession codes

Study Description	Experiment Accession Number	Designated Abbreviations	SRA Project Accession number
Certain species of vaginal bacteria can increase a woman's susceptibility to HIV	PRJEB15497	HIV	ERP017263
Diverse vaginal microbiomes in reproductive-age women with vulvovaginal candidiasis	PRJEB4606	CANDIDIASIS	ERP003902
Complementary seminovaginal microbiome in couples	PRJEB8658	SV	ERP009682
Characterization of the Vaginal Microbiota among Sexual Risk Behavior Groups of Women with Bacterial Vaginosis	PRJNA259744	BV	SRP045868
Distinct effects of the cervico-vaginal microbiota and herpes simplex type 2 infection on female genital tract immunology	PRJNA310998	HSV2	SRP071021
Vaginal microbiome of reproductive-age women *	PRJNA329618	SRP090242	SRP090242
Endometrial cancer microbiome *	PRJNA295859	SRP064295	SRP064295

Appendix 5:

Pipeline two downloaded sequence files from ENA database in a fastq format. Convert_fastaqual_fastq.py QIIME’s script converts all downloaded fastq files, to more compatible .fna files for all five selected studies.

```
#changing format for fastqs to .fna
import glob
import subprocess
from subprocess import call

files = glob.glob("./fastq_files/fastq_files_PRJNA329618/*.fastq")
list1 = []
for i in files:
    list1.append(i[38:])

for filename in list1:
    cmd_str = "convert_fastaqual_fastq.py -c fastq_to_fastaqual -f
" +
"/home/qiime/Desktop/Shared_Folder/fastq_files/fastq_files_PRJNA32
9618/"  + filename + " -o
/home/qiime/Desktop/Shared_Folder/fasta_files/fasta_files_PRJNA329
618/"
    call(cmd_str, shell=True)
```

Appendix 6:

Split libraries QIIME commands. Appendix 6a illustrates the script employed in the suggested pipeline one. Sequence de-multiplexing was initially based on sequence sample IDs and not Barcodes or Linker Primer sequences. –i argument instructs the input files, in this case all 16S rRNA sequences within a study. --sample_ids define an alias for all sample sequences and –-barcode_type defines whether the sequences contain barcodes or not. --phred_offset is the ascii offset used to decode phred scores. In other words, phred offset 33 value represents the possibility of substitution errors.

```
a) !split_libraries_fastq.py -i
   ERR341370.fastq,ERR341371.fastq,ERR341372.fastq,ERR341373.fas
   tq,ERR341374.fastq,ERR341375.fastq,ERR341376.fastq,ERR341377.
   fastq,ERR341379.fastq,ERR341380.fastq,ERR341381.fastq,ERR3413
   82.fastq,ERR341383.fastq,ERR341384.fastq,ERR341385.fastq,ERR3
   41386.fastq,ERR341387.fastq,ERR341388.fastq,ERR341389.fastq,E
   RR341390.fastq,ERR341391.fastq,ERR341392.fastq,ERR341393.fast
   q,ERR341394.fastq,ERR341395.fastq,ERR341396.fastq,ERR341397.f
   astq,ERR341398.fastq,ERR341399.fastq,ERR341400.fastq,ERR34140
   1.fastq,ERR341402.fastq,ERR341403.fastq,ERR341404.fastq,ERR34
   1405.fastq,ERR341406.fastq,ERR341407.fastq,ERR341408.fastq,ER
   R341409.fastq,ERR341410.fastq,ERR341411.fastq,ERR341412.fastq
   ,ERR341413.fastq,ERR341414.fastq,ERR341415.fastq,ERR341416.fa
   stq,ERR341417.fastq,ERR341467.fastq,ERR341468.fastq,ERR341469
   .fastq,ERR341470.fastq,ERR341471.fastq,ERR341472.fastq,ERR341
   473.fastq,ERR341474.fastq,ERR341475.fastq,ERR341476.fastq,ERR
   341477.fastq,ERR341478.fastq,ERR341479.fastq,ERR341480.fastq,
   ERR341481.fastq,ERR341482.fastq,ERR341483.fastq,ERR341484.fas
   tq,ERR341485.fastq,ERR341486.fastq,ERR341487.fastq,ERR341488.
   fastq,ERR341489.fastq,ERR341490.fastq,ERR341491.fastq,ERR3414
   92.fastq,ERR341493.fastq,ERR341494.fastq,ERR341495.fastq,ERR3
   41496.fastq,ERR341497.fastq,ERR341498.fastq,ERR341499.fastq,E
   RR341500.fastq,ERR341501.fastq,ERR341502.fastq,ERR341503.fast
   q,ERR341504.fastq,ERR341505.fastq,ERR341506.fastq,ERR341507.f
   astq,ERR341508.fastq,ERR341509.fastq,ERR341510.fastq,ERR34151
   1.fastq,ERR341512.fastq,ERR341513.fastq,ERR341514.fastq,ERR34
   1515.fastq,ERR341516.fastq,ERR341517.fastq,ERR341518.fastq,ER
   R341301.fastq,ERR341302.fastq,ERR341303.fastq,ERR341304.fastq
   ,ERR341305.fastq,ERR341306.fastq,ERR341307.fastq,ERR341308.fa
   stq,ERR341309.fastq,ERR341310.fastq,ERR341311.fastq,ERR341312
   .fastq,ERR341313.fastq,ERR341314.fastq,ERR341315.fastq,ERR341
   316.fastq,ERR341317.fastq,ERR341318.fastq,ERR341319.fastq,ERR
   341320.fastq,ERR341321.fastq,ERR341322.fastq,ERR341323.fastq,
   ERR341324.fastq,ERR341325.fastq,ERR341326.fastq,ERR341327.fas
   tq,ERR341328.fastq,ERR341329.fastq,ERR341330.fastq,ERR341331.
   fastq,ERR341332.fastq,ERR341333.fastq,ERR341334.fastq,ERR3413
   35.fastq,ERR341336.fastq,ERR341337.fastq,ERR341338.fastq,ERR3
   41339.fastq,ERR341340.fastq,ERR341341.fastq,ERR341342.fastq,E
   RR341343.fastq,ERR341344.fastq,ERR341345.fastq,ERR341346.fast
   q,ERR341347.fastq,ERR341348.fastq,ERR341349.fastq,ERR341350.f
```

```
astq,ERR341351.fastq,ERR341352.fastq,ERR341353.fastq,ERR34135
4.fastq,ERR341355.fastq,ERR341356.fastq,ERR341357.fastq,ERR34
1358.fastq,ERR341359.fastq,ERR341360.fastq,ERR341361.fastq,ER
R341362.fastq,ERR341363.fastq,ERR341364.fastq,ERR341365.fastq
,ERR341366.fastq,ERR341367.fastq,ERR341368.fastq,ERR341369.fa
stq,ERR341436.fastq,ERR341437.fastq,ERR341438.fastq,ERR341439
.fastq,ERR341440.fastq,ERR341441.fastq,ERR341442.fastq,ERR341
443.fastq,ERR341444.fastq,ERR341445.fastq,ERR341446.fastq,ERR
341447.fastq,ERR341448.fastq,ERR341449.fastq,ERR341450.fastq,
ERR341451.fastq,ERR341452.fastq,ERR341453.fastq,ERR341454.fas
tq,ERR341455.fastq,ERR341456.fastq,ERR341457.fastq,ERR341458.
fastq,ERR341459.fastq,ERR341460.fastq,ERR341461.fastq,ERR3414
62.fastq,ERR341463.fastq,ERR341464.fastq,ERR341465.fastq,ERR3
41466.fastq,ERR341418.fastq,ERR341419.fastq,ERR341420.fastq,E
RR341421.fastq,ERR341422.fastq,ERR341423.fastq,ERR341424.fast
q,ERR341425.fastq,ERR341426.fastq,ERR341427.fastq,ERR341428.f
astq,ERR341429.fastq,ERR341430.fastq,ERR341431.fastq,ERR34143
2.fastq,ERR341433.fastq,ERR341434.fastq,ERR341435.fastq,ERR34
1519.fastq,ERR341520.fastq,ERR341521.fastq,ERR341522.fastq,ER
R341523.fastq,ERR341524.fastq  --sample_ids
SRR1,SRR2,SRR3,SRR4,SRR5,SRR6,SRR7,SRR8,SRR9,SRR10,SRR11,SRR1
2,SRR13,SRR14,SRR15,SRR16,SRR17,SRR18,SRR19,SRR20,SRR21,SRR22
,SRR23,SRR24,SRR25,SRR26,SRR27,SRR28,SRR29,SRR30,SRR31,SRR32,
SRR33,SRR34,SRR35,SRR36,SRR37,SRR38,SRR39,SRR40,SRR41,SRR42,S
RR43,SRR44,SRR45,SRR46,SRR47,SRR48,SRR49,SRR50,SRR51,SRR52,SR
R53,SRR54,SRR55,SRR56,SRR57,SRR58,SRR59,SRR60,SRR61,SRR62,SRR
63,SRR64,SRR65,SRR66,SRR67,SRR68,SRR69,SRR70,SRR71,SRR72,SRR7
3,SRR74,SRR75,SRR76,SRR77,SRR78,SRR79,SRR80,SRR81,SRR82,SRR83
,SRR84,SRR85,SRR86,SRR87,SRR88,SRR89,SRR90,SRR91,SRR92,SRR93,
SRR94,SRR95,SRR96,SRR97,SRR98,SRR99,SRR100,SRR101,SRR102,SRR1
03,SRR104,SRR105,SRR106,SRR107,SRR108,SRR109,SRR110,SRR111,SR
R112,SRR113,SRR114,SRR115,SRR116,SRR117,SRR118,SRR119,SRR120,
SRR121,SRR122,SRR123,SRR124,SRR125,SRR126,SRR127,SRR128,SRR12
9,SRR130,SRR131,SRR132,SRR133,SRR134,SRR135,SRR136,SRR137,SRR
138,SRR139,SRR140,SRR141,SRR142,SRR143,SRR144,SRR145,SRR146,S
RR147,SRR148,SRR149,SRR150,SRR151,SRR152,SRR153,SRR154,SRR155
,SRR156,SRR157,SRR158,SRR159,SRR160,SRR161,SRR162,SRR163,SRR1
64,SRR165,SRR166,SRR167,SRR168,SRR169,SRR170,SRR171,SRR172,SR
R173,SRR174,SRR175,SRR176,SRR177,SRR178,SRR179,SRR180,SRR181,
SRR182,SRR183,SRR184,SRR185,SRR186,SRR187,SRR188,SRR189,SRR19
0,SRR191,SRR192,SRR193,SRR194,SRR195,SRR196,SRR197,SRR198,SRR
199,SRR200,SRR201,SRR202,SRR203,SRR204,SRR205,SRR206,SRR207,S
RR208,SRR209,SRR210,SRR211,SRR212,SRR213,SRR214,SRR215,SRR216
,SRR217,SRR218,SRR219,SRR220,SRR221,SRR222,SRR223 -o
./split_libraries_CANDIDIASIS --barcode_type 'not-barcoded' -
-phred_offset 33
```

Appendix 6b illustrates the split_libraries.py QIIME command utilised in pipeline two, which was applied in YARCC cluster computer in a terminal shell. The example below illustrates de-multiplexing executed on BV study. The sequence de-multiplexing was based on unique Barcodes and universal Linker Primer sequences. –m argument precedes the mapping_tableBV_corrected.txt; –f argument instructs the input sequence.fna files, and –o argument lists the output pathway. The mapping file utilised for split_libraries script followed the same format presented on Appendix 7.

```
b) python /usr/userfs/t/tef504/python/bin/split_libraries.py -m
   /scratch/tef504/BV/mapping_tableBV_corrected.txt -f
   SRR1561443_barcoded_linkedPrimer.fna,SRR1561444_barcoded_link
   edPrimer.fna,SRR1561445_barcoded_linkedPrimer.fna,SRR1561446_
   barcoded_linkedPrimer.fna,SRR1561447_barcoded_linkedPrimer.fn
   a,SRR1561448_barcoded_linkedPrimer.fna,SRR1561449_barcoded_li
   nkedPrimer.fna,SRR1561450_barcoded_linkedPrimer.fna,SRR156145
   1_barcoded_linkedPrimer.fna,SRR1561452_barcoded_linkedPrimer.
   fna,SRR1561453_barcoded_linkedPrimer.fna,SRR1561454_barcoded_
   linkedPrimer.fna,SRR1561455_barcoded_linkedPrimer.fna,SRR1561
   456_barcoded_linkedPrimer.fna,SRR1561457_barcoded_linkedPrime
   r.fna,SRR1561458_barcoded_linkedPrimer.fna,SRR1561459_barcode
   d_linkedPrimer.fna,SRR1561460_barcoded_linkedPrimer.fna,SRR15
   61461_barcoded_linkedPrimer.fna,SRR1561462_barcoded_linkedPri
   mer.fna,SRR1561463_barcoded_linkedPrimer.fna,SRR1561464_barco
   ded_linkedPrimer.fna,SRR1561465_barcoded_linkedPrimer.fna,SRR
   1561466_barcoded_linkedPrimer.fna,SRR1561467_barcoded_linkedP
   rimer.fna,SRR1561468_barcoded_linkedPrimer.fna,SRR1561469_bar
   coded_linkedPrimer.fna,SRR1561470_barcoded_linkedPrimer.fna,S
   RR1561471_barcoded_linkedPrimer.fna,SRR1561472_barcoded_linke
   dPrimer.fna,SRR1561473_barcoded_linkedPrimer.fna,SRR1561474_b
   arcoded_linkedPrimer.fna,SRR1561475_barcoded_linkedPrimer.fna
   ,SRR1561476_barcoded_linkedPrimer.fna,SRR1561477_barcoded_lin
   kedPrimer.fna,SRR1561478_barcoded_linkedPrimer.fna,SRR1561479
   _barcoded_linkedPrimer.fna,SRR1561480_barcoded_linkedPrimer.f
   na,SRR1561481_barcoded_linkedPrimer.fna,SRR1561482_barcoded_l
   inkedPrimer.fna,SRR1561483_barcoded_linkedPrimer.fna,SRR15614
   84_barcoded_linkedPrimer.fna,SRR1561485_barcoded_linkedPrimer
   .fna,SRR1561486_barcoded_linkedPrimer.fna,SRR1561487_barcoded
   _linkedPrimer.fna,SRR1561488_barcoded_linkedPrimer.fna,SRR156
   1489_barcoded_linkedPrimer.fna,SRR1561490_barcoded_linkedPrim
   er.fna,SRR1561491_barcoded_linkedPrimer.fna,SRR1561492_barcod
   ed_linkedPrimer.fna,SRR1561493_barcoded_linkedPrimer.fna,SRR1
   561494_barcoded_linkedPrimer.fna,SRR1561495_barcoded_linkedPr
   imer.fna,SRR1561496_barcoded_linkedPrimer.fna,SRR1561497_barc
   oded_linkedPrimer.fna,SRR1561498_barcoded_linkedPrimer.fna,SR
   R1561499_barcoded_linkedPrimer.fna,SRR1561500_barcoded_linked
   Primer.fna,SRR1561501_barcoded_linkedPrimer.fna,SRR1561502_ba
   rcoded_linkedPrimer.fna,SRR1561503_barcoded_linkedPrimer.fna,
   SRR1561504_barcoded_linkedPrimer.fna,SRR1561505_barcoded_link
   edPrimer.fna,SRR1561506_barcoded_linkedPrimer.fna,SRR1561507_
   barcoded_linkedPrimer.fna,SRR1561508_barcoded_linkedPrimer.fn
   a,SRR1561509_barcoded_linkedPrimer.fna,SRR1561510_barcoded_li
   nkedPrimer.fna,SRR1561511_barcoded_linkedPrimer.fna,SRR156151
```

```
2_barcoded_linkedPrimer.fna,SRR1561513_barcoded_linkedPrimer.
fna,SRR1561514_barcoded_linkedPrimer.fna,SRR1561515_barcoded_
linkedPrimer.fna,SRR1561516_barcoded_linkedPrimer.fna,SRR1561
517_barcoded_linkedPrimer.fna,SRR1561518_barcoded_linkedPrime
r.fna,SRR1561519_barcoded_linkedPrimer.fna,SRR1561520_barcode
d_linkedPrimer.fna,SRR1561521_barcoded_linkedPrimer.fna,SRR15
61522_barcoded_linkedPrimer.fna,SRR1561523_barcoded_linkedPri
mer.fna,SRR1561524_barcoded_linkedPrimer.fna,SRR1561525_barco
ded_linkedPrimer.fna,SRR1561526_barcoded_linkedPrimer.fna,SRR
1561527_barcoded_linkedPrimer.fna,SRR1561528_barcoded_linkedP
rimer.fna,SRR1561529_barcoded_linkedPrimer.fna,SRR1561530_bar
coded_linkedPrimer.fna,SRR1561531_barcoded_linkedPrimer.fna,S
RR1561532_barcoded_linkedPrimer.fna,SRR1561533_barcoded_linke
dPrimer.fna,SRR1561534_barcoded_linkedPrimer.fna,SRR1561535_b
arcoded_linkedPrimer.fna,SRR1561536_barcoded_linkedPrimer.fna
,SRR1561537_barcoded_linkedPrimer.fna,SRR1561538_barcoded_lin
kedPrimer.fna,SRR1561539_barcoded_linkedPrimer.fna,SRR1561540
_barcoded_linkedPrimer.fna,SRR1561541_barcoded_linkedPrimer.f
na,SRR1561542_barcoded_linkedPrimer.fna,SRR1561543_barcoded_l
inkedPrimer.fna,SRR1561544_barcoded_linkedPrimer.fna,SRR15615
45_barcoded_linkedPrimer.fna,SRR1561546_barcoded_linkedPrimer
.fna,SRR1561547_barcoded_linkedPrimer.fna,SRR1561548_barcoded
_linkedPrimer.fna,SRR1561549_barcoded_linkedPrimer.fna,SRR156
1550_barcoded_linkedPrimer.fna,SRR1561551_barcoded_linkedPrim
er.fna,SRR1561552_barcoded_linkedPrimer.fna,SRR1561553_barcod
ed_linkedPrimer.fna,SRR1561554_barcoded_linkedPrimer.fna -o
/scratch/tef504/BV/split_libraries_BV/
```

Appendix 7:

Compatible format of mapping files for QIIME analysis. Columns #SampleID, BarcodeSequence, LinkerPrimerSequence and Description are essential formats of the mapping file for QIIME analysis, whereas the Study column was an additional feature added, specific to our analysis.

```
#SampleID     BarcodeSequence  LinkerPrimerSequence  Study  Description
ERR1679399    CCGTTTACTCTA     ATGCTGCCTCCCGTAGGAGT  HIV    HIV
ERR1679400    CTGCGCCCAGGT     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017264
ERR1679401    GCACATATGATC     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017265
ERR1679402    CGGCGCTCAAAT     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017266
ERR1679403    TCTGTTCTCAAG     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017267
ERR1679404    GGAGTTATGTGA     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017268
ERR1679405    ACCCAGGGTCAT     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017269
ERR1679406    CCCGTAAGACGG     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017270
ERR1679407    ATGGAACATAGC     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017271
ERR1679408    GCTCACGCGTGT     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017272
ERR1679409    AATCAATGGTCG     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017273
ERR1679410    TTGCCTGCGATG     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017274
ERR1679411    CAAAAACAACCA     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017275
ERR1679412    GGAAACACGACG     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017276
ERR1679413    CACTCGGATGAG     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017277
ERR1679414    CTCAAGACCAAG     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017278
ERR1679415    AGATAAGCCTAG     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017279
ERR1679416    TGGTAGAGAATA     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017280
ERR1679417    ACCAAAGTTTAG     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017281
ERR1679418    ATCAACTTGTGG     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017282
ERR1679419    ACTGTCGCCGAT     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017283
ERR1679420    GAAACGAGTCGG     ATGCTGCCTCCCGTAGGAGT  HIV    ERP017284
```

Appendix 8:

Python script run in an ipython notebook shell in the Oracle VM Virtual Box, generating unique barcodes for each SRR sample file of a study.

```
from random import choice

random_string = String(2856)   #2856 because HIV has 238 samples ie that many barcodes needed with 12bases each so 238*12=2856

list1 = []

barcodes = []


n = 12

for i in range(0, len(random_string), n):

    list1.append(random_string[i:i+n])


for n,i in enumerate(list1):

    barcodes.append(i)



#Python script creating random barcodes. AND ADDING THEM TO FILES

import glob

import pandas as pd

import numpy as np

bardict = {}


table = pd.read_excel("mapping_tableSRP064295.xlsx")

samples = table["#SampleID"]

barcodes = table["BarcodeSequence"]


for i in range(len(samples)):

    bardict[samples[i][0:10]] = barcodes[i]

list1 = table["BarcodeSequence"].tolist()

len(list1)  !=  len(set(list1))  #if  True  means  that  barcodes  NOT UNIQUE

linkerprimesequence = "ATGCTGCCTCCCGTAGGAGT"
```

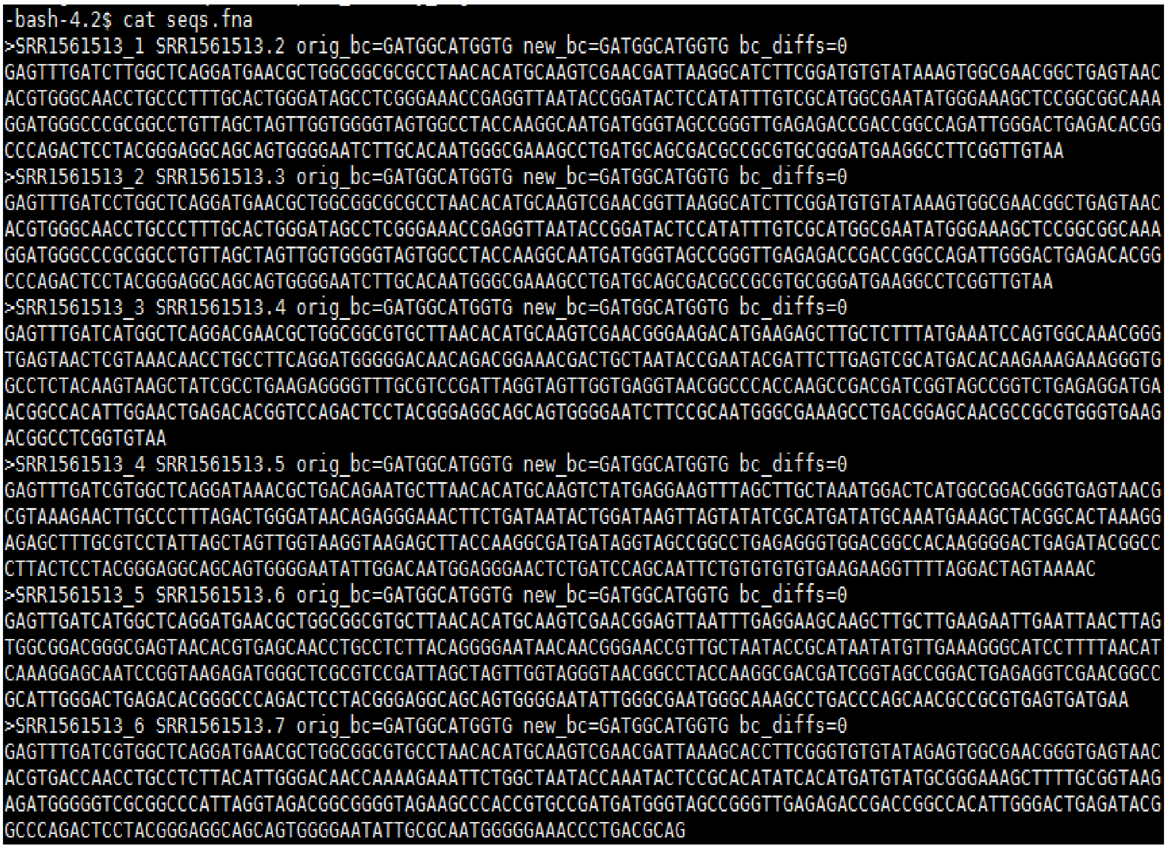
```
filepath =
("./fasta_files/fasta_files_PRJNA295859/SRR2533984_1.fna")

for filepath in glob.glob(filepath):

    file1 = open(filepath, "r")

#this section creates an empty output file for every input file

    output_filepath = filepath.split("/")

    output_filename    =    output_filepath[-1].split(".")[0]    +
"_barcoded_linkedPrimer." + output_filepath[-1].split(".")[1]

    output_filepath[-1] = output_filename

    output_filepath = "/".join(output_filepath)

    output_file = open(output_filepath, "w")

#actually writting in new file

    barcode = ""

    for n,line in enumerate(file1):

        newline = line

        if len(barcode) == 12:

#this is not the first thing that the computer reads. It doesn't
see any barcodes so the first thing it will do is find a line that
starts with ">" once this is done THEN it will define the barcodes,
however this needs to be written this way because it won't work
otherwise

            newline = barcode + linkerprimesequence + newline

            barcode = ""

        if line.startswith(">"):

            name = line[1:11]

            barcode = bardict[name]      #looks up barcode for the
name you gave in the dictionary

        output_file.write(newline)

    output_file.close()
```

Appendix 9:

pick_open_reference_otus.py Qiime command performing open reference OTU picking. Appendix 9a demonstrates the QIIME script running OTU picking for study SV through pipeline one. Appendix 9a was carried out in ipython notebook via the Oracle VM Virtual Box. –f and –i arguments instruct the input seq.fna file (a concatenated file of all 16S rRNA sequences of a study – shown in Appendix 9b). –r argument instructs the GreenGenes 2010 database, which was selected for the purpose of this study (gg_97_otus_6oct2010_aligned.fasta). –p argument instructs the parameters file, where the format is illustrated in Appendix 9c. Finally, the script provided in Appendix 9d, demonstrates QIIME's open reference OTU picking through the second designed pipeline, pipeline two, for study BV. This script utilises identically formatted seq.fna, mapping, parameter and database files as mentioned for pipeline one.

a)
```
!pick_open_reference_otus.py -f -i split_librariesSV/seqs.fna -r
current_Bacteria_aligned.fa -o otusSV/ -p params.txt --
suppress_align_and_tree
```

b) Format of seq.fna concatenated file:



```
-bash-4.2$ cat seqs.fna
>SRR1561513_1 SRR1561513.2 orig_bc=GATGGCATGGTG new_bc=GATGGCATGGTG bc_diffs=0
GAGTTTGATCTTGGCTCAGGATGAACGCTGGCGGCGCGCCTAACACATGCAAGTCGAACGATTAAGGCATCTTCGGATGTGTATAAAGTGGCGAACGGCTGAGTAAC
ACGTGGGCAACCTGCCCTTTGCACTGGGATAGCCTCGGGAAACCGAGGTTAATACCGGATACTCCATATTTGTCGCATGGCGAATATGGGAAAGCTCCGGCGGCAAA
GGATGGGCCCGCGGCCTGTTAGCTAGTTGGTGGGGTAGTGGCCTACCAAGGCAATGATGGGTAGCCGGGTTGAGAGACCGACCGGCCAGATTGGGACTGAGACACGG
CCCAGACTCCTACGGGAGGCAGCAGTGGGGAATCTTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGCGGGATGAAGGCCTTCGGTTGTAA
>SRR1561513_2 SRR1561513.3 orig_bc=GATGGCATGGTG new_bc=GATGGCATGGTG bc_diffs=0
GAGTTTGATCCTGGCTCAGGATGAACGCTGGCGGCGCGCCTAACACATGCAAGTCGAACGGTTAAGGCATCTTCGGATGTGTATAAAGTGGCGAACGGCTGAGTAAC
ACGTGGGCAACCTGCCCTTTGCACTGGGATAGCCTCGGGAAACCGAGGTTAATACCGGATACTCCATATTTGTCGCATGGCGAATATGGGAAAGCTCCGGCGGCAAA
GGATGGGCCCGCGGCCTGTTAGCTAGTTGGTGGGGTAGTGGCCTACCAAGGCAATGATGGGTAGCCGGGTTGAGAGACCGACCGGCCAGATTGGGACTGAGACACGG
CCCAGACTCCTACGGGAGGCAGCAGTGGGGAATCTTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGCGGGATGAAGGCCTCGGTTGTAA
>SRR1561513_3 SRR1561513.4 orig_bc=GATGGCATGGTG new_bc=GATGGCATGGTG bc_diffs=0
GAGTTTGATCATGGCTCAGGACGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGGGAAGACATGAAGAGCTTGCTCTTTATGAAATCCAGTGGCAAACGGG
TGAGTAACTCGTAAACAACCTGCCTTCAGGATGGGGGACAACAGACGGAAACGACTGCTAATACCGAATACGATTCTTGAGTCGCATGACACAAGAAAGAAAGGGTG
GCCTCTACAAGTAAGCTATCGCCTGAAGAGGGGTTTGCGTCCGATTAGGTAGTTGGTGAGGTAACGGCCCACCAAGCCGACGATCGGTAGCCGGTCTGAGAGGATGA
ACGGCCACATTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCAACGCCGCGTGGGTGAAG
ACGGCCTCGGTGTAA
>SRR1561513_4 SRR1561513.5 orig_bc=GATGGCATGGTG new_bc=GATGGCATGGTG bc_diffs=0
GAGTTTGATCGTGGCTCAGGATAAACGCTGACAGAATGCTTAACACATGCAAGTCTATGAGGAAGTTTAGCTTGCTAAATGGACTCATGGCGGACGGGTGAGTAACG
CGTAAAGAACTTGCCCTTTAGACTGGGATAACAGAGGGAAACTTCTGATAATACTGGATAAGTTAGTATATCGCATGATATGCAAATGAAAGCTACGGCACTAAAGG
AGAGCTTTGCGTCCTATTAGCTAGTTGGTAAGGTAAGAGCTTACCAAGGCGATGATAGGTAGCCGGCCTGAGAGGGTGGACGGCCACAAGGGGACTGAGATACGGCC
CTTACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGAGGGAACTCTGATCCAGCAATTCTGTGTGTGTGAAGAAGGTTTTAGGACTAGTAAAAC
>SRR1561513_5 SRR1561513.6 orig_bc=GATGGCATGGTG new_bc=GATGGCATGGTG bc_diffs=0
GAGTTGATCATGGCTCAGGATGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGGAGTTAATTTGAGGAAGCAAGCTTGCTTGAAGAATTGAATTAACTTAG
TGGCGGACGGGCGAGTAACACGTGAGCAACCTGCCTCTTACAGGGGAATAACAACGGGAACCGTTGCTAATACCGCATAATATGTTGAAAGGGCATCCTTTTAACAT
CAAAGGAGCAATCCGGTAAGAGATGGGCTCGCGTCCGATTAGCTAGTTGGTAGGGTAACGGCCTACCAAGGCGACGATCGGTAGCCGGACTGAGAGGTCGAACGGCC
GCATTGGGACTGAGACACGGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGGCGAATGGGCAAAGCCTGACCCAGCAACGCCGCGTGAGTGATGAA
>SRR1561513_6 SRR1561513.7 orig_bc=GATGGCATGGTG new_bc=GATGGCATGGTG bc_diffs=0
GAGTTTGATCGTGGCTCAGGATGAACGCTGGCGGCGTGCCTAACACATGCAAGTCGAACGATTAAAGCACCTTCGGGTGTGTATAGAGTGGCGAACGGGTGAGTAAC
ACGTGACCAACCTGCCTCTTACATTGGGACAACCAAAAGAAATTCTGGCTAATACCAAATACTCCGCACATATCACATGATGTATGCGGGAAAGCTTTTGCGGTAAG
AGATGGGGGTCGCGGCCCATTAGGTAGACGGCGGGGTAGAAGCCCACCGTGCCGATGATGGGTAGCCGGGTTGAGAGACCGACCGGCCACATTGGGACTGAGATACG
GCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCGCAATGGGGGAAACCCTGACGCAG
```

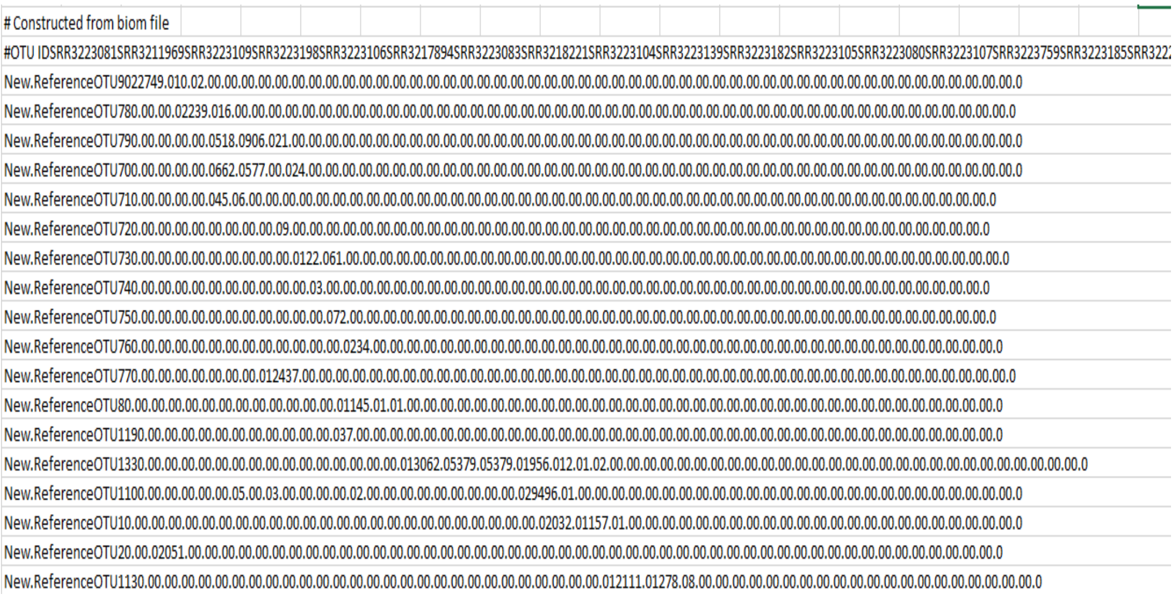
```
c) Params file:

        pick_otus:enable_rev_strand_match True
        assign_taxonomy:assignment_method blast
```

```
      pick_otus:similarity 0.97
      prefilter_identical_sequences:False
```

```
d) python
/usr/userfs/t/tef504/python/bin/pick_open_reference_otus.py -i
/scratch/tef504/BV/split_libraries_BV/seqs.fna -r
/scratch/tef504/qiime_scripts/gg_97_otus_6oct2010_aligned.fasta -o
/scratch/tef504/BV/OTUs_BV_export -p /scratch/tef504/BV/params.txt
```

Appendix 10:

.biom output file generated though pick_open_reference_otus.py QIIME script. The original .biom file consisted of a list of the assigned OTUs along with the corresponding abundance data for each sample within one study. Appendix 10 illustrates a segment of the .biom output file generated for study SRO071202.



```
# Constructed from biom file
#OTU IDSRR3223081SRR3211969SRR3223109SRR3223198SRR3223106SRR3217894SRR3223083SRR3218221SRR3223104SRR3223139SRR3223182SRR3223105SRR3223080SRR3223107SRR3223759SRR3223185SRR322
New.ReferenceOTU9022749.010.02.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.0
New.ReferenceOTU780.00.00.02239.016.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.0
New.ReferenceOTU790.00.00.00.00.0518.0906.021.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.0
New.ReferenceOTU700.00.00.00.00.0662.0577.00.024.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.0
New.ReferenceOTU710.00.00.00.00.045.06.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.0
New.ReferenceOTU720.00.00.00.00.00.00.00.00.09.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.0
New.ReferenceOTU730.00.00.00.00.00.00.00.00.00.0122.061.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.0
New.ReferenceOTU740.00.00.00.00.00.00.00.00.00.00.03.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.0
New.ReferenceOTU750.00.00.00.00.00.00.00.00.00.00.00.072.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.0
New.ReferenceOTU760.00.00.00.00.00.00.00.00.00.00.00.00.0234.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.0
New.ReferenceOTU770.00.00.00.00.00.00.00.012437.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.0
New.ReferenceOTU80.00.00.00.00.00.00.00.00.00.00.00.00.01145.01.01.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.0
New.ReferenceOTU1190.00.00.00.00.00.00.00.00.00.00.00.037.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.0
New.ReferenceOTU1330.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.013062.05379.05379.01956.012.01.02.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.0
New.ReferenceOTU1100.00.00.00.00.00.05.00.03.00.00.00.00.02.00.00.00.00.00.00.00.00.00.029496.01.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.0
New.ReferenceOTU10.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.02032.01157.01.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.0
New.ReferenceOTU20.00.02051.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.0
New.ReferenceOTU1130.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.012111.01278.08.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.00.0
```

Appendix 11:

rep_set_tax_assignment.txt output file generated for all selected studies, through pick_open_reference_otus.py QIIME script. The table below illustrates a section of the file generated for study HIV, which includes the assigned OTUs with their corresponding taxonomies. Additionally, the table includes the quality scores of the blasting identifiers, as well as a column of the confidence values for the deepest level of taxonomy shown. The file allows detection of the over assignment of unique OTUs performed through QIIME, as multiple unique OTUs characterise the same taxonomies.

New.ReferenceOTU2056	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella;s__	2e-130	2217
New.ReferenceOTU2724	k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Bifidobacteriales;f__Bifidobacteriaceae;g__Gardnerella;s__Gardnerella vaginalis	5e-128	531390
New.ReferenceOTU2054	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__;s__	5e-128	130750
New.ReferenceOTU2055	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Dialister;s__Dialister micraerophilus	8e-124	136415
New.ReferenceOTU2720	No blast hit	None	None
New.ReferenceOTU2053	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Shuttleworthia;s__	5e-128	137258
New.ReferenceOTU916	No blast hit	None	None
New.ReferenceOTU2051	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella;s__Prevotella melaninogenica	8e-118	471122
New.ReferenceOTU918	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Shuttleworthia;s__	5e-119	137258
New.ReferenceOTU2196	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__	3e-123	224390
New.ReferenceOTU3018	k__Bacteria;p__Fusobacteria;c__Fusobacteria (class);o__Fusobacteriales;f__Fusobacteriaceae;g__Sneathia;s__	9e-130	11298
New.ReferenceOTU3647	k__Bacteria;p__Actinobacteria;c__Actinobacteria (class);o__Bifidobacteriales;f__Bifidobacteriaceae;g__Gardnerella;s__Gardnerella vaginalis	8e-124	531390
New.ReferenceOTU2193	No blast hit	None	None
New.ReferenceOTU2192	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__;s__	1e-125	130750
New.ReferenceOTU2058	No blast hit	None	None
New.ReferenceOTU2059	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__;s__	1e-128	130750
New.ReferenceOTU887	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__	5e-54	114115
New.ReferenceOTU2727	No blast hit	None	None
New.ReferenceOTU3754	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__;s__	5e-128	130750

Appendix 12:

QIIME core_diversity_analyses.py command for diversity analysis was implemented in pipeline one as well as in the optimised pipeline two. Both scripts are similar, although formatted differently depending on the operating system they run on. –i argument instructs the input file, followed by –m listing the mapping file (see Appendix 7 for format) and finally the –e argument instructs a threshold of poor quality data (singletons etc.). Pipeline one run core_diversity_analyses.py script in a ipython notebook shell, in the format depicted in Appendix 12a. The script was a test run, performed for a study, which did not meet our selection requirements and thus was removed from the final analysis.

The core_diversity_analyses.py QIIME script used for diversity analysis via pipeline two is presented in Appendix 12b which illustrates the script for study BV. The script run through the YARCC computer cluster in a Linux shell.

a)

```
!core_diversity_analyses.py -i
swarm_otusSRR1823471/otu_table.biom -o cdoutSRR1823471.2/ -m
validate_mapping_file_outputSRR1823471.2/mapping_tableSRR1823471_c
orrected.txt -e 1 --nonphylogenetic_diversity
```

b)

```
python
/usr/userfs/t/tef504/python/bin/core_diversity_analyses.py -i
/scratch/tef504/BV/OTUs_BV_export/otu_table_mc2_w_tax.biom -o
/scratch/tef504/BV/Core_Div_BV/ -m
/scratch/tef504/BV/mapping_tableBV_corrected.txt -e 7000 --
nonphylogenetic_diversity
```

Appendix 13:

The table below represents the modified version of the .biom file, created though the "Modifying BIOM file script" discussed in Chapter section 2.3. The table includes unique taxonomies, which were fully characterised at genus level, instead of the numerous duplicated OTU assessments in the initial format presented in Appendix 11.

	SRR3223081	SRR3211969	SRR3223109	SRR3223198	SRR3223106	SRR3217894	SRR3223083	SRR3218221	SRR3223104
k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales f__Lactobacillaceae g__Lactobacillus	22799	8888	4	2324	22348	203	175	394	10
k__Bacteria p__Actinobacteria c__Actinobacteria o__Bifidobacteriales f__Bifidobacteriaceae g__Gardnerella	3949	4	10118	7339	5259	12853	6340	4376	12636
k__Bacteria p__Actinobacteria c__Coriobacteriia o__Coriobacteriales f__Coriobacteriaceae g__Atopobium	10	3	10041	5	219	3269	255	2080	4112
k__Bacteria p__Bacteroidetes c__Bacteroidia o__Bacteroidales f__Prevotellaceae g__Prevotella	34	1673	2683	12	0	3293	202	1729	1252
k__Bacteria p__Firmicutes c__Clostridia o__Clostridiales f__Clostridiaceae g__Clostridium	5	2	7	0	0	3354	5693	721	1077
k__Bacteria p__Firmicutes c__Clostridia o__Clostridiales f__Lachnospiraceae g__Shuttleworthia	2	1	0	0	0	0	152	0	2713
k__Bacteria p__Actinobacteria c__Actinobacteria o__Bifidobacteriales f__Bifidobacteriaceae g__Bifidobacterium	0	408	0	0	0	18	22	49	0
k__Bacteria p__Firmicutes c__Clostridia o__Clostridiales f__Veillonellaceae g__Megasphaera	2	4	0	0	0	693	766	40	1114

Appendix 14:

Python script programmed to run Pearson correlation analysis for all five selected studies utilising the modified .biom tables (see format in Appendix 13). The script run in an ipython notebook shell through Oracle VM Virtual Box.

```
#NOW MOVING TO CREATING PEARSON

new_biomfile = pd.read_excel("Final_biom_CANDIDIASIS.xlsx")

n = new_biomfile.shape[0]

#creates a numpy table full of zeros (x,z,y) which will then be
filled with the data bellow

output_data = np.zeros([n,n,2])

#.values changes Dataframe into numpy array

otutable_values = new_biomfile.values

for row1 in range(n):

    for row2 in range(row1,n):

        row = otutable_values[row1,1:]

        col = otutable_values[row2,1:]

        output_data[row1,row2,:] = scipy.stats.pearsonr(row, col)


np.save("PearsonTableCANDIDIASIS.npy", output_data)

#see numpy 2D table only with the pearson values

numpy3D = np.load("PearsonTableCANDIDIASIS.npy")

pearson2D_pears = pd.DataFrame(numpy3D[:,:,0])


#name columns and rows

b = output_sum_taxa.keys()

Pearson_data = pd.DataFrame(pearson2D_pears)

Pearson_data.columns = b

Pearson_data.index = b


#write pearson 2D file to excel

writer = pd.ExcelWriter("Pearsonvalues_table.xlsx")
```

```
Pearson_data.to_excel(writer, sheet_name="Sheet1")

writer.save()

#see numpy 2D table only with the p-values

numpy3D = np.load("PearsonTableCANDIDIASIS.npy")

pvalue2D_pears = pd.DataFrame(numpy3D[:,:,1])

#name columns and rows

b = output_sum_taxa.keys()

p_data = pd.DataFrame(pvalue2D_pears)

p_data.columns = b

p_data.index = b

#write pearson 2D file to excel

writer = pd.ExcelWriter("p_values_table.xlsx")

p_data.to_excel(writer, sheet_name="Sheet1")

writer.save()
```

Appendix 15:

Python script programmed in an ipython notebook shell performing Shapiro-Wilk test, to test whether the bacterial abundances from the biom files of each study were normalised. All values returned less than 0.055, suggesting that the abundance data of all five selected studies were not normally distributed.

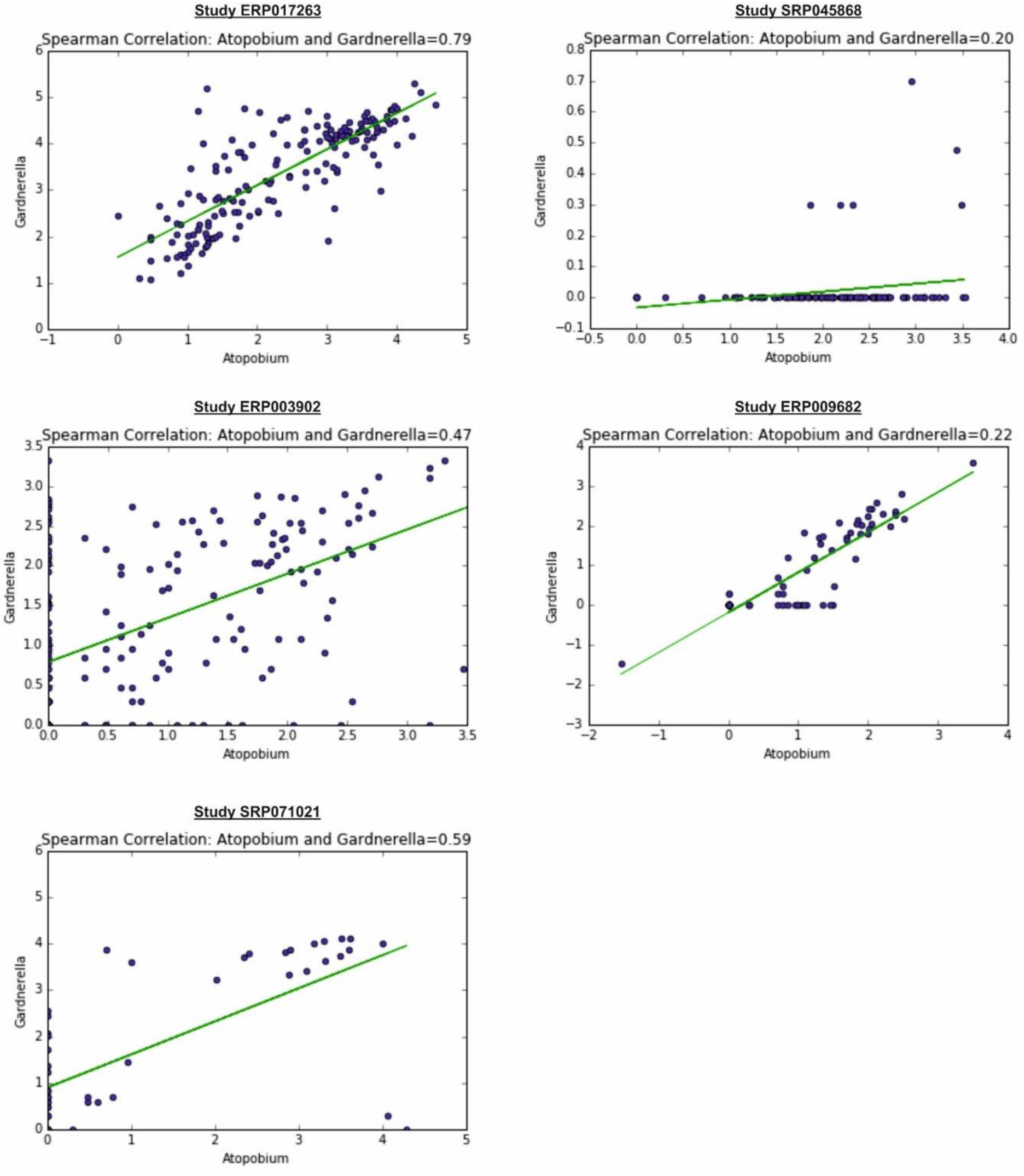
```
#Shapiro-Wilk test python

array = list(biom.values)

x2,p = scipy.stats.shapiro(array)

if(p < 0.055):                  #anything less than 0.055 is not normally distributed

    print "Not normal distribution," , "x2 =",x2, ", p_val =",p


a=biom.loc["k__Bacteriap__Firmicutesc__Bacillio__Lactobacillalesf__Lactobacillaceaeg__Lactobacillus"]

x2,p = scipy.stats.shapiro(a)

if(p < 0.055):

    print "Not normal distribution," , "x2=",x2, ", p_val=",p


a_log = np.log10(a.values[a.values>0])


plt.hist(a, bins=20)

plt.xlim(10, 18000)

plt.ylim(0,20)

#NO MY DATA ARE NOT NORMALISED
```

Appendix 16:

Python script followed to carry out Bonferroni correction for all corresponding p-values calculated through Spearman Rank Correlation Coefficient statistical test. The correction lowers the threshold at which a p value is considered significant (original p = 0.05).

```
Bonferroni correction:

S_p_values = pd.read_excel("2Spearman_P_values.xlsx")

bonferroni_array = S_p_values.values/1443 #38 taxa * 38 taxa -
1=1443

Bonferroni_data = pd.DataFrame(bonferroni_array,
index=new_biom.index, columns=new_biom.index)


writer = pd.ExcelWriter("2Sp_Bonferroni_P_values.xlsx")

Bonferroni_data.to_excel(writer, sheet_name="Sheet1")

writer.save()
```

Appendix 17:



Study ERP017263

Spearman Correlation: Atopobium and Gardnerella=0.79

Study SRP045868

Spearman Correlation: Atopobium and Gardnerella=0.20

Study ERP003902

Spearman Correlation: Atopobium and Gardnerella=0.47

Study ERP009682

Spearman Correlation: Atopobium and Gardnerella=0.22

Study SRP071021

Spearman Correlation: Atopobium and Gardnerella=0.59

Figure 28: Scatter plots of *Atopobium* and *Gardnerella* correlations for all five selected studies. Scatter plots for all five studies (HIV, BV, CANDIDIASIS, SV, HSV2) demonstrating correlations between *Atopobium* and *Gardnerella.* The blue data points represent the relative logged abundance data for *Atopobium* (x-axis) against the relative logged abundances of *Gardnerella* (y-axis). The green best fit line illustrates the correlation link between the two variables (*Atopobium* and *Gardnerella*). Each subplot contains the Spearman Rank Correlation Coefficient value for each study. All subplots suggest positive correlations (with 95% confidence) between *Atopobium* and *Gardnerella* with varying Spearman correlation values and steepness lines (with some studies illustrating higher and stronger correlations than others).

Appendix 18:

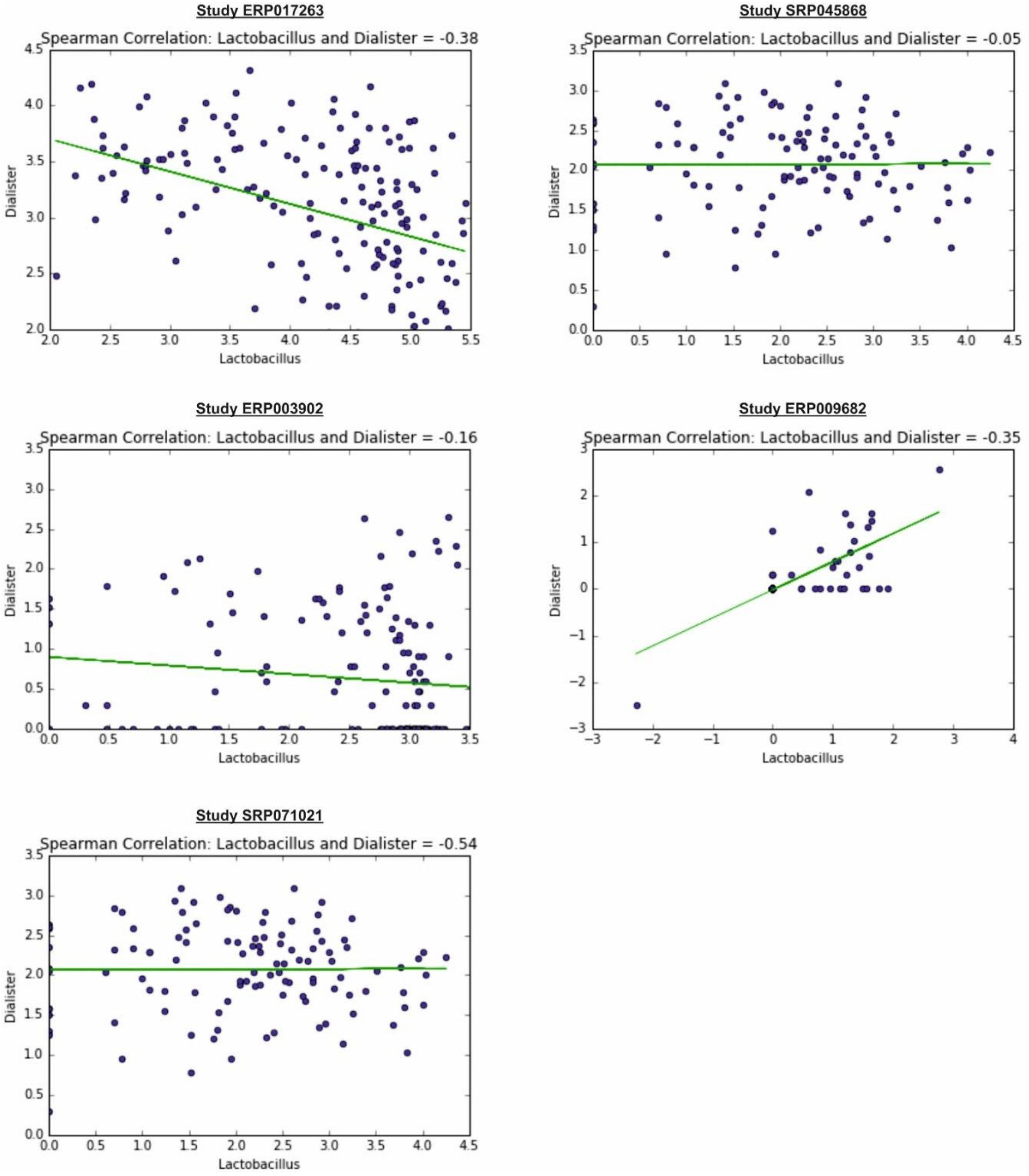
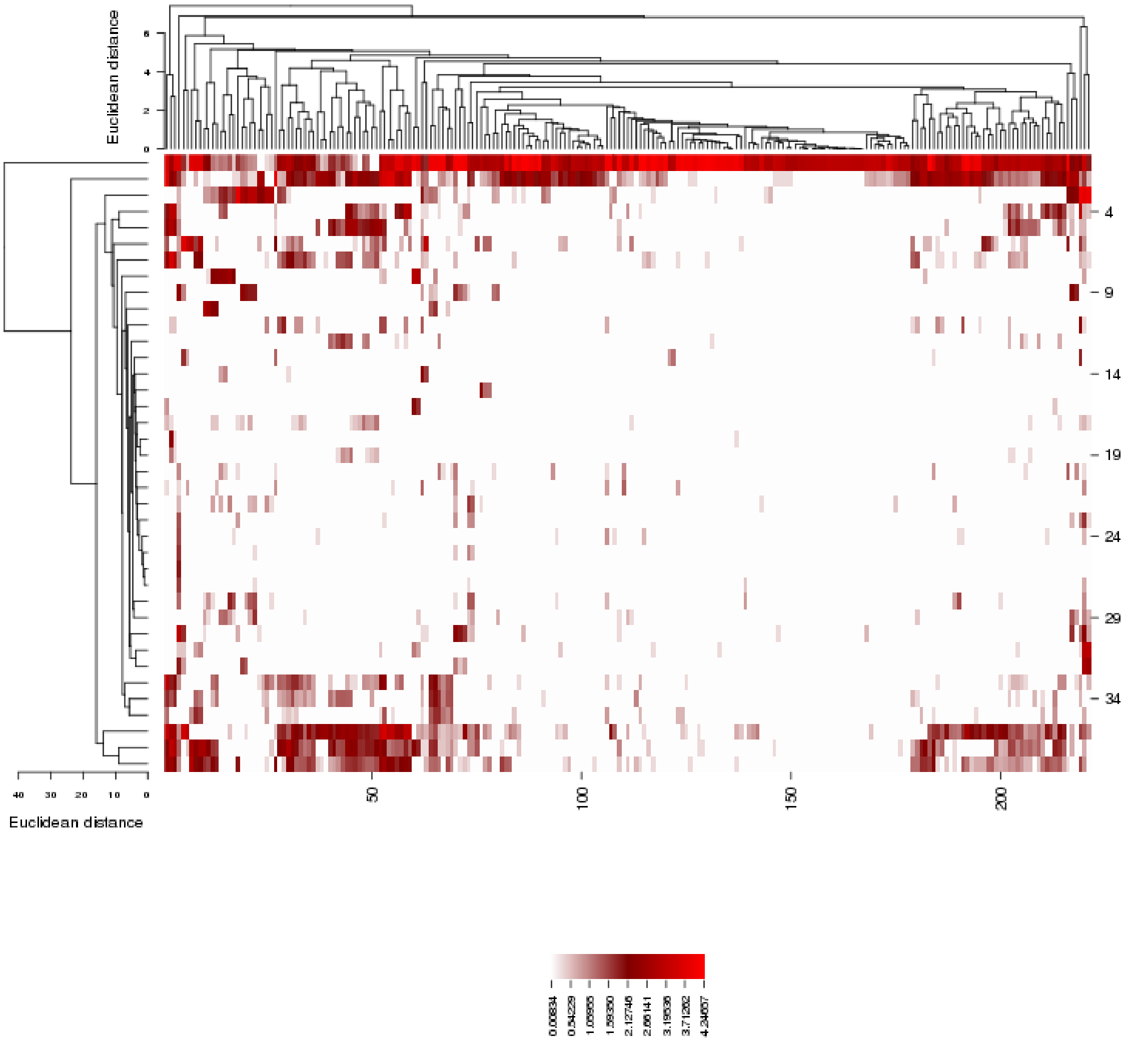


Figure 29: Scatter plots of *Lactobacillus* and *Dialister* correlations for all five selected studies. Scatter plots for all five studies (HIV, BV, CANDIDIASIS, SV, HSV2) demonstrating correlations between *Lactobacillus and Dialister*. Each subplot consists of blue data points representing the relative logged abundances for *Lactobacillus* (x-axis) against the relative logged abundances of *Dialister* (y-axis). The Spearman Rank Correlation Coefficient value for each study is shown. The green best fit line illustrates the correlation link between the two variables (*Lactobacillus* and *Dialister*). All subplots suggest low negative correlations between *Lactobacillus* and *Dialister* with varying negative correlation values and steepness lines. None of the correlations presented in Figure 29 meet the 95% confidence correlation threshold that ensures that the correlation is not a result of random chance.

Appendix 19:

Hierarchical clustering illustrated via a heatmap performed for study CANDIDIASIS through CIMminer open source tool. CIMminer applied Euclidean distance method and average linkage clustering algorithms for the analysis. The heatmap represents clustering analysis for both bacteria and sample data of study CANDIDIASIS presented via dendrograms on the top x-axis and the left y-axis. CIMminer is an interactive software allowing the corresponding sample and bacteria IDs to be visualised by placing the cursor over a data point. The data matrix of the logged bacteria taxonomy abundances was displayed as colour scales. Dark red represents the high abundance whereas white represents zero abundance. Hierarchical clustering via CIMminer did not propose any distinct clusters between sample or bacteria data thus our own python script was programmed to perform Hierarchical clustering as discussed in chapter 2.5.

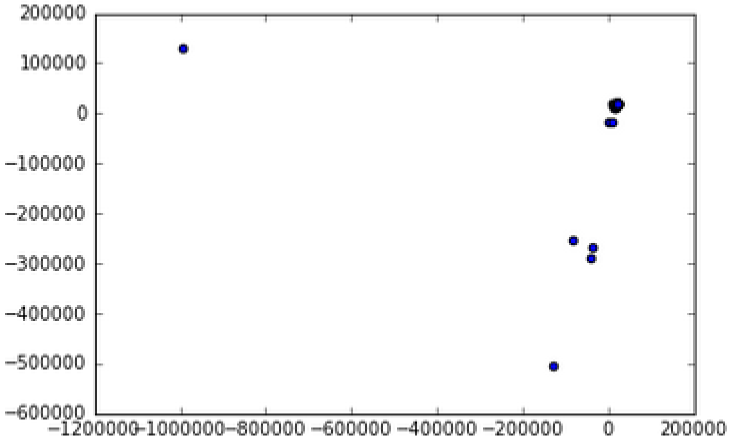


Appendix 20:

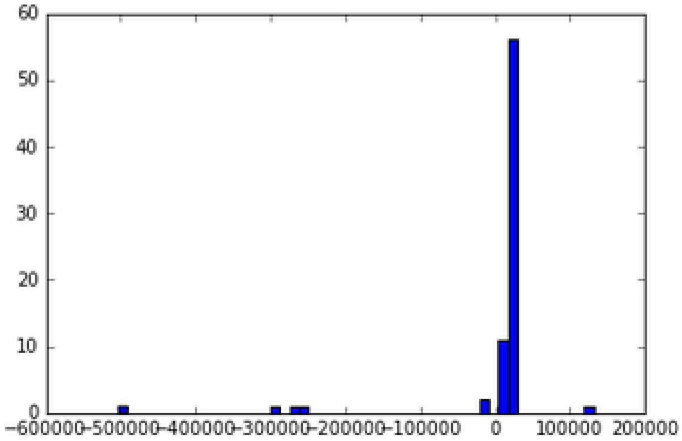
Python script programed in an ipython shell to allow two dimensional and three dimensional PCA analysis, as well as to permit focus on specific principal components.

PYTHON PCA:

```
import pandas as pd

import numpy as np

from sklearn.decomposition import PCA

from matplotlib.mlab import PCA as mpl_PCA

import matplotlib.pyplot as plt

%matplotlib inline


file1 = pd.read_excel("High_Abund_HIV.xlsx")

data = file1.values


#2 Principle Components - Bacteria

def doPCA(data):

    pca = PCA(n_components=2)

    pca.fit(data)

    return pca

pca = doPCA(data)

print pca.explained_variance_ratio_

first_pc = pca.components_[0]

second_pc = pca.components_[1]

#third_pc = pca.components_[2]

#print pca.components_

>> [ 0.60096727  0.29539025]


plt.scatter(transformed_data[:,0], transformed_data[:,1])
#principle component one and two grafting
```

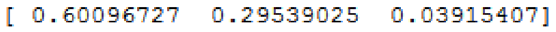


```
plt.hist(transformed_data[:,1], bins=50)

plt.show()
```

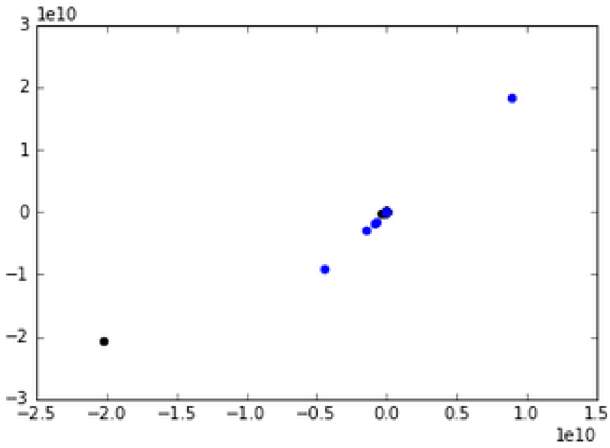


```
#3 Principle Components

#3 PCs

def doPCA(data):

    pca = PCA(n_components=3)

    pca.fit(data)

    return pca


pca = doPCA(data)

print pca.explained_variance_ratio_

first_pc = pca.components_[0]

second_pc = pca.components_[1]

third_pc = pca.components_[2]
```

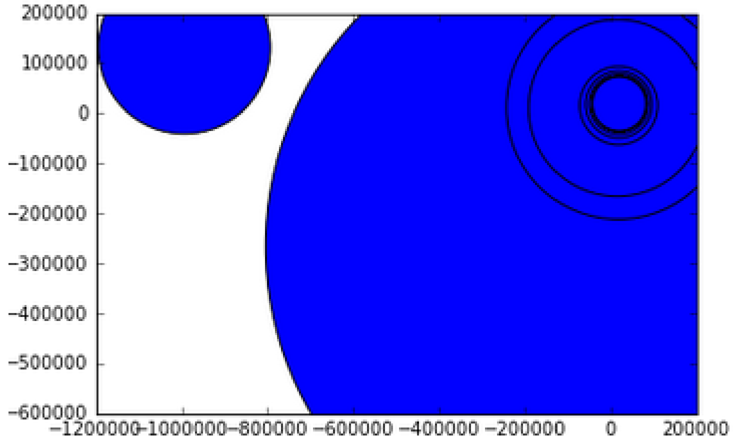
```
#print pca.components_

transformed_data = pca.transform(data)

for i, j in zip(transformed_data, data):

    plt.scatter(first_pc[0]*i[0]*i[0], first_pc[1]*i[0]*i[0],
color="k")

    plt.scatter(second_pc[0]*i[1]*i[0], second_pc[1]*i[1]*i[0],
color="b")

    #plt.scatter(third_pc[0]*i[0]*i[1], third_pc[1]*i[1]*i[1],
color="c")

    #plt.plot(j, "ro")

    #print data.shape

    #plt.xlim(-0.25,0.05)

    #plt.ylim(-0.2,0.2)
```



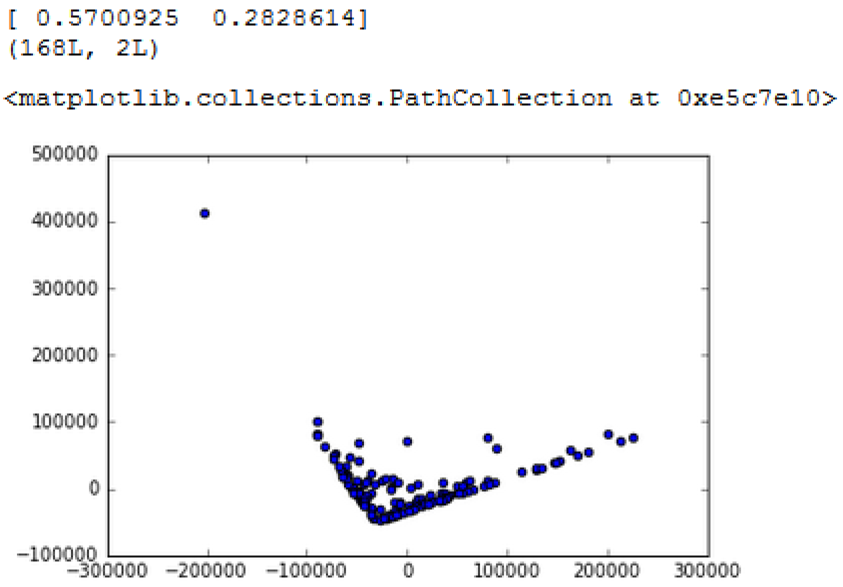
```
[ 0.60096727  0.29539025  0.03915407]
```



```
plt.scatter(transformed_data[:,0], transformed_data[:,1],
transformed_data[:,2])
```

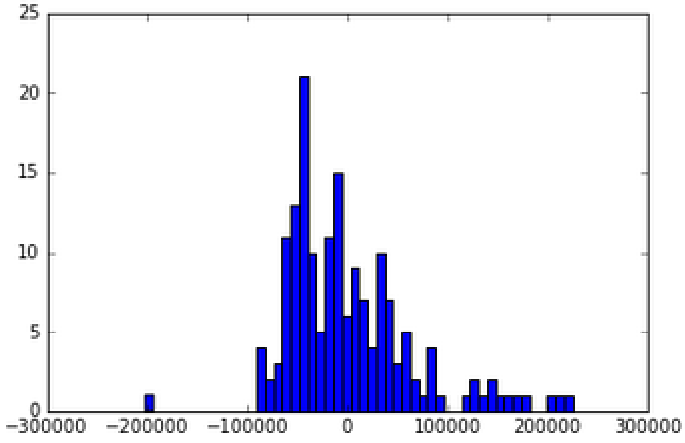


```
# 2 Principal Component - Samples

file1 = pd.read_excel("High_Abund_HIV.xlsx")

data = file1.transpose()

def doPCA(data):

    pca = PCA(n_components=2)

    pca.fit(data)

    return pca

pca = doPCA(data)

print pca.explained_variance_ratio_

first_pc = pca.components_[0]

second_pc = pca.components_[1]

#third_pc = pca.components_[2]

#print pca.components_

transformed_data = pca.transform(data)

print transformed_data.shape

plt.scatter(transformed_data[:,0], transformed_data[:,1])
#principal component one and two grafting
```



```
[ 0.5700925  0.2828614]
(168L, 2L)

<matplotlib.collections.PathCollection at 0xe5c7e10>
```

```
plt.hist(transformed_data[:,0], bins=50)

plt.show()
```



Appendix 21: List of Electronic files

Spearman rank correlation coefficient value tables:

1. Spearman_values_HSV2.xlsx
2. Spearman_values_BV.xlsx
3. Spearman_values_ERP017021.xlsx
4. Spearman_values_SV.xlsx
5. Spearman_values_CANDIDIASIS.xlsx

Modified OTU tables with taxonomies:

1. OTU_table_HSV2.xlsx
2. OTU_table_ BV.xlsx
3. OTU_table_ ERP017021.xlsx
4. OTU_table_ SV.xlsx
5. OTU_table_ CANDIDIASIS.xlsx

Figures:

Study ERP017021:

1. taxa-genus-HIV_legend.pdf
2. taxa-genus-HIV.pdf
3. taxa-family-HIV_legend.pdf
4. taxa-family-HIV.pdf
5. PCA-qiime-HIV.pdf
6. heatmap_HIV.png

Study BV:

1. taxa-order-BV_legend.pdf
2. taxa-order-BV.pdf
3. taxa-genus-BV_legend.pdf
4. taxa-genus-BV.pdf
5. PCA-QIIME-BV.pdf
6. HEATMAP_BV.jpg

Study HSV2:

1. taxa-family-SRP07102_legend.pdf
2. taxa-family-SRP07102.pdf
3. taxa-genus-SRP07102_legend.pdf
4. taxa-genus-SRP07102.pdf
5. PCA-Qiime-HSV2.pdf
6. Heatmap_HSV2.png

Study SV:

1. taxa-family-SV_legend.pdf
2. taxa-family-SV.pdf
3. taxa-genus-SV_legend.pdf
4. taxa-genus-SV_legend.pdf
5. PCA-QIIME-SV.pdf

6. Heatmap_SV.png

Study CANDIDIASIS:

1. taxa-order-CANDIDIASIS_legend.pdf
2. taxa-order-CANDIDIASIS.pdf
3. taxa-genus-CANDIDIASIS_legend.pdf
4. taxa-genus-CANDIDIASIS.pdf
5. PCA-QIIME-CANDIDIASIS.pdf
6. Heatmap_CANDIDIASIS.png

BIBLIOGRAPHY

[1] L. Dethlefsen, M. McFall-Ngai, and D. A. Relman, “An ecological and evolutionary perspective on humang-microbe mutualism and disease,” *Nature*, vol. 449, no. 7164, pp. 811–818, 2007.

[2] R. A. White, S. J. Callister, R. J. Moore, E. S. Baker, and J. K. Jansson, “The past, present and future of microbiome analyses,” *Nat. Protoc.*, vol. 11, no. 11, pp. 2049–2053, 2016.

[3] R. Lamendella, N. VerBerkmoes, and J. K. Jansson, “‘Omics’ of the mammalian gut--new insights into function.,” *Curr. Opin. Biotechnol.*, vol. 23, no. 3, pp. 491–500, 2012.

[4] C. A. Lozupone, J. I. Stombaugh, J. I. Gordon, J. K. Jansson, and R. Knight, “Diversity, stability and resilience of the human gut microbiota,” *Nature*, vol. 489, no. 7415, pp. 220–230, 2012.

[5] J. Kuczynski, C. L. Lauber, W. A. Walters, L. W. Parfrey, J. C. Clemente, D. Gevers, and R. Knight, “Experimental and analytical tools for studying the human microbiome,” *Nat. Rev. Genet.*, vol. 13, no. 1, pp. 47–58, 2011.

[6] S. Cribby, M. Taylor, and G. Reid, “Vaginal Microbiota and the Use of Probiotics,” *Interdiscip. Perspect. Infect. Dis.*, vol. 2008, pp. 1–9, 2008.

[7] E. A. Grice and J. A. Segre, “The skin microbiome,” *Nat. Rev. Microbiol.*, vol. 9, no. 8, pp. 626–626, 2011.

[8] V. D’Argenio and F. Salvatore, “The role of the gut microbiome in the healthy adult status,” *Clin. Chim. Acta*, vol. 451, pp. 97–102, 2015.

[9] A. M. O’Hara and F. Shanahan, “The gut flora as a forgotten organ,” *EMBO Rep.*, vol. 7, no. 7, pp. 688–693, 2006.

[10] C. Dunne, “Adaptation of bacteria to the intestinal niche: probiotics and gut disorder.,” *Inflamm. Bowel Dis.*, vol. 7, no. 2, pp. 136–145, 2001.

[11] X. Perret, C. Staehelin, and W. J. Broughton, “Molecular Basis of Symbiotic Promiscuity,” *Microbiol. Mol. Biol. Rev.*, vol. 64, no. 1, pp. 180–201, 2000.

[12] L. V. Hooper, “Commensal Host-Bacterial Relationships in the Gut,” *Science*, vol. 292, no. 5519, pp. 1115–1118, 2001.

[13] A. S. Neish, “Prokaryotic Regulation of Epithelial Responses by Inhibition of Ikappa B-alpha Ubiquitination,” *Science*, vol. 289, no. 5484, pp. 1560–1563, 2000.

[14] J. Xu, “A Genomic View of the Human-Bacteroides thetaiotaomicron Symbiosis,” *Science*, vol. 299, no. 5615, pp. 2074–2076, 2003.

[15] V. Pybus and B. Onderdonk, “Evidence for a commensal, symbiotic relationship between Gardnerella vaginalis and Prevotella bivia involving ammonia: potential significance for bacterial vaginosis.,” *J. Infect. Dis.*, vol. 175, pp. 406–413, 1997.

[16] M. C. E. Catenazzi, H. Jones, I. Wallace, J. Clifton, J. P. J. Chong, M. A. Jackson, S. Macdonald, J. Edwards, and J. W. B. Moir, “A large genomic island allows Neisseria meningitidis to utilize propionic acid, with implications for colonization of the human

nasopharynx," *Mol. Microbiol.*, vol. 93, no. 2, pp. 346–355, 2014.

[17] W. A. Claes, A. Pühler, and J. Kalinowski, "Identification of two prpDBC gene clusters in Corynebacterium glutamicum and their involvement in propionate degradation via the 2-methylcitrate cycle," *J. Bacteriol.*, vol. 184, no. 10, pp. 2728–2739, 2002.

[18] W. M. McCormack, C. H. Hayes, B. Rosner, J. R. Evrard, V. A. Crockett, S. Alpert, and S. H. Zinner, "Vaginal colonization with Corynebacterium vaginale (Haemophilus vaginalis)," *J. Infect. Dis.*, vol. 136, no. 6, pp. 740–745, 1977.

[19] D. M. Chu, J. Ma, A. L. Prince, K. M. Antony, M. D. Seferovic, and K. M. Aagaard, "Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery," *Nat. Med.*, vol. 23, no. 3, p. 314–326\t13, 2017.

[20] E. B. Hollister, K. Riehle, R. A. Luna, E. M. Weidler, M. Rubio-Gonzales, T.-A. Mistretta, S. Raza, H. V. Doddapaneni, G. A. Metcalf, D. M. Muzny, R. A. Gibbs, J. F. Petrosino, R. J. Shulman, and J. Versalovic, "Structure and function of the healthy pre-adolescent pediatric gut microbiome," *Microbiome*, vol. 3, no. 1, p. 36, 2015.

[21] D. B. DiGiulio, B. J. Callahan, P. J. McMurdie, E. K. Costello, D. J. Lyell, A. Robaczewska, C. L. Sun, D. S. A. Goltsman, R. J. Wong, G. Shaw, D. K. Stevenson, S. P. Holmes, and D. A. Relman, "Temporal and spatial variation of the human microbiota during pregnancy," *Proc. Natl. Acad. Sci.*, vol. 112, no. 35, pp. 11060–11065, 2015.

[22] N. Ottman, H. Smidt, W. M. de Vos, and C. Belzer, "The function of our microbiota: who is out there and what do they do?," *Front Cell Infect Microbiol*, vol. 2, p. 104, 2012.

[23] B. B. Oakley, T. L. Fiedler, J. M. Marrazzo, and D. N. Fredricks, "Diversity of Human Vaginal Bacterial Communities and Associations with Clinically Defined Bacterial Vaginosis," *Appl. Environ. Microbiol.*, vol. 74, no. 15, pp. 4898–4909, 2008.

[24] V. Redondo-Lopez, R. L. Cook, and J. D. Sobel, "Emerging role of lactobacilli in the control and maintenance of the vaginal bacterial microflora," *Rev. Infect. Dis.*, vol. 12, no. 5, pp. 856–872, 1990.

[25] K. Rea, T. G. Dinan, and J. F. Cryan, "The microbiome: A key regulator of stress and neuroinflammation," *Neurobiology of Stress*, vol. 4. pp. 23–33, 2016.

[26] Y. Taur and E. G. Pamer, "The intestinal microbiota and susceptibility to infection in immunocompromised patients," *Curr Opin Infect Dis*, vol. 26, no. 4, pp. 332–337, 2013.

[27] C. Gosmann, M. N. Anahtar, S. A. Handley, M. Farcasanu, G. Abu-Ali, B. A. Bowman, N. Padavattan, C. Desai, L. Droit, A. Moodley, M. Dong, Y. Chen, N. Ismail, T. Ndung'u, M. S. Ghebremichael, D. R. Wesemann, C. Mitchell, K. L. Dong, C. Huttenhower, B. D. Walker, H. W. Virgin, and D. S. Kwon, "Lactobacillus-Deficient Cervicovaginal Bacterial Communities Are Associated with Increased HIV Acquisition in Young South African Women," *Immunity*, vol. 46, no. 1, pp. 29–37, 2017.

[28] J. Ravel, P. Gajer, Z. Abdo, G. M. Schneider, S. S. K. Koenig, S. L. McCulle, S. Karlebach, R. Gorle, J. Russell, C. O. Tacket, R. M. Brotman, C. C. Davis, K. Ault, L. Peralta, and L. J. Forney, "Vaginal microbiome of reproductive-age women," *Proc. Natl. Acad. Sci.*, vol. 108, no. Supplement_1, pp. 4680–4687, 2011.

[29] R. E. Ley, F. Backhed, P. Turnbaugh, C. A. Lozupone, R. D. Knight, and J. I. Gordon, "Obesity

alters gut microbial ecology," *Proc. Natl. Acad. Sci.*, vol. 102, no. 31, pp. 11070–11075, 2005.

[30] R. B. Sartor, "Microbial Influences in Inflammatory Bowel Diseases," *Gastroenterology*, vol. 134, no. 2, pp. 577–594, 2008.

[31] M. J. G. Segovia, "Thumb-sucking, nail-biting, and atopic sensitization, asthma, and hay fever," *Acta Pediatrica Espanola*, vol. 74, no. 8. p. 202, 2016.

[32] G. Falony, S. Vieira-Silva, and J. Raes, "Microbiology Meets Big Data: The Case of Gut Microbiota–Derived Trimethylamine," *Annu. Rev. Microbiol.*, vol. 69, no. 1, pp. 305–321, 2015.

[33] A. L. Meditz, M. K. Haas, J. M. Folkvord, K. Melander, R. Young, M. McCarter, S. Mawhinney, T. B. Campbell, Y. Lie, E. Coakley, D. N. Levy, and E. Connick, "HLA-DR+ CD38+ CD4+ T Lymphocytes Have Elevated CCR5 Expression and Produce the Majority of R5-Tropic HIV-1 RNA In Vivo.," *J. Virol.*, vol. 85, no. 19, pp. 10189–10200, 2011.

[34] K. C. Anukam, E. O. Osazuwa, I. Ahonkhai, and G. Reid, "16S rRNA gene sequence and phylogenetic tree of lactobacillus species from the vagina of healthy Nigerian women," *African J. Biotechnol.*, vol. 4, no. 11, pp. 1222–1227, 2005.

[35] V. Jespers, J. van de Wijgert, P. Cools, R. Verhelst, H. Verstraelen, S. Delany-Moretlwe, M. Mwaura, G. F. Ndayisaba, K. Mandaliya, J. Menten, L. Hardy, and T. Crucitti, "The significance of Lactobacillus crispatus and L. vaginalis for vaginal health and the negative effect of recent sex: a cross-sectional descriptive study across groups of African women," *BMC Infect. Dis.*, vol. 15, no. 1, p. 115, 2015.

[36] B. Ma, L. J. Forney, and J. Ravel, "Vaginal Microbiome: Rethinking Health and Disease," *Annu. Rev. Microbiol.*, vol. 66, no. 1, pp. 371–389, 2012.

[37] S. Boris, J. E. Suárez, F. Vázquez, and C. Barbés, "Adherence of human vaginal lactobacilli to vaginal epithelial cells and interaction with uropathogens," *Infect. Immun.*, vol. 66, no. 5, pp. 1985–1989, 1998.

[38] R. Martín and J. E. Suárez, "Biosynthesis and degradation of H2O2 by vaginal lactobacilli," *Appl. Environ. Microbiol.*, vol. 76, no. 2, pp. 400–405, 2010.

[39] R. F. Lamont, J. D. Sobel, R. A. Akins, S. S. Hassan, T. Chaiworapongsa, J. P. Kusanovic, and R. Romeroa, "The vaginal microbiome: New information about genital tract flora using molecular based techniques," *BJOG: An International Journal of Obstetrics and Gynaecology*, vol. 118, no. 5. pp. 533–549, 2011.

[40] J. Atashili, C. Poole, P. Ndumbe, A. Adimora, and J. Smith, "Bacterial vaginosis and HIV acquisition: a meta-analysis of published studies," *AIDS*, vol. 22, no. 12, pp. 1493–1501, 2008.

[41] M. C. Stöppler, "Vaginal Infections (Vaginitis)," *EmedicineHealth*. [Online]. Available: http://www.emedicinehealth.com/vaginal_infections/article_em.htm. [Accessed: 04-Jul-2017].

[42] R. Mayfield-Blake, "Common vaginal infections," *Bupa*. [Online]. Available: https://www.bupa.co.uk/health-information/directory/v/vaginal-infections. [Accessed: 04-Jul-2017].

[43] P. E. Hay, “Chapter 7: Bacterial Vaginosis as a Mixed Infection,” in *Polymicrobial Diseases*, Washington DC: ASM Press, 2002.

[44] Rachita, “Differences between gonorrhoea and yeast infection,” *DifferenceBetween.net*, 2013. [Online]. Available: http://www.differencebetween.net/science/health/disease-health/differences-between-gonorrhoea-and-yeast-infection/. [Accessed: 04-Jul-2017].

[45] S. M. Fingerhuth, S. Bonhoeffer, N. Low, and C. L. Althaus, “Antibiotic-Resistant Neisseria gonorrhoeae Spread Faster with More Treatment, Not More Sexual Partners,” *PLoS Pathog.*, vol. 12, no. 5, p. e1005611, 2016.

[46] J. D. Sobel, “What’s new in bacterial vaginosis and trichomoniasis?,” *Infectious Disease Clinics of North America*, vol. 19, no. 2 SPEC. ISS. pp. 387–406, 2005.

[47] D. N. Fredricks, T. L. Fiedler, and J. M. Marrazzo, “Molecular identification of bacteria associated with bacterial vaginosis.,” *N. Engl. J. Med.*, vol. 353, no. 18, pp. 1899–911, 2005.

[48] S. L. Hillier, “Diagnostic microbiology of bacterial vaginosis.,” *Am. J. Obstet. Gynecol.*, vol. 169, no. 2 Pt 2, pp. 455–459, 1993.

[49] “Bacterial vaginosis,” *NHS Choices*, 2016. [Online]. Available: http://www.nhs.uk/conditions/bacterialvaginosis/Pages/Introduction.aspx. [Accessed: 24-Nov-2016].

[50] H. L. Martin, B. A. Richardson, P. M. Nyange, L. Lavreys, S. L. Hillier, B. Chohan, K. Mandaliya, J. O. Ndinya-Achola, J. Bwayo, and J. Kreiss, “Vaginal lactobacilli, microbial flora, and risk of human immunodeficiency virus type 1 and sexually transmitted disease acquisition.,” *J. Infect. Dis.*, vol. 180, no. 6, pp. 1863–1868, 1999.

[51] D. Saxena, Y. Li, L. Yang, Z. Pei, M. Poles, W. R. Abrams, and D. Malamud, “Human microbiome and HIV/AIDS,” *Current HIV/AIDS Reports*, vol. 9, no. 1. pp. 44–51, 2012.

[52] P. E. Hay, R. F. Lamont, D. Taylor-Robinson, D. J. Morgan, C. Ison, and J. Pearson, “Abnormal bacterial colonisation of the genital tract and subsequent preterm delivery and late miscarriage,” *BMJ*, vol. 308, no. 6924, pp. 295–298, 1994.

[53] J. C. Carey, M. A. Klebanoff, J. C. Hauth, S. L. Hillier, E. A. Thom, J. M. Ernest, R. P. Heine, R. P. Nugent, M. L. Fischer, K. J. Leveno, R. Wapner, M. Varner, W. Trout, A. Moawad, B. M. Sibai, M. Miodovnik, M. Dombrowski, M. J. O’Sullivan, J. P. VanDorsten, O. Langer, and J. Roberts, “Metronidazole to Prevent Preterm Delivery in Pregnant Women with Asymptomatic Bacterial Vaginosis,” *N. Engl. J. Med.*, vol. 342, no. 8, pp. 534–540, 2000.

[54] N. Sewankambo, R. H. Gray, M. J. Wawer, L. Paxton, D. McNaim, F. Wabwire-Mangen, D. Serwadda, C. Li, N. Kiwanuka, S. L. Hillier, L. Rabe, C. A. Gaydos, T. C. Quinn, and J. Konde-Lule, “HIV-1 infection associated with abnormal vaginal flora morphology and bacterial vaginosis.,” *Lancet*, vol. 350, no. 9077, pp. 546–50, 1997.

[55] X. Zhou, C. J. Brown, Z. Abdo, C. C. Davis, M. A. Hansmann, P. Joyce, J. A. Foster, and L. J. Forney, “Differences in the composition of vaginal microbial communities found in healthy Caucasian and black women,” *ISME J.*, vol. 1, no. 2, pp. 121–133, 2007.

[56] G. T. Spear, D. Gilbert, A. L. Landay, R. Zariffard, A. L. French, P. Patel, and P. M. Gillevet, “Pyrosequencing of the genital microbiotas of HIV-seropositive and -seronegative women reveals lactobacillus iners as the predominant lactobacillus species,” *Appl. Environ. Microbiol.*, vol. 77, no. 1, pp. 378–381, 2011.

[57] R. Hummelen, A. D. Fernandes, J. M. Macklaim, R. J. Dickson, J. Changalucha, G. B. Gloor, and G. Reid, "Deep sequencing of the vaginal microbiota of women with HIV," *PLoS One*, vol. 5, no. 8, p. e12078, 2010.

[58] H. Jousimies-Somer, "Recently described clinically important anaerobic bacteria: taxonomic aspects and update.," *Clin. Infect. Dis.*, vol. 25 Suppl 2, pp. S78-87, 1997.

[59] C. P. Tamboli, C. Neut, P. Desreumaux, and J. F. Colombel, "Dysbiosis in inflammatory bowel disease.," *Gut*, vol. 53, no. 1, pp. 1–4, 2004.

[60] C. Chang and H. Lin, "Dysbiosis in gastrointestinal disorders," *Best Pract. Res. Clin. Gastroenterol.*, vol. 30, no. 1, pp. 3–15, 2016.

[61] N. H. Salzman and C. L. Bevins, "Dysbiosis-A consequence of Paneth cell dysfunction," *Seminars in Immunology*, vol. 25, no. 5. pp. 334–341, 2013.

[62] L. F. Buttó and D. Haller, "Dysbiosis in intestinal inflammation: Cause or consequence," *International Journal of Medical Microbiology*, vol. 306, no. 5. pp. 302–309, 2016.

[63] X. C. Morgan and C. Huttenhower, "Chapter 12: Human Microbiome Analysis," *PLoS Comput. Biol.*, vol. 8, no. 12, p. e1002808, 2012.

[64] S. G. Tringe, C. von Mering, A. Kobayashi, A. A. Salamov, K. Chen, H. W. Chang, M. Podar, J. M. Short, E. J. Mathur, J. C. Detter, P. Bork, P. Hugenholtz, and E. M. Rubin, "Comparative metagenomics of microbial communities.," *Science*, vol. 308, no. 5721, pp. 554–7, 2005.

[65] P. J. Turnbaugh and J. I. Gordon, "An invitation to the marriage of metagenomics and metabolomics.," *Cell*, vol. 134, no. 5, pp. 708–13, Sep. 2008.

[66] M. J. Pallen, N. J. Loman, and C. W. Penn, "High-throughput sequencing and clinical microbiology: progress, opportunities and challenges.," *Curr. Opin. Microbiol.*, vol. 13, no. 5, pp. 625–31, 2010.

[67] J. Handelsman and J. Tiedje, *THE NEW SCIENCE OF METAGENOMICS: Revealing the Secrets of Our Microbial Planet*. Washington (DC): National Academies Press, 2007.

[68] J. Handelsman and W. Madison, "Metagenomics and Microbial Communities," *Life Sci.*, vol. 8, pp. 229–242, 2007.

[69] R. J. Hickey, X. Zhou, M. L. Settles, J. Erb, K. Malone, M. A. Hansmann, M. L. Shew, B. Van Der Pol, J. Dennis Fortenberry, and L. J. Forney, "Vaginal microbiota of adolescent girls prior to the onset of menarche resemble those of reproductive-age women," *MBio*, vol. 6, no. 2, pp. e00097-15, 2015.

[70] N. Segata, S. Haake, P. Mannon, K. P. Lemon, L. Waldron, D. Gevers, C. Huttenhower, and J. Izard, "Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples," *Genome Biol.*, vol. 13, no. 6, p. R42, 2012.

[71] C. Manichanh, L. Rigottier-Gois, E. Bonnaud, K. Gloux, E. Pelletier, L. Frangeul, R. Nalin, C. Jarrin, P. Chardon, P. Marteau, J. Roca, and J. Dore, "Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach.," *Gut*, vol. 55, no. 2, pp. 205–11, 2006.

[72] V. Kunin, A. Copeland, A. Lapidus, K. Mavromatis, and P. Hugenholtz, "A bioinformatician's

guide to metagenomics.," *Microbiol. Mol. Biol. Rev.*, vol. 72, no. 4, p. 557–78, Table of Contents, 2008.

[73] J. G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Peña, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, and R. Knight, "QIIME allows analysis of high-throughput community sequencing data.," *Nat. Methods*, vol. 7, no. 5, pp. 335–6, 2010.

[74] D. J. Lane, B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin, and N. R. Pace, "Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 82, no. 20, pp. 6955–6959, 1985.

[75] F. H. Karlsson, "Systems Biology of the Gut Microbiome in Metabolic Diseases," PhD dissertation, Dept. Chem. and Bio. Eng., Chalmers Uni. Of Tech., Gothenburg, Sweden, 2014.

[76] S. Chakravorty, D. Helb, M. Burday, N. Connell, and D. Alland, "A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria," *J. Microbiol. Methods*, vol. 69, no. 2, pp. 330–339, 2007.

[77] K. Becker, D. Harmsen, A. Mellmann, C. Meier, P. Schumann, G. Peters, and C. Von Eiff, "Development and evaluation of a quality-controlled ribosomal sequence database for 16S ribosomal DNA-based identification of Staphylococcus species," *J. Clin. Microbiol.*, vol. 42, no. 11, pp. 4988–4995, 2004.

[78] Y. Wang and P.-Y. Qian, "Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies.," *PLoS One*, vol. 4, no. 10, p. e7401, 2009.

[79] T. Z. DeSantis, P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen, "Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB.," *Appl. Environ. Microbiol.*, vol. 72, no. 7, pp. 5069–72, 2006.

[80] R. W. Hyman, M. Fukushima, L. Diamond, J. Kumm, L. C. Giudice, and R. W. Davis, "Microbes on the human vaginal epithelium," *Proc. Natl. Acad. Sci.*, vol. 102, no. 22, pp. 7952–7957, 2005.

[81] M. B. Liu, S. R. Xu, Y. He, G. H. Deng, H. F. Sheng, X. M. Huang, C. Y. Ouyang, and H. W. Zhou, "Diverse vaginal microbiomes in reproductive-age women with vulvovaginal CANDIDIASIS," *PLoS One*, vol. 8, no. 11, p. e79812, 2013.

[82] R. Mändar, M. Punab, N. Borovkova, E. Lapp, R. Kiiker, P. Korrovits, A. Metspalu, K. Krjutškov, H. Nlvak, J. K. Preem, K. Oopkaup, A. Salumets, and J. Truu, "Complementary seminovaginal microbiome in couples," *Res. Microbiol.*, vol. 166, no. 5, pp. 440–447, 2015.

[83] C. A. Muzny, I. R. Sunesara, R. Kumar, L. A. Mena, M. E. Griswold, D. H. Martin, E. J. Lefkowitz, and J. R. Schwebke, "Characterization of the vaginal microbiota among sexual risk behavior groups of women with bacterial vaginosis," *PLoS One*, vol. 8, no. 11, p. e80254, 2013.

[84] B. Shannon, P. Gajer, T. Yi, B. Ma, M. Humphrys, J. Thomas-Pavanel, L. Chieza, P. Janakiram, M. Saunders, W. Tharao, S. Huibner, K. Shahabi, J. Ravel, and R. Kaul, "Distinct

effects of the cervico-vaginal microbiota and herpes simplex type 2 infection on female genital tract immunology," *J. Infect. Dis.*, vol. 215, no. 9, pp. 1366–1375, 2017.

[85] C. J. Robinson, B. J. M. Bohannan, and V. B. Young, "From structure to function: the ecology of host-associated microbial communities.," *Microbiol. Mol. Biol. Rev.*, vol. 74, no. 3, pp. 453–76, 2010.

[86] H. Verstraelen, R. Verhelst, G. Claeys, M. Temmerman, and M. Vaneechoutte, "Culture-independent analysis of vaginal microflora: The unrecognized association of Atopobium vaginae with bacterial vaginosis," *Am. J. Obstet. Gynecol.*, vol. 191, no. 4, pp. 1130–1132, 2004.

[87] M. Morotomi, F. Nagai, H. Sakon, and R. Tanaka, "Dialister succinatiphilus sp. nov. and Barnesiella intestinihominis sp. nov., isolated from human faeces," *Int. J. Syst. Evol. Microbiol.*, vol. 58, no. 12, pp. 2716–2720, 2008.

[88] S. Srinivasan, N. G. Hoffman, M. T. Morgan, F. A. Matsen, T. L. Fiedler, R. W. Hall, F. J. Ross, C. O. McCoy, R. Bumgarner, J. M. Marrazzo, and D. N. Fredricks, "Bacterial communities in women with bacterial vaginosis: High resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria," *PLoS One*, vol. 7, no. 6, p. e37818, 2012.

[89] R. Poretsky, L. M. Rodriguez-R, C. Luo, D. Tsementzi, and K. T. Konstantinidis, "Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics," *PLoS One*, vol. 9, no. 4, p. e93827, 2014.

[90] N. Shah, H. Tang, T. G. Doak, and Y. Ye, "Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics," *Pac Symp Biocomput*, pp. 165–176, 2011.

[91] L. K. Ursell, J. C. Clemente, J. R. Rideout, D. Gevers, J. G. Caporaso, and R. Knight, "The interpersonal and intrapersonal diversity of human-associated microbiota in key body sites," *Journal of Allergy and Clinical Immunology*, vol. 129, no. 5. pp. 1204–1208, 2012.

[92] J. Kuczynski, J. Stombaugh, W. A. Walters, A. González, J. G. Caporaso, and R. Knight, "Using QIIME to analyze 16s rRNA gene sequences from microbial communities," *Curr. Protoc. Microbiol.*, no. Chapter: Unit 10.7, 2011.

[93] J. B. Hughes, J. J. Hellmann, T. H. Ricketts, and B. J. M. Bohannan, "Counting the Uncountable: Statistical Approaches to Estimating Microbial Diversity," *Applied and Environmental Microbiology*, vol. 67, no. 10. pp. 4399–4406, 2001.

[94] J. M. Janda and S. L. Abbott, "16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls," *Journal of Clinical Microbiology*, vol. 45, no. 9. pp. 2761–2764, 2007.

[95] M. De Hoon, "Cluster 3.0 for Windows, Mac OS X, Linux, Unix," *Human Genome Center, University of Tokyo.* 2002.

[96] N. W. H. Mason, D. Mouillot, W. G. Lee, and J. B. Wilson, "Functional richness, functional evenness and functional divergence: The primary components of functional diversity," *Oikos*, vol. 111, no. 1, pp. 112–118, 2005.

[97] P. J. Turnbaugh, R. E. Ley, M. A. Mahowald, V. Magrini, E. R. Mardis, and J. I. Gordon, "An obesity-associated gut microbiome with increased capacity for energy harvest," *Nature*, vol. 444, no. 7122, pp. 1027–131, 2006.

[98] J. R. Bray and J. T. Curtis, “An ordination of the upland forest communities of southern Wisconsin,” *Ecol. Monogr.*, vol. 27, no. 4, pp. 325–349, 1957.

[99] T. M. Nelson, J. L. C. Borgogna, R. M. Brotman, J. Ravel, S. T. Walk, and C. J. Yeoman, “Vaginal biogenic amines: Biomarkers of bacterial vaginosis or precursors to vaginal dysbiosis?,” *Front. Physiol.*, vol. 6, p. 253, 2015.

[100] C. J. Yeoman, S. M. Thomas, M. E. B. Miller, A. V. Ulanov, M. Torralba, S. Lucas, M. Gillis, M. Cregger, A. Gomez, M. Ho, S. R. Leigh, R. Stumpf, D. J. Creedon, M. A. Smith, J. S. Weisbaum, K. E. Nelson, B. A. Wilson, and B. A. White, “A Multi-Omic Systems-Based Approach Reveals Metabolic Markers of Bacterial Vaginosis and Insight into the Disease,” *PLoS One*, vol. 8, no. 2, p. e56111, 2013.

[101] J. Downes, M. A. Munson, D. R. Radford, D. A. Spratt, and W. G. Wade, “*Shuttleworthia* satelles gen. nov., sp. nov., isolated from the human oral cavity,” *Int. J. Syst. Evol. Microbiol.*, vol. 52, no. 5, pp. 1469–1475, 2002.

[102] M. Mimee, R. J. Citorik, and T. K. Lu, “Microbiome therapeutics - Advances and challenges,” *Advanced Drug Delivery Reviews*, vol. 105. pp. 44–54, 2016.

[103] M. Fujiya, N. Ueno, and Y. Kohgo, “Probiotic treatments for induction and maintenance of remission in inflammatory bowel diseases: A meta-analysis of randomized controlled trials,” *Clinical Journal of Gastroenterology*, vol. 7, no. 1. pp. 1–13, 2014.

[104] B. Ma, L. J. Forney, and J. Ravel, “Vaginal microbiome: rethinking health and disease.,” *Annu. Rev. Microbiol.*, vol. 66, pp. 371–89, 2012.

[105] N. Takahashi and T. Yamada, “Glucose metabolism by Prevotella intermedia and Prevotella nigrescens,” *Oral Microbiol. Immunol.*, vol. 15, no. 3, pp. 188–195, 2000.

[106] M. Aldunate, D. Srbinovski, A. C. Hearps, C. F. Latham, P. A. Ramsland, R. Gugasyan, R. A. Cone, and G. Tachedjian, “Antimicrobial and immune modulatory effects of lactic acid and short chain fatty acids produced by vaginal microbiota associated with eubiosis and bacterial vaginosis,” *Front. Physiol.*, vol. 6, p. 164, 2015.

[107] L. V Hill, “Anaerobes and Gardnerella vaginalis in non-specific vaginitis.,” *Genitourin. Med.*, vol. 61, no. 2, pp. 114–9, 1985.

[108] R. N. Fichorova, H. S. Yamamoto, M. L. Delaney, A. B. Onderdonk, and G. F. Doncel, “Novel vaginal microflora colonization model providing new insight into microbicide mechanism of action,” *MBio*, vol. 2, no. 6, pp. e00168-11, 2011.

[109] R. Verhelst, H. Verstraelen, G. Claeys, G. Verschraegen, J. Delanghe, L. Van Simaey, C. De Ganck, M. Temmerman, and M. Vaneechoutte, “Cloning of 16S rRNA genes amplified from normal and disturbed vaginal microflora suggests a strong association between Atopobium vaginae, Gardnerella vaginalis and bacterial vaginosis.,” *BMC Microbiol.*, vol. 4, p. 16, 2004.

[110] W.-T. Liu, T. L. Marsh, H. Cheng, and L. J. Forney, “Characterization of Microbial Diversity by Determining Terminal Restriction Fragment Length Polymorphisms of Genes Encoding 16S rRNA,” *Appl. Environ. Microbiol.*, vol. 63, no. 11, pp. 4516–4522, 1997.

[111] J. C. Fuller, P. Khoueiry, H. Dinkel, K. Forslund, A. Stamatakis, J. Barry, A. Budd, T. G. Soldatos, K. Linssen, and A. M. Rajput, “Biggest challenges in bioinformatics.,” *EMBO Rep.*, vol. 14, no. 4, pp. 302–304, 2013.

[112] A. L. Boulesteix, “Over-optimism in bioinformatics research,” *Bioinformatics*, vol. 26, no. 3, pp. 437–439, 2009.

[113] W. P. Hanage, “Microbiome science needs a healthy dose of scepticism,” *Nature*, vol. 512, pp. 247–248, 2014.

[114] H. Teeling and F. O. Glöckner, “Current opportunities and challenges in microbial metagenome analysis-A bioinformatic perspective,” *Brief. Bioinform.*, vol. 13, no. 6, pp. 728–742, 2012.