Impact of the mycobiome on the health of the female genital tract

by

Tamlyn Kirstey Gangiah

(GNGTAM001)

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN

In fulfilment of the requirements for the degree

Msc(Med) in Bioinformatics

Faculty of Health Sciences

UNIVERSITY OF CAPE TOWN

Date of Submission: 31 May 2022

Supervisor: Prof. Nicola Mulder

Co-supervisor: Dr. Lindi Masson

DECLARATION

I, *Tamlyn Kirstey Gangiah*, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I have used the Genetics convention for citation and referencing. Each contribution to, and quotation in, this thesis from the work(s) of other people has been attributed, and has been cited and referenced.

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

Signature: Signed by candidate

Date: 31 May 2022

Acknowledgements

First and foremost, I would like to thank my remarkable supervisors, Prof. Nicola Mulder and Dr. Lindi Masson, for their support, guidance, and input throughout my Master's. It has been an absolute blessing to work under two incredibly brilliant scientists who have inspired me to forge my path as a woman in science.

To my parents, Reggie and Evelyn Gangiah, your unconditional love, support, and prayers are what keep me going. Thank you for encouraging me to reach great heights and own my full potential. I could never repay you for working hard to give me the best life possible, but I hope this is a start.

To my family, Privan, Sherwin, and Zara Gia Gangiah: thank you for believing in me.

To Michael Ross Scott, this wouldn't have been possible without you. Your willingness to spare your extensive coding knowledge and help me work through challenges with data analysis is endlessly appreciated.

To my dearest friends: Annabel Fenton, Christen Da Costa, Daven Mauff, Iyra Maharaj, John Irandukunda, Lavania Rajah, and Rebecca Herbert - thank you for always being there to listen, offer sound advice, celebrate my highs, and be a constant source of motivation through my lows. I am so grateful to have such an incredible support structure.

To a great mentor - Thys Potgieter – thank you for taking the time to help, advise, and guide me along with data analysis.

My Honours supervisors, A/Prof Jean-Baptiste Ramond and Dr. Surendra Vikram, thank you for instilling your love for research (and the microbiome) in me and inspiring me to be a better scientist. It was an honour to start my postgrad with you both, and I carry the lessons you have taught me until now.

Finally, I would like to thank my funders, the National Research Foundation (NRF), for affording me the opportunity to pursue my dream of contributing to human microbiome research.

Table of Contents

List of abbreviations

ANOVA	Analysis of Variance
ART	Antiretroviral Therapy
BMI	Body Mass Index
BP	Biological Process
BV	Bacterial Vaginosis
C	Carbamidomethyl
CC	Cellular Component
COG	clusters of orthologous groups
CT	*Chlamydia trachomatis*
DMPA	Depot Medroxyprogesterone Acetate
DNA	Deoxy Nucleic Acid
eggNOG	evolutionary genealogy of genes: non-supervised orthologous
FA	Formic Acid
FDR	False Discovery Rate
FGT	Female Genital Tract
FMT	Faecal Microbiota Transplantation
FRT	Female Reproductive Tract
GO	Gene Ontology
GRP	Glucose-regulated Protein
HC	Hormonal Contraception
HIV	Human Immunodeficiency Virus
HSP	Heat Shock Protein
HSV	Herpes Simplex Virus
HPV	Human Papillomavirus
iBAQ	intensity Based Absolute Quantitation
ITS	Internal Transcribed Spacer

KEGG	Kyoto Encyclopedia of Genes and Genomes
KNN	K-nearest neighbor
KO	KEGG Ontology
LCA	Lowest Common Ancestor
MHC	Major Histocompatibility Complex
MF	Molecular Function
MG	*Mycoplasma genitalium*
ML	Machine Learning
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NCBI	National Center for Biotechnology Information
NMDS	Non-metric Multidimensional Scaling
NG	*Neiressia gonorrhea*
NGS	Next-generation Sequencing
nr	Non-redundant
OCP	Oral Contraceptive Pill
PCA	Principal Component Analysis
PEP	Posterior Error Probability
PERMANOVA	Permutational Multivariate Analysis of Variance
pH	Potential of Hydrogen
PSA	Prostate-specific antigen
PSM	Peptide Spectrum Match
PTM	Post Translational Modification
mRNA	Messenger Ribonucleic Acid
RDA	Redundancy Analysis
rRNA	Ribonucleic Acid
RNA	Ribonucleic Acid
RVVC	Recurrent Vulvovaginal Candidiasis

SA	South Africa
SCFA	Short Chain Fatty Acid
STI	Sexually Transmitted Infection
TLRs	Toll-like Receptors
TV	Trichomonas Vaginalis
VMB	Vaginal Microbiome
VVC	Vulvovaginal Candidiasis
WHO	World Health Organization

Abstract

The female genital tract microbiome comprises a community of microorganisms. Imbalances in the microbiome are associated with vaginal diseases such as bacterial vaginosis (BV) and sexually transmitted infections (STIs). These diseases greatly burden South Africa, and young women in this region are at an increased risk of contracting vaginal diseases. Consequently, it is vital to investigate the factors that influence FGT health. The fungal constituent of the microbiome (the mycobiome) has been demonstrated to play a role in regulating mucosal health, especially when the bacterial component is disturbed. However, we have a limited understanding of the vaginal mycobiome since many microbiome studies have focused on bacterial communities and have neglected low abundance taxonomic groups, such as fungi. To reduce this knowledge deficit, we present the first large-scale metaproteomic study to define the taxonomic composition and potential functional processes of the vaginal mycobiome in South African women. We examined vaginal fungal communities present in optimal and non-optimal states (BV, STIs, and genital inflammation) by collecting lateral vaginal wall swabs from 123 women for liquid chromatography-tandem mass spectrometry. Taxonomic analysis requires representative sequence databases; however, since mycobiome research is relatively new, fungal databases are still in their infancy. As a result, metaproteomic methods are not optimized for fungal research. Therefore, we optimized a metaproteomic approach to increase fungal protein group assignments. With this, 50 fungal proteins belonging to 39 different genera were identified post quality-control and analysed for taxonomic and functional distributions. Taxonomic analysis revealed that the vaginal mycobiome had a high relative abundance of *Candida* across optimal and non-optimal states. We observed changes in differential abundance at the genus and biological process level between optimal and non-optimal states for BV and *Mycoplasma genitalium*. In the BV positive state, most fungal proteins were significantly underabundant ($p<0.05$) compared to the BV negative state, with the exception of *Malassezia* and *Condiobolus*. Correspondingly, Nugent score was negatively associated with total fungal protein intensity, implying that the microenvironment during BV is less suitable for fungal growth. Furthermore, we assessed which clinical variables were associated with driving fungal community composition; results indicated that Nugent score, pro-inflammatory cytokines, chemokines, vaginal pH, *Chlamydia trachomatis*, and the presence of clue cells were involved. Lastly, we used publicly available vaginal proteome data to confirm our fungal identifications and suspect *Candida*, *Debaryomyces*, *Kluyveromyces*, *Malassezia*, *Penicillium*, *Yarrowia*, *Aspergillus*, *Cryptococcus, Wallemia, Trichosporon,* and *Saccharomyces* are likely true vaginal inhabitants. Thus, this study sets the groundwork for understanding the vaginal mycobiome and its association with prevalent vaginal diseases.

Chapter 1

Introduction

The female genital tract (FGT) microbiome compromises a community of microorganisms that play a role in maintaining vaginal health. The vagina is more prone to various infectious diseases when there are disruptions to the microbiome composition (Jones *et al.* 2014; Schwebke 2003; Soper *et al.* 1990). Such infections include bacterial vaginosis (BV), sexually transmitted infections (STIs), and Candidiasis (Jones *et al.* 2014; Pitt *et al.* 2005; Schwebke 2003; Soper *et al.* 1990). This is an issue particularly in Africa as the prevalence of BV and STIs is the highest in the world.

In a publication by Torrone *et al.* (2018), which analysed population studies from South African community-based regions, the prevalence of BV ranged between 35.8% to 52.4% among women in the 15-24 year old group. The prevalence of bacterial (*Chlymadia trachomatis*, *Neisseria gonorrhoeae*, and Syphilis) and parasitic STIs (*Trichomonas vaginalis*), were found to be higher among adolescents and young adults (15-24 years old age group) in comparison to older adults (24-49-year-old age group) (Torrone *et al.* 2018; Weiss *et al.* 2001). One highly prevalent STI in South Africa (SA) is the human immunodeficiency virus (HIV), which affects approximately 720,000 individuals in the adolescent and young adult population (Zanoni *et al.* 2016). Since young women in this region are at a higher risk of vaginal diseases, it is a public health priority to reduce the risk of these prevalent diseases by improving sexual and reproductive health efforts (van de Wijgert *et al.* 2008).

Another vaginal disease greatly burdening SA includes vulvovaginal candidiasis (VVC), which is the second most reported type of infectious vaginitis and affects an estimated 1 million women each year in SA (Denning *et al.* 2018; Schwartz *et al.* 2019). VVC is a common side effect of antibiotics used to treat BV, indicating that the vaginal microbiome (VMB) might be related to the colonization of yeast (Pirotta *et al*. 2003). The yeast species, *Candida albicans* accounts for 85% to 90% of cases of VVC, and the remaining cases are usually a result of *Candida glabrata* (Sobel 2007). However, clinical management of VVC is difficult as there is an increasing emergence of antifungal resistant species, such as *C. glabarata*, which restricts

treatment options (Antonovics *et al*. 2007; Cowen 2008; Cowen *et al.* 2015; Perlroth *et al.* 2007).

One contributing factor to the high prevalence of the abovementioned diseases is a result of the frequent asymptomatic nature of these diseases, in which females do not present symptoms and do not seek treatment (McClelland *et al.* 2009; Donbraye-Emmanuel *et al.* 2010; Masson *et al*. 2019). When these diseases are left untreated, a range of health complications can occur, including preterm delivery, pelvic inflammatory disease, upper female reproductive tract infections, and an increased likelihood of contracting STIs, including HIV (Fethers *et al.* 2008; Masson *et al.* 2015; Meizoso *et al.* 2008; Mølgaard-Nielsen *et al.* 2016; Peters *et al.* 2014b; Ray *et al.* 2011; Romoren *et al.* 2007; Simon *et al.* 2006; Srinivasan *et al.* 2012; Taha *et al.* 1998).

The high prevalence of BV, STIs, and VVC sheds light on the issue of vaginal disease-causing infections and how it is a problem that burdens SA. As a result, it is vital to investigate the factors that influence FGT health and their effect on the microbiome (Mirmonsef *et al.* 2011). One understudied area is the human mycobiome (Iliev and Leonardi 2017). Of late, fungi have been shown to play a greater role than previously described in the regulation of mucosal health, especially when the bacterial microbiome is disturbed (Huffnagle and Noverr 2013). Since not much is known about the role of the mycobiome during vaginal diseases, such as BV. More mycobiome research will prove to be essential as the vaginal mycobiome may have more effects on health and clinical outcomes than are currently known.

Literature review

The female genital tract

The FGT is made up of two compartments, namely the lower and upper genital tract. The upper FGT consists of the fallopian tubes, ovaries, uterus, and endocervix (Fig 1.1), while the lower FGT consists of the ectocervix and vagina (Fig 1.1) (Burgener *et al.* 2013).

The surface of the FGT contains both anatomical and biological defenses to inhibit pathogens and their associated particles (Burgener *et al.* 2013). The first line of physical defense includes epithelial cells, which line the internal surface of the FGT, forming a barrier (Burgener *et al.* 2013). This epithelial layer differs between the upper and lower FGT. In the upper FGT, there is only a single columnar epithelial cell barrier (Fig 1.1), whereas the lower region is protected by multiple layers of stratified squamous epithelium (Fig 1.1; Burgener *et al.* 2013). These epithelial cells secrete mucus, which surrounds the vagina and cervix, and aids in trapping infectious agents (Franklin and Kutteh 1999; Iwasaki and Medzhitov 2010).

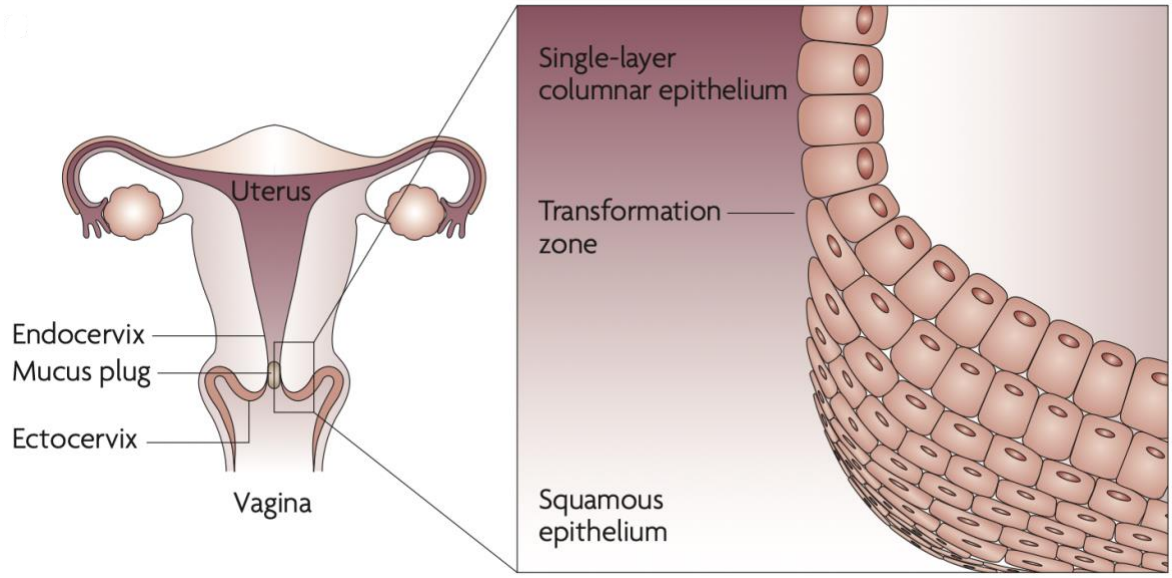


Figure 1.1. Diagram depicting the anatomy of the female genital tract (FGT) (Adapted from Hladik and McElrath 2008).

The first line of biological defense includes anti-microbial factors (*e.g.,* defensins), present in the mucus and immune cells (*e.g.,* macrophages and T cells), which populate epithelial cells (Orga *et al.* 1981; Shukair *et al.* 2013; Wira *et al.* 2002). The purpose of the biological defense is to fight off pathogens before they penetrate the FGT epithelial layer and circulate. The epithelial layer also assists in biological pathogen protection by expressing receptors (*e.g.,* Toll-like receptors (TLRs)) and major histocompatibility complex (MHC) molecules). Receptors aid in identifying pathogens and activating the cellular immune response (Wira *et al.* 2005). For example, activated TLRs stimulate epithelial cells to produce various cytokines and chemokines. Cytokines and chemokines assist in recruiting and activating immune cells to elicit an immune response (Wira *et al.* 2005b). Thus, the health of the lower FGT is determined in part by the interaction between host epithelial cells and other components of the immune system (Wira *et al.* 2005b).

Female genital tract microbiome

Another key component of the FGT that influences health is the microbiome. The microbiome is a physiological niche containing an aggregate group of microorganisms, which are generally known to include bacteria, archaea, fungi, and viruses (Cheng *et al.* 2018; Wu *et al.* 2010). The microbiome provides metabolic and physiological functions typically not characterized by the host. Such functions include the ability to modify inflammatory and immune responses (Pflughoeft and Versalovic 2012). Conversely, the host's reactions and immune responses are able to influence microbiome composition (Günther *et al.* 2016).

Vaginal inhabitants associated with an optimal microbiome play a vital role in the health and functioning of the vagina. These roles include sustaining a low pH ($\leq$4.5), preventing the entry of pathogenic microorganisms (*e.g.,* by influencing barrier function and competitive exclusion), and regulating the local innate immune system (Aldunate *et al.* 2015; Klebanoff *et al.* 1991).

The FGT microbiome is both highly dynamic and variable. Microbiome variation arises due to changes in the taxonomic composition, diversity, and physiological potential (Foster *et al.* 2008). Changes in taxonomic composition can include an increase in anaerobes and facultative anaerobes, which has the potential to disrupt the ability of the vaginal environment to perform its function. In the case where vaginal functioning is disrupted, diseases are likely to develop (Fig 1.2; Foster *et al.* 2008). FGT microbiome variation can occur as a result of many different factors which include, diet, race/ethnicity, sexual activity, hormonal contraception (HC), endogenous hormones, antimicrobial use, and personal hygiene practices (Fig 1.5; Brotman 2011; Fethers *et al.* 2008; Fettweis *et al.* 2014; Gajer *et al.* 2012; McClellend *et al.* 2008). Due to the influence of these different factors, the microbiome varies between females and within females at different time points (Carlson *et al.* 2017; Dworecka-Kas *et al.* 2016).

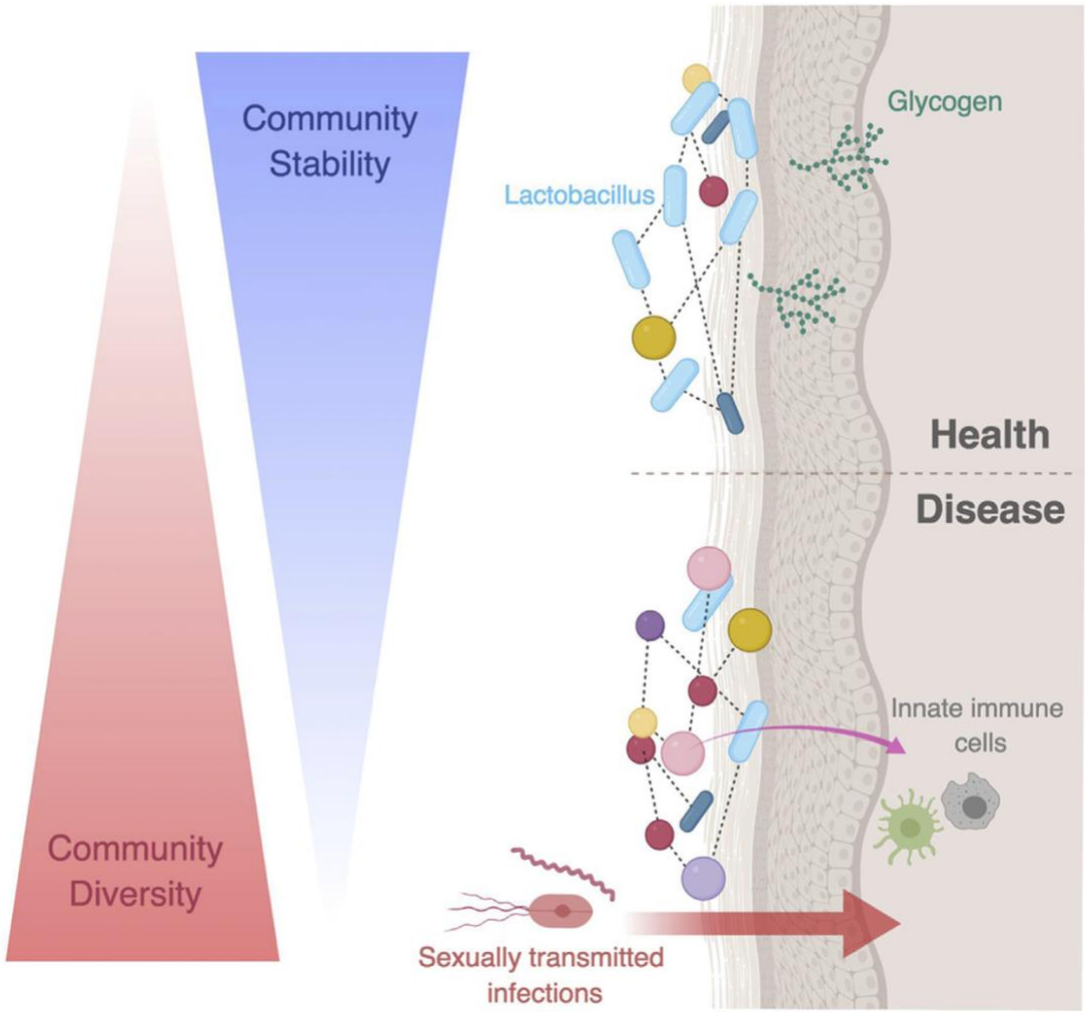


Figure 1.2. Diagram illustrating the relationship between community diversity and stability of the female genital tract (FGT) during the healthy and diseased states (Adapted from Greenbaum *et al.* 2019). In the vaginal microbiome, the healthy state is associated with low community diversity and increased prevalence of *Lactobacillus*. In disease states, the communities observed are more diverse and less stable. The increased diversity of disease-associated microorganisms can negatively modify the host's innate immune response and increase the likelihood of sexually transmitted infections (STIs).

Bacterial vaginal microbiome

The bacterial VMB can be sub-classified into a minimum of three categories, based on clinical diagnostic criteria. Namely, an optimal microbiota, a non-optimal microbiota with a predominance of bacteria linked to BV, and a third category known as intermediate BV (Hillier *et al.* 1992; Sha *et al.* 2005). Currently, little is known about the role of the intermediate vaginal microbiota, which can either be regarded as the transition from a healthy to a non-optimal state

or as the transition from a non-optimal back to a healthy VMB state (Donders 2007). Intermediate BV is a highly heterogeneous group that may or may not include *Lactobacillus,* and that may or may not be characterized by the presence of anaerobic bacteria (Hillier *et al.* 1992; Nugent *et al.* 1991).

A healthy VMB is characteristically a low-diversity environment dominated by lactic-acid producing *Lactobacillus* spp. (Fig 1.2; Fig 1.3; Fredericks *et al.* 2005). Lactic-acid producing species contribute to FGT immunity by providing non-specific defense against a variety of pathogenic microorganisms (Aroutcheva *et al.* 2001). Their defense mechanisms may include synthesizing lactic acid (which aids in maintaining an acidic pH) (Fig 1.3), antimicrobial hydrogen peroxide, and bacteriocins (Aroutcheva *et al.* 2001; Smith and Ravel 2017). These bacteria also assist by forming adherent colonies on epithelial cells and co-aggregating with pathogens. This protects against pathogens by providing a physical barrier that inhibits pathogens from binding to host cells, allowing for a more effective clearance (Mastromarino *et al.* 2002; Reid *et al.* 1998; Spurbeck *et al.* 2008). The ability of the host to allow lactic-acid producing bacteria to persist yet defend against invading pathogens is a result of the bi-directional relationship between the microbiome and the mucosal immune system (Fig 1.4; Ravel and Brotman 2016; Sansonetti 2011).

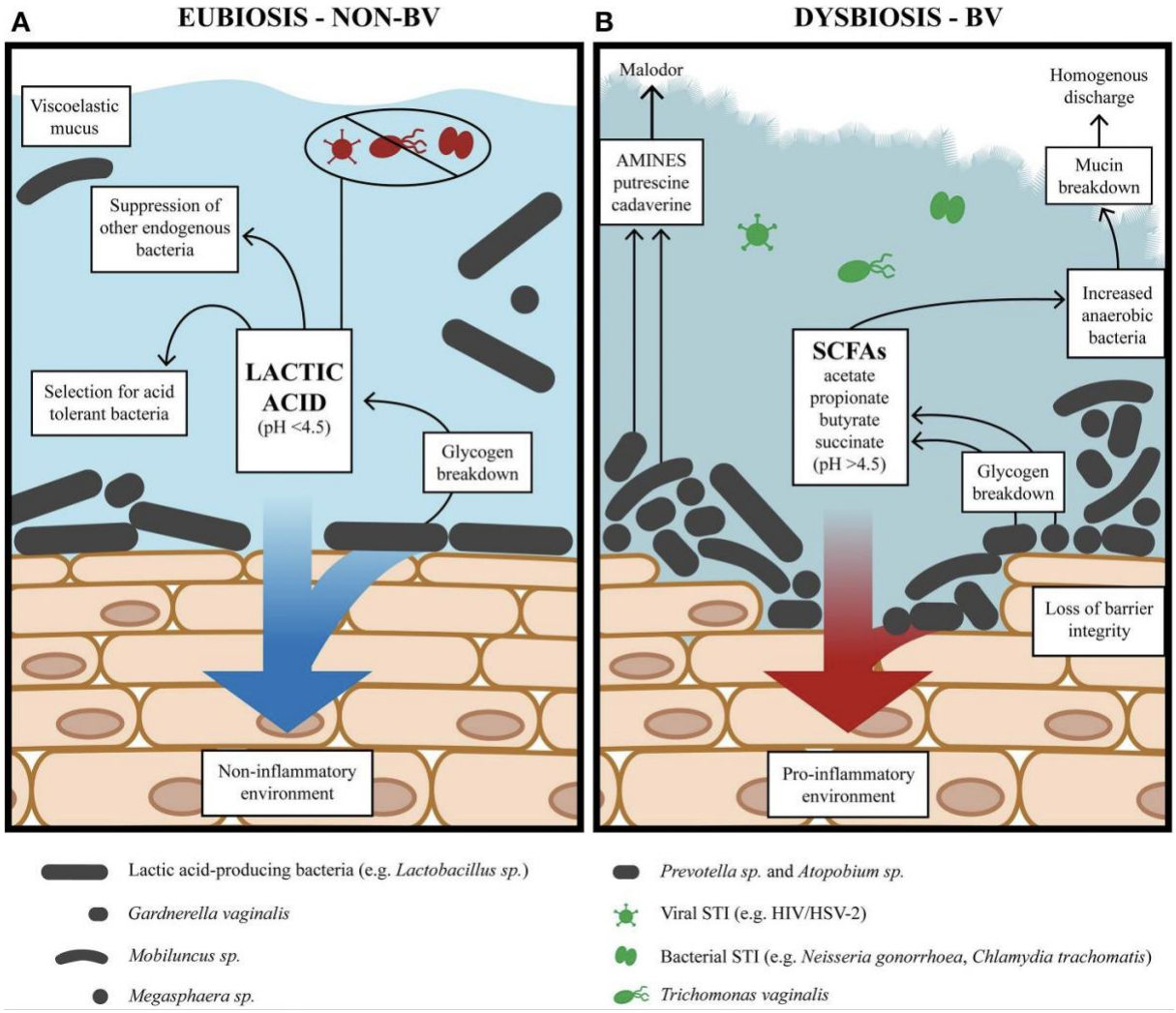


Figure 1.3. The vaginal environment during eubiosis (an optimal microbiota) and dysbiosis (bacterial vaginosis (BV)). Panel A: The optimal microbiome has lactic acid-producing bacteria, which contribute towards the acid nature of the vagina (pH <4.5), with lactic acid as the predominant metabolite. Lactic acid inactivates sexually transmitted infections, while lactic acid-producing species, such as Lactobacillus, promote a non-inflammatory environment. Panel B: During dysbiosis, bacterial diversity increases and the environment becomes more conducive for the growth of anaerobic bacteria. The concentration of mixed short chain fatty acids (SCFAs) and amines increase and the vaginal pH level increases (pH >4.5) reducing acidity. Anaerobic bacteria produce virulence factors which compromise epithelial barrier integrity, degrade mucin, and promote a pro-inflammatory environment (adapted from Aldunate *et al.* 2015).

Bacterial Vaginosis

BV is a condition associated with increased bacterial diversity and reduced stability (Fig 1.3; Fredericks *et al.* 2005). The bacterial composition of BV varies between females, but, in most cases, there is a reduction in *Lactobacillus* spp., and an increase in facultative anaerobic or

anaerobic bacteria (such as *Gardnerella vaginalis*) and vaginal pH (Fig 1.3; Fredricks *et al.* 2005; Ravel *et al.* 2011). Microbial communities dominated by pathogenic bacteria (such as *G. vaginalis* and *Prevotella*) elicit an immune response (Fig 1.3) by increasing the expression of certain receptor pathways (*e.g.,* TLR). This results in the upregulation of pro-inflammatory cytokines and chemokines (Fig 1.3; Anahtar *et al.* 2016). As a result, BV is associated with FGT inflammation (Fig 1.3; Fig 1.4; Lennard *et al.* 2018; Masson *et al.* 2014; Mitchell *et al.* 2014).

No single taxon has been implicated to cause BV as this disease can be found in women with widely varying vaginal microbiomes (Fredricks *et al.* 2005; Ferris *et al.* 2007; Lamont *et al.* 2011; Ravel *et al.* 2011). However, we do know that highly diverse communities are suggested to have negative implications for the FGT mucosal barrier (Anahtar *et al.* 2016). As previously mentioned, the mucosal barrier prevents attacking pathogens from FGT penetration (Burgener *et al.* 2015). Hence, VMB dysbiosis (changes in the amount, composition, distribution, function, and metabolic activity of physiological microbiome) may directly enhance the susceptibility of the host to infections, such as STIs (Fig 1.2), due to mucosal barrier disruption, wound repair inhibition, and inflammation (Zevin *et al.* 2017).

Although the majority of women with BV do not present clinical signs or symptoms, in some cases, BV may be characterized clinically by itching, pain, burning, odor, and/or discharge, and is often diagnosed based on a combination of symptoms, vaginal pH and cytological findings (Livengood *et al*. 1990). In a clinical environment, BV is characteristically diagnosed with Amsel's criteria (Table 1.1; Amsel *et al.* 1983). For a patient to be diagnosed with BV, at least three of the four criteria must be met. In research settings, BV is normally diagnosed by Gram-staining vaginal smears and viewing slides under a microscope. Gram-staining is used to distinguish bacterial communities dominated by *Lactobacillus* from communities dominated with small Gram-variable rods (*e.g., G. vaginalis*) and curved Gram-negative rods (*e.g., Mobiluncus* spp.) (Martin and Marrazzo 2016; Oakley *et al.* 2008; Sobel 2000). The Gram stain is scored on a scale from 1 to 10, known as the Nugent scoring system, and is regarded as the benchmark (Table 1.2; Datcu *et al.* 2014; Nugent *et al.* 1991).

Table 1.1 Amsel's Criteria used in the clinical diagnosis of bacterial vaginosis (BV) (Amsel *et al.* 1983).

Variable	Indicator

Vaginal discharge	Homogenous and white
Whiff test of vaginal discharge	Fishy odor
Clue cells (*i.e.,* squamous epithelial cells covered with adherent bacteria)	Present
Vaginal pH	>4.5

Table 1.2 Nugent's Scoring used to typically diagnose bacterial vaginosis (BV) in a research setting (Nugent *et al.* 1991).

Score	**BV-status**
≤3	BV-negative (*Lactobacillus*-dominated)
4-6	BV-intermediate
≥7	BV-positive

As mentioned earlier, a significant number of BV-infected women are asymptomatic (Mlisana *et al.* 2012). Since asymptomatic individuals do not show any symptoms, they do not receive medical treatment and live with BV unaware (Mlisana *et al.* 2012). This is a concern as BV is linked to several major health concerns, which include premature birth, cervicovaginal inflammation, upper FGT infections, and an increased likelihood of contracting STIs (Fig 1.2; Fig 4; Fethers *et al.* 2008; Srinivasan *et al.* 2012; Taha *et al.* 1998). These STIs include *Neisseria gonorrhoeae*, *Chlamydia trachomatis*, *Trichomonas vaginalis*, HSV, human papillomavirus (HPV), and HIV (Buvé *et al.* 2014; Fethers *et al.* 2008; Srinivasan *et al.* 2012). In the case where an individual is HIV-positive, BV correlates with an increased viral load and severity of immunodeficiency status (Borgdorff *et al.* 2014; Cohn *et al.* 2005).

Current treatment methods for BV include broad-spectrum antibiotics (Bradshaw *et al*. 2013). However, these methods have a high failure rate, since antibiotics cannot effectively penetrate bacterial biofilms implicated with BV and fail to promote recolonization of *Lactobacillus* to the optimal state, resulting in the frequent recurrence of BV (Hay 2009; Soto 2013; van de Wijgert *et al.* 2020). Furthermore, the efficacy of BV treatment methods has not been improved for decades (Bradshaw *et al.* 2013). As mentioned previously, this is a particular challenge in Sub-

Saharan Africa, as the prevalence of vaginal disease-causing conditions is the highest in the world (WHO 2012; Torrone *et al.* 2018). According to Torrone *et al.* (2018), the summary estimate of BV in SA is 42.1% in young women, affecting almost half the young adult population.

Sexually transmitted infections

STIs are associated with FGT inflammation (Haase 2011; Mitchell *et al.* 2014). While inflammatory responses can be beneficial and necessary to effectively eliminate some STIs, the presence of elevated genital inflammation in women is associated with an increased risk of HIV-acquisition and sexual transmission of HIV (Fig 1.4; Cohen *et al.* 2012; Hasse 2011; Martin *et al.* 1999; Masson *et al.* 2015; McKinnon *et al.* 2018; Morrison *et al.* 2014; Taha *et al.* 1999). In SA, HIV is extremely prevalent, with over 7.2 million individuals infected making this the largest HIV seropositive population in the world (Satoh and Boyer 2019). The high prevalence of HIV overlaps with high rates of BV, other STIs, and FGT inflammation.

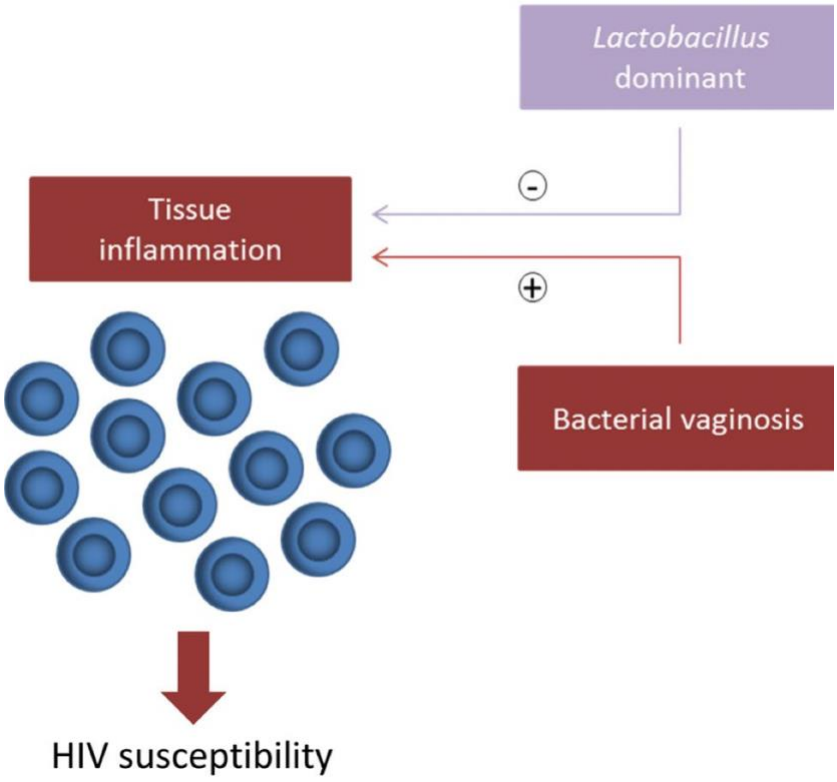


Figure 1.1. Diagram illustrating the effect of a *Lactobacillus*-dominant microbiome and the microbiome during bacterial vaginosis on tissue inflammation and HIV susceptibility (Adapted from Vitali *et al.* 2017).

With regard to bacterial STIs, several studies have reported a high prevalence of *C. trachomatis* infections in South African females ranging from 8.0% to 20.6% (Barnabas *et al.* 2017; Chirenje *et al.* 2017; Garrett *et al.* 2017; Kapiga *et al.* 2009; Torrone *et al.* 2018). The reported prevalence of other bacterial STIs in SA ranges between 1.4% and 8.9% for *N. gonorrhoeae*, 3.1% and 20.0% for *T. vaginalis,* and 9.6% for M. *genitalium* (Kaida *et al.* 2019; Torrone *et al.* 2018). The prevalence of the viral STI, HSV-2, was the highest and ranged between 31.9% and 53.7% (Torrone *et al.* 2018).

The current treatment methods used for bacterial STIs may place women at risk for other diseases. For example, antibiotics have been linked to fungal infection and overgrowth, and may increase the risk of symptomatic VVC in susceptible females (Dollive *et al.* 2013; Noverr *et al.* 2004; Pirotta and Garland 2006; Sobel 1985). On the other hand, viral STIs cannot be cured, instead, patients undergo treatment to manage symptoms. For example, HIV is treated with antiretroviral therapy (ART) to control HIV infection.

Vaginal mycobiome

The mycobiome (also referred to as the fungome) consists of the fungal species that inhabit the microbiome (Nash *et al.* 2017). Therefore, the fungal community inhabiting the lower FGT is termed the vaginal mycobiome. Human microbiome-colonizing fungi are capable of altering host function and behavior; metabolic function; energy acquisition; vitamin-cofactor availability; and immune system development and functioning (Dworecka-Kaszak *et al.* 2016). Fungi also likely function in maintaining microbial community structure (Dworecka-Kaszak *et al.* 2016; Wisecaver *et al.* 2014).

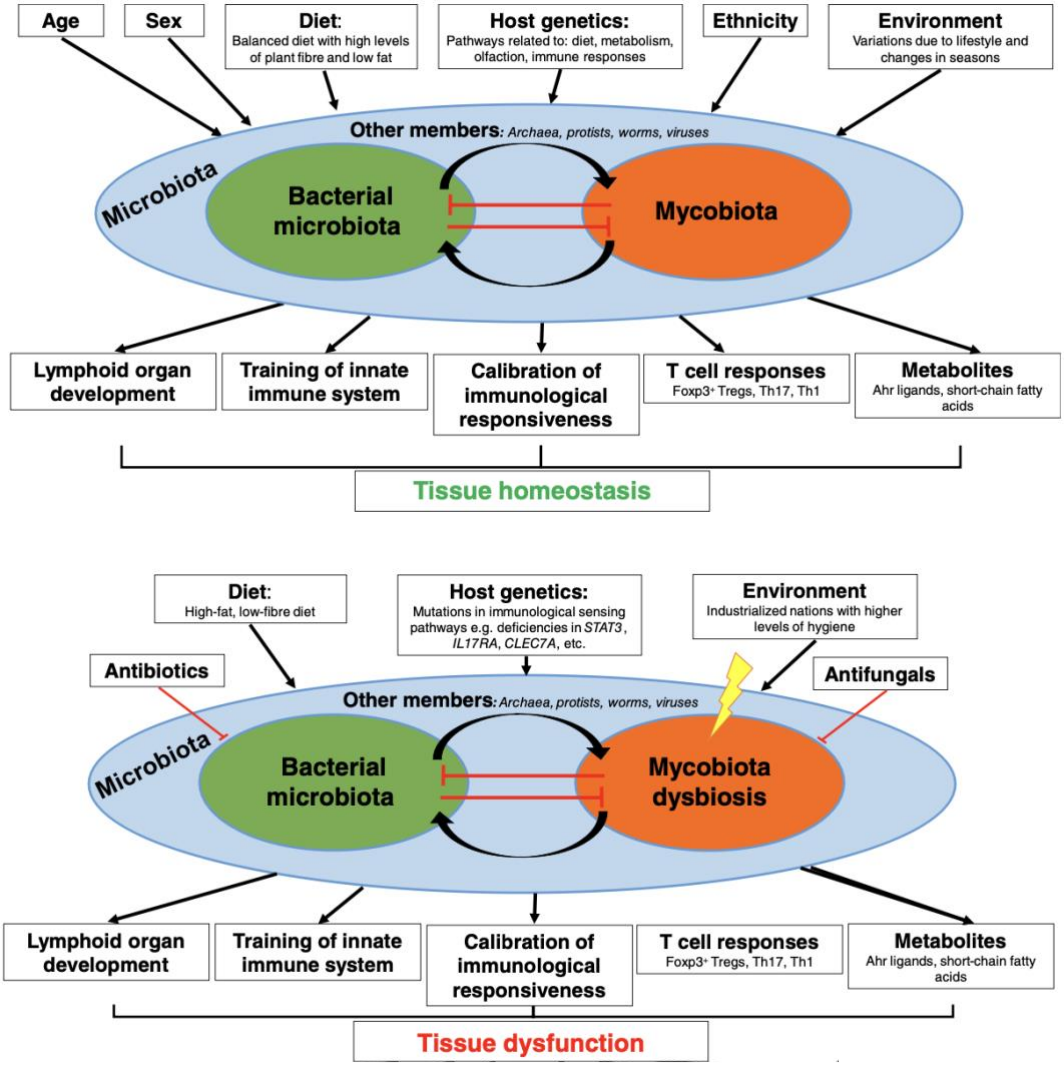


Figure 1.5. Host and environmental factors, including interactions between microorganisms across kingdoms, contribute to bacteriome and mycobiome diversity. Fungi, like bacteria, contribute to diverse facets of human health. In healthy individuals, the mycobiome modifies host physiology, such as host immunity, in a variety of ways, and contributes to tissue homeostasis (upper panel). Whereas

the composition and function of the mycobiome can be modulated by an assortment of factors, including genetic mutations, diet, the use of antibiotics or antifungals, and influences from the bacteriome. Some bacteria are able to inhibit certain fungi, and likewise, fungi are able to inhibit certain bacteria. Increased dietary fat intake has been associated with an increased risk of BV and severe BV. Of late, host genetics has been investigated for its influence on the VMB. Mycobiome dysbiosis results in pathological tissue function, excessive inflammation, and disease (lower panel) (Adapted from Lai *et al.* 2019).

Similar to the bacteriome, the mycobiome varies between individuals, body sites, and disease status (Fig 1.6; Nash *et al.* 2017; Nguyen *et al.* 2015; Witherden *et al.* 2017; Zhang *et al.* 2011). For instance, fungal diversity varies substantially between body sites located further apart, while similar constituents are observed at nearby body sites (Fig 1.6; Ghannoum *et al.* 2010; Sellart-Altisen *et al.* 2007). In addition, the mycobiome evolves with age, diet, antibiotic usage, and hygiene practices; consequently, the mycobiome also varies within individuals over time (Fig 1.5; Lai *et al.* 2017).

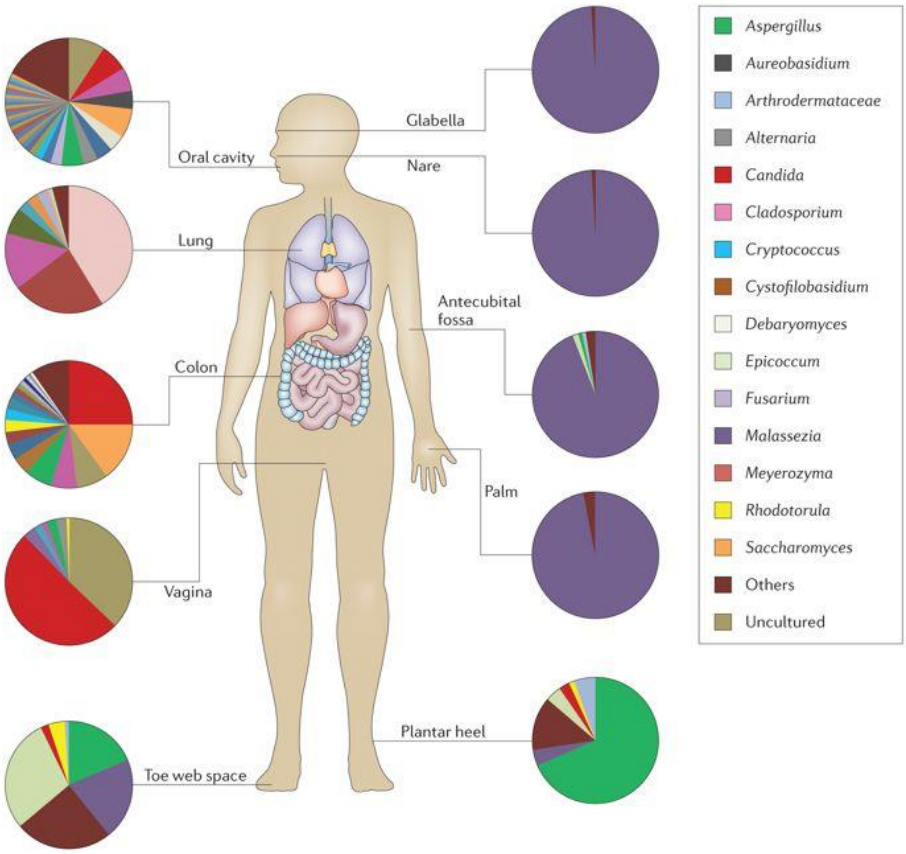


Figure 1.6. Diagram depicting the mycobiome makeup of various human body sites (Adapted from Underhill and Iliev 2014).

The yeast species, *Candida albicans* can be isolated in up to 80% of healthy individuals, owing to the ability of this species to inhabit a diverse range of host niches (Neder 1992). However, research using culture-independent analyses has revealed a more diverse population of resident fungi than just common yeast (Dworecka-Kaszak *et al.* 2016). In general, the phyla Ascomycota (including *Candida* spp., *Cladosporium* spp., and *Saccharomyces cerevisiae*) and Basidiomycota (including *Cryptococcus* spp., and *Malassezia* spp.) have been found to dominate most human body sites (Ghannoum *et al.* 2010; Hoffmann *et al.* 2013; van Woerden *et al.* 2013; Zhang *et al.* 2011).

A study conducted by Drell *et al.* (2013) analysed the vaginal mycobiome in healthy and asymptomatic females using Internal Transcribed 1 (ITS-1) next-generation sequencing (NGS). This study found that more than half of the fungal inhabitants (58%) were of the phylum Ascomycota, of which *C. albicans* was the most abundant species (34%). Followed by the phylum Basidiomycota (3%), of which *Rhodotorula sp LH51* was the most abundant species (1%) (Drell *et al.* 2013). However, a significant proportion of mycobiome sequences could not be matched to a specific taxon and remained unclassified (Cui *et al.* 2013; Rizzetto *et al.* 2014; Seed 2014). The high percentage of unclassified fungal sequences highlights how premature this field is.

Current mycobiome research is predominantly directed toward the fungal inhabitants of non-diseased versus diseased individuals (Bradford and Ravel 2017; Hager and Ghannoum 2018; Sam *et al.* 2017). The motivation behind this research is to obtain a better understanding of the influential role of the mycobiome on disease progression, from the beneficial functions provided by symbiotic fungi on host immunity to the deleterious effects of the same fungi during dysbiosis (Lai *et al.* 2018; Nash *et al.* 2017). Fungal dysbiosis, like bacterial, has been linked with several vaginal pathologic conditions, such as recurrent vaginal candidiasis (Guo *et al.* 2012). The majority of human-colonizing fungi are opportunistic pathogens that result in disease when the immune system is compromised. For example, immunosuppressed HIV-positive individuals have a greater susceptibility to contracting opportunistic fungal infections (Ackerman and Underhill 2017; Perfect and Casadevall 2006).

Vaginal fungal infections

VVC is an acute inflammatory condition affecting the vulva and vaginal mucosa induced by the overgrowth of *Candida* spp. (Fidel *et al.* 2004). VVC is the second most reported form of infectious vaginitis, as at least 30% of total vaginitis cases are attributed to VVC (Sobel *et al.* 1998). *C. albicans* accounts for approximately 70 to 90% of VVC cases, and the remaining infections are caused by *C. glabarata, C. parapsilosis, C. krusei*, *C. guilliermondii, C. kefyr*, *C. lusitaniae*, *C. viswanathii*, *C. dubliniensis*, *C. famata,* and *C. tropicalis* (Neder 1992; Richter *et al.* 2005; Sobel 2007; Workowski *et al.* 2010*). Candida* is a constituent of the healthy microbiota, and most individuals are asymptomatically colonized with this species (Beigi *et al.* 2004; Bradford and Ravel 2017; Drell *et al.* 2013). However, *Candida* has the potential to be pathogenic and proliferate (Mayer *et al.* 2013). Functions beneficial to Candida's survival include the ability to adhere to and attack host cells, morphologically transition, biofilm formation, and possess unique metabolic and fitness characteristics (Mayer *et al.* 2013). As a result, candidiasis is considered an opportunistic infection.

VVC can be symptomatic (ranging from mild to severe symptoms; Table 1.3) or asymptomatic (total absence of symptoms) (Sobel *et al*. 1998). While VVC may occur without any recognizable trigger, specific conditions that interfere with the balance of the normal vaginal microbiota may predispose individuals to symptomatic infection (Bingham 1999; Fidel 2004). These conditions include antibiotic usage, HC, high estrogen levels (caused by pregnancy and hormone replacement therapy), and medical diseases (*e.g.,* diabetes and HIV) (Edwards 2004; Helfgott *et al.* 2000; Sobel 1997).

Table 1.3 Clinical features of vulvovaginal candidiasis (VVC) (Sobel 1992).

Symptoms	**Indicator**
Vaginal discharge	White, thick, clumpy
Pruritus	Vaginal and vulva itching
Burning sensation	
Erythema	Red vaginal walls

To diagnose VVC, vaginal specimens are collected for laboratory confirmation using wet preparation microscopy. A droplet of potassium hydroxide is added to the wet mount to visualize hyphae or spores. As VVC is an inflammatory condition, during a pelvic exam inflammation of the vaginal and vulvar area is evident during a pelvic exam and white blood cells are frequently present on vaginal specimens. Nowadays, polymerase chain reaction (PCR) methods are readily available to detect the presence of the genus Candida (Cartwright *et al.* 2013; Sobel and Akins 2015; Weissenbacher *et al.* 2009).

Treatment of this condition is necessary when patients present symptoms consistent with VVC or present a laboratory confirmation of *Candida* from their vaginal specimen. Symptomatic patients can either be treated with over-the-counter antifungals if self-diagnosed or patients are prescribed oral or topical antifungal medications (Yano *et al.* 2019). However, the main symptoms of VVC often closely resemble those of BV, leading to incorrect diagnosis and treatment (Ferris *et al.* 2002). What differs between the two infections is that the vaginal pH is usually <4.5 in patients with VVC, and the whiff test is negative (Moodley *et al.* 2002). However, there is an exception when there is a coinfection with BV or *T. vaginalis*. If VVC is inaccurately diagnosed as BV, it creates an issue as fungi are entirely different organisms from bacteria and require fundamentally different drugs and treatment methods. Therefore, accurate diagnostic testing is crucial for VVC management.

As mentioned earlier, antibacterial treatments are thought to be able to trigger symptomatic candidiasis in females colonized with *C. albicans* or at risk for VVC (Pirotta and Garland 2006; Sobel 1985). This is likely due to the decrease in the resistance provided by colonizing bacteria against *Candida* (Fidel 2002; Yano *et al.* 2012). Conversely, Donders *et al.* (2011) indicated that BV is more likely to develop when *Candida* is eliminated by antifungals.

Since fungi are severely under-researched, especially in the context of infectious diseases, only a few antifungal drugs are currently accessible, making it easy for fungal resistance to emerge due to their extensive use (Antonovics *et al.* 2007; Cowen 2008; Low *et al.* 2006). Subsequently, the increasing emergence of antifungal resistant pathogens restricts already limited treatment options, making clinical management of candidiasis increasingly challenging and creating a dire need for improved drug therapies (Antonovics *et al*. 2007; Cowen 2008; Cowen *et al.* 2015; Perlroth *et al.* 2007). Focusing on biological factors that promote the emergence of antifungal resistance is essential to developing advanced therapeutics, as

improved diagnostics and intervention approaches are vital to overcome and prevent drug resistance (Cowen *et al.* 2015).

In summary, the overuse of antimicrobials to treat bacterial and fungal infections is creating a new breed of "superbugs", giving rise to one of the world's most irreversible health threats: drug-resistant infections. Without information on the role of the mycobiome, it is difficult to develop effective treatments, such as novel antimicrobials, to address and overcome these prevalent health concerns.

Importance of studying the mycobiome

Despite fungi being at a lower abundance than their bacterial counterparts across body sites, they are an essential component of the microbiome (Qin *et al.* 2010). It has been estimated that the total number of fungal cells is orders of magnitude lower than the number of bacterial cells in the human body (Qin *et al.* 2010). Shotgun metagenomics sequencing analysis has revealed that fungi consist of nearly 0.1% of the total microorganisms in the gut (Qin *et al.* 2010; Arumugam *et al.* 2011). In the vagina, the range has been found to be between 0.17 ± 0.04% if fungi are detected (Ma *et al.* 2020).

Yet the presence of fungal organisms that reside in the FGT has been extremely undervalued and understudied (Bradford and Ravel 2017). The earliest study ever to utilize NGS to examine the vaginal mycobiome was only published in 2013, inferring that this is still a new area of research (Drell *et al.* 2013). To date, the majority of human microbiome studies have focused primarily on bacterial communities (also referred to as the bacteriome) (Ackerman and Underhill 2017; Nash *et al.* 2017). The vast amount of bacterial research can be partly attributed to the extensive use of 16S rRNA sequencing methods, which do not detect fungi. Even though metagenomics is a powerful approach, it suffers from a number of inherent weaknesses. Since bacteria dominate the human microbiome, sequencing approaches are restricted in their ability to identify non-bacterial organisms present at a low relative abundance in the human microbiome, proving unsuitable for fungal studies (Morgan and Huttenhower 2014). Extremely deep and costly sequencing is required to extract fungal sequences in a collected sample (Limon *et al.* 2017). Furthermore, many reference genomes are available for bacteria found in the human microbiome, but only a few fungal genomes are available, which may partially contribute to this observed difference (Underhill and Iliev 2014).

Most of what we know about the FGT mycobiome is derived from studies focused on the key player, *C. albicans*. As mentioned previously, *C. albicans* has been implicated as one of the leading causes of vaginal infection (Drell *et al.* 2013). In the past, human mycobiome characterizations were solely reliant on culture-based techniques, which restricted the fungi researchers could access (Forbes *et al.* 2019). Therefore, analyses were exclusively limited to known human pathogens (*e.g., C. albicans*, *C. glabrata*, *Cryptococcus neoformans*, *Aspergillus fumigatus,* etc.) (Perfect and Casadevall 2006).

Candida spp. and *Malassezia* spp. are constituents of a healthy VMB; however, these fungi are able to live a dual lifestyle and function as pathogens (Romani 2011). *C. albicans*, for example, contains an arsenal of virulence and fitness attributes that contribute to both its ability to serve as a commensal and a pathogen (Mayer *et al.* 2013). Virulence factors that contribute to pathogenicity include adhesins, secreted hydrolytic enzymes, the formation of biofilms, and the ability to morph (Deorukhkar and Roushani 2017). The ability of *C. albicans* to morph allows it to reversibly transition from a yeast cell to a hyphal growing organism (Cassone 2015; Harriot *et al.* 2010; Majumdar *et al.* 2016; Noble *et al*. 2017). The hyphal state is required for vaginal infection, and studies have shown that when the key transcriptional regulators controlling the switch from the yeast to the hyphal state were deleted from *C. albicans* strains and inoculated into mice, the strain is typically non-pathogenic (Macdonald and Odds 1983). The fitness attributes of *C. albicans* allow it to adapt to pH and CO_2 changes, stress, and nutrient starvation (Nicholls *et al.* 2011). Therefore, commensal fungal populations that occupy the vagina are assumed to fluctuate in fungal species over time and in response to diseases (Sam *et al.* 2017; Underhill and Iliev 2014). Thus, during VVC, *Candida* spp. have the tendency to outgrow (Peters *et al.* 2014a)

One of the main issues with annotating fungal taxonomy using molecular-based methods is due to the underrepresentation of fungi in existing reference databases (Bradford and Ravel 2017). For instance, the gut microbiome study that proposed that fungi constitute 0.1% of the microbiome likely underestimated the abundance and significance of fungi (Arumugam *et al.* 2011; Qin *et al.* 2010). This approximation was estimated based on fungal sequences identified from annotated reference databases in which fungi are highly underrepresented. A simple text search performed on the NCBI nucleotide database reveals that there are 209,784 entries for vaginal-specific bacteria, whereas there are only 1,458 entries for fungi. These numbers highlight the obvious disparity between fungi and bacteria in sequence databases (Virtanen *et*

al. 2021). Furthermore, unculturable fungi make up a major proportion of the human mycobiome. This leads to a large percentage of sequences present in databases that have been annotated as uncultured or misclassified (Cui *et al.* 2013). Thus, it is essential to meticulously annotate sequences and curate databases in order to adequately detect fungi in microbiome analyses (Bull-Otterson *et al.* 2013; Devkota *et al.* 2012; Sonnenberg *et al.* 2012). An additional limitation of fungal analyses is that samples are prone to environmental contamination, such as airborne fungi, which further disadvantage taxonomic analysis (Drell *et al.* 2013). Contamination results in a considerable rate of false-positives that generates misleading results, requiring rigorous quality control processes (Khot and Fredericks 2009; Salter *et al.* 2014).

There is also limited information available on the mechanisms by which fungi interact with other microbiome constituents and the host. (Bradford and Ravel 2017; Klimesova *et al.* 2018; Ward *et al.* 2018; Witherden *et al.* 2017). These interactions are more complex than we once assumed as fungal growth has shown to be regulated in part by other microorganisms present in the microbiota (Shankar *et al.* 2015). Furthermore, immunosuppression that results from infections such as HIV has been shown to influence the fungal inhabitants of the microbiome (Chanda *et al.* 2017; Hager and Ghannoum 2018). Such interactions may contribute to disease development (Ward *et al.* 2018); therefore, more comprehensive research on the fungal constituents, functions, and compositional dynamics of the human mycobiome is needed.

To understand the function of the mycobiome and the mechanisms by which certain fungi contribute to or defend against diseases, researchers need a systems-level integrated method, rather than concentrating on certain disease-causing taxa (Huffnagle and Noverr 2013; Rizzetto *et al.* 2014). This information will assist researchers in classifying fungal species that contribute to disease and improve our understanding of fungal-bacterial and mycobiome-immune system relationships that are significant to our health (Cui *et al.* 2013; Hager and Ghannoum 2018).

The above research is also crucial for the development of new targeted, preventative, and personalized therapeutic options (Ma *et al.* 2012). This will allow us to move into translational and clinical arenas (Cui *et al.* 2013; Seed 2014). Future therapeutic options may include, antifungals, genetically manipulated strains (*i.e.,* delivery of targeted microbiome modulators), and biotherapeutic methods to control the abundance of certain microbiome components (Fig 1.7; Forbes *et al.* 2019; Vargason and Anselmo 2018). A biotherapeutic method could include

a probiotic that contains a single or a consortium of microbes selected based on evidence to prevent fungal infections (Fig 1.7; Kumar and Singhi 2016). Interestingly, in 2016, a patent outlining the potential of the mycobiome for a probiotic, diagnostic, and/or treatment tool was approved (Underhill and Iliev 2016).

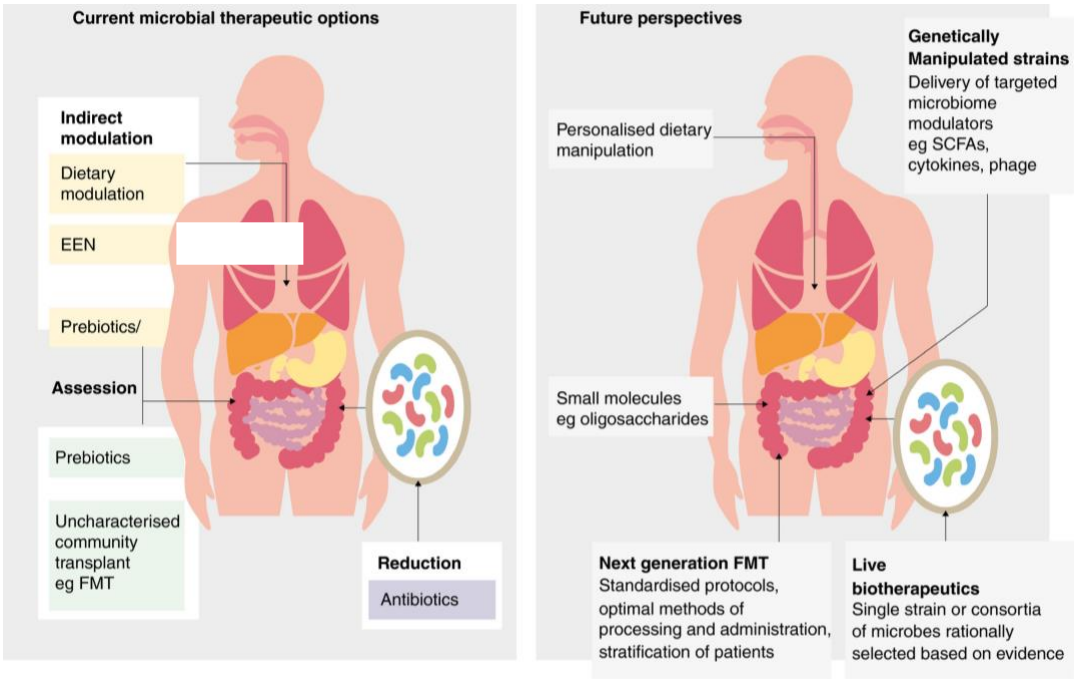


Figure 1.7. Diagram depicting current microbial therapeutic options and future perspectives for microbiome-modulating therapeutics. Existing therapeutic methods to modulate the microbiome are mostly nonspecific, falling into three broad categories: accession (faecal microbiota transplantation (FMT) and probiotics), reduction (antibiotics), or indirect modulation (prebiotics). Future therapeutic approaches should be more targeted and personalised to treat the underlying disease pathophysiology. Currently investigated approaches include live biotherapeutics containing single strains or consortia of bacteria that have been rationally selected and microbial strains that have been manipulated to deliver targeted microbiome modulators (short chain fatty acids (SCFAs), cytokines, and phage). Alongside this, developing personalised dietary manipulation strategies and small molecule delivery will also likely feature (Adapted from Mcilroy *et al.* 2017).

The development of new technology and software will allow researchers to solve most microbiome challenges. This will enable future microbiome studies to advance from merely partially describing taxonomic and functional composition to providing information on significant protein-centric molecular and functional processes (Rechenberger *et al.* 2019). Identification and detailed characterization of fungal species among the population will not only

help health professionals to choose suitable antifungal treatments but also reduce the development of drug resistance (Beed *et al.* 2014; Brown *et al.* 2012).

Studying the microbiome

Research on microbial communities was impeded in the past by the inability of the majority of microorganisms to grow in standard culture media under laboratory conditions. This hindered access to many microorganisms and as a result, culture-independent (multi-)omics methods have been developed to increase our understanding of microbial ecology (Gutleben *et al.* 2017). Multi-meta-omics studies comprise the analyses of total Deoxyribonucleic acid (DNA), Ribonucleic acid (RNA), proteins, and/or metabolites extracted directly from environmental samples (Fig 1.8); and are named metagenomics, metatranscriptomics, metaproteomics, and metabolomics, respectively (Fig 1.8). Multi-omic computational modelling enables a better understanding of microbial functions and their adaptive responses.

Studying “Who is present?” and “What are they doing?”

Microbiome research to date has mostly been done using metagenomics (*i.e.,* 16S rRNA gene sequencing and whole-genome sequencing) and metatranscriptomics. Metagenomics enables the discovery of potentially significant candidate species and genes associated with diseases and environmental conditions (Zoetendal *et al.* 2008). Metagenomics allows us to observe changes in species abundance. However, a limitation with this approach is that the output generated only allows us to predict function. Alternatively, Metatranscriptomics allows us to observe changes in gene expression at a specific time under certain conditions by capturing the total mRNA (Fig 1.8; Aguiar-Pulido *et al.* 2016). Metatranscriptomics helps obtain a functional profile that provides better confidence in the potential microbiome function. However, this approach is difficult, as prokaryotic mRNA is highly unstable, and body specimens are rapidly processed (Gosalbes *et al.* 2011). Both of the abovementioned approaches, in theory, offer comprehensive and unbiased data, but given the low abundance of fungi from human biological samples, these approaches prove inadequate for mycobiome characterization as it is both difficult and expensive to get an adequate sequencing depth as human and bacterial sequences overwhelm analyses (Ackerman and Underhill 2017).

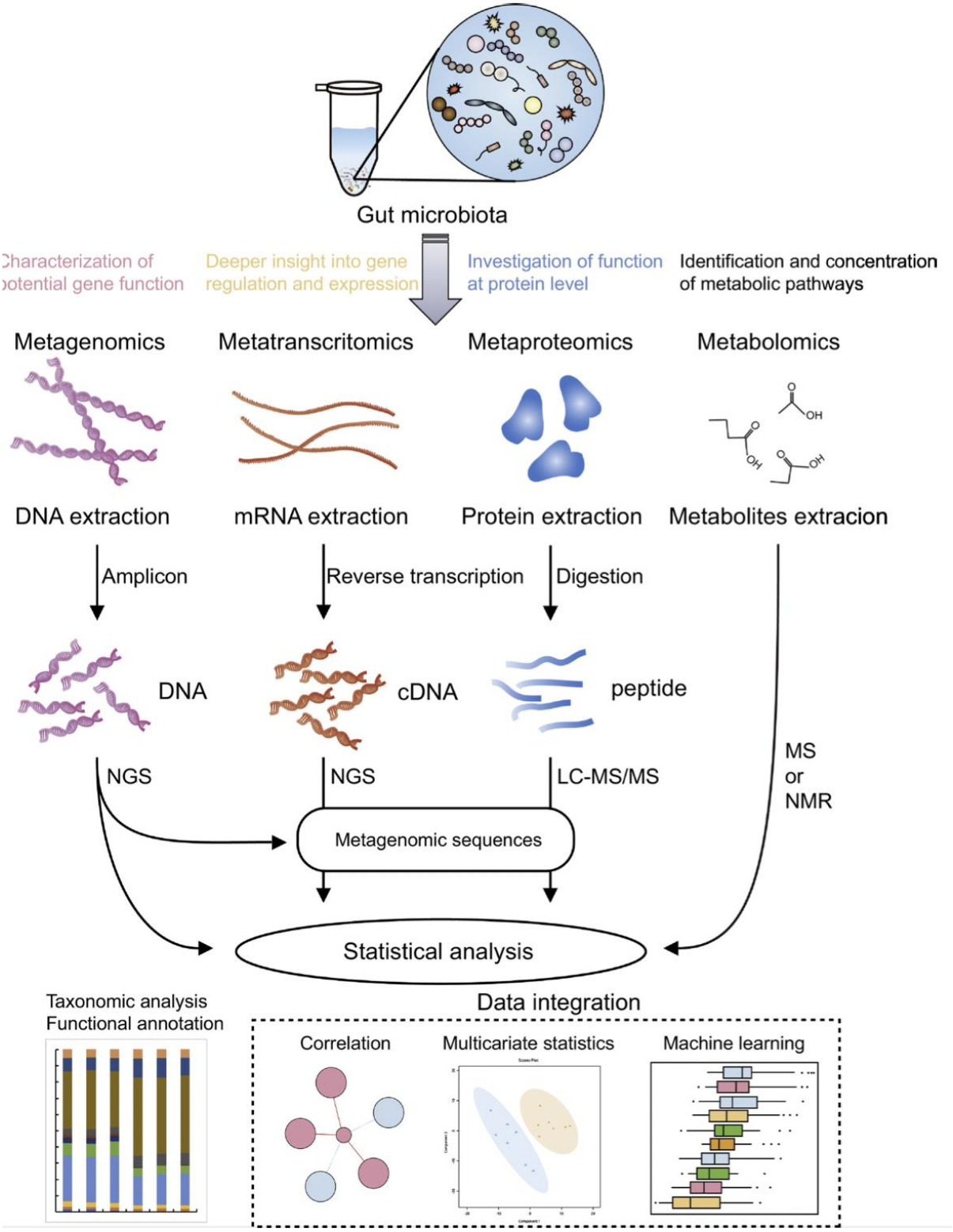


Figure 1.8. Illustration of using multi meta-omics approaches to analyse the microbiome and the data analysis steps involved (adapted from Wang *et al.* 2020). Microbiome studies can either involve the study of DNA, RNA, proteins, and/or metabolites extracted from the sample of interest, and are known as metagenomics, metatranscriptomics, metaproteomics, and metabolomics respectively.

Of late, metaproteomics and metabolomics are being used more often to analyze the microbiome, providing more functional information (Fig 1.8; Aguiar-Pulido *et al.* 2016). Metaproteomics applies the majority of the methods that were created for proteomics research (Kolmeder and de Vos 2013). To clarify, proteomics is the study of the protein composition of a single species or strain, whereas metaproteomics is the analysis of the total proteins in a microbial community (Wilmes and Bond 2006). This approach allows for the identification and quantification of proteins across a broad dynamic range, and detection of post-translational modifications (PTMs). Metaproteomics is an ideal platform to investigate the microbiome at a functional level and assess microbial protein expression in response to host, diet, or other environmental stimuli (Hettich *et al.* 2013; Kolmeder and de Vos 2014; Mao and Franke 2015). Thus, proteins are advantageous as they are proxies of activity, more stable than RNA, and protein analyses represent both information of on-going activities and those that have taken place in the recent past (Taverna and Goldstein 2002).

Metabolomics is the study of metabolites (Fig 1.8), which allows us to understand microbial metabolism and host-microbiota interactions. This omics approach is able to detect the state of microorganisms captured (*i.e*., whether they are active, dormant, or dead) by assessing microbial activity (Wang *et al.* 2020). The main limitation that comes with using metabolomics is that all body fluids interact with cells, and often metabolites are rapidly absorbed.

Both metabolite and metaproteome research are difficult as they require high throughput instruments. Furthermore, sequence databases are frequently incomplete and must be complemented by *de novo* detection (Jansson *et al.* 2009). In spite of this, functional studies aided by metaproteomics and metabolomics are expected to be essential in providing information on the mechanisms used by microbial species involved in disease (Bostanci *et al.* 2021). This may allow for the development of novel treatment methods for human microbiome-associated diseases.

Characterizing microbial communities with metaproteomics

Metaproteomics is a developing field aimed at understanding microbial community structure and functionality through mass spectrometry (MS) (Kleiner *et al.* 2017). This approach presents complementary data to NGS (*i.e.,* the study of genes) on microbial community structure and function. However, both approaches have limitations during taxonomic analysis,

with 16S we do not know if the species we are identifying are dead or alive, and there are copy number variations that create issues with relative abundance (Cosby and Criddle 2003). With metaproteomics, identified peptide sequences are often very short and have a close identity with multiple taxonomic species (Kleiner *et al.* 2017). However, unlike NGS, MS allows researchers to characterize the functional and metabolic activity of microorganisms present (Zevin *et al.* 2016). The metaproteome is not exclusive to the bacteriome as it provides information about host proteins, food residues, resident archaea and fungi, and possibly protozoa, helminths, and viruses (Cheng *et al.* 2018). Furthermore, this approach also provides information on host immune function, which helps our understanding of the interactions that exist between the host and microorganisms (Zevin *et al.* 2016). Affirmingly, Tanca *et al.* (2017) recommended the use of metaproteomics over metagenomics, as the former appeared to be more beneficial in functional annotation of the gut microbiome in healthy and diseased states. Another study by Cortes *et al.* (2019) showed that metaproteomics data yielded far more information than what was achievable from 16S rRNA gene sequencing on the same samples. As metaproteomics was able to identify more microorganisms with better resolution in each subject (Cortes *et al.* 2019).

The metaproteomics workflow typically entails protein extraction, fractionation and digestion, peptide separation, and tandem mass spectrometry (MS/MS) analysis (Fig 8; Heyer *et al.* 2015). Protein identification involves comparing the mass spectra obtained from experimental metaproteomics to theoretical mass spectra predicted from comprehensive protein databases (Heyer *et al.* 2015). After identification and quantification, peptide sequences are generally assigned to their source species using the NCBI Taxonomy database, UniProt, or other sequence databases (Wang *et al.* 2020). The functions of proteins are determined by their specific amino acid sequences, which are highly conserved, through databases such as clusters of orthologous groups (COG), Kyoto Encyclopedia of Genes and Genomes (KEGG), gene ontology (GO), and the evolutionary genealogy of genes: non-supervised orthologous (eggNOG) (Ashburner *et al.* 2000; Wang *et al.* 2020).

Of late, there has been an increase in the number of metaproteomic analyses owing to improved experimental and bioinformatic approaches. Experimental improvements include mass spectrometers with faster acquisition speeds, higher accuracy and resolution, and higher efficiency of protein and peptide separation (Schneider and Riedel 2010; Tanca *et al.* 2014; Zhang *et al.* 2016). Bioinformatics improvements include larger metagenomic databases and increased computing power (Schneider and Riedel 2010). These recent developments have

drastically helped increase microbial protein identifications in microbiome analyses (Zhang *et al.* 2016).

The majority of metaproteomics studies have identified the taxonomic and functional structure of complex microbial communities in their niches (Abram *et al.* 2011; Kan *et al.* 2005; Ram *et al.* 2005; Wilmes and Bond 2006). New research has also noted a link between microbiome variations at taxonomic and functional levels with specific disease and environmental conditions. This link emphasizes how metaproteomic analyses are useful for microbiome research (Brooks *et al*. 2015; Erickson *et al.* 2012; Greenblum *et al*. 2015; Heyer *et al.* 2016; Kolmeder *et al.* 2016). Metaproteomics is further advantageous as it can identify microbial proteins that are post-transcriptionally regulated and translated under specific environmental conditions (Singh and Reddy 2016). In addition, this method has the potential to identify protein biomarkers that are indicative of disease (Siggins *et al.* 2012).

In a functional metaproteomic study by Zevin *et al*. (2016), comparisons between females with BV and without BV were conducted. *Lactobacillus*-dominant individuals had a significantly increased abundance of proteins involved in transport and catabolism, folding, sorting and degradation, and energy metabolism. Whereas individuals with BV (*i.e., Lactobacillus*-scarce) had a greater abundance of *Gardnerella* proteins involved in starch and glycogen degradation, and transport. The enrichment of these proteins suggests that *Gardnerella* metabolism is focused on the catabolism and consumption of extracellular saccharides. As expected, vaginal pH negatively correlated with lactate dehydrogenase, which is responsible for lactic acid production. Consequently, this study provides evidence that metaproteomics may be a viable method of examining the female genital mycobiome between diseased and non-diseased individuals. That will allow us to better understand the activities of the FGT ecosystem and the host-microbe interactions.

Research performed by Alistoltani *et al.* (2020) aimed to utilize both metaproteomics and 16S rRNA gene sequencing to understand the relationship between microbial function and inflammatory profiles in the FGTs of South African women. The results of this study managed to identify a link between FGT microbial function and local inflammatory responses, suggesting that the presence of specific microbial taxa, their properties, and activities likely play a critical role in modulating FGT inflammation. However, a caveat of this study was that quality control measures implemented during taxonomic analyses removed *Candida* from the dataset. As aforementioned, *Candida* is expected to be present in the FGT microbiome, therefore, more

work needed to be done to achieve better fungal assignments using this dataset. Alistoni *et al.* (2020) suggested using an alternative methodology for sample processing to better resolve fungal assignments (Bianco and Perrotta 2015). Our study, therefore, sets out to use this dataset to increase fungal assignments using a different strategy to the primary study in order to better understand the composition of the FGT mycobiome and its function.

Limitations of metaproteomics

There are both benefits and limitations of metaproteomics in comparison to 16S rRNA gene sequence-based methods. For taxonomic annotation, both techniques depend on curated databases and how available and extensive these databases are. 16S rRNA gene repositories are seemingly more extensive, however, protein databases are starting to expand and become more readily accessible.

Metaproteomic data analysis is highly complex (Heyer *et al.* 2017). This is a result of the heterogeneous material surrounding microbes, the complexity of microbial ecosystems, and the high abundance of host proteins compared to microbial proteins (Kolmeder and de Vos 2014). This poses an issue as host proteins are able to overpower analyses, and result in the under-reporting of microbial species present in the microbiome as metaproteomics is biased towards dominant species. Most microbial proteins are secreted and extracellular, which leads to difficulties in distinguishing between human and prokaryotic proteins. In terms of MS/MS identification rate, metaproteomics usually has an MS/MS identification ranging between 10% to 30%; therefore, the majority of proteins present in a sample remain undetected (Siggins et al. 2012; Yuan *et al.* 2014). Furthermore, certain taxa have a much larger number of unique peptides, which biases the taxonomic profile towards these taxa (Heyer *et al.* 2017). The above issues must be considered when creating new metaproteomic bioanalytical methods (Cheng *et al.* 2018).

Metaproteomic data analysis is also computationally demanding and there is a shortage of high sensitivity methods (Kolmeder and de Vos 2014). Currently, the sensitivity of metaproteomics is still on the lower side and only proteins that have an abundance of over 1% are readily detected (Verberkmoes *et al.* 2009). A major challenge that remains is searching against comprehensive and large protein databases. Database searches require considerable computational time to utilize the complete reference search spaces (*e.g.,* NCBI), and there is

a risk that the analyses will fail because of software or hardware limitations. Thus, metaproteomic analysis requires more computational resources, such as memory, processors, and more powerful algorithms. Secondly, large sequence databases increase the false discovery rate (FDR), impairing the rate of protein identification (Muth *et al.* 2015; Vaudel *et al.* 2011).

Protein interference is a commonly occurring issue in metaproteomics due to many highly redundant protein sequences in databases (*i.e.,* identical peptides which belong to homologous proteins resulting in the identification of redundant proteins) (Herbst *et al.* 2016). Consequently, the results of taxonomic and functional analyses become uncertain. For example, a lactate dehydrogenase peptide ferments sugars to lactate in the genus *Lactobacillus*, however, this peptide also ferments lactate to acetate in the order Clostridiales. Therefore, depending on which taxa the peptide belongs to defines the function the peptide will perform (Kohrs *et al.* 2014).

An additional challenge is that a single mutation in any amino acid will result in a different tryptic peptide, which impedes peptide identification of the proteins being investigated. This issue is difficult to avoid as homologous proteins occur in various microorganisms with minor amino acid distinctions. Thus, protein assignment has to be thorough enough to not overestimate proteins, but not too stringent that protein possibilities are excluded (Cheng *et al.* 2018; Qin *et al.* 2010).

Protein identification also becomes increasingly challenging if the taxonomic makeup is undetermined or if protein databases are missing entries of unsequenced microorganisms (Cheng *et al.* 2018). Many fungal species cannot be classified due to the scarcity of fungal sequences in databases (Bradford and Ravel 2017; Drell *et al.* 2013). Therefore, for taxonomic assignments to be of high quality and accuracy, reference databases require high quality and representation of fungal species (Bradford and Ravel 2017).

Research aims and objectives

Mycobiome research is crucial as the vaginal mycobiome may have more effects on health and clinical outcomes than are currently known. Examining the microbiome as a whole (*i.e.,* including bacteria, viruses, archaea, and fungi) would provide a more comprehensive

understanding and would likely help us understand the associations between the microbiome and clinical features (Stout *et al.* 2020).

The first aim of this study was to compare the vaginal mycobiomes of black females residing in SA with and without a diagnosis of STIs or BV (diagnosed by Nugent scoring) using a metaproteomics approach. With this comparison, we evaluated correlations between the mycobiome and STIs and BV. We hypothesized that fungal communities differ between optimal and non-optimal states, with individuals diagnosed with STIs and/or BV having a greater fungal diversity. Secondly, we aimed to investigate whether a correlation exists between the vaginal mycobiome and genital inflammation. We hypothesized that inflamed FGTs will compromise of taxonomically and functionally distinct fungal communities compared to non-inflamed FGTs. Thirdly, we aimed to analyse the functional contribution of the mycobiome to the vaginal microbiome. We hypothesize that fungi, as eukaryotes probably contribute towards unique metabolic features compared to bacterial taxa.

Objectives of this study

1) To optimize a software pipeline to adequately analyse fungal data
2) To characterize which fungal species are present in the FGTs of a cohort of young women residing in Cape Town, SA
3) To functionally characterize fungal proteins present in the FGT mycobiome
4) To identify fungal genera, proteins, and functions associated with optimal and non-optimal states (*i.e.,* inflamed, BV, and STIs)
5) To determine which specific intrinsic and extrinsic factors could account for the differences in the mycobiomes of different females (including body mass index (BMI), hormonal contraceptive methods, and vaginal pH)
6) To identify fungal proteins associated with disease

Chapter 2

Refining a methodology

Introduction

A crucial step in metaproteomics is selecting the appropriate protein database for mass spectra identification. The selection of an appropriate database can significantly improve taxonomic and functional annotation results (Tanca *et al.* 2016). However, as mycobiome research is relatively new in the metaproteomics field, many fungal species cannot be classified due to the scarcity of fungi in sequence databases (Bradford and Ravel 2017; Drell *et al.* 2013). Fungal sequence databases are still in their infancy for both DNA and protein sequences. Accordingly, bioinformatic methods are not optimized for fungi, and improved methods are needed to increase the sensitivity of analyses to obtain the maximum number of fungal hits during taxonomic classification (Nilsson *et al.* 2006).

Several approaches have been proposed to increase the sensitivity of metaproteomic analyses. The first approach includes the use of more than one database and various annotation tools (Tanca *et al*. 2016). The second approach includes assembling a "pseudo-metaproteome", which is a database comprising of all publicly available sequences for microbial species that are predicted to form part of the microbiota of interest (Callister *et al.* 2010; Verberkmoes *et al.* 2009). However, there are several challenges to consider when choosing a database. Firstly, the human microbiome can comprise thousands of various microorganisms that differ significantly between individuals. Secondly, large sequence databases increase the false discovery rate (FDR), impairing the protein identification rate (Muth *et al.* 2015; Vaudel *et al.* 2011). Thus, limiting the number of protein sequences and taxa in a database is essential for good quality results.

Strategies to reduce the size of protein databases to increase specificity have been proposed. One approach entails an initial search conducted using a complex protein database, and a subsequent search conducted using a condensed protein database constructed from the unfiltered output of protein hits from the initial search. Thereafter, peptides identified from the second search are then subjected to FDR filtering (Jagtap *et al.* 2013). A second approach

involves an initial search utilizing the complete UniProtKB database; however, the second search is now conducted using genomes of taxa instead of just proteins that had a hit during the first search (Tanca *et al.* 2014).

Another useful approach is Metanovo, a novel open-source software pipeline that integrates existing tools (DeNovoGUI, DirecTag, PeptideMapper, and X!Tandem) with a custom algorithm to generate targeted protein sequence databases for mass spectrometry analysis. Metanovo works by mapping de novo sequence tags to a very large protein sequence database, generating a compact database prior to the target-decoy search. This pipeline is used to directly estimate the proteins and species present in complex mass spectrometry samples at the level of expressed proteomes in a FDR-controlled manner (Potgieter *et al.* 2019).

The pipeline performs three main steps based on the abovementioned open-source tools, generating de novo sequence tags (partial sequences from a tandem mass spectrum), mapping the tags to a protein sequence database, and probabilistic protein ranking and filtering based on estimated species and protein abundance (Potgieter *et al.* 2019). To elaborate, Metanovo matches raw spectra to database entries, and database matches are stored in an SQLite database. These matches are used to estimate protein abundance using sequence tag matches, accounting for protein length using a normalized spectral abundance factor estimation for each protein in each sample. The algorithm Metanovo uses ranks the data based on the protein abundance level estimation, generating a list of protein identifiers that can explain all the database matches (such that each spectral match maps to at least one protein in the non-redundant list). Representation of taxa in the generated list is calculated using the UniProt FASTA header "OS" entry, and a score for each organism is obtained. Database proteins are re-ranked based on the combined scores for spectral and organism abundance, and the database is exported. To complete the above steps, the pipeline requires raw mass spectrometry data in MGF format and a UniProt FASTA file to search as input.

Metanovo allows the characterization of complex mass spectrometry data with an extremely large search space to be possible. As a result, it allows improved accuracy of current metaproteomic analysis methods. An advantage of this method is the ability to search large databases, and identify potential pathogens and biomarkers that may be overlooked using limited or hand-selected databases (Potgieter *et al.* 2019).

A systematic evaluation of existing methods to increase peptide spectrum match (PSM) sensitivity in metaproteomics has not been conducted. Thus, this chapter aimed to perform a systematic comparison between different protein sequence databases and to select the most suitable database for FGT mycobiome research. Thus, our requirements for selecting a database for downstream analyses were detecting as many fungal proteins as possible, and of these fungal proteins, we wanted a large number to belong to the genus *Candida*.

Methods

Study participants

A total of 148 black, South African females were recruited as part of the Women's Initiative in Sexual Health (WISH) study; from a poverty-stricken, resource-poor, high-population-dense community in Cape Town (Barnabas *et al.* 2017). This community is known to have a large adolescent population with a high incidence of HIV and STIs. The aim of this study was to try understand why South African adolescent women are at such a high risk of HIV.

This sub-study included 123 of the women recruited in the main study between the ages of 16 and 22 years. Approval was obtained for the study from the Human Research Ethics Committee of the University of Cape Town (HREC REF #267/2013 and #121/2020). All participants older than 18 years of age provided written informed consent, while assent and parental consent were obtained for participants younger than 18 years of age. Females were enrolled in the main study if they were HIV-negative, in general good health, not pregnant or menstruating at the time of sampling, and not had unprotected sex or douched in the last 48 hours. Additional exclusion criteria included the use of antibiotics in the previous two weeks. Patient-related metadata was collected and recorded (Barnabas *et al.* 2017). Metadata included information about age, body mass index (BMI), and hormonal contraception (HC), cytokine measurements, STIs, and BV status.

Sample collection and processing

As previously described in Barnabas *et al.* (2017), one lateral vaginal wall swab was rolled onto a glass slide and Gram-stained for microscopy to evaluate BV status using Nugent's

criteria. Patients were classified as BV negative if Nugent scores ranged between 0 and 3, BV intermediate if scores ranged between 4 and 6, and BV positive if Nugent scores were between 7 and 10. Vaginal pH was measured from swabs using color-fixed indicator strips (Macherey-Nagel, Düren, Germany) (Barnabas *et al.* 2017). Swab eluants were analysed by PCR to identify women with STIs (*i.e., Chlamydia trachomatis*, *Neisseria gonorrhoeae*, *Trichomonas vaginalis*, *Mycoplasma genitalium*). To identify women with Herpes simplex virus 2 serum was collected for serology. Cytokines and chemokines were measured from MC secretions to identify the inflammation state of each participant (Dabee *et al.* 2019). Patient characteristics and metadata for this study can be found in supplementary table 1.

Table 2.1 Prevalence of diseased states in the sample population at visit 1 and visit 2.

	Visit 1 (*n* = 113) (%)	**Visit 2 (*n* = 89) (%)**	**In total (*n* = 202) (%)**	
Bacterial Vaginosis (BV) Status				
BV Positive	50.4	41.6	46.5	
BV Intermediate	6.2	5.6	5.9	
BV Negative	41.6	42.7	42.1	
No diagnosis done	1.8	10.1	5.5	
Sexually Transmitted Infections (STIs)				**= 1 STI**
>= 1 STI	42.5	29.2	36.6	-
N. gonorrhoeae	5.3	3.4	5.9	2.0
M. genitalium	6.2	6.7	6.4	2.0
T. vaginalis	5.3	3.4	5.9	3.0
C. trachomatis	32.7	18.0	26.2	18.3
Herpes Simplex Virus 2	3.5	3.4	3.5	1.0
No diagnosis done	0.9	5.6	3.0	
Inflammation Status				
High	24.8	38.2	30.1	
Moderate	43.4	31.5	38.1	

Low	31.9	24.7	28.7	
Not done	0	5.6	2.5	

Mass Spectrometry

For shotgun LC-MS/MS, lateral vaginal wall swabs were collected. Sample names were computationally randomized for experimental design, to have an equal representation of different comparison groups across batches. Each swab sample was eluted and clarified by centrifugation. The protein content of each supernatant was determined by the bicinchoninic acid assay (BCA). Equal protein amounts (100µg) were denatured with urea exchange buffer, filtered, reduced with dithiothreitol, alkylated with iodoacetamide, and washed with hydroxyethyl piperazineethanesulfonic acid (HEPES). Acetone precipitation/formic acid solubilization was added to further clean up samples. The samples were then incubated overnight with trypsin, and the peptides were eluted with HEPES and dried via vacuum centrifugation. Reversed-phase liquid chromatography was used for desalting using a step-function gradient. The eluted fractions from this step were dried via vacuum centrifugation. LC-MS/MS analysis was conducted on a Q-Exactive Quadrupole-Orbitrap MS coupled with a Dionex UltiMate 3000 nano-UPLC system (120 min per sample). The solvent system included solvent A: 0.1% formic acid (FA) in LC grade water, and solvent B: 0.1% FA in acetonitrile. The MS was operated in positive ion mode with a capillary temperature of 320 °C, and 1.95 kV electrospray voltage was applied. A more detailed MS/MS method was previously described in Alisotani *et al.* (2020).

Fungal database creation

Two manually curated fungal databases were created to search raw MS files against using the quantitative proteomics software package, MaxQuant, in order to increase fungal assignments. The Pan Proteome database was curated by downloading fungal pan proteome sequences from UniProtKB. Pan proteomes include the full set of proteins thought to be expressed by a group of highly related organisms (*e.g.,* multiple strains of the same related species). Pan proteomes provide a representative set of all the protein sequences within a taxonomic group and capture unique sequences not found in the group's reference proteome, encompassing all non-redundant (nr) proteins. The second database, the Proteome database, was curated using fungal proteome sequences from UniProtKB. Proteomes differ from pan proteomes as

they include the full set of proteins thought to be expressed by an organism that has a full sequenced genome. The fungal species that were chosen for the curated proteome and pan proteome database included those found to be present in vaginal research studies and opportunistic fungi that showed the ability to infect body sites in close proximity to the vagina (*i.e*., gut) (Table S2).

Two databases were also created using the Metanovo pipeline (Potgieter *et al.* 2019). For the first database, the Metanovo pipeline was used to filter the complete UniProtKB database (August 2017 release) for protein identifications with default settings (**Universal Metanovo DB**). For the second database, the Metanovo pipeline was used on a UniProtKB database that was filtered for human and microbial entries (**Host Metanovo DB**).

To account for FDR, the Host Metanovo database (encompassing human and bacterial proteins) was concatenated to both the fungal Pan Proteome (**Pan Proteome + Host Metanovo DB**) and Proteome fungal databases (**Proteome + Host Metanovo DB)**. The Host Metanovo database was selected as it outperformed the Universal Metanovo database in terms of MS/MS identifications, protein group/peptide assignments, fungal assignments, and had a lower number of contaminants.

The high-quality, curated, nr Swiss-Prot database (June 2020 release) was used to make comparisons to observe the effects of a full database. Swiss-Prot is nr as it contains a consensus sequence for each distinct protein, and known variants have been collapsed into a single entry. As nr databases encompass consensus sequences, the size of the database becomes smaller, allowing a faster search time and more concise results. The majority of the database contained bacteria, with 6% compromising of fungal sequences.

Data processing

Raw M/S files were processed with MaxQuant version 1.6.3.4 on Linux. The detailed parameters are provided in Supplementary Table 3. In summary, methionine oxidation and acetylation of the protein N-terminal amino acid were considered as variable modifications and carbamidomethyl (C) as a fixed modification. The digestion enzyme selected was trypsin with a maximum of two missed cleavages. Stochasticity between independent LC−MS/MS runs is a challenging problem in the field of proteomics, resulting in significant missing values among

observed peptides. To address this issue, the ‘match-between-runs’ function was selected for raw spectra processing.

Venn diagrams were produced using the online Venn diagram plotting tool Venny 2.1.0 to identify protein groups and peptide sequences shared between databases (Oliveros 2016).

Taxonomy was assigned to protein groups and peptides using UniProt and Unipept, respectively. To understand the differences in taxonomic information attainable with each database, we distinguished between taxa only present when using a specific database (*i.e.,* “qualitatively differential taxa”) and taxa that appeared across all databases (*i.e.*, “quantitatively differential taxa”).

Results

The main aim of this Chapter was to perform a systematic comparison of different protein databases that can be used in a metaproteomic analysis. For this purpose, we chose to conduct a preliminary investigation of the impact of several constructed databases using 202 raw M/S files. In total, we generated four databases. Two were generated using the Metanovo software filtered through UniProtKB and the other two databases were created by first generating "pseudo-metaproteomes" containing FGT fungal species. The "pseudo-metaproteomes", included 97 fungal proteomes and 37 fungal pan proteomes (Table S1) downloaded from UniProt and concatenated to a Metanovo database.

Protein and peptide identifications

12,299,698 MS/MS were submitted to MaxQuant for analysis of the different generated protein databases (Table 2.2). An FDR cut-off of 0.01 was used for both peptides and proteins.

The **Pan Proteome + Host Metanovo DB** resulted in 8.19% MS/MS identifications. This corresponded to the identification of 3,865 protein groups and 18,383 peptides. 18.2% (3333) were missed cleavages (Table 2.2; Fig 2.2). Data was of relatively good quality as indicated by a low percentage of contaminants (1.2% of total protein groups). This database identified 48 reverse peptides, 578 contaminant peptides, 9,268 non-reference peptides, and 8,489

reference peptides. The Posterior Error Probability (PEP) score is the probability that the observed Peptide Spectrum Match (PSM) is incorrect. After removing potential contaminants and reverse peptides, the mean PEP score was 0.0065, and the highest PEP score was 0.058 (Fig 2.1; Fig 2.2).

From the **Proteome + Host Metanovo DB** 5.81% of MS/MS were identified, corresponding to a total of 14,139 peptides and 3,133 total protein groups (Table 2.2; Fig 2.2). Contaminants made up 1.3% of protein groups, and 16% (2278) of peptide sequences were missed cleavages (Table 2.2). PEP Scores for all peptides < 0.048. However, this database had the highest mean PEP Score (0.007) (Fig 2.1; Fig 2.2).

The **Host Metanovo DB** had the highest MS/MS identification rate (9.9%) resulting in the identification of 21,400 peptides, and 4,353 protein groups (Table 2.2; Fig 2.2). A total of 1.2% of total protein groups were contaminants and 17.74% missed cleavages (Table 2.2). The Mean PEP score for this database was 0.00625, with the highest maximum score of 0.0590 (Fig 2.1; Fig 2.2).

The **Universal Metanovo DB** had an MS/MS identification rate of 9.44%, resulting in the identification of 23,224 total peptides, and 5,056 protein groups (Table 2.2; Fig 2.2). This database had the lowest percentage of contaminants (0.89% of total protein groups), and 17.50% missed cleavages (Table 2.2). This database had the lowest mean PEP score (PEP = 0.00619), with a max PEP score of 0.0580 (Fig 2.1; Fig 2.2).

The **Swiss-Prot DB** identified the lowest number of MS/MS assignments, protein groups, and peptide sequences. A total of 2.32% MS/MS, 1,134 protein groups, and 4,955 peptides were identified, respectively. This database also showed the highest number of contaminant protein groups (3.1%) and missed cleavages (21.4%) (Table 2.2).

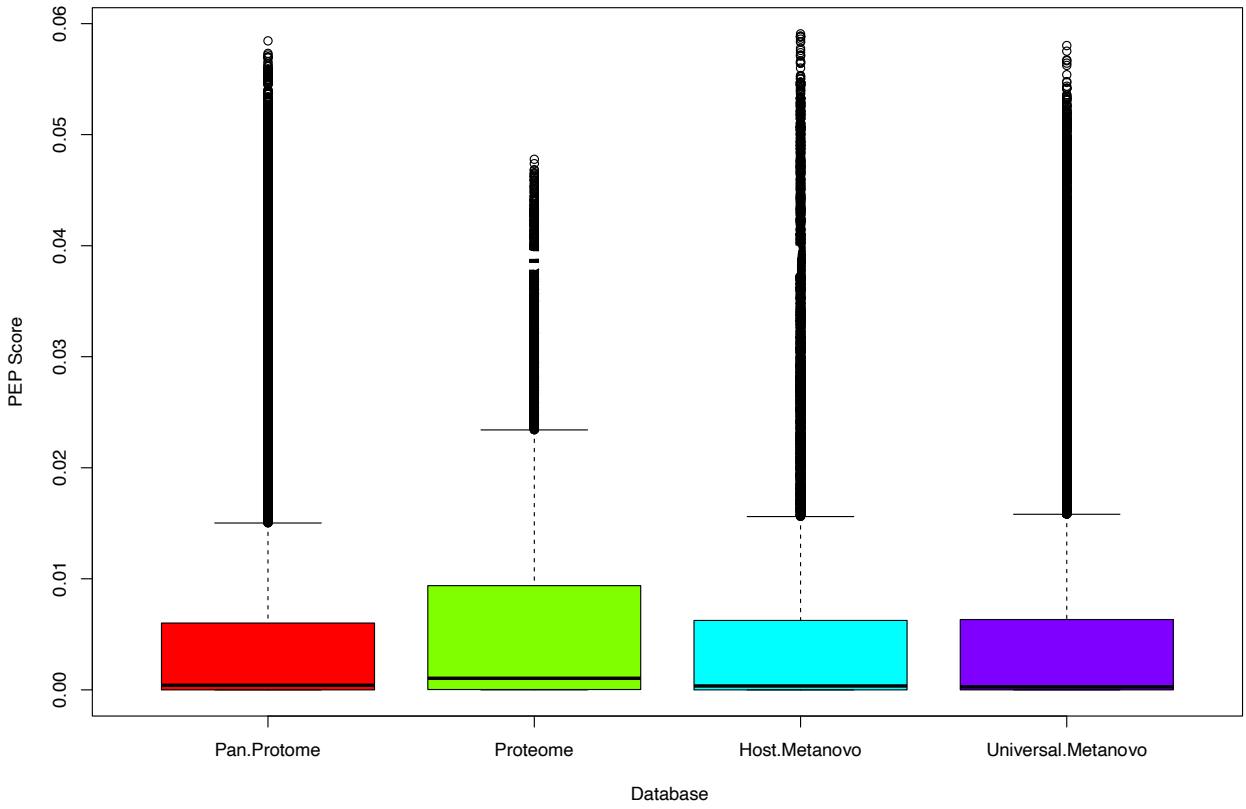


Figure 2.1. Posterior Error Probability (PEP score) across protein sequences databases searched in Maxqaunt. Universal Metanovo DB had the lowest mean PEP score. The Proteome + Host Metanovo DB had the lowest max PEP score, and the Host Metanovo DB had the highest PEP score.

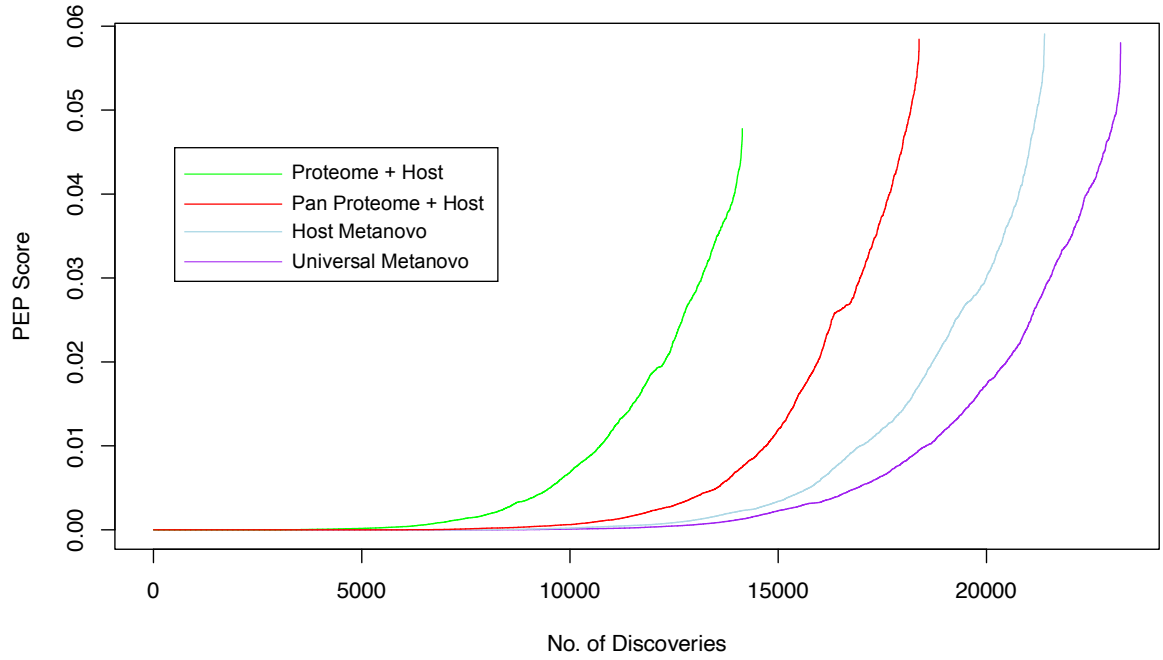


Figure 2.2. Posterior Error Probability (PEP score) plotted against number of peptide sequence discoveries across protein sequences databases searched in Maxqaunt. The Proteome + Host Metanovo DB had the lowest number of discoveries and lowest PEP Score range (Green). The Host Metanovo DB had the highest PEP Score (Light Blue). The Universal Metanovo DB had the highest number of peptide discoveries (Purple). The Pan Proteome + Host Metanovo DB had a comparable PEP Score range in comparison to the Host Metanovo DB and Universal Metanovo DB Score (Red).

Table 2.2. Overview of MS/MS spectra, peptide, and protein group identifications using different protein sequence database (DB) searches on the WISH data in Maxquant.

Database	MS/MS identified	Peptides Identified	Protein Groups	FDR	Contaminants	Missed Cleavages (%)
Pan Proteome + Host Metanovo DB	8.19%	18383	3865	0.01	1.2%	18.2 %
Proteome + Host Metanovo DB	5.81%	14139	3133	0.01	1.3%	16%
Swiss-Prot DB	2.32%	4955	1134	0.01	3.1%	21.4%
Host Metanovo DB	9.9%	21400	4353	0.01	1.2%	17.74%
Universal Metanovo DB	9.44%	23224	4885	0.01	0.89%	17.50%

Abbreviations: False discovery rate (FDR); Tandem mass spectrometry (MS/MS)

Taxonomic comparisons

The **Pan Proteome + Host Metanovo DB** shared 2722, 3400, and 2168 protein groups in common with the **Proteome + Host Metanovo DB**, **Host Metanovo DB**, and **Universal Metanovo DB**, respectively (Fig 2.3). Between all four databases, 1769 protein groups were shared in common. The **Pan Proteome + Host Metanovo DB** yielded 127 unique protein groups, in comparison to the other databases. When compared exclusively to the **Host Metanovo DB**, this database yielded 169 unique protein groups (Fig 2.3).

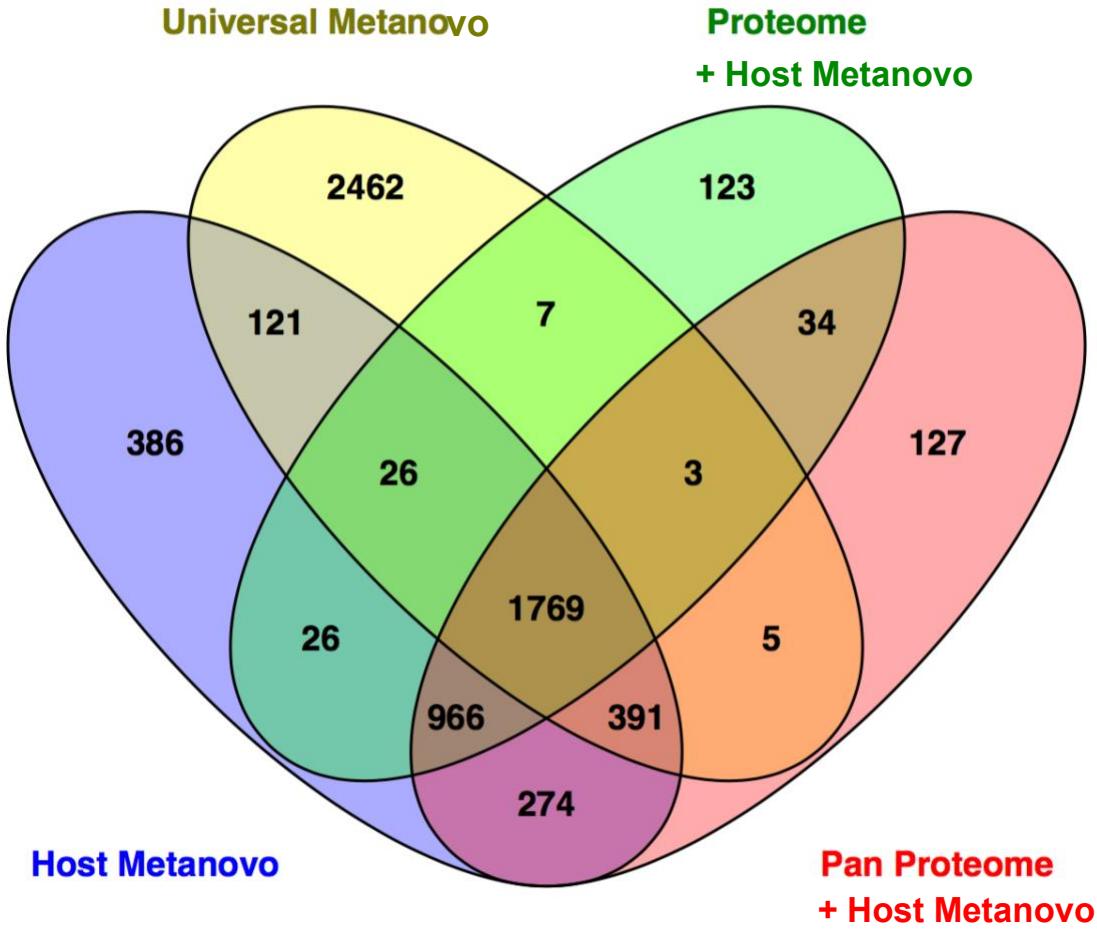


Figure 2.3. Venn diagram of protein groups identified from the four different database searches. 1769 protein groups were shared between protein sequence databases. The Pan proteome + Host Metanovo DB shared majority of protein groups with other databases, yielding only 127 protein groups.

After the removal of contaminants and reverse protein groups from the **Pan Proteome + Host Metanovo DB** dataset, of the protein groups that remained majority were bacterial (45.29%),

37.92% were mammalian, 4.31% fungal, 0.23% archaeal, and 0.17% of viral origin (Table 2.3). The **Pan Proteome + Host Metanovo DB** resulted in a 1.24% increase in fungal protein group assignments in comparison with the **Host Metanovo DB**.

Table 2.3. Abundance (%) of protein groups assigned to taxonomic kingdoms using each protein sequence databases (DBs) on the WISH dataset.

	Pan Proteome + Host Metanovo DB	**Proteome + Host Metanovo DB**	**Host Metanovo DB**	**Universal Metanovo DB**
Protein groups identified (n)	3502	2911	3904	4788
Fungi (%)	4.31	3.78	3.07	1.50
Bacteria (%)	45.29	42.87	48.26	32.83
Mammalia (%)	37.92	40.88	35.81	40.7
Archaea (%)	0.23	0.21	0.36	0.19
Viruses (%)	0.17	0.14	0.15	0.10

With regard to peptide sequences, 13415 of peptide sequences were shared between all four protein databases. When comparing the **Host Metanovo DB** and **Pan Proteome + Host Metanovo DB**, 84.96% (18180) of peptides were shared. Thus, the **Pan Proteome + Host Metanovo DB** yielded 205 unique peptide assignments (Fig 2.4).

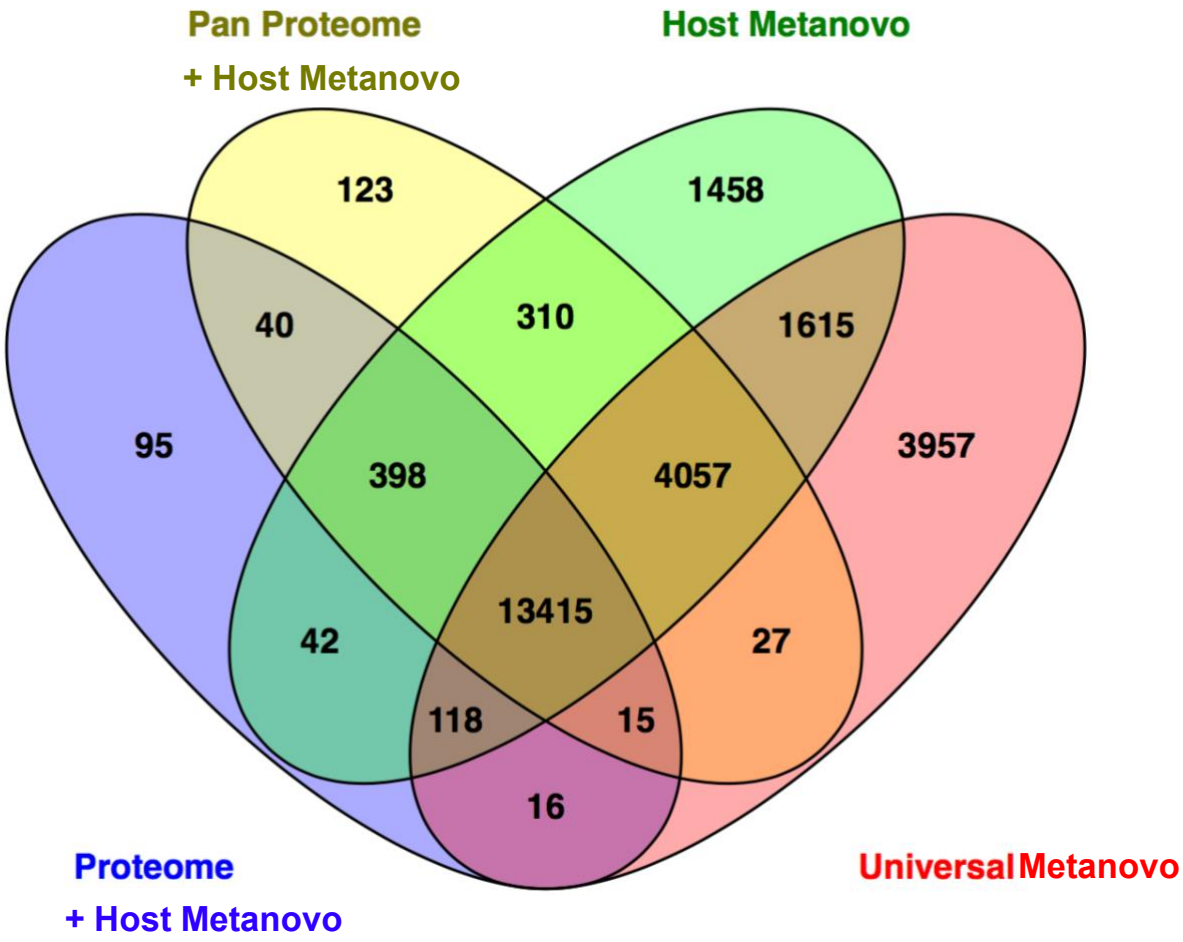


Figure 2.4. Venn diagram of peptide sequences identified from the four different database searches. Majority of peptide sequences (13145) were shared in common. The Pan Proteome + Host Metanovo db had a high percentage of identified peptides sequences shared, yielding only 123 unique peptide sequences.

According to Unipept 95 peptides were assigned to a fungal taxonomic lineage lowest common ancestor using the **Pan Proteome + Host Metanovo DB**. From the **Proteome + Host Metanovo DB** Unipept identified 67 fungal peptides, 88 peptides with **Host Metanovo**, and 64 peptides using the **Universal Metanovo database**.

Both Metanovo databases showed the lowest number of protein groups assigned to *Candida* spp. (Fig 2.5). The curated databases allowed for an increase in *Candida* assignments, with the **Pan Proteome + Host Metanovo DB** identifying the largest number of *Candida* spp. (*n* = 35). Notably this database also identified *3 C. albicans* proteins in comparison to both Metanovo databases which showed no protein hits to *C. albicans* (Fig 2.5).

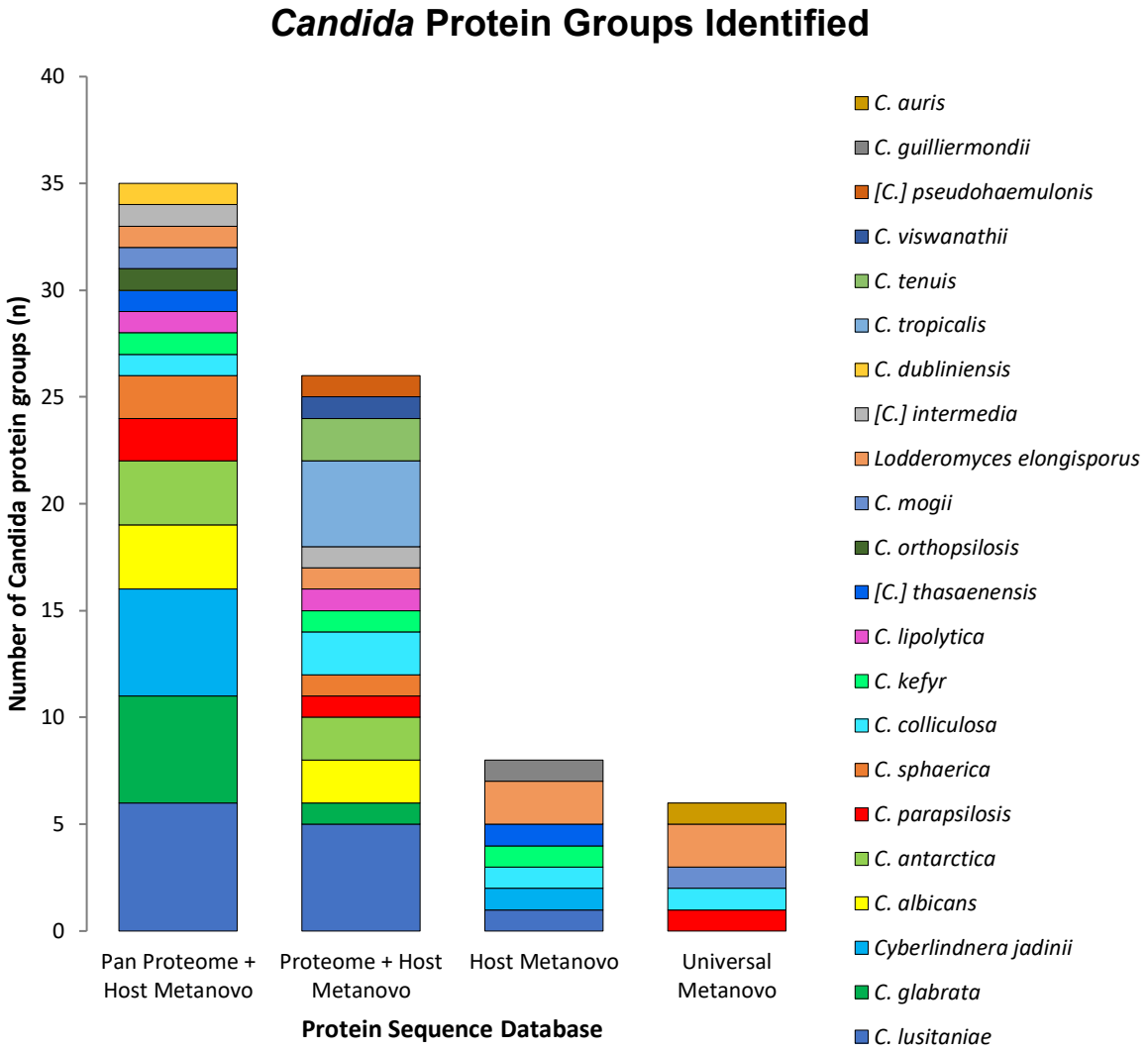


Figure 2.5. Number of protein groups that were assigned to the genera *Candida* using different protein sequence databases (DBs) searched in Maxquant on the WISH dataset. Taxonomic analysis at the *Candida* level revealed that the **Pan Proteome + Host Metanovo DB** yielded more *Candida* assignments in comparison to other DBS (28 more *Candida* assignments in comparison to **Host Metanovo DB**). Metanovo databases showed no hits to *Candida albicans* proteins, whereas the **Pan Proteome + Host Metanovo DB** and **Proteome + Host Metanovo DB** showed hits to *C. albicans* proteins.

Discussion

A primary aim of the work described in this thesis was to gain taxonomic and functional insight into the FGT mycobiome based on a metaproteomic analysis of human vaginal swab samples. Few studies have utilized metaproteomics to study the mycobiome, and as of 2021, none of

them have analysed the FGT mycobiome. As a result, many of the challenges this method faces have not been addressed.

One specific challenge in metaproteomics is the need for an adequate protein sequence database that sufficiently represents the diversity of proteins present in samples. For this purpose, the appropriate targeted protein sequence databases first had to be constructed.

The nr **Swiss-Prot DB** showed the lowest MS/MS identification rate. Two reasons explain this result. Firstly, Swiss-Prot is a smaller database dominated by bacterial sequences, and MS/MS can only assign to species present in the database. Secondly, if a lowly abundant protein is present in the MS/MS dataset and only represented by a few spectra, there is a high possibility that these critical sequences are absent from the nr database. A suggested approach to combat this issue is utilizing a complete, non-identical database, where every known protein sequence is represented, such as NCBI-nr or UniRef100. However, this method also has its disadvantages because as the database becomes larger, the search time will increase, and protein identification becomes more difficult. This is because the chances of the experimental masses randomly matching theoretical peptide masses increases, in so decreasing the confidence of protein identifications (Cottrell 2011).

The **Proteome + Host Metanovo DB** had the lowest range of PEP scores; this is expected since this database identified the least number of MS/MS. Therefore, the chances of identifying an incorrect PSM are lower. By this reasoning, the **Host Metanovo DB** had the highest maximum PEP score, as it identified the largest number of MS/MS. Therefore, the chances of this database identifying an incorrect PSM were higher. However, the **Pan Proteome + Host Metanovo DB**, **Host Metanovo DB**, and **Universal Metanovo DB** share similar mean PEP Scores for PSMs. These results indicate a set of reliable PSMs when using these databases, and only 1% of the data had PEP scores above 5%.

Of the total set of identified peptide sequences, 84.96% were shared between the **Host Metanovo DB** and the **Pan Proteome + Host Metanovo DB**. This percentage indicates a sensitive and specific identification rate using this method. The complementary nature of the results retrieved from each database, as well as the bias present in specific databases, implies that using several databases can result in a more comprehensive and unbiased picture of the function of microbial communities. However, we also observed that searching against different

databases produces produced varying fungal results, thus selecting a suitable database consequently can significantly strengthen taxonomic and functional fungal annotation of the metaproteome.

The Metanovo databases were not suitable for fungal assignments, as they produced few *Candida* hits. Furthermore, none of the *Candida* hits were *C. albicans*, which is widely known to be highly prevalent in the FGT (Drell *et al.* 2013). With the creation of a suitable fungal proteome database, it was possible to identify more fungal and *Candida* peptide sequences and protein groups. Thus, our requirements for selecting a database for downstream analyses was firstly detecting as many fungal proteins as possible, and secondly of these fungal proteins we wanted a large number to belong to the genus *Candida*.

Our findings from the **Pan Proteome + Host Metanovo DB** demonstrate the advantage of implementing a fungal-based sequence selection method to reduce the size of the database, as size-related FDR estimation problems are reduced (Tanca *et al.* 2016). This approach can be applied in cases where extensive fungal protein coverage is needed. However, the disadvantage of this database is that the MS/MS identification rate was reduced compared to the Metanovo database on its own. This is expected with a larger database as sensitivity will decrease. Taking the above results into account, the output from the **Pan Proteome + Host Metanovo DB** was selected for downstream analyses.

Conclusion

Metaproteomics is a young and developing field of research. Despite the progress that has been made in metaproteomic analysis approaches, major challenges still hinder this field. Thus, improvements to data analysis methods are needed, especially in the case of mycobiome research. The following search strategy to increase the number of identified fungal peptides is proposed: to use a selected collection of fungal pan proteomes known to be present in the microbiome of interest concatenated to a Metanovo fasta file as a database. This approach targets the problem of impaired peptide identification when the search sequence space becomes extremely large. We can conclude that the quality of fungal annotations depends on having a comprehensive, well-annotated database of fungal sequences.

Since there is still a large percentage of MS/MS that remains unidentified, and the majority of taxonomic data remains unexplored. Future work should try implementing the following approaches to increase identification yields using databases: 1) Multiple parallel searches rather than merging different types of sequences into a single database; 2) Using metagenome databases rather than UniProt-based databases (Tanca *et al.* 2016). However, since databases only allow us to uncover fungi that have already been identified and cultured, more research effort should be directed towards culturing more fungi in the laboratory, and the use of next generation sequencing to identify more fungi.

Supplementary work

Table S1. Clinical characteristics of patients from which vaginal swabs were collected that formed part of the WISH dataset.

PATIENT ID	VISIT	STI STATUS	BV STATUS	BV SCORE	INFLAMMATION CATEGORY	HC	AGE	PH	PH CATEGORY	BMI	BMI CATEGORY
100V2	1	+	-	0	LOW	NUR ISTERATE	17	4.4	NORMAL	17.90	UNDERWEIGHT
101V2	1	-	-	0	MED	DMPA	22	4.1	NORMAL	35.11	OBESE
102V2	1	+	-	3	HIGH	NUR ISTERATE	17	5	HIGH	31.56	OBESE
104V2	1	-	+	7	MED	NUR ISTERATE	21	NA	NA	20.55	HEALTHY
106V2	1	-	-	0	LOW	NUR ISTERATE	17	3.6	LOW	25.39	OVERWEIGHT
107V2	1	+	+	9	HIGH	NUR ISTERATE	19	6.1	HIGH	36.16	OBESE
108V2	1	+	+	10	MED	DMPA	16	3.6	LOW	21.05	HEALTHY
110V2	1	+	+	9	MED	IMPLANON	17	4.7	HIGH	25.22	OVERWEIGHT
10V2	1	+	+/-	5	HIGH	NUR ISTERATE	17	4.7	HIGH	26.04	OVERWEIGHT
114V2	1	+	+	10	MED	IMPLANON	20	5.6	HIGH	30.47	OBESE
116V2	1	-	+/-	5	MED	NUR ISTERATE	17	4.7	HIGH	19.15	HEALTHY
117V2	1	-	-	0	LOW	NUR ISTERATE	16	4.4	NORMAL	22.32	HEALTHY
118V2	1	+	+	9	LOW	NUR ISTERATE	20	4.4	NORMAL	25.65	OVERWEIGHT
12V2	1	-	+	10	HIGH	NUR ISTERATE	18	5.3	HIGH	32.08	OBESE
120V2	1	+	-	0	MED	NUR ISTERATE	19	4.4	NORMAL	25.15	OVERWEIGHT
121V2	1	-	+	7	HIGH	NUR ISTERATE	19	5.6	HIGH	21.45	HEALTHY
123V2	1	-	-	0	LOW	DMPA	19	NA	NA	22.49	HEALTHY
125V2	1	-	-	0	LOW	NUR ISTERATE	20	4.1	NORMAL	18.13	UNDERWEIGHT

126V2	1	-	-	3	HIGH	DMPA	20	5	HIGH	25.22	OVERWEIGHT
128V2	1	-	-	0	LOW	NUR ISTERATE	18	4.1	NORMAL	25.88	OVERWEIGHT
129V2	1	-	-	0	LOW	NUR ISTERATE	18	4.4	NORMAL	28.36	OVERWEIGHT
130V2	1	-	+	9	HIGH	NUR ISTERATE	16	5	HIGH	18.61	HEALTHY
131V2	1	+	+	9	MED	IMPLANON	18	5	HIGH	27.25	OVERWEIGHT
132V2	1	+	+	9	LOW	NUR ISTERATE	18	4.1	NORMAL	21.88	HEALTHY
135V2	1	-	-	0	LOW	NUR ISTERATE	18	5.3	HIGH	27.99	OVERWEIGHT
138V2	1	-	+	10	MED	NUR ISTERATE	22	5.3	HIGH	27.34	OVERWEIGHT
139V2	1	+	+	9	MED	IMPLANON	18	5	HIGH	NA	NA
13V2	1	-	+	9	MED	NUR ISTERATE	19	4.7	HIGH	33.06	OBESE
141V2	1	-	+	9	LOW	OCP	18	NA	NA	25.39	OVERWEIGHT
148V2	1	-	+	8	HIGH	NUR ISTERATE	18	4.7	HIGH	32.47	OBESE
149V2	1	-	NA	NA	LOW	NUR ISTERATE	18	4.1	NORMAL	28.44	OVERWEIGHT
150V2	1	-	+	9	MED	DMPA	19	4.7	HIGH	23.73	HEALTHY
154V2	1	-	+	10	MED	NUR ISTERATE	22	NA	NA	36.79	OBESE
156V2	1	+	+	7	HIGH	IMPLANON	16	4.7	HIGH	24.01	HEALTHY
158V2	1	-	-	0	LOW	NUR ISTERATE	17	4.7	HIGH	26.77	OVERWEIGHT
159V2	1	-	+	9	LOW	NUR ISTERATE	18	5	HIGH	19.57	HEALTHY
15V2	1	-	+	8	MED	NUR ISTERATE	20	4.7	HIGH	30.08	OBESE
161V2	1	-	-	0	LOW	DMPA	19	4.4	NORMAL	21.93	HEALTHY
164V2	1	+	+	8	MED	DMPA	19	5.3	HIGH	30.11	OBESE
168V2	1	+	+	9	HIGH	IMPLANON	22	5.3	HIGH	NA	NA
171V2	1	NA	NA	NA	HIGH	NUR ISTERATE	21	5.6	HIGH	28.69	OVERWEIGHT
172V2	1	+	-	0	MED	NUR ISTERATE	19	4.7	HIGH	20.45	HEALTHY

17V2	1	+	-	0	MED	NUR ISTERATE	18	4.4	NORMAL	23.68	HEALTHY
1V2	1	-	+	8	MED	DMPA	18	5.3	HIGH	27.48	OVERWEIGHT
21V2	1	+	-	0	LOW	DMPA	18	4.1	NORMAL	27.36	OVERWEIGHT
22V2	1	+	-	0	LOW	NUR ISTERATE	19	4.4	NORMAL	26.22	OVERWEIGHT
23V2	1	+	-	0	MED	NUR ISTERATE	18	3.6	LOW	20.08	HEALTHY
24V2	1	-	-	1	MED	NUR ISTERATE	19	4.1	NORMAL	22.31	HEALTHY
25V2	1	-	-	0	LOW	OCP	19	4.1	NORMAL	19.49	HEALTHY
26V2	1	-	+/-	6	LOW	DMPA	19	4.1	NORMAL	23.14	HEALTHY
27V2	1	+	-	0	LOW	DMPA	18	4.4	NORMAL	26.31	OVERWEIGHT
28V2	1	+	+	10	MED	NUR ISTERATE	18	4.4	NORMAL	21.72	HEALTHY
2V2	1	-	+	8	MED	NUR ISTERATE	22	4.7	HIGH	19.83	HEALTHY
30V2	1	-	+	9	MED	DMPA	21	5	HIGH	30.18	OBESE
31V2	1	+	-	0	MED	NUR ISTERATE	19	4.1	NORMAL	19.72	HEALTHY
32V2	1	+	+	7	HIGH	NUR ISTERATE	18	4.7	HIGH	22.77	HEALTHY
33V2	1	-	-	1	LOW	NUR ISTERATE	20	4.1	NORMAL	28.35	OVERWEIGHT
34V2	1	-	-	0	LOW	NUR ISTERATE	18	4.7	HIGH	NA	NA
35V2	1	+	+	7	HIGH	NUR ISTERATE	18	5.3	HIGH	23.42	HEALTHY
37V2	1	-	+	8	HIGH	NUR ISTERATE	18	5	HIGH	20.90	HEALTHY
39V2	1	+	-	1	MED	NUR ISTERATE	19	4.4	NORMAL	25.72	OVERWEIGHT
3V2	1	-	+	8	HIGH	OCP	19	5	HIGH	38.06	OBESE
40V2	1	+	-	1	MED	NUR ISTERATE	19	5.3	HIGH	19.47	HEALTHY
41V2	1	+	-	2	MED	NUR ISTERATE	16	4.1	NORMAL	26.49	OVERWEIGHT
43V2	1	+	+	8	MED	OCP	16	4.4	NORMAL	25.32	OVERWEIGHT
44V2	1	+	-	1	HIGH	NUR ISTERATE	16	4.7	HIGH	20.82	HEALTHY

45V2	1	+	-	0	LOW	NUR ISTERATE	17	4.4	NORMAL	22.83	HEALTHY
46V2	1	+	-	2	MED	NUR ISTERATE	20	4.4	NORMAL	35.11	OBESE
47V2	1	+	+	8	MED	OCP	21	4.7	HIGH	38.10	OBESE
48V2	1	+	+	10	HIGH	NUR ISTERATE	18	5	HIGH	25.71	OVERWEIGHT
4V2	1	+	-	0	MED	NUR ISTERATE	19	5	HIGH	NA	NA
51V2	1	-	-	0	MED	DMPA	18	4.7	HIGH	29.28	OVERWEIGHT
52V2	1	+	+/-	5	HIGH	DMPA	16	5.6	HIGH	25.36	OVERWEIGHT
53V2	1	-	-	0	MED	NUR ISTERATE	21	4.1	NORMAL	32.42	OBESE
54V2	1	+	+	7	MED	NUR ISTERATE	18	4.7	HIGH	28.91	OVERWEIGHT
55V2	1	-	-	0	MED	NUR ISTERATE	18	4.7	HIGH	27.92	OVERWEIGHT
56V2	1	-	+	8	LOW	NUR ISTERATE	20	4.1	NORMAL	20.83	HEALTHY
57V2	1	+	+	8	HIGH	NUR ISTERATE	18	4.7	HIGH	36.65	OBESE
59V2	1	-	-	2	MED	NUR ISTERATE	17	5	HIGH	21.40	HEALTHY
5V2	1	-	+/-	4	HIGH	NUR ISTERATE	20	6.1	HIGH	19.04	HEALTHY
60V2	1	-	-	2	MED	NUR ISTERATE	17	5	HIGH	18.26	UNDERWEIGHT
61V2	1	+	-	3	LOW	NUR ISTERATE	17	4.4	NORMAL	28.40	OVERWEIGHT
62V2	1	-	+	8	LOW	NUR ISTERATE	17	4.7	HIGH	30.43	OBESE
65V2	1	-	-	3	HIGH	NUVARING	17	5	HIGH	25.15	OVERWEIGHT
66V2	1	-	-	1	MED	NUR ISTERATE	19	4.1	NORMAL	24.39	HEALTHY
67V2	1	+	+	8	MED	NUR ISTERATE	17	4.4	NORMAL	19.78	HEALTHY
68V2	1	-	+	7	HIGH	NUR ISTERATE	16	5	HIGH	28.88	OVERWEIGHT
6V2	1	-	-	1	HIGH	NUR ISTERATE	20	4.7	HIGH	22.77	HEALTHY
70V2	1	+	+/-	4	LOW	NUR ISTERATE	17	4.4	NORMAL	17.36	UNDERWEIGHT
71V2	1	-	-	0	LOW	DMPA	18	4.1	NORMAL	21.08	HEALTHY

72V2	1	-	+	8	MED	NUR ISTERATE	17	4.4	NORMAL	26.13	OVERWEIGHT
73V2	1	+	+	9	HIGH	NUR ISTERATE	17	5	HIGH	27.83	OVERWEIGHT
74V2	1	-	+	8	MED	DMPA	19	4.1	NORMAL	NA	NA
76V2	1	-	+	10	MED	NUR ISTERATE	17	4.4	NORMAL	16.97	UNDERWEIGHT
77V2	1	-	+	9	MED	DMPA	18	5	HIGH	28.96	OVERWEIGHT
79V2	1	-	-	0	LOW	NUR ISTERATE	20	4.1	NORMAL	26.95	OVERWEIGHT
7V2	1	+	-	0	LOW	NUR ISTERATE	19	4.1	NORMAL	36.11	OBESE
80V2	1	-	-	0	LOW	DMPA	22	4.1	NORMAL	24.61	HEALTHY
81V2	1	-	+	8	LOW	NUR ISTERATE	18	4.7	HIGH	27.05	OVERWEIGHT
82V2	1	-	+	7	MED	NUR ISTERATE	18	5	HIGH	20.69	HEALTHY
84V2	1	+	+	10	HIGH	NUR ISTERATE	17	5.3	HIGH	21.22	HEALTHY
85V2	1	+	+/-	4	HIGH	DMPA	22	6.1	HIGH	22.50	HEALTHY
86V2	1	+	-	0	LOW	NUR ISTERATE	20	4.1	NORMAL	NA	NA
87V2	1	+	+	10	HIGH	NUR ISTERATE	19	5.3	HIGH	34.85	OBESE
88V2	1	-	+	7	MED	NUR ISTERATE	20	6.1	HIGH	25.34	OVERWEIGHT
91V2	1	-	+	8	MED	NUR ISTERATE	21	4.4	NORMAL	24.65	HEALTHY
94V2	1	-	+	10	LOW	OCP	20	4.7	HIGH	29.28	OVERWEIGHT
95V2	1	-	-	0	LOW	DMPA	19	4.1	NORMAL	28.55	OVERWEIGHT
96V2	1	-	+	9	HIGH	IMPLANON	20	5	HIGH	21.78	HEALTHY
97V2	1	-	+	10	LOW	IMPLANON	20	5	HIGH	29.52	OVERWEIGHT
98V2	1	-	+	9	MED	NUR ISTERATE	19	5.3	HIGH	19.96	HEALTHY
99V2	1	-	+	9	MED	NUR ISTERATE	19	5	HIGH	22.58	HEALTHY
9V2	1	+	+	10	MED	NUR ISTERATE	19	5.6	HIGH	22.98	HEALTHY
100V3	2	-	-	0	LOW	NUR ISTERATE	17	4.1	NORMAL	17.80	UNDERWEIGHT

10V3	2	-	+	9	HIGH	NUR ISTERATE	17	5.7	HIGH	26.04	OVERWEIGHT
114V3	2	-	-	0	LOW	IMPLANON	20	5	HIGH	30.47	OBESE
115V3	2	-	+	10	MED	NUR ISTERATE	20	4.7	HIGH	24.84	HEALTHY
11V3	2	+	+	8	HIGH	NUR ISTERATE	17	5.3	HIGH	36.33	OBESE
120V3	2	+	NA	NA	MED	NUR ISTERATE	19	5	HIGH	25.15	OVERWEIGHT
121V3	2	-	+	7	MED	NUR ISTERATE	19	NA	NA	21.45	HEALTHY
123V3	2	+	-	2	LOW	DMPA	19	4.1	NORMAL	22.49	HEALTHY
125V3	2	-	+	9	MED	NUR ISTERATE	20	4.7	HIGH	18.13	UNDERWEIGHT
126V3	2	-	+	7	HIGH	DMPA	20	NA	NA	25.22	OVERWEIGHT
128V3	2	-	-	0	LOW	NUR ISTERATE	18	3.6	LOW	25.88	OVERWEIGHT
130V3	2	-	NA	NA	HIGH	NUR ISTERATE	16	NA	NA	18.61	HEALTHY
138V3	2	-	+	9	MED	NUR ISTERATE	22	5.3	HIGH	27.34	OVERWEIGHT
139V3	2	-	-	0	LOW	IMPLANON	18	4.1	NORMAL	NA	NA
13V3	2	-	+	9	MED	NUR ISTERATE	19	4.1	NORMAL	33.06	OBESE
148V3	2	-	NA	NA	HIGH	NUR ISTERATE	18	NA	NA	32.47	OBESE
149V3	2	-	-	0	LOW	NUR ISTERATE	18	4.7	HIGH	28.44	OVERWEIGHT
150V3	2	+	+	9	HIGH	DMPA	19	5	HIGH	23.73	HEALTHY
152V3	2	NA	NA	NA	HIGH	NUR ISTERATE	17	5	HIGH	25.91	OVERWEIGHT
158V3	2	-	-	0	LOW	NUR ISTERATE	17	4.1	NORMAL	26.77	OVERWEIGHT
159V3	2	-	+	9	MED	NUR ISTERATE	18	5	HIGH	19.57	HEALTHY
15V3	2	+	-	3	LOW	NUR ISTERATE	20	4.1	NORMAL	30.08	OBESE
16V3	2	+	+/-	4	HIGH	NUR ISTERATE	18	5.3	HIGH	25.22	OVERWEIGHT
19V3	2	-	+/-	4	HIGH	DMPA	18	5.3	HIGH	22.81	HEALTHY
1V3	2	-	+/-	6	HIGH	DMPA	18	5.1	HIGH	27.48	OVERWEIGHT

21V3	2	-	-	0	HIGH	DMPA	18	4.1	NORMAL	27.36	OVERWEIGHT
22V3	2	+	-	0	LOW	NUR ISTERATE	19	4.1	NORMAL	26.22	OVERWEIGHT
23V3	2	+	-	0	MED	NUR ISTERATE	18	4.1	NORMAL	20.08	HEALTHY
24V3	2	-	-	0	LOW	NUR ISTERATE	19	4.1	NORMAL	22.31	HEALTHY
25V3	2	-	-	0	LOW	OCP	19	3.6	LOW	19.49	HEALTHY
26V3	2	-	-	0	LOW	DMPA	19	4.7	HIGH	23.14	HEALTHY
28V3	2	-	+	9	MED	NUR ISTERATE	18	6.1	HIGH	21.72	HEALTHY
2V3	2	-	+	7	MED	NUR ISTERATE	22	5.3	HIGH	19.83	HEALTHY
30V3	2	-	+	9	MED	DMPA	21	5.3	HIGH	30.18	OBESE
31V3	2	+	-	0	HIGH	NUR ISTERATE	19	4.4	NORMAL	19.72	HEALTHY
32V3	2	-	-	1	HIGH	NUR ISTERATE	18	4.7	HIGH	22.77	HEALTHY
33V3	2	-	+	10	HIGH	NUR ISTERATE	20	5	HIGH	28.35	OVERWEIGHT
34V3	2	-	-	0	MED	NUR ISTERATE	18	4.7	HIGH	NA	NA
37V3	2	-	+	8	HIGH	NUR ISTERATE	18	5.6	HIGH	20.90	HEALTHY
38V3	2	-	+	9	HIGH	NUR ISTERATE	16	4.7	HIGH	20.06	HEALTHY
40V3	2	-	-	0	MED	NUR ISTERATE	19	4.4	NORMAL	19.47	HEALTHY
41V3	2	+	-	3	HIGH	NUR ISTERATE	16	5	HIGH	26.49	OVERWEIGHT
43V3	2	-	+	9	MED	OCP	16	4.4	NORMAL	25.32	OVERWEIGHT
44V3	2	+	-	1	MED	NUR ISTERATE	16	4.1	NORMAL	20.82	HEALTHY
45V3	2	-	-	0	LOW	NUR ISTERATE	17	4.1	NORMAL	22.83	HEALTHY
46V3	2	+	+	8	HIGH	NUR ISTERATE	20	5.3	HIGH	35.11	OBESE
47V3	2	-	+	10	MED	OCP	21	5.7	HIGH	38.10	OBESE
48V3	2	-	+	10	MED	NUR ISTERATE	18	5.3	HIGH	25.71	OVERWEIGHT
4V3	2	-	-	2	MED	NUR ISTERATE	19	4.4	NORMAL	NA	NA

50V3	2	-	+	7	HIGH	DMPA	17	4.1	NORMAL	21.88	HEALTHY
51V3	2	-	-	1	MED	DMPA	18	5	HIGH	29.28	OVERWEIGHT
53V3	2	-	+	10	HIGH	NUR ISTERATE	21	4.4	NORMAL	32.42	OBESE
54V3	2	-	-	1	LOW	NUR ISTERATE	18	4.1	NORMAL	28.91	OVERWEIGHT
55V3	2	-	-	0	LOW	NUR ISTERATE	18	4.1	NORMAL	27.92	OVERWEIGHT
56V3	2	-	-	0	LOW	NUR ISTERATE	20	5.3	HIGH	20.83	HEALTHY
57V3	2	-	+	9	HIGH	NUR ISTERATE	18	4.7	HIGH	36.65	OBESE
59V3	2	-	-	0	HIGH	NUR ISTERATE	17	4.7	HIGH	21.40	HEALTHY
60V3	2	-	-	0	MED	NUR ISTERATE	17	4.7	HIGH	18.26	UNDERWEIGHT
61V3	2	+	-	1	MED	NUR ISTERATE	17	4.7	HIGH	28.40	OVERWEIGHT
63V3	2	+	+	7	MED	NUR ISTERATE	16	4.7	HIGH	23.83	HEALTHY
64V3	2	+	+/-	5	HIGH	NUR ISTERATE	18	4.7	HIGH	28.25	OVERWEIGHT
66V3	2	-	-	0	HIGH	NUR ISTERATE	19	5	HIGH	24.39	HEALTHY
67V3	2	+	+	8	HIGH	NUR ISTERATE	17	5	HIGH	19.78	HEALTHY
68V3	2	-	+	8	HIGH	NUR ISTERATE	16	6.1	HIGH	28.88	OVERWEIGHT
6V3	2	-	-	2	HIGH	NUR ISTERATE	20	4.7	HIGH	22.77	HEALTHY
70V3	2	-	NA	NA	LOW	NUR ISTERATE	17	4.7	HIGH	17.36	UNDERWEIGHT
72V3	2	+	+	10	MED	NUR ISTERATE	17	4.7	HIGH	26.13	OVERWEIGHT
73V3	2	+	-	1	LOW	NUR ISTERATE	17	4.1	NORMAL	27.83	OVERWEIGHT
74V3	2	-	+	8	MED	DMPA	19	4.7	HIGH	NA	NA
76V3	2	+	+	9	HIGH	NUR ISTERATE	17	5.6	HIGH	16.97	UNDERWEIGHT
79V3	2	-	-	0	LOW	NUR ISTERATE	20	NA	NA	26.95	OVERWEIGHT
7V3	2	+	-	0	LOW	NUR ISTERATE	19	4.1	NORMAL	36.11	OBESE
80V3	2	+	-	8	MED	DMPA	22	4.4	NORMAL	24.61	HEALTHY

81V3	2	-	+	9	MED	NUR ISTERATE	18	5.3	HIGH	27.05	OVERWEIGHT
82V3	2	+	-	0	HIGH	NUR ISTERATE	18	NA	NA	20.69	HEALTHY
84V3	2	+	+	9	MED	NUR ISTERATE	17	4.4	NORMAL	21.22	HEALTHY
85V3	2	-	+/-	5	HIGH	DMPA	22	4.7	HIGH	22.50	HEALTHY
86V3	2	+	+	8	MED	NUR ISTERATE	20	4.4	NORMAL	NA	NA
87V3	2	+	+	9	HIGH	NUR ISTERATE	19	4.4	NORMAL	34.85	OBESE
88V3	2	-	-	0	LOW	NUR ISTERATE	20	4.7	HIGH	25.34	OVERWEIGHT
8V3	2	+	-	2	LOW	NUR ISTERATE	18	4.7	HIGH	24.80	HEALTHY
97V3	2	-	+	10	MED	IMPLANON	20	5.3	HIGH	29.52	OVERWEIGHT
99V3	2	-	+	8	HIGH	NUR ISTERATE	19	5.6	HIGH	22.58	HEALTHY
9V3	2	-	+	9	HIGH	NUR ISTERATE	19	4.7	HIGH	22.98	HEALTHY

Abbreviations: bacterial vaginosis (BV); sexually transmitted infections (STIs); hormonal contraception (HC); body mass index (BMI).

Table S2. Research studies validating the presence of fungal species as part of the vagina or neighbouring body site (*i.e.*, gut) chosen for the manually curated pan proteome and proteome sequence databases.

Fungal species	Pan Proteome	Proteome	References
Candida albicans	**Pan proteome**	**Proteome**	Sobel 1997; Drell *et al.* 2013; Guo *et al.* 2012; Zheng *et al.* 2013; Guzel *et al.* 2011a; Guzel *et al.* 2011b; Goswami *et al.* 2000; Bentubo *et al.* 2015; Holanda *et al.* 2007; Abia-Bassey and Utsalo 2006; Ferrazza *et al.* 2005; Ozcan *et al.* 2006; Macura and Skóra 2012; Beigi *et al.* 2004; Lynch and Sobel 1994
Candida krusei (*Pichia kudriavzevii*)	**Pan proteome**	**Proteome**	Sobel 2007; Mahmoudi Rad *et al.* 2010; Odds and Bernaerts 1994; Drell *et al.* 2013; Singh *et al.* 2002; Nemes-Nikodém *et al.* 2015; Guo *et al.* 2012; Zheng *et al.* 2013; Guzel *et al.* 2011a; Guzel *et al.* 2011b; Bentubo *et al.* 2015; Abia-Bassey and Utsalo 2006; Ozcan *et al.* 2006; Macura and Skóra 2012; Lynch and Sobel 1994
Candida parapsilosis	**Pan proteome**	**Proteome**	Sobel 2007; Mahmoudi Rad *et al.* 2010; Odds and Bernearts 1994; Drell *et al.* 2013; Nemes-Nikodém *et al.* 2015; Guzel *et al.* 2011a; Guzel *et al.* 2011b; Bentubo *et al.* 2015; Holanda *et al.* 2007; Abia-Bassey and Utsalo 2006; Ferrazza *et al.* 2005; Macura and Skóra 2012; Beigi *et al.* 2004; Lynch and Sobel 1994
Candida tropicalis		**Proteome**	Sobel 2007; Mahmoudi Rad *et al.* 2010; Odds and Bernaerts 1994; Nemes-Nikodém *et al.* 2015; Paulitsch *et al.* 2006; Guo *et al.* 2012; Zheng *et al.* 2013; Guzel *et al.* 2011; Goswami *et al.* 2000; Guzel *et al.* 2011b; Bentubo *et al.* 2015; Holanda *et al.* 2007; Abia-Bassey and Utsalo 2006; Ferrazza *et al.* 2005; Ozcan *et al.* 2006; Macura and Skóra 2012; Lynch and Sobel 1994
Candida glabrata	**Pan proteome**		Sobel 2007; Mahmoudi Rad *et al.* 2010; Odds and Bernaerts 1994; Nemes-Nikodém *et al.* 2015; Paulitsch *et al.* 2006; Hong *et al.* 2016; Guo *et al.* 2012; Zheng *et al.* 2013; Guzel *et al.* 2011a; Guzel *et al.* 2011b; Goswami *et al.* 2000; Bentubo *et al.* 2015; Holanda *et al.* 2007; Abia-Bassey and Utsalo 2006; Ferrazza *et al.* 2005; Ozcan *et al.* 2006; Macura and Skóra 2012; Beigi *et al.* 2004; Lynch and Sobel 1994
Candida guilliermondii	**Pan proteome**		Mahmoudi Rad *et al.* 2010; Sobel 2007; Drell *et al.* 2013; Paulitsch *et al.* 2006; Guzel *et al.* 2011a; Guzel *et al.*

(*Meyerozyma guilliermondii*)			2011b; Bentubo *et al.* 2015; Holanda *et al.* 2007; Ferrazza *et al.* 2005; Macura and Skóra 2012; Abu-Elteen *et al.* 1997
Candida pseudotropicalis (*Candida kefyr*) (*Kluyveromyces marxianus*)	**Pan proteome**	**Proteome**	Fornari *et al.* 2016; Sobel 2007; Mahmoudi Rad *et al.* 2010; Odds and Bernaerts 1994; Corpus *et al.* 2004; Dennerstein *et al.* 2011; Nemes-Nikodém *et al.* 2015; Guzel *et al.* 2011a; Guzel *et al.* 2011b; Macura and Skóra 2012; Lynch and Sobel 1994
Zygosaccharomyces rouxii (*Candida mogii*)	**Pan proteome**		Drell *et al.* 2013
Candida dubliniensis	**Pan proteome**		Yang *et al.* 2003; Drell *et al.* 2013; Nemes-Nikodém *et al.* 2015; Paulitsch *et al.* 2006; Zheng *et al.* 2013; Guzel *et al.* 2011a; Macura and Skóra 2012; Beigi *et al.* 2004
Candida lusitaniae (*Clavispora lusitaniae*)	**Pan proteome**	**Proteome**	Sobel 2007; Mahmoudi Rad *et al.* 2010; Odds and Bernaerts 1994; Nemes-Nikodém *et al.* 2015; Paulitsch *et al.* 2006; Guo *et al.* 2012; Guzel *et al.* 2011a; Guzel *et al.* 2011b; Bentubo *et al.* 2015; Ferrazza *et al.* 2005; Lynch and Sobel 1994
Kluyveromyces lactis (*Candida sphaerica*)	**Pan proteome**		Fornari *et al.* 2016; Guzel *et al.* 2011a
Candida lipolytica (*Yarrowia lipolytica*)	**Pan proteome**		Rajagopalan *et al.* 1996; Guzel *et al.* 2011b
Torulaspora delbrueckii		**Proteome**	Fornari *et al*. 2016
Candida utilis (*Cyberlindnera jadinii*)	**Pan proteome**		Guzel *et al.* 2011b
[*Candida] intermedia*	**Pan proteome**	**Proteome**	Arechavala *et al.* 2007; Zheng *et al.* 2013
Candida auris	**Pan proteome**	**Proteome**	Welsh *et al.* 2017
Candida viswanathii		**Proteome**	Neder 1992; Guo *et al*. 2012
Candida inconspicua		**Proteome**	Dennerstein *et al.* 2011; Nemes-Nikodém *et al.* 2015; Paulitsch et al. 2006
Candida metapsilosis		**Proteome**	Miranda-Zapico *et al.* 2011
Candida orthopsilosis	**Pan proteome**	**Proteome**	Miranda-Zapico *et al.* 2011; Borman *et al.* 2009
Candida rugosa		**Proteome**	El-din *et al.* 2009; Adjapong *et al.* 2016; Huffnagle and Noverr 2013

Pseudozyma antarctica (*Candida antartica*)	**Pan proteome**	**Proteome**	Nash *et al.* 2017 (**GUT**)
Candida tenuis		**Proteome**	Nash *et al.* 2017 (**GUT**)
[Candida] haemulonii		**Proteome**	Fornari *et al.* 2016; Goswami *et al.* 2000
Candida maltosa		**Proteome**	Nash *et al.* 2017 (**GUT**)
Candida sake		**Proteome**	Guo *et al.* 2012
Candida solani		**Proteome**	Zheng *et al.* 2013
Candida norvegensis		**Proteome**	Abia-Bassey and Utsalo 2006
Candida zeylanoides		**Proteome**	Martens *et al.* 2004
Naumovozyma castellii		**Proteome**	Nash *et al.* 2017; Li *et al.* 2014 (**GUT**)
Saccharomyces cerevisiae	**Pan proteome**		Spinillo *et al.* 1995; Guo *et al.* 2012; Nemes-Nikodém *et al.* 2015; Sobel *et al.* 1993; Guo *et al.* 2012; Zheng *et al.* 2013; Macura and Skóra 2012; Lynch and Sobel 1994
Lodderomyces elongisporus (*Saccharomyces elongisporus*)		**Proteome**	Lockhart *et al.* 2008
Saccharomyces kudriavzevii		**Proteome**	Nash *et al.* 2017 (**GUT**)
Saccharomyces pastorianus		**Proteome**	Hong *et al.* 2016
Aspergillus fumigatus	**Pan proteome**		Pfaller and Diekema 2004; Drell *et al.* 2013
Aspergillus terreus		**Proteome**	Steinbach *et al.* 2004; Drell *et al.* 2013
Aspergillus niger	**Pan proteome**		Drell *et al.* 2013
Aspergillus versicolor		**Proteome**	Drell *et al.* 2013; Zheng *et al.* 2013
Aspergillus candidus	**Pan proteome**	**Proteome**	Drell *et al.* 2013
Aspergillus flavus			Gouba *et al*. 2014a; Nash *et al.* 2017 (**GUT**)
Aspergillus nomius		**Proteome**	Nash *et al.* 2017 (**GUT**)
Aspergillus clavatus			Nash *et al.* 2017; Li *et al.* 2014 (**GUT**)
Aspergillus sydowii		**Proteome**	Ukhanova *et al.* 2014
Aspergillus ochraceus		**Proteome**	Hong *et al.* 2016
Aspergillus penicillioides		**Proteome**	Guo *et al.* 2012; Drell *et al*. 2013
Eurotium amstelodami		**Proteome**	Drell *et al.* 2013

(*Aspergillus amstelodami*)			
Schizosaccharomyces pombe		**Proteome**	Fornari *et al.* 2016
Schizosaccharomyces octosporus	**Pan proteome**	**Proteome**	Nash *et al.* 2017
Blastomyces gilchristii	**Pan proteome**	**Proteome**	Muñoz *et al.* 2015
Ajellomyces dermatitidis	**Pan proteome**		Untereiner *et al.* 2004
Blastomyces percursus		**Proteome**	Sigler *et al.* 1996; Maphanga *et al.* 2020
Cryptococcus neoformans	**Pan proteome**		Mirza *et al.* 2003; El-din *et al.* 2009
Cryptococcus humicolus		**Proteome**	Holland *et al.* 2003
Cryptococcus albidus		**Proteome**	Guo *et al.* 2012
Trichosporon asahii	**Pan proteome**	**Proteome**	Ferrazza *et al.* 2005
Trichosporon inkin		**Proteome**	Bentubo *et al.* 2015; Abia-Bassey and Utsalo 2006
Trichosporon mucoides		**Proteome**	El-din *et al.* 2009
Trichosporon jirovecii		**Proteome**	Guo *et al.* 2012
Rhodotorula mucilaginosa		**Proteome**	De Almeida *et al.* 2008; Bentubo *et al.* 2015
Rhodotorula minuta		**Proteome**	Severo Gomes *et al.* 2011; El-din *et al.* 2009
Rhodotorula rubra		**Proteome**	Severo Gomes *et al.* 2011; El-din *et al.* 2009
Rhodotorula taiwanensis		**Proteome**	Nash *et al.* 2017 (**GUT**)
Rhodotorula sp.		**Proteome**	Drell *et al.* 2013; Holanda *et al.* 2007; Ferrazza *et al.* 2005
Rhizopus microspores		**Proteome**	Cano *et al.* 2014 (**GUT**)
Mucor circinelloides	**Pan proteome**	**Proteome**	Barizzi *et al.* 2016
Mucor piriformis		**Proteome**	Raimondi *et al*. 2019 (**GUT**)
Pichia fermentans		**Proteome**	Fornari *et al.* 2016
Pichia membranifaciens		**Proteome**	Fornari *et al.* 2016
Pichia segobiensis		**Proteome**	Fornari *et al.* 2016
Pichia ohmeri		**Proteome**	El-din *et al.* 2009
Pichia sorbitophila		**Proteome**	Nash *et al.* 2017 (**GUT**)
Alternaria Alternata	**Pan proteome**		Drell *et al.* 2013; Hong *et al.* 2016; Zheng *et al.* 2013

Fusarium solani		**Proteome**	Nucci and Anaissie 2007; Guo *et al.* 2012
Fusarium pseudograminearum	**Pan proteome**	**Proteome**	Nash *et al.* 2017 (**GUT**)
Malassezia globosa		**Proteome**	Kalus 2017 (**SKIN**); Hong *et al.* 2016
Malassezia sympodialis	**Pan proteome**		Hallen-Adams *et al.* 2015
Malassezia pachydermatis		**Proteome**	Chen *et al.* 2011; Gouba *et al.* 2013; Gouba *et al.* 2014a (**GUT**), Gouba *et al.* 2014b; Hamad *et al.* 2012
Malassezia furfur		**Proteome**	Theelen *et al.* 2018
Malassezia restricta		**Proteome**	Hong *et al.* 2016
Debaryomyces hansenii (*Candida famata*)	**Pan proteome**		El-din *et al.* 2009; Severo Gomes *et al.* 2011; Guzel *et al.* 2011; Abia-Bassey and Utsalo 2006; Macura and Skóra 2012; Beigi *et al.* 2004
Cladosporium perangustum		**Proteome**	Drell *et al.* 2013
Cladosporium silenes		**Proteome**	Hong *et al.* 2016
Cladosporium sp		**Proteome**	Drell *et al*. 2013
Trichoderma gamsii	**Pan proteome**	**Proteome**	Nash *et al.* 2017 (**GUT**)
Penicillium camemberti (part of *P. solitum*)	**Pan proteome**	**Proteome**	Nash *et al.* 2017 (**GUT**); Gouba *et al.* 2013
Penicillium digitatum	**Pan proteome**	**Proteome**	Nash *et al.* 2017 (**GUT**)
Penicillium marneffei	**Pan proteome**		Segretain 1959
Penicillium nordicum		**Proteome**	Nash *et al.* 2017 (**GUT**)
Penicillium expansum		**Proteome**	Nash *et al.* 2017 (**GUT**)
Penicillium italicum		**Proteome**	Ott *et al.* 2008 (**GUT**)
Penicillium solitum	**Pan proteome**	**Proteome**	Gouba *et al.* 2014a
Penicillium notatum	**Pan proteome**	**Proteome**	Cohen *et al.* 1969; Ukhanova *et al.* 2014 (**GUT**)
Penicillium griseofulvum		**Proteome**	Zheng *et al.* 2013
Phialemoniopsis curvata		**Proteome**	Hong *et al.* 2016
Thanatephorus cucumeris		**Proteome**	Hong *et al.* 2016
Kazachstania telluris		**Proteome**	Hong *et al.* 2016
Ambrosia artemisiifolia		**Proteome**	Hong *et al.* 2016

Stereum ostrea		**Proteome**	Hong *et al.* 2016
Davidiellaceae sp.		**Proteome**	Drell *et al.* 2013
Eladia saccula (*Penicillium sacculum*)		**Proteome**	Guo *et al.* 2012
Phytophthora sp.		**Proteome**	Guo *et al.* 2012
Dothideomycetes sp.		**Proteome**	Guo *et al.* 2012
Phoma gardeniae		**Proteome**	Guo *et al.* 2012
Saccothecium sepincola		**Proteome**	Guo *et al.* 2012
Plectosphaerella sp.		**Proteome**	Guo *et al.* 2012
Scopulariopsis brevicaulis		**Proteome**	Guo *et al.* 2012
Torrubiella sp.		**Proteome**	Guo *et al.* 2012
Simplicillium lanosoniveum		**Proteome**	Guo *et al.* 2012
Paecilomyces farinosus		**Proteome**	Guo *et al.* 2012
Coniochaeta velutina		**Proteome**	Guo *et al.* 2012
Filobasidium uniguttulatum		**Proteome**	Guo *et al.* 2012
Galactomyces geotrichum		**Proteome**	Drell *et al.* 2013; Zheng *et al.* 2013
Ophiocordyceps sp.		**Proteome**	Zheng *et al.* 2013
Chaetomium sp.		**Proteome**	Zheng *et al.* 2013
Scopulariopsis fusca		**Proteome**	Zheng *et al.* 2013
Wallemia sebi	**Pan proteome**	**Proteome**	Zheng *et al.* 2013
Galactomyces candidum		**Proteome**	Nash *et al.* 2017 (**GUT**); Arendrup *et al.* 2014

Table S3. Details of parameters used to search raw mass spectrometry (MS) files with MaxQuant version 1.6.3.4.

MaxQuant Parameter	Value/Description
Version	1.6.3.4
Fixed modifications	Carbamidomethyl (C)
Enzyme	Trypsin (P)
Include contaminants	TRUE
PSM FDR	0,01
Protein FDR	0,01
Site FDR	0,01
Use Normalized Ratios For Occupancy	TRUE
Min. peptides	1
Min. razor peptides	1
Min. unique peptides	0
Modifications included in protein quantification	Oxidation (M); Acetyl (Protein N-term)
Discard unmodified counterpart peptides	TRUE
iBAQ	TRUE
iBAQ log fit	TRUE
Decoy mode	revert
Match between runs	TRUE
Include contaminants	TRUE
Second peptides	TRUE
Stabilize large LFQ ratios	TRUE
Require MS/MS for LFQ comparisons	TRUE
Min. peptide length for unspecific search	8
Max. peptide length for unspecific search	25
Razor protein FDR	TRUE

Chapter 3

Taxonomic and Functional Analysis of the Vaginal Mycobiome

Introduction

The vaginal microbiome (VMB) is a complex, dynamic and diverse ecosystem of microorganisms (*i.e.*, bacteria, archaea, viruses, and fungi) that inhabit the vagina (Danielsson *et al.* 2011; Diaz *et al.* 2010). Taxonomic inhabitants of an optimal microbiome, such as *Lactobacillus* spp., are involved in maintaining a low vaginal pH (≤4.5), reducing the entry of pathogens, and regulating the local innate immune system (Aldunate *et al.* 2015; Donders *et al.* 2009; Klebanoff *et al.* 1991; Schwebke 2001). Thus, when the VMB is in an optimal state, it helps maintain vaginal health and functions to protect its host from diseases and prevent reproductive complications (Danielsson *et al.* 2011). However, when there are alterations in VMB characteristics, such as shifts in taxon composition and/or function, there is a disruption in the functioning of the vaginal environment. As a result, vaginal diseases are able to emerge (Foster *et al.* 2008).

Unfortunately, many microbiome studies have focused exclusively on the predominant bacteriome (Ackerman and Underhill 2017; Nash *et al.* 2017). Accordingly, research has overlooked minority kingdoms such as fungi and archaea (Drell *et al.* 2013; Hager and Ghannoum 2018; Huber *et al.* 2007; Witherden *et al.* 2017). In 2017, less than 0.4% of the microbiome-related literature referred to or examined fungal communities, and even fewer had investigated these communities in the female genital tract (FGT) (Drell *et al.* 2013; Hernández-Santos and Klein 2017). Furthermore, early publications of vaginal microbiology underestimated the composition, diversity, and complexity of fungal communities. This was in part due to the limitations of the culture-dependent techniques used, which restricted analyses to known pathogens (*e.g., Candida* and *Saccharomyces* species). Consequently, there is minimal data on which fungal species are present in the VMB (Drell *et al.* 2013).

The genus *Candida* typically dominates FGT fungal communities and is usually a non-pathogenic commensal inhabitant (Beigi *et al.* 2004; Bradford and Ravel 2017; Drell *et al.* 2013). However, *Candida* is also one of the leading causes of vaginal infection as it is able to also function as a pathogen (Rane *et al.* 2014). The pathogenicity of *Candida* spp. is attributed

to various fitness and virulence traits, which allows pathogens to establish host infection. Beneficial traits to *Candida* include adhesion to host tissues, biofilm formation, and secretion of extracellular hydrolases (Deorukhkar and Saini 2015; Sachin *et al.* 2012). *Candida* is responsible for the inflammatory condition vulvovaginal candidiasis (VVC), which affects approximately 75% of the female population at least once during their lives (Fidel 2004; Hurley and De Louvois 1979).

Drell *et al.* (2013) were the first to analyse the vaginal mycobiome using barcoded pyrosequencing technology. This study showed that the fungal component of the microbiome is more diverse than what was originally assumed. Their results confirmed that *Candida* spp. were the most prevalent (64.5%) colonizers of the vagina. Of this genus, *C. albicans* was identified as the most prevalent (82%) species, and the remaining non-albicans species belonged to *C. parapsilosis*, *C. dubliniensis*, *Candida* sp. VI04616, and *Pichia kudriavzevii* (*Candida krusei*). All aforementioned non-albicans species, with the exception of *Candida* sp. VI04616, had been linked to pathogenicity in the vagina in previous studies. However, the authors hypothesized that *C. parapsilosis*, C. *dubliniensis,* and *P. kudriavzevii* can also be affiliated with a normal VMB without causing disease. This hypothesis was based on the fact that all study participants were asymptomatic (without disease). This study also detected *Rhodotorula* sp LH51 as the most abundant species of the phylum Basidiomycota (Drell *et al.* 2013).

Another vaginal study by Zheng *et al.* (2013) used 18S rRNA sequencing to examine pregnant diabetic women. Their results found that the most predominant vaginal fungal species belonged to the genera *Candida* and *Saccharomyces.* The remaining taxa belonged to uncultured and low-abundance fungi that were unrecorded or underrepresented in previous studies using cultivation-dependent methods, such as *Aspergillus* spp. This study confirmed that the *C. albicans*, *C. glabarata*, *C. parapsilosis*, *C. tropicalis*, and *C. krusei* can be non-pathogenic as these species were present in both diabetic and non-diabetic groups.

A more recent vaginal study by Lehtoranta *et al.* (2021) identified more than 30 different fungal species using Internal Transcribed Spacers (ITS) sequencing. Fungal genera that were not identified in the previously mentioned studies included *Malassezia*, *Cladosporium*, and *Fusarium*. Interestingly, they identified a relatively high dominance of *Lodderomyces elongisporus*, which has not been previously identified in the vagina.

In general, the fungal community inhabiting the human microbiome performs many essential functions (Iliev and Leonardi 2017). Such functions include altering host and metabolic functioning, energy acquisition, vitamin-cofactor accessibility, immune system development and functioning, and disease progression (Dworecka-Kaszak *et al.* 2016; Iliev *et al.* 2012; Seed 2014). Fungi are also suggested to assist in maintaining microbial community composition (Dworecka-Kaszak *et al*. 2016; Wisecaver *et al*. 2014). From gut studies, alterations in the mycobiome have been observed during disease (Huseyin *et al*. 2017; Iliev and Leonardi 2017). These observations suggest that the mycobiome may act as a reservoir of potential opportunistic pathogens when the host is vulnerable (Chen *et al.* 2011; Polvi *et al.* 2015; Miceli *et al.* 2011). However, the role of fungal communities in non-diseased hosts remains largely unknown. Due to this, we remain uncertain on whether some fungi have beneficial effects on the host (Krüger *et al.* 2019).

Thus far, there has been conflicting evidence on whether *C. albicans* is associated with a BV negative or BV positive microbiome. In a study by Hong *et al.* (2016), samples that detected *C. albicans* had a low Nugent score (<4); and *Candida* was identified to be the most abundant fungal genus in the BV negative group, followed by the BV intermediate group. However, studies by Pramanick *et al.* (2019) and Lehtoranta *et al.* (2021), showed conflicting evidence. These studies showed that the prevalence of *C. albicans* increased in BV positive women, in comparison to BV negative women. Due to the conflicting evidence, more research would prove to be beneficial to better understand the relationship between BV and fungi.

Regarding genital inflammation, *Candida* is capable of altering host cytokine production and promoting anti-inflammatory signalling in a lactic acid-producing environment (*e.g.,* the BV negative state) (Bradford and Ravel 2017). Regarding STIs, studies have demonstrated that women infected with *Chlamydia trachomatis* are also colonized with *Candida*. *C. trachomatis* has been shown to bind *C. albicans* inhibiting chlamydial infectivity (Kruppa *et al.* 2019).

It is critical to improve our understanding of the role of the vaginal mycobiome in optimal and non-optimal states. To do this, we need to explore the functional and taxonomic compositional dynamics of fungal species, which requires a culture-independent approach. One viable method includes metaproteomics, which provides a deeper understanding of microbiomes by providing information on protein expression levels and active microbial species in a sample (Heyer *et al.* 2017). Metaproteomic data enables us to understand microbiome functions and

the metabolic pathways possibly implicated in disease pathogenesis (Potgieter *et al*. 2019). To date, metaproteomics has been used to characterize various human compartments, such as the gut microbiome (Ferrer *et al.* 2013; Long *et al.* 2020; Verberkmoes *et al.* 2009; Young *et al*. 2015), lung microbiome (Jagtap *et al.* 2018), oral microbiome (Rabe *et al.* 2019), and vaginal bacteriome (Afiuni-Zadeh *et al.* 2018). However, there have not been any metagenomic or metaproteomic studies examining the FGT fungal populations in African women. This research is essential since African women are at a high risk of HIV, BV, and other STIs; and fungi are known to be major contributors of opportunistic infections that affect patients with HIV. These opportunistic fungal infections are mainly caused by *Cryptococcus neoformans*, *Histoplasma capsulatum*, and *Talaromyces marneffei*.

From the few FGT mycobiome studies that have been performed, the majority have used DNA-based phylogenetic profiling (Drell *et al.* 2013; Guo *et al.* 2012). Thus, this study would be the first to examine FGT fungal communities using metaproteomics and determine the value of this method when conducting mycobiome research. The work conducted in this Chapter aims to taxonomically and functionally characterize the FGT mycobiome in non-optimal states, such as BV, genital inflammation, and STIs using metaproteomics.

Methods

Curation and fungal profiling

The cohort described in chapter 2 was used for downstream analyses, which included 123 women. In summary, some women provided samples over two visits, 113 women provided vaginal samples during visit 1 and 89 during visit 2. Of the samples collected, 94 samples were positive for BV, 12 samples were BV intermediate, 85 were BV negative, 62 were highly inflamed, 77 were moderately inflamed, and 58 were lowly inflamed.

To ensure identifications of protein groups were reliable, a protein group was only accepted for downstream analysis if it had >=2 peptides assigned to it. Potential contaminants, reverse peptides, and protein groups that were deleted from the UniProtKB database were removed from our dataset. Raw iBAQ protein intensity values were normalized and $\log_2$-transformed using the *MSnbase R* package (Gatto and Lilley 2012). For differential analysis of the vaginal protein profiles, missing iBAQ intensities were imputed using the k-nearest neighbour (KNN)

method using *the imputeLCMD R* package to not introduce any negative values into the dataset. Imputation allows the use of statistical methods that require a complete data matrix, such as t-tests and hierarchical clustering methods (Lazar *et al.* 2016).

Protein groups were separated according to their five taxonomic kingdoms: bacterial, mammalian, fungal, archaeal, and viral. The relative abundance value for each taxonomic kingdom was calculated by dividing the summed iBAQ protein intensity value of each kingdom by the total summed iBAQ protein intensity values across all kingdoms. Thereafter, fungal protein groups were extracted for downstream analysis.

A diversity plot of fungal genera was created to assess the relative contribution of each fungal genus to the total fungal protein intensity value. The genus distribution was calculated by summing all iBAQ protein intensity values mapped to a specific genus and dividing the result by the total summed protein intensity value across all fungal genera. An average fungal community group structure profile was created using the genus distribution for 1) BV categories (negative, intermediate, positive), 2) inflammation categories (low, moderate, high), and 3) different STIs (*Chlamydia trachomatis*, *Neisseria gonorrhoeae*, *Trichomonas vaginalis*, *Mycoplasma genitalium*, and Herpes Simplex Virus 2) using Excel.

Taxonomic and Functional Analysis

Reverse and potential contaminant peptide sequences were removed from the dataset. Peptide sequences were submitted to the *Unipept* website (http://unipept.ugent.be/; version 1.4; settings: I and L equated, advanced missed cleavage option, search trypsinized, and duplicate peptides were filtered) for taxonomic analysis. *Unipept* assigned peptide sequences to taxa by estimating the lowest common ancestor (LCA).

The functional potential of the vaginal mycobiome was estimated using a protein-centric approach. The list of protein IDs from the *proteinGroups.txt* file was assigned to *Kyoto Encyclopedia of Genes and Genomes* (*KEGG*) *Orthology* (*KO*) groups and *GO* IDs using *GhostKOALA* version 2.2 and UniProt, respectively (Kanehisa and Goto 2000; Kanehisa *et al.* 2016). *GO* identifiers (IDs) provide insight into biological processes (BPs), molecular functions, and cellular components. KO assignments were manually curated to remove categories associated with host-level functions.

The relative abundance of functional terms was calculated by summing the iBAQ protein intensities of all associated *KO*s and *GO* terms, including proteins belonging to multiple categories. These values were used to generate a functional profile. An average fungal functional profile was created for 1) BV categories, 2) inflammation categories, and 3) different STIs using *Excel*.

Comparisons

Statistical analyses were performed using the *R* v3.4.2 statistical software. The imputed iBAQ intensity values were used to identify significant differentially abundant fungal genera, proteins, and functional terms at different BV states, inflammation levels, and STI status using the *limma R* package (Smyth 2005). To perform fungal analysis at the genus level, proteins were aggregated if they were associated with a particular genus. To perform analysis at the *GO* BP level, iBAQ intensities were aggregated for protein IDs mapping to the same BP. *Limma* provided a list of significant differentially abundant proteins, genera, and functional terms by performing a moderated t-test (Karp and Lilley 2007). T-tests accounted for confounding factors, which were determined using an ANOVA test for BV and inflammation category comparisons. T-tests that compared STI and non-STI groups accounted for individuals with multiple STIs. Proteins with FDR adjusted p-values <0.05 were considered as differentially abundant.

To identify similarities and differences in taxonomic and functional expression profiles; and visualize relationships between optimal and non-optimal states, unsupervised hierarchical clustering of iBAQ intensities was performed for all protein and functional term assignments using the *heatmap.2 R* package (Warnes *et al.* 2005). Manhattan was used as the distance metric to calculate the absolute differences between samples. Differences in the fungal community structure between optimal and non-optimal individuals were visualized using a Non-metric Multi-dimensional Scaling (NMDS) plot. The Bray-Curtis dissimilarity matrix was used, and the plot was created using the *R vegan* and *ggplot2* package (Wickham 2009). To test for statistically significant differences in composition between and within optimal and non-optimal groups a Permutational Analysis of Variance (PERMANOVA) was performed using the *adonis* function in *R*'s *vegan* package (Oksanen *et al.* 2017).

Results

Taxonomic profile

We calculated the relative abundances of taxonomic kingdoms using iBAQ protein intensities. Prior to normalization and imputation, Metazoa was the most relatively abundant kingdom (91.1%), followed by Bacteria (8.5%), Fungi (0.4%), Viruses (0.06%), and Archaea (5.46e-05%) (Fig 3.1). After normalization and imputation, Bacteria had shifted to become the most relatively abundant kingdom (49.7%; Fig 3.1) as a result of there being more bacterial proteins (1289) in comparison to Metazoan (1202). From these results, it is clear that the imputation of zero values artificially inflates the relative abundance of taxonomic kingdoms (Fig 3.1).

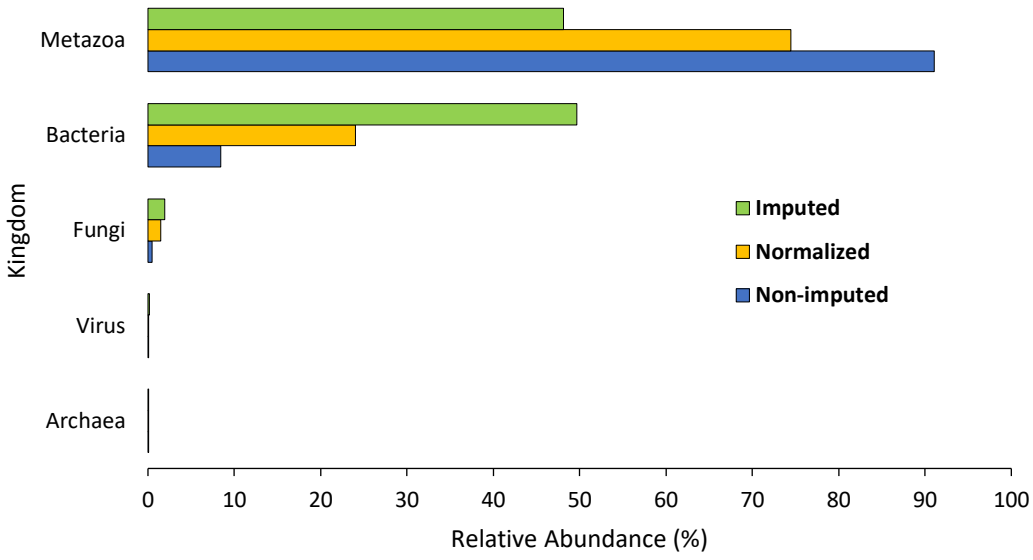


Figure 3.1. Relative abundance of taxonomic kingdoms pre- and post-normalization and imputation of the summed total iBAQ protein intensity values for each taxonomic kingdom determined using UniProt.

MS analysis identified 18,383 unique peptides, however, only 17,758 remained after QC. According to *Unipept's pept2lca,* 94 peptides were assigned to a fungal taxonomic lineage as its LCA. As a result, fungi represented 1.21% of the total Eukaryota peptides detected in samples and 0.53% of the total peptides identified (Fig 3.2). A total of 46.20% of the peptides were of bacterial origin, and 40.42% were of metazoan origin, while less than 1% could be assigned to other taxa (*i.e.*, Archaea and Viruses) (Fig 3.2). A total of 7.20% of peptide

sequences were assigned to the root, and 3.34% of identified peptide sequences could not be assigned to a taxonomic kingdom.

Figure 3.2 shows Metazoa and Bacteria to have a similar abundance since there is a comparable amount of metazoan (7177) and bacterial (8208) peptides. This abundance differs from the non-imputed relative abundance of taxonomic kingdoms for protein groups as those values were calculated using protein intensity values, whereas Unipept depicts this peptide abundance using the number of peptide sequences assigned to each taxonomic kingdom.

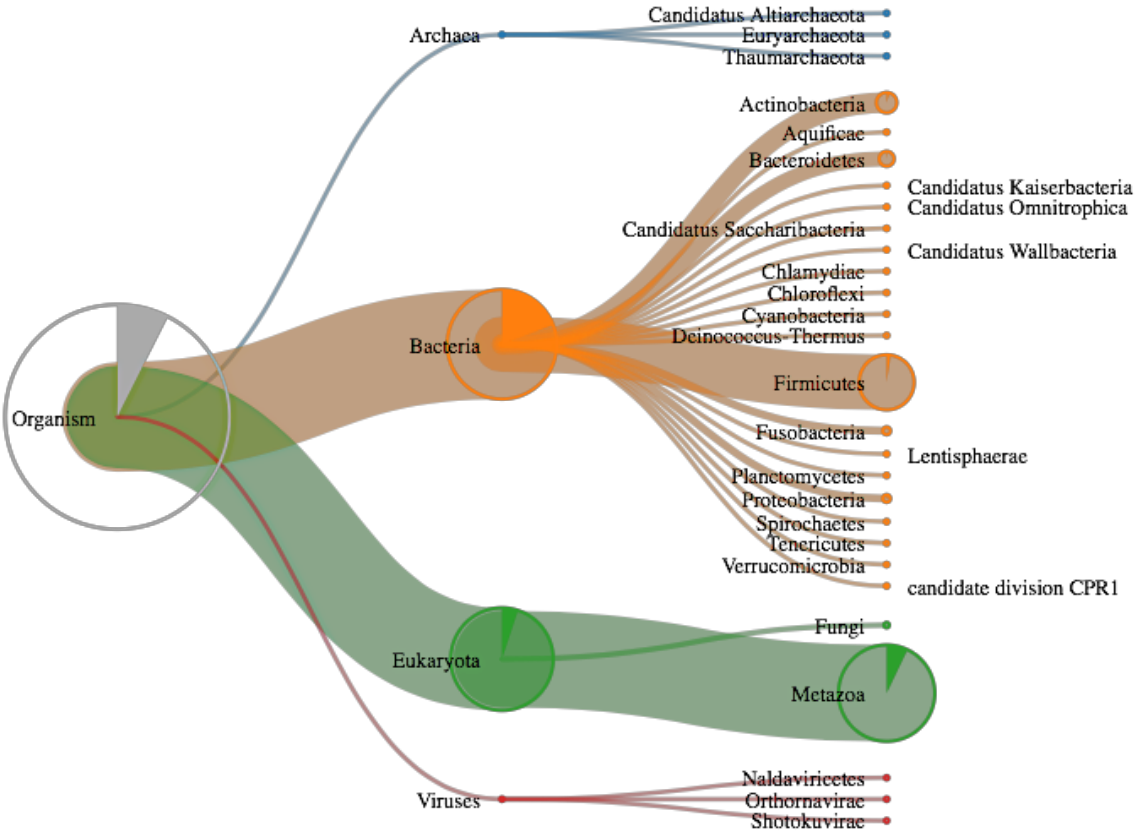


Figure 3.2. The abundance of peptide sequences that were assigned to various taxonomic kingdoms using Unipept. Peptide sequences were assigned to Archaea, Bacteria, Eukaryota, and Viruses. The majority of sequences were assigned to Bacteria (blue) and Eukaryota (green).

We investigated the percentage of peptides assigned to five major phyla. The majority of peptide sequences were assigned to the phyla Ascomycota (68.42%), followed by Basidiomycota (22.11%). A smaller percentage of peptide sequences were assigned to the phyla Mucoromycota, Blastocladiomycota, Microsporidia, Zoopagomycota, Microsporidia, and Cryptomycota (Fig 3.3).

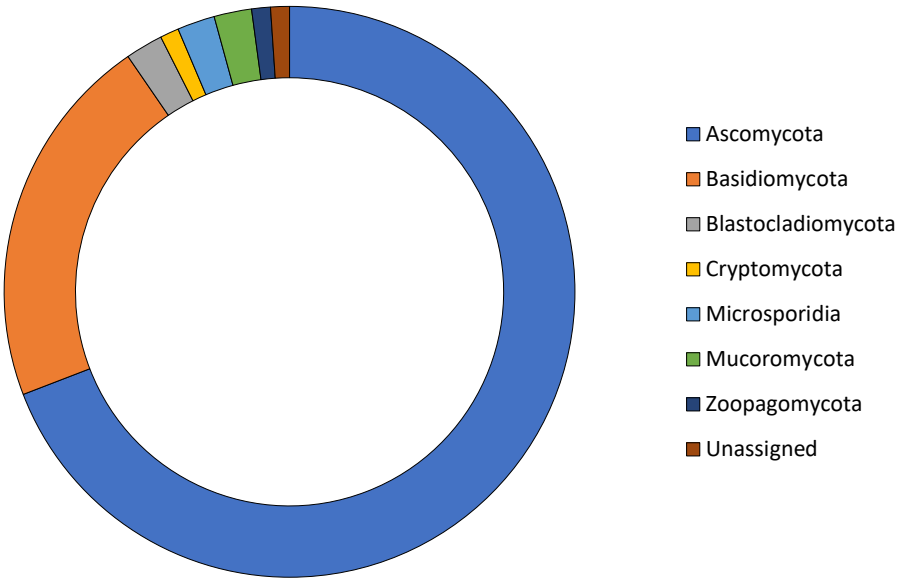


Figure 3.3. The proportion of peptide sequences assigned to fungal phyla using *Unipept.* A large proportion of fungal peptide sequences were assigned to the phylum Ascomycota (blue).

Unipept identified 46 distinct fungal genera using peptide sequences. Of those genera, the majority (7.4%) of sequences were assigned to *Aspergillus*, 6.4% to *Candida*, followed by 4.3% to *Alternaria*, *Cyberlindnera*, and *Trichosporon*. Of peptide sequences, 13.69% were unable to be assigned to the genus level. However, using protein groups (after implementing the exclusion criterion) 39 fungal genera were identified. Given the high number of zero values, non-imputed values were used to calculate relative abundances as to not artificially inflate the relative abundance of fungal genera. From these calculations, *Candida* (53.2%), *Conidiobolus* (20.9%), *Rhizopogon* (7.1%), and *Teratoramularia* (3.1%) were the most prevalent genera (Fig 3.4).

There is a difference in the number of genera identified using peptide sequences in comparison to protein groups, owing to multiple reasons, firstly protein groups underwent QC to remove protein groups that had less than two peptides assigned. Secondly, identified peptides are used to predict which proteins from the sequence database are present in the sample. However, mapping peptides to proteins and taxa is hard as many proteins cannot be identified using unique peptides but rather share peptides with other homologous proteins (Kleiner *et al*. 2017). These homologous proteins are sorted into one protein group and quantified together, which may account for the differences when using peptide sequences as opposed to protein

groups. Therefore, taxonomic analysis using peptide sequences is more accurate for the above reasons and is generally used in the taxonomic analyses in Metaproteomics.

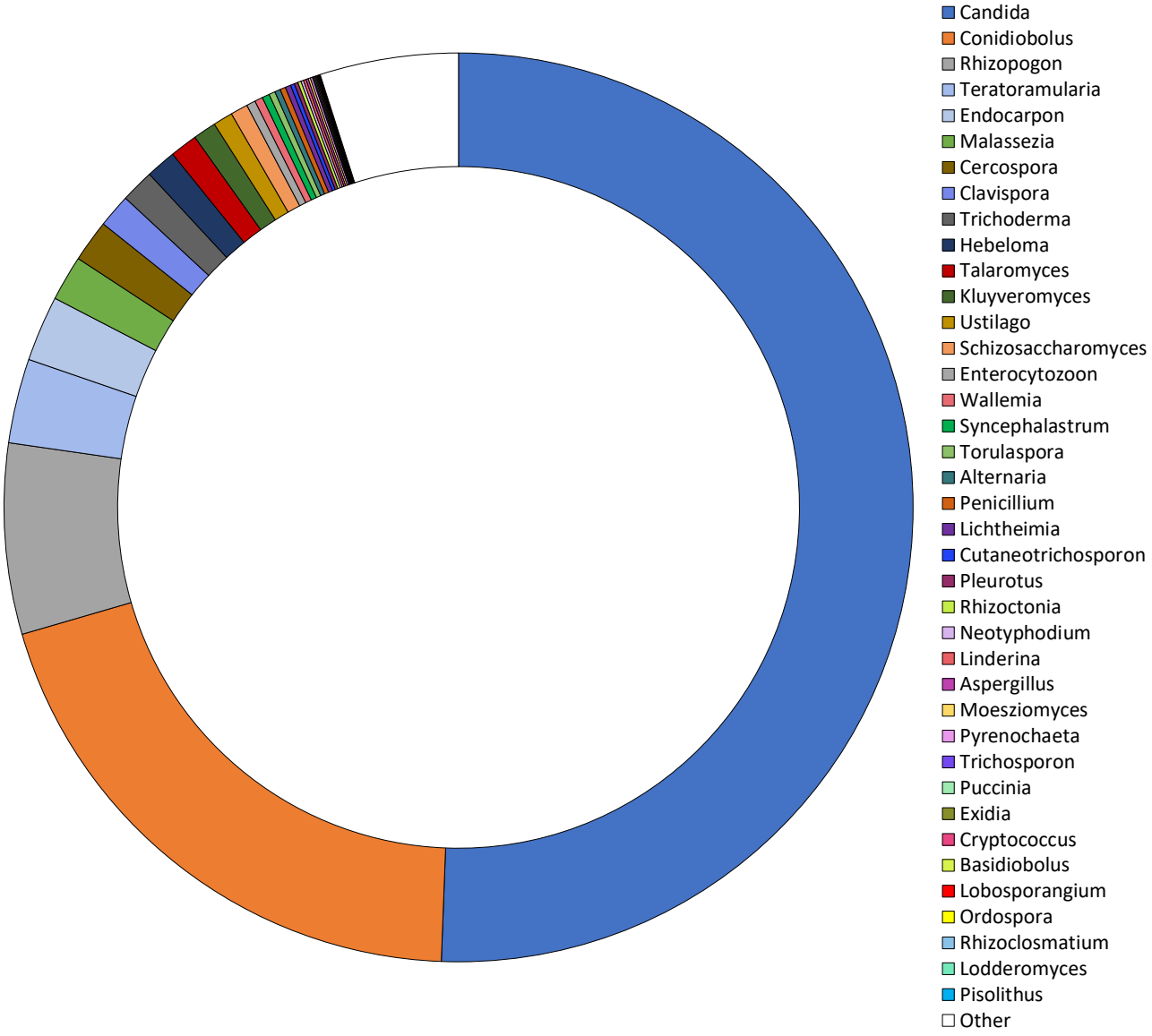


Figure 3.4. Relative abundance of fungal genera pre-normalization using summed total iBAQ protein intensity values for each genus. *Candida* was the most relatively abundant genus (blue).

Table 3.1. Characteristics and importance of the fungal species identified in this study using protein groups.

Fungal Species Found	Characteristics and Importance
Alternaria alternata	Commonly found species in the vaginal mycobiome (Drell *et al.* 2013; Guo *et al.* 2012).

Basidiobolus meristosporus	A filamentous fungus distributed worldwide (Kwon-Chung and Bennett 1992). It can saprophytically live in the intestines of mainly cold-blooded vertebrates and on decaying fruits and soil. The fungus prefers glucose as a carbon source and grows rapidly at room temperature (Yang 1962). This species is known to cause zygomycosis in humans (Gugnani 1992; Joe *et al.* 1956).
***Candida albicans* (Yeast)**	Commonly found species in the vaginal mycobiome (Drell *et al.* 2013; Guo *et al.* 2012; Zheng *et al.* 2013).
***Candida dubliniensis* (Yeast)**	Commonly found species in the vaginal mycobiome (Drell *et al.* 2013; Fornari et al. 2016; Guo *et al.* 2012; Lehtoranta *et al.* 2021).
Candida orthopsilosis	Commonly found species in the vaginal mycobiome (Drell *et al.* 2013; Guo *et al.* 2012).
Candida thasaenensis	Identified as a new anamorphic yeast species (Poomtein *et al.* 2013)
Cercospora beticola	A pathogen that typically infects plants, it causes Cercospora leaf spot disease in sugar beets, spinach, and swiss chard (Rangel *et al*. 2020).
***Clavispora lusitaniae* (*Candida lusitaniae*)**	Commonly found species in the vaginal mycobiome (Drell *et al.* 2013; Guo *et al.* 2012; Nejat *et al.* 2018).
Conidiobolus coronatus	An opportunistic pathogen that causes rhinofacial conidiobolomycosis in humans (Isa-Isa *et al.* 2012; Shabrawi *et al*. 2014). However, a case of *Condiobolus* was found in the vagina in a study by Subramanian and Sobel (2011).
Cryptococcus neoformans	This species has previously been isolated from the vagina (El-din *et al.* 2009). It is an opportunistic pathogen, which can infect the vagina, however, it is uncommon and usually occurs in immunocompromised patients (Chen *et al.* 1993).
***Cutaneotrichosporon oleaginosum* (*yeast*)**	This yeast can metabolize various carbohydrates including lactose (Bracharz *et al.* 2017) and is rarely isolated from clinical samples (Nash *et al.* 2017).
Endocarpon pusillum	A lichen-forming fungus (Wang *et al.* 2014)

Enterocytozoon bieneusi	An obligate intracellular parasite that is able to infect humans (Lanternier *et al.* 2009)
Exidia glandulosa	This genus usually occurs in dead wood (Spirin *et al.* 2018). *Exidia* has also been found in the air (Els *et al.* 2019).
Hebeloma cylindrosporum	A mushroom-forming fungus that seems to prefer sandy soils with little to no organic matter (Slot *et al.* 2010)
Kluyveromyces marxianus (Yeast) **(*Candida kefyr*)**	*Kluyveromyces* has been isolated from the vagina (Fornari *et al.* 2016; Nejat *et al*. 2018)
Lichtheimia corymbifera	The natural environment for *L. corymbifera* appears to be soil and decaying grasses. This fungus has optimal growth at human body temperature (35-37 °C). It normally lives as a saprotrophic mold, but can also be an opportunistic pathogen in humans (Piancastelli *et al.* 2009).
Linderina pennispora	Aerial hyphae (Young 1969)
Lobosporangium transversale	The genus is monotypic, containing the single species *L. transversale*, which has only been isolated from arid soils on three occasions between 1964 and 1968 and has not been reported again (Benny and Blackwell 2004)
Lodderomyces elongisporus (Saccharomyces elongisporus)	Lehtoranta *et al*. (2021) identified this species in their vaginal study, prior to this species had not been reported in the vagina. This species is usually mistaken for *C. albicans* in culture-based assays.
Malassezia sympodialis	Commonly found genus in the vaginal mycobiome (Lehtoranta *et al*. 2021).
Moesziomyces aphidis (***Pseudozyma aphidis***)	*M. aphidis* is sometimes found in clinical samples, human infections are rarely reported but this genus has been described as a human pathogen (Sugita *et al.* 2003). This species has been isolated from water (Boekhout 2011) and various other sources, including soil and human blood (Kruse *et al.* 2017). Members of this genus are close relatives of *Ustilago maydis* (also identified) (Joo *et al.* 2016).
Neotyphodium sp.	*Neotyphodium* is a genus of endophytic fungi symbiotic with grass (Christensen *et al.* 2012).

Ordospora colligata	An obligatory intracellular parasite, that has only been shown to infect the gut of the crustacean Daphnia magna thus far (Larsson *et al.* 1997).
Penicillium vulpinum	*Penicillium* is well known and one of the most common fungi found in a diverse range of habitats, including soil, air, extreme environments (temperature, salinity, water deficiency, and pH), and various food products. It also has been found in the vagina (Moraes *et al.* 2004).
***Petromyces alliaceus* (*Aspergillus alliaceus*)**	Although *Petromyces* spp. has been isolated from surfaces of plants and soil, *P. alliaceus* is rarely recovered from an invasive fungal infection (Balajee *et al.* 2007). The genus *Aspergillus* has been found to infect the vagina (Drell *et al.* 2013; Godoy-Vitorino *et al.* 2018; Zheng *et al.* 2013).
Pisolithus tinctorius	*Pisolithus tinctorius* is an ectomycorrhizal fungus that interacts with some of the most important tree genera from temperate forests (Sebastiana *et al.* 2020).
Pleurotus ostreatus	The common edible oyster mushroom (Rühl *et al.* 2008).
Puccinia striiformis	*P. striiformis* is a plant pathogen, that causes stripe rust on wheat, but has other hosts as well (Chen 2005).
Pyrenochaeta sp.	This species belongs to a group of filamentous fungi that inhabits the soil and plant debris worldwide, and are well-known known as pathogens for a variety of plants and, occasionally, humans (Ferrer *et a*l. 2009; Lević *et al*. 2009). This species belongs to the order Pleosporales, which has been found in the vagina (Godoy-Vitorino *et al.* 2018).
Rhizoclosmatium globosum	*R. globosum* is flagellated Chytridiomycota fungus that has been found in damp and aquatic environments and associated with mixed vegetation (Grossart *et al.* 2019)
Rhizoctonia solani	A soil-borne plant pathogen with a wide host range and worldwide distribution (Paulitz *et al.* 2006; Verma 1996)
Rhizopogon vinicolor	*Rhizopogon* species have been established as a common component in the diet of many small mammals (Izzo *et al.* 2005) and deer (Ashkannehhad and Horton 2006). The viability of *Rhizopogon* spores is and may even be increased after

	mammalian gut passage, making mammals an important dispersal vector for *Rhizopogon* (Colgan and Claridge 2002).
Schizosaccharomyces cryophilus	This yeast has been occasionally isolated in fermented drinks or derived products such as grapes, must, wine, and beer. Most of the isolates involving the genus *Schizosaccharomyces* have been achieved in foods having high sugar content, such as honey, sweets, molasses, and dried fruit (Benito *et al.* 2018).
Schizosaccharomyces japonicus	*S. japonicus* is nonpathogenic to humans. This species was initially isolated from strawberries and grapes (Yukawa and Maki 1931; Wickerham and Duprat 1945).
Syncephalastrum racemosum	Found mainly from soil and dung in tropical and sub-tropical regions (Barcoto *et al.* 2017), usually presenting as colonizers (Rao *et al.* 2007) and rarely causes human infections (Amatya *et al.* 2010; Garg *et al.* 2004). It can also be a difficult laboratory contaminant.
Talaromyces marneffei **(*Penicillium marneffei*)**	*T. marneffei* is a pathogenic, thermally dimorphic fungus that is a significant cause of morbidity and mortality among HIV-infected (and other immunocompromised patients) (Chan *et al.* 2016; Limper *et al.* 2017; Vanittanakom *et al.* 2006). This species grows as a yeast at 37°C in human tissue (Sirisanthana and Supparatpinyo 2012).
Teratoramularia persicariae	Infects leaves (Videira *et al.* 2016).
Torulaspora delbrueckii (Yeast) **(*Candida colliculosa*)**	*T. delbrueckii* is ubiquitous and has been frequently isolated from natural environments (plants, soil, and insects) and from various habitats associated with human activities, such as winemaking, kefir, olive, and tequila, and cider production (Albertin *et al.* 2014). Fornari *et al.* 2016 detected this species in a clinical isolate in their study analyzing patients with vulvovaginitis.
Trichoderma gamsii	Common in soil and rhizosphere of plants in natural and agricultural fields and forests and on decaying wood (Danielson and Davey 1973). They are also occurring in the air, settled dust, and different water-related habitats including marine environments and drinking water (Kohlmeyer 1974). Facultative pathogens of humans (Kredics *et al.* 2011).

Trichosporon asahii	Trichosporon is a yeast isolated from soil, water samples, vegetables, as well as being a member of the normal flora of mouth, skin, and nails, it is the causative agent of superficial and deep infections in humans. It has been found to infect genital skin (Moraes *et al.* 2004) and El-din *et al.* (2009) isolated this genus from the vagina.
Ustilago maydis	This dimorphic species is a member of the smut family that infects maize, it is an economically important pathogen (Lanver *et al.* 2017).
Wallemia ichthyophaga	One of the most halophilic fungi ever described is rarely isolated from nature. The existing isolates have been found in concentrated seawater, in air, and on food preserved with low water activity (Gostinčar *et al.* 2009). *Wallemia* spp. are commonly found as a minor component of the commensal gastrointestinal mycobiota in humans (Paterson *et al.* 2017). Scott et al. (2018) identified this genus in their analysis of the vaginal mycobiome.
Wallemia mellicola	It has a worldwide distribution and is often found in habitats with low accessibility of water, from food preserved with high concentrations of sugar, salt or with drying, to dried straw and house dust (Zajc and Gunde-Cimerman 2018). Scott *et al.* (2018) identified this genus in their analysis of the vaginal mycobiome.

For some fungal assignments that were not characteristically found in the human body, we examined the peptide and protein characteristics from the *peptides.txt* and *proteinGroups.txt* from the Maxqaunt output files to determine if these fungal assignments were false (Table 3.2).

Table 3.2. Peptide and protein characteristics of fungal species assignments not found in the human body.

	Rhizoclosmatium globosum	***Endocarpon pusillum***
PEP score	0.0002	0.0003
Peptide count	2 (razor + unique)	2 (razor + unique)

Peptide sequence length	15	11
Protein count	24 (not unique)	1 (unique)
Protein sequence length	398	544

Functional Mycobiome Profile

The 94 peptides were assigned to 50 fungal protein groups passed quality control filtering, the majority of protein groups were actin, and heat shock proteins (HSPs). Of the HSPs, 5 were heat shock protein 70kDa (HSP-70). The remainder of prevalent proteins were calmodulin, tubulin beta chain, 78kDa glucose-regulated proteins (GRP78), and elongation factor 2 proteins (Table 3.3).

Table 3.3. Fungal protein groups identified and the percentage of these protein groups in the dataset.

Protein name	Number of proteins (n)	Percentage (%)
Actin	9	18%
Heat shock proteins	9	18%
Calmodulin	3	6%
Tubulin beta chain	3	6%
78 kDa glucose-regulated protein	2	4%
Elongation factor 2	2	4%
AAA ATPase	1	2%
ATP-dependent RNA helicase SUB2	1	2%
DNA-binding protein HU	1	2%
Exocyst complex component SEC15	1	2%
Fructose-bisphosphate aldolase	1	2%
Gamma-actin	1	2%
Glutamate dehydrogenase	1	2%
Gp_dh_C domain-containing protein	1	2%
Histone H2A	1	2%
NAD-binding 6-phosphogluconate dehydrogenase	1	2%
p-loop containing nucleoside triphosphate hydrolase protein	1	2%
Proteasome subunit alpha type	1	2%

Proteasome subunit beta	1	2%
Rab small monomeric GTPase Rab7	1	2%
Sporulation protein Spo15	1	2%
Tr-type G domain-containing protein	1	2%
Tubulin alpha chain	1	2%
V-type ATPase	1	2%
Xap5-domain-containing protein	1	2%
Uncharacterized protein	3	6%

According to GO annotation, 48% of the fungal proteins could be assigned to known BPs, 46% of proteins to cellular components, and 76% of proteins to molecular functions using *Uniprot*'s ID Mapping Tool. Whereas 20% protein groups could not be assigned to a GO term. From the GO BPs, the most relatively abundant pathways were histone H4 acetylation (37.7%), membrane raft polarization (37.7%), and the glycolytic process (16.8%) (Fig 3.5). "Other" included GO biological processes with a relative abundance of less than 0.5% (see Supplementary Table S1).

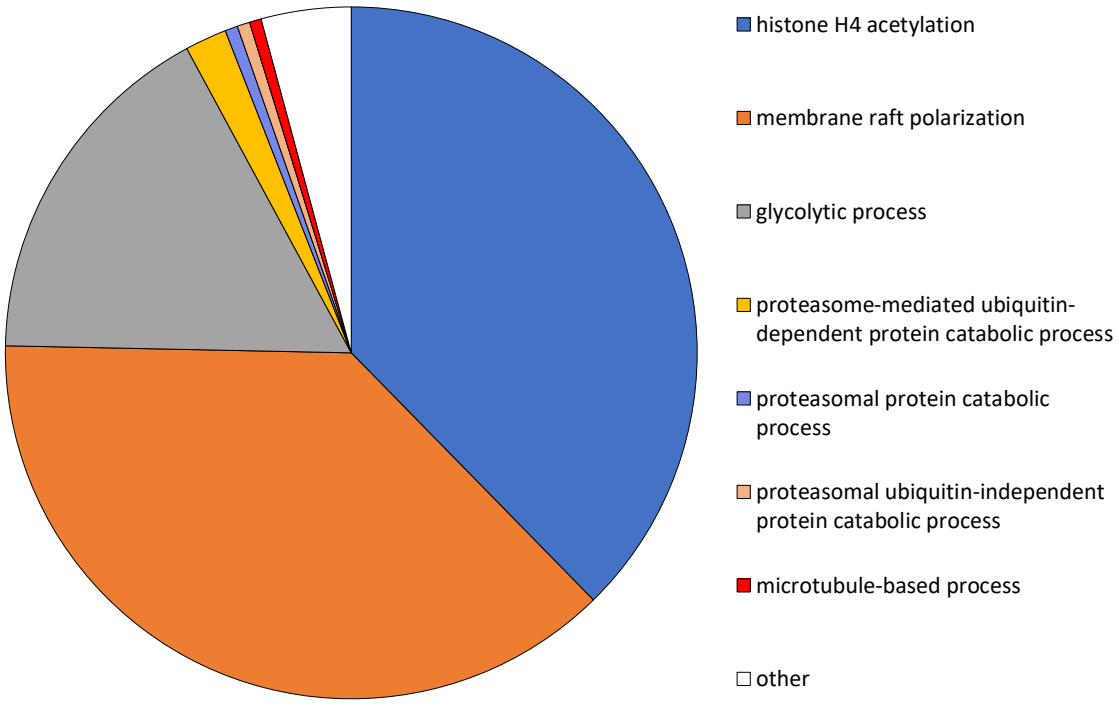


Figure 3.5. Relative abundance of fungal protein biological process (BP) function (determined by summing protein values assigned to the same gene ontology term; *n* = 24). The most relatively abundant BPs were histone H4 acetylation (blue) and membrane raft polarization (orange).

A total of 94% of fungal proteins were assigned to a *KEGG* term using *GhostKOALA*. KEGG term assignments revealed that the majority of protein sequences were assigned to Genetic Information Processing (31.9%). Other prevalent assignments included Protein families: Signaling and Cellular Processes (19.1%), Environmental Information Processing (17.0%), Cellular Processes (12.8%), and Protein families: Genetic Information Processing (8.5%), Carbohydrate Metabolism (6.4%), and Energy Metabolism (4.3%).

A relative abundance plot of *KO* terms showed that the detected fungal proteins consisted primarily of proteins involved in the Cytoskeleton (18.4%), Exosome (10.6%), Membrane trafficking (10.3%), and the Pentose phosphate pathway (9.3%) (Fig 3.6).

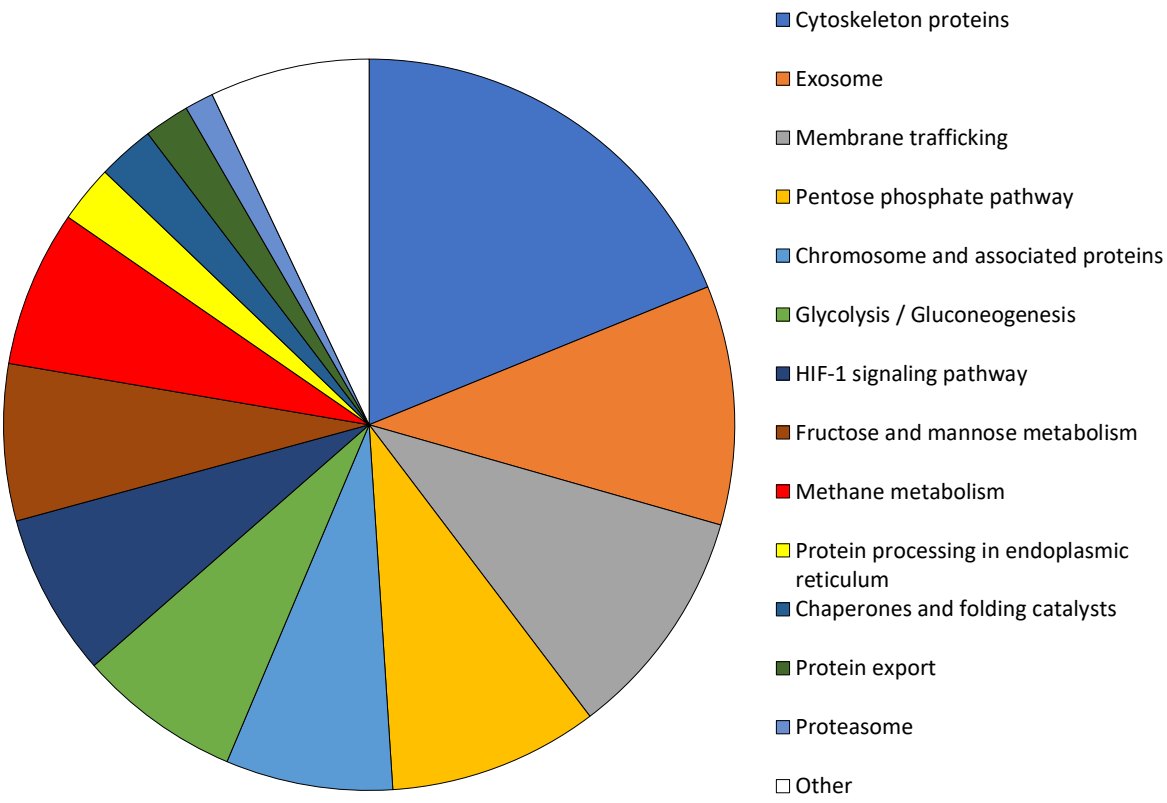


Figure 3.6. Relative Abundance of KEGG Ontology (KO) Terms (determined by summing protein values assigned to the same KO term; *n* = 52). The most relatively abundant KO were cytoskeleton proteins (blue) and exosome proteins (orange).

Bacterial Vaginosis States

An ANOVA test demonstrated that pro-inflammatory cytokines (*p*-value = 1.78e-12) and vaginal pH (*p*-value = 1.04e-08) were significantly associated with BV status and fungal communities. Thus, these variables were considered as confounding variables during t-tests comparing genera, proteins, and BPs between BV states.

Taxonomic Analysis

We compared the relative abundance of fungal genera across different BV states at two different time points. At visit 1, 57 patients were BV positive, 47 were BV negative and 7 were BV intermediate. At visit 2, 37 patients were BV positive, 38 were BV negative, and 5 were BV intermediate.

Candida and *Condiobolus* were the most relatively abundant fungal genera in each BV state (Fig 3.7a). The BV negative group had the largest relative abundance *of Candida* at 69.7% and 68.1% at visit 1 and visit 2, respectively (Fig 3.7a). In the BV intermediate category, the relative abundance of *Candida* was 59.9% at visit 1 and 48.7% at visit 2 (Fig 3.7a). In the BV positive category, the relative abundance of *Candida* was much lower at 36.5% at visit 1, and 32.2% at visit 2 (Fig 3.7a). The total iBAQ protein intensity for *Candida* showed that both BV negative groups had a higher total intensity for this genus compared to the BV positive groups (Fig 3.7b). This result is interesting as the BV negative group at visit 1 had a lower sample size than the BV positive group for visit 1 and both groups had a similar sample size at visit 2. A *limma* t-test established that *Candida* was significantly underabundant in the BV positive group compared to the BV negative group at visit 1 after adjusting for confounding variables (Table 3.4). However, at visit 2 this difference was not found to be significant (*adj p-value* = 0.08).

The BV positive group had a high relative abundance of *Condiobolus* at 33.6% at visit 1 and 33.3% at visit 2 (Fig 3.7a). In the BV negative group, the relative abundance of this genus was comparatively lower (10.7% at visit 1 and 10.4% at visit 2), similar to the BV intermediate group (11.4% at visit 1 and 17.5% visit 2) (Fig 3.7a). The total iBAQ protein intensity for *Condiobolus* confirmed that the BV positive group at both visits had a higher total intensity for this genus (Fig 3.7b). A *limma* test demonstrated that, while *Condiobolus* was overabundant in the BV

positive state (logFC = 1.66; logFC = 3.96) compared to the BV negative state for both visits, this difference was only significant during visit 2 (*adj. p-value* = 0.001) (Table 3.4; Table 3.5).

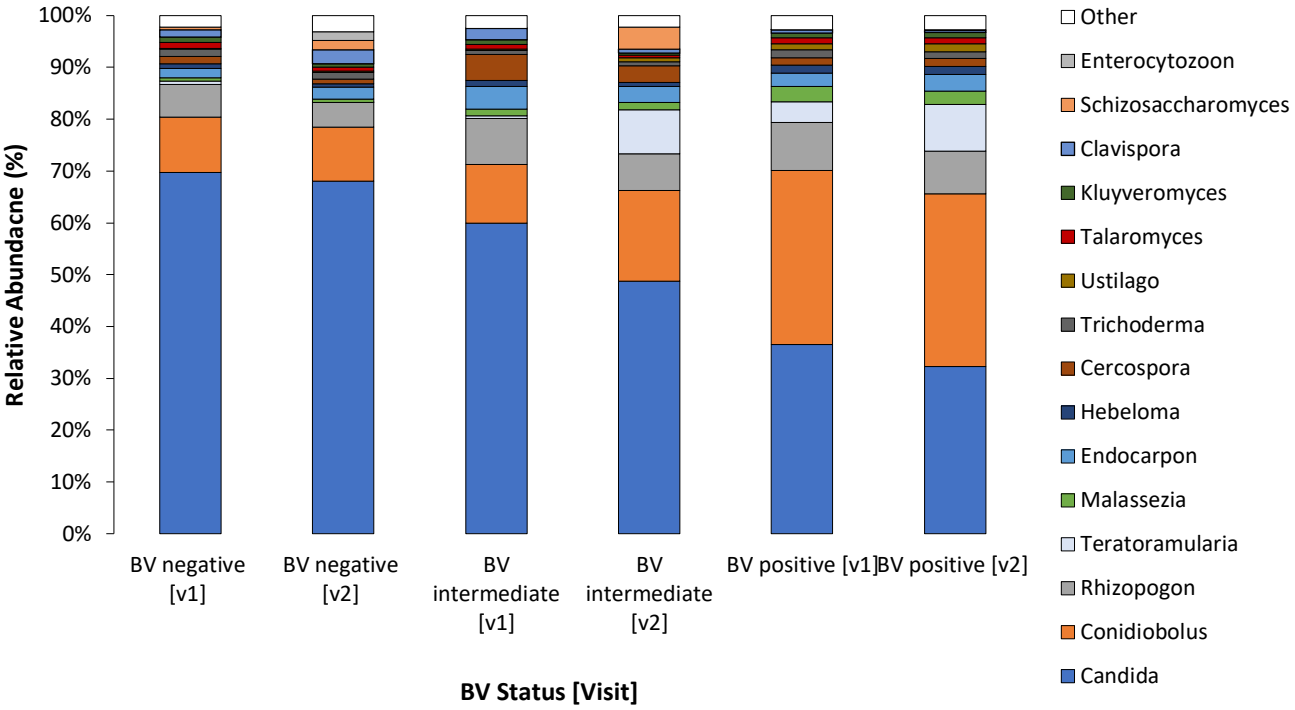
In the BV positive group, the relative abundance of *Rhizopogon* was at 9.3% in visit 1 and 8.2% in visit 2. The BV intermediate group showed a similar relative abundance of this genus (8.8% at visit 1 and 7.1% at visit 2), whereas in the BV negative group this genus was at a lower relative abundance (6.3% at visit 1 and 4.7% at visit 2) (Fig 3.7a). The total iBAQ protein intensity for *Rhizopogon* showed that this genus had the highest total intensity in the BV positive group at visit 1, however, the BV negative group at visit 1 and BV positive group at visit 2 showed a similar total intensity (Fig 3.7b). A *limma* test confirmed that *Rhizopogon* was only significantly underabundant in the BV positive at visit 1 when compared to the BV negative group (Table 3.5). This result may be due to the BV positive group having a larger sample size compared at visit 1 compared to the BV negative group.

In the BV positive group, *Malassezia* had a relative abundance of 3.0% at visit 1 and 2.7% at visit 2. This genus was a much lower relative abundance in the BV negative group at both visits (0.7%), similar to the BV intermediate group (1.3% at visit 1 and 1.4% at visit 2) (Fig 3.7a). The total iBAQ protein intensity for *Malassezia* demonstrated that this genus had higher total intensity in the BV positive groups in comparison to the BV negative groups (Fig 3.7b). According to *limma*, this difference was only found to be significantly overabundant in the BV positive group at visit 1 when compared to the BV negative group (Table 3.4).

The BV positive group (3.92% at visit 1, and 9.04% at visit 2) and BV intermediate group at visit 2 (8.5%) had a comparatively higher relative abundance of *Teratoramularia* than the BV negative (0.6% visit 1; 0.03% visit 2) and BV intermediate group at visit 1 (0.5%) (Fig 3.7a). From the total iBAQ intensity values, this genus was highest in BV positive group compared to the BV negative group (Fig 3.7b). However, the *limma* test did not find this difference to be significant when adjusting for confounding variables.

The remainder of the differentially abundant genera were those that constituted less than 1% of the total fungal relative abundance and were included under the category “Other” found in Supplementary Table S1 (Fig 3.7a). Of these, *Penicillium* and *Torulospora* were differentially overabundant in the BV negative state at visit 1 when compared to the BV intermediate group

(Table 3.3). Similarly, *Penicillium* was differentially overabundant in the BV positive group compared to the BV intermediate group at visit 2 (Table 3.5).



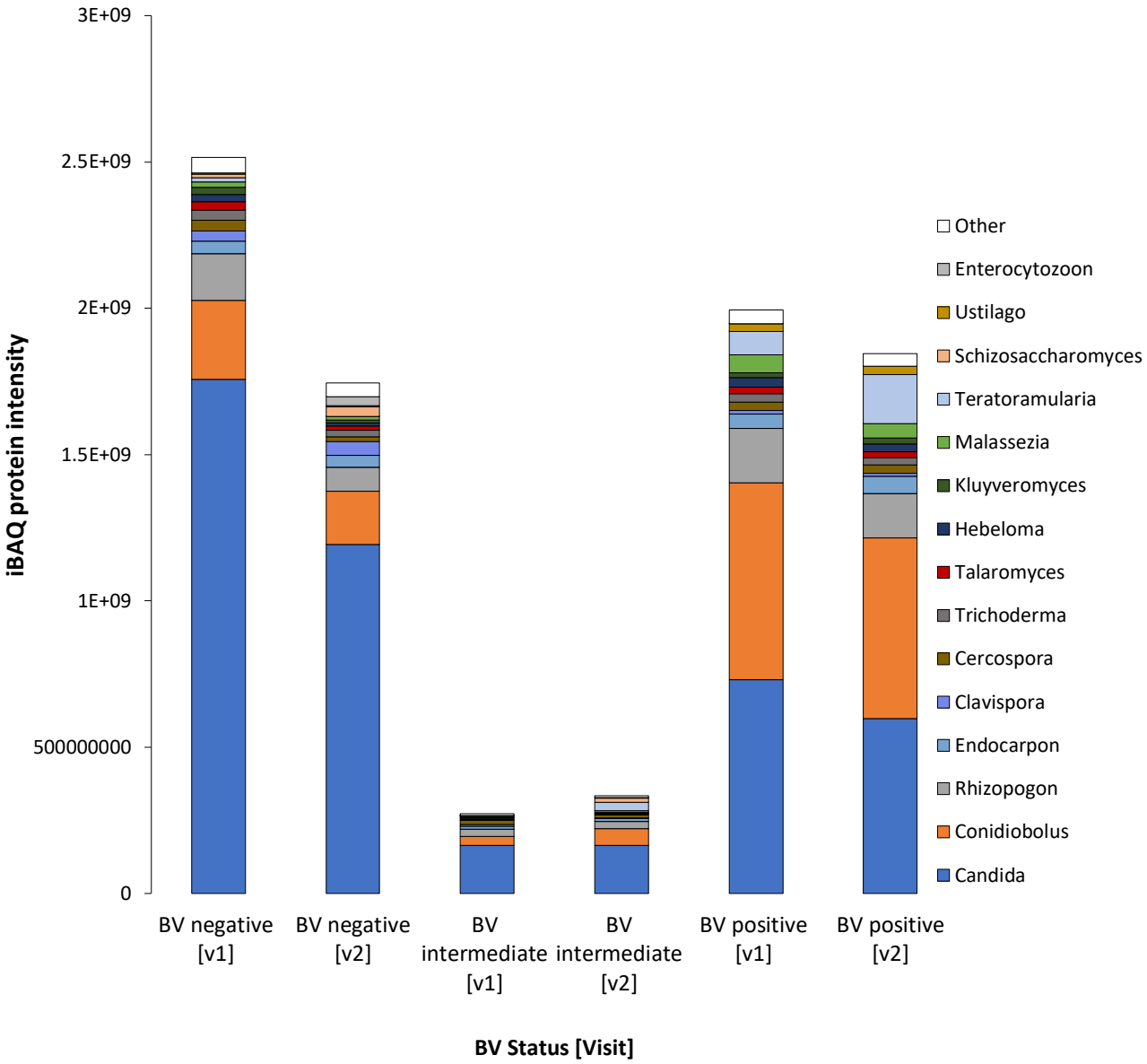


Figure 3.7. **a)** Fungal community composition at the genus level across BV states at different visits: BV negative at visit 1 (n = 47) and visit 2 (n =38); BV intermediate at visit 1 (n = 7) and visit 2 (n = 5); and BV positive at visit 1 (n = 57) and visit 2 (n =37); **b)** total amount of fungal genus by BV state at each visit by summing total fungal iBAQ intensity values of proteins assigning to each fungal genus. Genus assignments were determined using UniProt on protein groups.

Table 3.4. The fungal genera differentially abundant between different BV categories in visit 1 after adjusting for confounding variables (*viz.* vaginal pH and cytokines).

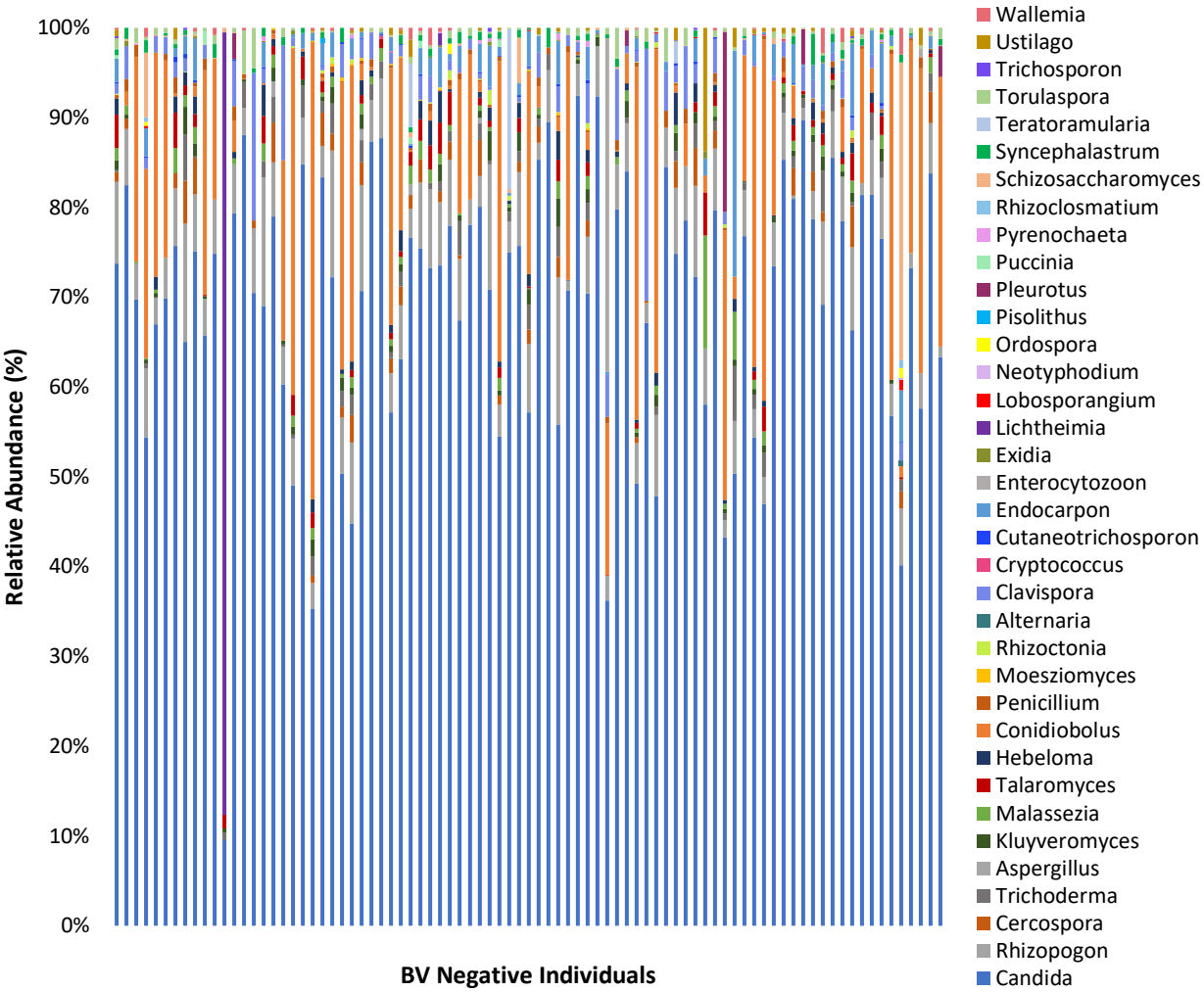
BV Positive – BV Negative (visit 1)			
Genera	**logFC**	***p*-value**	**adj. *p*-value**
Lodderomyces	-0.73	2.19e-05	0.00043
Pisolithus	-0.55	0.00053	0.0052
Alternaria	-1.39	0.00049	0.0052
Rhizopogon	-1.42	0.00069	0.0054
Malassezia	0.96	0.00094	0.0061
Talaromyces	-0.71	0.0013	0.0071
Kluyveromyces	-1.63	0.0017	0.0081
Wallemia	-0.58	0.010	0.045
Candida	-1.33	1.49e-06	5.81e-05
BV Negative – BV Intermediate (visit 1)			
Torulaspora	2.39	0.00026	0.010
Penicillium	2.00	0.0012	0.0234

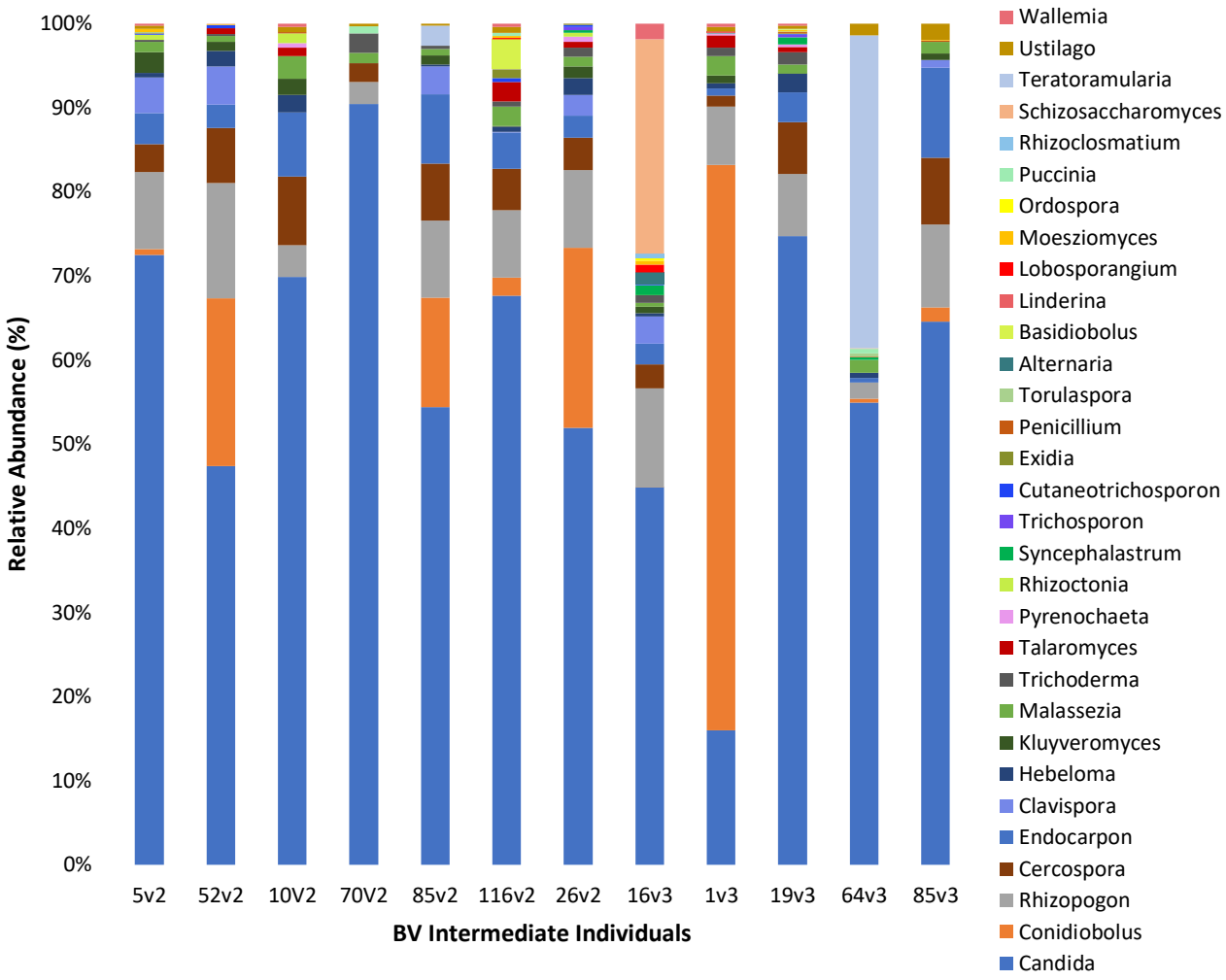
Table 3.5. Fungal genera differentially abundant between different BV categories in visit 2 after adjusting for confounding variables (*viz.* vaginal pH and cytokines).

BV Positive – BV Negative (Visit 2)			
	logFC	***p*-value**	**adj. *p*-value**
Conidiobolus	3.96	3.19e-05	0.0012
BV Positive – BV Intermediate (Visit 2)			
Penicillium	3.07	0.0010	0.039

According to the mycobiome profile of each patient, BV negative and BV intermediate individuals had a high relative abundance of *Candida* (Fig 3.8a). However, two BV negative individuals showed unusual mycobiome profiles, including individual 79v2, which had a high relative abundance of *Lichthiema,* and individual 41v3, which had a high relative abundance of *Schizosaccharomyces* (Fig 3.8a). These profiles were unusual as these genera were found at a low relative abundance in other BV negative individuals (Fig 3.8a). Similarly, two BV intermediate individuals also demonstrated uncommon mycobiome profiles, including individual 1v3 instead, who had a high relative abundance of *Condiobolus* and patient 64v3 had a relatively high abundance of *Teratoramularia,* as both genera were found at a much lower relative abundance across other BV intermediate individuals (Fig 3.8b).

In BV positive patients, the majority of patients have a high relative abundance of *Condiobolus*, but a large number of patients also showed a high relative abundance of *Candida*. Additionally, there were also a few individuals (30v2, 9 (at visit 1 and visit 2), and 2v2) who unusually showed a high relative abundance of *Teratoramularia* (Fig 3.8c).


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Relative Abundance (%)
BV Negative Individuals
Wallemia
Ustilago
Trichosporon
Torulaspora
Teratoramularia
Syncephalastrum
Schizosaccharomyces
Rhizoclosmatium
Pyrenochaeta
Puccinia
Pleurotus
Pisolithus
Ordospora
Neotyphodium
Lobosporangium
Lichtheimia
Exidia
Enterocytozoon
Endocarpon
Cutaneotrichosporon
Cryptococcus
Clavispora
Alternaria
Rhizoctonia
Moesziomyces
Penicillium
Conidiobolus
Hebeloma
Talaromyces
Malassezia
Kluyveromyces
Aspergillus
Trichoderma
Cercospora
Rhizopogon
Candida


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Relative Abundance (%)
5v2
52v2
10V2
70V2
85v2
116v2
26v2
16v3
1v3
19v3
64v3
85v3
BV Intermediate Individuals
Wallemia
Ustilago
Teratoramularia
Schizosaccharomyces
Rhizoclosmatium
Puccinia
Ordospora
Moesziomyces
Lobosporangium
Linderina
Basidiobolus
Alternaria
Torulaspora
Penicillium
Exidia
Cutaneotrichosporon
Trichosporon
Syncephalastrum
Rhizoctonia
Pyrenochaeta
Talaromyces
Trichoderma
Malassezia
Kluyveromyces
Hebeloma
Clavispora
Endocarpon
Cercospora
Rhizopogon
Conidiobolus
Candida

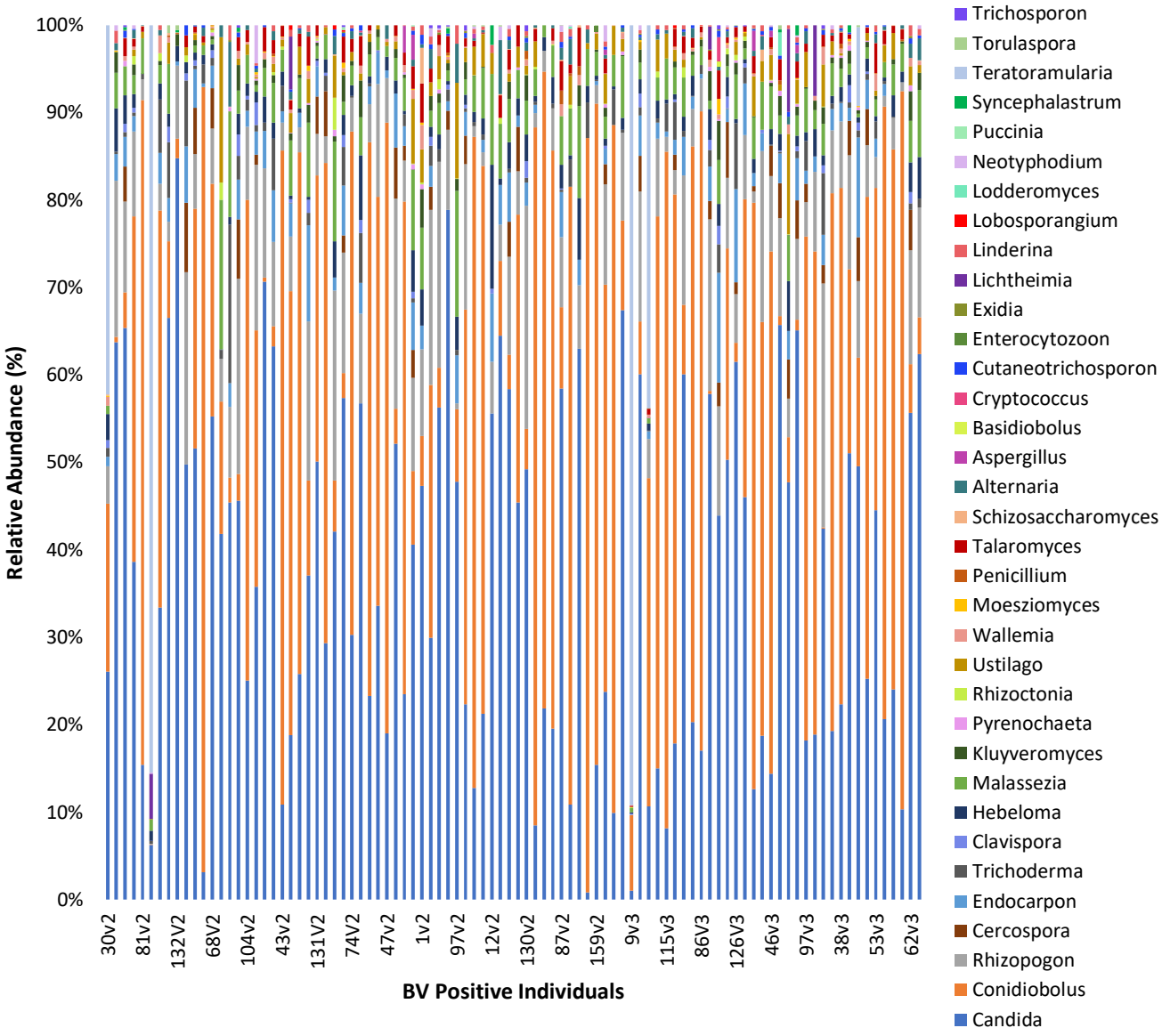


Figure 3.8. Stacked bar plot depicting the fungal composition of each participant in each BV group a) Stacked bar plot depicting the fungal composition of BV negative diagnosed patients at visit 1 and visit 2 (*n* = 95); b) Stacked bar plot depicting the fungal composition of BV intermediate diagnosed patients at visit 1 and visit 2 (*n* = 12); c) Stacked bar plot depicting the fungal composition of BV positive diagnosed patients at visit 1 and visit 2 (*n* = 94).

We performed the *limma* test using iBAQ protein intensities to identify significant differentially abundant (adj. *p-value* < 0.05) proteins between optimal and non-optimal patient groups. The comparison between BV positive and BV negative groups revealed that 12 fungal proteins were significantly different at visit 1 (Table 3.6). However, only one protein, M5EKD5 (Proteasome subunit alpha type, *Malassezia sympodialis*), was overabundant in the BV positive state. At visit 2, 4 fungal proteins were identified as significantly different, 3 fungal proteins were the same as those identified as differentially abundant in visit 1, however,

A0A137P973 (Fructose-bisphosphate aldolase; *Condiobolus coranatus*) was now identified as significantly overabundant in the BV positive state in visit 2.

The comparison between BV negative and BV intermediate groups revealed that 1 fungal protein was significantly different in visit 1 (Table 3.6). The protein G8ZUB2 **(**Exocyst complex component SEC15; *Torulaspora delbrueckii*) was identified as overabundant in the BV negative group.

The comparison between BV positive and BV intermediate samples identified 1 differentially abundant fungal protein at visit 1 and visit 2 (Table 3.6; Table 3.7). The overabundant fungal protein in the BV positive group at visit 1 was C4Y881 (Gp_dh_C domain-containing protein; *Clavispora lusitaniae*). Whereas the overabundant fungal protein in the BV positive group at visit 2 was A0A1V6RS99 (Uncharacterized protein; *Penicillium vulpinum*) (Table 3.7).

Table 3.6. Differentially abundant proteins between different bacterial vaginosis (BV) states at visit 1.

BV Positive – BV Negative (visit 1)					
Protein	**Protein Name**	**Fungal Species**	**logFC**	***p*-value**	**adj. *p*-value**
A5E2Z2	Uncharacterized	*Lodderomyces elongisporus*	-0.73	2.47e-05	0.00042
J7M8M3	Actin	*[Candida] thasaenensis*	-1.12	0.00024	0.0025
A0A177DGX3	Glutamate dehydrogenase	*Alternaria alternata*	-1.39	0.00049	0.0042
A0A0C3J1M2	Uncharacterized	*Pisolithus tinctorius*	-0.55	0.00058	0.0043
A0A1B7MJE7	NAD-binding 6-phosphogluconate dehydrogenase	*Rhizopogon vinicolor*	-1.42	0.00069	0.0044
R9APA2	Calmodulin	*Wallemia ichthyophaga*	-1.14	0.00089	0.0049
M5EKD5	Proteasome subunit alpha type	*Malassezia sympodialis*	0.96	0.00096	0.0049
W0T5M8	Heat shock protein SSA2	*Kluyveromyces marxianus*	-1.64	0.0016	0.0075

B6Q757	Translation elongation factor EF-2 subunit, putative	*Talaromyces marneffei*	-0.60	0.0026	0.011
B6Q4U4	Rab small monomeric GTPase Rab7, putative	*Talaromyces marneffei*	-0.82	0.0047	0.019
H8WXM9	Kar2 protein	*Candida orthopsilosis*	-1.79	4.20e-07	1.07e-05
A0A1D8PFR4	Actin	*Candida albicans*	-2.20	9.68e-08	4.94e-06
BV Negative – BV Intermediate					
G8ZUB2	Exocyst complex component SEC15	*Torulaspora delbrueckii*	2.39	0.00028	0.014
BV Positive – BV Intermediate					
C4Y881	Gp_dh_C domain-containing protein	*Clavispora lusitaniae*	2.16	7.99e-05	0.0041

Table 3.7. Differentially abundant proteins between different bacterial vaginosis (BV) states at visit 2.

BV Positive – BV Negative (Visit 2)					
Protein	**Protein Name**	**Fungal Species**	**logFC**	***p*-value**	**adj. *p*-value**
A0A137P973	Fructose-bisphosphate aldolase	*Conidiobolus coronatus*	3.96	3.20e-05	0.0016
A0A1D8PFR4	Actin	*Candida albicans*	-1.64	9.59e-05	0.0024
J7M8M3	Actin	*[Candida] thasaenensis*	-1.15	0.0029	0.042
M5EKD5	Proteasome subunit alpha type	*Malassezia sympodialis*	1.05	0.0033	0.042

BV Positive – BV Intermediate (Visit 2)					
A0A1V6RS99	Uncharacterized	*Penicillium vulpinum*	3.07	0.00099	0.051

The Nonmetric Multidimensional Scaling (NMDS) plot was used to observe how individuals cluster according to BV status using fungal proteins. The NMDS shows that BV positive and BV negative individuals tended to cluster into their respective groups with a few BV positive outliers (Fig 3.9a; Fig 3.9b). Individuals belonging to the BV intermediate group tended to cluster on the BV negative distribution of the plot (Fig 3.9a; Fig 3.9b). An ANOSIM test confirmed that fungal proteins were similar with some differences in visit 1 (R = 0.2413, p = 0.001) and visit 2 (R = 0.1695, p = 0.001) across BV groups.

The NMDS analysis revealed that 33 fungal proteins were driving the distribution of individuals according to BV status in visit 1 and 25 fungal proteins during visit 2. Of these proteins, A0A1B7MJE7 (NAD-binding 6-phosphogluconate dehydrogenase, *Rhizopogon vinicolor*), A0A1D8PFR4 (Actin, *Candida albicans*), H8WXM9 (kar2 protein; *Candida orthopsilosis*), W0T5M8 (Heat shock protein SSA2; Kluyveromyces marxianus (*Candida kefyr*)), and M5EKD5 (Proteasome subunit alpha type, Malassezia *sympodialis*) accounted for over 41% of the distribution.

An ANOVA test revealed that the p-value was not significant (p-value = 0.69), hence group dispersions did not differ between BV groups (average distance to the median is similar across all BV groups) and appeared to be similar. We then tested if compositions were different using an *Adonis* test, which revealed that BV states had significantly different compositions (p = 1e-04) at visits 1 and 2. Therefore, BV groups present homogeneity among group dispersions, while having significantly different compositions. Adonis revealed that approximately 9.1% and 7.8% of the variation in distances was explained by grouping the samples by BV status at visits 1 and 2, respectively.

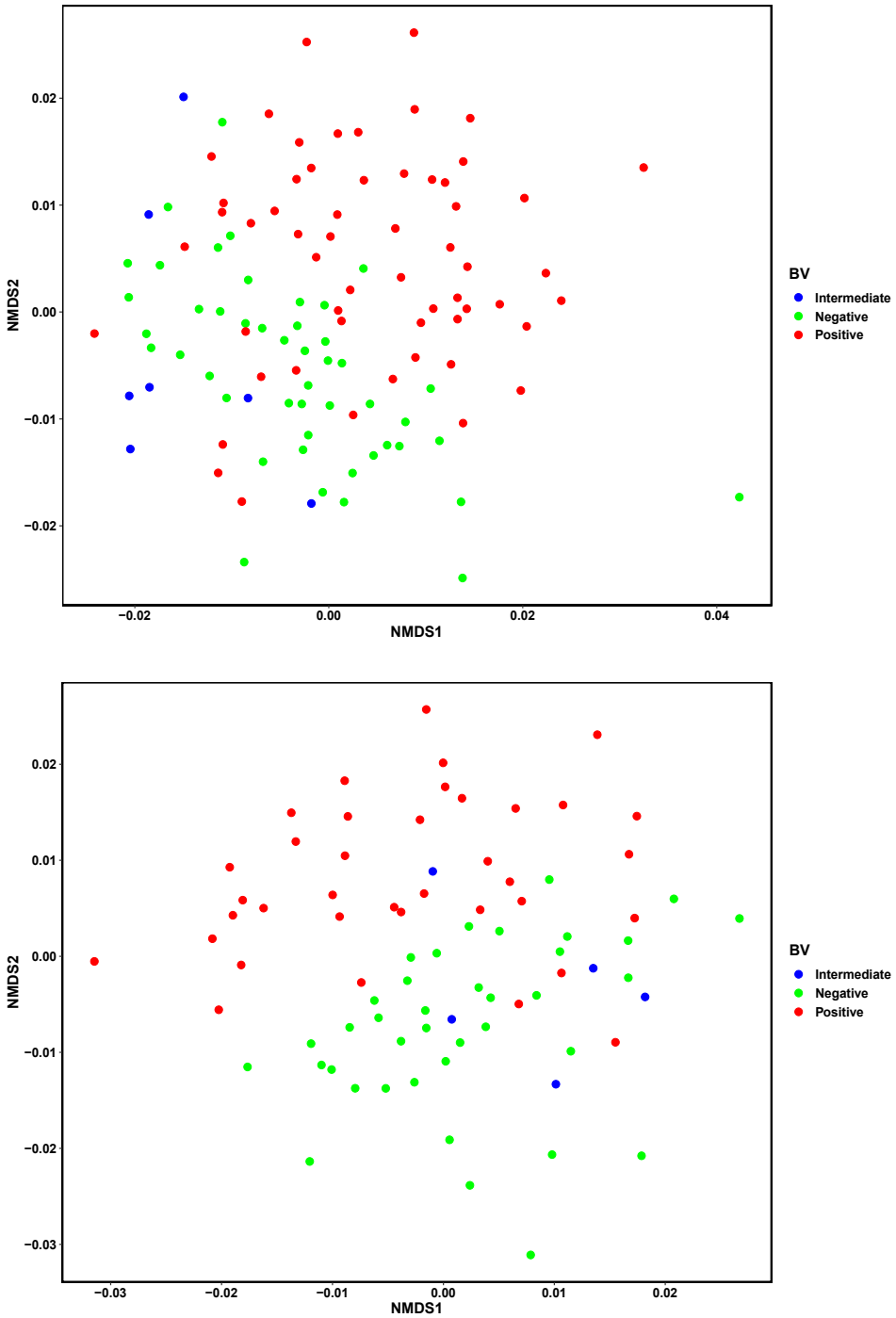


Figure 3.9. **a)** Nonmetric Multidimensional Scaling (NMDS) analysis was used to group individuals based on the log2-transformed imputed intensity-based absolute quantification (iBAQ) intensities of fungal proteins for visit 1; **b)** NMDS analysis was used to group individuals based on the log2-transformed imputed iBAQ intensities of fungal proteins for visit 2

Hierarchical clustering according to fungal protein iBAQ intensity values confirmed that individuals showed some clustering according to BV status (Fig 3.10). The first cluster was a mixture of BV states, the second cluster was majority BV positive, and the third cluster was

majority BV negative (Fig 3.10). Four proteins, A0A0C3J1M2 (Uncharacterized protein, *Pisolithus tinctorius*), A5E2Z2 (Uncharacterized protein, *Lodderomyces elongisporus*), B6Q757 (Translation elongation factor EF-2, *Talaromyces marneffei*), and A0A1D8PFR4 (Actin, *Candida albicans*), were consistently highly expressed across clusters (Fig 3.10). A0A1D8PFR4 showed the highest expression of all the detected fungal proteins, indicated by the dark red. The BV negative dominant cluster showed the highest expression of the majority of proteins in comparison to the other clusters evident by the red banding pattern. Whereas, two proteins, I4Y633 (ATPase, *Wallemia mellicola*) and A0A0J0XQ73 (Heat shock protein 70, *Cutaneotrichosporon oleaginosum*), were consistently detected at low levels across clusters (Fig 3.10). Changes in expression patterns are visible between fungal proteins that were differentially abundant across BV positive and BV negative states according to *limma* tests.

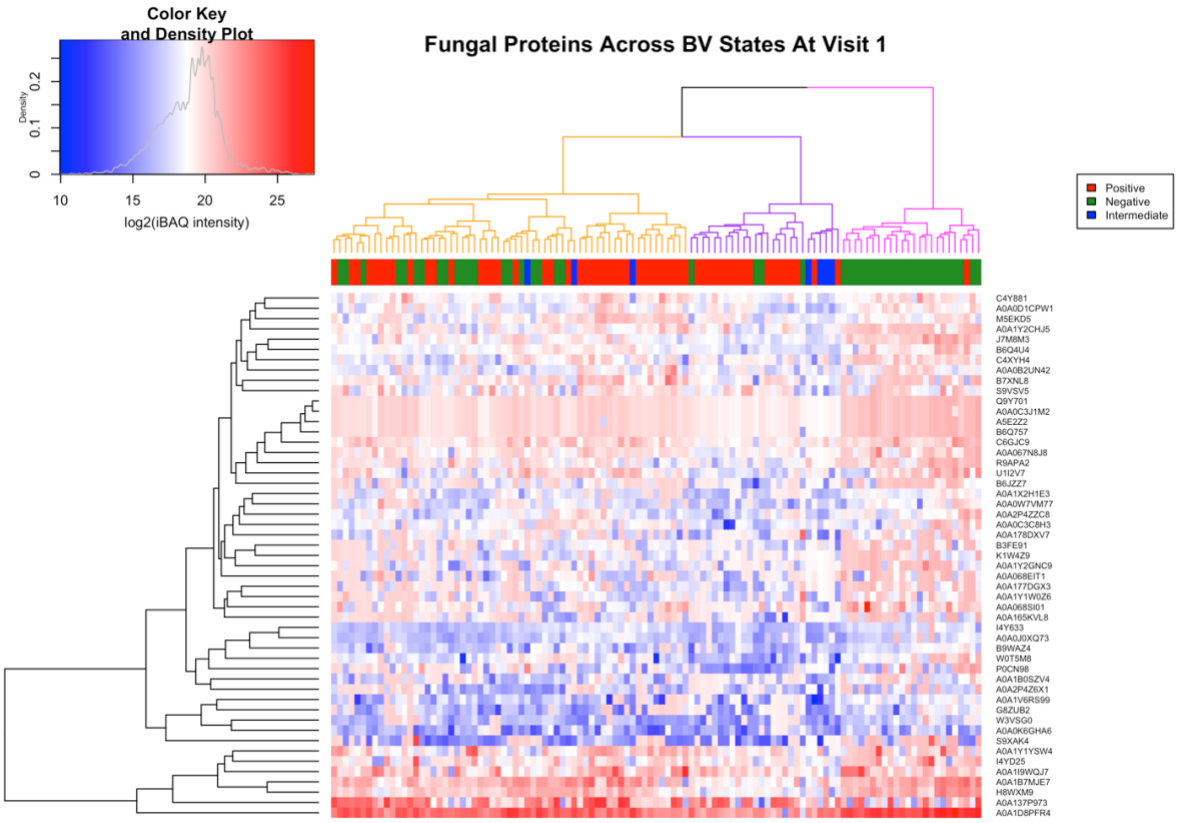


Figure 3.10. Hierarchical clustering was used to group samples based on the log2-transformed imputed intensity-based absolute quantification (iBAQ) values of fungal proteins at visit 1. Red indicated the highest expression, and blue indicated the lowest expression. The darker the red banding pattern, the higher the IBAQ intensity value. The darker the blue banding pattern, the lower the IBAQ intensity value. Samples showed some clustering according to BV status indicated by the top dendrogram.

Functional Analysis

A comparison of the relative abundance between *GO* BPs across different BV groups revealed that both histone H4 acetylation, and membrane raft polarization were the most relatively abundant BPs across BV negative and BV intermediate samples (Fig 3.11). In the BV negative group, both BPs cumulatively made up over 80% of the relative abundance (43.9% v1; 42.4% v2), similar to the BV intermediate group at visit 1 (41.1%) (Fig 3.11). Comparatively, the BV intermediate group at visit 2 (33.8%) and in the BV positive group (29.2% v1; 27.2% v2) had a lower relative abundance of this BP (Fig 3.11). A *limma* test confirmed that both histone H4 acetylation, and membrane raft polarization were significantly differentially underabundant in the BV positive group across both visits compared to the BV negative groups (Table 3.8).

In the BV positive group, the glycolytic process had the highest relative abundance (31.3% v1; 33.7% v2) (Fig 3.11). However, in the BV negative (7.6% v1; 7.2% v2) and BV intermediate (8.7% v1; 13.5% v2) groups this BP was found at a much lower relative abundance. A *limma* test confirmed that the glycolytic process was significantly differentially overabundant in the BV positive group at visit 2 (Table 3.9)

In the BV positive group, proteasome-mediated ubiquitin-dependent protein catabolic process had a relative abundance of 4.0% at visit 1 and 4.2% at visit 2 (Fig 3.11). However, comparatively in the BV negative (0.6% v1; 0.7% v2) and BV intermediate group (1.3% v1; 1.7% v2) the relative abundance was lower (Fig 3.11). A *limma* test revealed that the Proteasome-mediated ubiquitin-dependent protein catabolic process was significantly differentially overabundant in the BV positive category in comparison to BV negative category in visit 2 (Table 3.9).

In total, the *limma* test revealed that 9 BPs and 10BPs were significantly differentially abundant between the BV positive and BV negative group at visit 1 and visit 2, respectively. Of these BPs, only 2 were overabundant in the BV positive group during visit 2, which included the glycolytic process and Proteasome mediated ubiquitin dependent protein catabolic process (Table 3.9).

Two BPs were overabundant in the BV negative group compared to the BV intermediate group at visit 1, including Golgi to plasma membrane transport and vesicle docking involved in

exocytosis (Table 3.8). No BPs were found to be significantly differentially abundant between the BV positive and BV intermediate group.

BPs that constituted less than 1% of the relative abundance across BV states can be found in Supplementary Table S4.

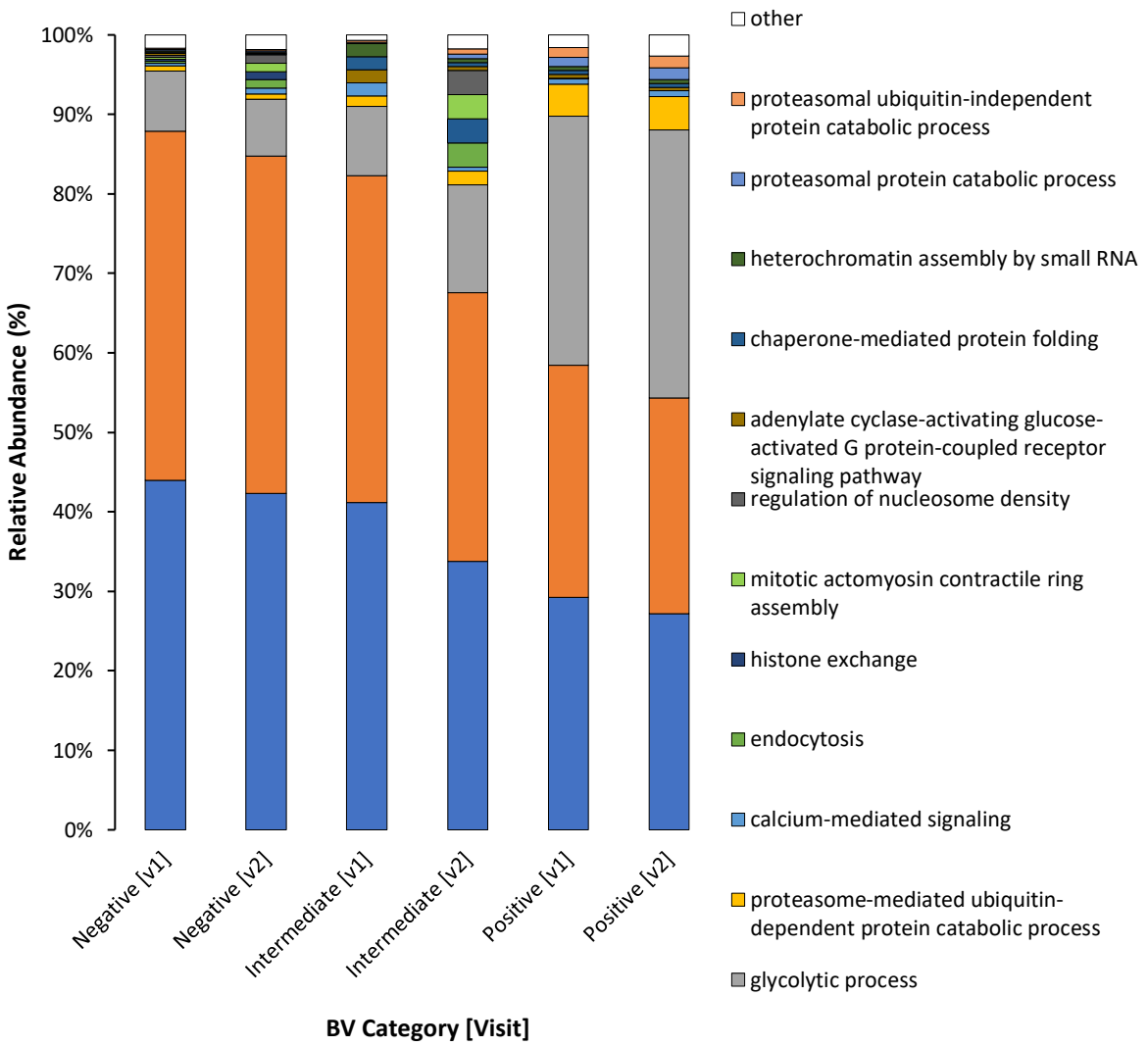


Figure 3.11. Relative abundance of biological processes (BP) according to Gene Ontology determined using UniProt across bacterial vaginosis (BV) states at different visits: BV negative at visit 1 (*n* = 47) and visit 2 (*n* =38); BV intermediate at visit 1 (*n* = 7) and visit 2 (*n* = 5); and BV positive at visit 1 (*n* = 57) and visit 2 (*n* =37).

Table 3.8. Differentially abundant biological processes between different bacterial vaginosis (BV) states at visit 1.

BV positive – BV negative (visit 1)			
Biological Process	**logFC**	***p*-value**	**adj. *p*-value**
Histone H4 acetylation	-2.20	9.40 e-08	1.41e-06
Membrane raft polarization	-2.20	9.40 e-08	1.41e-06
Cell adhesion	-0.73	1.03e-05	4.41e-05
Cellular response to heat	-0.73	1.03e-05	4.41e-05
Cellular response to osmotic stress	-0.73	1.03e-05	4.41e-05
Cellular response to oxidative stress	-0.73	1.03e-05	4.41e-05
Plasma membrane organization	-0.73	1.03e-05	4.41e-05
Cellular amino acid metabolic process	-1.39	0.00048	0.0018
Calcium mediated signaling	-0.82	0.00069	0.0023
BV negative – BV intermediate (visit 1)			
Golgi to plasma membrane transport	2.39	0.00024	0.0036
Vesicle docking involved in exocytosis	2.39	0.00024	0.0036

Table 3.9. Differentially abundant biological processes between the bacterial vaginosis (BV) positive and BV negative state at visit 2.

BV positive – BV negative (visit 2)			
Biological Process	**logFC**	***p*-value**	**adj. *p*-value**
Glycolytic process	3.96	3.93e-05	0.00083
Histone H4 acetylation	-1.64	8.33e-05	0.00083
Membrane raft polarization	-1.64	8.33e-05	0.00083
Calcium mediated signaling	-0.73	0.0082	0.03
Cell adhesion	-0.52	0.0085	0.03
Cellular response to heat	-0.52	0.0085	0.03
Cellular response to osmotic stress	-0.52	0.0085	0.03

Cellular response to oxidative stress	-0.52	0.0085	0.03
Plasma membrane organization	-0.52	0.0085	0.03
Proteasome mediated ubiquitin dependent protein catabolic process	0.81	0.016	0.05

BV categories showed some clustering using fungal BPs as indicated by a mixed cluster, a more dominant BV positive cluster, and a dominant BV negative cluster (Fig 3.12). Membrane raft polarization and histone H4 acetylation were highly expressed across all clusters, with the BV negative dominant cluster having the highest expression indicated by the darker red banding pattern (Fig 3.12). The microtubule-based process, IRE1 mediated unfolded protein response, and translation were expressed at low levels across all clusters. Changes in expression patterns are visible between BPs that were differentially abundant between BV positive and BV negative states according to *limma*.

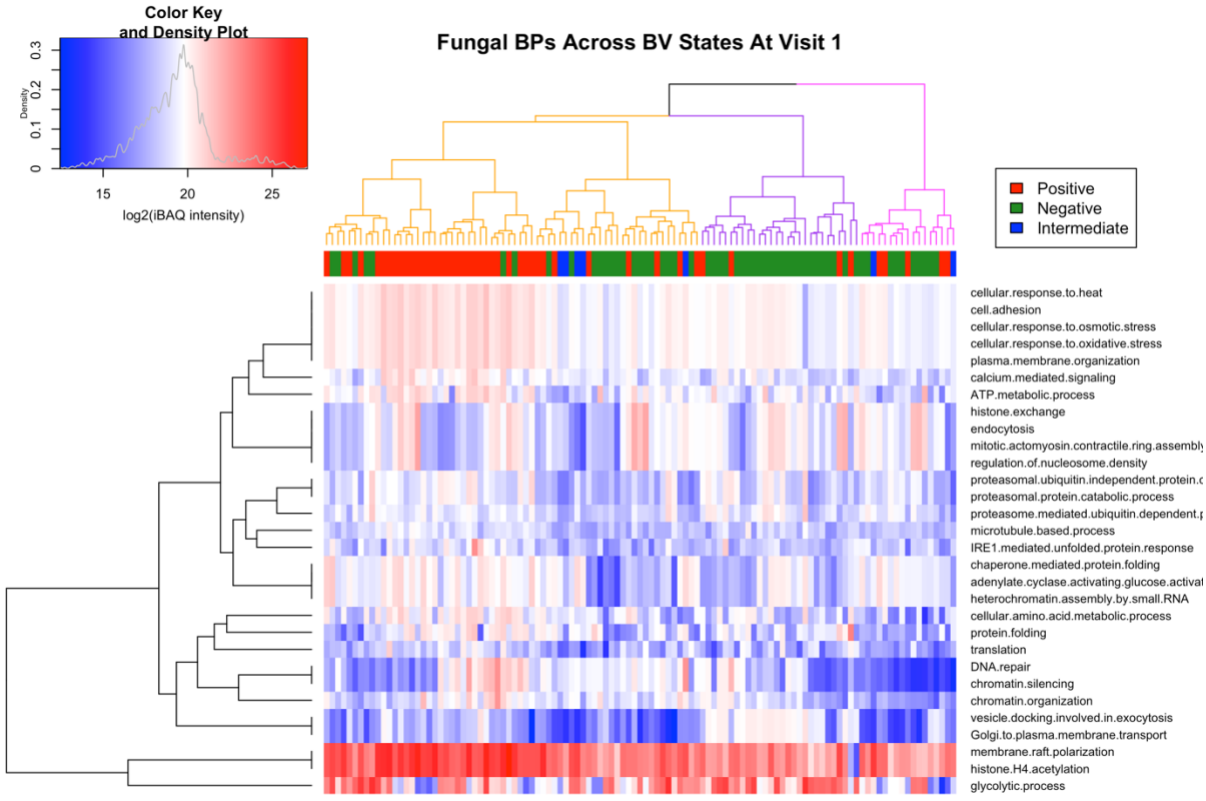


Figure 3.12. Unsupervised hierarchical clustering was used to group individuals using the log2-transformed imputed intensity-based absolute quantification (iBAQ) values of fungal biological processes according to gene ontologies (GO).

Differences between Inflammation Levels

An ANOVA test demonstrated that BV status (*p*-value = 4.96e-07) and vaginal pH (*p*-value = 2.49 e-10) had a significant associated with inflammation. These variables were considered as confounding variables during t-tests comparing inflammation categories.

We made comparisons between the relative abundance of fungal genera across different inflammation groups across two visits. In visit 1, 36 patients were classified as lowly inflamed, 49 patients as moderately inflamed, and 28 patients as highly inflamed. In visit 2, 22 patients were classified as lowly inflamed, 28 patients as moderately inflamed, and 34 patients as highly inflamed.

As seen with BV, *Candida* was the most relatively abundant genus in each inflammation category, followed by *Condiobolus*. The low inflammation state had the largest relative abundance of *Candida* at both visits (64.2% visit 1; 70.1% visit 2) (Fig 3.13a). In the moderate inflammation state, the relative abundance of *Candida* was at 52.0% at visit 1 and 38.9% at visit 2, and in the high inflammation state, the relative abundance of this genus was at 55.7% at visit 1 and 47.1% at visit 2 (Fig 3.13a). When looking at the total iBAQ protein intensity for *Candida* it was highest in the medium inflammation group at visit 1, however, this group had a larger sample size in comparison to other groups, which may explain the higher total iBAQ protein value (Fig 3.13b). Whereas the medium inflammation group at visit 2 had the lowest total iBAQ protein intensity for *Candida*, despite having a larger sample size than the low inflammation group at visit 1 and having the sample size as the high inflammation group at visit 1 (Fig 3.13b).

The moderate inflammation state had a large relative abundance of *Condiobolus* at both visits (24.0% visit 1; 38.7% visit 2) (Fig 3.13a). The low inflammation (17.3% visit 1; 16.6% visit 2) and high inflammation group (16.6% visit 1; 16.8% visit 2) had a comparatively lower abundance of this genus (Fig 3.13a). Correspondingly, the total iBAQ protein intensity for *Condiobolus* was highest in the medium inflammation group at both visits (Fig 3.13b).

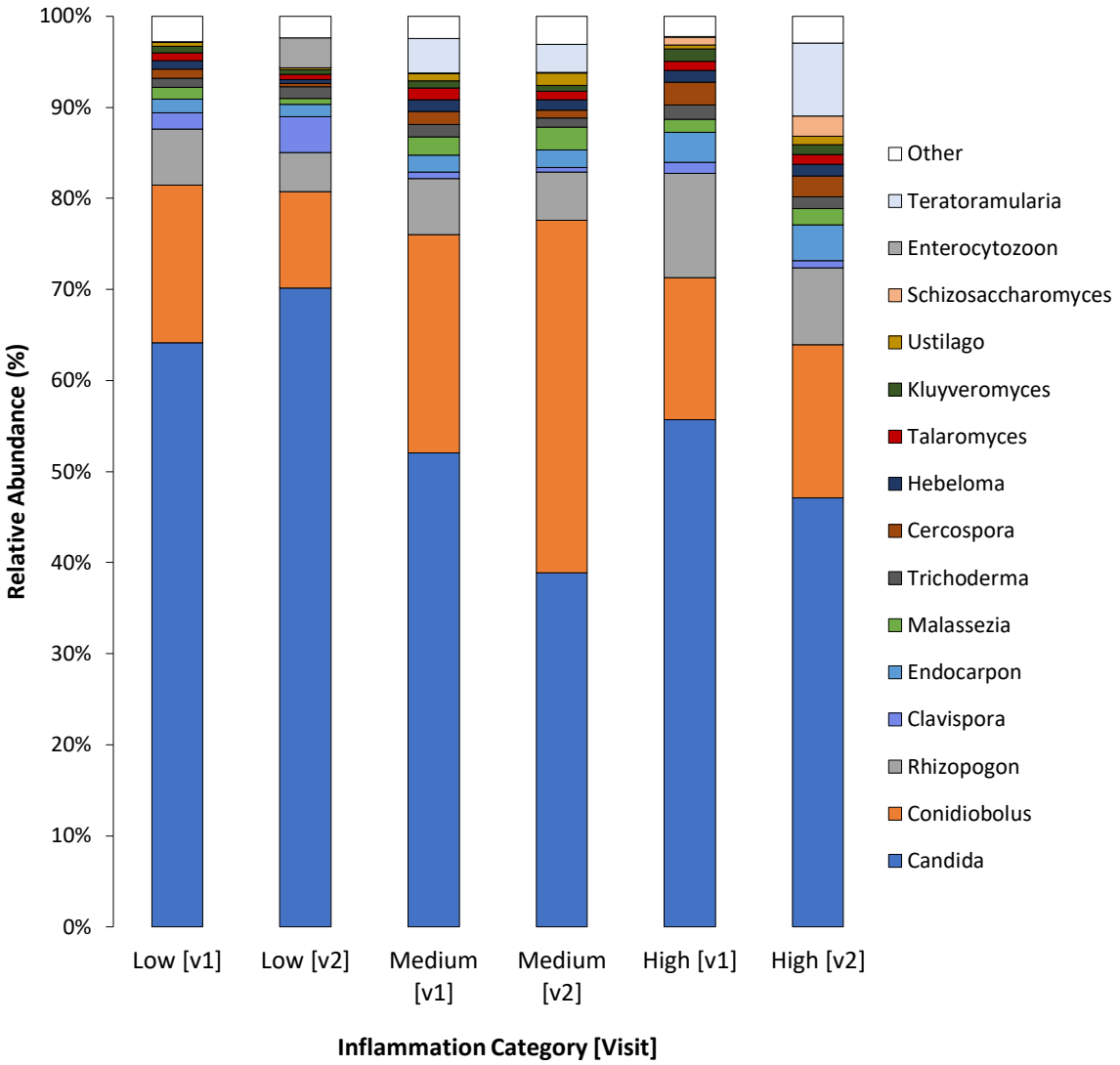
The low inflammation group had a high relative abundance of *Enterocytozoon* in visit 2 (3.3%; Fig 3.13a). Accordingly, this genus had the highest total iBAQ protein intensity value in

comparison to other inflammation categories (Fig 3.13b). Likewise, this group had the highest relative abundance of *Clavispora* (3.9%) at visit 2 (Fig 3.13a).

The high inflammation category had a high relative abundance of *Rhizopogon* (11.4% visit 1; 8.4% visit 2), whereas the low inflammation (6.1% visit 1; 4.3% visit 2) and medium inflammation (6.1% visit 1; 5.3% visit 2) groups had a comparatively lower relative abundance of this genus (Fig 3.13a). The total iBAQ protein intensity for *Rhizopogon* demonstrated this genus had the highest value in the high inflammation group at both visits (Fig 3.13b). In addition, the medium inflammation group at visit 1 had a similar total intensity value. However, the total iBAQ value of *Rhizopogon* for the low inflammation group at visit 1 was lower despite this genus having the same relative abundance as the medium inflammation group at visit 1. This is most likely due to the smaller sample size of the low inflammation group at visit 1. In addition, the high inflammation category at visit 2 had a high relative abundance of *Teratoramularia* (8.0%) (Fig 3.13a). The total iBAQ value for this genus was the highest in this category (Fig 3.13b).

No genera were found to be significantly different between inflammation categories after being adjusted for confounding variables.

Fungal genera that had a relative abundance of less than 1% across each inflammation group have been summed and depicted on the relative abundance plot under the category “Other” and these genera can be found in Supplementary Table S4 (Fig 3.13a).


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Relative Abundance (%)
Low [v1]
Low [v2]
Medium [v1]
Medium [v2]
High [v1]
High [v2]
Inflammation Category [Visit]
Other
Teratoramularia
Enterocytozoon
Schizosaccharomyces
Ustilago
Kluyveromyces
Talaromyces
Hebeloma
Cercospora
Trichoderma
Malassezia
Endocarpon
Clavispora
Rhizopogon
Conidiobolus
Candida

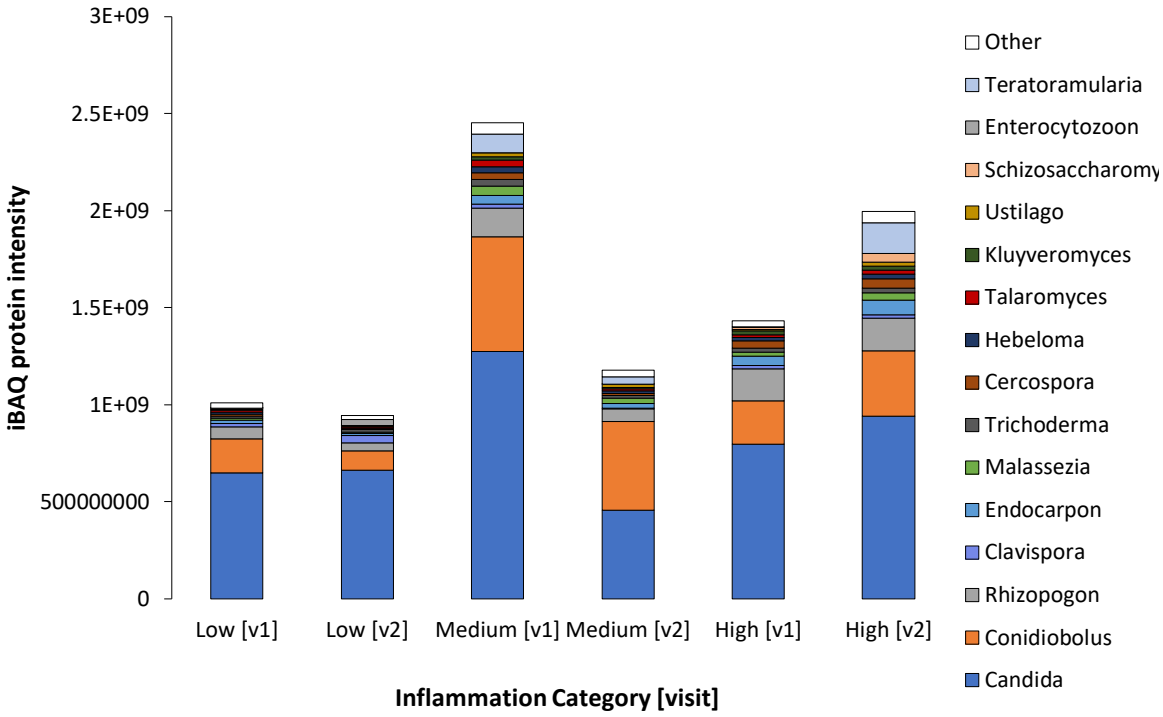
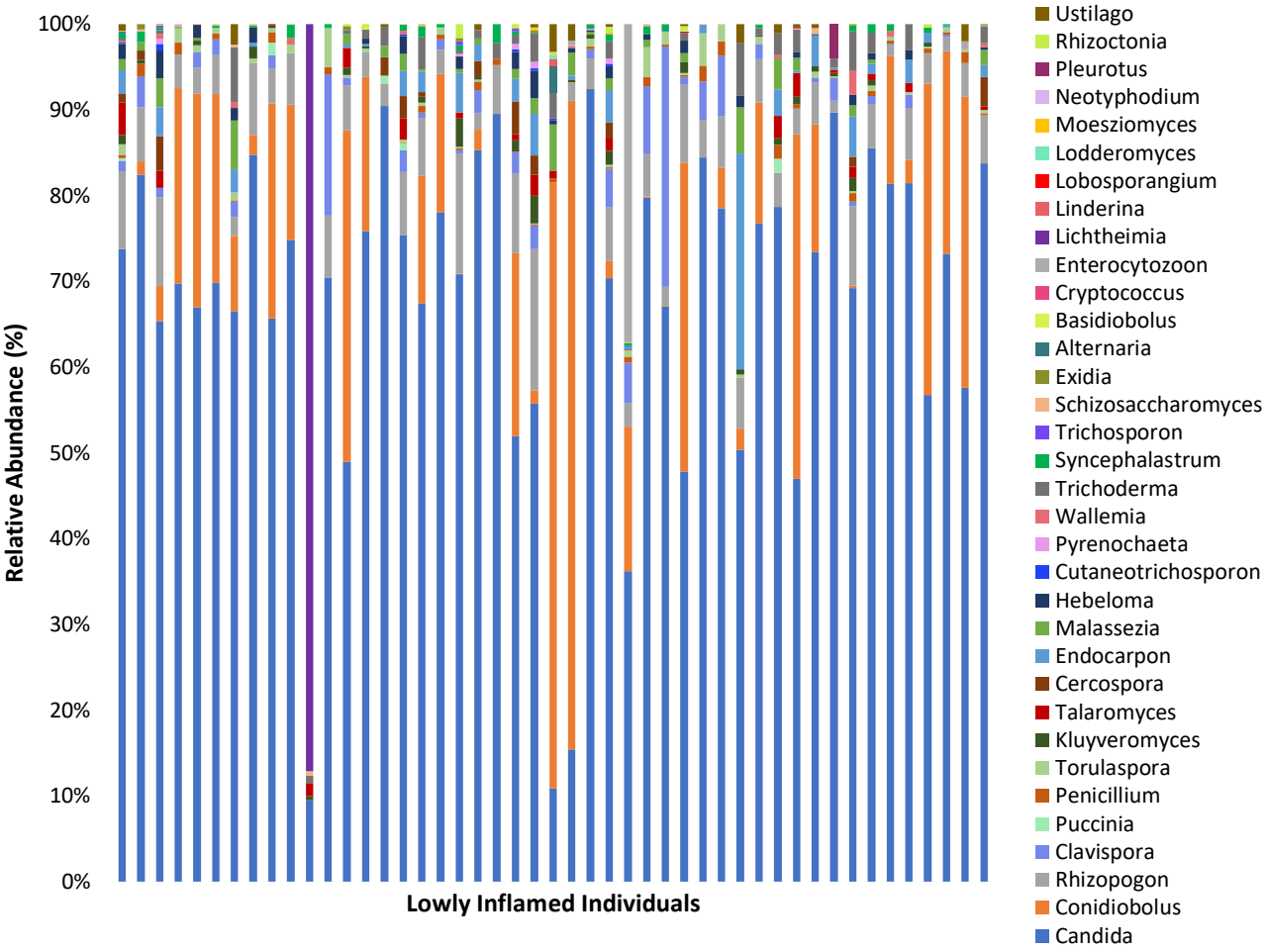
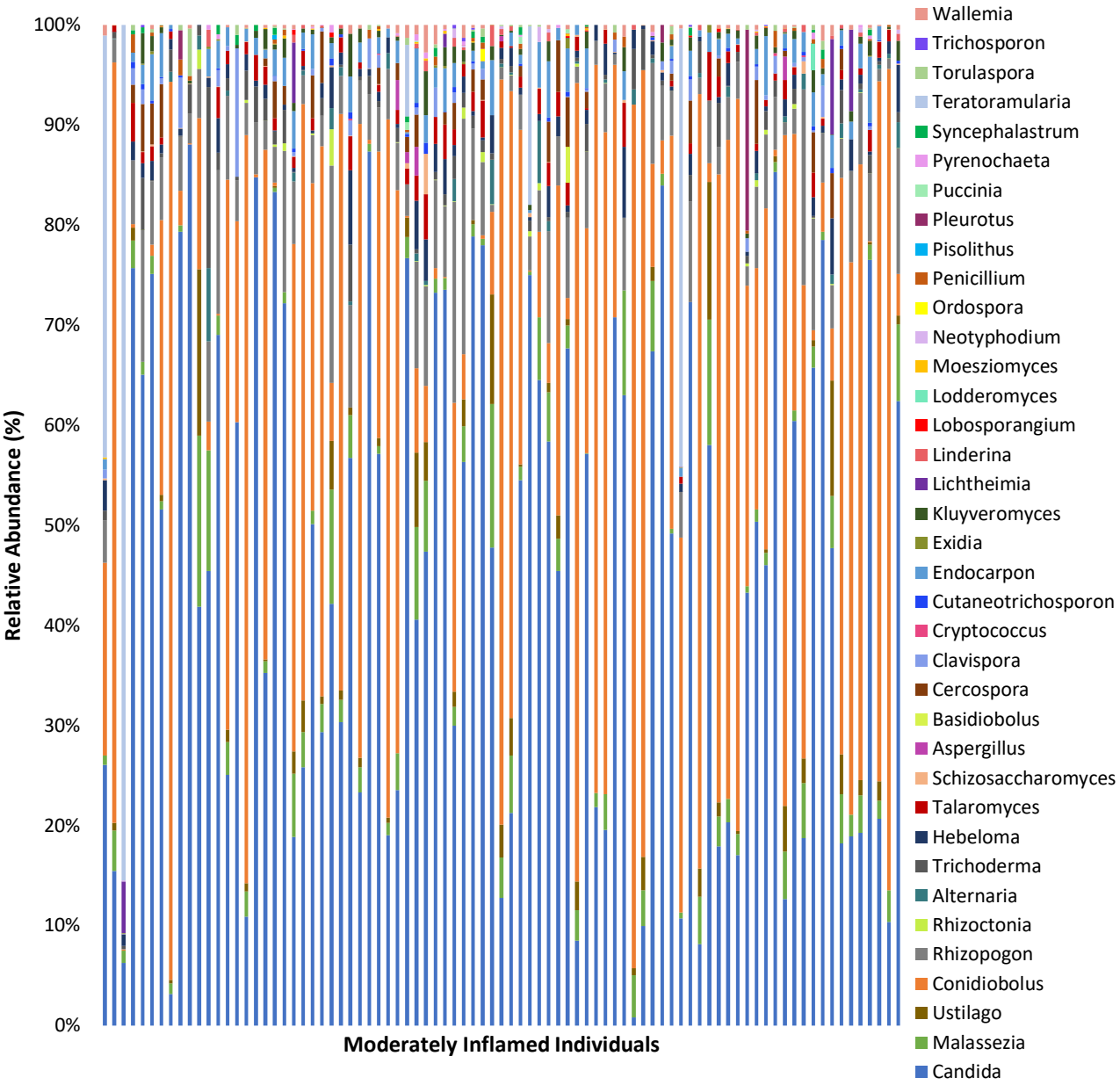


Figure 3.13. a) Fungal community composition at the genus level across Inflammation states at two different visits: lowly inflamed at visit 1 (*n* = 36) and visit 2 (*n* = 22); moderately inflamed at visit 1 (*n* = 49) and visit 2 (*n* = 28); highly inflamed at visit 1 (*n* = 28) and visit 2 (*n* = 34); b) Total fungal iBAQ protein intensity by summing fungal proteins assigned to each genus across different inflammation states at different visits.

According to the mycobiome profile of each patient, in lowly inflamed and highly inflamed individuals, the majority had a high relative abundance of *Candida* (Fig 3.14a). However, three lowly inflamed individuals showed unusual mycobiome profiles, including individual 79v2, which had a high relative abundance of *Lichthiema,* and individuals 94v2 and 158v2, which had a high relative abundance of *Condiobolus* (Fig 3.14a). This was unusual as these genera were found at a low relative abundance in other lowly inflamed individuals (Fig 3.14a).

In moderately inflamed individuals, *Candida* and *Condiobolus* were both highly relatively abundant, with the exception of individuals 30v2, 9v2, and 2v3, which had a high relative abundance of *Teratoramularia* (Fig 3.14b).


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Relative Abundance (%)
Lowly Inflamed Individuals
Ustilago
Rhizoctonia
Pleurotus
Neotyphodium
Moesziomyces
Lodderomyces
Lobosporangium
Linderina
Lichtheimia
Enterocytozoon
Cryptococcus
Basidiobolus
Alternaria
Exidia
Schizosaccharomyces
Trichosporon
Syncephalastrum
Trichoderma
Wallemia
Pyrenochaeta
Cutaneotrichosporon
Hebeloma
Malassezia
Endocarpon
Cercospora
Talaromyces
Kluyveromyces
Torulaspora
Penicillium
Puccinia
Clavispora
Rhizopogon
Conidiobolus
Candida


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Relative Abundance (%)
Moderately Inflamed Individuals
Wallemia
Trichosporon
Torulaspora
Teratoramularia
Syncephalastrum
Pyrenochaeta
Puccinia
Pleurotus
Pisolithus
Penicillium
Ordospora
Neotyphodium
Moesziomyces
Lodderomyces
Lobosporangium
Linderina
Lichtheimia
Kluyveromyces
Exidia
Endocarpon
Cutaneotrichosporon
Cryptococcus
Clavispora
Cercospora
Basidiobolus
Aspergillus
Schizosaccharomyces
Talaromyces
Hebeloma
Trichoderma
Alternaria
Rhizoctonia
Rhizopogon
Conidiobolus
Ustilago
Malassezia
Candida

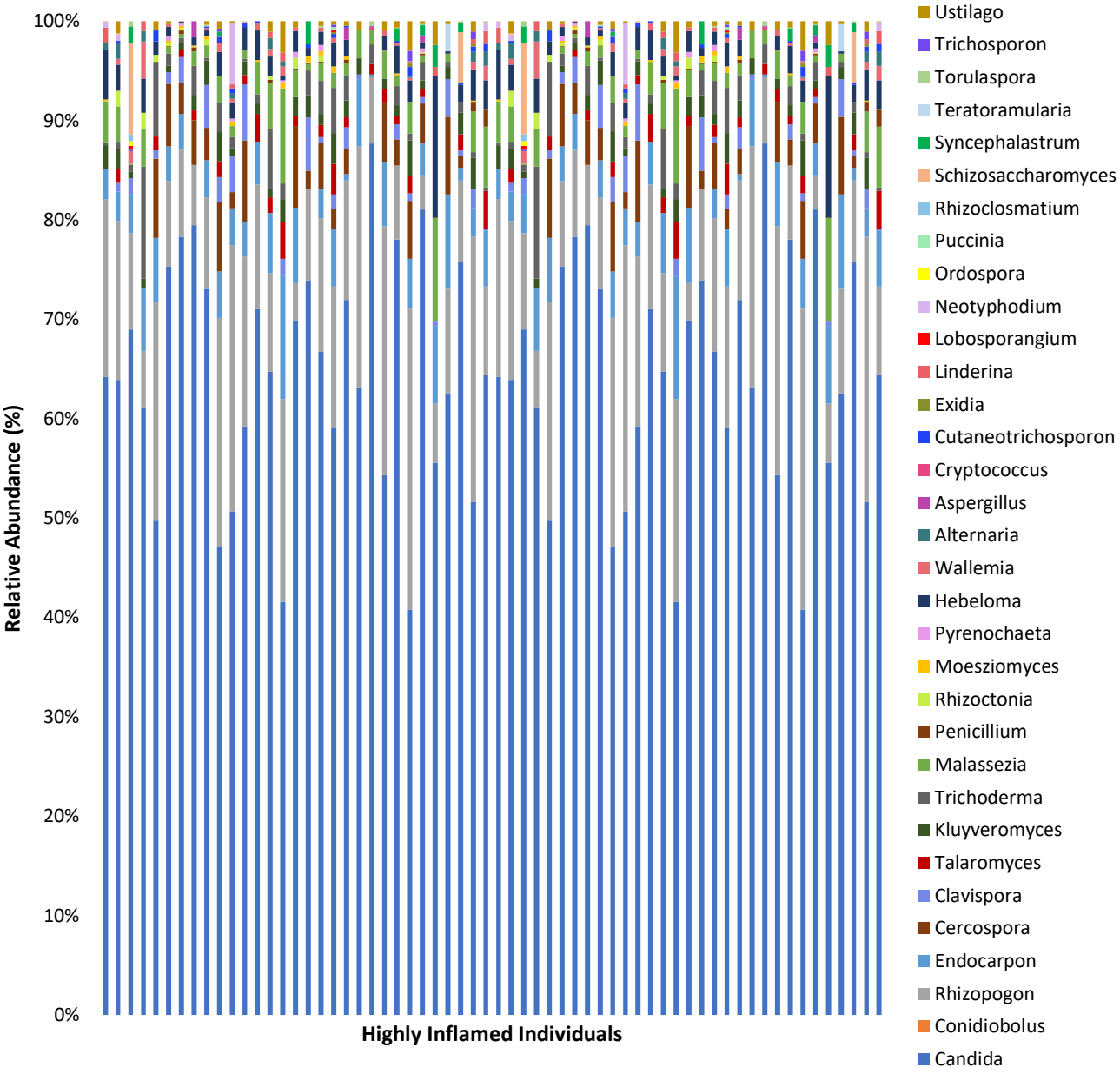


Figure 3.14. Stacked bar plot depicting the fungal composition of each participant in each inflammation group a) Stacked bar plot depicting the fungal composition of lowly inflamed individuals at visit 1 and visit 2 (*n* = 58); b) Stacked bar plot depicting the fungal makeup of moderately inflamed individuals at visit 1 and visit 2 (*n* = 77); c) Stacked bar plot depicting the fungal makeup of highly inflamed individuals at visit 1 and visit 2 (*n* = 62).

A *limma* test revealed that no proteins were differentially abundant between inflammation categories at both visits when adjusted for confounding variables (adj. *p*-value > 0.17 visit 1; adj. *p*-value > 0.08 visit 2).

The NMDS plot shows that individuals belonging to different inflammation categories overlapped substantially for both visits (Fig 3.15; Fig S3). An ANOSIM test confirmed that fungal protein relative abundance between inflammation states were similar (R = 0.0634; p-value = 0.09) in visit 1 and (R = 0.0418; p-value = 0.0089) visit 2. A total of 32 fungal proteins were implicated in significantly driving the distribution of individuals in visit 1, and 33 fungal proteins in visit 2. An *adonis* test revealed that approximately 3.5% (p-val = 0.0053) and 5.7% (p-val = 6e-04) of the variation in distances was explained by grouping the samples by inflammation category in visit 1 and visit 2, respectively.

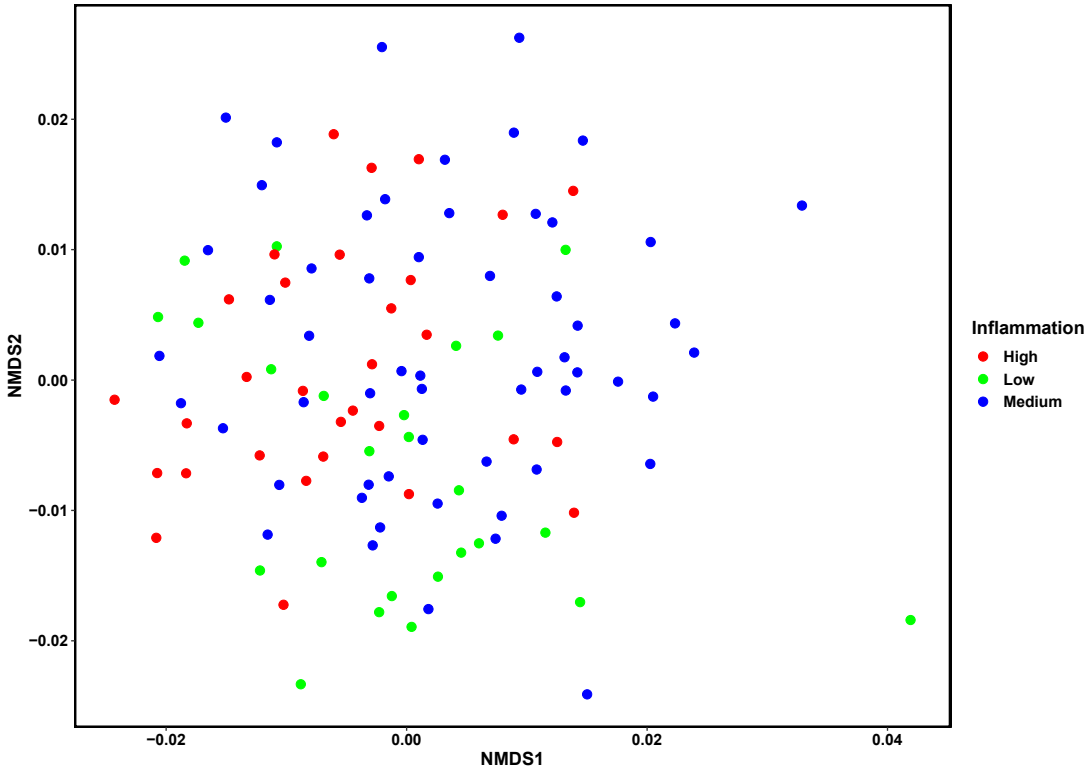


Figure 3.15. Non-metric Multi-dimensional Scaling (NMDS) analysis was used to group individuals based on the log2-transformed imputed intensity-based absolute quantification (iBAQ) values of fungal proteins in visit 1 according to inflammation categories.

The hierarchical clustering plot confirmed that individuals did not cluster according to inflammation categories using fungal proteins, as all three clusters contained a random mixture of inflammation categories (Fig 3.16).

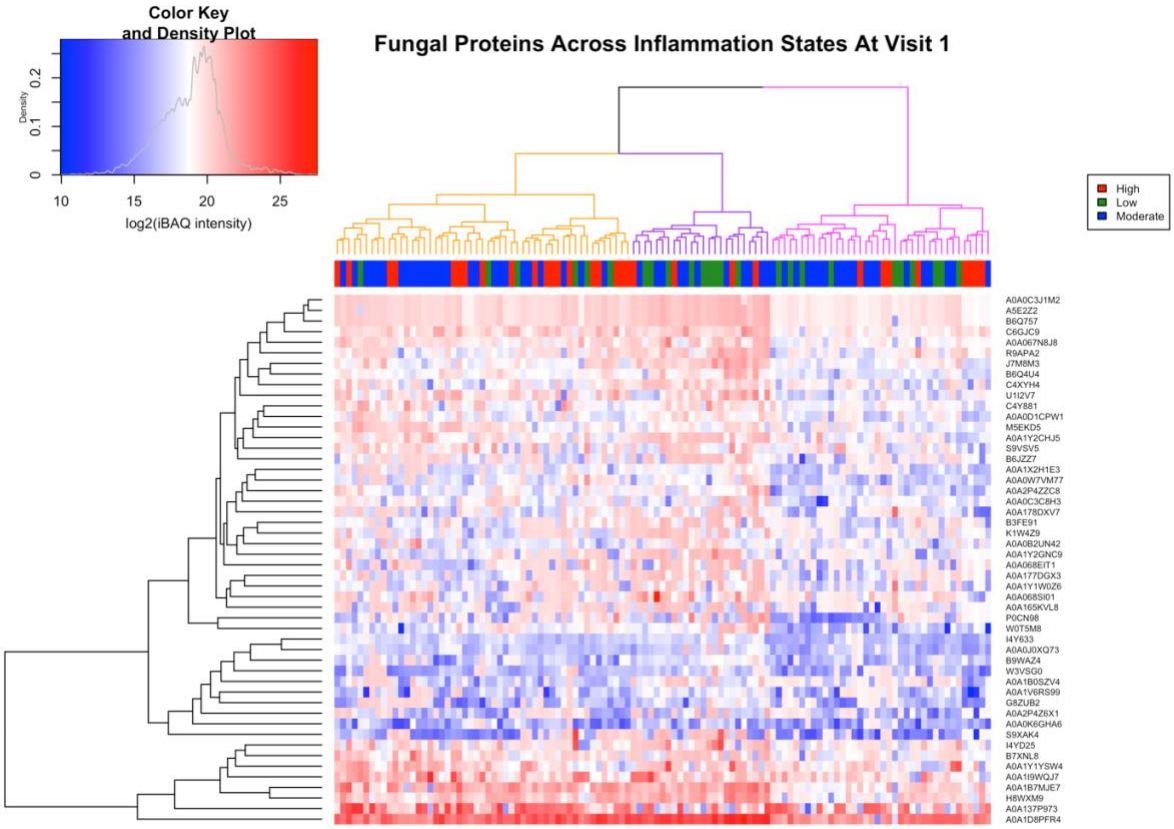


Figure 3.16. Hierarchical clustering was used to group individuals based on the log2-transformed intensity-based absolute quantification (iBAQ) values of fungal proteins across different inflammation categories at visit 1.

Functional Analysis

A comparison between inflammation states demonstrated that similar to BV, histone H4 acetylation and membrane raft polarization is the most relatively abundant BPs across inflammation states (Fig 3.17a). The high inflammation group had the highest relative abundance of these BPs only at visit 2 (54.4%), followed by the low inflammation group at both visits (41.3% visit 1 and 44,7% at visit 2). The high inflammation group at visit 1 had a relative abundance of 39.4%, and the medium inflammation group had a relative abundance of 37.4% at visit 1 and 28.9% at visit 2 (Fig 3.17a).

The high inflammation category at visit 2 had the highest relative abundance of the following BPs: endocytosis, histone exchange, mitotic actomyosin contractile ring assembly, regulation of nucleosome density in comparison to other inflammation groups (1.7%; Fig 3.17a).

The moderate inflammation category had the highest relative abundance of proteasome-mediated ubiquitin-dependent protein catabolic process during visit 2 compared to other inflammation groups (1.2%; Fig 3.17a).

BPs that constituted less than 1% across the Inflammation States can be found in Supplementary Table S6.

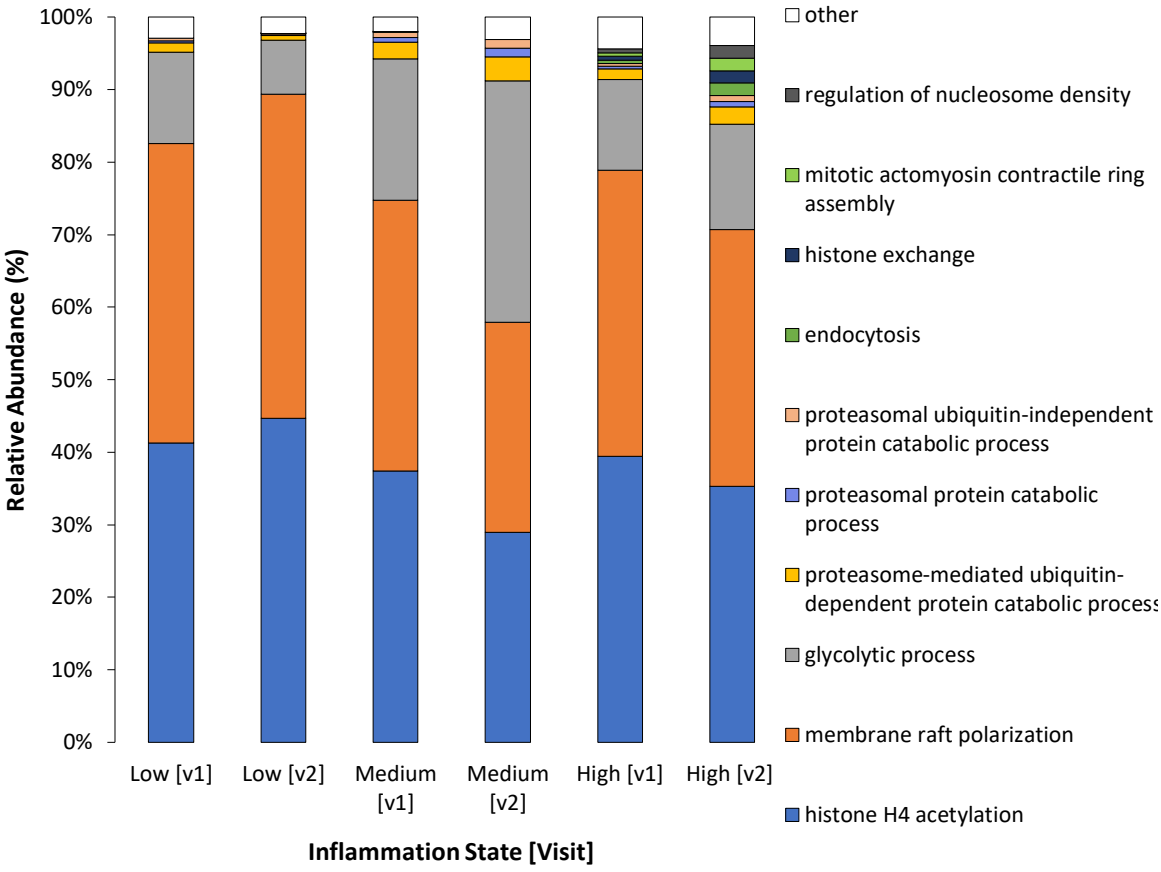


Figure 3.17. a) Functional composition at the biological process level according to gene ontology across inflammation categories at two different visits: lowly inflamed at visit 1 (*n* = 36) and visit 2 (*n* = 22); moderately inflamed at visit 1 (*n* = 49) and visit 2 (*n* = 28); highly inflamed at visit 1 (*n* = 28) and visit 2 (*n* = 34).

The comparison between inflammation states of BPs according to GO annotations revealed that similar to genera, no BPs were differentially abundant between inflammation states when adjusted for confounding variables.

Hierarchical clustering once again showed that individuals did not cluster into inflammation groups even with fungal BPs, with each of the three clusters including a mixture of inflammation states (Fig 3.18).

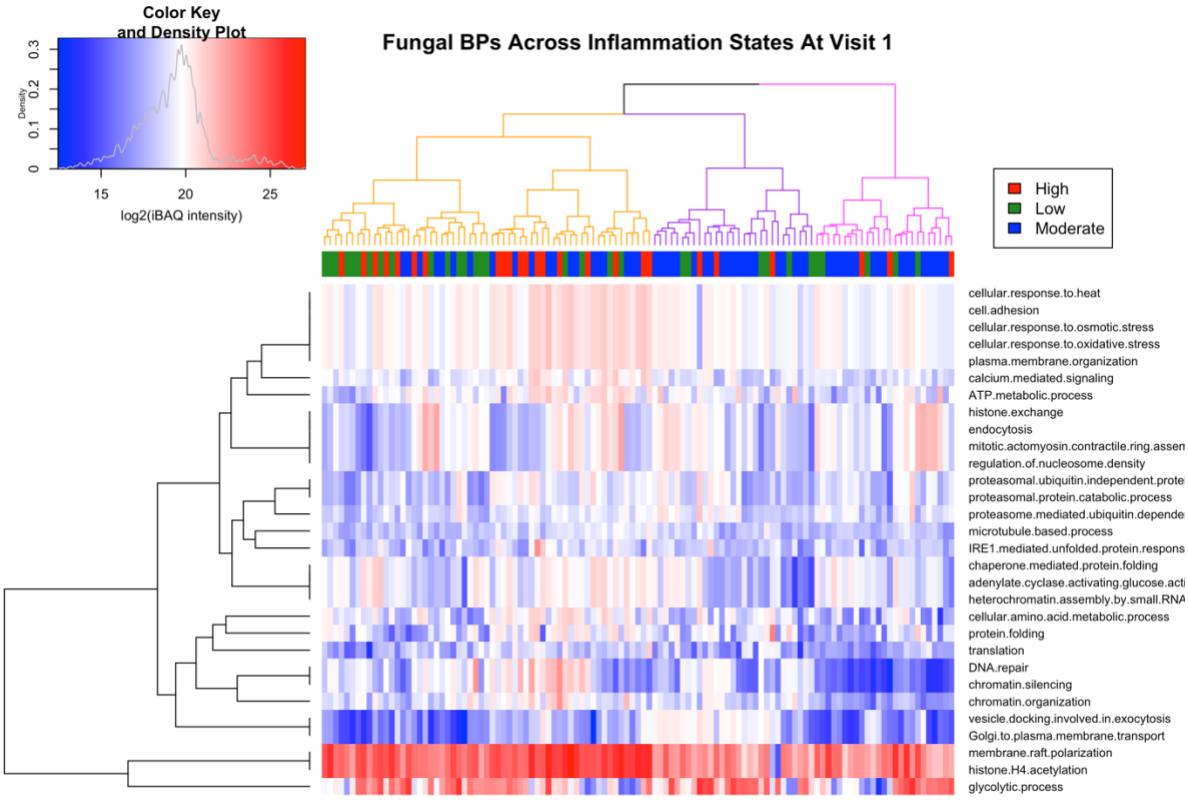


Figure 3.18. Hierarchical clustering was used to group individuals based on the log2-transformed intensity-based absolute quantification (iBAQ) values of biological processes of fungal proteins according to gene ontology.

Differences between STI status

Taxonomic Analysis

We made comparisons between the relative abundance of fungal genera across women infected with different STIs including *T. vaginalis* (*n* = 6 visit 1; *n* = 3 visit 2), *M. genitalium* (*n* = 7 visit 1; *n* = 6 visit 2), *N. gonorrhoeae* (*n* = 6 visit 1; *n* = 3 visit 2), *C. trachomatis* (*n* = 37 visit 1; *n* = visit 16), and Herpes Simplex Virus-2 (HSV-2) (*n* = 4 visit 1; *n* = 3 visit 2).

Candida was the most relatively abundant in each of the five STIs, with women who were positive for NG in visit 2 having the highest relative abundance of this genus (70.3%; Fig 26a), followed by women infected with TV in visit 1 (67.1%; Fig 3.19a). The total iBAQ protein intensity for *Candida* was highest in women infected with CT at both visits, however, this group had a higher number of individuals in comparison to other groups, which explains why the total iBAQ protein value is larger for each genus (Fig 3.19b).

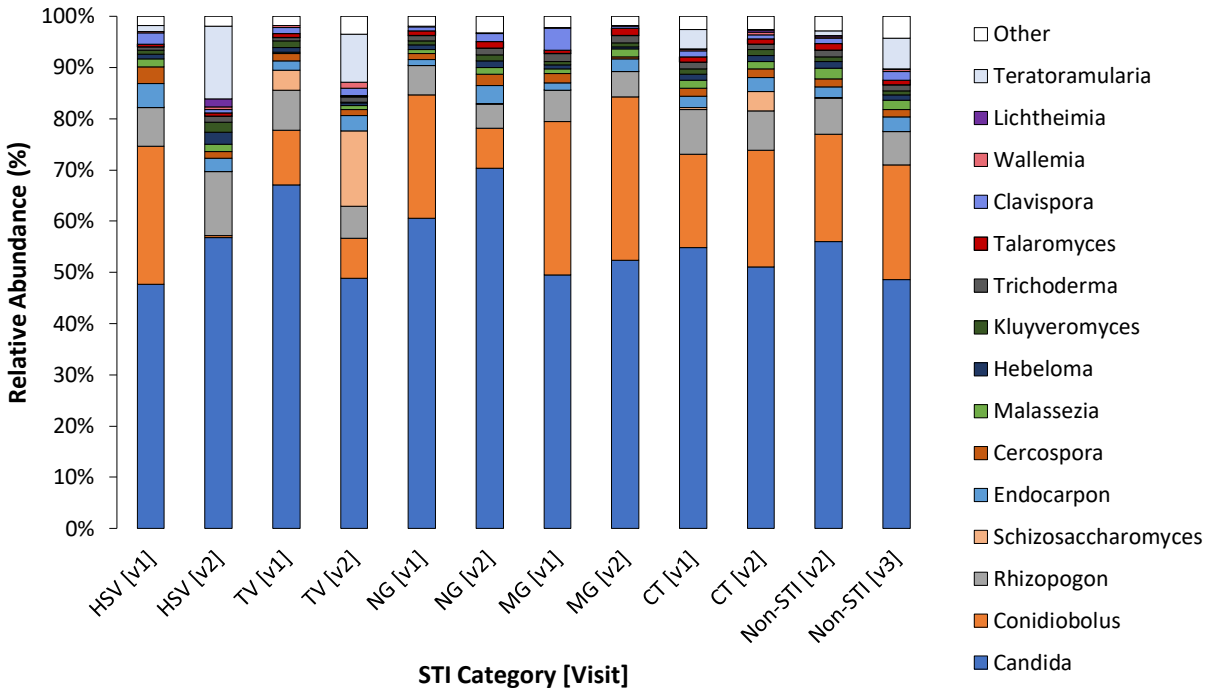
Conidiobolus was the second most relatively abundant genus in the majority of STI groups, with the exception of women infected with HSV in visit 2 and women infected with TV in visit 2 (Fig 3.19a). Women who were infected with MG at visit 2 had the highest relative abundance of *Condiobolous* (31.8%; Fig 3.19a).

In women infected with TV at visit 2, *Schizosaccharomyces* had a high relative abundance (14.6%) (Fig 3.19a). The total iBAQ value for this genus confirmed that it was higher in TV-infected women than in women infected with other STIs, despite women infected with CT having a larger total fungal iBAQ value (and larger sample size) (Fig 3.19b).

Women infected with HSV-2 had a high relative abundance of *Rhizopogon* (12.5%) and *Teratoramularia* (14.3%) at visit 2 (Fig 3.19a). The total iBAQ intensity value was the same for HSV-2 infected women and TV-infected women at visit 2, and the sample size for both groups was the same, however, we observe differences in relative abundances of this genus.

Limma tests were adjusted to account for individuals who were infected with multiple STIs and none of the above genera were significantly differentially abundant between each STI and the Non-STI group for both visits.

A *limma* test demonstrated that 1 protein (R9APA2 (Calmodulin, *Wallemia ichthyophaga*) (logFC = 3.17; *p-val* = 3.95e-05; *adj. p-val* = 0.0020)) was significantly overabundant in MG-infected women when compared to the Non-STI infected women during visit 2.



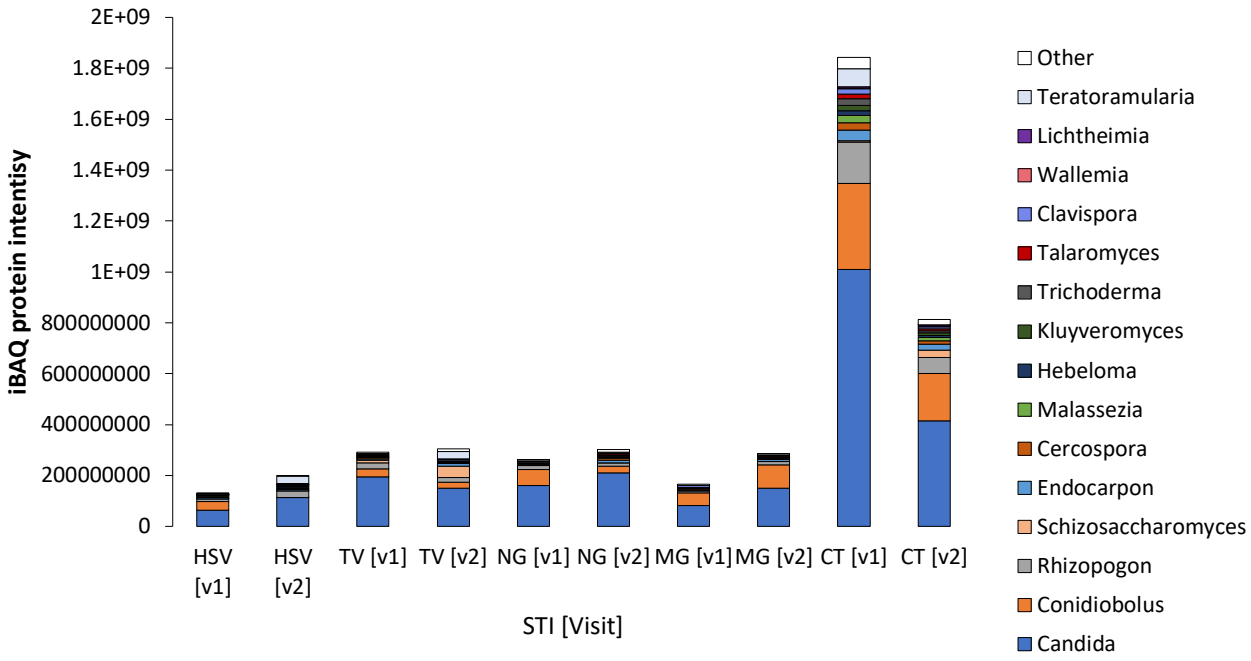


Figure 3.19. a) Fungal community composition at the genus level across different STIs at two different visits: Herpes simplex virus-2 (HSV-2) at visit 1 (*n* = 4) and visit 2 (*n* = 3); *T. vaginalis* (TV) at visit 1 (*n* = 6) and visit 2 (*n* = 3); *N. gonorrhoeae* (NG) at visit 1 (*n* = 6) and visit 2 (*n* = 3); M. *genitalium* at visit 1 (*n* = 7) and visit 2 (*n* = 6); *C. trachomatis* (CT) at visit 1 (*n* = 37) and visit 2 (*n* = 16); and Non-STI samples at visit 1 (*n* = 62) and visit 2 (*n* = 59). B) Total fungal iBAQ protein intensity of different STIs at two different visits.

The NMDS plot shows a high level of overlap between STI-positive and STI-negative samples as no clear grouping was observed (Fig 3.20). An ANOSIM test confirmed that fungal proteins between STI states were similar (R = -0.02, *p* = 0.88). An *adonis* test revealed that only 0.05% of the variation in distances was explained by grouping the samples by STI status (*p* = 0.4078).

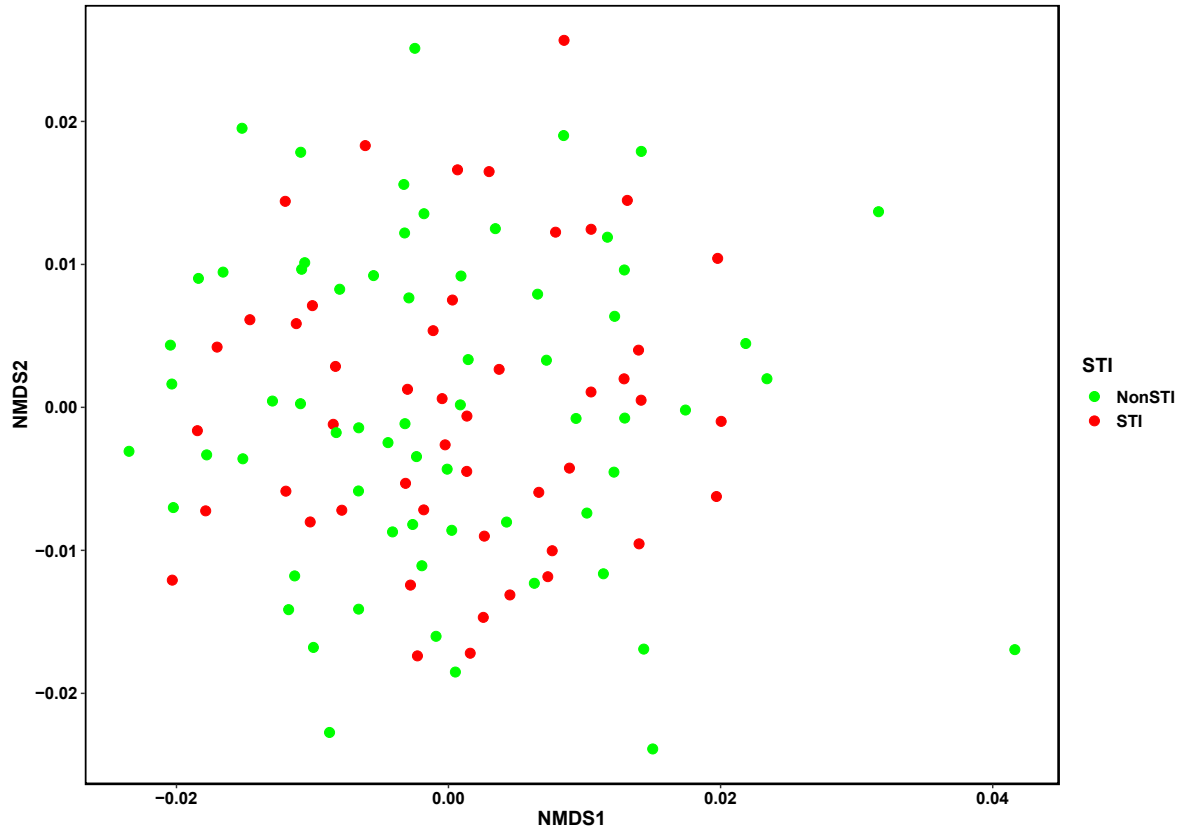


Figure 3.20. Non-metric Multi-dimensional Scaling (NMDS) analysis was used to group individuals based on the log2-transformed imputed intensity-based absolute quantification (iBAQ) intensities for fungal proteins in visit 1.

Hierarchical clustering confirmed that the study participants did not cluster by STI status based on fungal protein relative abundance, as both clusters contained a random mixture of STI states (Fig 3.21).

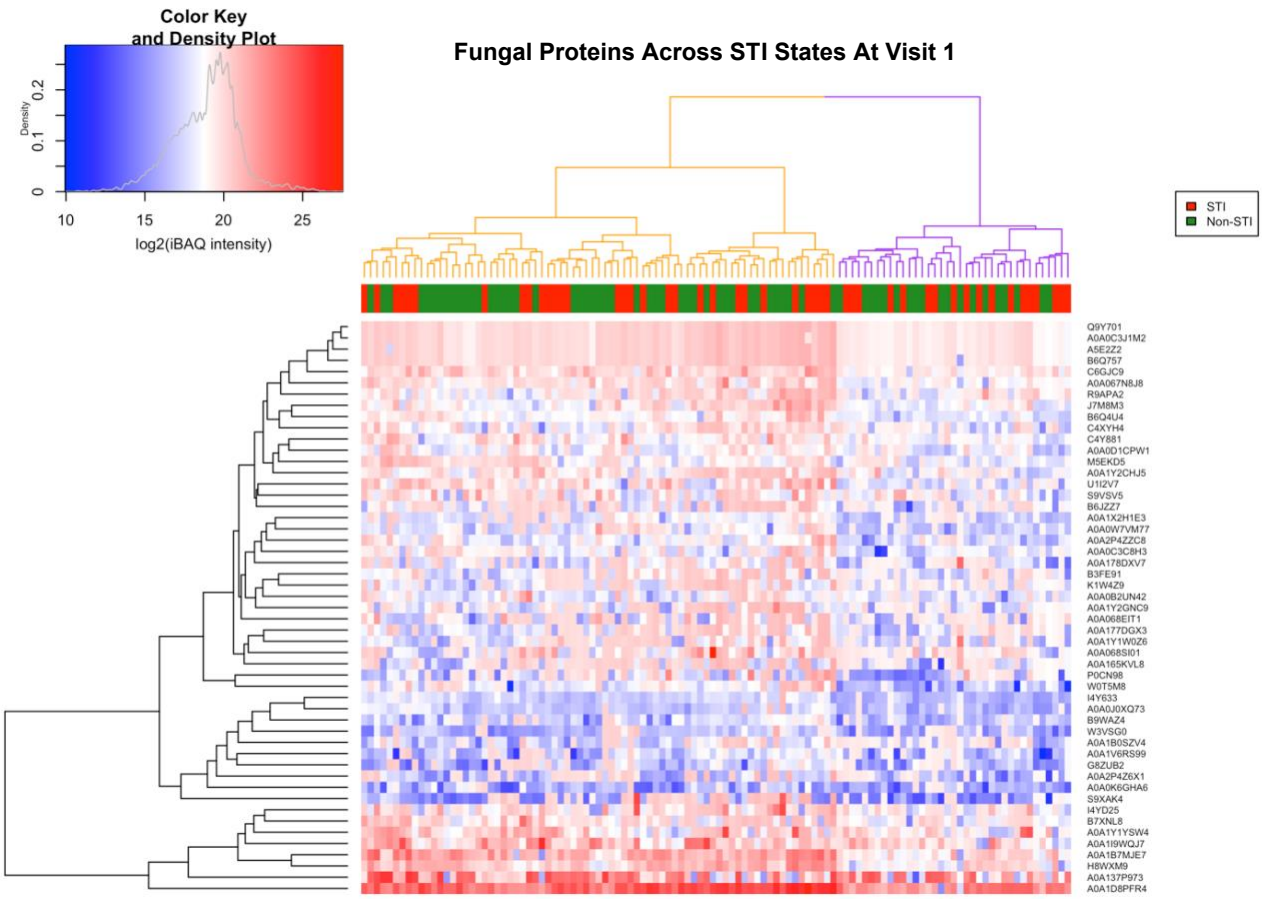


Figure 3.21. Hierarchical clustering was used to group individuals based on the log2-transformed intensity-based absolute quantification (iBAQ) values of fungal proteins.

Functional Analysis

As seen in inflammation groups, Histone H4 acetylation, and membrane raft polarization are highly relatively abundant across all STI groups (Fig 3.22a). Women infected with HSV (45.3%) and women infected with NG (44.2%) at visit 2 had the highest relative abundance of these BPs relative to women with other STIs (Fig 3.22a).

According to *limma*, one BP was found to be significantly overabundant in women infected with MG in comparison with Non-STI infected women. The significantly overabundant BP was Histone H4 acetylation (logFC = 2.79; *p-val* = 0.003; *adj. p-val* = 0.047) when adjusted for multiple STIs.

The glycolytic process was most relatively abundant in cases where women were infected with MG at both visits (23.3% visit 1; 24.4% visit 2). However, this BP was a lower abundance in

women infected with TV (7.3% visit 1; 5.0% visit 2), NG at visit 2 (5.6%), and HSV in visit 2 (0.4%) (Fig 3.22a).

Women infected with TV at visit 2 had the highest relative abundance of endocytosis, histone exchange, mitotic actomyosin contractile ring assembly, and regulation of nucleosome density in comparison to other STIs (8.2%; Fig 3.22a).

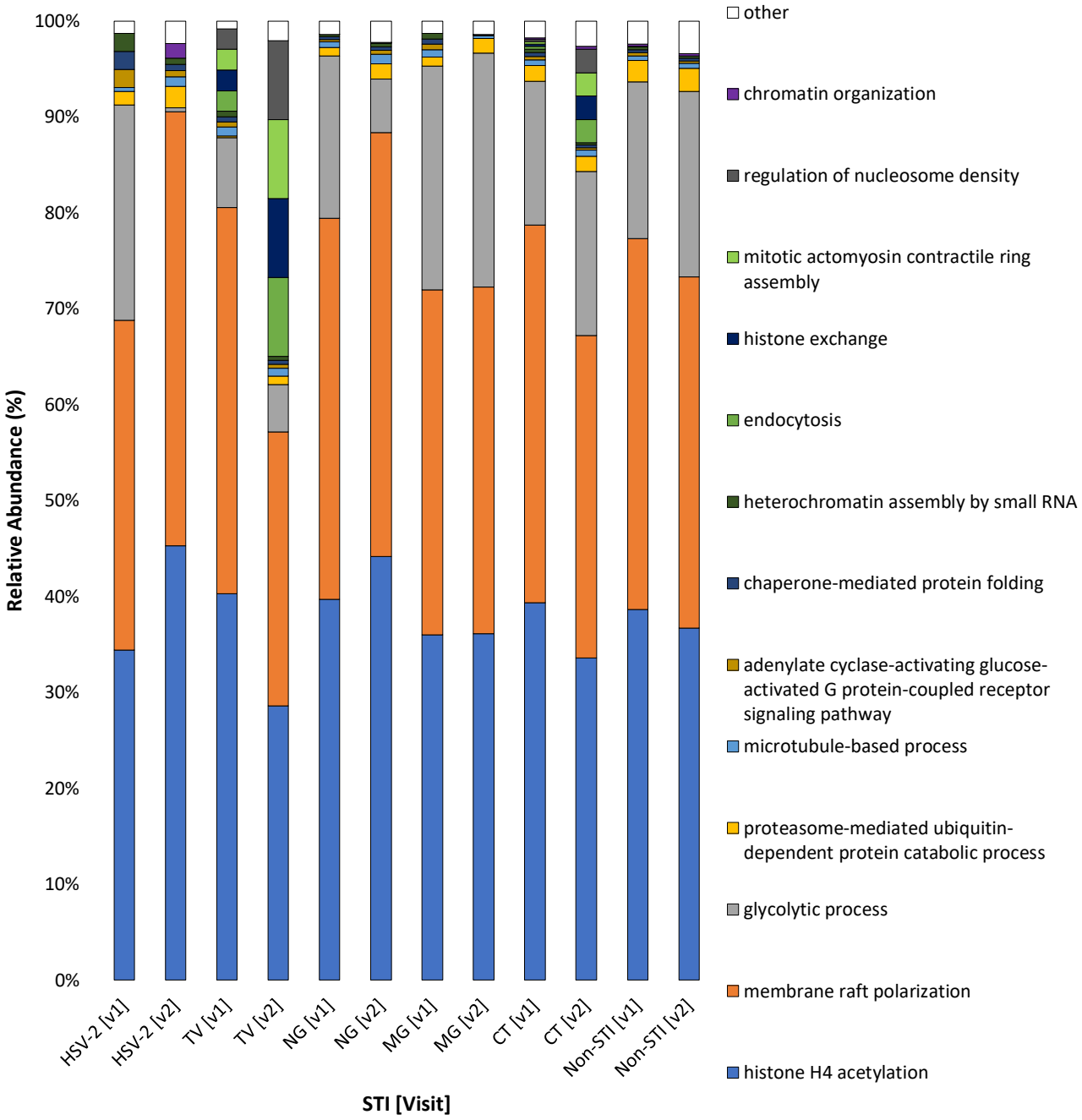


Figure 3.22. a) Relative abundance of biological processes according to Gene Ontology assignments across five STI groups and Non-STI at two different time points: HSV-2 at visit 1 (*n* = 4) and visit 2 (*n* = 3); TV at visit 1 (*n* = 6) and visit 2 (*n* = 3); NG at visit 1 (*n* = 6) and visit 2 (*n* = 3); MG at visit 1 (*n* = 7)

and visit 2 (*n* = 6); CT at visit 1 (*n* = 37) and visit 2 (*n* = 16) and non-STI at visit 1 (*n* = 62) and visit 2 (*n* = 59).

Discussion

Fungi play an essential role in the VMB, despite their low relative abundance. In our study, the relative abundance of fungi was found to be 0.4%. This is close to the range that is expected as Ma *et al.* (2020) reported that in the vagina the fungal prevalence range was between 0.17 ± 0.04%. However, this percentage was calculated based on fungal sequences identified from annotated reference databases in which fungi are highly underrepresented. Thus, this percentage likely underestimates the abundance and significance of fungi.

Taxonomic composition of the vaginal mycobiome

The phyla Ascomycota and Basidiomycota dominate most human anatomic sites, including the vagina, skin, gut, and oral cavity (Ghannoum *et al.* 2010; Hoffmann *et al.* 2013; van Woerden *et al.* 2013; Zhang *et al.* 2011). Not surprisingly, our study showed the same trend where Ascomycota dominated the vaginal mycobiome, followed by Basidiomycota. The dominance of these two phyla across different body sites suggests that they may be well-suited for life on mammalian hosts (Nash *et al.* 2017).

Previous FGT studies have shown that commonly detected fungi include *Candida*, *Saccharomyces*, *Aspergillus*, *Malassezia*, *Alternaria*, *Clavispora*, and *Trichosporon* (Drell *et al*. 2013; Fornari *et al.* 2016; Godoy-Vitorino *et al.* 2018; Guo *et al*. 2012; Nejat *et al.* 2018; Zheng *et al.* 2013). All of the aforementioned fungal genera were detected in our study, with the exception of *Saccharomyces* only being detected using peptide sequences. The genera *Aspergillus* and *Alternaria* were detected at relatively low abundances, which is in accordance with Zheng *et al.* (2013) and Guo *et al.* (2013). Other less commonly detected genera that have been identified in the literature, include *Kluyveromyces*, *Lodderomyces,* and *Penicillium* (Fornari *et al.* 2016; Lehtoranta *et al.* 2021; Nejat *et al.* 2018).

Without exception, *Candida* is the predominant member of the vaginal fungal community (often >70%; Barousse *et al.* 2004; Goldacre *et al.* 1981; Nowakowska *et al.* 2004). Within the genus

Candida, *C. albicans*, *C. glabrata*, *C. dubliniensis,* and *C. parapsilosis* are the most prevalent species in the vagina (Drell *et al.* 2013). In agreement with these studies, our study reconfirmed *Candida* as the predominant genus of the vaginal mycobiome and the presence of *C. albicans*, *C. dubliniensis*, *C. orthopsilosis*, and C. *thasaenensis*.

Some fungal species assignments such as *Rhizoclosmatium globosum* and *Endocarpon pusillum* have not been found in the human body previously. *Rhizoclosmatium* has characteristically been found in damp and aquatic environments, and *Endocarpon* has been found in soil (Belnap *et al.* 2001). Many opportunistic pathogens (such as *Aspergillus fumigatus*, *Cryptococcus neoformans*, and *Histoplasma caspulatum*) are naturally found in soil or other environmental niches. However, a limitation of *E. pusillum* assignment may be a result of the peptide/protein count and peptide sequence being too low. A BLAST search of the protein sequence for *E. pusillum* confirmed that this fungus was not incorrectly assigned during MS/MS assignment as the second highest percentage identity of other assigned taxa was 79.8% with *Exophiala aquamarine*. A BLAST search of the protein sequence for *R. globosum* confirmed that this fungus was not incorrectly assigned during MS/MS assignment as the second highest percentage identity of other assigned taxa was 88.9% with *Chytriomyces confervae*. Since these are actin and ATPase proteins (*i.e*., ubiquitous housekeeping proteins) sequences are similar across fungal species, and clear taxonomic annotation is not always feasible in metaproteomics. As discussed in chapter 1, peptides can be shared between different proteins of the same organism or between multiple organisms (Kleiner *et al.* 2017) or could represent potential human fungi that have not yet been sequenced.

Another possible hypothesis explaining the presence of fungi not commonly found in the human body may be a result of vaginal product use, hygiene habits, and water purity influencing the taxa that are detected in the FGT (Humphries *et al.* 2019). Lehtoranta *et al.* (2021) stated that surfaces and cavities in the human body are exposed to exogenous environmental fungi that can enter the vagina from various external sources. A South African study by Gafos *et al.* (2010) showed that women insert a range of products into their vaginas. Examples of products include crushed tobacco, Knorrox cubes (a food seasoning), rice water, and water after soaking jellyfish or tree bark. These practices may explain interesting fungal assignments usually not found in the vagina. However, more work is required to determine the influence of vaginal insertion and hygiene practices on the microbiome (Noyes *et al.* 2018). Just as the microbial underpinnings of BV are complex and varied, so too are the influences of a woman's sexual, sanitary, and other practices (Ma *et al.* 2012). It is also important to note

that VMB differs between different geographic areas. Therefore, we cannot say definitively whether these fungal species are normal commensals as no studies have examined the mycobiome in South Africa (McClelland *et al*. 2015; Ravel *et al.* 2011; Srinivasan *et al.* 2012).

Some fungal species found in the gut are transient and allochthonous (El Aila *et al.* 2009; El Aila *et al.* 2011; Suhr *et al.* 2016). Since the anus and the vagina are in close proximity, microorganisms can be translocated externally via wiping and/or during bathing. As a result, members of the gut microbiota may be shared with the vaginal microbiota (Suhr *et al.* 2016). This hypothesis may explain the presence of gut members, such as *Ustilago maydis* (a maize fungus), *Hebeloma cylindrosporum* (an edible mushroom), and *Cercospora beticola* (a sugar beet/spinach fungus) that were found in our study. Interestingly, a smut fungus (*Gjaerumia*) like *U. maydis* was found in a vaginal study by Godoy-Vitorino *et al.* (2018), and *U. maydis* has also been found to be closely related to *Malassezia* (Xu *et al.* 2007).

Clinical samples are highly susceptible to environmental contamination, such as airborne fungi, which get assigned during taxonomic analysis (Drell *et al.* 2013). Necessary precautions were taken to prevent the occurrence of contaminants, such as careful laboratory techniques. However, controlling all factors in a clinical environment proves difficult during sample collection. Therefore, more FGT mycobiome research is required to determine whether the fungal species identified in this study are true vaginal colonizers, transient members, or are merely contaminants.

Differences in fungi between optimal and non-optimal states

At least 17 out of the approximately 150 known *Candida* species, cause disease in humans (Neder 1992). The most common pathogenic species are *C. albicans*, *C. dubliniensis*, *C. glabrata*, *C. guilliermondii*, *C. krusei*, *C. lusitaniae*, *C. parapsilosis*, *C. tropicalis,* and *C. kefyr* (Neder 1992). In our study, *C. lusitaniae* was significantly overabundant in the BV positive state. From vaginal studies, this species is usually isolated from VVC patients (Fornari *et al.* 2016; Nejat *et al.* 2018). However, findings from studies such as Drell *et al.* (2013) and Zheng *et al.* (2013) suggest that several *Candida* spp. can also form part of an optimal microbiome, without causing a symptomatic infection (Beigi *et al.* 2004; Cotch *et al.* 1998; McClelland *et al.* 2009). Our study corroborates these results as *Candida* spp. (*C. albicans*, *C. thasanensis, C.*

orthopsilosis) were prevalent and significantly overabundant in samples from BV negative women.

C. albicans contains an arsenal of virulence and fitness attributes that contribute to both its ability to serve as a commensal and a pathogen, which includes adhesins, secreted hydrolytic enzymes, biofilm formation, and morphogenesis (Beigi *et al.* 2004; Deorukhkar and Roushani 2017). Biofilm formation protects *C. albicans* from external factors such as host immune system defenses and antifungal drugs (Katragkou *et al.* 2010; Xie *et al.* 2012). Furthermore, the fitness attributes of this species allow it to adapt to pH and CO_2 changes, and nutrient starvation (Nicholls *et al.* 2011). The ability of *C. albicans* to tolerate pH changes may explain the ability of *Candida* to co-exist with BV and its high relative abundance in each BV state in our study (Donders *et al.* 2011; Wei *et al.* 2012).

Some studies have shown that *C. albicans* co-colonizes the vagina with *Lactobacillus* (Beigi *et al.* 2004; Cotch *et al.* 1998; Förster *et al.* 2016; McClelland *et al.* 2009). Souza *et al.* (2009) illustrated this by showing that the most common bacterial flora associated with 3,125 instances of *Candida* spp. was *Lactobacilli*. Interestingly, one study showed that *L. crispatus* prevalence was higher in females diagnosed with VVC (Pendharkar *et al.* 2015). Thus, vaginal colonization with *Candida* spp. appears to be more common in females with a lactobacilli-dominated VMB (*e.g.,* BV negative) than in females with dysbiosis (*i.e.,* BV positive) (Biagi *et al.* 2009; Drell *et al.* 2013; Vitali *et al.* 2007; Zhou *et al.* 2009). However, *Candida* has also previously been correlated with the simultaneous presence of both lactobacilli and BV-associated bacteria (Donders *et al.* 2011; Wei *et al.* 2012). Thus, studies have suggested that there may be a co-occurrence between *Candida* and BV-associated bacteria, which may explain the high relative abundance of *Candida* in the BV positive group. In support of this Lehtoranta *et al.* (2021) and Pramanick *et al.* (2019) showed that *C. albicans* was the dominant fungal species even in BV-infected women.

From a clinical perspective, based on culture and microscopy, co-infections of *Candida* and BV-associated bacteria are common (approximately 20–30% of BV patients are co-infected). However, the patient's symptoms rarely reflect both infections (mixed vaginitis) (Sobel *et al.* 2013). In cases where individuals are infected with BV and VVC, it is suspected that the presence of Candidiasis may alter the structure of the non-optimal microbiome (*i.e.,* the BV

microbiome). This hypothesis was based on the finding from microbiomes where BV and VVC were both present showed an abundance of *Lactobacillus*, contrary to its depletion in individuals with only BV (Liu *et al.* 2013). This shift may indicate that *Candida* infections create an environment that promotes the growth of *Lactobacillus* (Liu *et al.* 2013). Especially since studies have found that *Lactobacillus* spp. are abundant in VVC (Cribby *et al*. 2008), and Zhou *et al.* (2009) found no significant difference in *Lactobacillus* abundance when comparing VVC to optimal VMBs.

From the literature the most common yeasts found associated with diseases in humans are *Candida*, *Histoplasma*, *Blastomyces*, *Cryptococcus*, and to a lesser extent, *Geotrichum*, *Malassezia*, *Pichia*, *Rhodotorula*, *Saccharomyces*, and *Trichosporon* (Enoch *et al.* 2006; Sanglard 2016). Of these, *Candida*, *Histoplasma*, *Blastomyces*, *Pichia*, *Saccharomyces*, and *Trichosporon* have been associated with causing disease in the FGT (Makela *et al.* 2003; Mouzin and Seilke 1996; Singh *et al.* 2002; Smith *et al*. 1997; Sobel *et al.* 1993). Our study identified the majority of these genera, however, protein assignments to other potentially pathogenic fungi including *Rhodotorula*, *Blastomyces*, *Geotrichum*, and *Pichia* were unreliable as they only had one peptide assigned to protein groups and removed during the exclusion process.

In our study, *Malassezia sympodialis* was overabundant in the BV positive state. *M. sympodialis* has been found in the FGT (Lehtoranta *et al*. 2021), although this species is commonly recognized as a skin fungus and can be found both in non-diseased individuals and individuals with skin diseases, such as atopic dermatitis (Oh *et al.* 2009; Sugita *et al.* 2001). From the literature, *Malassezia* is known to be a constituent of an optimal VMB; however, this fungus is able to live a dual lifestyle and can function as a pathogen. Several virulence factors have been identified that may contribute to the pathogenic potential of this fungus. Among them are lipolytic activities, distinct cell wall construction, hyphae formation, and the recently described tryptophan-dependent pigment production (Hort and Mayser 2011). In a vaginal study by Godoy-Vitorino *et al.* (2018), *Malassezia* was significantly dominant from high-risk Human papillomavirus (HPV) patients. Similarly, *M. sympodialis* has also been found in the gut mycobiome and has been shown to increase in prevalence during inflammatory bowel syndrome (IBS) in comparison to their non-IBS controls (Sokol *et al.* 2017). Therefore, the overabundance of *M. sympodialis* during BV is plausible. However, commensal fungal populations that occupy the vagina are assumed to change in response to disease (Sam *et al.*

2017; Underhill and Iliev 2014), but there has been little work done on understanding the changes in commensal communities.

Another overabundant fungus during BV was *Condiobolus coronatus* was also overabundant during BV; however, only one study has shown the involvement of this pathogenic species in the vagina and more work would be required to confirm this species involvement in the vagina (Subramanian and Sobel 2011).

From our results, there is limited evidence of qualitative changes in the taxonomic and functional makeup of the vaginal fungal communities. However, fungal communities undergo quantitative changes, therefore, there is an indication that BV and *M. genitalium* are associated with fungal community changes at the genus and functional level since the composition of the vaginal mycobiome varies under these different conditions. While the directionality of the relationship between the mycobiome and vaginal diseases cannot be determined based on the results of this study, fungal populations may be influenced by disease (*i.e.,* host–microbe interaction). However, clustering of a large proportion of the fungal proteins and functional terms into optimal and non-optimal groups was only demonstrated by BV status and not by inflammation category. The lack of clustering for inflammation categories indicates that fungi are not the major drivers of inflammation, and other clinical variables likely drive genital inflammation (such as vaginal pH and BV). However, further work is needed to confirm the factors that affect the mycobiome, BV status. and FGT inflammation, and to improve our understanding of the relationship between the vaginal mycobiome and vaginal diseases. This information could lead to effective, targeted interventions that eradicate high-risk pathogens and promote vaginal colonization with optimal microbial communities.

A caveat of the data analysis approach we used to examine microbiome communities (*i.e.,* NMDS, PERMANOVA, and ANOSIM) is that it reduces microbial community structure down to two or three dimensions. This reduction presents a challenge as microbiome data is inherently high-dimensional. By reducing the dimensionality, we conceal intra-microbiome interactions and interactions between metadata variables because statistical testing occurs on a variable-by-variable basis (Noyes *et al.* 2018). These methods do not reveal more nuanced and complex dynamics between and within the mycobiome, and host and environmental factors. Identifying such dynamics can be crucial for understanding multifactorial conditions such as BV (Shankar 2017). Thus, future work should incorporate Bayesian Networks (BN) to help elucidate some of the nuanced associations between the host, environment, and mycobiome.

BN analysis is able to handle mixed datasets for complex conditions as it does not reduce the data down into two dimensions.

Functional capacity of the mycobiome

In a metaproteomic study conducted by Zevin *et al.* (2016), the major functional categories represented in FGT bacterial communities included transport and catabolism, carbohydrate metabolism, nucleotide metabolism, and amino acid metabolism. Specifically, *Lactobacillus*-dominant (optimal-associated) communities showed enrichment of proteins involved in transport and catabolism, energy metabolism, and folding, sorting, and degradation. *Gardnerella*-dominant (BV-associated) communities were significantly enriched in membrane transport functions and showed increased factors involved in cytoskeletal-binding and proteasome activity (Zevin *et al.* 2016). Similarly, the major functional categories represented in our FGT fungal communities included carbohydrate metabolism (glycolysis/gluconeogenesis, pentose phosphate pathway, fructose and mannose metabolism), signal transduction (signaling pathways), energy metabolism, and folding, sorting, and degradation (protein export, and protein processing in endoplasmic reticulum). Correspondingly, the BV positive group showed an overabundance of proteins involved in proteasomal activity.

An FGT study by Li *et al.* (2018) used shotgun metagenomic sequencing to examine two different locations of the FGT, specifically the pouch of Douglas (PF) and the cervical canal (CV). This was done to examine the differences between the upper (PF) and lower (CV) FGT, as the upper FGT remains largely unexplored (Chen *et al.* 2018; Li *et al.* 2018). They functionally annotated genes using KEGG assignments. Functional annotation showed that PF samples had a higher proportion of microbial (*i.e.,* bacterial, fungal, archaeal, and viral) genes involved in carbohydrate metabolism, replication and repair (genetic information processing), and membrane transport (environmental processing). CV samples showed an enrichment of microbial genes involved in translation (genetic information processing), energy metabolism, and metabolism of cofactors and vitamins (Li *et al.* 2018).

Another supporting shotgun metagenomic study by Yang *et al.* (2020) examined vaginal samples to make comparisons between human papillomavirus (HPV)-infected women and non-HPV infected women. Their functional analysis showed that the majority of microbial (*i.e.,*

bacterial, fungal, viral, and archaeal) genes were assigned to carbohydrate metabolism, amino acid metabolism, translation, and membrane transport. Among the HPV16-positive women, the most prevalent pathways involved carbohydrate metabolism, amino acid metabolism, energy metabolism, membrane transport, and signal transduction (Yang *et al.* 2020). Our study showed the same trend in which the majority of our fungal proteins were assigned to genetic information processing (including translation and membrane trafficking), carbohydrate metabolism, cellular processes (including transport), energy metabolism, and signal transduction. In contrast, the fungal communities detected in our study showed few proteins assigned to amino acid metabolism. Therefore, fungal communities are probably involved in carbohydrate metabolism, energy metabolism, translation, and transport.

Adaption of fungal microorganisms

Nutrition is a fundamental requirement for the survival and growth of all living organisms. The metabolic adaptability of microorganisms enables the effective assimilation of nutrients available in a dynamic environment (Brown *et al.* 2012). Fungal pathogens have at least one shared challenge in common, their nutrition during pathogenesis. If a fungal species is not able to consume nutrients available in the host, it will not succeed as a pathogen (Brock 2009). For example, *Candida*'s is able to combat nutrient starvation by efficiently adapting their metabolism to the host microenvironment, rapidly tuning their metabolism to the available nutrients (Brown *et al.* 2014). As a result, *Candida* is able to proliferate under both nutrient-rich and nutrient-poor conditions. *C. albicans* cells are also able to induce glycolytic and tricarboxylic acid cycle genes during mucosal invasion (Wilson *et al*. 2009; Zakikhany *et al.* 2007). Thus, metabolic flexibility is integral for pathogenic fungi (Brock 2009; Fleck *et al.* 2011).

Most aerobic microorganisms (most fungi are aerobic) prefer the use of glucose as a carbon source rather than using gluconeogenic nutrients. This preference is a result of glucose requiring fewer catalytic steps to enter central metabolism via glycolysis. Glycolysis is thought to contribute to successful host colonization and pathogenesis (Brock 2009). For instance, the metabolism of glucose by *C. albicans* appears to be essential during infection because immune cells, such as macrophages, rely on glucose for survival (Tucey *et al.* 2014). Glycolysis not only provides building blocks for nucleic acids and cofactors but also generates NADPH for biosynthetic processes and for combating oxidative stresses (Thön *et al*. 2007). In order to favour glycolysis over gluconeogenesis, most fungi seem to contain a carbon catabolite

repressor system, which is mainly represented by the protein Mig1 in *S. cerevisiae* and *C. albicans,* and by CreA in filamentous fungi (Ruijter and Visser 1997). Correspondingly, Yang *et al.* (2020) hypothesized that active metabolism in the VMB provides a favorable microenvironment for pathogen survival. Thus, the ability of fungi to dynamically respond to host and pathogen-induced changes to nutrient availability in the VMB contributes to its success as a pathogen. However, more work needs to be done to understand these mechanisms.

The two stress proteins, Heat shock proteins (HSPs), and Glucose-regulated protein 78kDa (GRP78) were prevalent fungal proteins in our study. GRP78 is known to be a member of HSP70 (Tsai and Lee 2018). The Human Protein Atlas (proteinatlas.org) revealed that HSPs are elevated in the vagina compared to other tissue types. HSPs aid in cell survival under adverse environmental conditions, like microbial pathogen infection or pathogen-induced inflammation (Dasari *et al.* 2016; Sisti *et al.* 2015; Ventolini 2013). Induced hsp70 binds to its protein substrates and prevents their degradation or improper assembly and folding under adverse conditions (Mayer and Bukau 2005). The appearance of these stress proteins in the metaproteome may indicate a perturbation in the vaginal environment. Such perturbations could include changes in pH, oxygen, or nutrient concentrations. Thus, the level of hsp70 indicates the extent to which a cell senses threatening circumstances. HSP-70 has been shown to be expressed in women experiencing VVC (Giraldo *et al.* 1999a; Giraldo *et al.* 1999b). In our study, HSP was found to be underabundant in the BV negative state, in which *Candida* was significantly overabundant.

The high relative abundance of membrane raft polarization is due to the presence of actin proteins and cytoskeleton protein families since membrane rafts are enriched with actin cytoskeletons (Chichili and Rodgers 2009). Actin participates in many important cellular processes, including cell motility, cell division, cytokinesis, signal transduction, and the establishment and maintenance of cell junctions. The polarization of membrane rafts has been linked to morphogenesis and cell movement in diverse cell types. The ability to switch between budding and hyphal growth (*i.e.,* morphogenesis) is important for the pathogenesis of *C. albicans* (Sudbery 2011). The adaptation of *C. albicans* to acidic environments regulates key biological processes including morphogenesis (Buffo *et al.* 1984). Martin and Konopka (2004) suggest that membrane polarization may contribute to the pathogenesis of this species. This hypothesis is supported by the presence of the glutamate dehydrogenase protein which is

involved in the nitrogen assimilation pathway which is shown to be important in morphogenesis (Han *et al.* 2019).

Calmodulin proteins were also detected in this study. Calmodulin regulates intracellular calcium signaling (*i.e.,* signal transduction) and acts on several metabolic pathways and regulates gene expression in many eukaryotic organisms, including potential pathogens such as *C. albicans* (de Carvalho *et al.* 2003). Signal transduction is essential in regulating and eliciting physiological responses. Responses are elicited in response to stress (MAPK signaling pathway), oxygen deprivation (HIF-1 signaling pathway), metabolism, survival (Ras signaling pathway), differentiation (Hippo signaling pathway), cell-cell junction formation, and cell adhesion (Rap 1 signaling pathway). It is interesting to note that true hyphal formation is regulated by several proteins including those involved in the cAMP and MAP kinase pathway (Kumamoto and Vinces 2005). The MAPK pathway is able to contribute to cell wall stress resistance in *C. albicans* and involved its morphogenesis from yeast to hyphae, which is conserved in other fungi (Román *et al.* 2007).

Conclusion

We present the first large protein dataset to define the taxonomic composition and potential functional processes of the vaginal mycobiome in reproductive-age women in South Africa. While we reconfirmed *Candida* and *Malassezia* as dominant genera present in the vagina, this analysis identified several new prevalent genera, such as *Condiobolus, Rhizopogon*, and *Teratoramularia*. For these and other novel microorganisms (*e.g., Rhizoclosmatium* and *Endocarpon*) found in this study, future work should aim to confirm the presence of these novel species using longitudinal studies with extensive controls. With this, we can distinguish colonizing fungi from transient fungi and contaminants. Another advantageous method may include developing fractionation methods to increase the representation of samples prior to mass spectrometry analysis and reduce the interference between peptides. This will consequently increase the coverage of the metaproteome, and in turn, increase fungal identifications. (Mostovenko *et al.* 2013).

Our analysis of the mycobiome has yielded insight into changes in genus abundance and functional processes of fungal communities associated with BV and *M. genitalium*. This suggests that fungi may adapt to survive BV or that changes in the vaginal environment may

influence fungal functional profiles. However, we are only beginning to understand the potential role fungal communities play in the VMB, and more work is needed to confirm its function.

In conclusion, metaproteomics proves to be a valuable method as it provides information about protein expression and gets us closer to the potential function of the mycobiome, which cannot be done using methods such as NGS. However, routine application of metaproteomics is not common yet, which can be attributed to several challenges. Firstly, sample preparation is difficult due to the high complexity of samples that are vulnerable to contamination. Secondly, data analysis is challenging because well-annotated protein sequence databases are required, and protein inference causes uncertainty in protein annotations (Heyer *et al.* 2017; Mostovenko *et al.* 2013). Ideally, to increase the accuracy of metaproteomics, future work should combine longitudinal data collection with both NGS and mass spectrometry of fractionated samples.

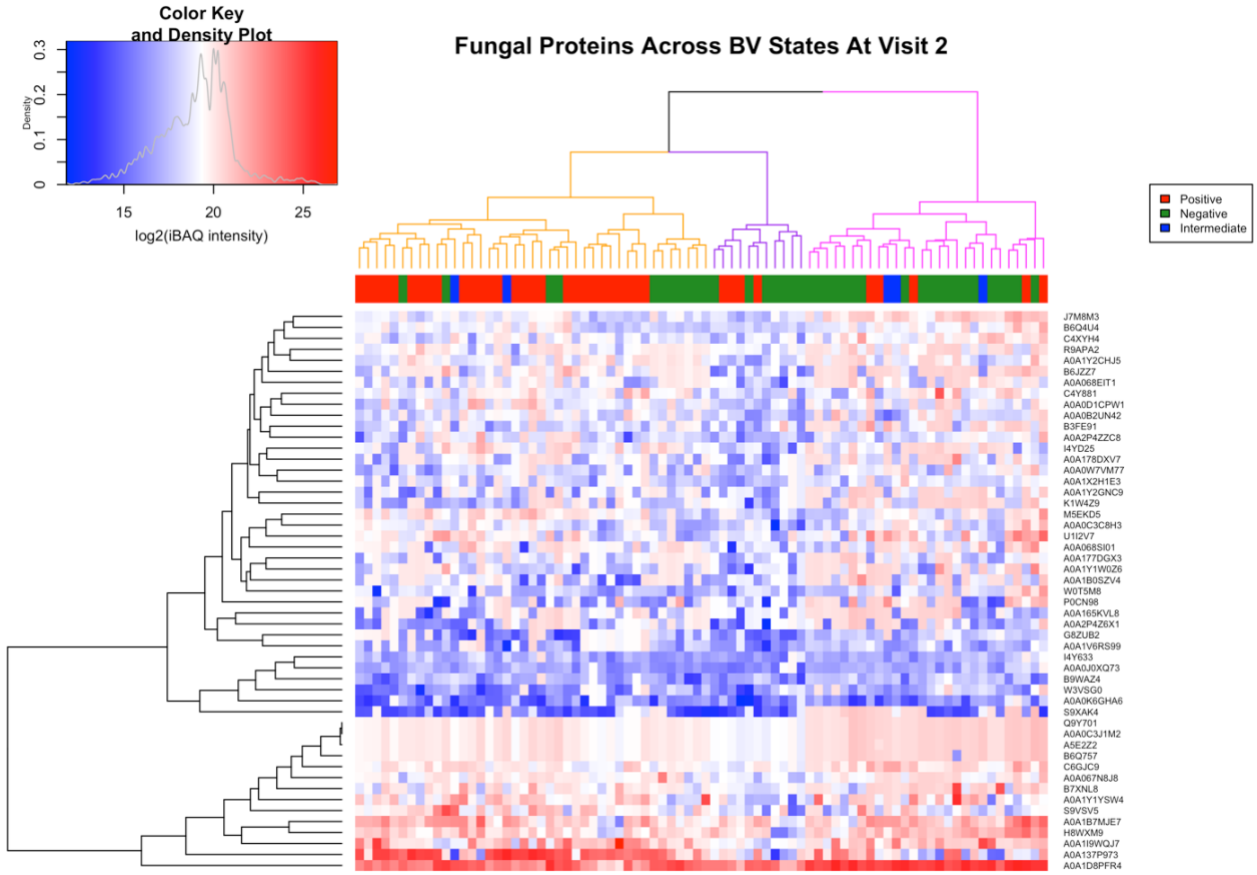


Figure S1. Hierarchical clustering was used to group samples based on the log2-transformed imputed intensity-based absolute quantification (iBAQ) values of fungal proteins at visit 2. Samples showed some clustering according to BV status indicated by the top dendrogram.

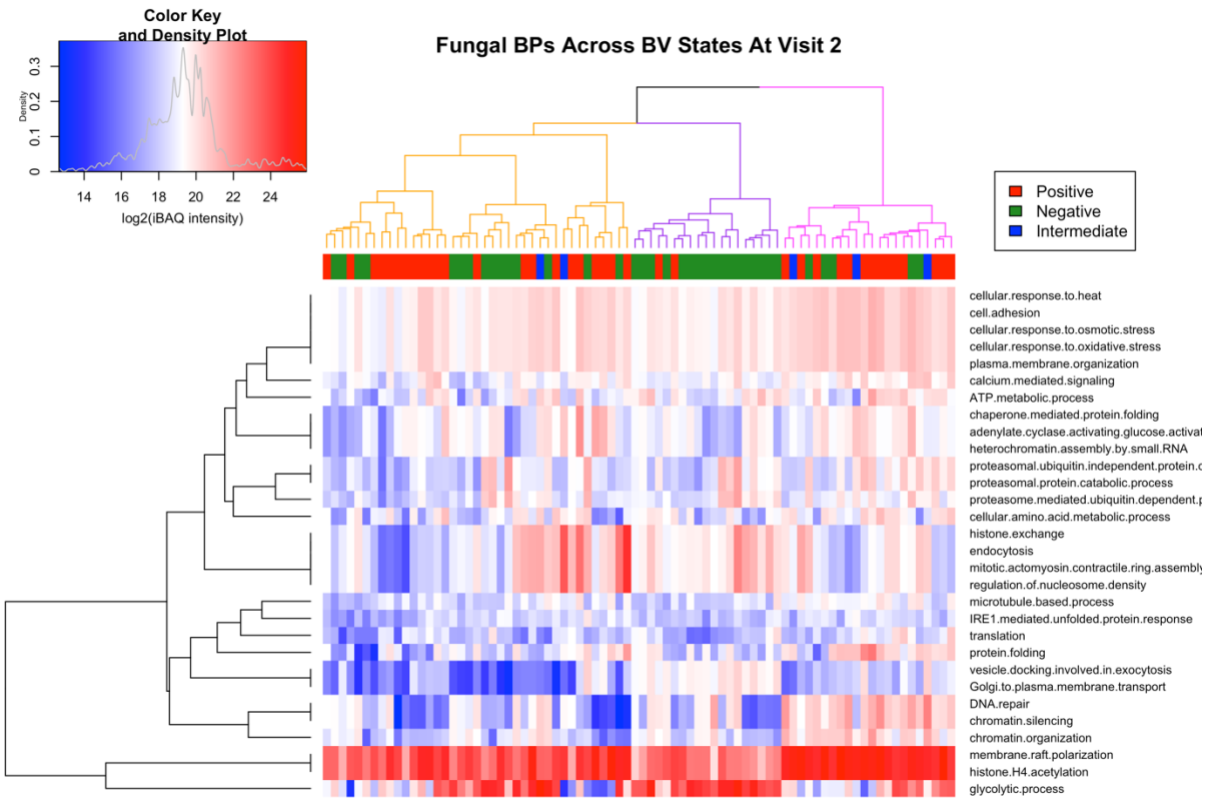


Figure S2. Hierarchical clustering was used to group samples based on the aggregated log2-transformed imputed intensity-based absolute quantification (iBAQ) values of fungal biological processes according to gene ontology at visit 2. Samples showed some clustering according to BV status indicated by the top dendrogram.

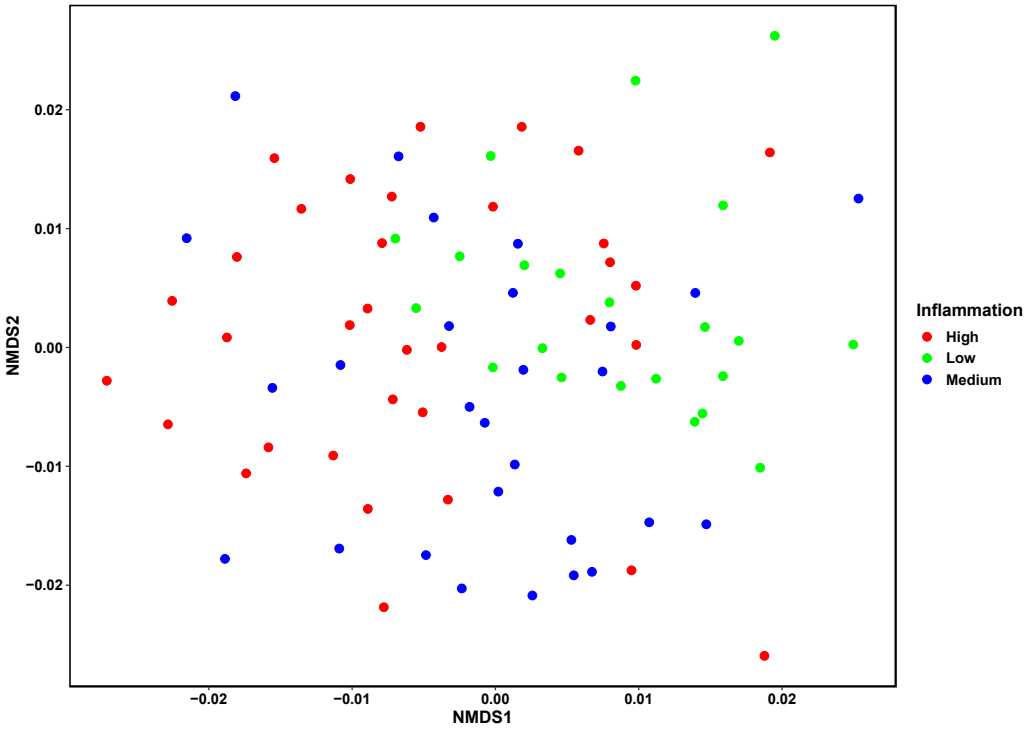


Figure S3. NMDS analysis was used to group individuals based on the log2-transfomed imputed intensity-based absolute quantification (iBAQ) values of fungal proteins at visit 2 according to inflammation categories.

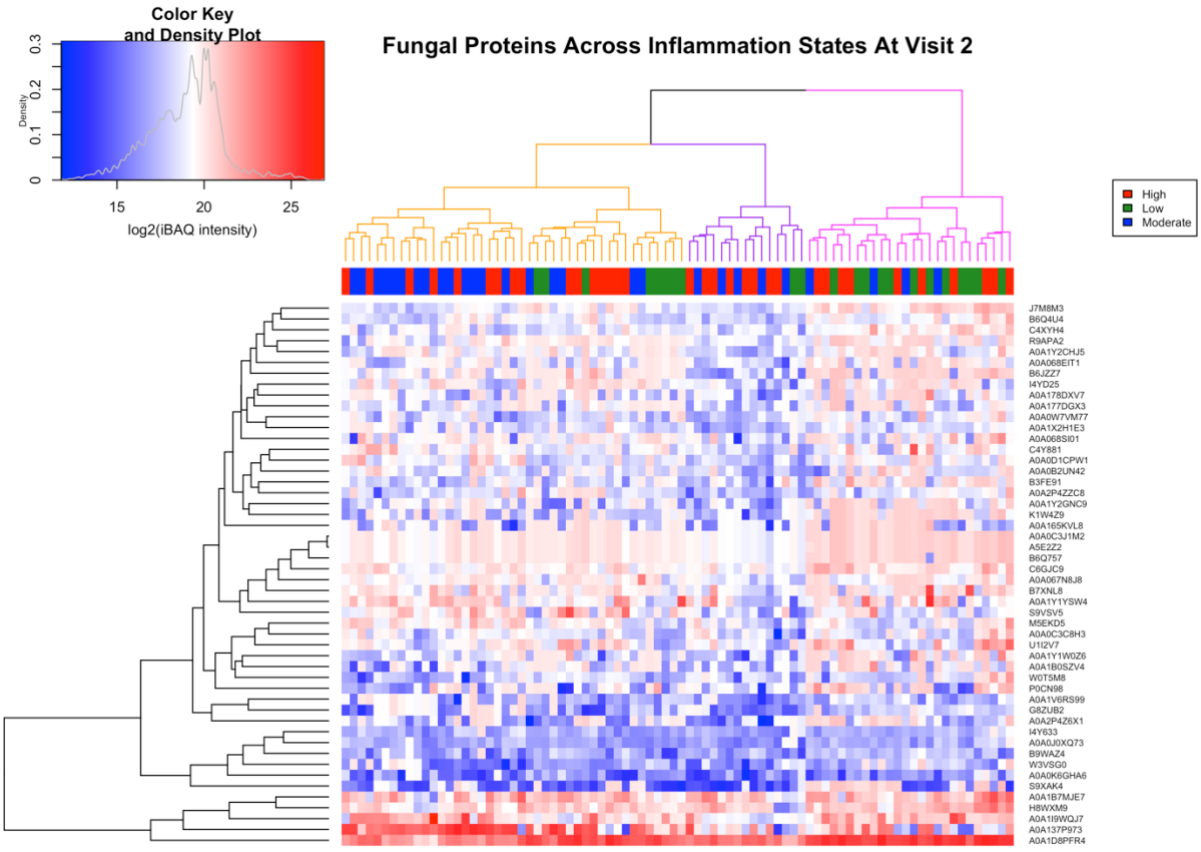


Figure S4. Hierarchical clustering was used to group individuals based on the log2-transformed intensity-based absolute quantification (iBAQ) values of fungal proteins across different inflammation categories at visit 2.

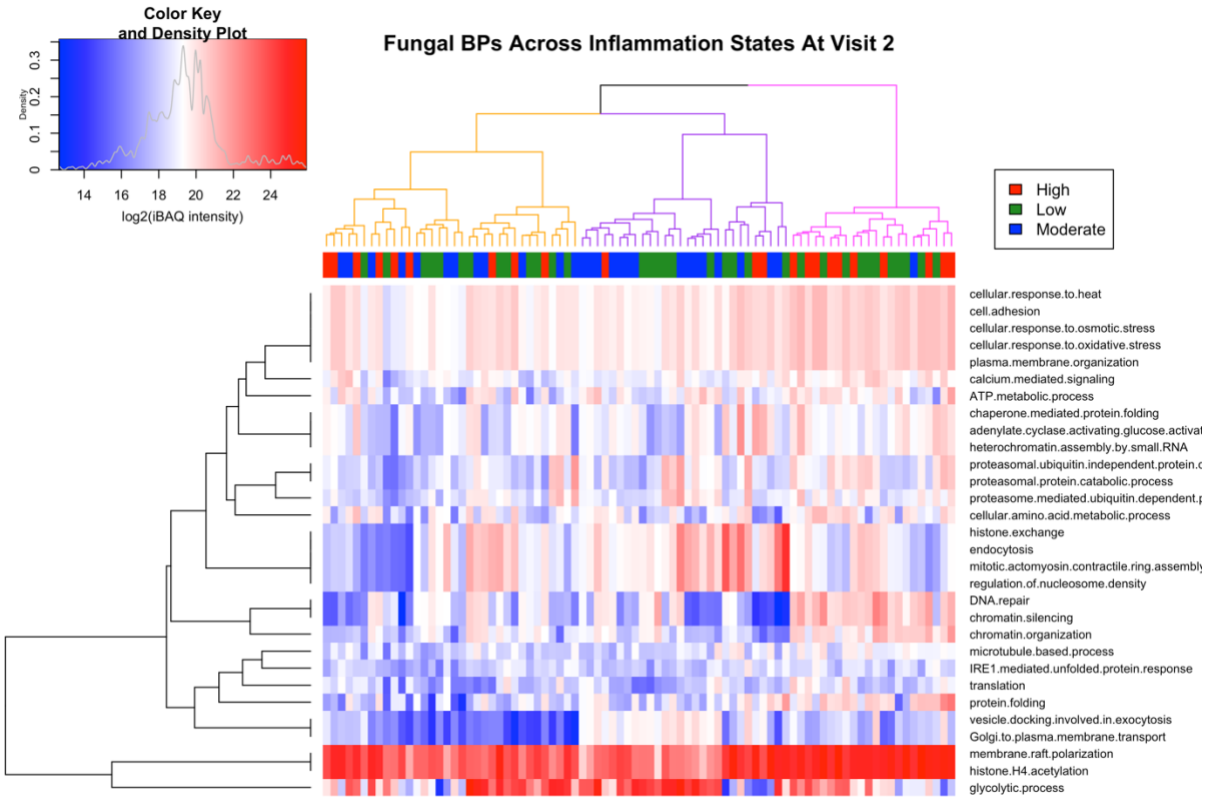


Figure S5. Hierarchical clustering was used to group individuals based on the log2-transformed intensity-based absolute quantification (iBAQ) values of biological processes of fungal proteins according to gene ontology across inflammation categories at visit 2.

Table S1. Fungal Biological Processes (according to Gene Ontology IDs) that had a relative abundance of less than 0.5%.

Biological Process	Relative Abundance (%)
histone exchange	0.43
mitotic actomyosin contractile ring assembly	0.43
regulation of nucleosome density	0.43
adenylate cyclase-activating glucose-activated G protein-coupled receptor signaling pathway	0.33
chaperone-mediated protein folding	0.33
heterochromatin assembly by small RNA	0.33

calcium-mediated signaling	0.24
chromatin organization	0.20
Golgi to plasma membrane transport	0.19
vesicle docking involved in exocytosis	0.19
cellular amino acid metabolic process	0.17
IRE1-mediated unfolded protein response	0.15
translation	0.13
protein folding	0.08
chromatin silencing	0.03
DNA repair	0.03
cell adhesion	0.01
cellular response to heat	0.01
cellular response to osmotic stress	0.01
cellular response to oxidative stress	0.01
plasma membrane organization	0.01

Table S2. Fungal Kyoto Encyclopaedia of Genes and Genomes (KEGG) term assignments that had a total relative abundance of less than 1%.

KEGG term	**Relative Abundance (%)**
Spliceosome	0.68
Endocytosis	0.67
Phagosome	0.58
Mitochondrial biogenesis	0.51
Messenger RNA biogenesis	0.47
Autophagy - yeast	0.34
GTP-binding proteins	0.34
MAPK signaling pathway	0.33
Ribosome biogenesis	0.33
mRNA surveillance pathway	0.28
RNA transport	0.28
Gap junction	0.23
Necroptosis	0.15
PI3K-Akt signaling pathway	0.14
Rap1 signaling pathway	0.11

DNA repair and recombination proteins	0.11
DNA replication proteins	0.11
AMPK signaling pathway	0.11
Translation factors	0.11
Apoptosis	0.10
Tight junction	0.10
Apelin signaling pathway	0.10
Calcium signaling pathway	0.10
cAMP signaling pathway	0.10
Cellular senescence	0.10
cGMP-PKG signaling pathway	0.10
Oocyte meiosis	0.10
Phohatidylinositol signaling system	0.10
Ras signaling pathway	0.10
Nitrogen metabolism	0,07
RNA degradation	0,03
Adherens junction	0.01
Focal adhesion	0.01
Hippo signaling pathway	0.01
Regulation of actin cytoskeleton	0.01
Transcription Machinery	0.01
mTOR signaling pathway	0.005
Oxidative phohorylation	0.005

Table S3. Fungal genera that constituted less than 1% of the relative abundance across bacterial vaginosis (BV)-states at different visits.

Genera	BV negative [v1] (%)	BV negative [v2] (%)	BV intermediate [v1] (%)	BV intermediate [v2] (%)	BV positive [v1] (%)	BV positive [v2] (%)
Alternaria	0	0.03	0	0.25	0.53	0.37
Wallemia	0.18	0.39	0.14	0.47	0.36	0.39
Lichtheimia	0.17	0	0	0	0.31	0.46
Neotyphodium	0	0.01	0	0	0.31	0.12
Linderina	0	0	0	0.04	0.24	0.2
Cutaneotrichosporon	0.11	0.1	0.25	0.08	0.21	0.22

Rhizoctonia	0.12	0.12	0.22	0.02	0.15	0.1
Aspergillus	0.09	0	0	0	0.14	0.18
Trichosporon	0.03	0.02	0.1	0.06	0.09	0.05
Moesziomyces	0.08	0.02	0.1	0.14	0.09	0.16
Pyrenochaeta	0.06	0.05	0.2	0.09	0.08	0.15
Penicillium	0.36	0.38	0.03	0.06	0.08	0.04
Syncephalastrum	0.48	0.56	0.1	0.45	0.05	0.06
Exidia	0.05	0.03	0.28	0.02	0.04	0.04
Torulaspora	0.41	0.53	0.03	0.11	0.04	0.03
Lobosporangium	0.02	0.08	0.04	0.15	0.03	0.02
Puccinia	0.08	0	0.14	0.15	0.03	0
Lodderomyces	0	0	0	0	0.01	0.07
Cryptococcus	0.02	0.02	0	0	0.01	0.18
Basidiobolus	0	0	0.91	0	0	0.04
Ordospora	0.05	0.06	0	0.09	0	0
Pisolithus	0.04	0	0	0	0	0
Pleurotus	0.03	0.7	0	0	0	0
Rhizoclosmatium	0.02	0.05	0	0.08	0	0

Table S4. Biological processes according to gene ontology terms which constituted less than 1% of the relative abundance across bacterial vaginosis (BV) states at visit 1 and visit 2.

Biological process	BV negative [v1] (%)	BV negative [v2] (%)	BV intermediate [v1] (%)	BV intermediate [v2] (%)	BV positive [v1] (%)	BV positive [v2] (%)
chromatin organization	0.13	0.01	0	0	0.29	0.64
translation	0.1	0.05	0.22	0.17	0.16	0.3
IRE1-mediated unfolded protein response	0.09	0.17	0.2	0.28	0.2	0.22
chromatin silencing	0.01	0.01	0	0	0.01	0.18
DNA repair	0.01	0.01	0	0	0.01	0.18
cell adhesion	0	0	0	0	0.01	0.07
cellular amino acid metabolic process	0	0.02	0	0.2	0.49	0.37
cellular response to heat	0	0	0	0	0.01	0.07

cellular response to osmotic stress	0	0	0	0	0.01	0.07
cellular response to oxidative stress	0	0	0	0	0.01	0.07
plasma membrane organization	0	0	0	0	0.01	0.07
protein folding	0	0	0	0.03	0.22	0.2
Golgi to plasma membrane transport	0.29	0.37	0.02	0.09	0.03	0.03
vesicle docking involved in exocytosis	0.29	0.37	0.02	0.09	0.03	0.03
microtubule-based process	0.82	0.85	0.27	0.94	0.17	0.19

Table S5. Fungal genera that constituted less than 1% of the relative abundance across inflammation states at visit 1 and visit 2.

Genera	**Low [v1] (%)**	**Low [v2] (%)**	**Medium [v1] (%)**	**Medium [v2] (%)**	**High [v1] (%)**	**High [v2] (%)**
Penicillium	0.43	0.53	0.2	0.18	0.11	0.07
Syncephalastrum	0.41	0.43	0.19	0.11	0.31	0.36
Lichtheimia	0.27	0	0.31	0.45	0	0.16
Alternaria	0.21	0	0.27	0.34	0.14	0.23
Rhizoctonia	0.17	0.1	0.15	0.11	0.11	0.12
Puccinia	0.14	0.08	0.05	0	0.03	0.03
Pyrenochaeta	0.13	0.06	0.06	0.07	0.08	0.13
Cutaneotrichosporon	0.11	0.05	0.13	0.15	0.22	0.21
Wallemia	0.1	0.3	0.27	0.26	0.3	0.51
Trichosporon	0.07	0	0.03	0.02	0.09	0.08
Exidia	0.06	0.02	0.07	0.03	0.05	0.04
Linderina	0.06	0	0.11	0.12	0.11	0.13
Lodderomyces	0.02	0	0	0.11	0	0
Neotyphodium	0.02	0	0.05	0.03	0.34	0.1
Basidiobolus	0.01	0	0.11	0	0	0.04
Aspergillus	0	0	0.14	0.16	0.11	0.08
Cryptococcus	0	0.01	0.01	0.02	0.01	0.17
Lobosporangium	0	0.01	0.03	0.03	0.04	0.09

Ordospora	0	0	0.04	0	0.02	0.07
Pisolithus	0	0	0.04	0	0	0
Pleurotus	0	0.17	0.03	0.8	0	0.07
Rhizoclosmatium	0	0	0	0	0.03	0.06
Moesziomyces	0.03	0.02	0.07	0.02	0.16	0.17
Torulaspora	0.6	0.66	0.17	0.14	0.07	0.11

Table S6. Biological processes according to gene ontology terms which constituted less than 1% of the relative abundance across all inflammation states at visit 1 and visit 2.

Biological process	Low [v1] (%)	Low [v2] (%)	Medium [v1] (%)	Medium [v2] (%)	High [v1] (%)	High [v2] (%)
microtubule-based process	0.65	0.77	0.42	0.21	0.66	0.73
Golgi to plasma membrane transport	0.44	0.47	0.14	0.12	0.06	0.1
vesicle docking involved in exocytosis	0.44	0.47	0.14	0.12	0.06	0.1
adenylate cyclase-activating glucose-activated G protein-coupled receptor signaling pathway	0.25	0.07	0.13	0.13	0.98	0.48
chaperone-mediated protein folding	0.25	0.07	0.13	0.13	0.98	0.48
heterochromatin assembly by small RNA	0.25	0.07	0.13	0.13	0.98	0.48
chromatin organization	0.2	0.01	0.26	0.4	0.01	0.28
cellular amino acid metabolic process	0.16	0	0.22	0.3	0.11	0.2
translation	0.11	0.06	0.1	0.07	0.19	0.25
IRE1-mediated unfolded protein response	0.08	0.03	0.11	0.13	0.21	0.34
protein folding	0.05	0	0.09	0.1	0.09	0.12
cell adhesion	0.02	0	0	0.09	0	0
cellular response to heat	0.02	0	0	0.09	0	0
cellular response to osmotic stress	0.02	0	0	0.09	0	0
cellular response to oxidative stress	0.02	0	0	0.09	0	0
plasma membrane organization	0.02	0	0	0.09	0	0
calcium-mediated signaling	0.02	0.31	0.19	0.85	0.13	0.2
chromatin silencing	0	0.01	0.01	0.02	0.01	0.14
DNA repair	0	0.01	0.01	0.02	0.01	0.14

Table S7. Fungal genera that constituted less than 1% of the total fungal relative abundance across five sexually transmitted infections (STIs) at visit 1 and visit 2.

Genera	HSV-2 [v1] (%)	HSV-2 [v2] (%)	TV [v1] (%)	TV [v2] (%)	NG [v1] (%)	NG [v2] (%)	MG [v1] (%)	MG [v2] (%)	CT [v1] (%)	CT [v2] (%)
Aspergillus	0	0	0	0	0.2	0.19	0	0	0.17	0.06
Exidia	0	0.11	0.04	0	0	0	0	0.04	0.06	0.04
Rhizoctonia	0	0	0.04	0	0.01	0.27	0.11	0.01	0.17	0.04
Neotyphodium	0.12	0	0	0.1	0	0.05	0	0	0.28	0.03
Moesziomyces	0.09	0.02	0	0.1	0.09	0.1	0.05	0	0.1	0.03
Linderina	0.19	0.09	0	0.05	0	0.12	0	0	0.08	0.03
Cryptococcus	0	0	0	0	0.05	0	0	0.07	0.01	0.02
Trichosporon	0	0	0	0	0	0.01	0	0	0.09	0.01
Basidiobolus	0	0	0	0	0	0	0	0	0	0
Enterocytozoon	0	0	0	0	0	0	0	0.08	0	0
Pisolithus	0	0	0	0	0	0	0	0	0.05	0
Puccinia	0	0.22	0	0.14	0.12	0	0.1	0	0.1	0
Syncephalastrium	0.34	0.13	0.89	0.57	0.34	0.65	0.4	0.17	0.23	0.33
Pyrenochaeta	0	0.31	0.02	0.03	0.03	0.22	0.05	0.11	0.06	0.2
Torulaspora	0	0.16	0.34	0.17	0.19	0.23	0.19	0.16	0.17	0.19
Penicillium	0.28	0.02	0.09	0.02	0.26	0.14	0.09	0.48	0.22	0.17
Pleurotus	0	0	0	0	0	0	0	0	0	0.17
Cutaneotrichosporon	0.17	0	0.17	0.06	0.06	0.21	0.12	0.13	0.18	0.16
Lodderomyces	0	0	0	0	0	0	0	0	0	0.16
Lobosporangium	0	0	0.08	0.54	0	0	0	0.02	0.02	0.15
Ordospora	0	0	0.09	0.42	0	0	0	0	0	0.13
Rhizoclosmatium	0	0	0.13	0.35	0	0	0	0	0	0.1
Alternaria	0.51	0	0	0.47	0.15	0.18	0	0	0.15	0.09
Ustilago	0.2	0.86	0.03	0.52	0.47	0.87	0.17	0.56	0.5	0

Table S8. Biological processes according to gene ontology terms which constituted less than 1% of the total fungal relative abundance across all five sexually transmitted infections (STIs) at visit 1 and visit 2.

Biological Process	CT [v1] (%)	CT [v2] (%)	MG [v1] (%)	MG [v2] (%)	NG [v1] (%)	NG [v2] (%)	TV [v1] (%)	TV [v2] (%)	HSV-2 [v1] (%)	HSV-2 [v2] (%)
calcium-mediated signalling	0.19	0	0.04	0	0.18	0.14	0	0	0.18	0.17
IRE1-mediated unfolded protein response	0.14	0	0.25	0.78	0.04	0.15	0.17	0.1	0.15	0.36
Golgi to plasma membrane transport	0.01	0.16	0.23	0.11	0.13	0.16	0.27	0.13	0.14	0.14
vesicle docking involved in exocytosis	0.01	0.16	0.23	0.11	0.13	0.16	0.27	0.13	0.14	0.14
translation	0.08	0.32	0.02	0.08	0.09	0.22	0.12	0.09	0.13	0.17
cellular amino acid metabolic process	0.42	0	0	0.3	0.11	0.13	0	0	0.13	0.07
protein folding	0.16	0.08	0	0.03	0	0.09	0	0	0.07	0.02
chromatin silencing	0	0	0	0	0.04	0	0	0.05	0.01	0.02
DNA repair	0	0	0	0	0.04	0	0	0.05	0.01	0.02
cell adhesion	0	0	0	0	0	0	0	0	0	0.12
cellular response to heat	0	0	0	0	0	0	0	0	0	0.12
cellular response to osmotic stress	0	0	0	0	0	0	0	0	0	0.12
cellular response to oxidative stress	0	0	0	0	0	0	0	0	0	0.12
plasma membrane organization	0	0	0	0	0	0	0	0	0	0.12
proteasomal protein catabolic process	0.17	0.83	0.02	0.33	0.33	0.62	0.24	0.43	0.41	0.5
proteasomal ubiquitin-independent protein catabolic process	0.17	0.83	0.02	0.33	0.33	0.62	0.24	0.43	0.41	0.5

Chapter 4

Understanding the vaginal mycobiome

Introduction

The vaginal microbiome (VMB) is a dynamic environment influenced by many factors, such as hormonal changes, sexual activity, antimicrobial drug usage, and hygiene (Gajer *et al.* 2012; Mulder *et al.* 2019; Plummer *et al.* 2019). However, some factors that may affect the VMB, and particularly fungal communities, such as hormonal contraception (HC) and body mass index (BMI), are not fully understood (Stout *et al.* 2020).

From research studies conducted thus far, estrogen-containing HCs, such as oral contraceptives, have been associated with an optimal *Lactobacillus*-containing microbiome (van de Wijgert *et al*. 2013). However, oral contraceptives are also capable of increasing the risk of vulvovaginal candidiasis (VVC) (Linhares *et al.* 2001; Spinillo *et al*. 1995; van de Wijgert *et al*. 2013). This claim is supported by Grigoriou *et al.* (2006), which showed *C. albicans* and non-albicans *Candida* species were isolated at a higher rate in women who used oral contraceptives than non-contraceptive users. Similarly, women who used a vaginal ring as a form of HC were found to have a higher number of *Candida* infections (Balle *et al*. 2020). On the contrary, several studies have shown that progestin-only contraceptive usage has been associated with an altered vaginal microbial profile (Balle *et al.* 2020; Song *et al.* 2020). One study suggested that the progestin-only depot medroxyprogesterone acetate (DMPA) HC may cause a hypoestrogenic state, which may impair colonization of *Lactobacilli* (Miller *et al.* 2000). In chapter 3, we discussed that vaginal colonization with *Candida* spp. appears to be more prevalent in females with a lactobacilli-dominated VMB (*i.e*., BV negative) than in females with dysbiosis (*i.e*., BV positive). Therefore, it would be beneficial to gain more information on the effects of HC on fungi present in the microbiome, especially since progestin-only injectables are the most frequently used form of contraception in sub-Saharan Africa, where HIV burden is highest (Jacobstein *et al.* 2014).

Regarding vaginal pH, VVC is not generally associated with increased pH levels (usually the vaginal pH < 4.5) (Moodley *et al.* 2002; Frobenius and Bogdan 2015). Previous studies have suggested that organic acid production and the decrease in pH by lactobacilli growth are some

of the mechanisms that inhibit *Candida* growth (Jiang *et al.* 2015). However, *C. albicans* is able to tolerate fluctuations in environmental pH (Biswas *et al.* 2007) and Wang *et al.* (2017) also proposed that anti-*Candida* activities are not largely attributed to low pH but rather directly due to the presence of *Lactobacillus crispatus*.

Looking at BMI, using the same cohort as the one used in this study, Lennard *et al.* (2018) showed that females with a higher BMI had a higher likelihood of having a non-*Lactobacillus* dominant VMB. Interestingly, Donders *et al.* (2002) found that women with recurrent VVC (RVVC) had a significantly greater mean BMI than the non-VVC infected control group.

Clinical and experimental observations suggest that local immunity is important in host defense against candidiasis (Fidel *et al.* 2003; Taylor *et al.* 2000). Accordingly, cytokines and chemokines are present in the vaginal mucosa during *C. albicans* infections (Steele and Fidel 2020), as a result, most patients with VVC are diagnosed by the presence of vulvar inflammation (Yano *et al.* 2019).

Given the above, more research is required to understand how the vaginal mycobiome changes in response to the host environment and vice versa. This information would likely clarify associations between the mycobiome and clinical features. Since fungi are also suggested to assist in regulating the composition of microbial communities (Dworecka-Kaszak *et al.* 2016; Ma *et al.* 2020; Wisecaver *et al.* 2014), a better understanding of factors that maintain stable fungal vaginal communities will help develop strategies to establish and sustain sexual and reproductive health (Gajer *et al.* 2012).

The pathogenesis of numerous single fungal species has been well studied, whereas the correlation between fungal diversity and pathogenicity is less clearly defined (Cui *et al.* 2013). It is assumed that fungal diversity should be higher in more severe cases of a disease (Cui *et al.* 2013). In corroboration, a vaginal study by Godoy-Vitorino *et al.* (2018) found that fungal diversity is higher in high-risk Human papillomavirus (HPV) individuals compared to low-risk individuals. However, complex interspecies dynamics need to be studied to understand the relationships that exist within the mycobiome during optimal and non-optimal states. Health outcomes are usually determined by polymicrobial interactions between microorganisms and the human host; rather than via one-on-one host-microbe relationships (Lai *et al.* 2019).

Therefore, when understanding the role that a fungal species or a fungal community plays in diseases, it is essential to consider that fungi do not exist in isolation (Witherden *et al.* 2017).

Recent research has suggested that the bacteriome and mycobiome directly impact each other (Oever and Netea 2014; Krüger *et al.* 2019). Thus, shifts in one community modulate the community structure of the other. However, there is minimal data available on fungal interactions and possible coexistence patterns with bacterial communities (Drell *et al.* 2013). Of the research conducted, the fungal genera best studied for their association with bacteria is *Candida*. One example includes the study by Hermann *et al.* (1999), which reported that of their clinical specimens (oral swabs, throat swabs, respiratory secretions, gastric juices, faeces samples, urine samples, genital swabs, blood samples, wound secretions, and skin swabs) isolated for *Candida*, many were accompanied by bacteria. Yet the consequences of such bacterial-fungal interactions on the host are largely unknown. The lack of this research leaves some key questions unanswered: Are fungi a threat that needs to be controlled by the host and bacteriome? Do some fungi have beneficial effects on the bacteriome? In the context of dysbiosis, does co-localization of fungi and bacteria affect disease risk? (Krüger *et al.* 2019). Therefore, consideration of fungi along with other kingdoms would provide a more comprehensive understanding of the microbiome as a whole. With this information, it may be possible to improve current STI and vaginal dysbiosis management strategies for females and enable therapeutic advances, particularly in locations where vaginal diseases are prevalent (Masson *et al.* 2019).

Methods

Clustering

Unsupervised machine learning of fungal protein intensity data was performed by *k*-means clustering using the *factoextra* package in *R*. To identify the number of optimal clusters (*k*) to split the dataset into, the *cluster R* package was used (Charrad *et al.* 2015). The silhouette method was chosen to identify the optimal number of clusters (*k*). The silhouette method computes silhouette coefficients of each object that measure how much an object is similar to its own cluster compared to other clusters. For this, a range of candidate values of *k* are chosen, we chose a range between 2 and 5, and the silhouette method trains *k*-Means clustering for each value of *k*. The Euclidean distance was used to classify observations into clusters.

Effect of clinical variables

To investigate the effects of HC, vaginal swab pH, BMI, Nugent Score, pro-inflammatory cytokines, and chemokines on vaginal fungal communities, a redundancy analysis (RDA) plot was used. RDA utilized the *EnvFit* function within the *Vegan* package in *R* to determine covariates significantly associated with the mycobiome profile (Oksanen *et al.* 2017).

Correlations

Relationships between fungal species identified during taxonomic analysis in chapter 3 were interrogated using a correlation analysis. Spearman's rank correlation was chosen to measure the relationship between log2 normalized iBAQ intensity abundances. Fungal-bacterial relationships were also investigated using fungal species identified to be differentially abundant in chapter 3 and selecting bacterial species found to be associated with BV and inflammation in the literature from our metaproteomic dataset (Gardner and Dukes 1954; Schwebke and Lawing 2001; Fethers *et al.* 2012; McMillan *et al.* 2015; Ceccarani *et al.* 2019). In addition, we investigated the correlation between fungal species and clinical variables recorded in this study. Correlation plots were generated in *R* using the *Hmisc* and *Corrplot* packages (Harrell and Dupont 2008, Wei and Wei 2016).

Bayesian Network construction

Fungal-bacterial relationships were further interrogated using the log2 normalized iBAQ intensity abundances of fungal species identified to be differentially abundant in chapter 3 and bacterial species found to be associated with BV and inflammation in the literature from our metaproteomic dataset (Gardner and Dukes 1954; Schwebke and Lawing 2001; Fethers *et al.* 2012; McMillan *et al.* 2015; Ceccarani *et al.* 2019) using a Bayesian network (BN). In addition, we investigated the correlation between fungal species and clinical variables recorded in this study. BNs were constructed in *R* using the *bnlearn* package (Scutari 2010).

Disease-associated fungi and clinical variables

To predict fungal proteins/species and clinical variables associated with non-optimal states (BV and vaginal inflammation), an ENSEMBLE bagging decision tree was used. The *caret*

(Classification And REgression Training) package was used to perform ENSEMBLE bagging in *R* (Kuhn *et al.* 2020). We selected 80% of our data for the training dataset, and the remaining 20% was used as the test dataset. We modified the resampling method using 10 repeated cross-validations.

We created a bagging pipeline with Python's *sklearn*'s *BaggingClassifier* (Pedregosa *et al.* 2012). For protein data, default parameters were used. For clinical variables data, the n_estimators parameter was set to 200 (200 classification trees) and the max_samples parameter to 100 (number of samples drawn to train each estimator) in order to obtain the best accuracy for the test data set.

Results

Clustering

With unsupervised *k*-means clustering, we aimed to determine how many optimal groups individuals clustered into using iBAQ intensity data for fungal proteins. We identified the number of optimal clusters (*k*) to be used as the hyperparameter during *k*-means clustering using the Silhouette method (Fig 4.1). A high Silhouette Score is desirable, as the score reaches its global maximum at the optimal *k*. The Silhouette Score reached its peak at *k* =2 for both visits, thus the optimal number of clusters was 2 (Fig 4.1). The parameters we chose for *k*-means clustering analysis were *k* = 2, *iter.max* = 10, and *nstart* = 25. The *nstart* parameter attempts 25 initial configurations and reports on the best one. Clustering results assigned 44 individuals to cluster 1 and 69 individuals to cluster 2 for visit 1 (*n* = 113). For visit 2 (*n* = 89), cluster 1 had 56 individuals assigned, and cluster 2 had 33 individuals assigned.

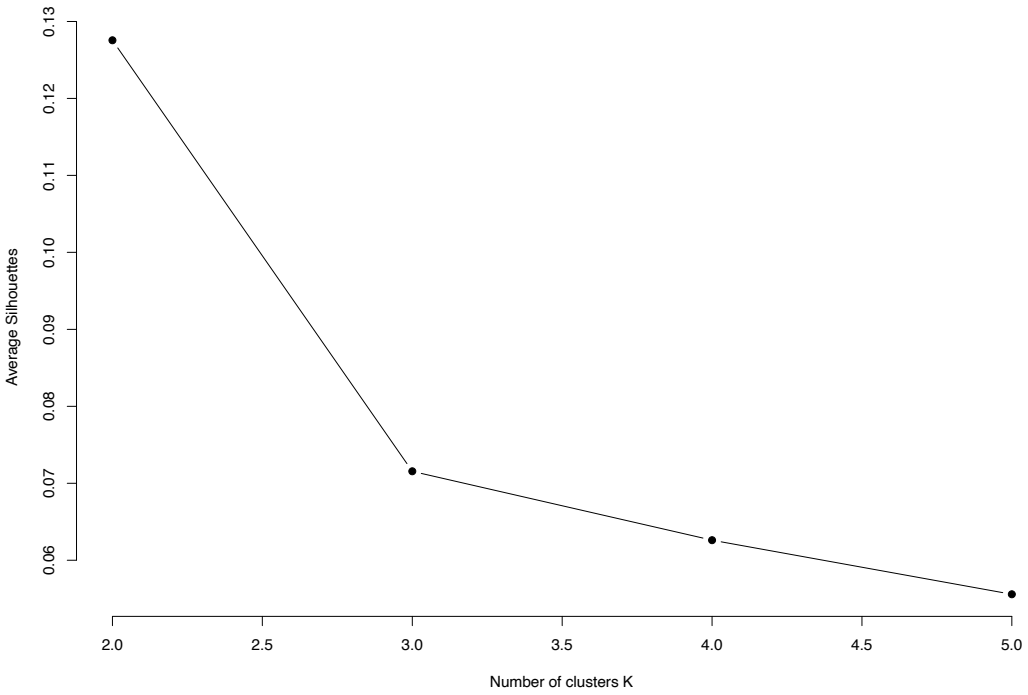
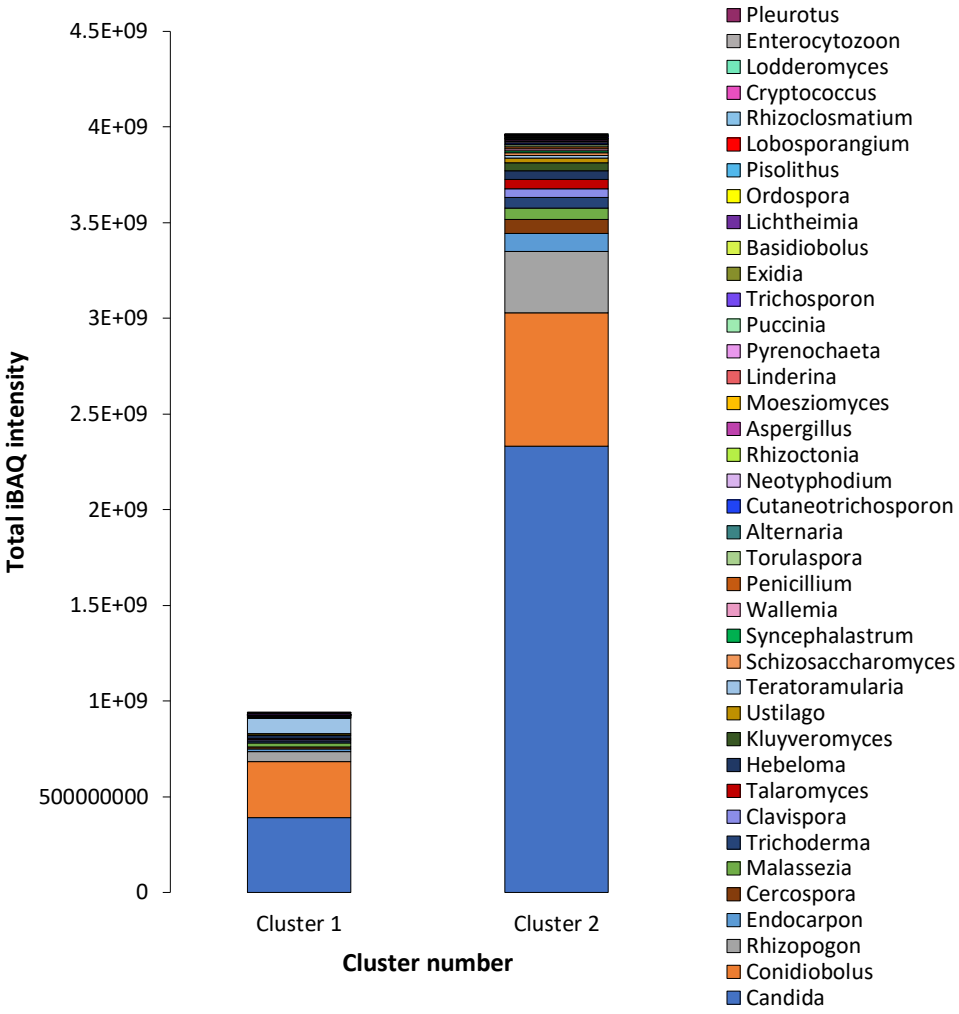


Figure 4.1. Investigating fungal protein intensity data using the silhouette method for an optimal number of clusters (*k*) to perform *k*-means clustering on the samples collected at visit 1. Individuals belonging to visit 1 showed an optimal clusters (*k*) value of 2.

Individuals in Cluster 1 were primarily BV positive, negative for *T. vaginalis*, and no individuals had a low vaginal pH (pH < 3.8). Individuals in Cluster 2 were majority BV negative, with the majority of individuals having a high pH, and positive for *C. trachomatis*. Both clusters had a similar percentage of each cytokine factor.

Both clusters included women with a high relative abundance for *Candida* (Fig 4.2b). In cluster 1, the proportion of *Candida* to *Condiobolus* is comparatively similar, where *Condiobolus* is relatively lower in cluster 2 in comparison to *Candida* (Fig 4.2a; 4.2b). Cluster 2 had a higher total fungal iBAQ intensity value for fungi, and correspondingly a larger total iBAQ intensity value for *Candida*, which is likely a result of 25 more individuals assigning to cluster 2 (Fig 4.2a). Regardless of the lower sample size cluster 1 had a larger total iBAQ intensity for *Teratoramularia* (Fig 4.2a).


4.5E+09
4E+09
3.5E+09
3E+09
2.5E+09
2E+09
1.5E+09
1E+09
500000000
0
Total iBAQ intensity
Cluster 1
Cluster 2
Cluster number
Pleurotus
Enterocytozoon
Lodderomyces
Cryptococcus
Rhizoclosmatium
Lobosporangium
Pisolithus
Ordospora
Lichtheimia
Basidiobolus
Exidia
Trichosporon
Puccinia
Pyrenochaeta
Linderina
Moesziomyces
Aspergillus
Rhizoctonia
Neotyphodium
Cutaneotrichosporon
Alternaria
Torulaspora
Penicillium
Wallemia
Syncephalastrum
Schizosaccharomyces
Teratoramularia
Ustilago
Kluyveromyces
Hebeloma
Talaromyces
Clavispora
Trichoderma
Malassezia
Cercospora
Endocarpon
Rhizopogon
Conidiobolus
Candida

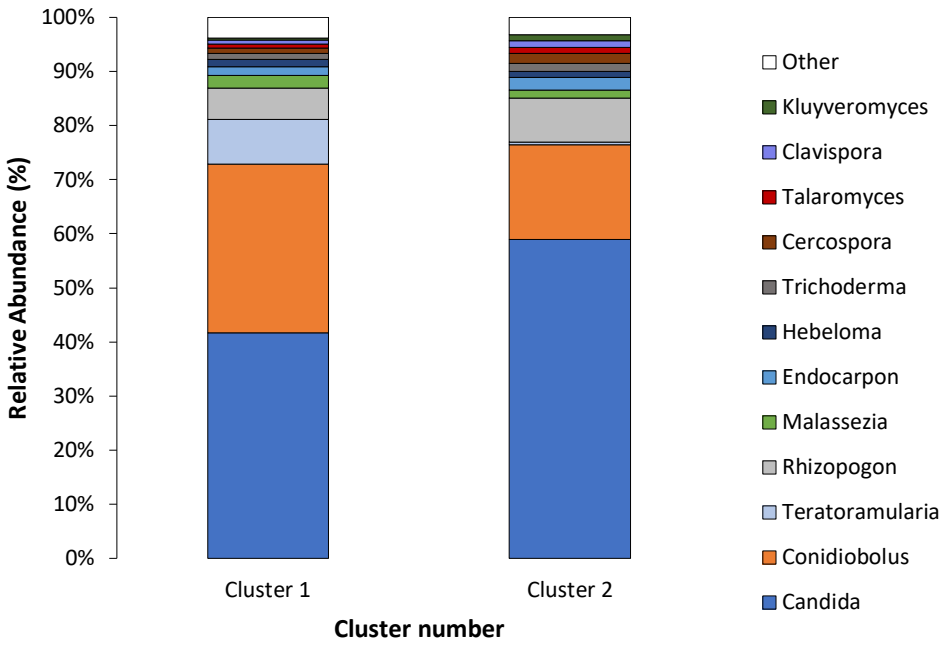


Figure 4.2. a) Total amount of fungal genus of each cluster determined by summing total fungal iBAQ intensity values of proteins assigning to each fungal genus; **b)** Relative abundance of fungal genera in each cluster that was determined using the k-means clustering method (*k* = 2), cluster 1 had 44 samples, and cluster 2 had 69 samples.

Table 4.1. Characteristics of individuals belonging to the two optimal clusters (*k* = 2) in visit 1.

	Characteristics	Cluster 1 (n = 44) (%)	Cluster 2 (n = 69) (%)
BV Status	**BV positive**	63.64	42.03
	BV negative	25	52.18
	BV intermediate	9.1	4.35
Cytokine Category	**High cytokines**	27.28	30.44
	Low cytokines	20.46	20.29
	Medium cytokines	52.28	49.28
Vaginal pH	**High pH**	25	62.32
	Normal pH	25	33.34
	Low pH	0	2.9
BMI	**Healthy**	18.19	39.14
	Obese	9.1	13.05
	Overweight	20.46	37.69
	Underweight	0	4.35

STI	***C. trachomatis***	11.37	37.69
	N. gonorrhoeae	2.28	5.8
	M. genitalium	6.82	2.9
	Herpes simplex virus 2	4.55	2.9
	T. vaginalis	0	4.35

Relationship between fungal populations and clinical variables

A Redundancy analysis (RDA) was performed to determine which predictor variables (*i.e.*, Nugent score, vaginal pH, STIs, BMI, HC, clue cells, pro-inflammatory cytokines, and chemokines) significantly explained the variation in fungal community composition (Fig 4.3). Five variables were associated with being significant ($r^2 > 0.05$; p-value < 0.04) drivers of fungal community composition. The proposed significant predictor variables included vaginal pH, Nugent score, clue cells, *C. trachomatis*, pro-inflammatory cytokines, and chemokines (Table 4.2; Fig 4.3). Among the predictor variables, Nugent Score had the most significant effect on the distribution (28%) (Table 4.2).

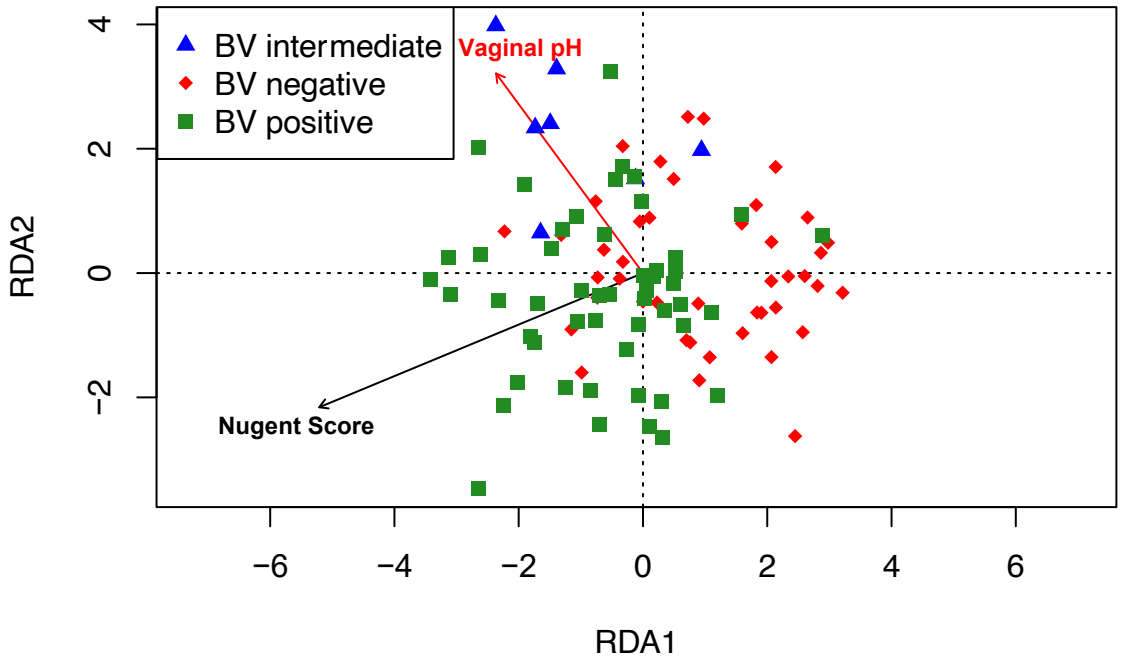


Figure 4.3. Redundancy analysis (RDA) showing fungal community composition in response to clinical factors at visit 1 (*n* = 113). Arrows represent the degree of the contribution of the factor to the fungal community structure. Factors are not depicted on the RDA plot as they are not quantifiable like vectors.

Table 4.2. Effect of extrinsic factors on the distribution of samples according to the RDA based on fungal proteins.

Clinical variable	R^2	*p*-val
Nugent Score	0.28	0.001
Vaginal pH	0.14	0.002
C. trachomatis	0.04	0.020
Cytokines	0.05	0.044
Chemokines	0.15	0.001
Clue cells	0.14	0.001

Correlations

Fungi-fungi relationships

We conducted Spearman's correlation analyses to explore fungal-fungal associations among fungal species found in our study using the samples collected from visit 1 (*n* = 113) (Fig 4.4). From the correlation plot, there were a greater number of positive correlations than negative correlations observed (Fig 4.4). Our results revealed that the strongest positive correlation occurred between *Ordospora colligata* and *Rhizoclosmatium globosum* (Fig 4.4; Table 4.3). Significant moderate positive correlations were observed between *Alternaria alternata* and *Linderina pennispora*; and *Malassezia sympodialis* and *Ustilago maydis* (Fig 4.4; Table 4.3). *Rhizopogon vinicolor* also had several significant moderate positive correlations with fungal species, which included *Candida orthopsilosis*, *Kluyveromyces marxianus*, and *Endocarpon pusillum* (Fig 4.4; Table 4.3). Many positive correlations were also observed between *Candida* species, including *C. orthopsilosis*, *C. lusitaniae*, and *C. thasaenensis* with *C. albicans*; and *C. orthopsilosis* with *K. marxianus* and *C. dubliniensis* (Fig 4.4; Table 4.3).

In contrast, negative correlations observed were much weaker in association than positive correlations (Fig 4.4). The strongest negative correlation was moderate and occurred between *A. alternata and Candida albicans* (Fig 4.4; Table 4.4). *Torulaspora delbrueckii* had several

significant moderate correlations with fungal species, including *M. sympodialis* and *Hebeloma cylindrosum* (Fig 4.4; Table 4.4). The two fungal species, *C. albicans* and *M. sympodialis* had a significant moderate negative correlation (Fig 4.4; Table 4.4).

Visit 2 samples (*n* = 89) confirmed the positive correlations that were observed between fungal species in visit 1 (Supplementary Figure S2). However, the strongest negative correlation was now observed between *M. sympodialis* and *T. delbrueckii* (cor = -0.71; *p* = 3.77e-15).

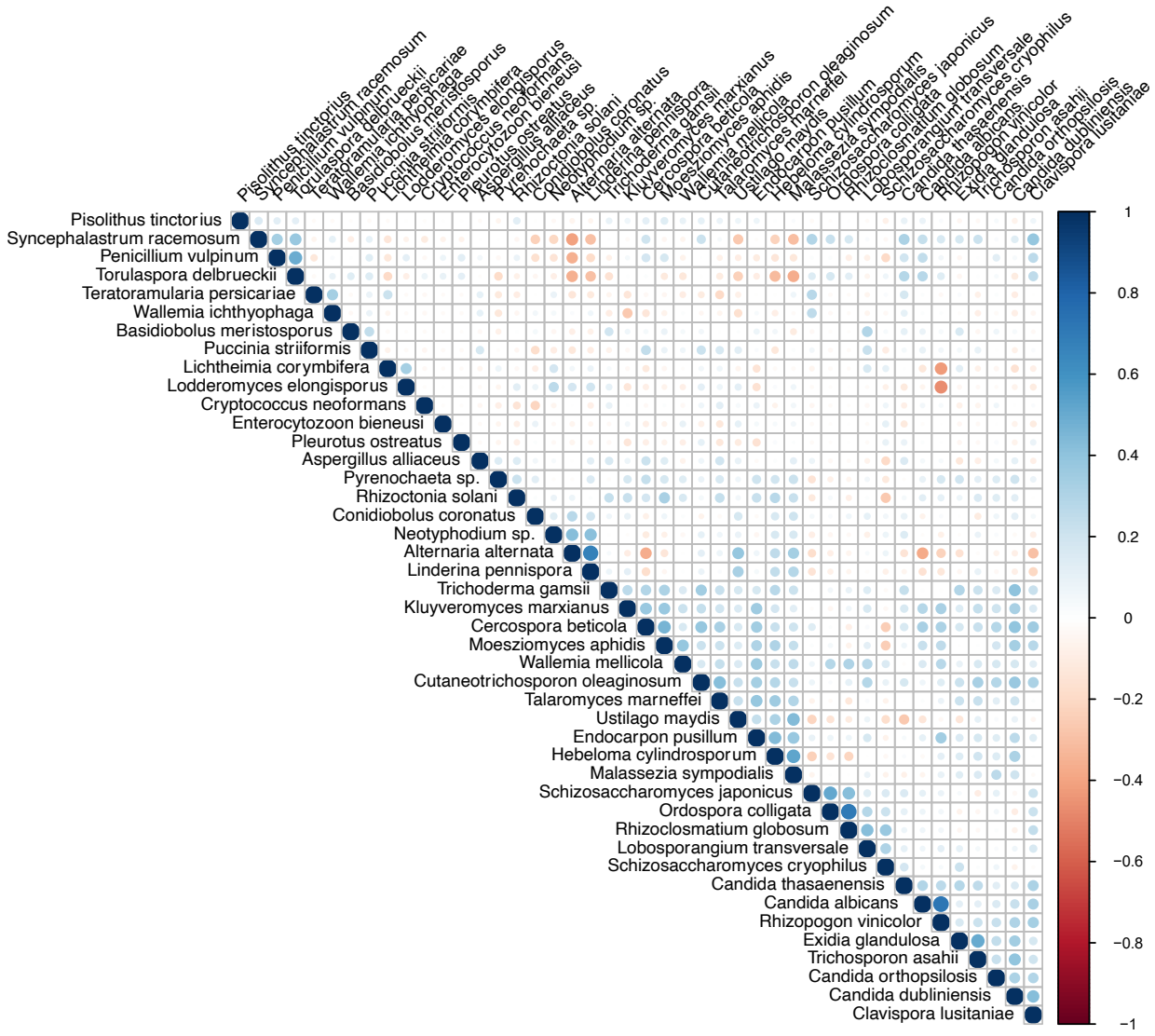


Figure 4.4. Correlations between fungal species were identified in this study in visit 1 (*n* = 113). The size of the circle represents the strength of the correlation (r-value). Red circles represent negative correlations (negative r-value), and blue circles represent positive correlations (positive r-value). More positive correlations between fungal species are observed in comparison to negative correlations. The

strongest positive correlation occurred between *Ordospora colligata* and *Rhizoclosmatium globosum.* The strongest negative correlation was between *Malassezia sympodialis* and *Torulaspora delbrueckii.*

Table 4.3. Significant positive correlations between fungal species from samples collected in visit 1 (*n* = 113) as identified from Spearman's rank correlation.

Fungal species	Fungal species	Correlation (r)	*p-value*
Ordospora colligata	***Rhizoclosmatium globosum***	0.70	0 (<1e-314)
Alternaria alternata	***Linderina pennispora***	0.68	0 (<1e-314)
Candida orthopsilosis	***Rhizopogon vinicolor***	0.68	2.2e-16
Malassezia sympodialis	***Ustilago maydis***	0.66	8.88e-16
Endocarpon pusillum	***Rhizopogon vinicolor***	0.65	4.88e-15
Kluyveromyces marxianus	***Rhizopogon vinicolor***	0.652	6.22e-15
Candida orthopsilosis	***Kluyveromyces marxianus***	0.61	1.17e-12
Candida albicans	***Candida orthopsilosis***	0.59	3.80e-12
Candida dubliniensis	***Candida orthopsilosis***	0.53	2.15e-09
Candida thasaenensis	***Candida albicans***	0.52	2.31e-09
Candida albicans	***Clavispora lusitaniae***	0.52	2.48e-09

Table 4.4. Significant negative correlations between fungal species from samples collected in visit 1 (*n* = 113) as identified from Spearman's rank correlation.

Fungal species	Fungal species	Correlation (r)	p-value
Alternaria alternata	***Candida albicans***	-0.52	3.61e-09
Malassezia sympodialis	***Torulaspora delbrueckii***	-0.51	7.35e-09
Candida albicans	***Linderina pennispora***	-0.41	7.50e-06
Hebeloma cylindrosporum	***Torulaspora delbrueckii***	-0.40	8.88e-06
Candida albicans	***Malassezia sympodialis***	-0.40	1.36e-05

Fungi-bacterial relationships

We conducted correlation analyses to explore the relationships between differentially expressed fungal species identified in chapter 3 (*n* = 14) and bacteria (identified using our metaproteomics data) found to be associated with BV states in the literature (*n* = 11) (Fig 4.5).

Spearman's correlation demonstrated that *M. sympodialis* had the strongest positive correlation with the Gram-variable bacterium *Gardnerella vaginalis* (Fig 4.5; Table 4.5). In addition, both *M. sympodialis* and *A. alternata* had many positive associations with various Gram-variable and Gram-negative bacteria including, *G. vaginalis*, *Mobiluncus mulieris*, *Prevotella amnii*, *Prevotella sp. S7-1-8*, *Megasphaera genomosp. type_1*, *Escherichia coli*, *Megasphaera sp. UPII 135-E*, and *Sneathia amnii* (Fig 4.5; Table 4.5; Table 4.6). *Candida* species (*C. albicans*, *C. thasaenensis, Clavispora lusitaniae*) had a moderate positive association with Gram-positive bacteria, *L. crispatus*, *L. iners*, and *L. jensenii* (Fig 4.5; Table 4.7; Table 4.9; Table 4.12).

Negative correlations could also be observed between many bacterial and fungal species (Fig 4.5). The strongest negative correlation was observed between *G. vaginalis* and *C. albicans* (Fig 4.5; Table 4.7). Furthermore, *T. delbrueckii, C. albicans*, *Penicillium vulpinum* had many negative associations with *G. vaginalis*, *P. amnii*, *Megasphaera genomosp. type_1, Megasphaera sp. UPII 135-E,* and *S. amnii* (Fig 4.5; Table 4.7; Table 4.8; Table 4.10). Conversely, *M. sympodialis* and *A. alternata* had negative correlations with Gram-positive bacteria, *Lactobacillus crispatus*, L. iners, and L. jensenii (Fig 4.5; Table 4.5; Table 4.6).

Correlations between fungal and bacterial species using visit 2 samples (*n* = 89) confirm many of the correlations seen in visit 1 results (Fig S3). However, the strongest negative correlation was observed between *G. vaginalis* and *T. delbrueckii* (cor = -0.70; *p* = 2.09e-14).

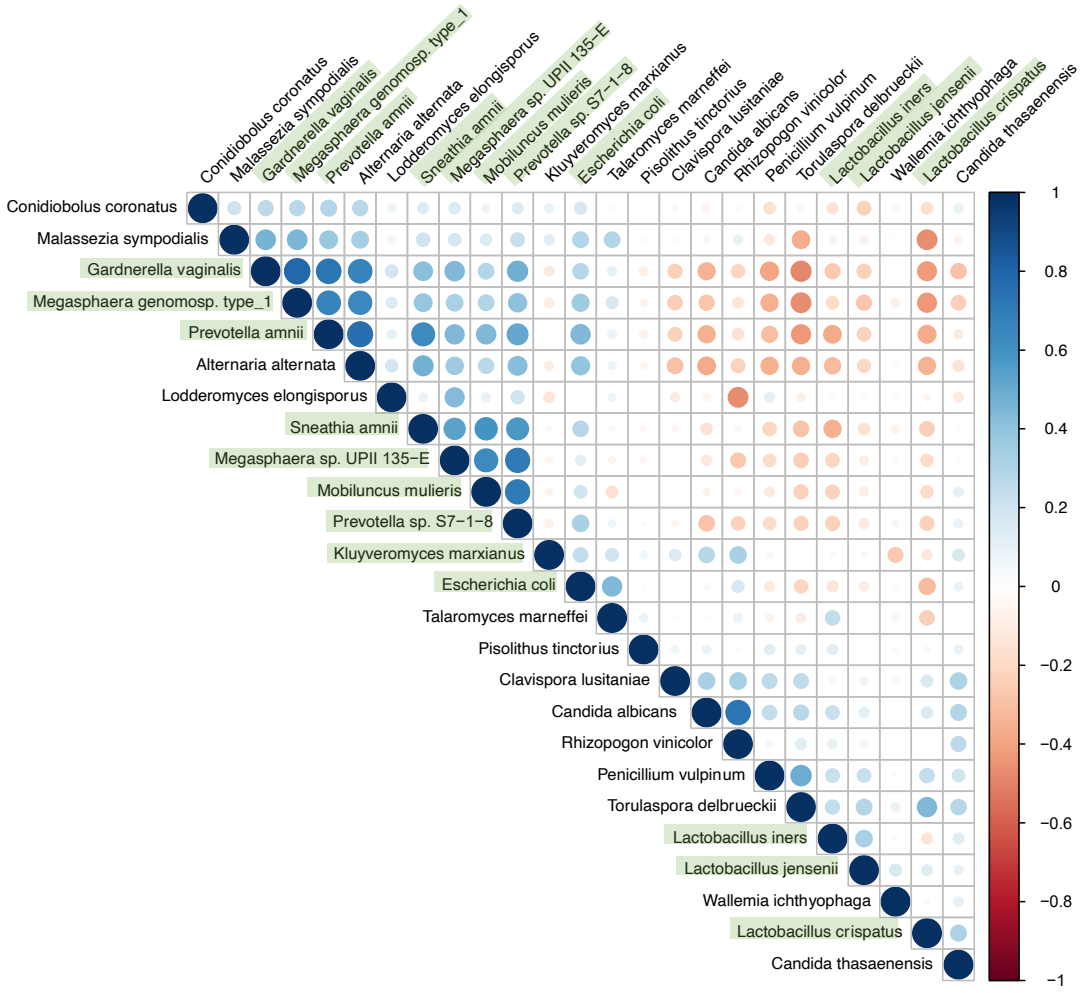


Figure 4.5. Correlations between differentially expressed fungal species (*n* = 14) and bacterial species associated with BV and Inflammation from visit 1 samples (*n* = 113). Bacterial species are highlighted in green. Red circles represent negative correlations, whereas blue circles represent positive correlations. Darker colours and circle size represent stronger correlations. *M. sympodialis* had the strongest positive correlation with the bacterium *G. vaginalis.* The strongest negative correlation was observed between *C. albicans* and *G. vaginalis*.

Table 4.5. Correlations between bacterial species and *Malassezia sympodialis.*

Bacterial Species	Fungal Species	Correlation (r)	*p*-value
Gardnerella vaginalis	*Malassezia sympodialis*	0.80	0
Prevotella amnii	*Malassezia sympodialis*	0.74	0
Megasphaera genomosp. type_1	*Malassezia sympodialis*	0.72	0
Megasphaera sp. UPII 135-E	*Malassezia sympodialis*	0.48	5.87e-08
Prevotella sp. S7-1-8	*Malassezia sympodialis*	0.48	1.01e-07

Sneathia amnii	*Malassezia sympodialis*	0.46	2.67e-07
Escherichia coli	*Malassezia sympodialis*	0.32	0.0004
Lactobacillus crispatus	*Malassezia sympodialis*	-0.49	4.95e-08

Table 4.6. Correlations between bacterial species and *Alternaria alternata.*

Bacterial Species	Fungal Species	Correlation (r)	*p-value*
Prevotella amnii	*Alternaria alternata*	0.70	0
Gardnerella vaginalis	*Alternaria alternata*	0.62	2.08e-13
Prevotella sp. S7-1-8	*Alternaria alternata*	0.62	2.27e-13
Megasphaera genomosp. type_1	*Alternaria alternata*	0.59	6.00e-12
Megasphaera sp. UPII 135-E	*Alternaria alternata*	0.49	4.998e-08
Sneathia amnii	*Alternaria alternata*	0.47	1.42e-07
Mobiluncus mulieris	*Alternaria alternata*	0.25	0.007
Lactobacillus crispatus	*Alternaria alternata*	-0.42	2.64e-06
Lactobacillus iners	*Alternaria alternata*	-0.30	0.0014
Lactobacillus jensenii	*Alternaria alternata*	-0.25	0.0076

Table 4.7. Correlations between bacterial species and *Candida albicans.*

Bacterial Species	Fungal Species	Correlation (r)	p-value
Lactobacillus crispatus	*Candida albicans*	0.38	3.04e-05
Lactobacillus iners	*Candida albicans*	0.28	0.002
Lactobacillus jensenii	*Candida albicans*	0.27	0.003
Gardnerella vaginalis	*Candida albicans*	-0.58	1.65e-11
Alternaria alternata	*Candida albicans*	-0.52	3.61e-09
Prevotella amnii	*Candida albicans*	-0.49	3.60 e-08
Megasphaera genomosp. type_1	*Candida albicans*	-0.46	2.16e-07
Megasphaera sp. UPII 135-E	*Candida albicans*	-0.40	1.25e-05

Prevotella sp. S7-1-8	*Candida albicans*	-0.37	4.86e-05
Sneathia amnii	*Candida albicans*	-0.25	0.0085

Table 4.8. Correlations between bacterial species and *Torulaspora delbrueckii.*

Bacterial Species	Fungal Species	Correlation	*p*-value
Lactobacillus crispatus	*Torulaspora delbrueckii*	0.56	1.04e-10
Lactobacillus jensenii	*Torulaspora delbrueckii*	0.38	3.11e-05
Lactobacillus iners	*Torulaspora delbrueckii*	0.24	0.011
Gardnerella vaginalis	*Torulaspora delbrueckii*	-0.55	2.07e-10
Malassezia sympodialis	*Torulaspora delbrueckii*	-0.51	7.35e-09
Prevotella amnii	*Torulaspora delbrueckii*	-0.50	1.21e-08
Megasphaera genomosp. type_1	*Torulaspora delbrueckii*	-0.50	1.73e-08
Megasphaera sp. UPII 135-E	*Torulaspora delbrueckii*	-0.38	2.90e-05
Sneathia amnii	*Torulaspora delbrueckii*	-0.35	0.00014
Escherichia coli	*Torulaspora delbrueckii*	-0.23	0.014
Prevotella sp. S7-1-8	*Torulaspora delbrueckii*	-0.22	0.016

Table 4.9. Correlations between bacterial species and *Candida thasaenensis.*

Bacterial Species	Fungal Species	Correlation (r)	p-value
Lactobacillus crispatus	*Candida thasaenensis*	0.44	8.57e-07
Lactobacillus jensenii	*Candida thasaenensis*	0.30	0.001
Gardnerella vaginalis	*Candida thasaenensis*	-0.35	0.0001
Megasphaera genomosp. type_1	*Candida thasaenensis*	-0.32	0.0006
Prevotella amnii	*Candida thasaenensis*	-0.21	0.02

Table 4.10. Correlations between bacterial species and *Penicillium vulpinum.*

Bacterial Species	Fungal Species	Correlation (r)	p-value

Gardnerella vaginalis	*Penicillium vulpinum*	-0.45	7.05e-07
Prevotella amnii	*Penicillium vulpinum*	-0.34	0.00022
Megasphaera genomosp. type_1	*Penicillium vulpinum*	-0.33	0.0003
Megasphaera sp. UPII 135-E	*Penicillium vulpinum*	-0.32	0.0005
Prevotella sp. S7-1-8	*Penicillium vulpinum*	-0.28	0.002
Sneathia amnii	*Penicillium vulpinum*	-0.26	0.005

Table 4.11. Correlations between bacterial species and *Conidiobolus coronatus.*

Bacterial Species	Fungal Species	Correlation (r)	p-value
Gardnerella vaginalis	*Conidiobolus coronatus*	0.37	6.19e-05
Megasphaera genomosp. type_1	*Conidiobolus coronatus*	0.36	7.84e-05
Prevotella amnii	*Conidiobolus coronatus*	0.34	0.00025
Megasphaera sp. UPII 135-E	*Conidiobolus coronatus*	0.30	0.0011
Lactobacillus crispatus	*Conidiobolus coronatus*	-0.36	7.57e-05
Lactobacillus jensenii	*Conidiobolus coronatus*	-0.29	0.002
Lactobacillus iners	*Conidiobolus coronatus*	-0.22	0.02

Table 4.12. Correlations between bacterial species and *Clavispora lusitaniae.*

Bacterial Species	Fungal Species	Correlation (r)	p-value
Gardnerella vaginalis	*Clavispora lusitaniae*	-0.30	0.001
Prevotella amnii	*Clavispora lusitaniae*	-0.27	0.0039
Megasphaera genomosp. type_1	*Clavispora lusitaniae*	-0.25	0.0069
Prevotella sp. S7-1-8	*Clavispora lusitaniae*	-0.23	0.016

A Bayesian Network (BN) analysis allowed us to confirm many of the fungal-bacterial associations we identified from the Spearman's correlation plot (Fig 4.6). Our BN showed the strongest associations were between the Gram-negative bacteria *P. amnii* and *A. alternata*

(strength = 0.95) (Fig 4.6; Table S1), and between the Gram-positive bacteria *L. iners* with *T. marneffei* (strength = 0.83) (Fig 4.6; Table S1). Both the Gram-positive bacteria *L. crispatus* and the Gram-variable *G. vaginalis* showed an association with the fungal species *M. sympodialis* (Fig 4.6; Table S1). Moreover, *L. crispatus* also had an association with *T. delbrueckii* and *C. thasanensis* (Fig 4.6; Table S1). Another fairly strong association was noted between *S. amnii* and *R. vinicolor* (Fig 4.6; Table S1).

The BN also allowed us to examine the relationship between fungal species. From this, the strongest associations were observed between *C. albicans* and *R. vinicolor* (strength = 1), and *T. delbrueckii* with *P. vulpinum* (0.95) (Fig 4.6; Table S1). The *Candida* species, *C. albicans and C. thasanensis showed an association* (Fig 4.6; Table S1). *W. ichthyophaga* and *R. vinicolor* both showed an association with *K. marxianus* (strength = 0.79; strength = 0.59) (Fig 4.6; Table S1). Moreover, *R. vinicolor* also showed an association with *C. lusitaniae* (strength = 0.54) (Fig 4.6; Table S1).

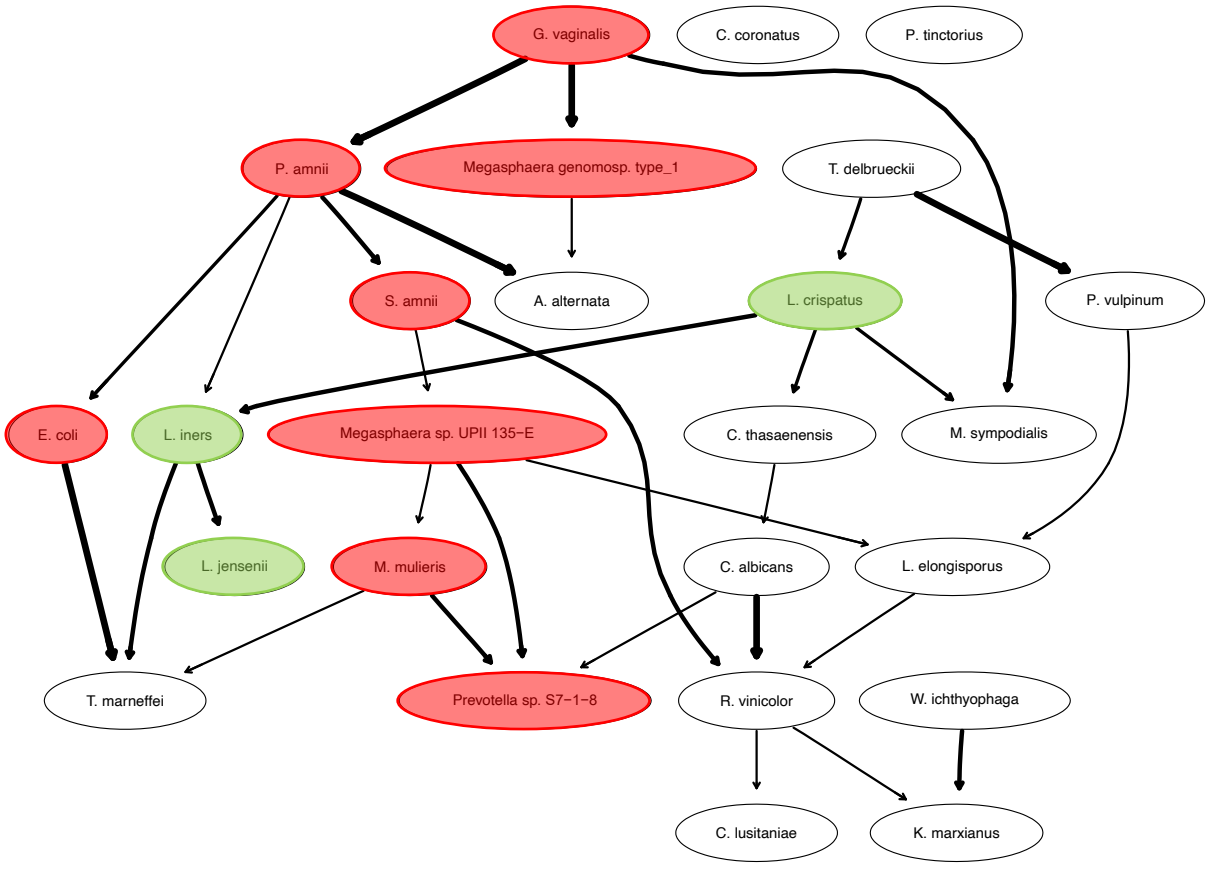


Figure 4.6. Bayesian Network depicting the associations between differentially abundant fungal species (*n* = 14) and optimal (green) and non-optimal bacterial (red) species linked with BV and

inflammation from the literature (*n* = 11) using metaproteomic data at visit 1 (*n* = 113). Arrows represent the microbial species that have an association with other microbial species. The thickness of the arrow represents the strength of the association.

Correlations between fungi and clinical variables

We also quantified the associations between fungal species (*n* = 45) and clinical factors at visit 1 using Spearman's correlation tests.

As can be seen from the correlation plot, *M. sympodialis*, *A. alternata*, and *C. coronatus* had a positive correlation with Nugent score (Fig 4.7; Table 4.13). Of these species, *M. sympodialis* had the strongest negative correlation with Nugent score (Fig 4.7; Table 4.13). Conversely, yeast species (*C. abicans*, *C. thasaenensis*, *C. lusitaniae*, *C. orthopsilosis*, and *T. delbrueckii*) were negatively correlated with Nugent score and positively correlated with clue cells (Fig 4.7; Table 4.13; Table 4.14).

When looking at associations between chemokines, the *Candida* species, *C. albicans* and *K. marxianus*, had a weak positive correlation with low chemokines, whereas *C. coronatus* and *A. alternata* had weak negative correlations with this clinical variable (Fig 4.7; Table 4.17). *R. vinicolor*, *E. pusillum*, *C. dubliniensis*, *Moesziomyces aphidis*, *Trichoderma gamsii* all had negative correlations with high chemokines (Fig 4.7; Table 4.18).

H. cylindrosporum and *E. pusillum* had a weak positive correlation with low cytokines and vaginal pH, whereas *T.delbrueckii* and *P. vulpinum* had weak negative correlations with these variables (Fig 4.7; Table 4.16; Table 4.19). Looking at only pH, *M. sympodialis*, *L. pennispora*, *Wallemia mellicola*, *A. alternata*, and *R. vinicolor* had positive correlations with vaginal pH (Fig 4.7; Table 4.19).

Fungal species were also examined for their associations with STIs. *M. sympodialis* had a positive correlation with *T. vaginalis*, whereas the fungal species *Rhizoclosmatium globosum*, *Schizosaccharomyces japonicus*, *O. colligata*, and *Schizosaccharomyces cryophilus* had a negative correlation with this STI (Fig 4.7; Table 4.20). The only significantly correlated fungal species with *N. gonorrhoea* was *Cryptococcus neoformans*. Likewise, the only significantly

correlated species with *M. genitalium* was *C. orthopsilosis*. The STIs, *C. trachomatis,* and HSV-2 had no significantly correlated fungal species.

Regarding HC methods, 5 different HC methods were evaluated, including Nuvaring (*n* = 1), Nur isterate (*n* = 68), DMPA (*n* = 15), oral contraceptive pills (OCP) (*n* = 5), and Implanon (*n* = 6). *Aspergillus alliaceus* was the only fungal species to show a significant correlation with Nuvaring, which was identified to be negative (Fig 4.7; Table 4.21). Similarly, *Rhizoctonia solani* was the only fungal species that had a significant correlation with Nur isterate, and was identified to be positive (Fig 4.7; Table 4.21). Some *Candida* spp. (*C. albicans* and *Kluyveromyces marxianus*) had a positive correlation with OCP, whereas *Lodderomyces elongisporus* had a negative correlation with OCP (*n* = 5) (Fig 4.7; Table 4.21). Implanon had a positive correlation with *T. delbrueckii* and negative correlations with *Rhizooctania solani*, and *Ustilago maydis* (Fig 36; Table 4.21). DMPA was the only HC that did not show any significant correlations with fungal species.

Fungal species that showed positive correlations with the presence of yeast cells (determined by Gram staining) were predominantly identified to be yeast species, with the exception of *Trichoderma gamsii* (Fig 4.7; Table 4.23). Interestingly, *C. albicans* did not show a significant positive correlation with yeast cells. However, the presence of yeast cells was positively correlated with total fungal intensity value, but this correlation was not found to be significant (Fig 4.7; Table 4.24).

The total fungal iBAQ intensity showed a negative correlation with Nugent score (Fig 4.7; Table 4.24). Whereas this variable showed significant positive correlations with *Candida* species (*C. albicans*, *C. orthopsilosis*, *C. thasaenensis*, *C. lusitaniae*, *K. marxianus*, *C. dubliniensis*). Of these, *C. albicans* had the strongest correlation with the total fungal iBAQ intensity value (Fig 4.7; Table 4.22). On the contrary, *A. alternata* and *L. pennispora* showed weak negative correlations with total fungal iBAQ intensity values (Fig 4.7; Table 4.22).

In terms of BMI, only one significant moderate correlation was noted between BMI and *Wallemia ichthyophaga* (cor = 0.34; *p*-value = 0.0007).

All correlations observed between fungal species and clinical variables were confirmed using samples for visit 2 (*n* = 89) in supplementary figure S4.

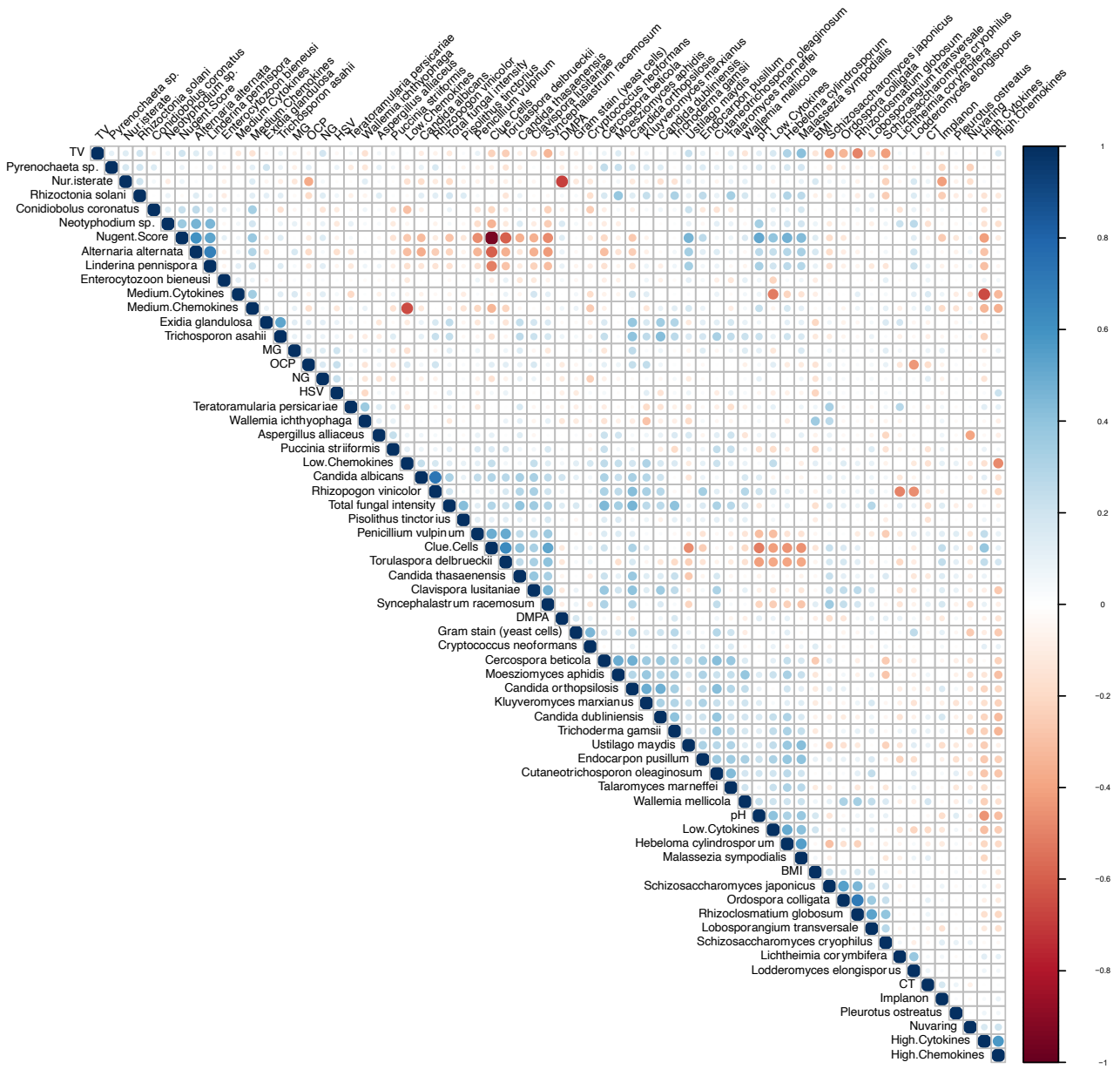


Figure 4.7. Correlations between fungal species identified (*n* = 45) and clinical variables at visit 1 (*n* = 113). Red circles represent negative correlations, whereas blue circles represent positive correlations. Darker colours and increased circle size represent stronger correlations.

Table 4.13. Significant correlations between fungal species and Nugent score.

Fungal Species	Clinical Variable	Correlation (r)	p-value
Malassezia sympodialis	**Nugent score**	0.75	0
Alternaria alternata	**Nugent score**	0.59	4.82e-10

Linderina pennispora	**Nugent score**	0.54	1.89e-08
Ustilago maydis	**Nugent score**	0.50	3.12e-07
Hebeloma cylindrosporum	**Nugent score**	0.43	1.55e-05
Neotyphodium sp.	**Nugent score**	0.38	0.00013
Conidiobolus coronatus	**Nugent score**	0.32	0.0015
Torulaspora delbrueckii	**Nugent score**	-0.60	1.38e-10
Candida albicans	**Nugent score**	-0.51	1.19e-07
Penicillium vulpinum	**Nugent score**	-0.47	1.34e-06
Clavispora lusitaniae	**Nugent score**	-0.42	2.78e-05
Syncephalastrum racemosum	**Nugent score**	-0.41	3.89e-05
Candida thasaenensis	**Nugent score**	-0.36	0.00036
Candida orthopsilosis	**Nugent score**	-0.23	0.024

Table 4.14. Significant correlations between fungal species and clue cells.

Fungal Species	Clinical Variable	Correlation (r)	p-value
Torulaspora delbrueckii	**Clue cells**	0.66	3.94e-13
Candida albicans	**Clue cells**	0.56	3.83e-09
Syncephalastrum racemosum	**Clue cells**	0.52	5.82e-08
Penicillium vulpinum	**Clue cells**	0.52	9.35e-08
Candida thasaenensis	**Clue cells**	0.47	1.72e-06
Clavispora lusitaniae	**Clue cells**	0.37	0.00024
Candida orthopsilosis	**Clue cells**	0.31	0.0023
Malassezia sympodialis	**Clue cells**	-0.73	0
Alternaria alternata	**Clue cells**	-0.58	6.28e-10
Ustilago maydis	**Clue cells**	-0.50	2.028e-07
Linderina pennispora	**Clue cells**	-0.50	3.30e-07
Hebeloma cylindrosporum	**Clue cells**	-0.39	8.44e-05
Conidiobolus coronatus	**Clue cells**	-0.35	0.000589

Table 4.15. Significant correlations between fungal species and high pro-inflammatory cytokines.

Fungal Species	Clinical Variable	Correlation (r)	p-value
Torulaspora delbrueckii	**High cytokines**	0.27	0.008

Endocarpon pusillum	**High cytokines**	-0.40	6.49e-05
Rhizopogon vinicolor	**High cytokines**	-0.33	0.001
Cutaneotrichosporon oleaginosum	**High cytokines**	-0.31	0.002
Hebeloma cylindrosporum	**High cytokines**	-0.30	0.003
Wallemia mellicola	**High cytokines**	-0.30	0.003
Linderina pennispora	**High cytokines**	-0.30	0.003

Table 4.16. Significant correlations between fungal species and low pro-inflammatory cytokines.

Fungal Species	Clinical Variable	Correlation (r)	p-value
Hebeloma cylindrosporum	**Low cytokines**	0.36	0.00039
Endocarpon pusillum	**Low cytokines**	0.28	0.0057
Torulaspora delbrueckii	**Low cytokines**	-0.42	2.58e-05
Penicillium vulpinum	**Low cytokines**	-0.33	0.001

Table 4.17. Significant correlations between fungal species and low factor chemokines.

Fungal species	Clinical Variable	Correlation (r)	p-value
Candida albicans	**Low chemokines**	0.32	0.001
Kluyveromyces marxianus	**Low chemokines**	0.31	0.002
Conidiobolus coronatus	**Low chemokines**	-0.32	0.002
Alternaria alternata	**Low chemokines**	-0.30	0.003

Table 4.18. Significant correlations between fungal species and high factor chemokines.

Fungal Species	Clinical Variable	Correlation (r)	p-value
Rhizopogon vinicolor	**High chemokines**	-0.39	0.00012
Endocarpon pusillum	**High chemokines**	-0.35	0.0005
Candida dubliniensis	**High chemokines**	-0.34	0.00066
Moesziomyces aphidis	**High chemokines**	-0.32	0.0014
Trichoderma gamsii	**High chemokines**	-0.31	0.0024

Table 4.19. Significant correlations between fungal species and vaginal pH.

Fungal Species	Clinical Variable	Correlation (r)	p-value
Malassezia sympodialis	**pH**	0.47	2.23e-06
Hebeloma cylindrosporum	**pH**	0.36	0.0004
Linderina pennispora	**pH**	0.34	0.0007
Neotyphodium sp.	**pH**	0.31	0.002
Endocarpon pusillum	**pH**	0.29	0.004
Wallemia mellicola	**pH**	0.29	0.004
Alternaria alternata	**pH**	0.28	0.005
Rhizopogon vinicolor	**pH**	0.27	0.008
Torulaspora delbrueckii	**pH**	-0.45	4.50e-06
Penicillium vulpinum	**pH**	-0.30	0.004

Table 4.20. Correlations between fungal species and sexually transmitted infections (STIs).

Fungal Species	Clinical Variable	Correlation (r)	p-value
Malassezia sympodialis	***T. vaginalis***	0.30	0.0033
Rhizoclosmatium globosum	***T. vaginalis***	-0.49	4.79e-07
Schizosaccharomyces japonicus	***T. vaginalis***	-0.36	0.00033
Syncephalastrum racemosum	***T. vaginalis***	-0.36	0.00043
Ordospora colligata	***T. vaginalis***	-0.33	0.0011
Schizosaccharomyces cryophilus	***T. vaginalis***	-0.32	0.0018
Torulaspora delbrueckii	***T. vaginalis***	-0.28	0.0057
Candida albicans	***N. gonorrhoea***	-0.16	0.13
Cryptococcus neoformans	***N. gonorrhoea***	-0.23	0.028
Candida orthopsilosis	***M. genitalium***	0.21	0.040

Table 4.21. Significant correlations between fungal species and hormonal contraception methods.

Fungal Species	Clinical Variable	Correlation (r)	p-value
Rhizoctonia solani	**Nur isterate**	0.24	0.022
Lodderomyces elongisporus	**OCP**	-0.44	1.03e-05
Kluyveromyces marxianus	**OCP**	0.23	0.024

Candida albicans	**OCP**	0.20	0.058
Rhizoctonia solani	**OCP**	-0.20	0.05
Torulaspora delbrueckii	**Implanon**	0.23	0.023
Rhizoctonia solani	**Implanon**	-0.26	0.010
Ustilago maydis	**Implanon**	-0.20	0.049
Aspergillus alliaceus	**Nuvaring**	-0.38	0.00014

Table 4.22. Significant correlations between fungal species and total fungal protein intensity.

Fungal species	Clinical Variable	Correlation (r)	p-value
Candida albicans	**Total fungal intensity**	0.65	2.00e-12
Candida orthopsilosis	**Total fungal intensity**	0.53	2.84e-08
Candida thasaenensis	**Total fungal intensity**	0.46	3.37e-06
Rhizopogon vinicolor	**Total fungal intensity**	0.44	1.05e-05
Clavispora lusitaniae	**Total fungal intensity**	0.42	2.69e-05
Cercospora beticola	**Total fungal intensity**	0.41	3.43e-05
Trichoderma gamsii	**Total fungal intensity**	0.39	8.08e-05
Talaromyces marneffei	**Total fungal intensity**	0.37	0.00021
Kluyveromyces marxianus	**Total fungal intensity**	0.36	0.0003
Syncephalastrum racemosum	**Total fungal intensity**	0.33	0.001
Candida dubliniensis	**Total fungal intensity**	0.30	0.003
Alternaria alternata	**Total fungal intensity**	-0.26	0.01
Linderina pennispora	**Total fungal intensity**	-0.21	0.038

Table 4.23. Significant correlations between clinical variables and total fungal protein intensity.

Clinical variable	Clinical Variable	Correlation (r)	*p*-value
Clue cells	**Total fungal intensity**	0.33	0.001
Gram stain (yeast cells)	**Total fungal intensity**	0.13	0.20

Nugent score	**Total fungal intensity**	-0.26	0.01

Table 4.24. Correlations between fungal species and yeast cells identified using Gram staining.

Fungal species	Clinical Variable	Correlation (r)	p-value
Cryptococcus neoformans	**Gram stain (yeast cells)**	0.45	4.51e-06
Candida orthopsilosis	**Gram stain (yeast cells)**	0.33	0.0012
Trichoderma gamsii	**Gram stain (yeast cells)**	0.27	0.0079
Cutaneotrichosporon oleaginosum	**Gram stain (yeast cells)**	0.27	0.0088
Lodderomyces elongisporus	**Gram stain (yeast cells)**	0.26	0.012
Kluyveromyces marxianus	**Gram stain (yeast cells)**	0.24	0.02
Talaromyces marneffei	**Gram stain (yeast cells)**	0.23	0.025
Candida dubliniensis	**Gram stain (yeast cells)**	0.22	0.034
Candida albicans	**Gram stain (yeast cells)**	0.14	0.17

Once again, a BN analysis allowed us to confirm most of the fungal-clinical variable relationships we identified from the above correlation plots (Fig 4.8). Our BN found the strongest associations were between BMI and *W. ichthyophaga*, and *M. sympodialis* with *T. vaginalis* (Fig 4.8; Table 25). *M. sympodialis* also showed an association with Nugent score, in addition to *A. alternata* (Fig 4.8; Table 25). Clue cells had associations with three species, including *T. delbrueckii*, *C. thasanensis*, and *P. vulpinum* (Fig 4.8; Table 25). The two fungal species, *T. marneffei* and *C. thasanensis* had an association with total fungal iBAQ intensity (Fig 4.8; Table 25). The yeast species, *L. elongisporus* had an association with yeast cells and OCP also showed an association with this species. (Fig 4.8; Table 25). Furthermore, the high cytokine factor had an association with *R. vinicolor* (Fig 4.8; Table 25).

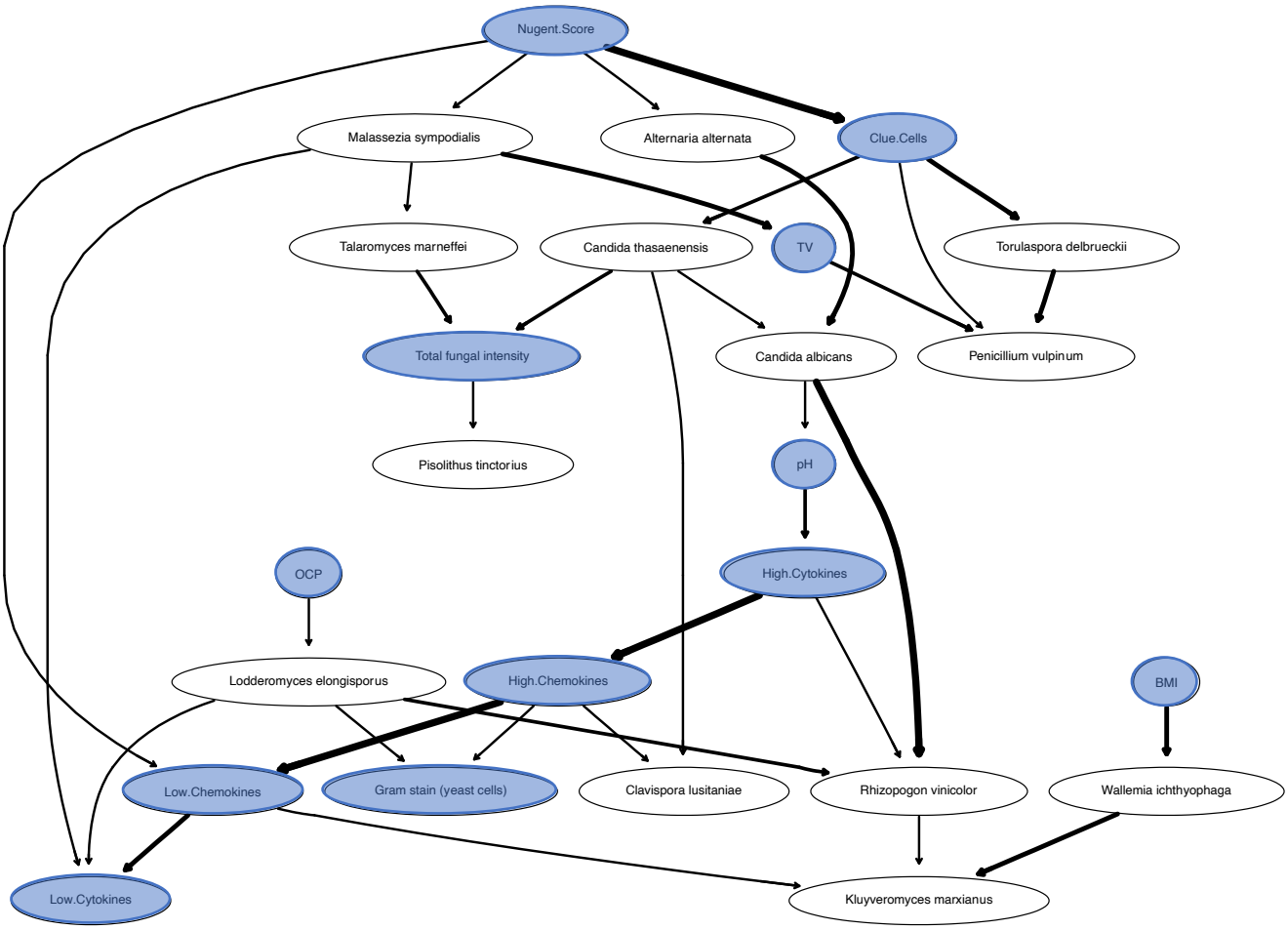


Figure 4.8. Bayesian Network depicting the associations between differentially abundant fungal species (*n* = 14) and clinical variables (blue) visit 1 (*n* = 113). Arrows represent associations between fungal species and clinical variables. Thickness of arrow represents strength of the association.

Table 4.25. Strength of associations between significantly differentially abundant fungal species and clinical variables as determined by a Bayesian Network analysis.

from	to	strength
BMI	*Wallemia ichthyophaga*	0.885
Malassezia sympodialis	TV	0.855
Clue cells	*Torulaspora delbrueckii*	0.85
Talaromyces marneffei	Total fungal intensity	0.715
Clue cells	*Candida thasaenensis*	0.71
Candida thasaenensis	Total fungal intensity	0.7
T. vaginalis (TV)	*Penicillium vulpinum*	0.685
Lodderomyces elongisporus	Gram stain (yeast cells)	0.655

OCP	*Lodderomyces elongisporus*	0.645
pH	*Candida albicans*	0.615
Malassezia sympodialis	Low cytokines	0.6
High cytokines	*Rhizopogon vinicolor*	0.54
Nugent score	*Alternaria alternata*	0.605
Clue cells	*Penicillium vulpinum*	0.6
High chemokines	*Clavispora lusitaniae*	0.575
Nugent score	*Malassezia sympodialis*	0.515

Fungal Proteins Associated with BV Status

To investigate fungal taxonomic signatures discriminating BV positive from BV negative individuals, we used machine learning approaches. The ENSEMBLE Bagging Decision Tree gave the best results when employing a classification approach on protein intensity data. For this analysis, we removed the BV intermediate state due to the low sample size (n = 7).

Our training classification model incorporated 80% of the dataset, which included 84 samples, 50 predictors (proteins), and 2 classes (BV positive and BV negative). We evaluated the performance of the classification model and the accuracy of the training dataset model was found to be a 100%, with a sensitivity and specificity of 100%, and a positive and negative prediction rate of 100%.

The test dataset included the remaining 20% of the dataset (n = 20) and had an accuracy of 90% (p-value = 0.01), this is how often the model predicts a result that will be correct. The model had a sensitivity of 100%, a specificity of 85%, and the positive prediction rate was 78%, and a negative prediction rate of 100%. The area under the curve (AUC) was determined to be 0.99.

The most important fungal species distinguishing the BV positive from the BV negative category were *Malassezia sympodialis* (M5EKD5), *Candida albicans* (A0A1D8PFR4), and *Candida thasaenensis* (J7M8M3) (Fig 4.9).

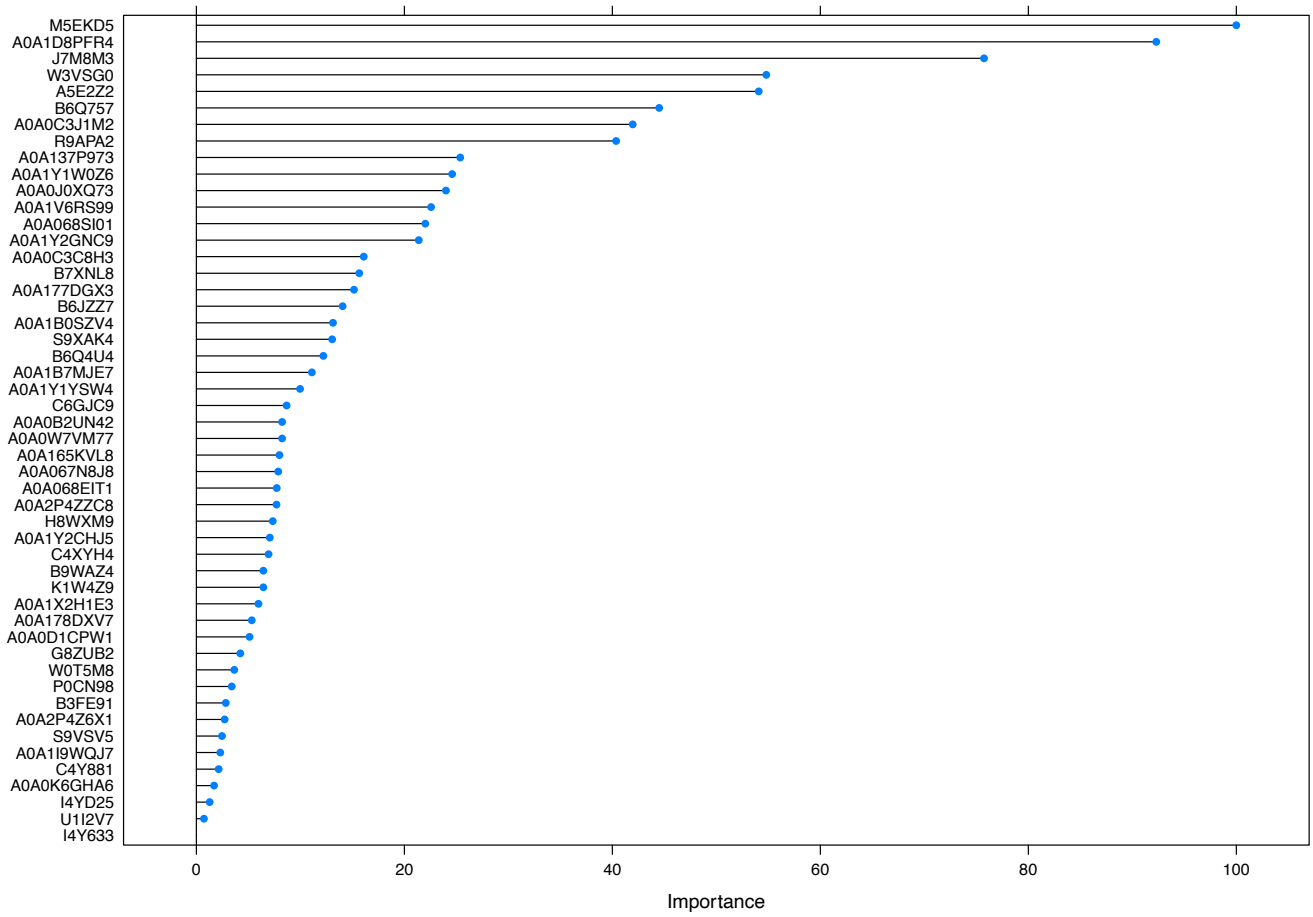


Figure 4.9. Identified fungal proteins and their importance in discriminating the BV positive group from the BV negative group. M5EKD5 (*Malassezia sympodialis*) was most important fungal species distinguishing the BV positive from the BV negative category.

Our decision tree elected A0A1D8PFR4 (*C. albicans*) as the root feature associated with BV status, thereafter the decision tree splits using the protein M5EKD5 (*M. sympodialis*), and then J7M8M3 (*C. thasaenensis*) (Fig 4.10). Individuals that an iBAQ value of < 23 for the *C. albicans* protein (A0A1D8PFR4) were immediately assigned as BV positive.

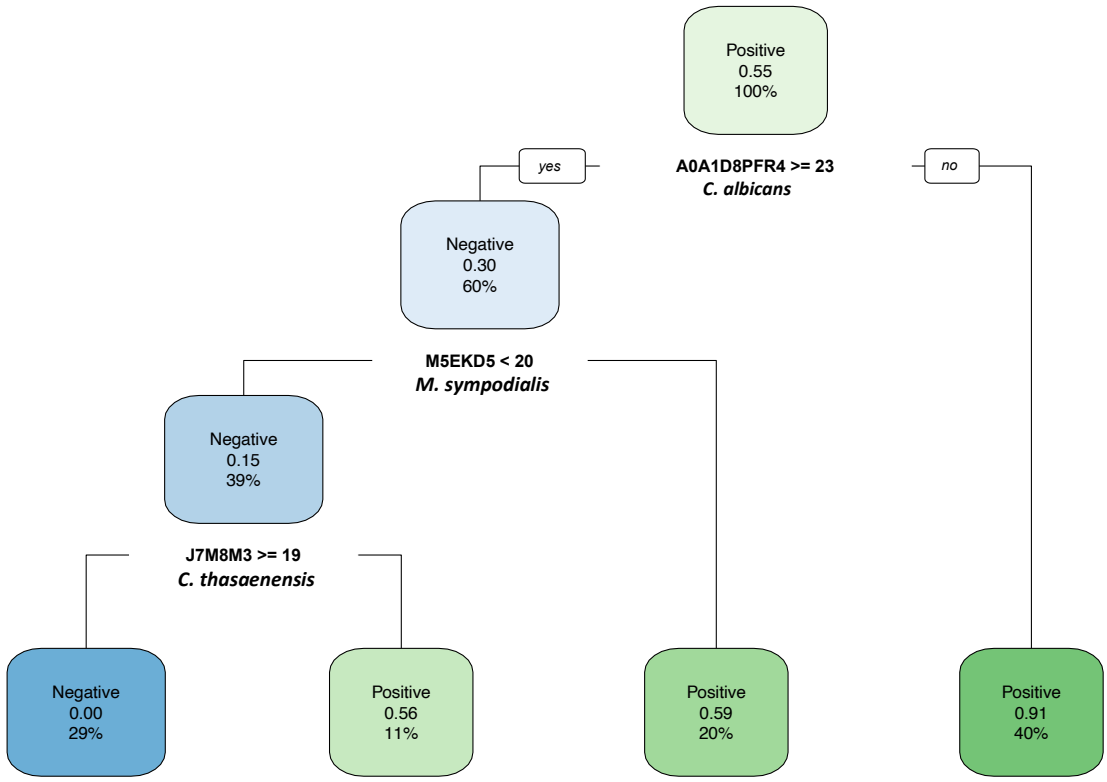


Figure 4.10. Decision tree using differentially expressed fungal proteins between the BV positive and the BV negative state to predict bacterial vaginosis (BV) status. *Candida albicans* (A0A1D8PFR4) was selected as the root feature of the decision tree. Each fungal species on the tree has an iBAQ intensity value unit for cutoff. Each node provides the percentage of the satisfied condition at the node and provides the probability of it not being the BV status of the node.

A Confusion matrix compares the actual target values with those predicted by the machine learning model. The classifier tool made a total of 20 predictions using the test dataset, out of those cases the classifier predicted a negative BV status 7 times and a positive BV status 13 times. However, in reality, 9 samples are BV negative and 11 are BV positive. Thus, we have 2 false BV positives (10%), and our model has a 10% misclassification rate. The two false positives were identified to belong to the individuals 129v2 and 34v2. According to the classification decision tree, these false positives were a result of individual 129v2 having an A0A1D8PFR4 iBAQ value $\leq$ 23 (22.8) and individual 34v2 having an M5EKD5 iBAQ value < 20 (19.0).

Clinical Variables as Predictors of BV Status

The training classification model like the above incorporated 80% of the dataset, which included 76 samples, 16 predictors (clinical variables), and 2 classes (BV positive and BV negative). Nugent score was removed as a clinical variable as it was used to assign BV status. The accuracy of the training dataset model was 98.57%, with a sensitivity and specificity of a 100%, and a positive and negative prediction rate of 100%. The test dataset included the remaining 20% of the dataset and had an accuracy of 70.83% and a precision of 72.72%.

From the training set, the clinical factors identified to contribute to BV included the presence of clue cells, vaginal pH, chemokines, pro-inflammatory cytokines, BMI, HC (Implanon, Nur isterate), *C. trachomatis*, and *T. vaginalis* (Fig 4.11).

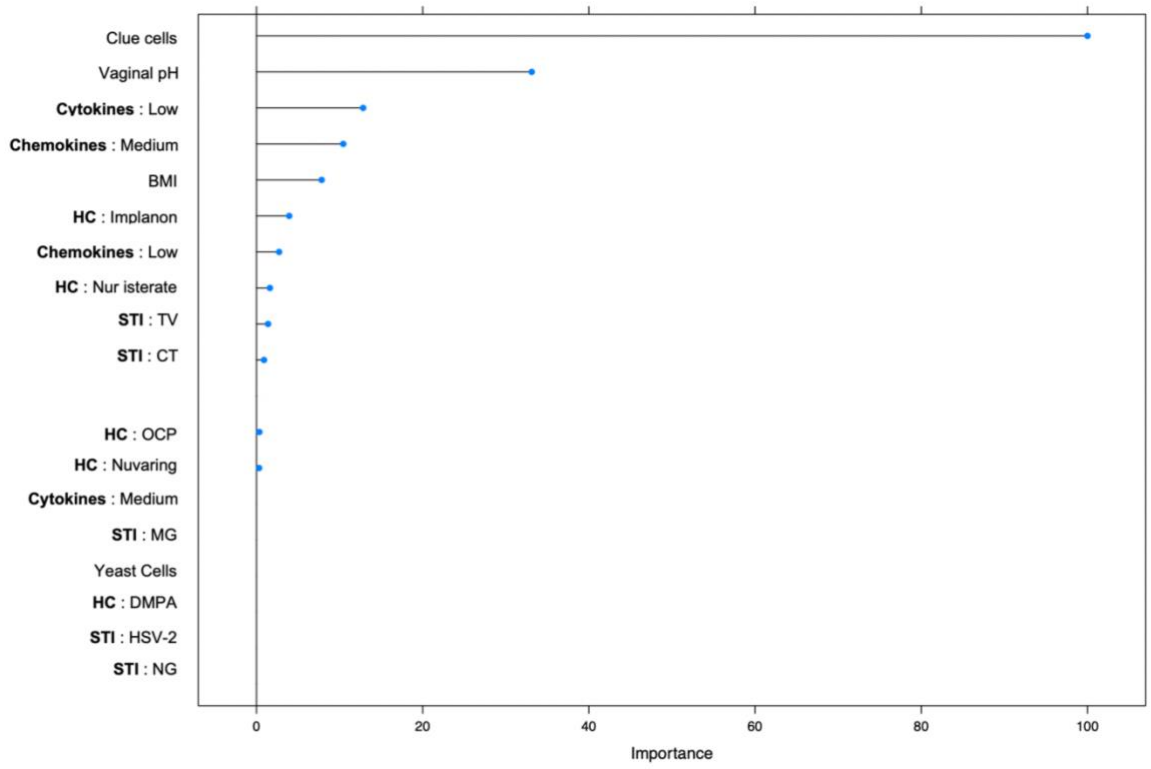


Figure 4.11. Clinical variables identified to contribute to bacterial vaginosis (BV) status and their importance. Clue cells was the important clinical factor contributing to BV status.

In order to investigate why there was no clear clustering of individuals according to BV status in chapter 3 for the hierarchical clustering plot and NMDS plot, it was important to check that individuals had the correct BV status assignment. To investigate this, we used a decision tree,

in which the dataset was removed for Nugent score, and clue cells, to determine whether Nugent score was an accurate classifier of BV status (Figure 4.12).

A decision tree classifier was used to determine why certain individuals were falsely assigned (Fig 4.13). Our decision tree elected vaginal pH as the root feature (vaginal pH < 4.6). From the literature, a vaginal pH of ≤ 4.5 is considered optimal, while a pH > 4.5 is considered to contribute to BV, and is used to classify BV in a clinical setting. Furthermore, the presence of clue cells is also used to classify BV according to Amsel's criteria in a clinical setting. According to these parameters, there were 3 false-positives and 5 false-negatives predicted (Fig 4.12; Table 4.26; Table 4.27). Of these, 2 false-negatives had a pH > 4.5. Three false-positives had a pH < 4.5 (4.1), of which one had an absence of clue cells. Of the false results, three individuals could not be identified for misleading clinical variables according to the classification decision, and were most likely purely a false result, due to the test accuracy of 70.8% (Fig 4.13).

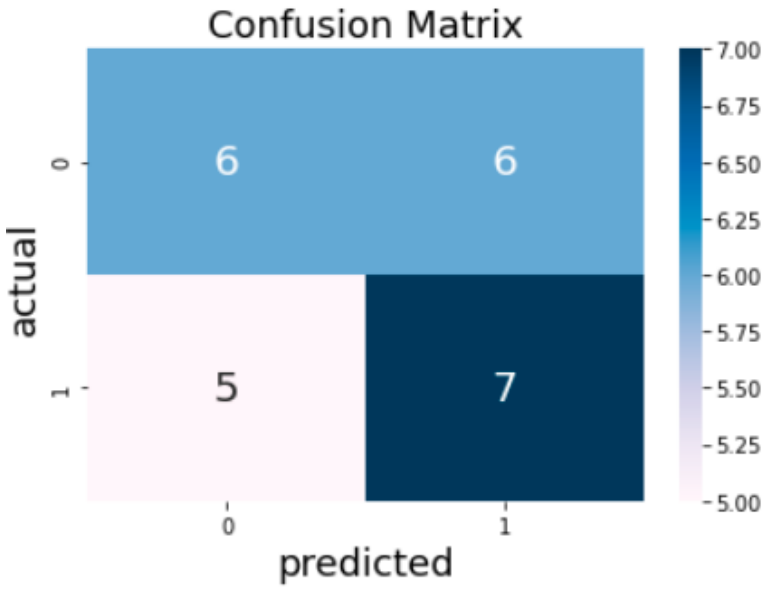


Figure 4.12. Confusion matrix used to predict BV status using clinical variables. There were 5 false-positives and 6 false-negative predicted.

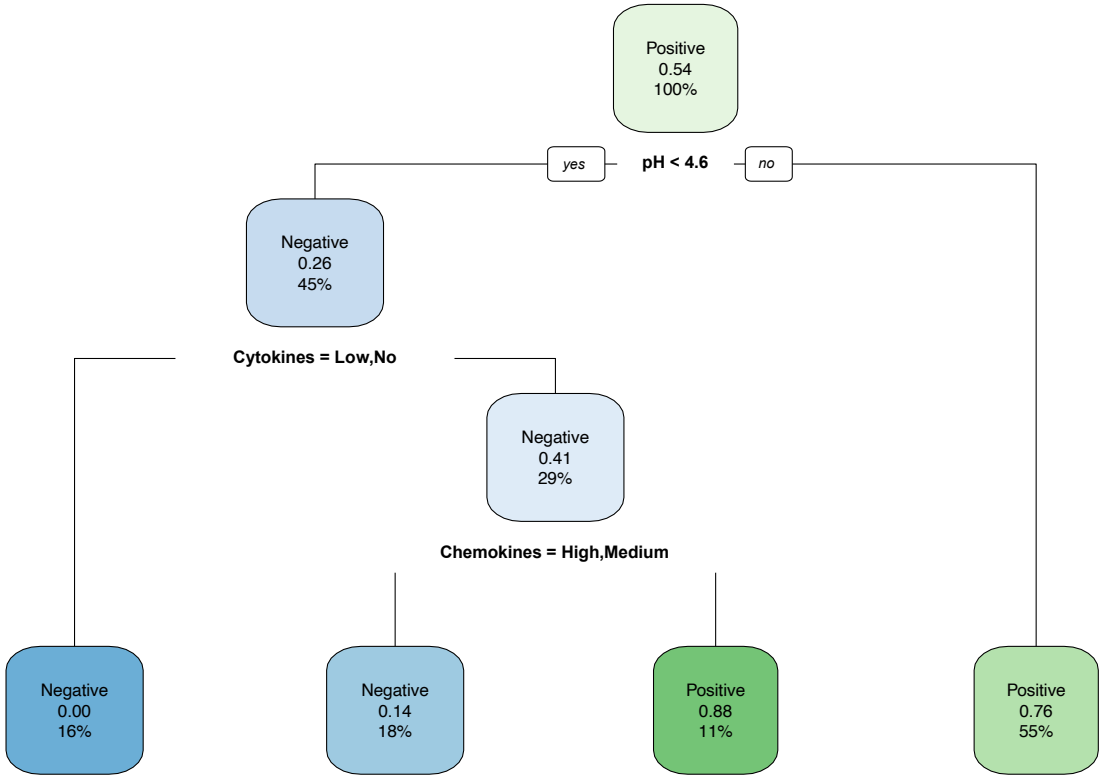


Figure 4.13. Decision tree using clinical variables predict BV status. Vaginal pH was selected as the root feature of the decision tree. Each clinical variable on the tree has a value for cutoff. Each node provides the percentage of the satisfied condition at the node and provides the probability of it not being the BV status of the node.

Table 4.26. Individuals that were identified as false positives using a decision tree (characteristics highlighted in yellow are purely false according to the decision tree (Fig 45), and green identifies individuals neither wrong by clinical criteria nor decision tree).

Sample ID	STI	Nugent Score	Cytokines	Chemokine	HC	pH	BMI	Clue cell	Gram stain
72v2	None	8	Medium	Low	Nur isterate	4.4	26.13	Clue	No
132V2	CT	9	NA	Low	Nur isterate	4.1	21.88	No	Yeast
98v2	No	9	Medium	Medium	Nur isterate	5.3	19.96	Clue	No
43v2	MG	8	Medium	Low	OCP	4.4	25.3	Clue	No
2v2	None	8	Medium	Low	Nur isterate	4.7	19.83	Clue	No

Table 4.27. Individuals that were identified as false-negative using a decision tree (characteristics highlighted in yellow are purely false according to the decision tree, and green identifies individuals neither wrong by clinical criteria nor decision tree).

Sample ID	STI	Nugent Score	Cytokines	Chemokine	HC	pH	BMI	Clue cell	Gram stain
6v2	No	1	High	High	Nur isterate	4.7	22.77	No	Yeast
125v2	No	0	Low	Medium	Nur isterate	4.1	18.13	No	No
126V2	No	3	High	High	Implanon	5.0	25.224	No	No
100V2	CT	0	Low	Medium	Nur isterate	4.4	17.80	No	No
31v2	CT	0	Medium	Medium	Nur isterate	4.1	19.72	No	No
59v2	No	2	Medium	Medium	Nur isterate	5	21.40	No	No

After further investigation, in total, 10 individuals in this dataset were BV positive as determined by the Nugent criteria but showed a pH <4.5, and 13 individuals had a pH >4.5 who were classified as BV negative.

Clinical Variables as Predictors of Inflammation Status

To investigate clinical variables discriminating inflammation categories, we once again used an ENSEMBLE decision tree. Proinflammatory cytokine level was not considered as a clinical variable to discriminate between inflammation states as it was used to diagnose samples with an inflammation category.

The training classification model incorporated 80% of the dataset, which included 83 samples, 17 predictors (clinical variables), and 3 classes (low, medium, and high). We optimized the fit of the classification tree and the accuracy was found to be 65.41%.

We evaluated the performance of the classification model and the accuracy of the training dataset was found to be a 100% with a sensitivity and specificity of a 100%, and a positive and negative prediction rate of a 100%. The test dataset included the remaining 20% of the dataset and had an accuracy of 72.22% (p-value = 0.048). The multi-class AUC was 89.91%.

Table 4.28. Statistics generated when predicting inflammation class as determined by an ENSEMBLE bagging tree using the test dataset.

Class	Low Inflammation	Medium Inflammation	High Inflammation
Sensitivity	0.5000	0.7778	0.8000
Specificity	0.8571	0.7778	0.9231
Positive Pred Value	0.5000	0.7778	0.8000
Negative Pred Value	0.8571	0.7778	0.9231

From the training set, the clinical factors identified to contribute to inflammation included Nugent score, BMI, vaginal pH, chemokines, the presence of clue cells, HC (Nur isterate, DMPA, and Implanon), and STIs (*C. trachomatis*, *T. vaginalis, M. genitalium, N. gonorrohoea,* and HSV-2) (Fig 4.14).

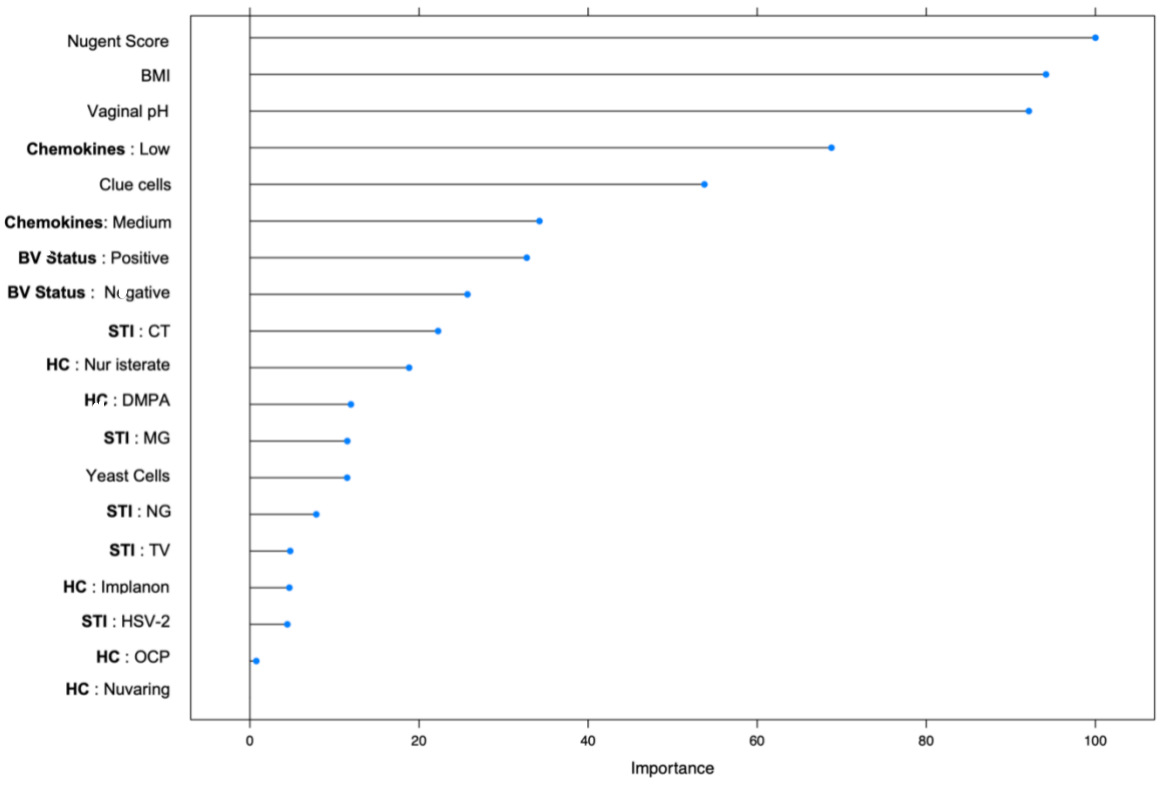


Figure 4.14. Clinical variables identified to contribute to vaginal inflammation and their importance. Nugent score was demonstrated to be the most important factor discriminating between inflammation states.

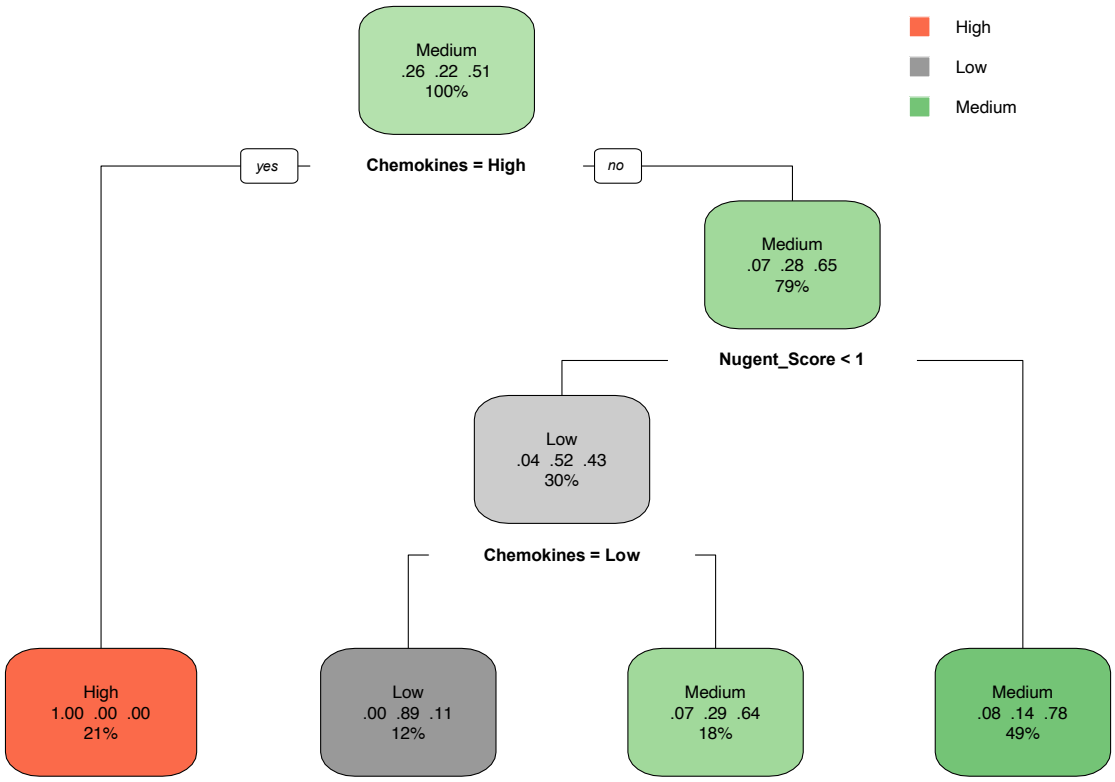


Figure 4.15. Decision tree using chemokines and clinical variables to predict inflammation category. High chemokine was selected as the root feature of the decision tree. Each clinical variable has an association condition for cutoff. Each node provides the percentage of the satisfied condition at the node and provides the probability of it not being the inflammation category of the node.

Discussion

Our fungal data split into two clusters; clusters differed on BV status, as one cluster was majority BV positive and the other BV negative. In the cluster prevalent for BV positive individuals, *Condiobolus* and *Teratoramularia* were highly relatively abundant, as previously seen in chapter 3.

Effect of clinical variables

Our study showed that Nugent score, pro-inflammatory cytokines, chemokines, vaginal pH, *C. trachomatis*, and presence of clue cells are associated with driving differences in fungal community composition. These results suggest that the vaginal mycobiome is either affected

by BV or BV is affected by the mycobiome. This suggestion is supported by the clustering of a large proportion of fungal proteins into BV status groups. However, the degree to which pro-inflammatory cytokines account for the variability in fungal community structure is low (R^2 = 0.05) and not enough to infer that inflammation affects the mycobiome or vice versa.

From the above, our results also suggest that vaginal pH influences mycobiome composition. Correlation analyses showed that *Candida* spp. were negatively associated with vaginal pH, and fungi identified to be overabundant in the BV positive state from chapter 3 (*i.e., Malassezia*, *Alternaria*, and *Condiobolus*) were positively correlated with vaginal pH. Thus, when vaginal pH is low (*i.e.,* optimal), *Candida* is more abundant, and when vaginal pH is high (*e.g.,* during the BV positive state), BV-overabundant genera are more prevalent. For that reason, further examination of the relationship between the vaginal pH and mycobiome is imperative.

C. trachomatis was demonstrated to influence fungal communities, and our correlation analysis revealed that the association between this STI and total fungi was negative. A study by Kruppa *et al.* (2019) studied the interaction between *C. albicans* and *C. trachomatis* and found that *C. trachomatis* binds to *C. albicans* in both its yeast and hyphal form, which inhibits chlamydial infectivity. This finding may explain the negative correlation observed between *C. albicans* and *C. trachomatis* (although the correlation was not significant). Likewise, we also found that *Candida* spp. were negatively correlated with *Neisseria gonorrhoeae* (Ridgway and Oriel 1997). The literature suggests that *C. albicans* inhibits the growth of *N. gonorrhoeae* via an unknown mechanism (Kaye and Levison 1977).

The lack of associations between HC and fungal community structure can be explained by longitudinal studies which evaluated female's pre-and post-initiation of HC's and found no increase in yeast colonization in the vagina (Eschenbach *et al.* 2000; Miller *et al.* 2000; Gupta *et al.* 2000). However, we did find a few significant associations between HC methods and fungal species. Of which, Nur isterate and Implanon showed a positive correlation with *Rhizoctonia solani*. Interestingly, DMPA does not significantly correlate with any fungal species. However, the correlations observed between some HC methods (namely OCP, Implanon, and Nuvaring) and fungal species may be due to the small sample size. Likewise, there was a lack of association between BMI and fungal community structure, and the only significant association was observed with *W. ichthyophaga*.

From our machine learning analyses, we question whether Gram-staining is an effective method to diagnose BV, due to its error rate (Rand and Tillan 2006). This uncertainty is prompted by false assignments of BV status and clustering that is not fully discrete between BV states. In most cases, vaginal pH seems to confirm our false BV state assignments since a vaginal pH > 4.5 has been characteristically used as one of the four Amsel criteria to diagnose BV in a clinical setting (Amsel 1983). However, factors that affect BV but did not demonstrate to affect the mycobiome, such as BMI and HC methods, could also explain the differences in clustering. Likewise, the clinical factors BMI, HC methods, and STIs (*T. vaginalis*, *M. genitalium*, *N. gonorrhea*, and HSV-2) all affected vaginal inflammation but not fungal communities. Since the aforementioned clinical variables drive vaginal inflammation, it may explain why fungal species did not cluster according to inflammation categories. The lack of clustering could also result from BV (according to Nugent score) being a major driver of inflammation.

We can conclude that several clinical factors influence fungal communities. Since the microbiome is a complex environment, many factors are able to drive change and dysbiosis. In the future, it might be beneficial to examine the contribution of other environmental factors, such as hygiene and geography, to the vaginal mycobiome.

Microbial correlations

This mycobiome study considered the full context in which disease takes place. We observed positive correlations between *Candida* species, which may be a result of non-albicans species being able to bind to *C. albicans* hyphae and hitchhiking (Tati *et al.* 2016). Conversely, we observed negative correlations between *C. albicans* and fungal species overabundant in the BV positive state (*Malassezia sympodialis*, *Condiobolus coronatus*, and *Alternaria alternata*). One potential hypothesis explaining this correlation may be through a quorum-sensing molecule known as farnesol. *C. albicans* secretes farnesol, which is able to control the morphology of other fungi, inhibiting hyphal growth, and early biofilm formation (Lorek *et al.* 2008; Semighini *et al.* 2008). By controlling morphology, this molecule can control the pathogenesis of fungi by inhibiting their transition to hyphae (Lorek *et al.* 2008; Semighini *et al.* 2008). Hyphal transition is required for biofilm formation (Lynch and Robertson 2008; Noverr and Huffnaggle 2004). Thus, *C. albicans* may use farnesol in antagonistic relationships with

other fungi. However, more work needs to be done to understand the interactions that occur between fungi in the VMB.

C. albicans is able to undergo a dimorphic transition from a yeast to a pathogenic hyphal form (Cassone 2015; Harriot *et al.* 2010; Majumdar *et al.* 2016; Noble *et al*. 2017). As a result, *C. albicans* may not have shown a significant positive correlation with yeast cells if this species was present in the vagina in its hyphal form, which is possible when patients have VVC.

To date, the dynamics between fungi and bacteria inhabiting the human body have received little attention, and there have been a limited number of studies that have analysed bacteria and fungi from the same sample. Of all interactions described between fungi and bacteria in the vagina, *Candida* in combination with different bacterial strains is best described. Our results demonstrate that *Candida* spp. are positively correlated with *Lactobacillus* spp. However, there has been conflicting evidence describing the relationship between *Lactobacillus* and *C. albicans* colonization. Some studies suggest that *Lactobacillus* spp. compete with *C. albicans* for adhesion sites and secrete substances that inhibit fungal attachment to control *C. albicans* invasion and disease (Jang *et al.* 2009; Li *et al.* 2019; Parolin *et al.* 2015). One such substance *Lactobacillus* secretes is lactic acid, which helps maintain the acidic environment of a healthy vagina and generally does not permit the growth of many potential pathogens (Donati *et al.* 2010; O'Hanlon *et al*. 2013). However, as discussed earlier, *Candida* is highly tolerable of changes in vaginal pH. As a result, other studies have shown that *C. albicans* co-colonizes the vagina with *Lactobacillus* (Beigi *et al.* 2004; Cotch *et al.* 1998; Förster *et al.* 2016; McClelland *et al.* 2009). Several studies have even shown that vaginal colonization with *Candida* spp. is more common in females with a *Lactobacilli*-dominated VMB than in females with dysbiosis (Biagi *et al.* 2009; Drell *et al.* 2013; Zhou *et al.* 2009).

Research has suggested an apparent inverse association between BV and the presence of vaginal yeast, implying that yeast colonization occurs more frequently in an optimal VMB (Biagi *et al*. 2009; Brown *et al*. 2019; Eastment *et al.* 2021; van de Wijgert *et al.* 2014). In confirmation, this chapter showed that yeast species (*e.g., Candida* spp.) were negatively correlated with BV-associated bacteria and Nugent score. Therefore, as BV-associated bacteria become more prevalent, the colonization of yeast species likely decreases. Supporting evidence from Hong *et al.* (2016) showed that their samples that detected *C. albicans* had a low Nugent score

(<4), and Moodley *et al.* (2002) found that the presence of yeasts on microscopy was inversely related to the level of ecologic disturbance. Consequently, Moodley *et al.* (2002) suggest that the BV environment is not conducive to yeast multiplication. Wei *et al.* (2012) found that were a greater percentage of BV cases that coexisted with *C. albicans* pseudohyphae (35.4%) compared to blastospores (14.2%). Their results suggested that pseudohyphae may lead to an obvious disturbance of bacterial flora. However, blastospores and pseudohyphae were found to be most prevalent in BV negative women.

From the above, we hypothesize that yeast interactions with BV-associated bacteria are likely antagonistic. One explanation to support this claim includes BV-associated bacteria harbouring several potential mechanisms that are capable of inhibiting vaginal yeast growth and lowering the risk of detecting yeast. For example, *Mageeibacillus indolicus* produces indole, a compound that inhibits *Candida in vitro*, and *Megasphaera* spp. and *M. indolicus* out-compete *Candida* for essential nutrients, which reduces yeast colonization (Austin *et al.* 2015; Basson 2000; Rajput and Karuppayil 2013). Another possible hypothesis explaining this negative association includes BV-associated bacteria producing acetic acid. *C. albicans* appears to be more sensitive to acetic acid than lactic acid (Chaudry *et al.* 2004; Lourenço *et al.* 2018).

From chapter 3 more fungal species were significantly underabundant in the BV positive when compared to the BV negative state. Accordingly, fungi underabundant in the BV positive state were shown to be negatively correlated with BV-associated bacteria and Nugent score. Whereas the two fungal species (*M. sympodialis* and *C. coronatus*) that were overabundant during the BV positive state were found to be positively correlated with BV-associated bacteria and likely have a synergistic effect with BV-associated bacteria (Krüger *et al.* 2019). Furthermore, our BN showed a likelihood that *L. iners* and *M. sympodialis* affect each other. Affirmingly, Godoy-Vitorino *et al.* (2018) showed that *Lactobacillus* and *Malassezia* have a low co-occurrence. We know *Lactobacillus* is one of the primary bacteria colonising an optimal VMB and assists in keeping the vaginal environment acidic, which helps prevent pathogens from thriving (D'ippolito *et al.* 2018; Vásquez *et al.* 2002; Witkin *et al.* 2007). Therefore, the negative correlation between these two species could either be from the reduction in *Lactobacillus* that gives *Malassezia* an opportunity to colonize the VMB; or specific fungal inhabitants could create an environment that inhibits *Lactobacilli* growth.

Overall, Nugent score is negatively associated with total fungal intensity. An inverse correlation between fungal diversity and BV is likely to occur because the BV-associated microenvironment becomes less suitable for fungal growth as bacterial diversity increases, and fungi likely compete for resources (Underhill and Iliev 2014).

Correlations between fungal and bacterial species can be explained by biophysical and metabolic interactions between species. These interactions include quorum sensing (already discussed), biofilm formation, production of secondary metabolites, and cellular signal transduction (O'Toole 2016; Polke and Jacobsen 2017; Papenfort and Bassler 2016; Rio 2017; Whiteley *et al.* 2017;). Fungi frequently flourish as biofilms and from the fungi identified in this study, several have been shown to form biofilms, including *Candida*, *Aspergillus*, *Cryptococcus*, *Fusarium*, *Malassezia*, *Trichosporon*, and *Rhizopus* (Beauvais *et al.* 2007; Cannizzo *et al.* 2007; Di Bonaventura *et al.* 2006; Imamura *et al.* 2008; Martinez *et al.* 2005; Nobile and Johnson 2015). One such biophysical interaction allows bacteria to attach to fungal hyphae and redistribute themselves within discrete layers of biofilms (Hogan *et al.* 2009).

Bacterial-fungal interactions can result in a diverse range of interactions, from antagonism, mutualism, synergistic, to neutral (Nogueira *et al.* 2019). Hence, bacteria and fungi can mutually support one another's growth or exert competitive effects, which may lead to the suppression of one microorganism and the growth of another. These interactions can influence the biology and ecology of species involved with regard to growth, reproduction, transport, competition for nutrients or adhesion sites, stress resistance, and pathogenicity (formation of mixed-species biofilms) (Bergeron *et al*. 2017; Neely *et al*. 1986; Nogueira *et al*. 2019). However, co-occurrence between species may also not be representative of any causal relationship, but rather a result of random 'mixing' within the microbiome.

In the past, bacterial-fungal interactions focused merely on the interaction between two species. However, recent evidence shows that interactions may exist as complex networks of multiple interacting organisms. Our BN analysis showed that for example, *C. albicans* associated with *R. vinicolor*, but *R. vinicolor* also associates with *Kluyveromces marxianus.* Future work should prioritize understanding the complexity of these networks and incorporating environmental and clinical variables.

The many positive and negative correlations that were observed between identified fungi and members of the bacterial microbiome suggest that interactions between the mycobiome and the bacteriome likely play a role in optimal and non-optimal states. Since we have established from the literature that fungal and bacterial communities undoubtedly have the potential to influence each other. Emerging studies also suggest that changes in commensal fungi may be relevant in diseases that are not primarily fungal (Underhill and Iliev 2014). By understanding the mechanisms of adhesion and signaling involved in fungal–bacterial interactions we can try to prevent bacterial and fungal infections. This may lead to the development of novel therapeutic strategies to combat microbial colonization and polymicrobial diseases. Antimicrobial therapies that target specific pathogens, in contrast to broad-spectrum antibiotics, may help prevent secondary problems that arise upon perturbation of beneficial bacterial-fungal interactions (Morales and Hogan 2010). However, more longitudinal studies are needed to describe the association between bacterial and fungal communities (Rechenberger *et al*. 2019).

Disease-associated fungi

An advantage of using mass spectrometry is the identification of disease-associated species, which are molecules that are consistently modified or present at abnormal concentrations in specific diseases (Diamandis 2004). It is also important to note that current diagnostic approaches are not designed to assess polymicrobial infections and will possibly miss the presence of bacterial-fungal co-infections (Nogueira *et al.* 2019).

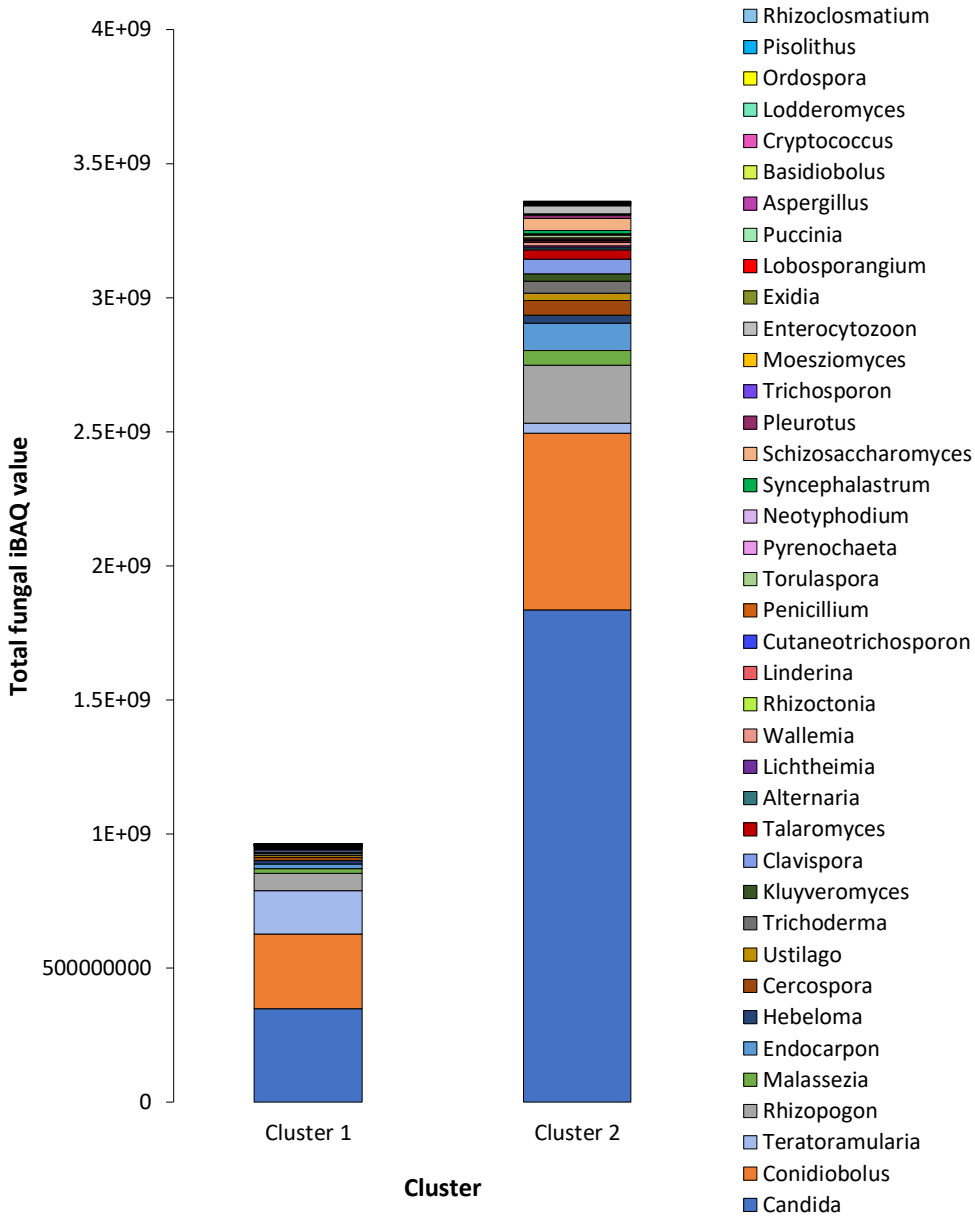
The most abundant fungal species found in this study that were associated with BV were *C. albicans*, *M. sympodialis*, and *C. thasaenensis*. From chapter 3, we know that *Candida* and *Malassezia* are frequently isolated from the vagina. A study by Hong *et al.* (2016) confirms this association, as *Malassezia* was associated with a high Nugent score (=8) and *C. albicans* associated with a low Nugent score (<4) in their study. Furthermore, Godoy-Vitorino *et al.* (2018) showed that *Candida* and *Malassezia* were associated with Human papillomavirus (HPV) infection. From our results, our BN showed a strong association between *Malassezia* and *T. vaginalis,* and Spearman's correlation showed positive correlations between *Malassezia* and all STIs. From these observations, it appears that *Malassezia* is associated with the non-optimal state. In contrast, a negative correlation was noted between *T. vaginalis* and *C. albicans* (Ridgway and Oriel 1977). An optimal VMB is proposed to inhibit infection with

T. vaginalis, as the prevalence of this STI is low when the Nugent score is low (Moodley *et al.* 2002). This association may explain the negative correlation between *T. vaginalis* and *C. albicans*, as the relative abundance of C. albicans is highest in the BV negative state.

Conclusion

We illustrated that several clinical factors were major drivers explaining the distinct community structure. While the directionality of the relationship between the mycobiome and clinical factors cannot be determined based on the results of this study.

There is a growing awareness that many human diseases are not driven by a single organism but are the product of multiple microbial populations. Machine learning (ML) methods offer the ability to incorporate the structure of the microbial communities as a whole and identify associations between community structure and disease state (Topçuoğlu *et al.* 2020). If optimal and non-optimal fungal communities are reliably classified, then ML methods offer the ability to identify microbial species within communities that are responsible for disease classification (Topçuoğlu *et al.* 2020). In our work, by characterizing fungal species differentially expressed between BV states, we could identify fungal species most closely associated with BV. Although on a small scale, we illustrate the association between the mycobiome and the bacteriome in the FGT during BV. Fungi appear to be inversely correlated to BV. Thus, fungal communities likely affect and are affected by bacterial communities. Additionally, with the identification of disease-associated fungi, we set the groundwork for determining the influence of these species in disease and understanding of vaginal community dynamics, which may lead to interventions aimed at shifting to (and maintaining) more protective microbial communities. In the future, to further understand the role of the vaginal mycobiome in vaginal diseases, the development of novel *in vitro* models of polymicrobial communities may prove useful (Bradford and Ravel 2017).


4E+09
3.5E+09
3E+09
2.5E+09
2E+09
1.5E+09
1E+09
500000000
0
Total fungal iBAQ value
Cluster 1
Cluster 2
Cluster
Rhizoclosmatium
Pisolithus
Ordospora
Lodderomyces
Cryptococcus
Basidiobolus
Aspergillus
Puccinia
Lobosporangium
Exidia
Enterocytozoon
Moesziomyces
Trichosporon
Pleurotus
Schizosaccharomyces
Syncephalastrum
Neotyphodium
Pyrenochaeta
Torulaspora
Penicillium
Cutaneotrichosporon
Linderina
Rhizoctonia
Wallemia
Lichtheimia
Alternaria
Talaromyces
Clavispora
Kluyveromyces
Trichoderma
Ustilago
Cercospora
Hebeloma
Endocarpon
Malassezia
Rhizopogon
Teratoramularia
Conidiobolus
Candida

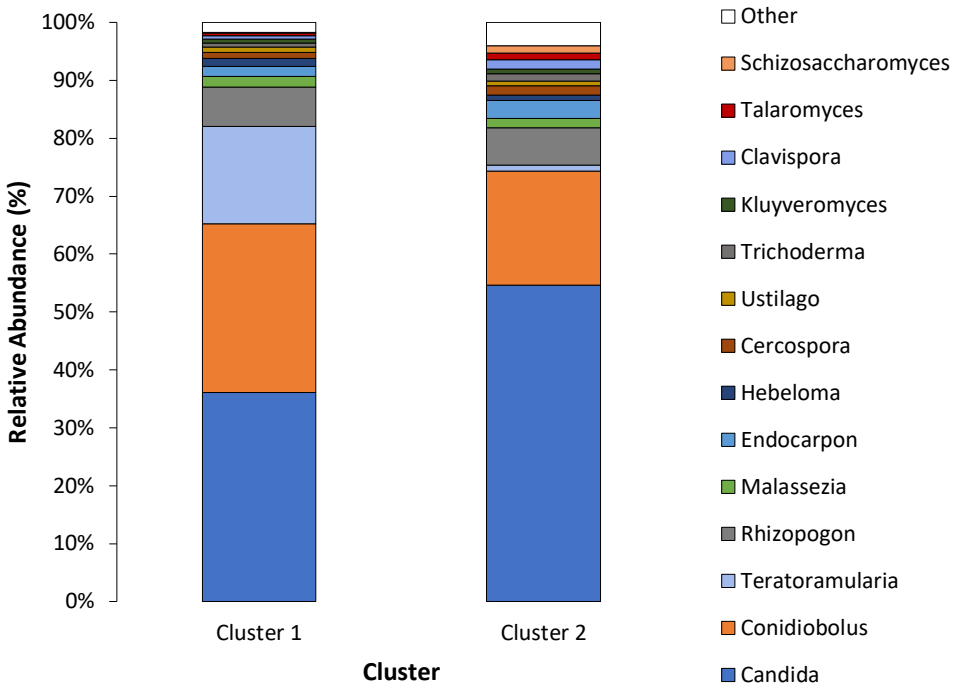


Figure S1. Relative abundance of fungal genera in each cluster using the k-means clustering method (*k* = 2), cluster 1 had 33 samples, and cluster 2 had 56 samples.

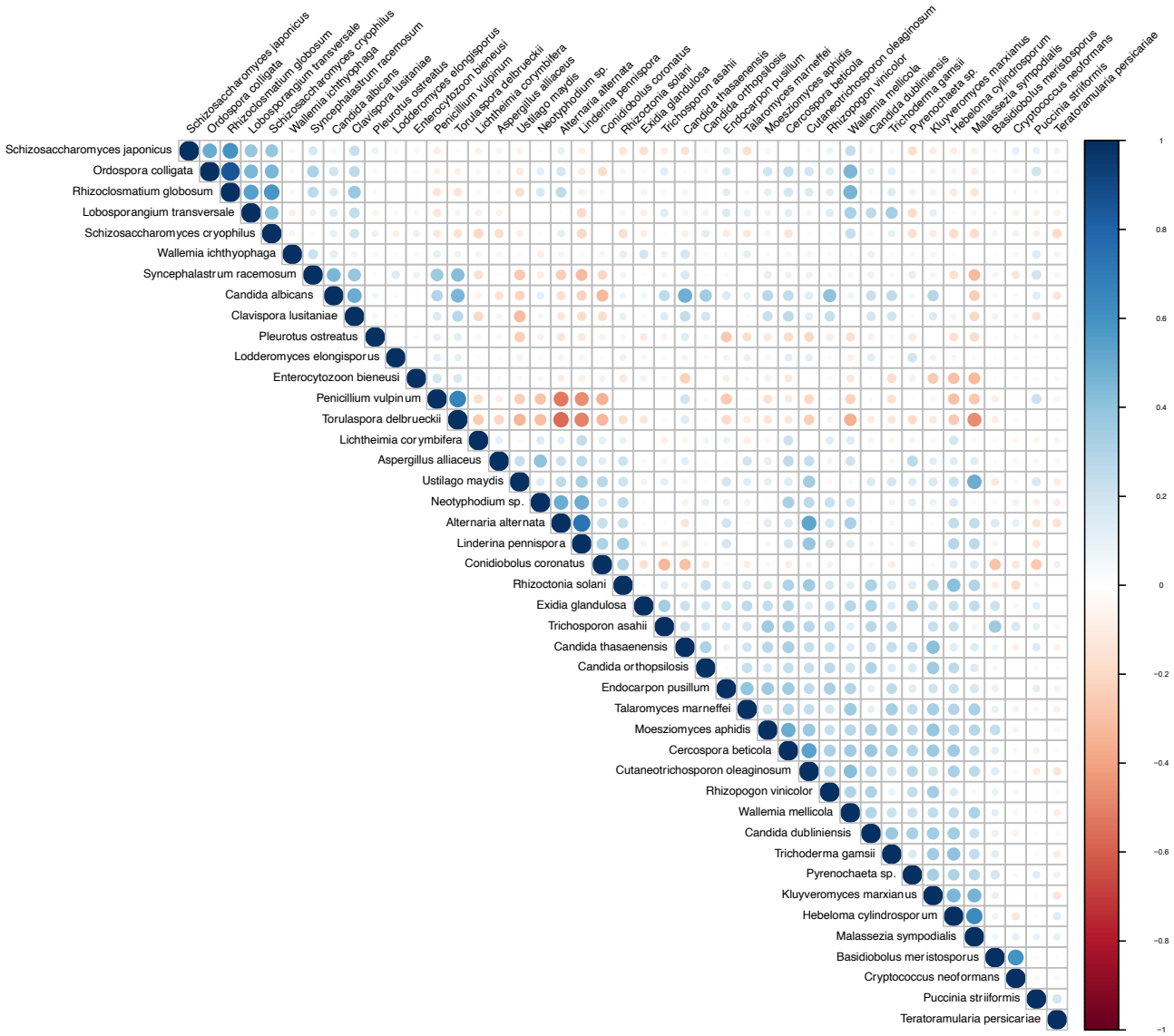


Figure S2. Correlations between fungal species identified in this study in visit 2 (n = 89). Red represents negative correlations (negative r-value), whereas blue represents positive correlations (positive r-value). The larger the circle the stronger the correlation. More positive correlations between fungal species are observed in comparison to negative correlations. The strongest positive correlation occurred between *Ordospora colligata* and *Rhizoclosmatium globosum.* The strongest negative correlation was between *Malassezia sympodialis* and *Torulaspora delbrueckii.*

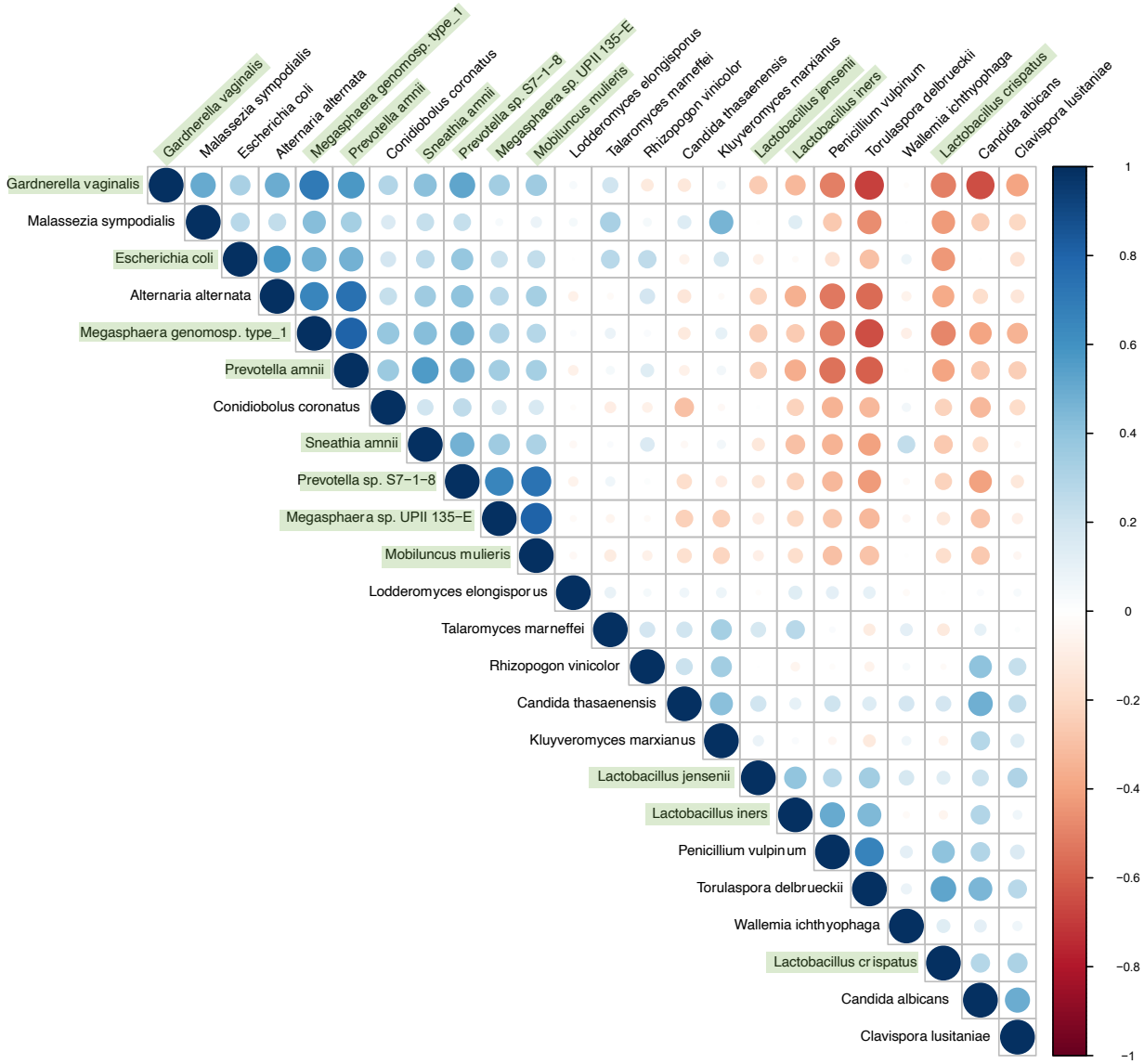


Figure S3. Correlations between differentially expressed fungal species (*n* = 14) and bacterial species associated with BV and Inflammation from visit 2 samples (*n* = 89). Bacterial species are highlighted in green. Red circles represent negative correlations, whereas blue circles represent positive correlations. Darker colours and increased circle size represent stronger correlations. *M. sympodialis* had the strongest positive correlation with the bacterium *G. vaginalis.* The strongest negative correlation was observed between *T. delbrueckii* and *G. vaginalis*.

Table S1. Strength of associations in Bayesian network between fungal species, optimal bacterial species, and non-optimal bacterial species.

from	to	strength	direction
P. amnii	*A. alternata*	0.99	0.84
A. alternata	*P. amnii*	0.99	0.16

P. vulpinum	*T. delbrueckii*	0.95	0.43
T. delbrueckii	*P. vulpinum*	0.95	0.57
L. iners	*T. marneffei*	0.83	0.73
T. marneffei	*L. iners*	0.83	0.26
W. ichthyophaga	*K. marxianus*	0.79	0.92
K. marxianus	*W. ichthyophaga*	0.79	0.076
G. vaginalis	*M. sympodialis*	0.78	0.73
S. amnii	*R. vinicolor*	0.78	0.85
M. sympodialis	*G. vaginalis*	0.78	0.27
R. vinicolor	*S. amnii*	0.78	0.15
L. crispatus	*T. delbrueckii*	0.74	0.34
T. delbrueckii	*L. crispatus*	0.74	0.66
L. crispatus	*C. thasaenensis*	0.72	0.52
C. thasaenensis	*L. crispatus*	0.72	0.48
L. crispatus	*M. sympodialis*	0.68	0.65

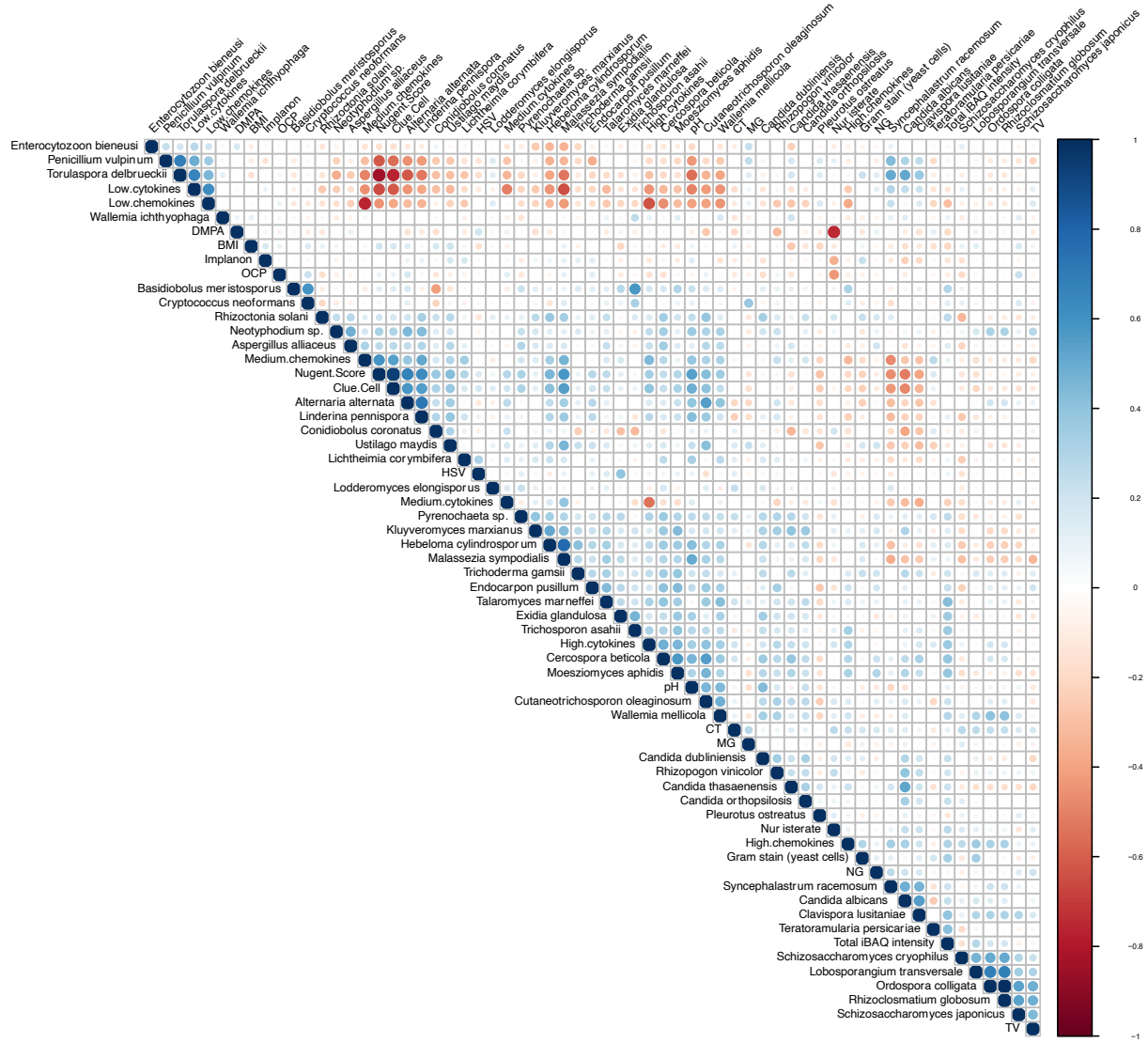


Figure S4. Correlations between fungal species identified (*n* = 45) and clinical variables at visit 2 (*n* = 89). Red circles represent negative correlations, whereas blue circles represent positive correlations. Darker colours and increased circle size represent stronger correlations.

Chapter 5

Fungal Validation with Publicly Available Data

Introduction

The analysis of existing data is a cost-efficient way to make maximum use of data that is already collected to address fundamental new research questions and provide a different assessment of the primary aims of the original study. Therefore, secondary data analysis allows us to test new hypotheses not initially addressed in the original research (Cheng and Philips 2014). Since a vast amount of microbiome data is currently available, but most studies have primarily focused on the bacterial component, the remainder of microbiome data remains unexplored.

The advantage of using metaproteomic approaches is that we are able to generate global hypotheses about the proteome. Validation of results generated using metaproteomics supports global hypotheses by (i) providing evidence for the method that generated significant results and (ii) ensuring that downstream taxonomic and functional analyses that were based on significant features had an initial accurate list of features (Dougherty 2007; Doughtery *et al.* 2007; Subramanian *et al.* 2005). The increasing availability of public data readily available online encourages cross-linking information from different data sources (Cheng and Philips 2014). However, there is still no gold standard for culture-independent mycobiome analyses, as many studies use various techniques and genomic regions to study fungi (Araujo 2014; Iliev and Leonardi 2017; Nilsson *et al.* 2019; Scanlan and Marchesi 2008; Tang *et al.* 2015).

Metaproteomics provides unique information on both the composition and functional status of the microbiome (Hettich *et al.* 2012). In Chapters 2 and 3, we concluded that using the appropriate protein sequence database enables us to taxonomically and functionally analyse fungi present in metaproteomic data. Therefore, metaproteomics does provide value when conducting mycobiome research.

In previous chapters, we highlighted that the fungal organisms that reside in the female genital tract (FGT) are understudied. The lack of FGT mycobiome research necessitates the validation of fungal species that are identified during analyses (Bradford and Ravel 2017). Since several fungal species we identified in Chapter 3 were new and novel, it is important to validate the

presence of likely commensals and pathogens. By using public data, we will be able to remove the biases from specific technology and account for factors such as sample population (*e.g.*, ethnicity and race). Data quality will differ between research studies utilized for validation, so we can expectantly have datasets that have accounted for environmental contaminants. With this, we should be able to confirm the fungal inhabitants of the FGT. Thus, we set out to validate our taxonomic and functional annotation results and methodology using publicly available proteome data. Validation will enable more confidence in the fungal inhabitants we observed in the FGT and provides a complementary approach to DNA-based methods.

Methods

Publicly available datasets used in this chapter to examine the VMB belonged to studies performed by Borgdorff *et al.* (2016), Afiuni-Zadeh *et al.* (2018), and Månberg *et al.* (2019). Datasets were downloaded from ProteomeXchange (http://proteomecentral.proteomexchange.org) using the following project IDs: PXD003176 (Borgdorff *et al.* 2016), PXD009596 (Afiuni-Zadeh *et al.* 2018), and PXD009723 (Månberg *et al.* 2019).

To create databases for each dataset, the Metanovo pipeline was used to filter the complete UniProtKB database (June 2020 release) for protein identifications for each publicly available dataset selected using default parameters. Using the Månberg *et al.* (2019) dataset we performed an additional Metanovo analysis using the mass tolerance parameters specified in the original study (Mass tolerance for precursor ions was 10 ppm and 0.5 Da mass tolerance for fragment ions). With this, we could compare how mass tolerance affected the MS/MS identification rate and the taxonomic and functional annotations obtained.

The output database (.fasta) files from Metanovo for each publicly available dataset were concatenated to the manually curated fungal Pan Proteome database that was created in Chapter 2. Raw files were also processed with the Metanovo + Pan Proteome database created for the WISH dataset in Chapter 2 (now referred to as the **FGT curated database**) to make comparisons between databases. Raw files for each dataset were processed with MaxQuant version 1.6.3.4 on Linux using various databases with the parameters specified in Chapter 2.

To taxonomically and functionally analyse publicly available datasets, we employed the method described in Chapter 3. In summary, quality control methods included removing protein groups from the dataset that had >=2 peptides assigned to it, potential contaminants, reverse hits, and entries deleted from the UniProtKB database. For quality control on peptide sequence identifications, potential contaminants and reverse hits were also removed. For quality control on functional assignments, KO assignments were manually curated to remove categories associated with host-level functions.

Results

Månberg *et al.* (2019) dataset

Taxonomic profile

Månberg *et al.* (2019) obtained vaginal samples using two cotton-tipped swabs that collected secretions from the cervical os and posterior vaginal fornix, from women residing in Nairobi, Kenya. Study participants (*n* = 102) included women in heterosexual HIV-serodiscordant relationships in which their male partner was HIV-infected, but they were uninfected. Eligible participants were older than 18 years of age, not pregnant, and reported sexual intercourse with their study partner at least three times in the three months prior to screening. Cervicovaginal secretions (CVS) samples were analysed with label-free tandem MS/MS using the Linear Trap Quadrupole Orbitrap Velos mass spectrometer (Thermo Fisher) coupled to a nano-flow Easy nLC II (Thermo Fisher).

Table 5.1. Information on the storage and processing of samples collected by Månberg *et al.* (2019).

Buffer solution	Phosphate-buffered saline (PBS)
Transport	Samples transferred on ice to the laboratory within 2 hours of collection.
Centrifugation	Centrifuged at 800 g for 10 min at 4 °C.
Supernatant preservation	Supernatant was aliquoted and cryopreserved in cryovials at -80 °C.
Denaturation and digestion	Protein from each sample was denatured in urea and digested with trypsin enzyme using filter-aided sample preparation.

Peptide cleaning	Peptides were cleaned of salts and detergents with reverse-phase liquid chromatography.

The database generated from Metanovo for the Månberg *et al.* (2019) dataset using default parameters was concatenated to the fungal Pan Proteome database created in Chapter 2, and the database was termed the **HIV curated DB**. The database generated from Metanovo for this dataset using adjusted mass tolerance parameters was termed the **adjusted HIV curated DB**.

The results from Maxquant revealed that between the **FGT curated DB** and the **adjusted HIV curated DB,** 412 protein groups were shared in common (Fig 5.1), of which 20 protein groups were assigned to bacteria, 388 to human, and 4 to fungi. Those fungal species included *Clavispora lusitaniae*, *Kluyveromyces lactis*, *Pichia kudriavzevii*, and *Talaromyces marneffi*. When comparing the **HIV curated DB** to the **FGT curated DB**, 62 protein groups were shared (Fig 5.1), of which all protein groups matched to human taxa. The comparison between the adjusted **HIV curated DB** and the **FGT curated DB** revealed that 82 protein groups were shared (Fig 5.1), of these 73 protein groups were assigned to human proteins. Overall, between all three databases, 52 protein groups overlapped (Fig 5.1), all of which assigned to human proteins.

The FGT curated DB provided a higher MS/MS identification rate, but the **adjusted HIV curated DB** identified a larger number of protein groups. As a result, we chose to perform downstream taxonomic and functional analyses on all three databases to compare results (Table 5.2).

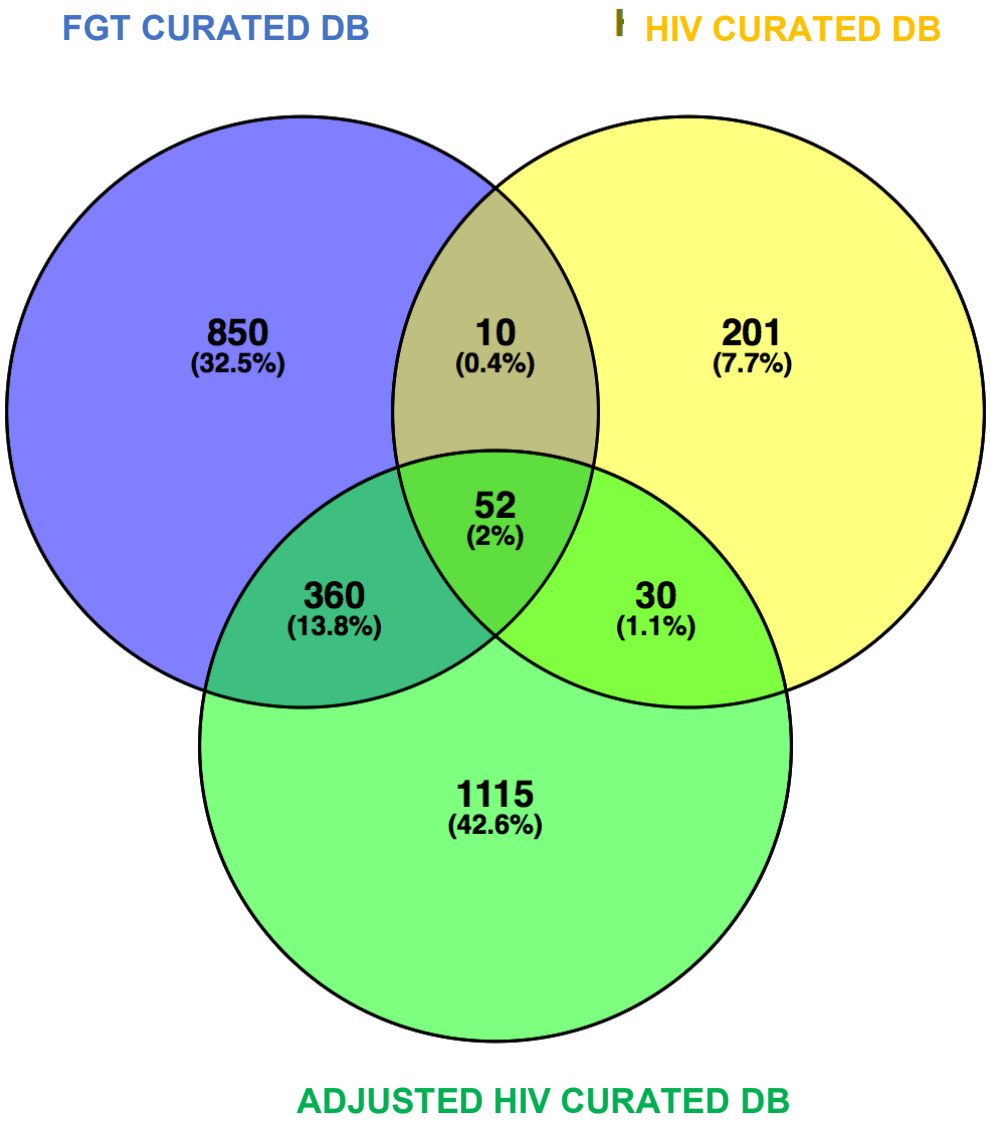


Figure 5.1. Venn diagram depicting the number of protein groups identified using three different protein sequence databases, *viz.* the HIV curated database (DB), the FGT curated DB, and the adjusted HIV curated DB with altered mass tolerance settings. The HIV curated DB identified 293 protein groups, the FGT curated DB identified 1272 protein groups, and the adjusted HIV curated DB identified 1557 protein groups. Between the three protein sequences databases, 52 protein groups were shared.

Table 5.2. Overview of the MS/MS identification rate, protein group. and peptide sequence identifications across each protein sequence database search in MaxQuant for the Manberg *et al.* (2019) dataset.

Database	MS/MS Identification Rate (%)	Protein Group Identifications	Peptide Sequence Identification	Peptides after QC
HIV curated DB	10.14	293	2,252	1,698
Adjusted HIV curated database	15.9	1,557	7,801	7,312
FGT curated DB	16.32	1,272	7,911	6,152

Abbreviations: DB: database; MS/MS: Mass Spectrum; FGT: Female genital tract; QC: quality control

From the protein group identifications, we calculated the relative abundances of taxonomic kingdoms using unnormalized iBAQ protein intensities after implementing the exclusion criterion. The results showed that across each of the three databases, Metazoa had the greatest relative abundance (>98.3%; Table 5.3). Of the three databases, the **FGT curated DB** was the only database to identify bacteria (1.57%) at a greater relative abundance than fungi (0.09%) (Table 5.3). Interestingly, the **adjusted HIV curated DB** identified kingdoms/subkingdoms that were not present in the other two databases, such as Protozoa, Archaea, and Sar (Table 5.3).

Taxonomic assignments for kingdoms using peptide sequences showed that across databases, Metazoa had the highest number of sequences assigned (>69.7%), with the **HIV curated DB** having the largest number of Metazoan assignments (80.9%) (Table 5.4). The **HIV curated DB**, however, had the lowest number of bacterial assignments (0.9%; Table 5.4) in comparison to the other two databases (>9.5%) (Table 5.4). The percentage of sequences assigned to fungi seem to be in the similar range across databases (0.2%-0.7%) (Table 5.4). As seen with protein groups, the **adjusted HIV curated DB** was the only database to detect Archaeal peptide sequences (Table 5.4).

Table 5.3. Relative abundance of taxonomic kingdoms using the summed total iBAQ protein intensity values for each taxonomic kingdom/subkingdom identified from each protein sequence database used on the Manberg *et al.* (2019) dataset.

	Relative abundance (%)		
Kingdom/ Subkingdom	**HIV curated DB**	**Adjusted HIV curated DB**	**FGT curated DB**
Metazoa	99.3	99.5	98.3
Fungi	0.6	0.05	0.09
Bacteria	0.09	0.3	1.57
Viridiplantae	0.0001	0.00009	0
Protozoa	0	0.02	0
Archaea	0	0.05	0
Sar	0	0.06	0

Abbreviations: DB: database; FGT: Female genital tract; HIV: Human immunodeficiency virus

Table 5.4. Abundance of peptide sequences assigned to taxonomic kingdoms identified using Unipept on the results from each protein sequence database for the Månberg *et al.* (2019) dataset.

	Peptide sequences assigned (%)		
Kingdom	**HIV curated DB** (n=17)	**Adjusted HIV curated DB** (n=17)	**FGT curated DB** (n=47)
Metazoa	80.9	69.7	77.8
Fungi	0.7	0.2	0.6
Bacteria	0.9	12.1	9.5
Virdiplantae	0.2	0.2	0
Archaea	0	0.09	0
Root	11.4	9.6	8.2
Unassigned	3.6	4.5	1.7

Abbreviations: DB: database; FGT: Female genital tract; HIV: Human immunodeficiency virus

From the fungal peptide sequence assignments, we calculated the percentage of peptides assigned to the phylum level. Across databases, the majority of peptide sequences were assigned to two fungal phyla, namely Ascomycota and Basidiomycota as seen previously seen in chapter 3 for the WISH dataset. For the **HIV curated DB**, only one peptide sequence was assigned to Basidiomycota, and the remainder of sequences (94.1%) were assigned to Ascomycota. The **adjusted HIV curated DB** identified a third phylum, Chytridiomycota, which had 5.9% of peptide sequences assigned.

The **HIV curated DB** identified the highest number of genera (18) using protein groups, of which *Zygosaccharomyces* (45.3%), *Wallemia* (24.0%), *Candida* (11.2%), *Cyberlindnera* (6.6%), and *Schizosaccharomyces* (3.2%) were the most relatively abundant genera (Fig 5.2). The **adjusted HIV curated DB** identified the lowest number of fungal genera (10) using protein groups. Of these genera, *Kluyveromyces* (34.2%), *Schizosaccharomyces* (26.3%), *Rhizoclosmatium* (7.1%), and *Wallemia* (5.2%) were the most relatively abundant (Fig 5.2). The **FGT curated DB** identified 13 fungal genera, of which *Candida* was the most relatively abundant (69.0%), followed by *Kluyveromyces* (19.3%) (Fig 5.2). Supplementary Table 1 contains the fungal genera that constituted less than 1% of the total relative abundance of fungi in each database.

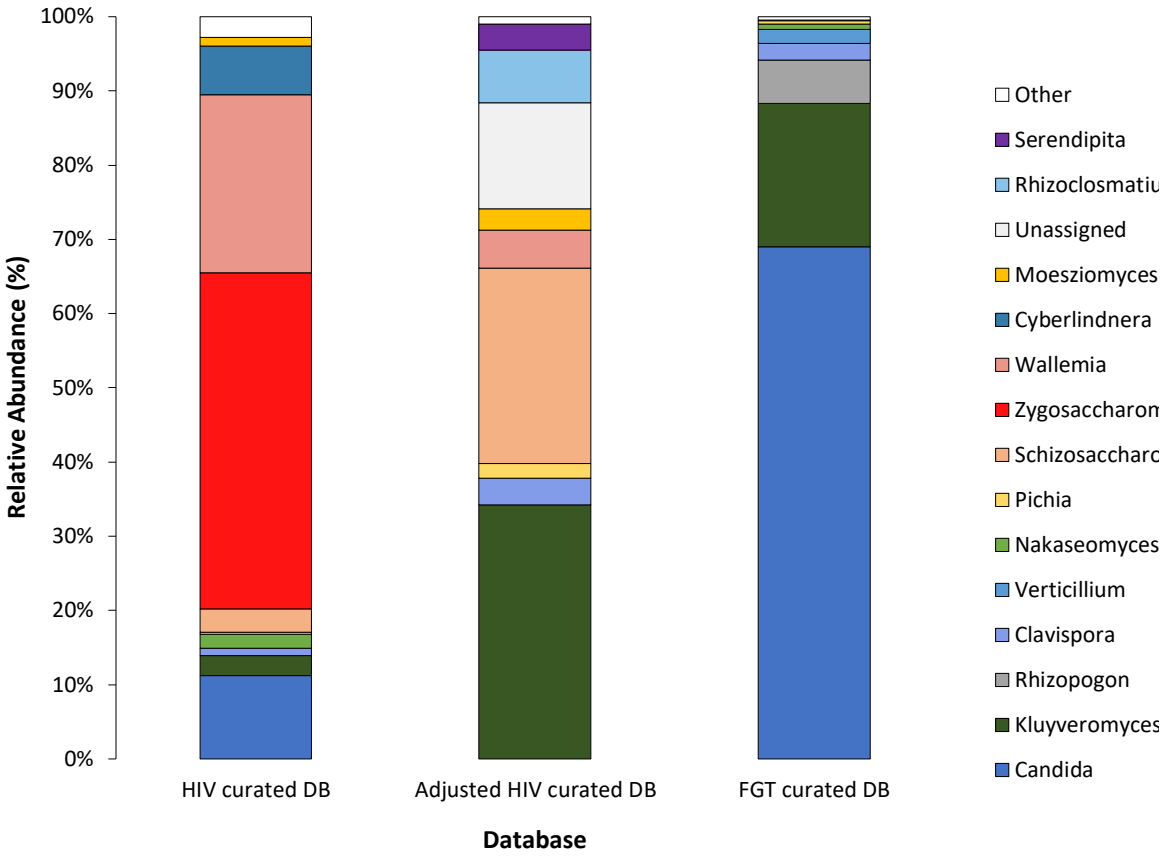


Figure 5.2. Relative abundance of fungal genera using the summed total iBAQ intensity values of the proteins that assigned to each fungal genus *(n* = 18) in Manberg *et al.* (2019) dataset using 3 different protein sequence databases. The FGT curated DB detected the largest relative abundance of *Candida*.

At the genus level, the **FGT curated DB** identified the largest number of distinct genera (*n* = 15) using peptide sequences, whereas the **HIV curated DB** and **adjusted HIV curated DB** identified 13 and 9 fungal genera, respectively (Fig 5.3).

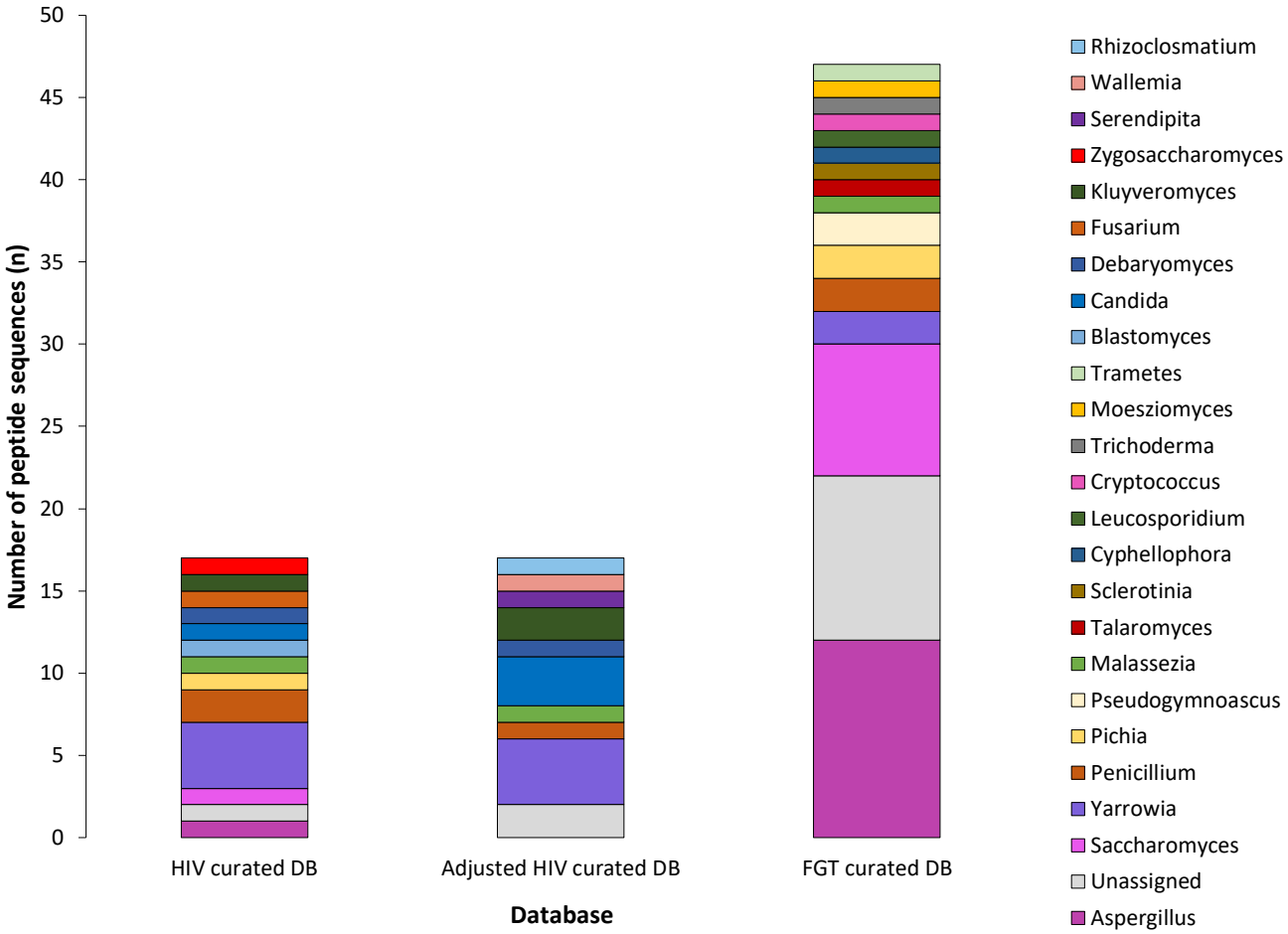


Figure 5.3. Number of peptide sequences assigned to each fungal genus using each protein sequence database using *Unipept*. The FGT curated DB identified the largest number of fungal peptide sequences (*n* = 47).

Fungal species level assignments using protein groups demonstrated that between the **FGT curated DB** and **HIV curated DB**, 6 fungal species are common; between the **adjusted HIV curated DB** and the **FGT curated DB**, 4 fungal species are common; and between the **adjusted HIV curated DB** and **t**he **HIV curated DB** 6 fungal species are common (Table 5.5).

Table 5.5. Fungal species level assignments using the protein groups identified in each protein sequence database. Gaps in table indicate species not detected in database

Adjusted HIV DB	FGT curated DB	HIV curated DB
	Alternaria tenuissima	*Alternaria tenuissima*
	Candida albicans	*Candida albicans*

	Candida glabrata	*Candida glabrata*
		Candida orthopsilosis
	Cercospora beticola	
Clavispora lusitaniae (*Candida lusitaniae*)	Clavispora lusitaniae (*Candida lusitaniae*)	*Clavispora lusitaniae* (*Candida lusitaniae*)
		Cryptococcus neoformans
		Cyberlindnera jadinii (*Pichia jadinii*)
		Fusarium culmorum
		Gibberella zeae (*Fusarium graminearum*)
Kluyveromyces lactis (*Candida sphaerica*)	Kluyveromyces lactis (*Candida sphaerica*)	*Kluyveromyces lactis* (*Candida sphaerica*)
	Lichtheimia corymbifera	
Moesziomyces aphidis		*Moesziomyces aphidis*
Mucoromycotina sp, EMW-2019b		
		Neosartorya fumigata (*Aspergillus fumigatus*)
		Penicillium digitatum
	Penicillium rubens	
Pichia kudriavzevii (*Candida krusei)*	*Pichia kudriavzevii* (*Candida krusei)*	*Pichia kudriavzevii* (*Candida krusei)*
		Pseudozyma antarctica (*Candida antarctica*)
Rhizoclosmatium globosum		
	Rhizopogon vinicolor	
Schizosaccharomyces cryophilus	*Schizosaccharomyces cryophilus*	*Schizosaccharomyces cryophilus*
Serendipita indica		
Talaromyces marneffei (*Penicillium marneffei*)	Talaromyces marneffei (*Penicillium marneffei*)	

		Trichoderma gamsii
		Trichosporon asahii
Wallemia mellicola		*Wallemia mellicola*
	Verticillium incurvum	
Yarrowia lipolytica (*Candida lipolytica*)		*Yarrowia lipolytica* (*Candida lipolytica)*
		Zygosaccharomyces rouxii (*Candida mogii*)

Abbreviations: DB: database; FGT: Female genital tract

Functional Mycobiome Profile

In all databases, protein identifications included heat shock proteins (HSPs), Actin, HATPase_c domain-containing protein, and 6-phosphogluconate dehydrogenase (Table 5.6). The majority of proteins were assigned to HSPs (>2) in each dataset (Table 5.6).

Table 5.6. Number of protein identifications of commonly detected proteins across protein sequence databases and/or proteins that had multiple hits in a single database.

Protein names	HIV curated DB (n =32)	Adjusted HIV curated DB (n =13)	FGT curated DB (n =13)
Heat shock proteins	3	3	2
Actin	1	3	2
Serine/threonine-protein phosphatase	0	0	2
Endoplasmic reticulum chaperone BiP proteins	3	0	0
Glyceraldehyde-3-phosphate dehydrogenase	3	0	0
ATP-dependent RNA helicase	2	0	1

HATPase_c domain-containing protein	1	1	1
6-phosphogluconate dehydrogenase	1	1	1
Protein kinase domain-containing protein	1	0	1
Uncharacterized	7	0	0

Abbreviations: DB: database; FGT: Female genital tract;

According to Gene Ontology (GO) annotation for each DB, at least 50% of fungal proteins mapped to biological processes (BPs), at least 46.2% proteins mapped to cellular components (CCs), and over 76.9% proteins to molecular functions (MFs) using *UniProt*'s ID Mapping Tool across databases (Fig 5.4).

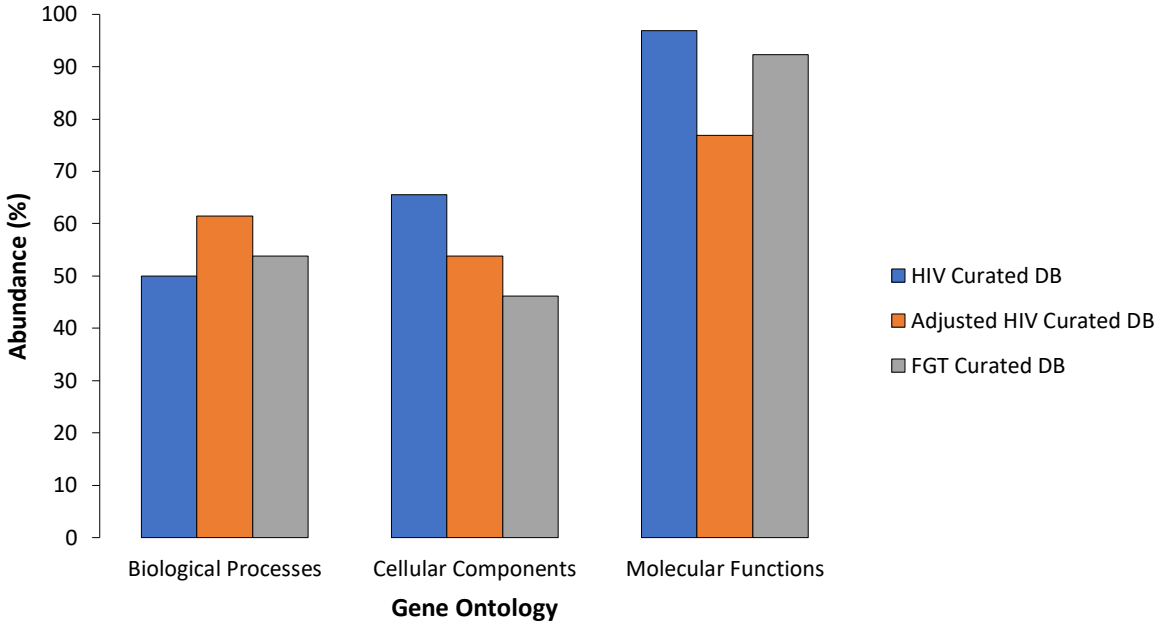


Figure 5.4. Relative abundance of proteins that assigned to a Gene Ontology (GO) for each protein sequence database. The majority of protein were able to assign to molecular function.

Using the **HIV curated DB**, the most relatively abundant biological processes were maturation of SSU-rRNA from tricistronic rRNA transcript (25.0%), ribosomal small subunit assembly

(25.0%), and translation (25.0%) (Fig 5.5). Using the **adjusted HIV curated DB**, the BPs, the most relative abundant pathways were pathways were the proteasome-mediated ubiquitin-dependent protein catabolic process (25.5%), and endocytosis (7.8%) (Fig 5.5). Lastly, using the **FGT curated DB**, the most relatively abundant pathways were histone H4 acetylation and membrane raft polarization (46.4%) (Fig 5.5). "Other" constituted BPs with a relative abundance of less than 1% (see Supplementary Table S2).

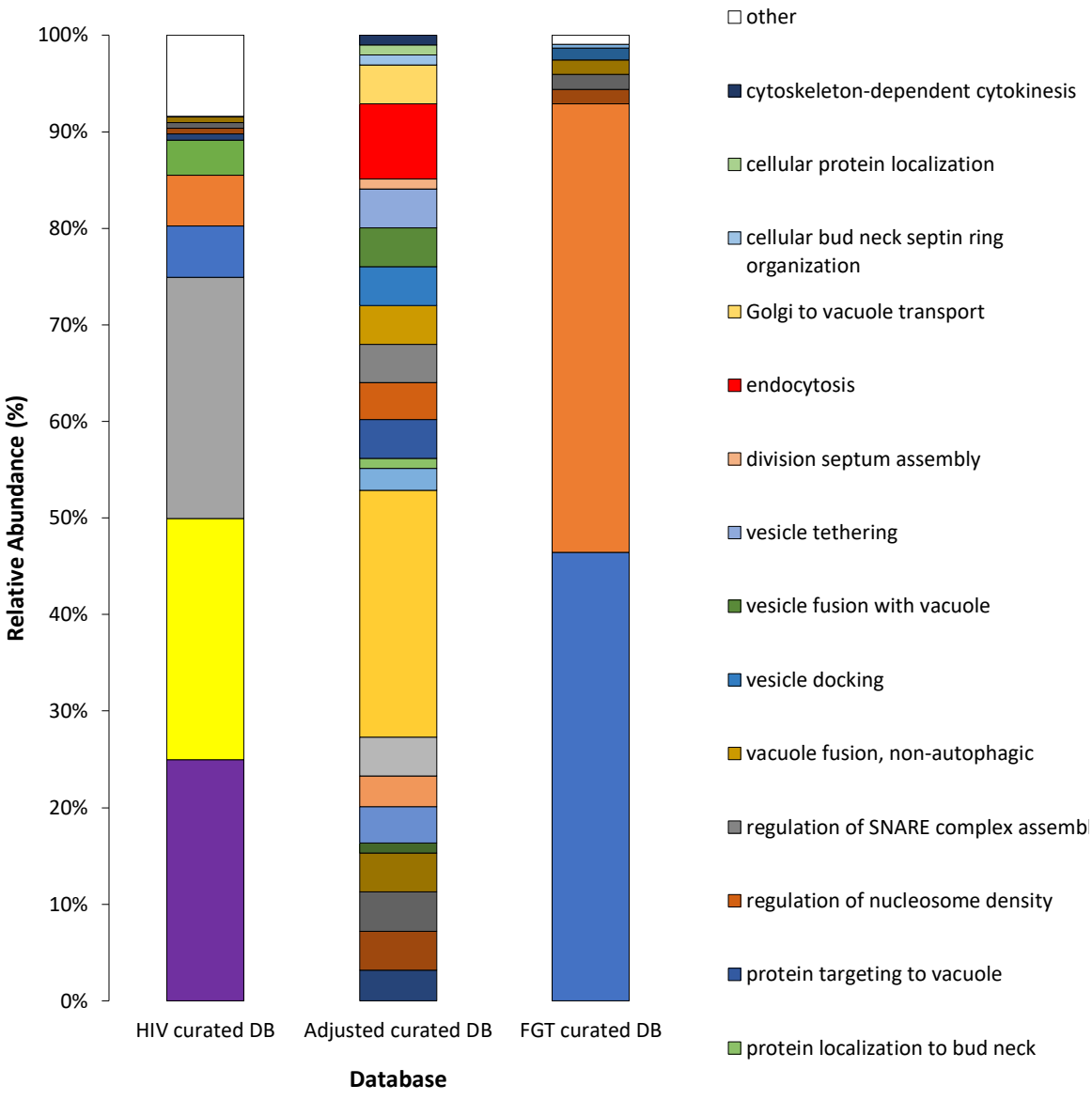


Figure 5.5. Relative Abundance of Biological Processes (BP) assigned to fungi according to Gene Ontology (GO) using summed iBAQ protein intensity values of each BP from Uniprot using each protein sequence database on the Månberg *et al.* (2019) dataset.

The **adjusted HIV curated DB** was the only database that did not have all fungal proteins assign to *KEGG* terms using *GhostKOALA,* as only 12 fungal proteins had assignments.

Across each of the three databases, the majority of protein sequences were assigned to genetic information processing. Other assignments that were present across databases included protein families: signaling and cellular processing, carbohydrate metabolism, and cellular processes. The **HIV curated DB** was the only database that had an assignment to energy metabolism (Fig 5.6).

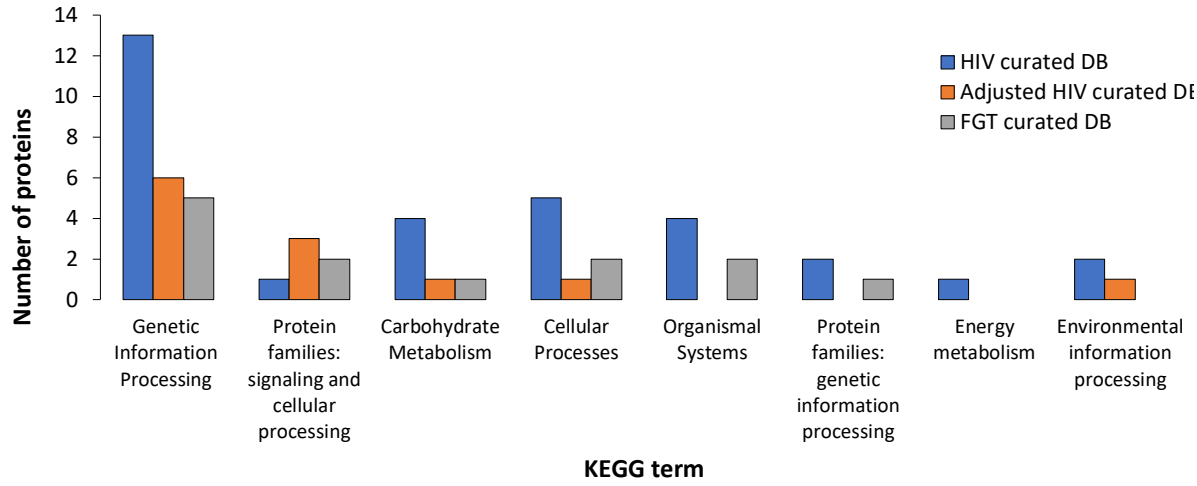


Figure 5.6. Number of protein sequences assigned to KEGG functional categories using GhostKOALA for protein sequence database. The majority of proteins assigned to genetic information processing.

The relative abundance plot of *KEGG* Ontology (*KO*) terms showed differing relative abundant KO terms using each protein sequence database. Using the **HIV curated DB**, there was a high relative abundance of proteins involved in the ribosome (translation) (32.2%), and ubiquitin-mediated proteolysis (folding, sorting, and degradation) (32.2%) (Fig 5.7). With the **adjusted HIV curated DB** there was a high relative abundance of proteins involved in protein processing in the endoplasmic reticulum (folding, sorting, and degradation) (12.0%), spliceosome (transcription) (11.5%), MAPK signaling pathway (10.1), and endocytosis (10.1%) (Fig 5.7). Lastly, using the **FGT curated** DB, there was high relative abundance of proteins were involved in the protein processing in the PI3K-Akt signaling pathway (17.5%), Hippo signaling pathway (15.3%), cGMP-PKG signaling pathway (15.3%), and Regulation of actin cytoskeleton (15.6%) (Fig 5.7). The category “Other” included KO terms identified to have a relative abundance of less than 1% (Supplementary Table S3).

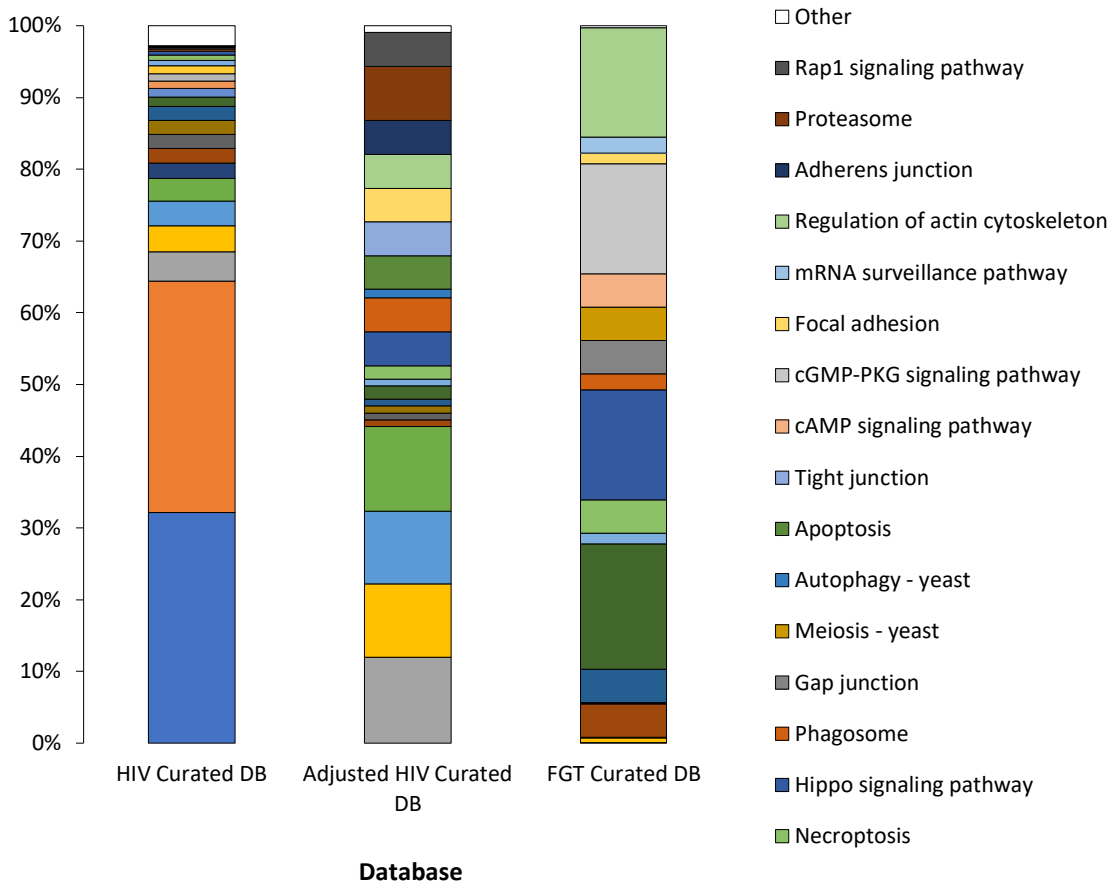


Figure 5.7. Relative Abundance of KEGG Ontology (KO) terms using summed iBAQ protein intensity of proteins assigned to each KO term using GhostKOALA for each protein sequence database.

Afiuni-Zadeh *et al.* (2018) dataset

Afiuni-Zadeh *et al.* (2018) obtained de-identified residual (waste) PAP tests that were collected per routine procedures using a liquid-based PAP test at the University of Minesotta. Cervical cells were collected from the ectocervix of healthy women by a physician using a broom-like device specifically designed for this purpose. Individuals selected for this study were at least 50 years old (median age of 58 years; ranging from 50–76 years) with normal cytology and without visible blood contamination. Vials were stored for one month at room temperature after the PAP test sample had been processed. The fractions of pooled or individual PAP test samples were run on an LTQ Orbitrap Velos mass spectrometer.

Table 5.7. Information on the storage and processing of samples collected by Afiuni-Zadeh *et al.* (2018).

Buffer solution	Mixture consisting of 21.7% ethanol, 1.2% methanol, 1.1% isopropanol, and formaldehyde
After collection of discarded samples	Vials were vortexed to resuspend proteins and to release cells/proteins from the cervical sampling device that remained in each vial
Digestion	Trypsin digestion
Precipitation and MS preparation	Proteins were precipitated using acetone were prepared for MS by Filter Aided Sample Preparation
Fractionation	Peptides were 2D fractionated offline by high pH reverse phase chromatography into 32 fractions. Fractions were concatenated into 16, and each of the concatenated fractions was submitted to the LTQ Orbitrap Velos spectrometer.

The generated database from Metanovo using default parameters for the Afiuni-Zadeh *et al.* (2018) dataset concatenated to the fungal Pan Proteome database was termed the **PAP curated DB**.

In the **PAP curated DB**, 491 (12.7%) of protein groups overlapped with the **FGT curated DB** (Fig 5.8). Of the shared protein groups, 2 were from the bacteria *L. crispatus*, 478 were from human proteins, and 11 were from fungi. Thus, we can confirm the likely presence of six fungal species identified, which include *Wallemia mellicola* (3), *M. sympodialis* (2), *Zycosaccharomyces rouxii* (2), *Cryptococcus neoformans* (2), *C. albicans*, and *Trichosporon asahii*.

Since the **PAP curated DB** had a higher MS/MS, protein group, and peptide sequence identification rate, it proved to be a suitable database for downstream taxonomic and functional analyses (Table 5.8). Therefore, QC was not performed on the **FGT curated DB** results.

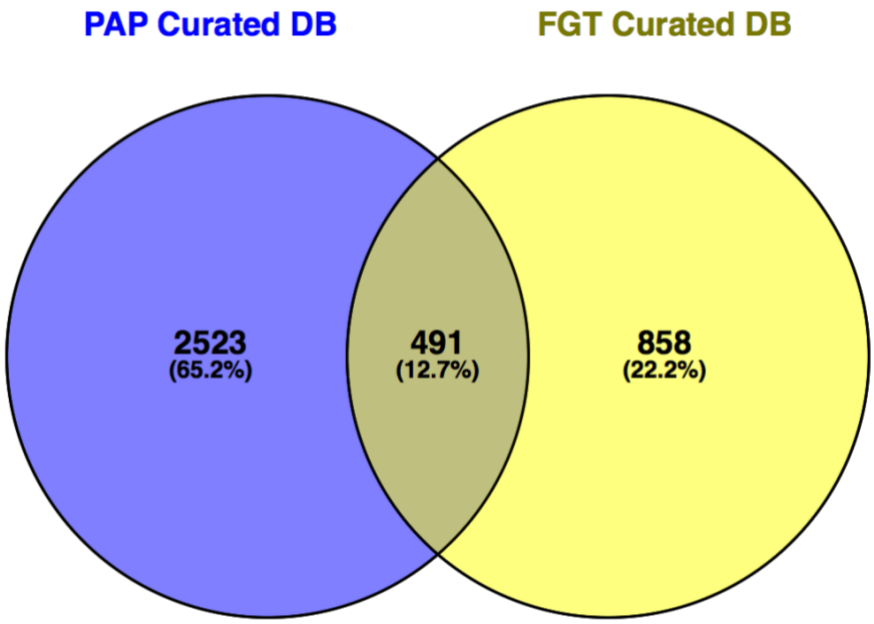


Figure 5.8. Venn diagram depicting the number of protein groups identified using two different protein sequence databases, namely the PAP curated database (DB) and the FGT curated DB. The PAP curated DB identified 3014 protein groups, and the FGT curated DB identified 1349 protein groups. Between the two protein databases 491 protein groups were shared.

Table 5.8. Overview of MS/MS identification rate, protein group and peptide sequence identifications from the two protein sequence databases searched in MaxQuant using the Afiuni-Zadeh *et al.* (2018) dataset.

Database	MS/MS Identification Rate (%)	Protein Group Identifications	Peptides Sequence Identifications	Peptides After QC
PAP curated DB	19.29	3,014	14,910	14,279
FGT curated DB	11.58	1,349	7,207	-

Abbreviations: DB: database; MS/MS: Mass Spectrum; FGT: Female genital tract; QC: quality control

Taxonomic profile

From the protein group identifications, we calculated the relative abundances of taxonomic kingdoms in the Afiuni-Zadeh *et al.* (2018) dataset. As seen previously, Metazoa had the highest relative abundance. The remainder of kingdoms in order of relative abundance included Fungi, Bacteria, Viridiplantae, Archaea, and Sar (Table 5.9).

Table 5.9. Relative abundance of taxonomic kingdoms using the summed total iBAQ protein intensity values for each taxonomic kingdom and subkingdom identified using the PAP curated DB on the Afiuni-Zadeh *et al.* (2018) dataset.

Kingdom/ Subkingdom	Relative Abundance (%)
Metazoa	97.3
Fungi	0.6
Bacteria	0.6
Viridiplantae	1.4
Archaea	0.0002
Sar	0.009

From peptide sequence identifications after QC, once again Metazoa made up the majority of peptide assignments (76.2%), the remainder of assignments included those of bacterial origin (2.1%), fungi (0.7%), virdiplantae (0.2%), and archaea (0.1%) (Fig 5.9). Of the remaining peptide sequences, 13.6% were assigned to root, and 2.4% could not be assigned to a taxonomic kingdom.

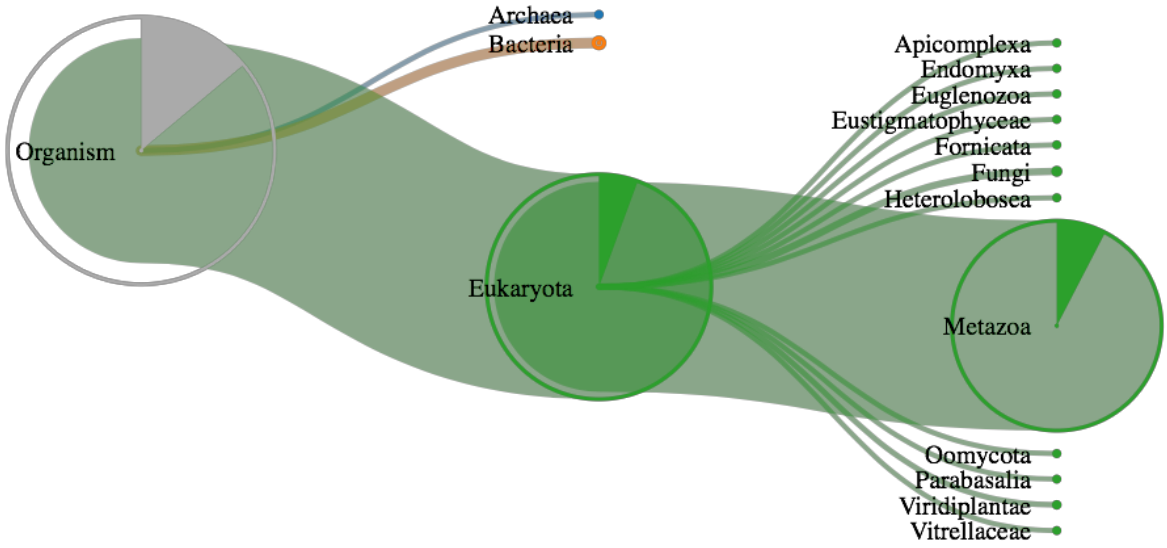


Figure 5.9. Visual representation of the proportion of peptide sequences assigned to each taxonomic kingdom identified in the Afiuni-Zadeh et al. (2018) dataset using the PAP curated database. Taxonomic assignment and visualization was done using Unipept.

.

Fungal peptide sequences were assigned to three phylum levels, the majority (79.8%) were assigned to the phyla Ascomycota, followed by 18.3% of sequences to Basidiomycota as previously seen, and 1.9% of sequences to Mucoromycota.

This study identified 20 fungal genera using protein group assignments. The relative abundance of Cryptococcus (80.5%) was the highest, followed by Candida (8.3%), and Armillaria (3.5%) (Fig 5.10). The category "Other" included fungal genera that had a relative abundance of less than 1% (see supplementary table S4). Zoning in on the species level, the majority of protein groups assigned to *Malassezia sympodialis*, *Wallemia mellicola*, and *Zygosaccharomyces rouxii* (Table 5.10).

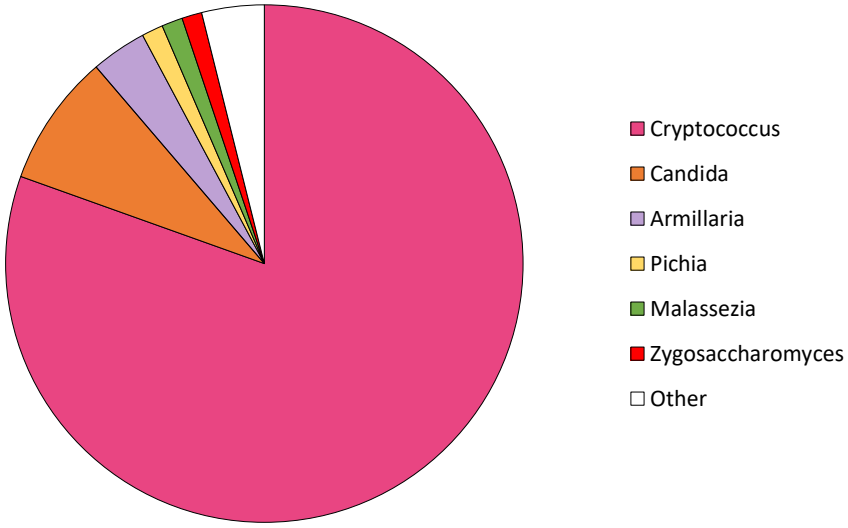


Figure 5.10. Relative abundance of fungal genera using the summed total iBAQ intensity values of proteins assigned to each fungal genus (*n* = 20) identified in the Afiuni-Zadeh *et al.* (2018) dataset using the PAP curated database.

Table 5.10. Fungal species identified using protein group assignments in the Afiuni-Zadeh *et al.* (2018) dataset using the PAP curated DB.

Fungal species	Number of proteins
Malassezia sympodialis	3
Wallemia mellicola	3
Zygosaccharomyces rouxii (Candida mogii)	3
Cryptococcus neoformans	2

Gibberella zeae (Fusarium graminearum)	2
Trichosporon asahii	2
Ajellomyces dermatitidis (Blastomyces dermatitidis)	1
Armillaria gallica	1
Aspergillus candidus	1
Aspergillus taichungensis	1
Auriculariopsis ampla	1
Candida albicans	1
Candida glabrata	1
Clathrospora elynae	1
Clavispora lusitaniae (Candida lusitaniae)	1
Moesziomyces aphidis	1
Monosporascus sp. MG133	1
***Pichia kudriavzevii* (*Candida krusei)**	1
Sugiyamaella lignohabitans	1
Talaromyces marneffei (Penicillium marneffei)	1
Yarrowia lipolytica (Candida lipolytica)	1

In total, we identified 46 distinct fungal genera using peptide sequences. Of those genera, 13.5% of sequences assigned genera were to *Aspergillus*, 5.8% to *Penicillium*, followed by 4.8% to *Clavispora*, *Moesziomyces*, *Saccharomyces*, *Trichoderma*, and 3.8% to *Pichia*. The remaining peptide sequences 9.6% were not assigned to the genus level.

Functional Mycobiome Profile

Of the 30 fungal protein groups that passed QC filtering by filtering out proteins that had < 2 peptides, contaminants, or reverse hits, 13.3% were Actin proteins, 10% were HSPs, of which

6.7% were HSP70, and 6.7% of protein groups were Proteasome subunit alpha type proteins. Of the remaining proteins, 6.7% were uncharacterized.

According to GO annotation, 56.7% of the fungal proteins mapped to BPs, 46.7% proteins to CCs, and 80% proteins to MFs using *UniProt*'s ID Mapping Tool. From the BPs, a high relative abundance of proteins were involved in histone exchange; chromatin silencing; negative regulation of CENP-A containing nucleosome assembly; nuclear-transcribed mRNA catabolic process, non-stop decay; positive regulation of transcription involved in G1/S transition of mitotic cell cycle; regulation of transcription by RNA polymerase II; and silent mating-type cassette heterochromatin assembly (Fig 5.11). The category "Other" included BPs that constituted a relative abundance of less than 1% (see Supplementary Table S5).

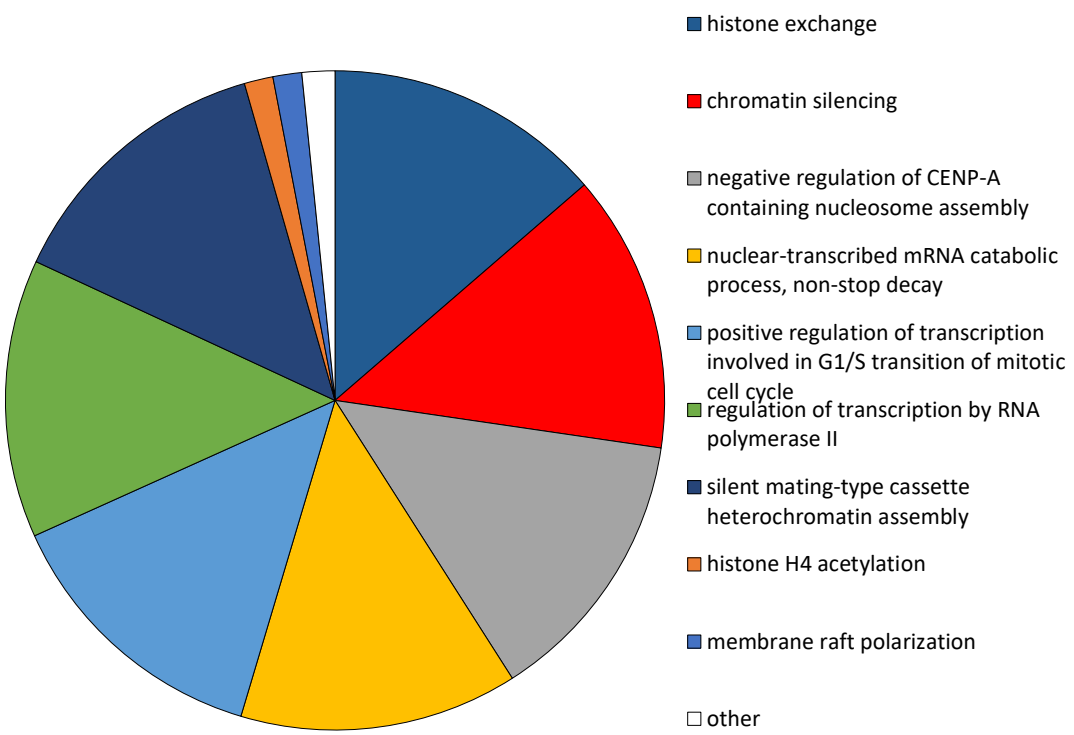


Figure 5.11. Relative Abundance of fungal Biological Processes (BP) according to Gene Ontology (GO) using summed iBAQ protein intensity values for each BP identified (*n* = 17) using the PAP curated DB on the Afiuni-Zadeh *et al.* (2018) dataset.

Of fungal proteins, 96.7% were assigned to a *KEGG* term using *GhostKOALA*. KEGG term assignments revealed that the majority of protein sequences were assigned to genetic

information processing (51.7%), protein families: signaling and cellular processes (17.2%), cellular processes (6.9%), and environmental information processing (6.9%) (Fig 5.12).

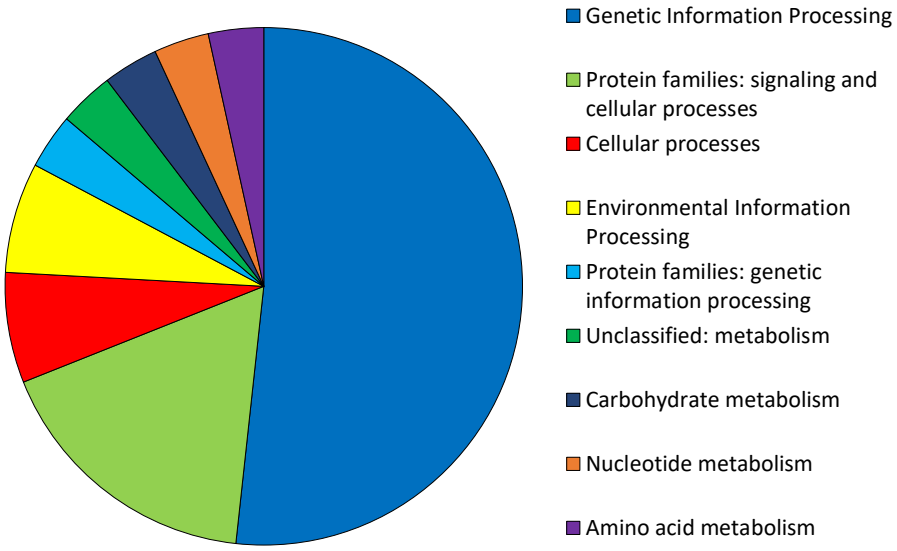


Figure 5.12. Number of protein sequences assigned to KEGG functional categories using GhostKOALA (*n* = 29). Genetic information processing had the most proteins assigned (*n* =15).

A relative abundance plot of *KO* terms showed that proteins involved in necroptosis (cell growth and death) was abundant (82.7%) (Fig 5.13). *KO's* that constituted less than 1% of the relative abundance were added to the "Other" category (see Supplementary Table S6).

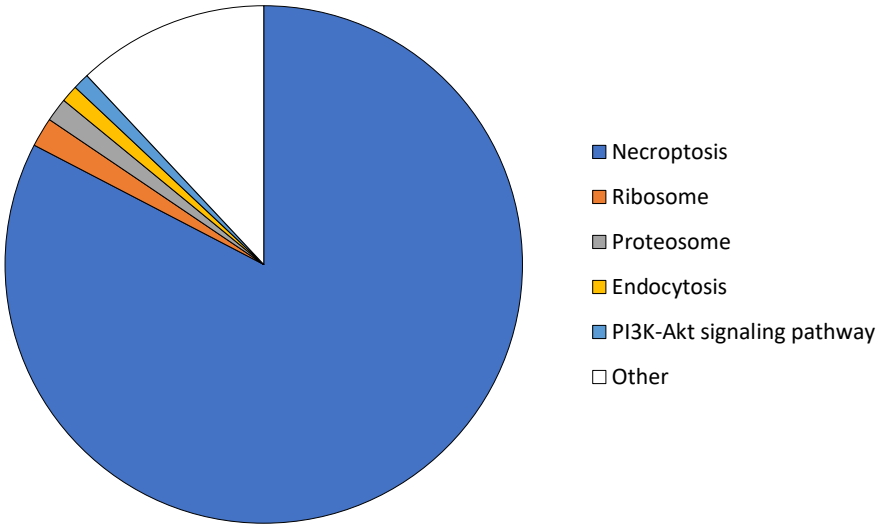


Figure 5.13. Relative Abundance of KEGG Ontology (KO) terms using iBAQ protein intensity of proteins (*n* = 30) assigned to each KO term using GhostKOALA. Necroptosis was the most relatively abundant KO term.

Borgdorff *et al.* 2016

Taxonomic profile

Borgdorff *et al.* (2016) collected samples by irrigating the left and right fornix and cervical os with normal saline, which was aspirated after 30 seconds to collect cervicovaginal lavage (CVL) fluid. The study cohort included Rwandan female sex workers between 2006 and 2009. Study participants (*n* = 50) were between 19-45 years of age, not pregnant, and their CVLs were not macroscopically bloody. Peptide mixtures were analyzed by online nanoflow liquid chromatography using the nanoACQUITY-nLC system (Waters) coupled to an LTQ-Orbitrap Velos (ThermoFisher Scientific, Bremen, Germany) mass spectrometer equipped with the manufacturer's nanospray ion source.

Table 5.11. Information on the storage and processing of samples collected by Borgdorff *et al.* (2016).

Post-collection storage	The CVL fluid was immediately placed on ice or at 41°C.
Centrifugation	The CVLs were centrifuged at 1,000rpm for 10min within 4h of collection.

Filtration	Supernatants were filtered using a sterile 0.2 mm cellulose acetate membrane
Storage	Cell pellets and aliquots of supernatant were stored at -80^{o}C until testing.
Heat inactivation	CVL supernatants were heat inactivated for 30 min at 56 ^{o}C before further processing.
Precipitation and centrifugation	Soluble proteins were precipitated using ice-cold 30% trichloroacetic acid in acetone and incubated. Samples were centrifuged at 12,000 g for 10 min to pellet proteins. Pellets were washed three times with ice-cold acetone and allowed to air dry.
Resuspension, reduction, and alkylation	Protein pellets were resuspended in 25mM ammonium bicarbonate, 0.05% rapigest, reduced, and alkylated.
Digestion	Proteomic-grade trypsin at a protein:trypsin ratio of 50:1
Rapigest precipitation	Rapigest was precipitated by addition of trifluoroacetic acid

The generated database from Metanovo using default parameters for Borgdorff *et al.'s* (2016) dataset concatenated to the fungal Pan Proteome database was termed the **CVL curated DB**.

In the **CVL curated DB** 10 protein groups overlapped with the **FGT curated DB** (Fig 5.14). Of those protein groups that overlapped, 9 belonged to fungi to 8 fungal taxa. This confirms the likelihood of the 8 fungal species including: *M. sympodialis* (2), *Ustilago maydis*, *Rhizopogon vinicolor*, *C. albicans*, *Batrachohytrium salamandrivorans*, *Cercospora beticola*, *Pseudozyma antartica* (*Candida antarctica*), *Schizosaccharomyces cryophilus*, and *Endocarpon pusillum* as part of the vaginal mycobiome.

Due to the low MS/MS, protein group, and peptide sequence identification rate for the **CVL curated DB** (Table 5.12), the **FGT curated DB** was used for downstream taxonomic and functional analyses to obtain more representative results.

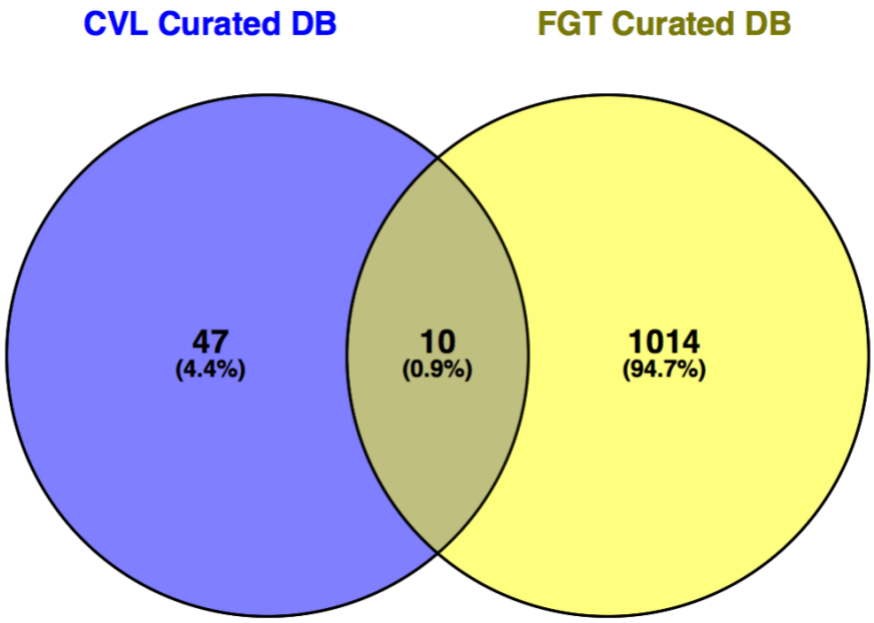


Figure 5.14. Venn diagram depicting the number of protein groups identified using MaxQuant with the CVL curated database (DB) and the FGT curated DB. The CVL curated DB identified 57 protein groups, and the FGT curated DB identified 1114 protein groups. Between the two protein databases 10 protein groups overlapped between databases.

Table 5.12. Overview of MS/MS identification rate, protein group, and peptide sequence identifications of each protein sequence database search in MaxQuant for the Borgdorff *et al.* (2016) dataset.

Database	MS/MS Identification Rate (%)	Protein Groups	Unique Peptides	Unique Peptides after QC
CVL curated DB	2.85	57	962	-
FGT curated DB	21.3	1,114	8,493	7,910

Abbreviations: DB: database; MS/MS: Mass Spectrum; FGT: Female genital tract; QC: quality control

Taxonomic Profile

We calculated the relative abundances of taxonomic kingdoms in the Borgdorff *et al.* (2016) dataset using iBAQ protein intensities. As seen in the previous two analyses, Metazoa had the highest relative abundance. The remaining relatively abundant kingdoms included Fungi followed by Bacteria, and Fungi (Table 5.13). This dataset was the only one to not detect Virdiplantae.

Table 5.13. Relative abundance of taxonomic kingdoms identified in the Borgdoff *et al.* (2016) dataset using the summed iBAQ protein intensity values of each taxonomic kingdom assigned using UniProt.

Kingdom	Relative Abundance (%)
Metazoa	98.8
Fungi	0.9
Bacteria	0.3

From the peptides identified from MS analysis fungi represented 0.6% of the total identifications (Fig 5.15). Of the remaining kingdoms, 9.5% of the identified peptides were of bacterial origin, and 77.7% were of Metazoan origin (Fig 5.15). The remaining peptide sequences were assigned to root (8.1%) and 1.7% could not be assigned to a taxonomic kingdom.

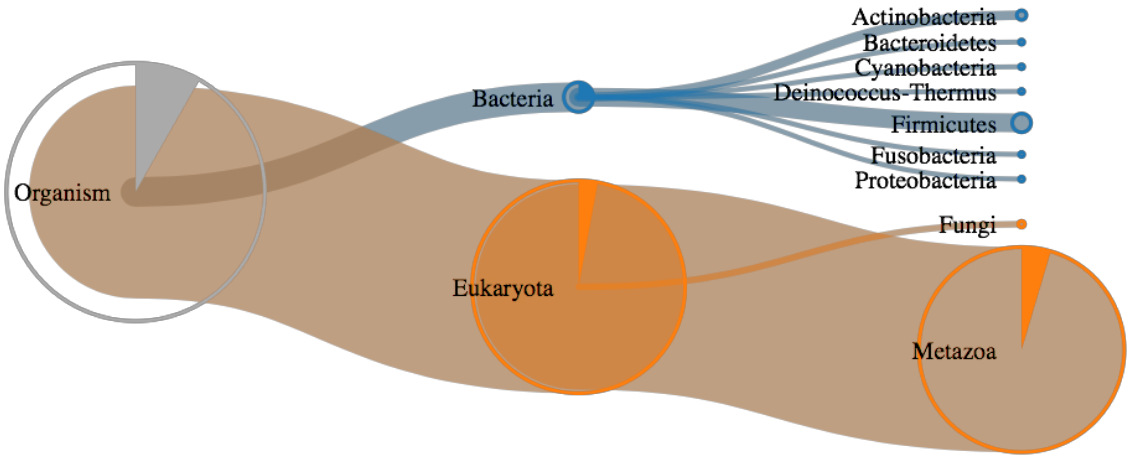


Figure 5.15. Visual representation of the proportion of peptide sequences assigned to each taxonomic kingdom in the Borgdorff *et al.* 2016 dataset using the FGT curated database. Taxonomic assignment and visualization was done using Unipept.

Once again peptide sequences assigned to two phyla, *viz.* Ascomycota (85.1%) and Basidiomycota (10.6%). In total, we identified 15 distinct fungal genera using peptide sequences. Of those genera, 25.5% of sequences were assigned to *Aspergillus*, 17.0% to *Saccharomyces*, followed by 4.3% to *Penicillium*, *Pichia*, *Pseudogymnoascus*, *Yarrowia*. This dataset had the large amount of peptide sequences (21.31%) that were not assigned to the genus level.

This study identified 15 fungal genera using protein groups after the exclusion criterion was implemented. The relative abundance of *Saccharomyces* (93.1%) was the highest, followed by *Aspergillus* (1.4%), and *Endocarpon* (1.09%) (Fig 5.16). The category "other" constituted fungal genera that had a relative abundance of less than 1% (see supplementary table S7). Of fungal protein species-level assignments, the majority assigned to *Saccharomyces cerevisiae*, followed by *Aspergillus niger*, and *M. sympodialis* (Table 5.14).

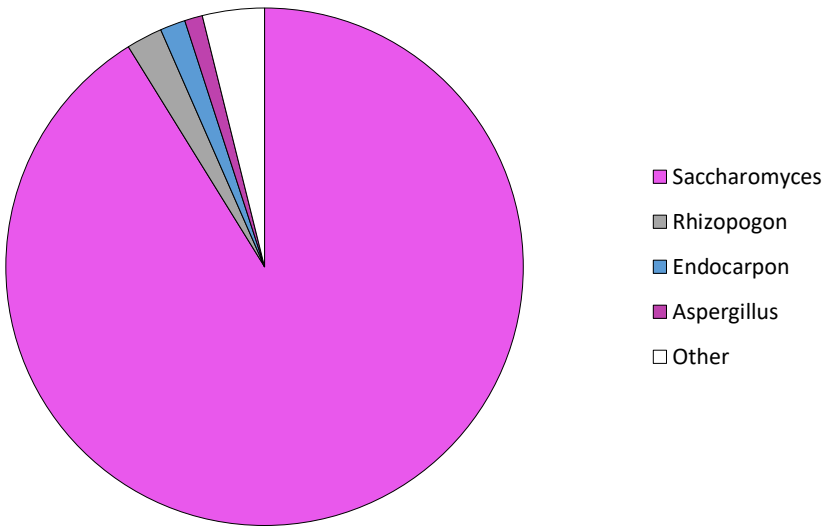


Figure 5.16. Relative abundance of fungal genera using summed total iBAQ intensity values of proteins assigned to each fungal genus (*n* = 15) identified in Borgdorff *et al.* (2016) dataset using the FGT curated database. *Saccharomyces* was the most relatively abundant genus.

Table 5.14. Number of protein groups assigned to each fungal species Borgdorff *et al.* (2016) dataset using the FGT curated databases.

Fungal species	Number of proteins (n)
Saccharomyces cerevisiae	3
Aspergillus niger	2
Malassezia sympodialis (Atopic eczema-associated yeast)	2
Candida albicans	1
Cercospora beticola	1
Cyberlindnera jadinii (Pichia jadinii)	1
Endocarpon pusillum	1

Kluyveromyces lactis (Candida sphaerica)	1
Kluyveromyces marxianus (Candida kefyr)	1
***Pichia kudriavzevii* (*Candida krusei*)**	1
Pseudogymnoascus sp. 24MN13	1
Pseudozyma antarctica (Candida antarctica)	1
Rhizoclosmatium globosum	1
Rhizopogon vinicolor	1
Schizosaccharomyces cryophilus	1
***Yarrowia lipolytica (Candida lipolytica*)**	1

Functional Mycobiome Profile

Of the 21 fungal protein groups that passed QC filtering, 19.0% were Actin proteins and Alcohol dehydrogenase, 9.5% were Proteosome subunit alpha type proteins, and 4.8% were HSPs. Of the remaining proteins, 4.8% was uncharacterized.

According to GO annotation, 47.6% of the fungal proteins mapped to BPs, 47.6% proteins to CCs, and 57.1% proteins to MFs. From the BPs, this database identified new relatively abundant GO annotations, including amino acid catabolic process to alcohol via Elrlich pathway, and NADH oxidation, ethanol biosynthetic process involved in glucose fermentation ethanol, and regulation of viral genome replication (Fig 5.17). “Other” constituted BPs with a relative abundance of less than 1% (see Supplementary Table S8).

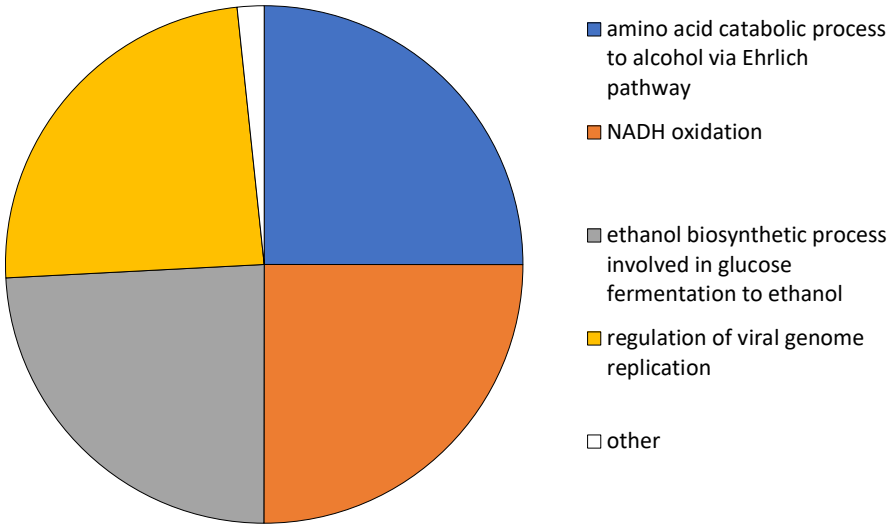


Figure 5.17. Relative Abundance of fungal biological processes (BP) according to Gene Ontology using summed iBAQ protein intensity values for each BP identified (*n* = 10) using the FGT curated DB on the Borgdorff *et al.* (2016) dataset.

A large proportion of fungal proteins (70%) were assigned to a *KEGG* term, revealing that the majority of protein sequences were assigned to Carbohydrate Metabolism (23.8%), Protein families: signaling and cellular processes (19.0%), and Genetic Information Processing (19.0%) (Fig 5.18).

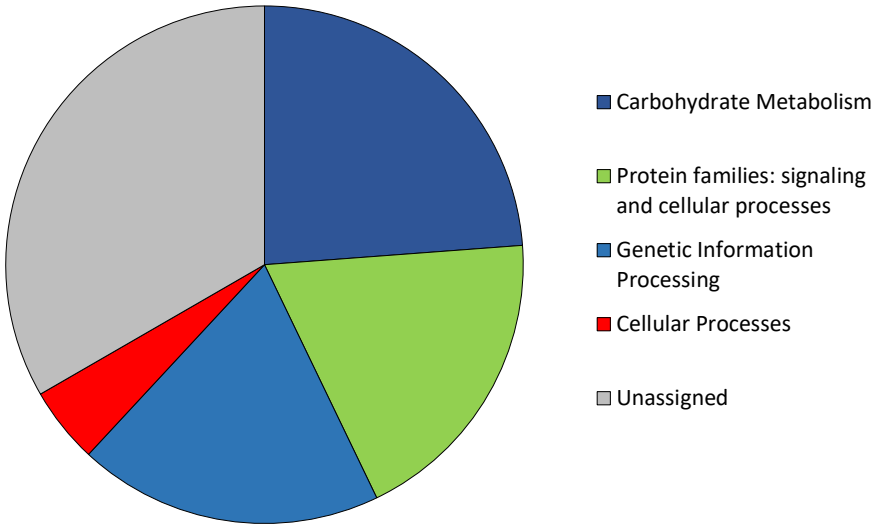


Figure 5.18. Number of protein sequences (*n* = 21) assigned to KEGG functional categories using GhostKOALA on the Borgdorff *et al.* (2016) dataset using the FGT curated DB. The majority of protein sequences were unassigned (*n* = 7), and a large proportion assigned to carbohydrate metabolism (*n* = 5).

A plot of *KO* terms showed that the most relatively abundant pathways were the Biosynthesis of secondary metabolites, Microbial metabolism in diverse environments (7.77%), Degradation of aromatic compounds, Glycolysis/Gluconeogenesis, Pyruvate metabolism, Fatty acid degradation, Tyrosine metabolism, Retinol metabolism, Chloralkane and chloroalkene degradation, Naphthalene degradation, Metabolism of xenobiotics, and Drug metabolism (7.58%) (Fig 5.19). “Other” constituted KO terms that had a relative abundance of less than 1% and can be found in Supplementary Table S9.

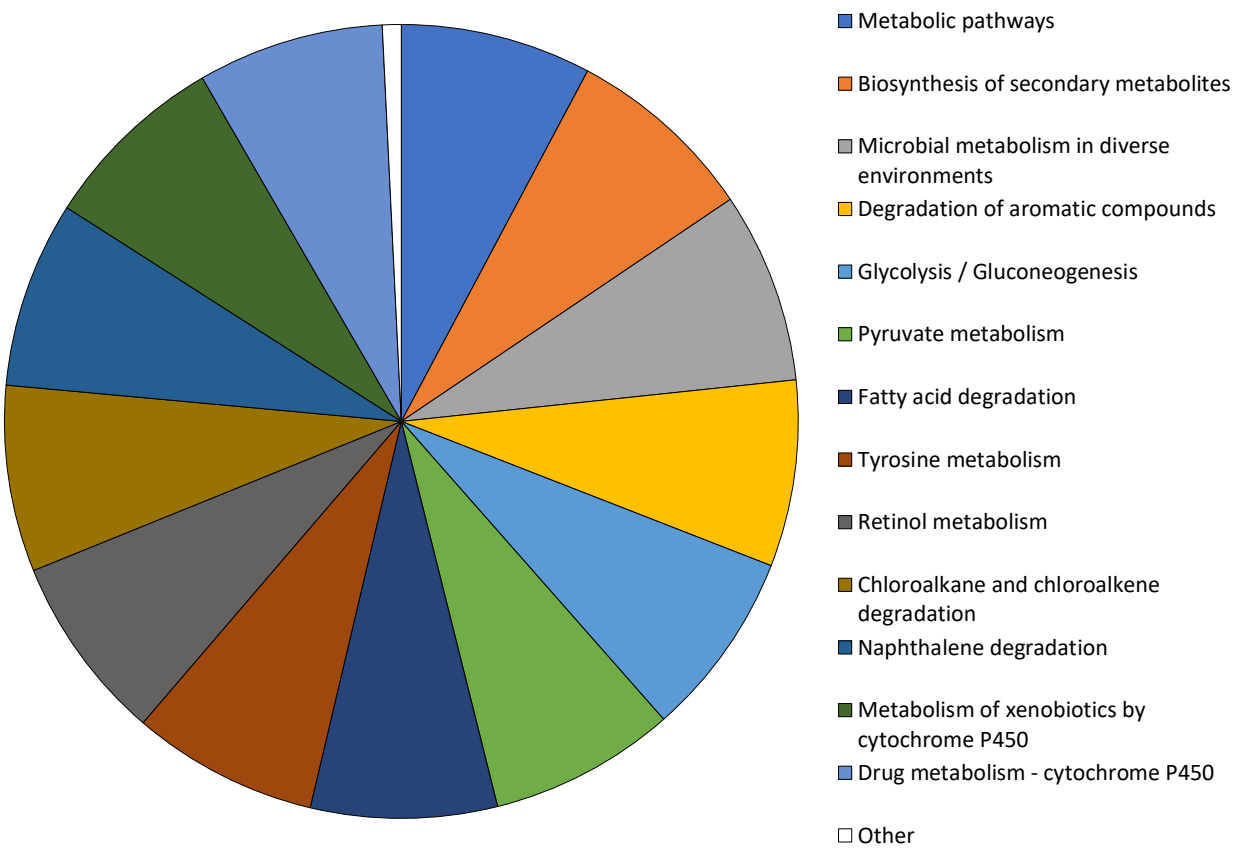


Figure 5.19. Relative Abundance of KEGG Ontology (KO) Terms using summed iBAQ protein intensity of proteins (*n* = 21) assigned to each KO term (*n* = 25) according to GhostKOALA. Metabolic pathways, biosynthesis of secondary metabolities, and microbial metabolism in diverse environments were the most relative abudant KO terms.

Comparisons

This section makes comparisons between the results of the publicly available datasets and the results we identified using our WISH dataset in Chapter 3, to identify any overlaps and differences.

Fungal genera that appeared in all datasets using peptide sequence identifications included, *Malassezia*, *Penicillium*, *Yarrowia*, *Aspergillus*, *Cryptococcus, Saccharomyces*, *Trametes*, and *Sclerotina* (Table 5.12). Interestingly, commonly detected fungal genera from the literature, *Candida* and *Kluyveromyces,* were not identified at genus level in the Borgdorff *et al*. (2016) dataset (Table 5.15).

Table 5.15. Fungal peptide sequences that assigned to the genus level in two or more datasets. The fungal genera that were also identified in the Chapter 2 WISH dataset are highlighted in green.

Genera	Dataset	Number of datasets (n)
Malassezia	All datasets	4
Penicillium	All datasets	4
Yarrowia	All datasets	4
Aspergillus	All datasets	4
Cryptococcus	All datasets (not detected using adjusted HIV curated and HIV curated databases)	4
Sclerotinia	All datasets (not detected using adjusted HIV curated and HIV curated databases)	4
Trametes	All datasets (not detected using adjusted HIV curated and HIV curated databases)	4
Saccharomyces	All datasets (not detected using adjusted HIV curated DB and FGT curated databases)	4
Candida	All datasets except Borgdorff *et al.* (2016)	3
Debaryomyces	All datasets except Borgdorff *et al.* (2016)	3

Kluyveromyces	All datasets except Borgdorff *et al.* (2016)	3
Moesziomyces	Borgdorff *et al.* (2016); Afiuni-Zadeh *et al.* (2018); Chapter 3 WISH dataset	3
Talaromyces	Borgdorff *et al.* (2016); Afiuni-Zadeh *et al.* (2018); Chapter 3 WISH dataset	3
Trichoderma	Borgdorff *et al.* (2016); Afiuni-Zadeh *et al.* (2018), Chapter 3 WISH dataset	3
Fusarium	Manberg *et al*. (2019) (not detected using HIV curated); Afiuni-Zadeh *et al.* (2018); Chapter 3 WISH dataset	3
Blastomyces	Manberg et al. (2019) (not detected using adjusted HIV curated DB and FGT curated DB); Afiuni-Zadeh *et al.* (2018); Chapter 3 WISH dataset	3
Pichia	Manberg *et al.* (2019) (not detected using adjusted HIV curated DB and FGT curated DB); Borgdorff *et al.* (2016); Afiuni-Zadeh *et al*. (2018)	3
Zygosaccharomyces	Manberg *et al.* (2019) (not detected using adjusted HIV curated DB and FGT curated DB); Afiuni-Zadeh *et al.* (2018); Chapter 3 WISH dataset	3
Puccinia	Manberg *et al.* (2019) (not detected using adjusted HIV curated DB and HIV curated DB); Afiuni-Zadeh *et al*. (2018); Chapter 3 WISH dataset	3
Alternaria	Afiuni-Zadeh *et al.* (2018); Chapter 3 WISH dataset	2
Cordyceps	Afiuni-Zadeh *et al*. (2018); Chapter 3 WISH dataset	2
Cyberlindnera	Afiuni-Zadeh *et al*. (2018); Chapter 3 WISH dataset	2
Nakaseomyces	Afiuni-Zadeh *et al.* (2018); Chapter 3 WISH dataset	2
Schizosaccharomyces	Afiuni-Zadeh et al. (2018); Chapter 3 WISH dataset	2
Trichosporon	Afiuni-Zadeh *et al.* (2018); Chapter 3 WISH dataset	2
Serendipita	Manberg *et al.* (2019) (not detected using HIV curated DB and FGT curated DB); Afiuni-Zadeh *et al.* (2018)	2
Wallemia	Manberg *et al.* (2019) (not detected using HIV curated DB); Afiuni-Zadeh *et al.* (2018)	2

Ascochyta	Manberg et al. (2019) (not detected using adjusted HIV curated DB and HIV curated DB); Chapter 3 WISH dataset	2
Clavispora	Manberg et al. (2019) (not detected using adjusted HIV curated DB and HIV curated DB); Afiuni-Zadeh *et al.* (2018)	2
Lodderomyces	Manberg et al. (2019) (not detected using adjusted HIV curated DB and HIV curated DB); Chapter 3 WISH dataset	2

Fungal species level assignments that were detected in 3 or more datasets using peptide sequences included, *Yarrowia lipolytica (Candida lipolytica)*, *Trichoderma gamsii*, *Candida parapsilosis*, *M. sympodialis*, *Aspergillus fumigatus*, *Penicillium digitatum*, *Pichia kudriavzevii* (*Candida krusei*), *Trametes cinnabarina,* and *Puccinia striiformis* (Table 5.16).

Table 5.16. Fungal peptides that assigned to species level in two or more datasets. Fungal species that were also in the Chapter 3 WISH dataset are highlighted in green.

Fungal species	Dataset	Number of datasets (n)
Yarrowia lipolytica (*Candida lipolytica)*	All datasets	4
Trichoderma gamsii	Borgdorff *et al.* (2016); Afiuni-Zadeh *et al.* (2018); Chapter 3 WISH dataset	3
Candida parapsilosis	Manberg *et al.* (2019) (except HIV curated DB search); Afiuni-Zadeh *et al.* (2018); Chapter 3 WISH dataset	3
Malassezia sympodialis	Manberg *et al.* (2019) (except FGT curated DB search); Borgdorff *et al.* (2016); Afiuni-Zadeh *et al.* (2018)	3
Aspergillus fumigatus	Manberg *et al.* (2019) (except adjusted HIV curated DB and FGT curated DB search); Afiuni-Zadeh *et al.* (2018); Chapter 3 WISH dataset	3
Penicillium digitatum	Manberg *et al.* (2019) (except adjusted HIV curated DB and FGT curated DB search); Afiuni-Zadeh *et al.* (2018); Chapter 3 WISH dataset	3

Pichia kudriavzevii (*Candida krusei)*	Manberg *et al.* (2019) (except adjusted HIV curated DB and FGT curated DB search); Borgdorff *et al.* (2016); Afiuni-Zadeh *et al.* (2018)	3
Trametes cinnabarina	Manberg *et al.* (2019) (except adjusted HIV curated DB and HIV curated DB search); Borgdorff *et al.* (2016); Chapter 3 WISH dataset	3
Puccinia striiformis	Manberg *et al.* (2019) (except adjusted HIV curated DB and HIV curated DB search); Afiuni-Zadeh *et al.* (2018); Chapter 3 WISH dataset	3
Wallemia mellicola	Borgdorff *et al.* (2016); Afiuni-Zadeh *et al.* (2018)	2
Aspergillus niger	Borgdorff *et al.* (2016); Afiuni-Zadeh *et al.* (2018)	2
Moesziomyces antarcticus	Borgdorff *et al.* (2016); Afiuni-Zadeh *et al.* (2018)	2
Talaromyces marneffei	Borgdorff *et al.* (2016); Afiuni-Zadeh *et al.* (2018)	2
[Candida] glabrata	Afiuni-Zadeh *et al.* (2018); Chapter 3 WISH dataset	2
Alternaria alternata	Afiuni-Zadeh *et al.* (2018); Chapter 3 WISH dataset	2
Cyberlindnera jadinii	Afiuni-Zadeh *et al.* (2018); Chapter 3 WISH dataset	2
Zygosaccharomyces rouxii	Afiuni-Zadeh *et al.* (2018); Chapter 3 WISH dataset	2
Candida intermedia	Manberg *et al.* (2019) (except adjusted HIV curated DB and HIV curated DB search); Afiuni-Zadeh *et al.* (2018)	2
Ascochyta rabiei	Manberg *et al.* (2019) (except adjusted HIV curated DB and HIV curated DB search); Afiuni-Zadeh et al. (2018)	2
Aspergillus candidus	Manberg *et al.* (2019) (except adjusted HIV curated DB and HIV curated DB search); Afiuni-Zadeh *et al.* (2018)	2
Fusarium culmorum	Manberg *et al.* (2019) (except adjusted HIV curated DB and HIV curated DB search); Chapter 3 WISH dataset	2
Lodderomyces elongisporus	Manberg *et al. (*2019) (except adjusted HIV curated DB and HIV curated DB search); Chapter 3 WISH dataset	2
Sclerotinia borealis	Manberg *et al.* (2019) (except adjusted HIV curated DB and HIV curated DB search); Borgdorff *et al.* (2016)	2
Blastomyces dermatitidis	Manberg *et al.* (2019) (except adjusted HIV curated DB and FGT curated DB search); Afiuni-Zadeh *et al.* (2018)	2

Saccharomyces eubayanus	Manberg *et al.* (2019) (except adjusted HIV curated DB and FGT curated DB search); Chapter 3 WISH dataset	2

Table 5.17 lists peptide sequences that assigned to the genus level in the chapter 3 WISH dataset but were not detected in the publicly available datasets.

Table 5.17. Fungal genera found in chapter 3 WISH dataset but were not detected in public datasets.

Fungal Genera
Allomyces
Conidiobolus
Cutaneotrichosporon
Diaporthe
Exidia
Moniliophthora
Neolecta
Ordospora
Phialophora
Psathyrella
Rhinocladiella
Rhizoctonia
Rhizopus
Rozella
Sporothrix
Spraguea
Stachybotrys
Teratoramularia
Torulaspora

Wickerhamomyces

Table 5.18. Fungal protein groups assigned to species levels and the databases they were detected in. Fungal species that showed up in the Chapter 3 WISH dataset are highlighted in green.

Fungal species	**Dataset**	**Number of datasets (n)**
Candida albicans	All datasets (except search using adjusted HIV DB)	4
Cercospora beticola	All datasets, except Afiuni-Zadeh *et al.* (2018)	3
Malassezia sympodialis	All datasets, except Manberg *et al.* (2019)	3
Pichia kudriavzevii (*Candida krusei)*	All datasets, except chapter 3 WISH dataset	3
Schizosaccharomyces cryophilus	All datasets, except Afiuni-Zadeh *et al.* (2018)	3
Yarrowia lipolytica (*Candida lipolytica)*	All datasets, except chapter 3 WISH dataset	3
Clavispora lusitaniae (*Candida lusitaniae)*	All datasets, except Afiuni-Zadeh *et al.* (2018)	3
Moesziomyces aphidis	All datasets, except Borgdorff *et al.* (2016)	3
Talaromyces marneffei	All datasets, except Borgdorff *et al.* (2016)	3
Wallemia mellicola	All datasets, except Borgdorff *et al.* (2016)	3
Cryptococcus neoformans	Afiuni-Zadeh *et al.* (2018); Manberg *et al.* (2019) (except adjusted HIV curated DB and FGT curated DB search); Chapter 3 WISH dataset	3
Trichosporon asahii	Afiuni-Zadeh *et al.* (2018); Manberg *et al.* (2019) (except adjusted HIV curated DB and FGT curated DB search); Chapter 3 WISH dataset	3
Rhizoclosmatium globosum	Borgdorff *et al.* (2016); Manberg *et al.* (2019) (except HIV curated DB and FGT curated DB search); Chapter 3 WISH dataset	3

Rhizopogon vinicolor	Borgdorff *et al.* (2016); Manberg *et al.* (2019) (except adjusted HIV curated DB and HIV curated DB search); Chapter 3 WISH dataset	3
Endocarpon pusillum	Borgdorff *et al.* (2016); Chapter 3 WISH dataset	2
Kluyveromyces lactis (Candida sphaerica)	Borgdorff et al. (2016); Manberg *et al.* (2019)	2
Kluyveromyces marxianus	Borgdorff *et al.* (2016); Chapter 3 WISH dataset	2
Candida glabrata	Afiuni-Zadeh *et al.* (2018); Manberg *et al.* (2019) (except adjusted HIV curated DB)	2
Zygosaccharomyces rouxii	Afiuni-Zadeh *et al.* (2018); Manberg *et al.* (2019) (except adjusted HIV curated DB and FGT curated DB search)	2
Cyberlindnera jadinii	Borgdorff *et al.* (2016); Manberg *et al.* (2019) using the HIV curated DB	2
Pseudozyma antarctica (Candida antartica)	Borgdorff *et al.* (2016); Manberg *et al.* (2019) using the HIV curated DB	2
Gibberella zeae	Afiuni-Zadeh *et al.* (2018); Manberg *et al.* (2019) (except adjusted HIV curated DB and FGT curated DB search)	2
Lichtheimia corymbifera	Manberg *et al.* (2019) (except adjusted HIV curated DB and HIV curated DB search); Chapter 3 WISH dataset	2
Candida orthopsilosis	Manberg *et al.* (2019) (except adjusted HIV curated DB and FGT curated DB search); Chapter 3 WISH dataset	2
Trichoderma gamsii	Manberg *et al.* (2019) (except adjusted HIV curated DB and FGT curated DB search); Chapter 3 WISH dataset	2

Discussion

This chapter evaluates publicly available metaproteome datasets to validate taxonomic and functional characterizations of fungi identified in chapter 3 and evaluate the performance of different protein databases on different datasets.

Database Selection

The Afiuni-Zadeh *et al.* (2018) dataset performed the best in terms of MS/MS identifications using the Metanovo concatenated to the fungal pan proteome method in comparison with the Chapter 2 WISH dataset and the remainder of the publicly available datasets, as we were able to characterize 1/5 of the data. However, using this method on the remainder of the datasets only managed to identify a maximum of 10.1% of MS/MS. In general, the identification rate of MS/MS spectra is usually low across searches conducted (*e.g.,* in most searches, only 10–30% of the spectra can be identified) (Yuan *et al.* 2014). The low number of MS/MS identifications is mostly due to the majority of microorganisms in the FGT not being cultured, identified, or characterized in the laboratory. The large number of MS/MS identifications for the Afiuni-Zadeh *et al.* (2018) dataset is most likely a result of this study conducting fractionation prior to submitting their samples to the mass spectrometer. In other datasets, the FGT curated DB performed better, most likely a result of it being more representative of the fungal species present in the vagina, as this database was created based on samples that were not enriched for human or bacterial analysis allowing for more fungal species to be identified, or the mass spectrometer used is more ideal for detecting fungi.

The vast heterogeneity and high dynamic range of microbial proteins present in any given sample may also interfere with the deep coverage of the metaproteome (Blackburn and Martens 2016). Advancements in metaproteomics software will be able to combat this issue. However, at present, the appropriate reference genomes may not exist to support metaproteomic data analyses (Blackburn and Martens 2016). The generation of more suitable metagenomes for protein identification will increase the amount of identified spectra significantly (Nesvizhskii 2014).

For the Månberg *et al.* (2019) dataset, we observed a higher number of protein groups and peptide assignments using the FGT curated database. This database provided a more representative microbiome composition than the HIV curated database, as it is usually not observed that fungi make up a greater relative abundance than bacteria. As a result, a lower number of fungal proteins were identified using the HIV curated database. The FGT curated database enabled the identification of *Candida*, which is expected in the FGT mycobiome. Furthermore, the potential function of fungal proteins identified seemed to corroborate the results shown in chapter 3 by showing a high relative abundance of Histone H4 acetylation

and membrane raft polarization. Peptide assignments across databases identified the majority of the same genera, including *Yarrowia*, *Penicillium*, *Kluyveromyces*, *Candida*, and *Debaryomyces*. Also, when comparing the three databases at the *KO* level, many of the same functions were conserved (*i.e.,* Protein processing in ER, Antigen processing, Estrogen signaling, Spliceosome, MAPK signaling, and Endocytosis). In conclusion, it appears that the major functions prevalent in these samples included transcription and translation (genetic information processing).

When mass tolerance parameters were adjusted to those specified in the original study of Månberg *et al.* (2019), we were able to increase MS/MS, protein group, and peptide identifications. The precursor mass tolerance parameter filters the set of candidate peptides that get compared with each experimental spectrum (Eng *et al.* 2011). As a result, the mass tolerance parameter acts as a strict peptide filter. Setting the peptide mass tolerance too narrow for a search will result in the correct peptide sequence being omitted from comparison to a tandem mass spectrum. Conversely, setting the mass tolerance to a value larger than deemed necessary may result in decreased search sensitivity because each putative peptide match must now compete against a larger pool of candidate peptides and may randomly result in a score higher than the correct peptide (Eng *et al.* 2011). It is important to note that these parameters were chosen to maximize the identification of human proteins, which explains why the relative abundance of Metazoan proteins was highest using this database. As mentioned in chapter 1, if human proteins are able to overpower analyses it results in the under-reporting of microbial proteins present, as it will be biased to the most prevalent species (Xiong *et al.* 2015). It could explain why this database showed uncharacteristic results, such as the absence of Candida protein groups.

From the results obtained using different databases, we can conclude that using the right parameters will improve the quality and representativeness of the results obtained. However, there is no single set of search parameters that is optimal for all types of analysis across all search engines. Different datasets, scientific questions, search methods, and search tools all play a role in defining what search parameters are optimal for analyses. Any parameter setting or search methodology can be optimal in one case and suboptimal in another (Eng *et al.* 2011).

For the Borgdorff *et al.* (2016) dataset, we observed a poor MS/MS identification rate using the Metanovo concatenated to Pan Proteome method. It is likely that Metanovo is not the appropriate tool for this dataset, or as mentioned above, we have used the incorrect

parameters to analyse the data. This may be due to samples being enriched in a specific way or the MS methods being less sensitive, especially since the purpose of this study was to examine human proteins. Future work may try implementing a different tool to examine this dataset or using a metagenome database. Since different search algorithms identify overlapping but non-identical sets of peptides and differences in peptide identification impact the quantification of proteins in samples (Degroeve *et al.* 2012), a search methodology can be optimal in one case and suboptimal in another (Eng *et al.* 2011).

From these analyses, we can conclude that database selection plays a critical role in the taxonomic and functional characterizations obtained and that a database that works for one dataset may not work as effectively for another dataset. Thus, choosing the appropriate database with suitable parameters is vital when performing microbiome analyses using metaproteomics.

Taxonomic composition of the vaginal mycobiome

Across all studies, the prevalence of fungi was found to range between 0.09% and 0.7%. As discussed in chapter 3, we know this is in the range that is expected for fungal microbiome analyses (Ma *et al.* 2020). However, we do see that mammalian protein groups overwhelm analyses (>97%). Phylum level assignments corroborate the trend seen in chapter 3 and previous studies confirmed the results of previous studies, which showed that Ascomycota and Basidiomycota dominate the vaginal mycobiome and most body sites (Ghannoum *et al.* 2010; Hoffmann *et al.* 2013; van Woerden *et al.* 2013; Zhang *et al.* 2011;). The original purpose of the Afiuni-Zadeh *et al.* (2018) study was to evaluate the use of residual PAP test samples for taxonomic and functional characterization of bacteria present in the FGT. Interestingly, their original analysis did manage to identify reads assigned to Ascomycota during taxonomic profiling.

Most previous mycobiome vagina studies have detected up to 20 different genera in their analysis, in contrast to gut mycobiome studies which have that the gut is home to >50 genera of fungi (Iliev *et al.* 2012; Hoffman *et al.* 2013). Our validation datasets were able to detect fungal genera in a range between 12 and 46. Our dataset fell into this range, as we were able to detect up to 39 different genera.

Across all public datasets, peptide sequences assigned to *Candida*, *Debaryomyces*, *Kluyveromyces*, *Malassezia*, *Penicillium*, *Yarrowia*, *Aspergillus*, *Fusarium*, *Cryptococcus*, and *Saccharomyces*. These results corroborate our chapter 3 results and indicate that these genera are most likely true inhabitants of the vagina. As mentioned in chapter 3, earlier FGT studies have shown that commonly detected fungi include *Candida*, *Saccharomyces*, *Aspergillus*, *Malassezia*, *Alternaria*, *Clavispora*, *Trichosporon*, and *Pichia* (Drell *et al*. 2013; Fornari *et al*. 2016; Godoy-Vitorino *et al.* 2018; Guo *et al*. 2012; Nejat *et al*. 2018; Zheng *et al.* 2013). In addition, *Penicillium*, *Fusarium,* and *Kluyveromyces* have also been found in the vagina (Fornari *et al.* 2016; Lehtoranta *et al*. 2021; Moraes *et al.* 2004; Nejat *et al.* 2018). The genus *Nakaseomyces* detected in our chapter 3 dataset and Afiuni-Zadeh *et al.* (2018) have been seen in a previous vaginal study by Godoy-Vitorino *et al.* (2018). Surprisingly, *Pichia* was not detected in the chapter 3 dataset but was found across all publicly available datasets.

From the validation datasets, we can likely confirm the following species identified in our study: *C. albicans*, *M. sympodialis*, *C. lusitaniae*, *T. marneffei*, *W. mellicola*, *C. neoformans*, and *Trichosporon asahii*. Furthermore, we were also able to detect novel species that were detected across the majority of datasets including *Cercospora beticola*, *Schizosaccharomyces cryophilus*, *Moesziomyces aphidis*, *Rhizsoclostmatium globosum*, and *Rhizopogon vinicolor*.

Using protein groups *Wallemia* was identified in all datasets except Borgdorff *et al*. (2016), this genus was identified previously in research by Scott *et al.* (2018), consequently, we present this genus as a commensal of the vaginal mycobiome.

In the Månberg et al. (2019) dataset, the FGT curated DB was the only database to show a high relative abundance of *Candida* as observed in chapter 3. In contrast to these results, Afiuni-Zadeh *et al.* (2016) and Borgdorff *et al.* (2016) showed a high relative abundance of *Cryptococcus* and *Saccharomyces*, respectively. The high relative abundance of *Saccharomyces* is probable as one vaginal study by Zheng *et al.* (2013) found this genus to be predominant. A likely hypothesis explaining the differences observed between datasets is that the total mass of fungal-derived proteins in any given sample may be insufficient for proteomic analysis and may be restricted by human-derived proteins that interfere with the analysis. A second hypothesis includes *Candida* not being the predominant fungal member of different populations. Evidence shows that microbiome makeup differs between geographical locations, ethnic groups, and age groups (Anahtar *et al.* 2018; Chu *et al.* 2018; Human Microbiome Project Consortium 2012; Lennard *et al.* 2019; Serrano *et al.* 2019). The

characteristics of each study cohort included in this analysis differed, for example, the Afiuni-Zadeh *et al.* (2018) dataset included a much older cohort at a different geographical location (America), Manberg *et al.* (2019) used a Kenyan cohort, and Borgdorff *et al.* (2016) used a Rwandan cohort. However, no record of ethnicity for each cohort could be found. A third hypothesis is that taxa detected depends on the type of sample collected. In this chapter, Borgdorff *et al.* (2016) collected CVL samples, Afiuni-Zadeh *et al.* (2018) used cells collected from the ectocervix of healthy women undergoing a routine PAP test, and Månberg *et al.* (2019) collected CVS. A drawback of FGT microbiome profiling is the difficulty in obtaining a sufficient amount of vaginal secretion via swabbing to allow accurate assessment of the mucosal microbiome. However, Schmidt *et al.* (2019) showed that the microbiome profile captured by CVL is comparable to that of vaginal swabs. A final theory is that due to differences in study methodologies in each original study, the fungal proteins identified differ (Nash *et al.* 2017). These differences may include storage buffer, sample processing (*e.g.,* centrifugation), lysis of cells or secretome only, MS/MS sensitivities, etc.

Candida species are the common fungal species detected in humans (Farr *et al.* 2015; Drell *et al.* 2013; Merenstein *et al.* 2013). In agreement with these studies, we detected the presence of *C. albicans*, *P. kudriavzevii* (*Candida krusei*), *Yarrowia lipolytica* (*Candida lipolytica*), *Clavispora lusitaniae* (*Candida lusitaniae),* and *C. parapsilosis* using protein group and peptide sequence identifications all found in previous FGT studies (Drell *et al.* 2013; Guo *et al*. 2012; Lehtoranta *et al.* 2021; Zheng *et al.* 2013). *P. kudriavzevii*, *Y. lipolytica*, *C. lusitaniae* are teleomorphs (*i.e.,* the sexually reproductive state) of their Candida counterparts. However, a caveat of taxonomic analysis is that peptide sequences are mapped to the lowest common ancestor, which results in the loss of information and limits its applicability at lower phylogenetic levels (such as the genus and species level) (Rechenberger *et al.* 2019). Moreover, clear taxonomic annotation is not always possible using metaproteomics since peptides can be shared between different proteins of the same organism, or even between multiple organisms (Kleiner *et al.* 2017). This may explain why *C. albicans* was not found at species level across all databases using peptide sequences even though this species is expected at a high prevalence in the vagina. This is especially apparent in the Borgdorff *et al*. (2016) dataset, where over 20% of fungal peptide sequences could not be assigned to the genus level.

In Chapter 3, we mentioned that some fungal species assignments such as *Rhizoclosmatium globosum* and *Endocarpon pusillum* have not previously been shown to inhabit the human

body. *R. globosum* was found in both the Borgdorff *et al.* (2016) and Månberg *et al.* (2019) datasets, whereas *E. pusillum* was only identified in the Borgdorff *et al.* (2016) dataset. It is probable that the Borgdorff et al. (2016) assignment may be a false positive as we used the same protein database on our chapter 3 dataset. Therefore, more research is still required to confirm such fungal species.

Several fungal genera identified in chapter 3 were not detected in public datasets. A few reasons may explain this result, maybe a more stringent quality control is needed, or other datasets did contain these genera, but peptide sequences were not assigned to genus level. Another hypothesis was specified in chapter 3, which states that the presence of fungi not commonly found in the human body may be a result of vaginal product use, hygiene habits, vaginal insertions, and water purity influencing the taxa that are detected in the FGT (Humphries *et al.* 2019). This is probable since the chapter 3 dataset is the only South African dataset as other datasets included Kenyan, Rwandan, and American participants. Another reason may directly be a result of the mass spectrometry analysis, such as the sensitivity of the mass spectrometer or the run time.

To recap chapter 3*, Torulaspora* and *Condiobolus* have been previously identified in the vagina (Fornari *et al.* 2016; Subramanian and Sobel 2011). The genus *Cutaneotrichosporon* is a rarely isolated yeast from clinical samples (Nath *et al.* 2017). Interestingly, the genus *Diaporthe* not identified using protein groups was found to inhabit the vaginas of giant pandas (Ma *et al.* 2017). *Rhizopus* has been detected in the gut (Cheng *et al.* 2009) and could likely be a transient fungus. *Phialophora* is a dematiaceous fungus associated with several human skin diseases, such as chromoblastomycosis, phaeohyphomycosis, and mycetoma (Turiansky *et al.* 1995; Tong *et al.* 2013). In addition, *Sporothrix also* affects the skin but this genus is characteristically found in soil and living plant material (Flournoy *et al.* 2000). *Exidia* has been found in the air and is likely an environmental contaminant (Els *et al.* 2019; Table 3.1). *Wickerhamomyces anomalus* can be isolated from different habitats such as soil, plants, flowers, fruit peels, food, dairy products, wastewater, human tissues, and marine environments (Walker 2011). *Rhinocladiella* is a yeast-like fungus, that can be associated with skin infections (Badali *et al.* 2010). These genera may be either contaminants from the skin or environment, or as mentioned above from vaginal insertions (Lehtoranta *et al.* 2021; Gafos *et al.* 2010).

Functional capacity of the mycobiome

As seen in chapter 3, both HSPs and Actin were relatively abundant proteins across all studies, which likely confirms that these proteins may have an essential role in the FGT. We have previously discussed in chapter 3 that HSPs aid in cell survival under adverse environmental conditions, like microbial pathogen infection or pathogen-induced inflammation (Dasari *et al.* 2016; Sisti *et al.* 2015; Ventolini 2013). The appearance of these stress proteins may once again indicate a perturbation in the vaginal environment since HSP70 is a sensitive indicator of stress. HSP70 synthesis is induced when a cell encounters physiological stresses such as those mentioned above (Craig 1985). In support of this, Borgdorff *et al.* (2016) and Manberg *et al.* (2019) specimens were collected from individuals with a wide range of vaginal diseases, including HIV, BV, and STIs. Several other indicators provide evidence that the vagina must be under stress. These include the finding of ER chaperone BiP proteins in the Månberg *et al.* (2019) dataset as BiP proteins are also members of the hsp70 family.

As mentioned in chapter 3, Actin is one of the core components of the eukaryotic cytoskeleton. Actin participates in cellular processes, including cell motility, cell division, cytokinesis, signal transduction, and the establishment and maintenance of cell junctions in fungi (Dominguez and Holmes 2011). Actin is crucial to the growth of hyphae, which is required for the morphogenesis of yeast species such as *C. albicans* (Lichius *et al.* 2011), which may explain why it was prevalent in all datasets.

The presence of glyceraldehyde-3-phosphate dehydrogenase (G3P dehydrogenase) proteins indicates that glycolysis is likely taking place, as this enzyme is known to be involved in the sixth step of glycolysis. From chapter 3, we suspect that fungi likely adapt their nutrient source for survival during infection. Supporting evidence includes the identification of three unique peptides from G3P dehydrogenase in the original analysis by Afiuni-Zadeh *et al.* (2018). In addition, over 33% of the top 20 protein families were associated with glycolysis and lactic acid production. These findings suggest that the fungi present in these women were metabolically active.

Across all vaginal studies analysed, the majority of fungal proteins were involved in genetic information processing and cellular processes. Genetic information processing includes transcription, translation, folding, sorting and degradation, and replication and repair.

Furthermore, the most prevalent BPs were involved in transcription and translation (maturation of SSU-rRNA from tricistronic rRNA transcript, ribosomal small subunit assembly, and histone H4 acetylation). When using metaproteomics, the more relatively abundant proteins are usually detected; commonly observed proteins are involved in translation, energy production, and carbohydrate metabolism (Rooijers *et al*. 2011; Verberkmoes *et al.* 2009). Histone exchange features prominently during the process of transcription initiation and elongation. The transcriptional activity of a gene is dependent on the local composition and organization of its chromatin environment (Ehrenhofer-Murray 2004).

It is important to note that the most relatively abundant functional terms will change depending on which fungal species are most abundant, which explains why there is a high abundance of different functional terms using public datasets in comparison to chapter 3 results. The majority of fungal species identified using public datasets did not follow the same relative abundance trend as those identified in chapter 3, with the exception of the FGT curated DB that was used to search the Månberg *et al.* (2019) dataset.

Conclusion

In conclusion, these studies demonstrate that using metaproteomics to study the mycobiome can be an effective method when the right database and search parameters are used, despite the majority of proteins being of human origin (Afiuni-Zadeh *et al.* 2018). From analyses using public studies, we could confirm the presence of the following fungal genera identified in chapter 3 analyses of the mycobiome: *Candida*, *Debaryomyces*, *Kluyveromyces*, *Malassezia*, *Yarrowia*, *Fusarium*, *Aspergillus*, *Penicillium*, *Trametes,* and *Saccharomyces*. A genus not found in our study but found across other datasets included *Pichia.*

We were unable to definitively confirm the presence of novel fungi (*e.g., Rhizopogon* and *Endocarpon*) identified in chapter 3, therefore, more work is required to confirm these findings using the suggestions mentioned in chapter 3, *viz.* next generation sequencing (NGS) and development of fractionation methods, and better controls including environmental swabs and swabs from surrounding body sites. Furthermore, there is a need to isolate and sequence more fungi to add to the existing databases.

From protein identifications, we suggest that fungal communities are most likely responding to stress and/or pathogens in the vagina by expressing the relevant proteins (*e.g.,* heat shock proteins). Furthermore, across all datasets, fungal proteins were identified that function in genetic information processing, such as transcription and translation, and carbohydrate metabolism (*e.g.,* glycolysis).

This work lays the groundwork for vaginal mycobiome research and will become even more significant as the sequences of new fungal genomes, become available and we are able to identify more mass spectra (Afiuni-Zadeh *et al.* 2018). Furthermore, using public data where disease status is known could provide additional insight into the role of fungi during disease.

Supplementary Tables

Table S1. Fungal genera that had a relative abundance than 1% across each protein sequence database used to analyse the Månberg *et al.* (2018).

Genera	HIV curated DB	Adjusted HIV curated DB	FGT curated DB
Trichoderma	0.80	0	0
Aspergillus	0.77	0	0
Fusarium	0.67	0	0
Cryptococcus	0.29	0	0
Yarrowia	0.14	0.93	0
Trichosporon	0.07	0	0
Alternaria	0.03	0	0.18
Penicillium	0.003	0	0.11
Cercospora	0	0	0.07
Talaromyces	0	0.10	0.055
Lichtheimia	0	0	0.030

Table S2. Fungal Biological Processes that had a relative abundance of less than 1% across each protein sequence database used to analyse the Månberg *et al.* (2019) dataset.

Biological Process	HIV curated DB	Adjusted HIV curated DB	FGT curated DB
glucose metabolic process	0.82	0	0
glycolytic process	0.82	0	0
adhesion of symbiont to host	0.78	0	0

biological process involved in interaction with host	0.78	0	0
cell-matrix adhesion	0.78	0	0
induction by symbiont of host defense response	0.78	0	0
pentose-phosphate shunt	0.62	0	0
negative regulation of DNA replication origin binding	0.13	0	0.48
regulation of meiotic nuclear division	0.13	0	0.48
cytoplasmic translational initiation	0.13	0	0
isocitrate metabolic process	0.12	0	0
tricarboxylic acid cycle	0.12	0	0
antimicrobial humoral response	0.08	0	0
cellular response to unfolded protein	0.08	0	0
chaperone cofactor-dependent protein refolding	0.08	0	0
peptide transport	0.08	0	0
protein refolding	0.08	0	0
response to toxic substance	0.08	0	0
SRP-dependent cotranslational protein targeting to membrane, translocation	0.08	0	0
ATP metabolic process	0.08	0	0
cellular protein complex disassembly	0.08	0	0
cytoplasm protein quality control by the ubiquitin-proteasome system	0.08	0	0
DNA replication termination	0.08	0	0
endoplasmic reticulum membrane fusion	0.08	0	0
ER-associated misfolded protein catabolic process	0.08	0	0
mitochondria-associated ubiquitin-dependent protein catabolic process	0.08	0	0
mitotic spindle disassembly	0.08	0	0
negative regulation of telomerase activity	0.08	0	0
nonfunctional rRNA decay	0.08	0	0
nuclear protein quality control by the ubiquitin-proteasome system	0.08	0	0
piecemeal microautophagy of the nucleus	0.08	0	0
positive regulation of histone H2B ubiquitination	0.08	0	0
positive regulation of mitochondrial fusion	0.08	0	0
positive regulation of protein localization to nucleus	0.08	0	0
protein transport to vacuole involved in ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway	0.08	0	0
retrograde protein transport, ER to cytosol	0.08	0	0
ribophagy	0.08	0	0
ribosome-associated ubiquitin-dependent protein catabolic process	0.08	0	0

SCF complex disassembly in response to cadmium stress	0.08	0	0
sister chromatid biorientation	0.08	0	0
stress-induced homeostatically regulated protein degradation pathway	0.08	0	0
ubiquitin-dependent ERAD pathway	0.08	0	0
microtubule cytoskeleton organization	0.04	0	0
mitotic cell cycle	0.04	0	0
nuclear division	0.04	0	0
nuclear migration by microtubule mediated pushing forces	0.04	0	0
regulation of transcription, DNA-templated	0	0	0.03

Table S3. Fungal KEGG functions that had a relative abundance of less than 1% across each protein sequence database used to analyse the Månberg *et al.* 2019 dataset.

KEGG Functional Term	HIV Curated DB	Adjusted HIV Curated DB	FGT Curated DB
Glutathione metabolism	0.94	0.95	0.09
Cell cycle	0.55	0	0
Oocyte meiosis	0.55	0	0
Phospholipase D signaling pathway	0.21	0	0
2-Oxocarboxylic acid metabolism	0.15	0	0
Citrate cycle (TCA cycle)	0.15	0	0
Peroxisome	0.15	0	0
Protein export	0.08	0	0
Oxidative Phosphorylation	0.05	0	0
Biosynthesis of secondary metabolites	0	0	0.09
Cellular senescence	0	0	0.09

Table S4. Fungal genera that had a relative abundance less than 1% in Afiuni-Zadeh *et al.* (2018) dataset identified using the PAP curated database.

Genera	Relative Abundance (%)
Nakaseomyces	0.82
Yarrowia	0.67
Wallemia	0.48
Monosporascus	0.30
Auriculariopsis	0.29
Trichosporon	0,27
Fusarium	0.23
Blastomyces	0.23
Aspergillus	0.22

Talaromyces	0.13
Clathrospora	0.13
Clavispora	0.05
Moesziomyces	0.03
Sugiyamaella	0.03

Table S5. Fungal Biological Processes according to gene ontology terms that had a relative abundance of less than 1% in Afiuni-Zadeh *et al.* (2018) dataset using the PAP curated database.

Biological Process	Relative Abundance (%)
translation	0.30
ubiquitin-dependent protein catabolic process	0.15
nucleocytoplasmic transport	0.14
protein transport	0.14
intracellular protein transport	0.11
vesicle-mediated transport	0.11
adenylate cyclase-activating glucose-activated G protein-coupled receptor signaling pathway	0.10
chaperone-mediated protein folding	0.10
heterochromatin assembly by small RNA	0.10
proteasome-mediated ubiquitin-dependent protein catabolic process	0.07
retrograde transport, endosome to Golgi	0.05
methionine metabolic process	0.04
one-carbon metabolic process	0.04
S-adenosylmethionine biosynthetic process	0.04
nitrogen compound metabolic process	0.04
maintenance of translational fidelity	0.009
ribosomal large subunit assembly	0.009
rRNA processing	0.009
D-gluconate metabolic process	0.006
pentose-phosphate shunt	0.006
endocytosis	0.006
mitotic actomyosin contractile ring assembly	0.006
regulation of nucleosome density	0.006
mRNA splicing, via spliceosome	0.0002

Table S6. Fungal KEGG functional terms that had a relative abundance of less than 1% in Afiuni-Zadeh *et al.* (2018) dataset using the PAP curated database.

KEGG functional term	Relative Abundance (%)
RNA transport	0.92
Protein processing in endoplasmic reticulum	0.91

Ribosome biogenesis in eukaryotes	0.85
Fc gamma R-mediated phagocytosis	0.79
Phospholipase D signaling pathway	0.79
Ras signaling pathway	0.79
Metabolic pathways	0.53
Biosynthesis of cofactors	0.49
AMPK signaling pathway	0.4
Autophagy	0.4
Cell cycle	0.4
Hippo signaling pathway	0.4
Meiosis - yeast	0.4
mRNA surveillance pathway	0.4
Oocyte meiosis	0.4
Sphingolipid signaling pathway	0.4
TGF-beta signaling pathway	0.4
Tight junction	0.4
Biosynthesis of secondary metabolites	0.3
Spliceosome	0.29
MAPK signaling pathway	0.29
Biosynthesis of amino acids	0.27
Cysteine and methionine metabolism	0.27
Alanine, aspartate and glutamate metabolism	0.23
Pyrimidine metabolism	0.23
Carbon metabolism	0.04
Glutathione metabolism	0.04
Microbial metabolism in diverse environments	0.04
Pentose phosphate pathway	0.04

Table S7. Fungal genera that had a relative abundance less than 1% in Borgdorff *et al.* (2016) dataset identified using the FGT curated database.

Genera	Relative Abundance (%)
Kluyveromyces	0.78
Candida	0.75
Cyberlindnera	0.74
Malassezia	0.54
Yarrowia	0.33
Rhizoclosmatium	0.27
Cercospora	0.24
Moesziomyces	0.075
Pichia	0.06

Schizosaccharomyces	0.05
Pseudogymnoascus	0.04

Table S8. Fungal biological processes according to gene ontology terms that had a relative abundance of less than 1% in Borgdorff *et al.* (2016) dataset using the FGT curated database.

Biological Process	Relative Abundance (%)
ethanol metabolic process	0.86
histone H4 acetylation	0.19
membrane raft polarization	0.19
cellular protein localization	0.088
establishment or maintenance of actin cytoskeleton polarity	0.088
proteasome-mediated ubiquitin-dependent protein catabolic process	0.053
protein targeting to membrane	0.034
nucleocytoplasmic transport	0.033
protein transport	0.033
ubiquitin-dependent protein catabolic process	0.020
endocytosis	0.019
histone exchange	0.019
regulation of nucleosome density	0.019
fatty acid biosynthetic process	0.014
secondary metabolic process	0.014

Table S9. Fungal KEGG functional terms that had a relative abundance of less than 1% in Borgdorff *et al.* (2016) dataset using the FGT curated database.

KEGG Ontology Term	Relative Abundance (%)
Carbon metabolism	0.18
Pentose phosphate pathway	0.18
Glutathione metabolism	0.18
Proteasome	0.044
Endocytosis	0.032
Spliceosome	0.026
Protein processing in endoplasmic reticulum	0.026
MAPK signaling pathway	0.026
Antigen processing and presentation	0.026
Estrogen signaling pathway	0.026
Longevity regulating pathway - multiple species	0.026
RNA transport	0.022

Ribosome biogenesis in eukaryotes	0.022
Phagosome	0.0062
Autophagy - yeast	0.0062

Chapter 6

Conclusion

The female genital tract (FGT) microbiome is a complex ecosystem constituted by assemblages of bacteria, fungi, archaea, and viruses. However, microbiome studies have focused on the most abundant kingdom that is bacteria, and their role in vaginal health and disease (Ackerman and Underhill 2017; Nash *et al.* 2017). Consequently, the extent of fungal diversity is largely unknown. This knowledge deficit is surprising given the importance of fungi as pathogens in vaginal diseases such as vulvovaginal candidiasis (Fidel 2004; Hurley and De Louvois 1979).

To contribute to the existing knowledge of the vaginal mycobiome, we investigated the fungal communities of women residing in Cape Town, South Africa, using liquid chromatography-tandem mass spectrometry. However, as few studies have utilized metaproteomics to study the mycobiome, many of the challenges that hinder this approach have not been addressed. Due to this, we set out to optimize the Proteomics software pipeline, Metanovo, to more accurately analyse fungal microbiome data. We concluded that achieving quality fungal assignments depends on having a comprehensive, well-annotated sequence database that includes fungal sequences known to occupy the microbiome of interest. This optimization allowed us to present the first large metaproteome dataset to define the taxonomic composition and potential functional processes of the vaginal mycobiome in a cohort of women at high risk of vaginal diseases. In the future, to further increase the number of identified mass spectra, we suggest the use of a metagenome database rather than a UniProt-based database that will allow us to further examine more fungal taxa present in samples.

From our analyses, fungi constituted 0.4% of the vaginal microbiome. As seen in other mycobiome studies, Ascomycota and Basidiomycota dominated the vaginal mycobiome of this cohort. From our genus assignments, we were able to detect commonly identified vaginal fungal genera, such as *Candida*, *Aspergillus*, *Alternaria*, *Clavispora*, *Trichosporon*, and *Malassezia* (Drell *et al*. 2013; Zheng *et al*. 2013; Guo *et al*. 2012; Zheng *et al.* 2013; Godoy-Vitorino *et al.* 2018; Fornari *et al*. 2016; Nejat *et al*. 2018), of which *Candida* was the most relatively abundant genus across optimal and non-optimal states. Other less commonly detected genera included *Cryptococcus*, *Penicillium*, *Torulaspora*, and *Lodderomyces* (Lehtoranta *et al.* 2021; El-din *et al*. 2009; Moraes *et al*. 2004; Fornari *et al.* 2016). Notably,

this study also detected unique fungal genera, such as *Teratoramularia*, *Endocarpon*, and *Rhizoclosmatium,* that have not been detected in other human mycobiome studies. However, unique identifications require further analysis to validate their presence as part of the FGT mycobiome to ensure they are not contaminants. In this study, at the time of sample collection, we were unable to collect controls to account for environmental contaminants (such as air samples from clinics and laboratories, and skin samples from clinicians and laboratory technicians), but this may prove advantageous for future studies.

This study collected samples at two visits, but for vaginal community composition to be accurately estimated, it is more beneficial to use a longitudinal study design in which a large cohort of women is frequently sampled. Longitudinal studies expose the limitations of cross-sectional studies as significant fluctuations in abundance and composition of microbial communities are overlooked if samples are infrequently collected. Therefore, the above is important to consider when conducting future clinical metaproteomic research (Rechenberger *et al.* 2019).

From functional profiling, the main functional categories represented in FGT fungal communities included carbohydrate metabolism, signal transduction, energy metabolism, genetic information processing, cellular processes, and folding, sorting, and degradation. Furthermore, from the fungal proteins identified (*e.g.*, heat shock proteins), we infer that fungal communities respond to stress in the vagina and may use functions such as carbohydrate metabolism (*e.g.*, glycolysis) to contribute to successful host colonization and pathogenesis.

Metaproteomics is a beneficial tool to compare microbial protein expression in response to different conditions (Salvato *et al.* 2021). Differential changes in fungal abundance and functional processes were observed between optimal and non-optimal states, including BV and *Mycoplasma genitalium.* These differences provided essential knowledge of the changes involved in the mycobiome when the vagina is in a diseased state. Furthermore, by characterizing fungal species differentially abundant between optimal and non-optimal states, we could identify fungal species which specifically change abundance during the BV diseased state, namely *Candida albicans*, *Malassezia sympodialis*, and *Candida thasanensis*.

We also determined which clinical factors were associated with significant differences in fungal community composition. Our results revealed that Nugent score, pro-inflammatory cytokines,

chemokines, vaginal pH, *C. trachomatis*, and the presence of clue cells were among these factors. Nugent score as a major correlate allows us to hypothesize that BV affects fungal community structure, or fungal communities affect BV. Interestingly, very few significant associations between HC methods and the mycobiome were identified; however, larger sample sizes are required to further evaluate this, as oral contraceptives have been shown to increase the risk of candidiasis (Linhares *et al.* 2001; Spinillo *et al*. 1995; van de Wijgert *et al*. 2013). Our examination of the mycobiome provided a more comprehensive understanding of the association between fungal communities and clinical factors. The influence of many clinical factors on fungal communities is expected as the microbiome is a complex environment and many factors are able to drive change and dysbiosis. In the future, it might be beneficial to examine other environmental factors, such as hygiene and geographical location, which we also suspect may contribute to vaginal community composition (Humphries *et al.* 2019; Lennard *et al.* 2019; Noyes *et al.* 2018).

In terms of fungal-bacterial relationships, there appears to be an inverse correlation between BV and vaginal yeast identified using metaproteomics. Similarly, the majority of fungal species were negatively correlated with BV-associated bacteria, and many fungal species were underabundant during the BV positive state. Additionally, Nugent score was negatively associated with total fungal intensity. The bacterial STI, *C. trachomatis* also associated with fungal community relative abundance and demonstrated a negative correlation with total fungal intensity. Inverse correlations between fungal diversity and bacterial diseases (*i.e.,* BV and *C. trachomatis*) are likely to occur because the microenvironment probably becomes less suitable for fungal growth as bacterial diversity increases and fungi likely compete for resources (Underhill and Iliev 2014).

This study illustrates on a small scale the complex relationship between the mycobiome and the bacteriome in the FGT. Fungal communities likely affect and/or are affected by bacterial communities, their metabolites, and host responses. Future work may include using host and fungal protein expression data to test for correlations between the host and mycobiome, as this information is lacking. These correlations will help infer host-mycobiome interactions that possibly underlie vaginal diseases (Salvato *et al.* 2021). In addition, future work may include utilizing other advantageous methods such as the use of NGS and developing fractionation methods to increase the representation of samples prior to mass spectrometry analysis, by increasing the coverage of the metaproteome, fungal protein identifications will increase (Mostovenko *et al.* 2013).

Some clinical factors, such as BMI, HC methods (Nur isterate, DMPA, and Implanon), and STIs (*C. trachomatis*, *T. vaginalis*, *M. genitalium*, *N. gonorrhoeae*, and HSV-2), BV, and vaginal pH were shown to affect vaginal inflammation. Additionally, BV and vaginal pH were identified as confounding factors. The above, and the lack of clustering between inflammation categories based on fungal communities, suggest that vaginal inflammation does not influence fungal species or vice versa.

The use of publicly available vaginal proteome studies allowed us to confirm the presence of previously identified mycobiome members such as *Candida*, *Debaryomyces*, *Kluyveromyces*, *Malassezia*, *Penicillium*, *Yarrowia*, *Aspergillus*, *Fusarium*, *Cryptococcus*, *Wallemia*, *Trichosporon*, and *Saccharomyces* (Scott *et al.* 2018). The validation analyses conducted in Chapter 6 also detected unique fungal assignments, *Rhizoclosmatium globosum,* and *Endocarpon pusillum,* providing more evidence that these species form part of the vaginal mycobiome.

We can conclude that metaproteomics is a suitable method to analyze the mycobiome as it is able to yield detailed information on the fungal composition, abundance, and likely function. However, it was not possible to infer definitive insights on the functionality of these microbiomes. Moreover, we were able to examine the association between fungi and bacteria. Consequently, this research provided a basis for understanding the vaginal mycobiome and shows a clear association between BV and fungal communities.

References

Abia-Bassey, L. N., and S. J. Utsalo, 2006 Yeast associated with human infections in south-eastern Nigeria. Mycoses 49: 510–515.

Abram, F., A. M. Enright, J. O'Reilly, C. H. Botting, G. Collins *et al*., 2011 A metaproteomic approach gives functional insights into anaerobic digestion. J Appl. Microbiol. 110: 1550–1560.

Abu-Elteen, K. H., A. M. Abdul Malek, and N. A. Abdul Wahid, 1997 Prevalence and susceptibility of vaginal yeast isolates in Jordan. Mycoses 40: 179–185.

Ackerman, A. L., and D. M. Underhill, 2017 The mycobiome of the human urinary tract: potential roles for fungi in urology. Ann. Transl. Med. 5: 31–31.

Adjapong, G., M. Hale, and A. Garrill, 2016 Population Structure of Candida albicans from Three Teaching Hospitals in Ghana. Med. Mycol. 54: 197–206.

Afiuni-Zadeh, S., K. L. M. Boylan, P. D. Jagtap, T. J. Griffin, J. D. Rudney *et al*., 2018 Evaluating the potential of residual PAP test fluid as a resource for the metaproteomic analysis of the cervical-vaginal microbiome. Sci. Rep. 8: 10868.

Aguiar-Pulido, V., W. Huang, V. Suarez-Ulloa, T. Cickovski, K. Mathee *et al*., 2016 Approaches for Microbiome Analysis. 12: 5–16.

Albertin, W., L. Chasseriaud, G. Comte, A. Panfili, A. Delcamp *et al*., 2014 Winemaking and Bioprocesses Strongly Shaped the Genetic Diversity of the Ubiquitous Yeast Torulaspora delbrueckii. PLoS One 9: e94246.

Aldunate, M., D. Srbinovski, A. C. Hearps, C. F. Latham, P. A. Ramsland *et al*., 2015 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. 6: 164.Alisoltani, A., M. T. Manhanzva, M. Potgieter, C. Balle, L. Bell *et al*., 2020 Microbial function and genital inflammation in young South African women at high risk of HIV infection. Microbiome 8: 165.

Amatya, R., B. Khanal, and A. Rijal, 2010 Syncephalastrum species producing mycetoma-like lesions. Indian J. Dermatol. Venereol Leprol. 76:284-286.

Amsel, R., P. A. Totten, C. A. Spiegel, K. C. Chen, D. Eschenbach *et al*., 1983 Nonspecific vaginitis. Diagnostic criteria and microbial and epidemiologic associations. Am. J. Med. 74:14–22.

Anahtar, M. N., E. H. Byrne, K. E. Doherty, B. A. Bowman, S. Yamamoto *et al*., 2016 Inflammatory Responses in the Female Genital Tract. Immunity 42: 965–976.

Anahtar, M. N., D. B. Gootenberg, C. M. Mitchell, and D. S. Kwon, 2018 Cervicovaginal Microbiota and Reproductive Health: The Virtue of Simplicity. Cell Host Microbe 23: 159–168.

Antonovics, J., J. L. Abbate, C. H. Baker, D. Daley, M. E. Hood *et al*., 2007 Evolution by any other name: antibiotic resistance and avoidance of the E-word. PLoS Biol. 5: e30–e30.

Araujo, R., 2014 Towards the Genotyping of Fungi: Methods, Benefits and Challenges. Curr. Fungal Infect. Rep. 8: 203–210.

Arechavala, A. I., M. H. Bianchi, A. M. Robles, G. Santiso, and R. Negroni, 2007 Identificación y sensibilidad frente a fluconazol y albaconazol de 100 cepas de levaduras aisladas de flujo vaginal. Rev. Iberoam. Micol. 24: 305–308.

Arendrup, M. C., T. Boekhout, M. Akova, J. F. Meis, O. A. Cornely *et al*., 2014 ESCMID† and ECMM‡ joint clinical guidelines for the diagnosis and management of rare invasive yeast infections. Clin. Microbiol. Infect. 20: 76–98.

Aroutcheva, A., D. Gariti, M. Simon, S. Shott, J. Faro *et al*., 2001 Defense factors of vaginal lactobacilli. Am. J. Obstet. Gynecol. 185: 375–379.

Arumugam M., J. Raes, E. Pelletier, D. Le Paslier, T. Yamada *et al*., 2011 Enterotypes of the human gut microbiome. Nature 473: 174.

Ashburner, M., C. Ball, J. Blake, D. Botsein, H. Butler *et al*., 2000 Gene ontology: Tool for the unification of biology. Nature Genet. 25: 25–29.

Austin, M. N., L. K. Rabe, S. Srinivasan, D. N. Fredricks, H. C. Wiesenfeld *et al*., 2015 Mageeibacillus indolicus gen. nov., sp. nov.: a novel bacterium isolated from the female genital tract. Anaerobe. 32:37-42.

Badali, H., A. Bonifaz, T. Barrón-Tapia, D. Vázquez-González, L. Estrada-Aguilar *et al*., 2010 Rhinocladiella aquaspersa, proven agent of verrucous skin infection and a novel type of chromoblastomycosis. Med. Mycol. 48: 696–703.

Balajee, S. A., M. D. Lindsley, N. Iqbal, J. Ito, P. G. Pappas *et al*., 2007 Nonsporulating clinical isolate identified as Petromyces alliaceus (anamorph Aspergillus alliaceus) by morphological and sequence-based methods. J. Clin. Microbiol. 45: 2701–2703.

Balle, C., I. N. Konstantinus, S. Z. Jaumdally, E. Havyarimana, K. Lennard *et al.*, 2020 Hormonal contraception alters vaginal microbiota and cytokines in South African adolescents in a randomized trial. Nat. Commun. 11: 5578.

Barcoto, M. O., F. Pedrosa, O. C. Bueno, and A. Rodrigues, 2017 Pathogenic nature of Syncephalastrum in Atta sexdens rubropilosa fungus gardens. Pest Manag. Sci. 73: 999–1009.

Barnabas, S. L., S. Dabee, J-A. S. Passmore, H. B. Jaspan, D. A. Lewis *et al.*, 2017 Converging epidemics of sexually transmitted infections and bacterial vaginosis in southern African female adolescents at risk of HIV. Int. J. STD AIDS 29: 531–539.

Barousse, M., B. Van Der Pol, D. Fortenberry, D. Orr, and P. L. Fidel Jr, 2004 Vaginal yeast colonisation, relative abundance of vaginitis, and associated local immunity in adolescents. Sex. Transm. Infect. 80: 48–53.

Barizzi, J., E. Merlo, P. Grassi, B. Togni, V. Bruderer *et al.*, 2016 Vaginal colonisation by Mucor circinelloides . Case report with cytopathology, molecular sequencing and epidemiology. Cytopathology 27: n/a-n/a.

Basson, R., 2000 The female sexual response: a different model. J. Sex. Marital. Ther. 26:51-65.

Beauvais, A., C. Schmidt, S. Guadagnini, P. Roux, E. Perret *et al*., 2007 An extracellular matrix glues together the aerial-grown hyphae of Aspergillus fumigatus. Cell. Microbiol. 9: 1588–1600.Beed, M., R. Sherman, and S. Holden, 2014 Fungal infections and critically ill adults. Contin. Educ. Anaesth. Crit. Care Pain 14: 262–267.

Beigi, R. H., L. A. Meyn, D. M Moore, M. A. Krohn, and S. L. Hillier, 2004 Vaginal yeast colonization in nonpregnant women: a longitudinal study. Obstet. Gynecol. 104: 926–930.

Belnap, J., B. Büdel, and O. Lange, 2001 Biological Soil Crusts: Characteristics and Distribution. In: Biological Soil Crusts: Structure, Function and Management Ecological Studies, pp. 3–30.

Benito, Á., F. Calderón, and S. Benito 2018 Schizosaccharomyces pombe Biotechnological Applications in Winemaking. Methods Mol. Biol. 1721: 217–226.

Benny, G. L., and M. Blackwell, 2004 Lobosporangium, a new name for Echinosporangium Malloch, and Gamsiella, a new genus for Mortierella multidivaricata. Mycologia 96: 143–149.

Bentubo, H. D. L., A. Mantovani, J. T. Yamashita, W. Gambale, and O. Fischman, 2015 Yeasts of the genital region of patients attending the dermatology service at Hospital São Paulo, Brazil. Rev. Iberoam. Micol. 32: 229–234.

Bergeron, A. C., B. G. Seman, J. H. Hammond, L. S. Archambault, D. A. Hogan *et al.*, 2017 Candida albicans and Pseudomonas aeruginosa Interact To Enhance Virulence of Mucosal Infection in Transparent Zebrafish. Infect. Immun. 85: e00475-17.

Biagi, E., B. Vitali, C. Pugliese, M. Candela, G. G. Donders *et al.*, 2009 Quantitative variations in the vaginal bacterial population associated with asymptomatic infections: a real-time polymerase chain reaction study. Eur. J. Clin. Microbiol. Infect. Dis. 28:281–285.

Bianco, L., and G. Perrotta, 2015 Methodologies and perspectives of proteomics applied to filamentous fungi: from sample preparation to secretome analysis. Int. J. Mol. Sci. 16: 5803–5829.

Bingham, J. S., 1999 What to do with the patient with recurrent vulvovaginal candidiasis. Sex. Transm. Infect. 75: 225–227.

Blackburn, J. M., and L. Martens, 2016 The challenge of metaproteomic analysis in human samples. Expert Rev. Proteomics 13: 135–138.

Boekhout, T., 2011 Pseudozyma Bandoni emend. Boekhout (1985) and a comparison with the yeast state of Ustilago maydis (De Candolle) Corda (1842), pp. 1857–1868 in: *The yeasts, a taxonomic study*, edited by C. P. Kurtzman, J. W. Fell, and T. Boekhout. Elsevier, London.

Borgdorff, H., E. Tsivtsivadze, R. Verhelst, M. Marzorati, S. Jurriaans *et al*., 2014 Lactobacillus-dominated cervicovaginal microbiota associated with reduced HIV/STI prevalence and genital HIV viral load in african women. ISME J. 8: 1781–1793.

Borgdorff, H., R. Gautam, S. D. Armstrong, D. Xia, G. F. Ndayisaba *et al.*, 2016 Cervicovaginal microbiome dysbiosis is associated with proteome changes related to alterations of the cervicovaginal mucosal barrier. Mucosal Immunol. 9: 621–633.

Borman, A. M., C. J. Linton, D. Oliver, M. D. Palmer, A. Szekely *et al.*, 2009 Pyrosequencing analysis of 20 nucleotides of internal transcribed spacer 2 discriminates Candida parapsilosis, Candida metapsilosis, and Candida orthopsilosis. J. Clin. Microbiol. 47: 2307–2310.

Bostanci, N., M. Grant, K. Bao, A. Silbereisen, F. Hetrodt *et al.*, 2021 Metaproteome and metabolome of oral microbial communities. Periodontol. 2000 85: 46–81.

Bracharz, F., T. Beukhout, N. Mehlmer, and T. Brück, 2017 Opportunities and challenges in the development of Cutaneotrichosporon oleaginosus ATCC 20509 as a new cell factory for custom tailored microbial oils. Microb. Cell Fact. 16: 178.

Bradford, L. L., and J. Ravel, 2017 The vaginal mycobiome: A contemporary perspective on fungi in women's health and diseases. Virulence 8: 342–351.

Bradshaw, C. S., L. A. Vodstrcil, J. S. Hocking, M. Law, M. Pirotta *et al.*, 2013 Recurrence of bacterial vaginosis is significantly associated with posttreatment sexual activities and hormonal contraceptive use. Clin. Infect. Dis. 56: 777–86.

Brinkmann, V., U. Reichard, C. Goosmann, B. Fauler, Y. Uhlemann *et al.*, 2004 Neutrophil extracellular traps kill bacteria. Science 303: 1532–1535.

Brock, M., 2009 Fungal metabolism in host niches. Curr. Opin. Microbiol. 12: 371-376.

Brooks, B., R. S. Mueller, J. C. Young, M. J. Morowitz, R. L. Hettich *et al.*, 2015 Strain-resolved microbial community proteomics reveals simultaneous aerobic and anaerobic function during gastrointestinal tract colonization of a preterm infant. Front. Microbiol. 6: 654.

Brotman R. M., 2011 Vaginal microbiome and sexually transmitted infections: an epidemiologic perspective. J. Clin. Invest. 121: 4610–4617.

Brown, A. J. P., K. Haynes, N. A. R. Gow, and J. Quinn, 2012 Stress Responses in Candida, pp. 225-242 in *Candida and Candidiasis*, edited by R. A. Calderone and C. J. Clancy. ASM Press, Washington, DC.

Brown, A. J. P., G. D. Brown, M. G. Netea, and N. A. R. Gow, 2014 Metabolism impacts upon Candida immunogenicity and pathogenicity at multiple levels. Trends Microbiol. 22: 614–622.

Brown, S. E., J. A. Schwartz, C. K. Robinson, D. E. O'Hanlon, L. L. Bradford *et al.*, 2019 The Vaginal Microbiota and Behavioral Factors Associated With Genital Candida albicans Detection in Reproductive-Age Women. Sex. Transm. Dis. 46: 753–758.

Buffo, J., M. A. Herman, and D. R. Soll, 1984 A characterization of pH-regulated dimorphism in Candida albicans. Mycopathologia 85: 21–30.

Bull-Otterson, L., W. Feng, I. Kirpich, Y. Wang, X. Qin *et al.*, 2013 Metagenomic Analyses of Alcohol Induced Pathogenic Alterations in the Intestinal Microbiome and the Effect of Lactobacillus rhamnosus GG Treatment. PLoS One 8: e53028.

Burgener, A., I. McGowan, and N. R. Klatt, 2015 HIV and mucosal barrier interactions: consequences for transmission and pathogenesis. Curr. Opin. Immunol. 36: 22–30.

Buvé, A., V. Jespers, T. Crucitti, and R. N. Fichorova, 2014 The vaginal microbiota and susceptibility to HIV. AIDS 28.

Callister, S. J., M. J. Wilkins, C. D. Nicora, K. H. Williams, J. F. Banfield *et al.*, 2010 Analysis of Biostimulated Microbial Communities from Two Field Experiments Reveals Temporal and Spatial Differences in Proteome Profiles. Environ. Sci. Technol. 44: 8897–8903.

Cannizzo, F. T., E. Eraso, P. A. Ezkurra, M. Villar-Vidal, E. Bollo *et al.*, 2007 Biofilm development by clinical isolates of Malassezia pachydermatis. Med. Mycol. 45: 357–361.

Cano, R. J., J. Rivera-Perez, G. A. Toranzos, T. M. Santiago-Rodriguez, Y. M. Narganes-Storde *et al.*, 2014 Paleomicrobiology: Revealing Fecal Microbiomes of Ancient Indigenous Cultures. PLoS One 9: e106833.

Carlson, R. D., A. N. Sheth, T. D. Read, M. B. Frisch, C. C. Mehta *et al.*, 2017 The Female Genital Tract Microbiome Is Associated with Vaginal Antiretroviral Drug Concentrations in Human Immunodeficiency Virus-Infected Women on Antiretroviral Therapy. J. Infect. Dis. 216: 990–999.

Cartwright, C. P., B. D. Lembke, K. Ramachandran, B. A. Body, M. B. Nye *et al.*, 2013 Comparison of nucleic acid amplification assays with BD affirm VPIII for diagnosis of vaginitis in symptomatic women. J. Clin. Microbiol. 51: 3694–3699.

Cassone, A., 2015 Vulvovaginal Candida albicans infections: pathogenesis, immunity and vaccine prospects. BJOG 122:785-794.

Ceccarani, C., C. Foschi, C. Parolin, A. D'Antuono, V. Gaspari *et al.*, 2019 Diversity of vaginal microbiome and metabolome during genital infections. Sci. Rep. 9: 14095.

Chan, J. F. W., S. K. P. Lau, K.-Y. Yuen, and P. C. Y. Woo, 2016 Talaromyces (Penicillium) marneffei infection in non-HIV-infected patients. Emerg. Microbes Infect. 5: e19.

Chanda, W., T. P. Joseph, W. Wang, A. A. Padhiar, and M. Zhong, 2017 The potential management of oral candidiasis using anti-biofilm therapies. Med. Hypotheses 106: 15–18.

Charrad, M., N. Ghazzali, V. Boiteau, and A. Niknafs, 2014 NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. J. Stat. Softw. 61: 1–36.

Chaudry, A. N., P. J. Travers, J. Yuenger, L. Colletta, P. Evans *et al.*, 2004 Analysis of vaginal acetic acid in patients undergoing treatment for bacterial vaginosis. J. Clin. Microbiol. 42: 5170–5175.

Chen, C. K., D. Y. Chang, S. C. Chang, E. F. Lee, S. C. Huang *et al.*, 1993 Cryptococcal infection of the vagina. Obstet. Gynecol. 81: 867–869.

Chen, Y., Z. Chen, R. Guo, N. Chen, H. Lu *et al.,* 2011 Correlation between gastrointestinal fungi and varying degrees of chronic hepatitis B virus infection. Diagn Micr. Infec. Dis. 70:492–498.

Cheng, V. C. C., J. F.W. Chan, A. H. Y. Ngan, K. K. W. To, S. Y. Leung *et al.*, 2009 Outbreak of intestinal infection due to Rhizopus microsporus. J. Clin. Microbiol. 47: 2834–2843.

Cheng, H. G., and M. R. Phillips, 2014 Secondary analysis of existing data: opportunities and implementation. Shanghai Arch. psychiatry 26: 371–375.

Cheng, K., Z. Ning, X. Zhang, J. Mayne, and D. Figeys, 2018 Trends in Analytical Chemistry Separation and characterization of human microbiomes by metaproteomics. Trends Anal. Chem. 108: 221–230.

Chichili, G. R., and W. Rodgers, 2009 Cytoskeleton-membrane interactions in membrane raft structure. Cell. Mol. Life Sci. 66: 2319–2328.

Chirenje, Z. M., H. M. Gundacker, B. Richardson, L. Rabe, Z. Gaffoor *et al.*, 2017 Risk Factors for Incidence of Sexually Transmitted Infections Among Women in a Human Immunodeficiency Virus Chemoprevention Trial: VOICE (MTN-003). Sex. Transm. Dis. 44: 135–140.

Chu, D. M., M. Seferovic, R. M. Pace, and K. M. Aagaard, 2018 The microbiome in preterm birth. Best Pract. Res. Clin. Obstet. Gynaecol. 52: 103–113.

Cohen, R., F. J. Roth, E. Delgado, D. G. Ahearn, and M. H. Kalser, 1969 Fungal Flora of the Normal Human Small and Large Intestine. N. Engl. J. Med. 280: 638–641.

Cohen, C. R., J. R. Lingappa, J. M. Baeten, M. O. Ngayo, C. A. Spiegel *et al.*, 2012 Bacterial vaginosis associated with increased risk of female-to-male HIV-1 transmission: a prospective cohort analysis among African couples. PLoS Med. 9: e1001251–e1001251.

Cohn, J. A., F. B. Hashemi, M. Camarca, F. Kong, J. Xu *et al.*, 2005 HIV-inducing factor in cervicovaginal secretions is associated with bacterial vaginosis in HIV-1-infected women. J. Acquir. Immune Defic. Syndr. 39: 340–346.

Colgan, W., and A. W. Claridge, 2002 Mycorrhizal effectiveness of Rhizopogon spores recovered from faecal pellets of small forest-dwelling mammals. Mycol. Res. 106: 314–320.

Corpus, K., R. Hegeman-Dingle, and I. Bajjoka, 2004 Candida kefyr, an Uncommon but Emerging Fungal Pathogen: Report of Two Cases. Pharmacother. J. Hum. Pharmacol. Drug Ther. 24: 1084–1088.

Cortes, L., H. Wopereis, A. Tartiere, J. Piquenot, J. W. Gouw *et al.*, 2019 Metaproteomic and 16S rRNA gene sequencing analysis of the infant fecal microbiome. Int. J. Mol. Sci. 20: 9–12.

Cotch, M. F., S. L. Hillier, R. S. Gibbs, and D. A. Eschenbach, 1998 Epidemiology and outcomes associated with moderate to heavy Candida colonization during pregnancy. Am. J. Obstet. Gynecol. 178:374–380.

Cottrell, J. S., 2011 Protein identification using MS/MS data. J. Proteomics 74: 1842–1851.

Cowen, L. E., 2008 The evolution of fungal drug resistance: modulating the trajectory from genotype to phenotype. Nat. Rev. Microbiol. 6: 187–198.

Cowen, L. E., D. Sanglard, S. J. Howard, P. D. Rogers, and D. S. Perlin, 2015 Mechanisms of Antifungal Drug Resistance. Cold Spring Harb. Perspect. Med. 5: a019752–a019752.

Craig, E. A., 1985 The heat shock response. CRC Crit. Rev. Biochem. 18:239-280.

Cribby, S., M. Taylor, and G. Reid, 2008 Vaginal microbiota and the use of probiotics. Interdiscip. Perspect. Infect. Dis. 2008: 256490.

Crosby, L. D., and C. S. Criddle, 2003 Understanding bias in microbial community analysis techniques due to rrn operon copy number heterogeneity. Biotechniques 34: 790–4, 796, 798 passim.

Cui, L., A. Morris, and E. Ghedin 2013: The human mycobiome in health and disease. Genome Med. 5:63.

D'ippolito, S., F. Di Nicuolo, A. Pontecorvi, M. Gratta,and G. Scambia, 2018 Endometrial microbes and microbiome: recent insights on the inflammatory and immune "players" of the human endometrium, Am. J. Reprod. Immunol. 80: e13065.

Dabee, S., S. L. Barnabas, K. S. Lennard, S. Z. Jaumdally, H. Gamieldien *et al.*, 2019 Defining characteristics of genital health in South African adolescent girls and young women at high risk for HIV infection. PLoS One 14: e0213975–e0213975.

Danielson, R. M., and C. B. Davey, 1973. The abundance of Trichoderma propagules and the distribution of species in forest soils. Soil Biol. Biochem. 5: 485–494.

Danielsson, D, P. K. Teigen, and H. Moi, 2011 The genital econiche: focus on microbiota and bacterial vaginosis. Ann. N. Y. Acad. Sci. 1230:48–58.

Dasari, S., S. K. Anandan, W. Rajendra, and L. Valluru, 2016 Role of microbial flora in female genital tract: A comprehensive review. Asian Pacific J. Trop. Dis. 6: 909–917.

De Almeida, D. G. M., S. Figueiredo Costa, M. Melhem, A. L. Motta, M. Walderez Szeszs *et al.*, 2008 Rhodotorula spp. isolated from blood cultures: clinical and microbiological aspects. Med. Mycol. 46: 547–556.

Degroeve, S., A. Staes, P.-J. De Bock, and L. Martens, 2012 The effect of peptide identification search algorithms on MS2-based label-free protein quantification. OMICS 16: 443–448.

Dennerstein, G. J., D. H. Ellis, C. S. Reed, and C. M. Bennett, 2011 Pathogenicity of Non-albicans Yeasts in the Vagina. J. Low. Genit. Tract Dis. 15.

Denning, D. W., M. Kneale, J. D. Sobel, and R. Rautemaa-Richardson, 2018 Global burden of recurrent vulvovaginal candidiasis: a systematic review. Lancet. Infect. Dis. 18: e339–e347.

Deorukhkar ,S. C., and S. Roushani, 2017 Virulence Traits Contributing to Pathogenicity of Candida Species. J Microbiol Exp 5(1): 00140.

Deorukhkar, S. C., and S. Saini, 2015 Candidiasis: Past, present, future Int. J. Infect. Trop. Dis 2: 12-24.

Deruaz, M., and A. D. Luster, 2015 Chemokine-mediated immune responses in the female genital tract mucosa. Immunol. Cell Biol. 93: 347–354.

Devkota, S., Y. Wang, M. W. Musch, V. Leone, H. Fehlner-Peach *et al.*, 2012 Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10−/− mice. Nature 487: 104.

Diamandis E. P., 2004 Mass spectrometry as a diagnostic and a cancer biomarker discovery tool. Mol. Cell. Proteomics 3: 367– 378.

Diaz, N., D. Dessì, S. Dessole, P. L. Fiori, and P. Rappellia, 2010 Rapid detection of co-infections by Trichomonas vaginalis, Mycoplasma hominis, and Ureaplasma urealyticum by a new multiplex polymerase chain reaction. Diagn. Microbiol. Infect. Dis. 67:30-36

Dollive, S., Y.-Y. Chen, S. Grunberg, K. Bittinger, C. Hoffmann *et al*., 2013 Fungi of the Murine Gut: Episodic Variation and Proliferation during Antibiotic Treatment. PLoS One 8: e71806.

Donati, L., A. Di Vico, M. Nucci, L. Quagliozzi, T. Spagnuolo *et al*., 2010 Vaginal microbial flora and outcome of pregnancy. Arch. Gynecol. Obstet. 281:589–600.

Donbraye-Emmanuel, O. O. B., E. Donbraye, I. Okonko, A. Adeolu, M. Ojezele *et al*., 2010 Detection and prevalence of Candida among pregnant women in Ibadan, Nigeria. World Appl. Sci. J. 10(9) 986-991.

Donders, G. G., A. Vereecken, E. Bosmans, A. Dekeersmaecker, G. Salembier *et al.*, 2002 Definition of a type of abnormal vaginal flora that is distinct from bacterial vaginosis: Aerobic vaginitis. BJOG 109: 34–43.

Donders, G. G., 2007 Definition and classification of abnormal vaginal flora. Best Pract Res Clin Obstet Gynaecol. 21: 355–373.

Donders, G. G., K. van Calsteren, G. Bellen, R. Reybrouck, T. van den Bosch T *et al.*, 2009 Predictive value for preterm birth of abnormal vaginal flora, bacterial vaginosis and aerobic vaginitis during the first trimester of pregnancy. BJOG 116: 1315–1324.

Donders, G. G., G. Bellen, J. Ausma, L. Verguts, J. Vaneldere *et al*., 2011 The effect of antifungal treatment on the vaginal flora of women with vulvo-vaginal yeast infection with or without bacterial vaginosis. Eur. J. Clin. Microbiol. Infect. Dis. 30: 59–63.

Dougherty, E. R., 2007 Validation of inference procedures for gene regulatory networks. Curr. Genomics 8: 351–359.

Drell, T., T. Lillsaar, L. Tummeleht, J. Simm, A. Aaspollu *et al.*, 2013 Characterization of the vaginal micro- and mycobiome in asymptomatic reproductive-age Estonian women. PloS One 8:e54379.

Dworecka-Kaszak, B., I. Dabrowska, and I. Kaszak, 2016 The mycobiome–a friendly cross-talk between fungal colonizers and their host. Ann. Parasitol. 62. 175–184.

Eastment, M. C., J. E. Balkus, B. A. Richardson, S. Srinivasan, J. Kimani *et al.*, 2021 Association Between Vaginal Bacterial Microbiota and Vaginal Yeast Colonization. J. Infect. Dis. 223: 914–923.

Ehrenhofer-Murray A. E., 2004 Chromatin dynamics at DNA replication, transcription and repair. Eur. J. Biochem. 271: 2335–2349.

El Aila, N. A., I. Tency, B. Saerens, E. De Backer, P. Cools *et al*., 2011 Strong correspondence in bacterial loads between the vagina and rectum of pregnant women. Res Microbiol. 162:506–13.

El Aila, N. A., I. Tency, G. Claeys, H. Verstraelen, B. Saerens *et al.*, 2009 Identification and genotyping of bacteria from paired vaginal and rectal samples from pregnant women indicates similarity between vaginal and rectal microflora. BMC Infect Dis. 9:167.

El-Din, A. Z. K., F. Habib, N. Abd-Allah, and O. Khorshid, 2009 Mycotic vulvovaginitis: Epidemiology, pathogenesis and profile of antifungal agents. J. Taibah Univ. Med. Sci. 4: 123–136.

Els, N., C. Larose, K. Baumann-Stanzer, R. Tignat-Perrier, C. Keuschnig *et al*., 2019 Microbial composition in seasonal time series of free tropospheric air and precipitation reveals community separation. Aerobiologia (Bologna). 35.

Eng, J. K., B. C. Searle, K. R. Clauser, and D. L. Tabb, 2011 A face in the crowd: recognizing peptides through database search. Mol. Cell. Proteomics 10: R111.009522-R111.009522.

Enoch, D. A., H. A. Ludlam, and N. M. Brown, 2006 Invasive fungal infections: a review of epidemiology and management options. J. Med. Microbiol. 55:809-818.

Erickson, A. R., B. L. Cantarel, R. Lamendella, Y. Darzi, E. F. Mongodin *et al*., 2012 Integrated Metagenomics/Metaproteomics Reveals Human Host-Microbiota Signatures of Crohn's Disease. PLoS One 7: e49138.

Eschenbach, D. A., D. L. Patton, A. Meier, S.S. Thwin, J. Aura *et al*., 2000 Effects of oral contraceptive pill use on vaginal flora and vaginal epithelium. Contraception 62:107–112.

Farr, A., H. Kiss, I. Holzer, P. Husslein, M. Hagmann *et al*., 2015 Effect of asymptomatic vaginal colonization with Candida albicans on pregnancy outcome. Acta Obstet. Gynecol. Scand. 94: 989–996.

Ferrazza, M. H. S. H., M. L. F. Maluf, M. E. L. Consolaro, C. S. Shinobu, T. I. E. Svidzinski *et al.*, 2005 Caracterização de leveduras isoladas da vagina e sua associação com candidíase vulvovaginal em duas cidades do sul do Brasil. Rev. Bras. Ginecol. e Obs. 27: 58–63.

Ferrer, C., J. J. Pérez-Santonja, A. E. Rodríguez, M. F. Colom, J. Gené *et al*., 2009 New pyrenochaeta species causing keratitis. J. Clin. Microbiol. 47: 1596–1598.

Ferrer, M., A. Ruiz, F. Lanza, S.-B. Haange, A. Oberbach, H. Till *et al.*, 2013 Microbiota from the distal guts of lean and obese adolescents exhibit partial functional redundancy besides clear differences in community structure, Environ. Microbiol. 15:211–226.

Ferris, D., P. Nyirjesy, J. Sobel, D. Soper D *et al.*, 2002 Over-the-Counter Antifungal Drug Misuse Associated With Patient-Diagnosed Vulvovaginal Candidiasis. Obstet. Gynecol. 99: 419–425.

Ferris, M. J., J. Norori, M. Zozaya-Hinchliffe, and D. H. Martin, 2007 Cultivation-independent analysis of changes in bacterial vaginosis flora following metronidazole treatment. J. Clin. Microbiol. 45: 1016–1018.

Fethers, K. A., Fairley C. K., Hocking J. S., Gurrin L. C., and Bradshaw C. S., 2008 Sexual Risk Factors and Bacterial Vaginosis: A Systematic Review and Meta-Analysis. Clin. Infect. Dis. 47: 1426–1435.

Fethers, K., J. Twin, C. K. Fairley, F. J. I. Fowkes, S. M. Garland *et al.*, 2012 Bacterial Vaginosis (BV) Candidate Bacteria: Associations with BV and Behavioural Practices in Sexually-Experienced and Inexperienced Women. PLoS One 7: e30633.

Fettweis, J. M., J. P. Brooks, M. G. Serrano, N. U. Sheth, P. H. Girerd *et al*., 2014 Differences in vaginal microbiome in African American women versus women of European ancestry. Microbiology 160: 2272–2282.

Fidel Jr, P. L., 2002 Immunity to Candida. Oral Dis. 8: 69–75.

Fidel Jr, P. L., M. Barousse, V. Lounev, T. Espinosa, R. R. Chesson *et al.*, 2003 Local immune responsiveness following intravaginal challenge with Candida antigen in adult women at different stages of the menstrual cycle. Med. Mycol. 41: 97–109.

Fidel Jr, P. L., 2004 History and new insights into host defense against vaginal candidiasis. Trends Microbiol. 12: 220–227.

Fidel Jr, P. L., M. Barousse, T. Espinosa, M. Ficarra M, J. Sturtevant *et al.*, 2004 An intravaginal live Candida challenge in humans leads to new hypotheses for the immunopathogenesis of vulvovaginal candidiasis. Infect Immun. 72:2939–46.

Fleck, C. B., F. Schöbel, and M. Brock 2011 Nutrient acquisition by pathogenic fungi: nutrient availability, pathway regula- tion, and differences in substrate utilization. Int. J. Med. Microbiol. 301: 400-407.

Flournoy, D. J., J. B. Mullin, and R. J. McNeal, 2000 Isolation of fungi from rose bush thorns. J. Okla. State Med. Assoc. 93: 271–274.

Forbes, J. D., C. N. Bernstein, H. Tremlett, G. Van Domselaar, and N. C. Knox 2019 A fungal world: Could the gut mycobiome be involved in neurological disease? Front. Microbiol. 10: 1–13.

Fornari, G., V. A. Vicente, R. R. Gomes, M. D. Muro, R. L. Pinheiro *et al.*, 2016 Susceptibility and molecular characterization of candida species from patients with vulvovaginitis. Brazilian J. Microbiol. 47: 373–380.

Förster, T. M., S. Mogavero, A, Dräger, K. Graf, M. Polke *et al.*, 2016 Enemies and brothers in arms: Candida albicans and gram-positive bacteria. Cell. Microbiol. 18: 1709–1715.

Foster, J. A., S. M. Krone, and L. J. Forney, 2008 Application of ecological network theory to the human microbiome. Interdiscip. Perspect. Infect. Dis. 2008: 839501.

Franklin, R. D., and W. H. Kutteh, 1999 Characterization of immunoglobulins and cytokines in human cervical mucus: influence of exogenous and endogenous hormones. J. Reprod. Immunol. 42: 93–106.

Franklin, T. L., and G. R. Monif, 2000 Trichomonas vaginalis and bacterial vaginosis. Coexistence in vaginal wet mount preparations from pregnant women. J Reprod. Med. 45(2):131-134.

Fredricks, D. N., T. L. Fiedler, and J. M. Marrazzo, 2005 Molecular identification of bacteria associated with bacterial vaginosis. N. Engl. J. Med. 353: 1899–1911.

Frobenius, W., and C. Bogdan, 2015 Diagnostic Value of Vaginal Discharge, Wet Mount and Vaginal pH - An Update on the Basics of Gynecologic Infectiology. Geburtshilfe Frauenheilkd. 75 4: 355–366.

Frohner, I. E., C. Bourgeois, K. Yatsyk, O. Majer, and K. Kuchler, 2009 Candida albicans cell surface superoxide dismutases degrade host-derived reactive oxygen species to escape innate immune surveillance. Mol. Microbiol. 71: 240–252.

Gafos, M., M. Mzimela, S. Sukazi, R. Pool, C. Montgomery *et al.*, 2010 Intravaginal insertion in KwaZulu-Natal: sexual practices and preferences in the context of microbicide gel use. Cult. Health Sex. 12: 929–942.

Gajer, P., R. M. Brotman, G. Bai, J. Sakamoto, U. M. E. Schütte *et al.*, 2012 Temporal dynamics of the human vaginal microbiota. Sci. Transl. Med. 4:132ra52.

Gardner, H. L., and C. D. Dukes, 1954 New etiologic agent in nonspecific bacterial vaginitis. Sci. 120: 853.

Garg, N., O. Prakash, B. K. Pandey, B. P. Singh, and G. Pandey, 2004 First Report of Black Soft Rot of Indian Gooseberry Caused by Syncephalastrum racemosum. Plant Dis. 88: 575.

Garrett, N. J., F. Osman, B. Maharaj, N. Naicker, A. Gibbs *et al.*, 2018 Beyond syndromic management: Opportunities for diagnosis-based treatment of sexually transmitted infections in low- and middle-income countries. PLoS One 13: e0196209.

Gatto, L., and K. S. Lilley, 2012 MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. Bioinformatics 28: 288–289.

Ghannoum, M. A., R. J. Jurevic, P. K. Mukherjee, F. Cui, M. Sikaroodi *et al.*, 2010 Characterization of the Oral Fungal Microbiome (Mycobiome) in Healthy Individuals. PLOS Pathog. 6: e1000713.

Giraldo, P. C., A. D. Ribeiro-Filho, J. A. Simões, A. Neuer, S. B. N. Feitosa *et al.*, 1999a Circulating heat shock proteins in women with a history of recurrent vulvovaginitis. Infect. Dis. Obstet. Gynecol. 7: 128–132.

Giraldo, P., A. Neuer, I. L. Korneeva, A. D. Ribeiro-Filho, J. A. Simões *et al.*, 1999b Vaginal heat shock protein expression in symptom-free women with a history of recurrent vulvovaginitis. Am. J. Obstet. Gynecol. 180: 524–529.

Godoy-Vitorino, F., J. Romaguer, C. Zhao, D. Vargas-Robles, G. Ortiz-Morales *et al.*, 2018 Cervicovaginal Fungi and Bacteria Associated With Cervical Intraepithelial Neoplasia and High-Risk Human Papillomavirus Infections in a Hispanic Population. Front. Microbiol. 9:2533

Goldacre, M. J., L. J. R. Milne, B. Watt, N. Loudon, and M. P. Vessey ,1981 Relative abundance of yeasts and fungi other than Candida albicans in the vagina of normal young women. BJOG An Int. J. Obstet. Gynaecol. 88: 596–600.

Gosalbes, M. J., A. Durbán, M. Pignatelli, J. J. Abellan, N. Jiménez-Hernández *et al.*, 2011 Metatranscriptomic Approach to Analyze the Functional Human Gut Microbiota. PLoS One 6: e17447.

Gostinčar, C., X. Sun, J. Zajc, C. Fang, Y. Hou *et al*., 2019 Population Genomics of an Obligately Halophilic Basidiomycete Wallemia ichthyophaga. Front. Microbiol. 10: 2019.

Goswami, R., V. Dadhwal, S. Tejaswi, K. Datta, A. Paul *et al.*, 2000 Species-specific prevalence of vaginal candidiasis among patients with diabetes mellitus and its relation to their glycaemic status. J. Infect. 41: 162–166.

Gouba, N., D. Raoult, and M. Drancourt, 2013 Plant and fungal diversity in gut microbiota as revealed by molecular and culture investigations. PLoS One 8: e59474.

Gouba, N., D. Raoult, and M. Drancourt, 2014a Eukaryote Culturomics of the Gut Reveals New Species. PLoS One 9: e106994.

Gouba, N., D. Raoult, M. Drancourt, 2014b Gut microeukaryotes during anorexia nervosa: a case report. BMC Res. Notes 7: 33.

Greenblum, S., R. Carr, and E. Borenstein, 2015 Extensive Strain-Level Copy-Number Variation across Human Gut Microbiome Species. Cell 160: 583–594.

Grigoriou, O., S. Baka, E. Makrakis, D. Hassiakos, G. Kapparos *et al.*, 2006 Prevalence of clinical vaginal candidiasis in a university hospital and possible risk factors. Eur. J. Obstet. Gynecol. Reprod. Biol. 126: 121–125.

Grossart, H.-P., S. van den Wyngaert, M. Kagami, C. Wurzbacher, M. Cunliffe *et al.*, 2019 Fungi in aquatic ecosystems. Nat. Rev. Microbiol. 17: 339–354.

Gugnani, H. C., 1992 A review of zygomycosis due to basidiobolus ranarum. Eur. J. Epi. 15: 923–929.

Günther, J., M. Koy, A. Berthold, H. J. Schuberth, and H. M. Seyfert, 2016 Comparison of the pathogen species-specific immune response in udder derived cell types and their models. Vet. Res. 47: 1–11.

Guo, R., N. Zheng, H. Lu, H. Yin, J. Yao, and Y. Chen, 2012 Increased diversity of fungal flora in the vagina of patients with recurrent vaginal candidiasis and allergic rhinitis. Microb. Ecol. 64:918-27.

Gupta, K., S. L. Hillier, T. M. Hooton, P. L. Roberts, and W. E. Stamm, 2000 Effects of contraceptive method on the vaginal microbial flora: a prospective evaluation. J. Infect. Dis. 181:595–601.

Guzel, A. B., M. Ilkit, R. Burgut, I. F. Urunsak, and F. T. Ozgunen, 2011a An evaluation of risk factors in pregnant women with Candida vaginitis and the diagnostic value of simultaneous vaginal and rectal sampling. Mycopathologia 172: 25–36.

Guzel A. B., M. Ilkit, T. Akar, R. Burgut, and S. C. Demir, 2011b Evaluation of risk factors in patients with vulvovaginal candidiasis and the value of chromID Candida agar versus CHROMagar Candida for recovery and presumptive identification of vaginal yeast species. Med. Mycol. 49: 16–25.

Haase, A. T., 2011 Early Events in Sexual Transmission of HIV and SIV and Opportunities for Interventions. Annu. Rev. Med. 62: 127–139.

Hager, C. L., and M. A. Ghannoum 2018 The mycobiome in HIV. Current opinion in HIV and AIDS. 13:69–72.

Hallen-Adams, H., S. Kachman, J. Kim, R. Legge, and I. Martínez, 2015 Fungi inhabiting the healthy human gastrointestinal tract: A diverse and dynamic community. Fungal Ecol. 15.

Hamad, I., C. Sokhna, D. Raoult, and F. Bittar, 2012 Molecular Detection of Eukaryotes in a Single Human Stool Sample from Senegal. PLoS One 7: e40888.

Han, T.-L., R. D. Cannon, S. M. Gallo, and S. G. Villas-Bôas, 2019 A metabolomic study of the effect of Candida albicans glutamate dehydrogenase deletion on growth and morphogenesis. npj Biofilms Microbiomes 5: 13.

Harrell Jr, F. E., and C. Dupont, 2008. Hmisc: harrell miscellaneous. R Packag version 3.

Hay, P., 2009 Recurrent bacterial vaginosis. Curr. Opin. Infect. Dis. 22.

Hermann, C., J. Hermann, U. Munzel, and R. Ruchel, 1999 Bacterial flora accompanying Candida yeasts in clinical specimens. Mycoses 42: 619–627.

Hernández-Santos, N., and B. S. Klein, 2017 Through the Scope Darkly: The Gut Mycobiome Comes into Focus. Cell Host Microbe 22: 728–729.

Hettich, R., R. Sharma, K. Chourey, and R. Giannone, 2012 Microbial metaproteomics: Identifying the repertoire of proteins that microorganisms use to compete and cooperate in complex environmental communities. Curr. Opin. Microbiol. 15: 373–380.

Hettich, R. L., C. Pan, K. Chourey, R. J. Giannone, 2013 Metaproteomics: Harnessing the Power of High Performance Mass Spectrometry to Identify the Suite of Proteins That Control Metabolic Activities in Microbial Communities. Anal. Chem. 85: 4203–4214.

Heyer, R., F. Kohrs, U. Reichl, and D. Benndorf, 2015. Metaproteomics of complex microbial communities in biogas plants. Microb. Biotechnol. 8:749–763.

Heyer, R., D. Benndorf, F. Kohrs, J. De Vrieze, N. Boon *et al.*, 2016 Proteotyping of biogas plant microbiomes separates biogas plants according to process temperature and reactor type. Biotechnol. Biofuel 9:155.

Heyer, R., K. Schallert, R. Zoun, B. Becher, G. Saake *et al.*, 2017 Challenges and perspectives of metaproteomic data analysis. J. Biotechnol. 261: 24–36.

Hillier, S. L., M. A. Krohn, R. P. Nugent, and R. S. Gibbs, 1992 Characteristics of three vaginal flora patterns assessed by gram stain among pregnant women. Vaginal Infections and Prematurity Study Group. Am. J. Obstet. Gynecol. 166: 938–944.

Hoffmann, C., S. Dollive, S. Grunberg, J. Chen, H. Li *et al.*, 2013 Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. PLoS ONE 8:e66019.

Hogan, D., M. Wargo, and N. Beck, 2009 Bacterial Biofilms on Fungal Surfaces, pp. 235–245 in *In The Biofilm Mode of Life: Mechanisms and Adaptations*, edited by S. Kjelleberg, and M. Givskov. Horizon Scientific Press, Norfolk.

Holland, J., M. L. Young, O. Lee, and C. A. S. Chen, 2003 Vulvovaginal carriage of yeasts other than Candida albicans. Sex. Transm. Infect. 79: 249–50.

Holanda A. A. R., A. C. S. Fernandes, C. M. Bezerra, M. A. F. Ferreira, M. R. R. Holanda *et al.*, 2007 Candidíase vulvovaginal: sintomatologia, fatores de risco e colonizac¸ãoanal concomitante. Rev Bras Ginecol Obstet. 29:3–9.

Holland, J., M. L. Young, O. Lee, and C.-A. S. Chen, 2003 Vulvovaginal carriage of yeasts other than Candida albicans. Sex. Transm. Infect. 79: 249–250.

Hong, K. H., S. K. Hong, S. I. Cho, E. Ra, K. H. Han *et al.*, 2016 Analysis of the Vaginal Microbiome by Next-Generation Sequencing and Evaluation of its Performance as a Clinical Diagnostic Tool in Vaginitis. Ann. Lab. Med. 2016 36(5):441-449.

Hort, W., and P. Mayser, 2011 Malassezia virulence determinants. Curr. Opin. Infect. Dis. 24.

Huber, J. A., D. B. Mark Welch, H. G. Morrison, S. M. Huse, P. R. Neal *et al.*, 2007 Microbial Population Structures in the Deep Marine Biosphere. Sci. 318: 97-100.

Huffnagle, G. B., and M. C. Noverr, 2013 The emerging world of the fungal microbiome. Trends Microbiol. 21: 334–341.

Human Microbiome Project Consortium, 2012 Structure, function and diversity of the healthy human microbiome. Nature 486: 207–214.

Hurley, R., and J. De Louvois, 1979 Candida vaginitis. Postgrad. Med. J. 155: 645–47.

Huseyin, C. E., P. W. O'Toole, P. D. Cotter, and P. D. Scanlan, 2017 Forgotten fungi-the gut mycobiome in human health and disease. FEMS Microbiol. Rev. 41: 479–511.

Iliev, I. D., V. A. Funari, K. D. Taylor, Q. Nguyen, C. N. Reyes *et al.*, 2012 Interactions between commensal fungi and the C-type lectin receptor Dectin-1 influence colitis. Sci. 336: 1314–1317.

Iliev, I. D., and I. Leonardi, 2017 Fungal dysbiosis: immunity and interactions at mucosal barriers. Nat. Rev. Immunol. 17: 635–646.

Isa-Isa, R., R. Arenas, R. F. Fernández, and M. Isa, 2012 Rhinofacial conidiobolomycosis (entomophthoramycosis). Clin. Dermatol. 30: 409–412.

Iwasaki, A., and R. Medzhitov, 2010 Regulation of adaptive immunity by the innate immune system. Science 327: 291–295.

Izzo, A. D., M. Meyer, J. M. Trappe, M. North, and T. D. Bruns, 2005 Hypogeous Ectomycorrhizal Fungal Species on Roots and in Small Mammal Diet in a Mixed-Conifer Forest. For. Sci. 51: 243–254.

Jacobstein, R., and C. B. Polis, 2014 Progestin-only contraception: Injectables and implants. Best Pract. Res. Clin. Obstet. Gynaecol. 28: 795–806.

Jagtap, P. D., J. Goslinga, J. A. Kooren, T. McGowan, M. S. Wroblewski *et al.*, 2013 A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies. Proteomics 13: 1352–1357.

Jagtap, P. D., K. J. Viken, J. Johnson, T. McGowan, K. M. Pendleton *et al.*, 2018 BAL Fluid Metaproteome in Acute Respiratory Failure. Am. J. Respir. Cell Mol. Biol. 59: 648–652.

Jang, S. J., K. Lee, B. Kwon, H. J. You, and G. Ko, 2019 Vaginal lactobacilli inhibit growth and hyphae formation of Candida albicans. Sci. Rep. 9: 8121.

Jansson, J., B. Willing, M. Lucio, A. Fekete, J. Dicksved *et al*., 2009 Metabolomics Reveals Metabolic Biomarkers of Crohn's Disease. PLoS One 4: e6386.

Joe, L. K., N. I. T. Eng, H. van der Muileen, and E. W. Emmons, 1956 Basidiobolus ranarum as a cause of subcutaneous mycosis in Indonesia. Arch. Dermatol. 74: 378 – 383.

Jones, M. L., C. J. Martoni, J. G. Ganopolsky, and A. Labbe, 2014 Prakash The human microbiome and bile acid metabolism: dysbiosis, dysmetabolism, disease and intervention. Expert Opin. Biol. Ther. 14:467–482.

Joo, H., Y.-G. Choi, S.-Y. Cho, J.-K. Choi, D.-G. Lee *et al.*, 2016 Pseudozyma aphidis fungaemia with invasive fungal pneumonia in a patient with acute myeloid leukaemia: case report and literature review. Mycoses 59: 56–61.

Kaida, A., J. J. Dietrich, F. Laher, M. Beksinska, M. Jaggernath *et al.*, 2018 A high burden of asymptomatic genital tract infections undermines the syndromic management approach among adolescents and young adults in South Africa: implications for HIV prevention efforts. BMC Infect. Dis. 18: 499.

Kalia, N., J. Singh, and M. Kaur, 2019 Immunopathology of Recurrent Vulvovaginal Infections: New Aspects and Research Directions. Front. Immunol. 10: 2034.

Kalus, A., 2017 Fungal Skin Infections. In:, pp. 488–500.

Kan, J., T. E. Hanson, J. M. Ginter, K. Wang, and F. Chen, 2005 Metaproteomic analysis of Chesapeake Bay microbial communities. Saline Systems 1: 7

Kanehisa, M., and S. Goto, 2000 KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28: 27–30.

Kanehisa, M., Y. Sato, and K. Morishima, 2016 BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. 428: 726-731.

Kapiga, S., C. Kelly, S. Weiss, T. Daley, L. Peterson *et al.*, 2009 Risk Factors for Incidence of Sexually Transmitted Infections Among Women in South Africa, Tanzania, and Zambia: Results From HPTN 055 Study. Sex. Transm. Dis. 36.

Karp, N. A., and K. S. Lilley, 2007 Design and Analysis Issues in Quantitative Proteomics Studies. Proteomics 7: 42–50.

Katragkou, A., M. J. Kruhlak, M. Simitsopoulou, A. Chatzimoschou, A. Taparkou *et al.*, 2010 Interactions between human phagocytes and Candida albicans biofilms alone and in combination with antifungal agents. J. Infect. Dis. 201: 1941–1949.

Kaye, D., and M. E. Levison, 1977 In vitro inhibition of growth of neisseria gonorrhoeae by genital microorganisms. Sex. Transm. Dis. 4: 1–3.

Khot, P. D., and D. N. Fredricks, 2009 PCR-based diagnosis of human fungal infections. Expert Rev. Anti. Infect. Ther. 7: 1201–1221.

Klebanoff, S. J., S. L. Hillier, D. A. Eschenbach, and A. M. Waltersdorph, 1991 Control of the microbial flora of the vagina by H2O2-generating lactobacilli. J. Infect. Dis. 164:94–100.

Kleiner, M., E. Thorson, C. E. Sharp, X. Dong, D. Liu *et al.,* 2017 Assessing species biomass contributions in microbial communities via metaproteomics. Nat. Commun. 8: 1558.

Klimesova, K., Z. Jiraskova Zakostelska, and H. Tlaskalova-Hogenova, 2018 Oral Bacterial and Fungal Microbiome Impacts Colorectal Carcinogenesis. Front. Microbiol. 9: 774.

Kohlmeyer, J., 1974. On the Definition and Taxonomy of Higher Marine Fungi. vol. 5. Veröffentl. Inst. Meeresforsch, Bremerhaven S263–S286.

Kohrs, F., R. Heyer, A. Magnussen, D. Benndorf, T. Muth *et al.*, 2014 Sample prefractionation with liquid isoelectric focusing enables in depth microbial metaproteome analysis of mesophilic and thermophilic biogas plants. Anaerobe 29: 59–67.

Kolmeder, C. A., W. M. de Vos, 2014 Metaproteomics of our microbiome - Developing insight in function and activity in man and model systems. J. Proteomics.

Kolmeder, C. A., J. Salojärvi, J. Ritari, M. de Been, J. Raes *et al.*, 2016 Faecal Metaproteomic Analysis Reveals a Personalized and Stable Functional Microbiome and Limited Effects of a Probiotic Intervention in Adults. PLoS One 11: e0153294.

Kredics, L., Z. Antal, I. Dóczi, L. Manczinger, F. Kevei, and E. Nagy, 2003. Clinical importance of the genus Trichoderma. Acta Microbiol. Immunol. Hung. 50: 105–117.

Krüger, W., S. Vielreicher, M. Kapitan, I. D. Jacobsen, and M. J. Niemiec, 2019 Fungal-Bacterial Interactions in Health and Disease. Pathog. (Basel, Switzerland) 8.

Kruppa, M. D., J. Jacobs, K. King-Hook, K. Galloway, A. Berry *et al.*, 2019 Binding of Elementary Bodies by the Opportunistic Fungal Pathogen Candida albicans or Soluble β-Glucan, Laminarin, Inhibits Chlamydia trachomatis Infectivity. Front. Microbiol. 9: 3270.

Kruse, J., G. Doehlemann, E. Kemen, and M. Thines, 2017 Asexual and sexual morphs of Moesziomyces revisited. IMA Fungus 8: 117–129.

Kuhn, M., 2008 Building Predictive Models in R Using the caret Package. J. Stat. Software 28:5.

Kumamoto, C. A., and M. D. Vinces, 2005 Contributions of hyphae and hypha-co-regulated genes to Candida albicans virulence. Cell. Microbiol. 7: 1546–1554.

Kumar, S., and S. Singhi, 2016 Role of Probiotics in Prevention of Candida Colonization and Invasive Candidiasis. J Matern Fetal Neonatal Med. 29:818–9.

Kwon-Chung, K. J., and J. E. Bennett, 1992 Philadelphia: Lea and Febiger. Med. Myc.

Lai, G. C., T. G. Tan, and N. Pavelka, 2019 The mammalian mycobiome: A complex system in a dynamic relationship with the host. Wiley Interdiscip. Rev. Syst. Biol. Med11(1):e1438.

Lamont, R. F., J. D. Sobel, R. A. Akins, S. S. Hassan, T. Chaiworapongsa *et al.*, 2011 The vaginal microbiome: new information about genital tract flora using molecular based techniques. BJOG 118: 533–549.

Lanternier, F., D. Boutboul, J. Menotti, M. O. Chandesris, C. Sarfati *et al.*, 2009 Microsporidiosis in solid organ transplant recipients: two Enterocytozoon bieneusi cases and review. Transpl. Infect. Dis. 11: 83–88.

Lanver, D., M. Tollot, G. Schweizer, L. Lo Presti, S. Reissmann *et al.*, 2017 Ustilago maydis effectors and their impact on virulence. Nat. Rev. Microbiol. 15: 409–421.

Lazar, C., L. Gatto, M. Ferro, C. Bruley, and T. Burger, 2016 Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies.

Lehtoranta, L., A. A. Hibberd, N. Yeung, A. Laitila, J. Maukonen *et al.*, 2021 Short communication: Characterization of vaginal fungal communities in healthy women and women with bacterial vaginosis (BV); a pilot study. Microb. Pathog.: 105055.

Lennard, K., S. Dabee, S. Barnabas, E. Havyarimana, A. Blakney *et al.*, 2018 Microbiome composition and function predict genital tract inflammation and persistent bacterial vaginosis in adolescent South African women. Infect. Immun. 86(1):e00410–17.

Lennard, K., S. Dabee, S. Barnabas, E. Havyarimana, A. Blakney *et al.*, 2019 Vaginal microbiota varies by geographical location in South African women.

Li, Q., C. Wang, C. Tang, Q. He, N. Li *et al.*, 2014 Dysbiosis of gut fungal microbiota is associated with mucosal inflammation in Crohn's disease. J. Clin. Gastroenterol. 48: 513–523.

Li, F., C. Chen, W. Wei, Z. Wang, J. Dai *et al.*, 2018 The metagenome of the female upper reproductive tract. Gigascience 7: giy107.

Li, T., Z. Liu, X. Zhang, X. Chen, and S. Wang, 2019 Local pobiotic Lactobacillus crispatus and Lactobacillus delbrueckii exhibit strong antifungal effects against vulvovaginal candidiasis in a rat model. Front Microbiol. 10:1033.

Lichius, A., A. Berepiki, and N. D. Read, 2011 Form follows function – The versatile fungal cytoskeleton. Fungal Biol. 115: 518–540.

Limper, A. H., A. Adenis, T. Le, and T. S. Harrison, 2017 Fungal infections in HIV/AIDS. Lancet. Infect. Dis. 17: e334–e343.

Linhares, I. M., S. S. Witkin, S. D. Miranda, A. M. Fonseca, J. A. Pinotti *et al.*, 2001 Differentiation between women with vulvovaginal symptoms who are positive or negative for Candida species by culture. Infect. Dis. Obstet. Gynecol. 9: 221–225.

Liu, M.-B., S.-R. Xu, Y. He, G.-H. Deng, H.-F. Sheng *et al.*, 2013 Diverse vaginal microbiomes in reproductive-age women with vulvovaginal candidiasis. PLoS One 8: e79812–e79812.

Livengood, C. H. 3rd, J. L. Thomason, and G. B. Hill, 1990 Bacterial vaginosis: diagnostic and pathogenetic findings during topical clindamycin therapy. Am. J. Obstet. Gynecol. 163: 515–520.

Lockhart, S. R., S. A. Messer, M. A. Pfaller, and D. J. Diekema, 2008 Lodderomyces elongisporus masquerading as Candida parapsilosis as a cause of bloodstream infections. J. Clin. Microbiol. 46: 374–376.

Long, S., Y. Yang, C. Shen, Y. Wang, A. Deng *et al.*, 2020 Metaproteomics characterizes human gut microbiome function in colorectal cancer. npj Biofilms Microbiomes 6: 14.

Lorek, J., S. Pöggeler, M. R. Weide, R. Breves, and D. P. Bockmühl, 2008 Influence of farnesol on the morphogenesis of Aspergillus niger. J. Basic Microbiol. 48: 99–103.

Lourenço, A., N. A. Pedro, S. B. Salazar, and N. P. Mira, 2018 Effect of Acetic Acid and Lactic Acid at Low pH in Growth and Azole Resistance of Candida albicans and Candida glabrata. Front. Microbiol. 9: 3265.

Low, N., N. Broutet, Y. Adu-Sarkodie, P. Barton, M. Hossain, and S. Hawkes, 2006 Global control of sexually transmitted infections. Lancet 368: 2001–16.

Lynch, M. E., and J. D.Sobel, 1994 Comparative in vitro activity of antimycotic agents against pathogenic vaginal yeast isolates. J. Med. Vet. Mycol. bi-monthly Publ. Int. Soc. Hum. Anim. Mycol. 32: 267–274.

Lynch, A. S., and G. T. Robertson, 2008 Bacterial and fungal biofilm infections. Annu. Rev. Med. 59: 415–428.

Ma, B., L. J. Forney, and J. Ravel, 2012 Vaginal microbiome: rethinking health and disease. Annu. Rev. Microbiol. 66: 371–389.

Ma, X., C. Li, J, Hou, and Y. Gu, 2017 Isolation and identification of culturable fungi from the genitals and semen of healthy giant pandas (Ailuropoda melanoleuca). BMC Vet. Res. 13: 344.

Ma, B., M. T. France, J. Crabtree, J. B. Holm, M. S. Humphrys *et al.*, 2020 A comprehensive non-redundant gene catalog reveals extensive within-community intraspecies diversity in the human vagina. Nat. Commun. 11:1–13.

Macdonald, F., and F. C. Odds, 1983. Virulence for mice of a proteinase-secreting strain of Candida albicans and a proteinase-deficient mutant. J. Gen. Microbiol. 129: 431-438.

Macura, A. B., and M. Skóra, 2012 [Fungi isolated from the vagina and their susceptibility to antifungals]. Ginekol. Pol. 83: 433–438.

Mahmoudi Rad, M., S. Zafarghandi, B. Abbasabadi, and M. Tavallaee, 2011 The epidemiology of Candida species associated with vulvovaginal candidiasis in an Iranian patient population. Eur. J. Obstet. Gynecol. Reprod. Biol. 155: 199–203.

Mahmoudi Rad, M., A. Zafarghandi, M. Zabihi, M. Tavallaee, and Y. Mirdamadi, 2012 Identification of Candida Species Associated with Vulvovaginal Candidiasis by Multiplex PCR. Infect. Dis. Obstet. Gynecol. 872169.

Majumdar, T., J. Mullick, R. Bir, J. Roy, and S. Sil, 2016 Determination of virulence factors and biofilm formation among isolates of vulvovaginal candidiasis. J. Med. Sci. 36:53.

Makela, P., D. Leaman, and J. D. Sobel, 2003 Vulvovaginal trichosporonosis. Infect. Dis. Obstet. Gynecol. 11: 131–133.

Månberg, A., F. Bradley, U. Qundos, B. L. Guthrie, K. Birse *et al.*, 2019 A High-throughput Bead-based Affinity Assay Enables Analysis of Genital Protein Signatures in Women At Risk of HIV Infection. Mol. Cell. Proteomics 18: 461–476.

Mao, L., and J. Franke, 2015 Symbiosis, dysbiosis, and rebiosis-The value of metaproteomics in human microbiome monitoring. Proteomics 15: 1142–1151.

Maphanga, T. G., M. Birkhead, J. F. Muñoz, M. Allam, T. G. Zulu *et al.*, 2020 Human Blastomycosis in South Africa Caused by Blastomyces percursus and Blastomyces emzantsi sp. nov., 1967 to 2014. J. Clin. Microbiol. 58.

Martin, S. W., J. B. Konopka, 2004 Lipid raft polarization contributes to hyphal growth in Candida albicans. Eukaryot. Cell 3: 675–684.

Martin, D. H., and J. M. Marrazzo, 2016 The Vaginal Microbiome: Current Understanding and Future Directions. J. Infect. Dis. 214: S36–S41.

Martin, H. L., B. A. Richardson, P. M. Nyange, L. Lavreys, S. L. Hillier *et al.*, 1999 Vaginal lactobacilli, microbial flora, and risk of human immunodeficiency virus type 1 and sexually transmitted disease acquisition. J. Infect. Dis. 180: 1863–1868.

Martinez, L. R., and A. Casadevall, 2005 Specific antibody can prevent fungal biofilm formation and this effect correlates with protective efficacy. Infect. Immun. 73: 6350–6362.

Masson, L., K. Mlisana, F. Little, L. Werner, N. N. Mkhize *et al.*, 2014 Defining genital tract cytokine signatures of sexually transmitted infections and bacterial vaginosis in women at high risk of HIV infection: a cross-sectional study. Sex. Transm. Infect. 90: 580–587.

Masson, L., J.-A. S. Passmore, L. J. Liebenberg, L. Werner, C. Baxter *et al.*, 2015 Genital inflammation and the risk of HIV acquisition in women. Clin. Infect. Dis. 61: 260–269.

Masson, L., S. Barnabas, J. Deese, K. Lennard, S. Dabee *et al.*, 2019 Inflammatory cytokine biomarkers of asymptomatic sexually transmitted infections and vaginal dysbiosis: a multicentre validation study. Sex. Transm. Infect. 95: 5–12.

Mastromarino, P., P. Brigidi, S. Macchia, L. Maggi, F. Pirovano *et al.*, 2002 Characterization and selection of vaginal Lactobacillus strains for the preparation of vaginal tablets. J. Appl. Microbiol. 93: 884–893.

Mayer, M. P., and B. Bukau, 2005 Hsp70 chaperones: cellular functions and molecular mechanism. Cell. Mol. Life Sci. 62: 670–684.

Mayer, F. L., D. Wilson, and B. Hube, 2013 Candida albicans pathogenicity mechanisms. Virulence 4: 119–128.

McClelland, R., B. Richardson, W. Hassan, S. Graham, J. Kiarie *et al.*, 2009 A Prospective Study of Vaginal Bacterial Flora and Other Risk Factors for Vulvovaginal Candidiasis. J. Infect. Dis. 199: 1883–1890.

McClelland, R. S., J. E. Balkus, J. Lee, O. Anzala, J. Kimani *et al.*, 2015 Randomized Trial of Periodic Presumptive Treatment With High-Dose Intravaginal Metronidazole and Miconazole to Prevent Vaginal Infections in HIV-negative Women. J. Infect. Dis. 211: 1875–1882.

Mcilroy, J., G. Ianiro G., I. Mukhopadhya, R. Hansen, and G. L. Hold, 2018 Review article: the gut microbiome in inflammatory bowel disease—avenues for microbial management. Aliment. Pharmacol. Ther. 47: 26–42.

McKinnon, L. R., L. J. Liebenberg, N. Yende-Zuma, D. Archary, S. Ngcapu *et al.*, 2018 Genital inflammation undermines the effectiveness of tenofovir gel in preventing HIV acquisition in women. Nat. Med. 24: 491—496.

McMillan, A., S. Rulisa, M. Sumarah, J. M. Macklaim, J. Renaud *et al.*, 2015 A multi-platform metabolomics approach identifies highly specific biomarkers of bacterial diversity in the vagina of pregnant and non-pregnant women. Sci. Rep. 5: 14174.

Meizoso, T., T. Rivera, M. J. Fernández-Aceñero, M. J. Mestre, M. Garrido *et al.*, 2008 Intrauterine candidiasis: report of four cases. Arch. Gynecol. Obstet. 278: 173–176.

Merenstein, D., H. Hu, C. Wang, P. Hamilton, M. Blackmon et al., 2013 Colonization by Candida species of the oral and vaginal mucosa in HIV-infected and noninfected women. AIDS Res. Hum. Retroviruses 29: 30–34.

Miceli, M. H., J. A. Díaz, S. A. Lee 2011 Emerging opportunistic yeast infections. Lancet Infect. Dis. 11:142–51.

Miller, L., D. L. Patton, A. Meier, S. S. Thwin, T. M. Hooton *et al.*, 2000 Depomedroxyprogesterone-induced hypoestrogenism and changes in vaginal flora and epithelium. Obstet. Gynecol. 96:431–9.

Miranda-Zapico, I., E. Eraso, J. L. Hernández-Almaraz, L. M. López-Soria, A. J. Carrillo-Muñoz *et al.*, 2011 Prevalence and antifungal susceptibility patterns of new cryptic species inside the species complexes Candida parapsilosis and Candida glabrata among blood isolates from a Spanish tertiary hospital. J. Antimicrob. Chemother. 66: 2315–2322.

Mirmonsef, P., D. Gilbert, M. R. Zariffard, B. R. Hamaker, A. Kaur *et al.*, 2011 The Effects of Commensal Bacteria on Innate Immune Responses in the Female Genital Tract. Am. J. Reprod. Immunol. 65: 190–195.

Mirza, S. A., M. Phelan, D. Rimland, E. Graviss, R. Hamill *et al.*, 2003 The Changing Epidemiology of Cryptococcosis: An Update from Population-Based Active Surveillance in 2 Large Metropolitan Areas, 1992–2000. Clin. Infect. Dis. 36: 789–794.

Mitchell, C., and J. Marrazzo, 2014 Bacterial vaginosis and the cervicovaginal immune response. Am. J. Reprod. Immunol. 71(6):555–563.

Mlisana, K., N. naicker, L. Werner, L. Roberts, F. van Loggerenberg *et al.*, 2012 Symptomatic vaginal discharge is a poor predictor of sexually transmitted infections and genital tract inflammation in high-risk women in South africa. J Infect. Dis. 206:6–14.

Mølgaard-Nielsen, D., H. Svanström, M. Melbye, A. Hviid, and B. Pasternak, 2016 Association Between Use of Oral Fluconazole During Pregnancy and Risk of Spontaneous Abortion and Stillbirth. JAMA 315: 58–67.

Moodley, P., C. Connolly, and A. W. Sturm, 2002 Interrelationships among human immunodeficiency virus type 1 infection, bacterial vaginosis, trichomoniasis, and the presence of yeasts. J. Infect. Dis. 185: 69–73.

Moraes, I. A., J. S. P. Stussi, W. Lilenbaum, A. Pissinatti, F. P. Luz *et al.*, 2004 Isolation and identification of fungi from vaginal flora in three species of captive Leontopithecus. Am. J. Primatol. 64: 337–343.

Morales, D. K., and D. A. Hogan, 2010 Candida albicans Interactions with Bacteria in the Context of Human Health and Disease. PLOS Pathog. 6: e1000886.

Morrison, C., R. N. Fichorova, C. Mauck, P.-L. Chen, C. Kwok *et al.*, 2014 Cervical Inflammation and Immunity Associated With Hormonal Contraception, Pregnancy, and HIV-1 Seroconversion. JAIDS J. Acquir. Immune Defic. Syndr. 66.

Mostovenko, E., C. Hassan, J. Rattke, A. M. Deelder, P. A. van Veelen *et al.*, 2013 Comparison of peptide and protein fractionation methods in proteomics. EuPA Open Proteomics 1: 30–37.

Mouzin, E. L., and M. A. Seilke, 1996 Female Genital Blastomycosis: Case Report and Review. Clin. Infect. Dis. 22: 718–719.

Mulder, M., D. Radjabzadeh, R. J. Hassing, J. Heeringa, A. G. Uitterlinden *et al.*, 2019 The effect of antimicrobial drug use on the composition of the genitourinary microbiota in an elderly population. BMC Microbiol. 19:9.

Muñoz, J. F., G. M. Gauthier, C. A. Desjardins, J. E. Gallo, J. Holder *et al.*, 2015 The Dynamic Genome and Transcriptome of the Human Fungal Pathogen Blastomyces and Close Relative Emmonsia. PLOS Genet. 11: e1005493.

Muth, T., C. A. Kolmeder, J. Salojärvi, S. Keskitalo, M. Varjosalo *et al.*, 2015 Navigating through metaproteomics data: A logbook of database searching. Proteomics 15: 3439–3453.

Naglik, J. R., D. L. Moyes, B. Wachtler, and B. Hube, 2011 Candida albicans interactions with epithelial cells and mucosal immunity. Microbes Infect. 13:963-976.

Naglik, J. R., 2014 Candida albicans pathogenecity and epithelial immunity. PLoS Pathog. 10:1371.

Nash, A. K., T. A. Auchtung, M. C. Wong, D. P. Smith, J. R. Gesell *et al.*, 2017 The gut mycobiome of the Human Microbiome Project healthy cohort. Microbiome 5: 153.

Nath, R., P. Sargiary, B. Borkakoty, and P. Parida, 2017 Cutaneotrichosporon (Trichosporon) debeurmannianum: A Rare Yeast Isolated from Blood and Urine Samples. Mycopathologia 183.

Neder, R. N., 1992 Microbiologia: manual de laboratório. Nobel.

Neely, A. N., E. J. Law, and I. A. Holder, 1986 Increased susceptibility to lethal Candida infections in burned mice preinfected with Pseudomonas aeruginosa or pretreated with proteolytic enzymes. Infect. Immun. 52: 200–204.

Nejat, Z., S. Farahyar, M. Falahati, M. Ashrafi Khozani, A. F. Hosseini *et al.*, 2017 Molecular Identification and Antifungal Susceptibility Pattern of Non-albicans Candida Species Isolated from Vulvovaginal Candidiasis. Iran. Biomed. J. 22: 33–41.

Nemes-Nikodém, É., A. Brunner, D. Pintér, N. Mihalik, G. Lengyel *et al.*, 2014 Antimicrobial susceptibility and genotyping analysis of Hungarian Neisseria gonorrhoeae strains in 2013. Acta Microbiol. Immunol. Hung. 61: 435–445.

Nesvizhskii, A. I., 2014 Proteogenomics: concepts, applications and computational strategies. Nat. Methods 11: 1114–1125.

Nett, J. E., 2016 The Host's Reply to Candida Biofilm. Pathog. 5:1.

Nguyen, L. D. N., E. Viscogliosi, and L. Delhaes, 2015 The lung mycobiome: an emerging field of the human respiratory microbiome . Front. Microbiol. 6: 89.

Nicholls, S., D. M. MacCallum, F. A. R. Kaffarnik, L. Selway, S. C. Peck *et al.*, 2011 Activation of the heat shock transcription factor Hsf1 is essential for the full virulence of the fungal pathogen Candida albicans. Fungal Genet. Biol. 48: 297–305.

Nilsson, R. H., M. Ryberg, E. Kristiansson, K. Abarenkov, K.-H. Larsson *et al.*, 2006 Taxonomic Reliability of DNA Sequences in Public Sequence Databases: A Fungal Perspective. PLoS One 1: e59.

Nilsson, R. H., S. Anslan, M. Bahram, C. Wurzbacher, P. Baldrian *et al.*, 2019 Mycobiome diversity: high-throughput sequencing and identification of fungi. Nat. Rev. Microbiol. 17: 95–109.

Nobile, C. J., and A. D. Johnson, 2015 Candida albicans Biofilms and Human Disease. Annu. Rev. Microbiol. 69: 71–92.

Nogueira, F., S. Sharghi, K. Kuchler, and T. Lion, 2019 Pathogenetic Impact of Bacterial-Fungal Interactions. Microorganisms 7: 459.

Noverr, M. C., and G. B. Huffnagle, 2004 Regulation of Candida albicans morphogenesis by fatty acid metabolites. Infect. Immun. 72: 6206–6210.

Noverr, M. C., R. M. Noggle, G. B. Toews, and G. B. Huffnagle, 2004 Role of Antibiotics and Fungal Microbiota in Driving Pulmonary Allergic Responses. Infect. Immun. 72: 4996– 5003.

Nowakowska, D., A. Kurnatowska, B. Stray-Pedersen, and J. Wilczynski, 2004 Relative abundance of fungi in the vagina, rectum and oral cavity in pregnant diabetic women: Relation to gestational age and symptoms. Acta Obstet. Gynecol. Scand. 83: 251–256.

Noyes, N., K. C. Cho, J. Ravel, L. J. Forney, and Z. Abdo, 2018 Associations between sexual habits, menstrual hygiene practices, demographics and the vaginal microbiome as revealed by Bayesian network analysis. PLoS One 13: e0191625.

Nucci, M., and E. Anaissie, 2007 Fusarium infections in immunocompromised patients. Clin. Microbiol. Rev. 20: 695–704.

Nugent, R. P., M. A. Krohn, and S. L. Hillierm 1991 Reliability of diagnosing bacterial vaginosis is improved by a stan- dardized method of gram stain interpretation. J Clin Microbiol. 29: 297–301.

Odds, F. C., and R. Bernaerts, 1994 CHROMagar Candida, a new differential isolation medium for presumptive identification of clinically important Candida species. J. Clin. Microbiol. 32: 1923–1929.

Oever, J. T., and M. G. Netea, 2014 The bacteriome-mycobiome interaction and antifungal host defense. Eur. J. Immunol. 44: 3182–3191.

O'Hanlon, D. E., T. R. Moench, and R. A. Cone, 2013 Vaginal pH and Microbicidal Lactic Acid When Lactobacilli Dominate the Microbiota. PLoS One 8: e80074.

O'Toole, G. A., 2016 Classic Spotlight: Quorum Sensing and the Multicellular Life of Unicellular Organisms. J. Bacteriol. 198: 601.

Oakley, B. B., T. L. Fiedler, J. M. Marrazzo, and D. N. Fredricks, 2008 Diversity of human vaginal bacterial communities and associations with clinically defined bacterial vaginosis. Appl. Environ. Microbiol. 74:4898-4909.

Oh, B. H., Y. C. Song, Y. W. Lee, Y. B. Choe, and K. J. Ahn, 2009 Comparison of Nested PCR and RFLP for Identification and Classification of Malassezia Yeasts from Healthy Human Skin. Ann. Dermatol. 21: 352–357.

Oksanen, F. J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre *et al.*, 2017 Vegan: Community Ecology Package. R Package Version 2.4-3. Available online at: https://CRAN.R-project.org/package=vegan

Oliveros, J. C., 2016 "Venny 2.1.0.," Venny. An Interactive Tool for Comparing Lists with Venn's Diagrams. (2007-2015). Available online at: http://bioinfogp.cnb.csic.es/tools/venny/.

Onderdonk, A. B., M. L. Delaney, and R. N. Fichorova, 2016 The Human Microbiome during Bacterial Vaginosis. Clin. Microbiol. Rev. 29: 223–238.

Ott, S. J., T. Kühbacher, M. Musfeldt, P. Rosenstiel, S. Hellmig *et al.*, 2008 Fungi and inflammatory bowel diseases: Alterations of composition and diversity. Scand. J. Gastroenterol. 43: 831–841.

Ozcan, K., M. Ilkit, A. Ates, A. Turac-Bicer, and H. Demirhindi, 2010 Performance of Chromogenic Candida Agar and CHROMagar Candida in recovery and presumptive identification of monofungal and polyfungal vaginal isolates. Med. Mycol. 48: 29–34.

Papenfort, K., and B. L. Bassler, 2016 Quorum sensing signal-response systems in Gram-negative bacteria. Nat. Rev. Microbiol. 14: 576–588.

Parolin, C., A. Marangoni, L. Laghi, C. Foschi, R. A. Ñahui Palomino *et al.*, 2015 Isolation of vaginal lactobacilli and characterization of anti-Candida activity. PLoS ONE 10: e0131220.

Paulitsch, A., W. Weger, G. Ginter-Hanselmayer, E. Marth, and W. Buzina, 2006 A 5-year (2000–2004) epidemiological survey of Candida and non-Candida yeast species causing vulvovaginal candidiasis in Graz, Austria. Mycoses 49: 471–475.

Paulitz, T. C., R. W. Smiley, and R. J. Cook, 2002 Insights into the prevalence and management of soilborne cereal pathogens under direct seeding in the Pacific Northwest, U.S.A. Can. J. Plant Pathol. 24: 416–428.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion *et al.*, 2012 Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12.

Pendharkar, S., E. Brandsborg, L. Hammarström, H. Marcotte, and P.-G. Larsson, 2015 Vaginal colonisation by probiotic lactobacilli and clinical outcome in women conventionally treated for bacterial vaginosis and yeast infection. BMC Infect. Dis. 15: 255.

Perfect, J. R., and A. Casadevall, 2006 A Fungal molecular pathogenesis: what can it do and why do we need it? In Molecular Principles of Fungal Pathogenesis. Edited by Heitman J, Filler SG, Edwards JE, Mitchell AP. Washington DC: ASM Press 3-12.

Perlroth, J., B. Choi B and B. Spellberg, 2007 Nosocomial fungal infections: epidemiology, diagnosis, and treatment. Med Mycol 45: 321–346.

Peters, B. M., J. Yano M. C. Noverr, and P. L. Fidel Jr, 2014a Candida vaginitis: when opportunism knocks, the host responds. PLoS Pathog. 10:e1003965.

Peters, R. P. H., J. H. Dubbink, L. van der Eem, S. P. Verweij, M. L. A. Bos *et al.*, 2014b Cross-sectional study of genital, rectal, and pharyngeal Chlamydia and gonorrhea in women in rural South Africa. Sex. Transm. Dis. 41: 564—569.

Peters, B. M., G. E. Palmer, A. K. Nash, E. A. Lilly, P. L. Fidel Jr *et al.*, 2014c Fungal morphogenetic pathways are required for the hallmark inflammatory response during Candida albicans vaginitis. Infect. Immun. 82: 532–543.

Pfaller, M. A., and D. J. Diekema, 2004 Rare and emerging opportunistic fungal pathogens: concern for resistance beyond Candida albicans and Aspergillus fumigatus. J. Clin. Microbiol. 42: 4419–4431.

Pflughoeft, K. J., and J. Versalovic, 2012 Human microbiome in health and disease. Annu. Rev. Pathol. 7: 99–122.

Piancastelli, C., F. Ghidini, G. Donofrio, S. Jottini, S. Taddei *et al.*, 2009 Isolation and characterization of a strain of Lichtheimia corymbifera (ex Absidia corymbifera) from a case of bovine abortion. Reprod. Biol. Endocrinol. 7: 138.

Pirotta, M. V., J. M. Gunn, and P. Chondros, 2003 “Not thrush again!” Women’s experience of post-antibiotic vulvovaginitis. Med. J. Aust. 179: 43–46.

Pirotta, M. V., and S. M. Garland, 2006 Genital Candida species detected in samples from women in Melbourne, Australia, before and after treatment with antibiotics. J Clin. Microbiol. 44:3213–3217.

Pitt, S. S., R. S. Garfein, C. A. Gaydos, S. A. Strathdee, S. G Sherman *et al.*, 2005 Prevalence and correlates of Chlamydia trachomatis, Neisseria gonorrhoeae, Trichomonas vaginalis infections, and bacterial vaginosis among a cohort of young injection drug users in Baltimore, Maryland. Sex. Transm. Dis. 32:446– 453.

Plummer, E. L., L. A. Vodstrcil, C. K. Fairley, S. N. Tabrizi, S. M. Garland *et al.*, 2019 Sexual practices have a significant impact on the vaginal microbiota of women who have sex with women. Sci. Rep. 9:19749.

Polke, M., and I. D. Jacobsen, 2017 Quorum sensing by farnesol revisited. Curr. Genet. 63: 791–797.

Polvi, E. J., X. Li, T. R. O’Meara, M. D. Leach, and L. E. Cowe, 2015 Opportunistic yeast pathogens: reservoirs, virulence mechanisms, and therapeutic strategies. Cell Mol. Life Sci. 72:2261–87.

Potgieter, T., A. J. M. Nel, D. L. Tabb, S. Fortuin, S. Garnett et al. 2019 MetaNovo: a probabilistic approach to peptide and polymorphism discovery in complex mass spectrometry datasets. bioRxiv: 605550.

Qin, J., R. Li, J. Raes, M. Arumugam, K. S. Burgdorf *et al.*, 2010 A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464: 59-65.

Rabe, A., M. Gesell Salazar, S. Michalik, S. Fuchs, A. Welk *et al.*, 2019 Metaproteomics analysis of microbial diversity of human saliva and tongue dorsum in young healthy individuals. J. Oral Microbiol. 11: 1654786.

Raimondi, S., A. Amaretti, C. Gozzoli, M. Simone, L. Righini *et al.*, 2019 Longitudinal survey of fungi in the human gut: ITS profiling, phenotyping, and colonization. Front. Microbiol. 10: 1–12.

Rajput, S. B., and S. M. Karuppayil ,2013 Small molecules inhibit growth, viability and ergosterol biosynthesis in Candida albicans. SpringerPlus 2:26.

Ram, R. J., N. C. Verberkmoes, M. P. Thelen, G. W. Tyson, B. J. Baker *et al.*, 2005 Community Proteomics of a Natural Microbial Biofilm. Science 308: 1915–1920.

Rand, K. H., and M. Tillan 2006 Errors in interpretation of Gram stains from positive blood cultures. Am. J. Clin. Pathol. 126: 686–690.

Rane, H. S., S. Hardison, C. Botelho, S. M. Bernardo, F. J. Wormley *et al.*, 2014 Candida albicans VPS4 contributes differentially to epithelial and mucosal pathogenesis. Virulence 5: 810–818.

Rangel, L. I., R. E. Spanner, M. K. Ebert, S. J. Pethybridge, E. H. Stukenbrock *et al.*, 2020 Cercospora beticola: The intoxicating lifestyle of the leaf spot pathogen of sugar beet. Mol. Plant Pathol. 21: 1020–1041.

Rao, C. Y., C. Kurukularatne, J. B. Garcia-Diaz, S. A. Kemmerly, D. Reed *et al.*, 2007 Implications of Detecting the Mold Syncephalastrum in Clinical Specimens of New Orleans Residents After Hurricanes Katrina and Rita. J. Occup. Environ. Med. 49.

Ravel, J., P. Gajer, Z. Abdo, G. M. Schneider, S. S. K. Koenig *et al.*, 2011 Vaginal microbiome of reproductive-age women. Proc. Natl. Acad. Sci. 108: 4680 LP – 4687.

Ravel, J., and R. M. Brotman, 2016 Translating the vaginal microbiome: gaps and challenges. Genome Med. 8: 35.

Ray, A., S. Ray, A. T. George, and N. Swaminathan, 2011 Interventions for prevention and treatment of vulvovaginal candidiasis in women with HIV infection. Cochrane database Syst. Rev. 8: CD008739.

Rechenberger, J., P. Samaras, A. Jarzab, J. Behr, M. Frejno *et al.*, 2019 Challenges in Clinical Metaproteomics Highlighted by the Analysis of Acute Leukemia Patients with Gut Colonization by Multidrug-Resistant Enterobacteriaceae.

Reid, G., J. A. McGroarty, R. Angotti, and R. L. Cook, 1988 Lactobacillus inhibitor production against escherichia coli and coaggregation ability with uropathogens. Can. J. Microbiol. 34(3):344–51.

Richter, S. S., R. P. Galask, S. A. Messer, R. J. Hollis, D. J. Diekema *et al.*, 2005 Antifungal susceptibilities of Candida species causing vulvovaginitis and epidemiology of recurrent cases. J. Clin. Microbiol. 43: 2155–62.

Ridgway, G. L., and J. D. Oriel, 1977 Interrelationship of Chlamydia trachomatis and other pathogens in the female genital tract. J. Clin. Pathol. 30: 933–936.

Rio, R. V. M., 2017 Don't Bite the Hand that Feeds You. Cell Host Microbe 21: 552–554.

Rizzetto, L., C. Filippo, and D. Cavalieri, 2014 Richness and diversity of mammalian fungal communities shape innate and adaptive immunity in health and disease: Highlights. Eur. J. Immunol. 44.

Román, E., I. Correia, D. Priet, R. Alonso, and J. Pla, 2020 The HOG MAPK pathway in Candida albicans: more than an osmosensing pathway. Int. Microbiol. Off. J. Spanish Soc. Microbiol. 23: 23–29.

Romani, L., 2011 Immunity to fungal infections. Nat. Rev. Immunol. 11:275-88.

Romoren, M., J. Sundby, M. Velauthapillai, M. Rahman, E. Klouman *et al.*, 2007 Chlamydia and gonorrhoea in pregnant Batswana women: time to discard the syndromic approach? BMC Infect. Dis. 7: 27.

Rooijers, K., C. Kolmeder, C. Juste, J. Doré, M. de Been *et al.*, 2011 An iterative workflow for mining the human intestinal metaproteome. BMC Genomics 12: 6.

Rühl, M., C. Fischer, and U. Kües, 2008 Ligninolytic enzyme activities alternate with mushroom production during industrial cultivation of Pleurotus ostreatus on wheat straw-based substrate. Curr. Trends Biotechnol. Pharm. 2: 478–492

Ruijter, G. J., and J. Visser, 1997 Carbon repression in Aspergilli. FEMS Microbiol. Lett. 151: 103–114.

Sachin, C. D., K. Ruchi, and S. Santosh, 2012 In vitro evaluation of proteinase, phospholipase and haemolysin activities of Candida species isolated from clinical specimens. Int. J. Med. Biomed. Res. 1: 153-137.

Salter, S. J., M. J. Cox, E. M. Turek, S. T. Calus, W. O. Cookson *et al.*, 2014 Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 12: 87.

Salvato, F., R. L. Hettich, and M. Kleiner, 2021 Five key aspects of metaproteomics as a tool to understand functional interactions in host-associated microbiomes. PLOS Pathog. 17: e1009245.

Sam, Q. H., M. W. Chang, and L. Y. A. Chai, 2017 The Fungal Mycobiome and Its Interaction with Gut Bacteria in the Host. Int. J. Mol. Sci. 18:330.

Sanglard, D., 2016 Emerging Threats in Antifungal-Resistant Fungal Pathogens. Front Med. (Lausanne) 3:11.

Sansonetti, P. J., 2011 To be or not to be a pathogen: that is the mucosally relevant question. Mucosal Immunol. 4: 8–14.

Satoh, S., and E. Boyer, 2019 HIV in South Africa. Lancet 394: 467.

Scanlan, P. D., and J. R. Marchesi, 2008 Micro-eukaryotic diversity of the human distal gut microbiota: qualitative assessment using culture-dependent and -independent analysis of faeces. ISME J. 2: 1183–1193.

Schneider, T., and K. Riedel, 2010 Environmental proteomics: analysis of structure and function of microbial communities. Proteomics 10:785–798.

Schwartz, I. S., T. H. Boyles, C. R. Kenyon, J. C. Hoving, G. D. Brown *et al.*, 2019 The estimated burden of fungal disease in South Africa. South African Med. J. 109: 885.

Schwebke, J. R., 2001 Role of Vaginal Flora As a Barrier to HIV Acquisition. Curr. Infect. Dis. Rep. 3: 152–155.

Schwebke, J. R., and L. F. Lawing, 2001 Prevalence of Mobiluncus spp among women with and without bacterial vaginosis as detected by polymerase chain reaction. Sex. Transm. Dis. 28:195–199.

Schwebke, J. R., 2003 Gynecologic consequences of bacterial vaginosis. Obstet. Gynecol. Clin. North. Am. 30:685–94.

Scott, V., J. Tang., T. Drell, J. Simm, A. Salumets *et al.*, 2018 PD19-01 Evaluation of the Vaginal Mycobiome in Asymptomatic Pre-menopausal Women. J. Urol. 199: e394–e395.

Sebastiana, M., A. Corrêa, P. Castro, and M. Ramos 2020, Pisolithus pp. 707–726 in *Beneficial Microbes in Agro-Ecology*, edited by N. Amaresan, M. Senthil Kumar, K. Annapurna, K. Kumar, and A. Sankaranarayanan. Academic Press, London.

Seed, P. C., 2014 The human mycobiome. Cold Spring Harb. Perspect. Med. 5: a019810–a019810.

Segretain, G., 1959 [Penicillium marneffei n.sp., agent of a mycosis of the reticuloendothelial system]. Mycopathol. Mycol. Appl. 11: 327–353.

Semighini, C. P., N. Murray, AND S. D. Harris, 2008 Inhibition of Fusarium graminearum growth and development by farnesol. FEMS Microbiol. Lett. 279: 259–264.

Serrano, M. G., H. I. Parikh, J. P. Brooks, D. J. Edwards, T. J. Arodz *et al.*, 2019 Racioethnic diversity in the dynamics of the vaginal microbiome during pregnancy. Nat. Med. 25: 1001–1011.

Severo Gomes, B., C. M. Souza Motta, A. N. Lima, and A. Porto, 2010 Pathogenic characteristics of yeasts isolated from vaginal secretion preserved under mineral oil. J. Venom. Anim. Toxins Incl. Trop. Dis. 17: 460–466.

Sha, B. E., H. Y. Chen, Q. J. Wang, M. R. Zariffard, M. H. Cohen *et al.*, 2005 Utility of Amsel criteria, Nugent score, and quantitative PCR for Gardnerella vaginalis, Mycoplasma hominis, and Lactobacillus spp. for diagnosis of bacterial vaginosis in human immunodeficiency virus-infected women. J. Clin. Microbiol. 43: 4607–4612.

Shankar, J., N. V. Solis, S. Mounaud, S. Szpakowski, H. Liu *et al.*, 2015. Using Bayesian Modelling to Investigate Factors Governing Antibiotic-induced Candida Albicans Colonization of the GI Tract. Sci Rep. 2015;5:8131.

Shankar, J., 2017 Insights into study design and statistical analyses in translational microbiome studies. Ann. Transl. Med. 5.

Shevchenko, A., H. Tomas, J. Havli, J. V. Olsen, AND M. Mann, 2006 In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat. Protoc. 1: 2856–2860.

Shin, L. Y., and R. Kaul, 2008 Stay It with Flora: Maintaining Vaginal Health as a Possible Avenue for Prevention of Human Immunodeficiency Virus Acquisition. J. Infect. Dis. 197: 1355–1357.

Siggins, A., E. Gunnigle, F. Abram, 2012 Exploring mixed microbial community functioning: Recent advances in metaproteomics. FEMS Microbiol. Ecol. 80: 265–280.

Sigler, L., 1996 Ajellomyces crescens sp. nov., taxonomy of Emmonsia spp., and relatedness with Blastomyces dermatitidis (teleomorph Ajellomyces dermatitidis). J. Med. Vet. Mycol. 34: 303–314.

Simon, V., D. D. Ho, and Q. Abdool Karim, 2006 HIV/AIDS epidemiology, pathogenesis, prevention, and treatment. Lancet (London, England) 368: 489–504.

Singh, S., J. D. Sobel, P. Bhargava, D. Boikov, and J. A. Vazquez, 2002 Vaginitis Due to Candida krusei: Epidemiology, Clinical Aspects, and Therapy. Clin. Infect. Dis. 35: 1066–1070.

Singh, R. P., and C. R. K. Reddy, 2016 Unraveling the Functions of the Macroalgal Microbiome . Front. Microbiol. 6: 1488.

Sirisanthana, T., and K. Supparatpinyo, 2012 Infection due to Penicillium marneffei, pp. 389–395 in *Sande's HIV/AIDs Medicine* edited by P. A. Volberding, W. C. Greene, J. M. A. Lange, J. E. Gallant, and N. Sewankambo. Elsevier, London.

Sisti, G., T. T. Kanninen, I. Ramer, and S. S. Witkin, 2015 Interaction between the inducible 70-kDa heat shock protein and autophagy: effects on fertility and pregnancy. Cell Stress Chaperones 20: 753–758.

Smith, M. B., V. J. Schnadig, P. Zaharopoulos, and C. Van Hook, 1997 Disseminated Histoplasma capsulatum infection presenting as genital ulcerations. Obstet. Gynecol. 89: 842–844.

Smith, S. B., J. Ravel, 2017 The vaginal microbiota, host defence and reproductive physiology. J. Physiol. 595: 451–463.

Smyth, G. K., 2005 limma: Linear Models for Microarray Data. Bioinforma. Comput. Biol. Solut. Using R Bioconductor: 397–420.

Sobel, J. D., 1985 Epidemiology and pathogenesis of recurrent vulvovaginal candidiasis. Am J Obstet Gynecol. 152:924–35.

Sobel, J. D., 1992 Pathogenesis and Treatment of Recurrent Vulvovaginal Candidiasis. Clin. Infect. Dis. 14: S148–S153.

Sobel, J. D., J. Vazquez, M. Lynch, C. Meriwether, and M. J. Zervos, 1993 Vaginitis due to Saccharomyces cerevisiae: epidemiology, clinical aspects, and therapy. Clin. Infect. Dis. an Off. Publ. Infect. Dis. Soc. Am. 16: 93–99.

Sobel, J. D., 1997 Vaginitis. N Engl. J Med. 337:1896–903.

Sobel, J. D., S. Faro, R. W. Force, B. Foxman, W. J. Ledger *et al.*, 1998 Vulvovaginal candidiasis: epidemiologic, diagnostic, and therapeutic considerations. Am J Obstet Gynecol. 178:203–11.

Sobel, J. D., 2000. Bacterial vaginosis. Annu. Rev. Med. 51: 349–356.

Sobel, J. D., 2007 Vulvovaginal candidosis. Lancet 369: 1961–1971.

Sobel, J. D., C. Subramanian, B. Foxman, M. Fairfax, and S. E. Gygax, 2013 Mixed Vaginitis—More Than Coinfection and With Therapeutic Implications. Curr. Infect. Dis. Rep. 15: 104–108.

Sobel, J. D., and R. A. Akins, 2015 The role of PCR in the diagnosis of Candida vulvovaginitisea new gold standard? Curr. Infect. Dis. Rep. 17:488.

Sokol, H., V. Leducq, H. Aschard, H.-P. Pham, S. Jegou *et al.*, 2017 Fungal microbiota dysbiosis in IBD. Gut 66: 1039–1048.

Song, S. D., K. D. Acharya, J. E. Zhu, C. M. Deveney, M. R. S. Walther-Antonio *et al*., 2020 Daily Vaginal Microbiota Fluctuations Associated with Natural Hormonal Cycle, Contraceptives, Diet, and Exercise. mSphere 5: e00593-20.

Sonnenberg, G. F., L. A. Monticelli, T. Alenghat, T. C. Fung, N. A. Hutnick *et al.*, 2012 Innate lymphoid cells promote anatomical containment of lymphoid-resident commensal bacteria. Science 336: 1321–1325.

Soper, D. E., R. C. Bump, and W. G. Hurt, 1990 Bacterial vaginosis and trichomonssis vaginitis are risk factors for cuff cellulitis after abdominal hystectyomy. Am. J. Obstet. Gynecol. 163:1016–1021.

Soto, S. M., 2013 Role of efflux pumps in the antibiotic resistance of bacteria embedded in a biofilm. Virulence 4: 223–229.

Souza, P. C., A. Storti-Filho, R. J. S. Souza, E. Damke, I. C. J. Mello *et al.*, 2009 Prevalence of Candida sp. in the cervical–vaginal cytology stained by Harris–Shorr. Arch. Gynecol. Obstet. 279: 625–629.

Spinillo, A., E. Capuzzo, S. Nicola, F. Baltaro, A. Ferrari *et al.*, 1995 The impact of oral contraception on vulvovaginal candidiasis. Contraception 51: 293–297.

Spirin, V., V. Malysheva, and K.-H. Larsson, 2018 On some forgotten species of Exidia and Myxarium (Auriculariales, Basidiomycota). Nord. J. Bot. 36: e01601: 1-11.

Spurbeck, R. R., and C. G. Arvidson, 2008 Inhibition of neisseria gonorrhoeae epithelial cell interactions by vaginal lactobacillus species. Infect. Immun. 76(7):3124–3130.

Srinivasan, S., N. G. Hoffman, M. T. Morgan, F. A. Matsen, T. L. Fiedler *et al.*, 2012 Bacterial communities in women with bacterial vaginosis: high resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. PLoS One 7: e37818.

Steele, C., and P. L. Fidel Jr, 2002 Cytokine and chemokine production by human oral and vaginal epithelial cells in response to Candida albicans. Infect. Immun. 70: 577–583.

Steinbach, W. J., D. K. Benjamin Jr., D. P. Kontoyiannis, J. R. Perfect, I. Lutsar *et al.*, 2004 Infections Due to Aspergillus terreus: A Multicenter Retrospective Analysis of 83 Cases . Clin. Infect. Dis. 39: 192–198.

Stout, M. J., T. N. Wylie, H. Gula, A. Miller, and K. M. Wylie, 2020 The microbiome of the human female reproductive tract. Curr. Opin. Physiol. 13: 87–93.

Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert *et al.*, 2005 Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. 102: 15545 LP – 15550.

Subramanian, C., and J. D. Sobel, 2011 A case of Conidiobolus coronatus in the vagina. Med. Mycol. 49: 427–429.

Sudbery, P. E., 2011 Growth of Candida albicans hyphae. Nat. Rev. Microbiol. 9: 737–748.

Sugita, T., H. Suto, T. Unno, R. Tsuboi, H. Ogawa et al., 2001 Molecular analysis of Malassezia microflora on the skin of atopic dermatitis patients and healthy subjects. J. Clin. Microbiol. 39: 3486–3490.

Sugita, T., M. Takashima, N. Poonwan, N. Mekha, K. Malaithao *et al.*, 2003 The first isolation of ustilaginomycetous anamorphic yeasts, Pseudozyma species, from patients' blood and a description of two new species: P. parantarctica and P. thailandica. Microbiol. Immunol. 47: 183–190.

Suhr, M. J., N. Banjara, and H. E. Hallen-Adams, 2016 Sequence-based methods for detecting and evaluating the human gut mycobiome. Lett. Appl. Microbiol. 62: 209–215.

Taha, T. E., D. R. Hoover, G. A. Dallabetta, N. I. Kumwenda, L. A. Mtimavalye *et al.*, 1998 Bacterial vaginosis and disturbances of vaginal flora: association with increased acquisition of HIV. AIDS 12(13):1699–1706.

Taha, T. E., R. H. Gray, N. I. Kumwenda, D. R. Hoover, L. A. Mtimavalye *et al.*, 1999 HIV infection and disturbances of vaginal flora during pregnancy. J. Acquir. immune Defic. Syndr. Hum. retrovirology Off. Publ. Int. Retrovirology Assoc. 20: 52–59.

Tanca, A., A. Palomba, S. Pisanu, M. Deligios, C. Fraumene *et al.*, 2014 A straightforward and efficient analytical pipeline for metaproteome characterization. Microbiome 2: 49.

Tanca, A., A. Palomba, C. Fraumene, D. Pagnozzi, V. Manghina *et al.*, 2016 The impact of sequence database choice on metaproteomic results in gut microbiota studies. Microbiome 4: 51.

Tanca, A., M. Abbondio, A. Palomba, C. Fraumene, V. Manghina *et al*., 2017 Potential and active functions in the gut microbiota of a healthy human cohort. Microbiome 5: 79.

Tang, J., I. D. Iliev, J. Brown, D. M. Underhill, and V. A. Funari, 2015 Mycobiome: Approaches to analysis of intestinal fungi. J. Immunol. Methods 421: 112–121.

Tati, S., P. Davidow, A. McCall, E. Hwang-Wong, I. G. Rojas *et al.*, 2016 Candida glabrata Binding to Candida albicans Hyphae Enables Its Development in Oropharyngeal Candidiasis. PLoS Pathog 12:e1005522.

Taverna, D. M., and R. A. Goldstein, 2002 Why are proteins marginally stable? Proteins 46:105-9 .

Taylor, B. N., M. Saavedra, and P. L. Fidel Jr, 2000 Local Th1/Th2 cytokine production during experimental vaginal candidiasis: potential importance of transforming growth factor-beta. Med. Mycol. 38: 419–431.

Teng, T.-S., A.-L. Ji, X.-Y. Ji, and Y.-Z. Li, 2017 Neutrophils and Immunity: From Bactericidal Action to Being Conquered. J. Immunol. Res. 2017: 9671604.

Theelen, B., C. Cafarchia, G. Gaitanis, I. D. Bassukas, T. Boekhout *et al.*, 2018 Malassezia ecology, pathophysiology, and treatment. Med. Mycol. 56: S10–S25.

Thön, M., Q. Al-Abdallah, P. Hortschansky, and A. A. Brakhage, 2007 The thioredoxin system of the filamentous fungus Aspergillus nidulans: impact on development and oxidative stress response. J. Biol. Chem. 282: 27259–27269.

Tong, Z., S. C.-A. Chen, L. Chen, B. Dong, R. Li *et al.*, 2013 Generalized subcutaneous phaeohyphomycosis caused by Phialophora verrucosa: report of a case and review of literature. Mycopathologia 175: 301–306.

Topçuoğlu, B. D., N. A. Lesniak, M. T. Ruffin, J. Wiens, and P. D. Schloss, 2020 A Framework for Effective Application of Machine Learning to Microbiome-Based Classification Problems. MBio 11: e00434-20.

Torrone, E. A., C. S. Morrison, P.-L. Chen, C. Kwok, S. C. Francis *et al.*, 2018 Prevalence of sexually transmitted infections and bacterial vaginosis among women in sub-Saharan Africa: An individual participant data meta-analysis of 18 HIV prevention studies. PLoS Med. 15: e1002511.

Tsai Y.-L., and A. S. Lee, 2018 Cell Surface GRP78: Anchoring and Translocation Mechanisms and Therapeutic Potential in Cancer pp. 41–62 in *Cell Surface GRP78*, *a New Paradigm in Signal Transduction Biology*, edited by S. V. Pizzo. Academic Press, London.

Tucey, T. M., J. Verma, P. F. Harrison, S. L. Snelgrove, T. L. Lo *et al.*, 2018 Glucose Homeostasis Is Important for Immune Cell Viability during Candida Challenge and Host Survival of Systemic Fungal Infection. Cell Metab. 27: 988-1006.e7.

Turiansky, G. W., P. M. Benson, L. C. Sperling, P. Sau, I. F. Salkin *et al.*, 1995 Phialophora verrucosa: a new cause of mycetoma. J. Am. Acad. Dermatol. 32: 311–315.

Ukhanova, M., X. Wang, D. J. Baer, J. A. Novotny, M. Fredborg *et al.*, 2014 Effects of almond and pistachio consumption on gut microbiota composition in a randomised cross-over human feeding study. Br. J. Nutr. 111: 2146–2152.

Underhill, D. M.,and I. D. Iliev, 2014 The mycobiota: interactions between commensal fungi and the host immune system. Nat. Rev. Immunol. 14:405-16.

Underhill, D. M., and I. D Iliev, 2016 Fungal mycobiome as Probiotics, Diagnostics and Therapeutics. Available online at: https://patents.google.com/patent/US9782389B2/en (Accessed March 8, 2018).

Untereiner, W. A., J. A. Scott, F. A. Naveau, L. Sigler, J. Bachewich *et al.*, 2004 The Ajellomycetaceae, a new family of vertebrate-associated Onygenales. Mycologia 96: 812–821.

van de Wijgert, J. H. H. M., C. S. Morrison, P. G. A. Cornelisse, M. Munjoma, J. Moncada *et al.*, 2008 Bacterial vaginosis and vaginal yeast, but not vaginal cleansing, increase HIV-1 acquisition in African women. J. Acquir. Immune Defic. Syndr. 48: 203–210.

van de Wijgert, J. H. H. M., M. C. Verwijs, A. N. Turner, and C. S. Morrison, 2013 Hormonal contraception decreases bacterial vaginosis but oral contraception may increase candidiasis: implications for HIV transmission. AIDS 27: 2141–2153.

van de Wijgert, J. H. H. M., H. Borgdorff, R. Verhelst, T. Crucitti, S. Francis *et al.*, 2014 The Vaginal Microbiota: What Have We Learned after a Decade of Molecular Characterization? PLOS ONE 9(8): e105998.

van de Wijgert, J. H. H. M., M. C. Verwijs, S. K. Agaba, C. Bronowski, L. Mwambarangwe *et al.*, 2020 Intermittent Lactobacilli-containing Vaginal Probiotic or Metronidazole Use to Prevent Bacterial Vaginosis Recurrence: A Pilot Study Incorporating Microscopy and Sequencing. Sci. Rep. 10: 3884.

van Woerden, H. C., C. Gregory, R. Brown, J. R. Marchesi, B. Hoogendoorn *et al.*, 2013 Differences in fungi present in induced sputum samples from asthma patients and non-atopic controls: a community based case control study. BMC Infect. Dis. 13:69.

Vanittanakom, N., C. R. Cooper, M. C. Fisher, and T. Sirisanthana, 2006 Penicillium marneffei infection and recent advances in the epidemiology and molecular biology aspects. Clin. Microbiol. Rev. 19: 95–110.

Vargason, A. M., and A. C. Anselmo, 2018 Clinical Translation of Microbe-based Therapies: Current Clinical Landscape and Preclinical Outlook. Bioeng. Transl. Med. 6:124-137.

Vásquez, A., T. Jakobsson, S. Ahrné, U. Forsum, and G. Molin, 2002 Vaginal Lactobacillus flora of healthy Swedish women. J. Clin. Microbiol. 40:2746–2749.

Vaudel, M., J. M. Burkhart, A. Sickmann, L. Martens, and R. P. Zahedi, 2011 Peptide identification quality control. Proteomics 11: 2105–2114.

Ventolini, G., 2013 New Insides on Vaginal Immunity and Recurrent Infections. J. Genit. Syst. Disord. 02.

Verberkmoes, N. C., A. L. Russell, M. Shah, A. Godzik, M. Rosenquistf *et al.*, 2009 Shotgun metaproteomics of the human distal gut microbiota. ISME J. 3: 179–189.

Verma, P. R., 1996 Biology and control of Rhizoctonia solani on rapeseed : A Review. Phytoprotection 77: 99–111.

Videira, S. I. R., J. Z. Groenewald, U. Braun, H. D. Shin, and P. W. Crous, 2016 All that glitters is not Ramularia. Stud. Mycol. 83: 49–163.

Virtanen, S., S. Saqib, T. Kanerva, P. Nieminen, I Kalliala *et al.*, 2021 Metagenome-validated Parallel Amplicon Sequencing and Text Mining-based Annotations for Simultaneous Profiling of Bacteria and Fungi: Vaginal Microbiome and Mycobiota in Healthy Women.

Vitali, B., C. Pugliese, E. Biagi, M. Candela, S. Turroni *et al.*, 2007 Dynamics of vaginal bacterial communities in women developing bacterial vaginosis, candidiasis, or no infection, analyzed by PCR-denaturing gradient gel electrophoresis and realtime PCR. Appl. Environ. Microbiol 73:5731–5741.

Vitali, D., J. M. Wessels, and C. Kaushic, 2017 Role of sex hormones and the vaginal microbiome in susceptibility and mucosal immunity to HIV-1 in the female genital tract. AIDS Res. Ther. 14: 1–5.

Walker, G. M., 2011 Pichia anomala: cell physiology and biotechnology relative to other yeasts. Antonie Van Leeuwenhoek 99: 25–34.

Wang, Y.-Y., B. Liu, X.-Y. Zhang, Q.-M. Zhou, T. Zhang *et al.*, 2014 Genome characteristics reveal the impact of lichenization on lichen-forming fungus Endocarpon pusillum Hedwig (Verrucariales, Ascomycota). BMC Genomics 15: 34.

Wang, S., Q. Wang, E. Yang, L. Yan, T. Li *et al.*, 2017 Antimicrobial Compounds Produced by Vaginal Lactobacillus crispatus Are Able to Strongly Inhibit Candida albicans Growth, Hyphal Formation and Regulate Virulence-related Gene Expressions. Front. Microbiol. 8: 564.

Wang, Y., Y. Zhou, X. Xiao, J. Zheng, H. Zhou, 2020 Metaproteomics: A strategy to study the taxonomy and functionality of the gut microbiota. J. Proteomics 219: 103737.

Ward, T. L., M. G. Dominguez-Bello, T. Heisel, G. Al-Ghalith, D. Knights *et al.*, 2018 Development of the Human Mycobiome over the First Month of Life and across Body Sites. mSystems 3: 1-12.

Warnes, G., B. Bolker, L. Bonebakker, R. Gentleman, W. Huber *et al.*, 2005 gplots: Various R programming tools for plotting data.

Wei, Q., B. Fu, J. Liu, Z. Zhang, and T. Zhao, 2012 Candida albicans and bacterial vaginosis can coexist on Pap smears. Acta Cytol. 56: 515–519.

Wei, T., and M. T. Wei, 2016 Package “corrplot.” Statistician 56:316–324.

Weiss, H. A., A. Buvé, N. J. Robinson, E. Van Dyck, M. Kahindo *et al.*, 2001 The epidemiology of HSV-2 infection and its association with HIV infection in four urban African populations. AIDS 15.

Weissenbacher, T., S. S. Witkin, W. J. Ledger, V. Tolbert, A. Gingelmaier *et al.*, 2009 Relationship between clinical diagnosis of recurrent vulvovaginal candidiasis and detection of Candida species by culture and polymerase chain reaction. Arch. Gynecol. Obstet. 279: 125–129.

Whiteley, M., S. P. Diggle, and E. P. Greenberg, 2017 Progress in and promise of bacterial quorum sensing research. Nature 551: 313–320.

Wickerham, L. J., and E. Duprat, 1945 A Remarkable Fission Yeast, Schizosaccharomyces versatilis NOV. SP. J. Bacteriol. 50: 597–607.

Wickham, H., 2009 *Ggplot2: Elegant Graphics for Data Analysis*. Springer, New York.

Wilmes, P., and P. L. Bond, 2006 Metaproteomics: studying functional gene expression in microbial ecosystems. Trends Microbiol. 14: 92–97.

Wilson, D., S. Thewes, K. Zakikhany, C. Fradin, A. Albrecht *et al.*, 2009 Identifying infection-associated genes of Candida albicans in the postgenomic era. FEMS Yeast Res. 9: 688–700.

Wisecaver, J. H., J. C. Slot, and A. Rokas, 2014 The Evolution of Fungal Metabolic Pathways. PLOS Genet. 10: e1004816.

Wira, C., J. Fahey, H. D. White, G. Yeaman, A. Givan *et al.*, 2002 The Mucosal Immune System in the Human Female Reproductive Tract: Influence of Stage of the Menstrual Cycle and Menopause on Mucosal Immunity in the Uterus. In:, pp. 359–391.

Wira C. R., J. V. Fahey, C. L. Sentman, P. A. Pioli, and L. Shen, 2005 Innate and adaptive immunity in female genital tract: cellular responses and interactions. Immunol. Rev. 206: 306–335.

Wira C. R., K. S. Grant-Tschudy, and M. A. Crane-Godreau, 2005b Epithelial cells in the female reproductive tract: a central role as sentinels of immune protection. Am. J. Reprod. Immunol. 53: 65–76.

Witherden, E. A., S. Shoaie, R. A. Hall, and D. L. Moyes 2017 The Human Mucosal Mycobiome and Fungal Community Interactions. J. Fungi 3: 56.

Witkin, S. S., I. M. Linhares, and P. Giraldo, 2007 Bacterial flora of the female genital tract: function and immune regulation. Best Pract. Res. Clin. Obstet. Gynaecol. 21:347–354.

World Health Organization, 2012 *Global incidence and prevalence of selected curable sexually transmitted infections -2008.* World Health Organization, Geneva.

Xie, Z., A. Thompson, T. Sobue, H. Kashleva, H. Xu *et al*., 2012 Candida albicans biofilms do not trigger reactive oxygen species and evade neutrophil killing. J. Infect. Dis. 206: 1936–1945.

Xiong, W., P. E. Abraham, Z. Li, C. Pan, and R. L. Hettich, 2015 Microbial metaproteomics for characterizing the range of metabolic functions and activities of human gut microbiota. Proteomics 15: 3424–3438.

Yang, B. Y., 1962 Basidiobolus meristosporus of Taiwan. Taiwania 8: 17-27.

Yang, C. W., T. M. S. Barkham, F. Y. Chan, Y. Wang, 2003 Prevalence of Candida species, including Candida dubliniensis, in Singapore. J. Clin. Microbiol. 41: 472–474.

Yang, Q., Y. Wang, X. Wei, J. Zhu, X. Wang X *et al.*, 2020 The Alterations of Vaginal Microbiome in HPV16 Infection as Identified by Shotgun Metagenomic Sequencing. Front. Cell. Infect. Microbiol. 10: 1–14.

Yano, J., M. C. Noverr, and P. L. Fidel Jr, 2012 Cytokines in the host response to Candida vaginitis: Identifying a role for non-classical immune mediators, S100 alarmins. Cytokine 58: 118–128.

Yano, J., J. D. Sobel, P. Nyirjesy, R. Sobel, V. L. Williams *et al.*, 2019 Current patient perspectives of vulvovaginal candidiasis: incidence, symptoms, management and post-treatment outcomes. BMC Womens. Health 19: 48.

Young, T. W. K., 1969 Ultrastructure of Aerial Hyphae in Linderina pennispora. Ann. Bot. 33: 211–216.

Young, J. C., C. Pan, R. Adams, B. Brooks, J. F. Banfield, M. J. Morowitz *et al.*, 2015 Metaproteomics reveals functional shifts in microbial and human proteins during a preterm infant gut colonization case, Proteomics 15: 3463–3473.

Yuan, Z.-F., S. Lin, R. C. Molden, and B. A. Garcia, 2014 Evaluation of Proteomic Search Engines for the Analysis of Histone Modifications. J. Proteome Res. 13: 4470–4478.

Yukawa, M., and T. Maki, 1931 Schizosaccharomyces japonicus nov. sp. La Bul Sci Fakultat Terkultura Kjusu Imp Univ Fukuoka Japan 4: 218– 226.

Zakikhany, K., J. R. Naglik, A. Schmidt-Westhausen, G. Holland, M. Schaller *et al.*, 2007 In vivo transcript profiling of Candida albicans identifies a gene essential for interepithelial dissemination. Cell. Microbiol. 9: 2938–2954.

Zanoni, B. C., M. Archary, S. Buchan, I. T. Katz, and J. E. Haberer, 2016 Systematic review and meta-analysis of the adolescent HIV continuum of care in South Africa: the Cresting Wave. BMJ Glob. Heal. 1: e000004.

Zevin, A. S., I. Y. Xie, K. Birse, K. Arnold, L. Romas *et al.*, 2016 Microbiome Composition and Function Drives Wound-Healing Impairment in the Female Genital Tract. PLoS Pathog. 12: 1–20.

Zhang, E., T. Tanaka, M. Tajima, R. Tsuboi, A. Nishikawa *et al.*, 2011 Characterization of the skin fungal microbiota in patients with atopic dermatitis and in healthy subjects. Microbiol. Immunol. 55:625–632.

Zhang, X., Z. Ning, J. Mayne, J. I. Moore, J. Li *et al.*, 2016 MetaPro-IQ: a universal metaproteomic approach to studying human and mouse gut microbiota. Microbiome 4: 31.

Zheng, N. N., X. C. Guo, W. Lv, X. X. Chen, G. F. Feng 2013 Characterization of the vaginal fungal flora in pregnant diabetic women by 18S rRNA sequencing. Eur. J. Clin. Microbiol. Infect. Dis. 32: 1031–1040.

Zhou, X., R. Westman, R. Hickey, M. A. Hansmann, C. Kennedy *et al.*, 2009 Vaginal microbiota of women with frequent vulvovaginal candidiasis. Infect. Immun. 77: 4130–4135.

Zoetendal, E. G., M. Rajilic-Stojanovic, W. M. de Vos 2008 High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. Gut 57: 1605–1615.