

RESEARCH

Open Access



# Exploring the female genital tract mycobiome in young South African women using metaproteomics

Tamlyn K. Gangiah<sup>1,2</sup>, Arghavan Alisoltani<sup>3</sup>, Matthys Potgieter<sup>1,4</sup>, Liam Bell<sup>5</sup>, Elizabeth Ross<sup>5</sup>, Arash Iranzadeh<sup>1</sup>, Zac McDonald<sup>5</sup>, Imane Allali<sup>1,6</sup>, Smritee Dabee<sup>3,7</sup>, Shaun Barnabas<sup>3</sup>, Jonathan M. Blackburn<sup>4,8</sup>, David L. Tabb<sup>8,9,10</sup>, Linda-Gail Bekker<sup>8,11</sup>, Heather B. Jaspan<sup>7,8,12</sup>, Jo-Ann S. Passmore<sup>3,8,13,14</sup>, Nicola Mulder<sup>1,8,15†</sup> and Lindi Masson<sup>3,8,13,16,17\*†</sup>

## Abstract

**Background** Female genital tract (FGT) diseases such as bacterial vaginosis (BV) and sexually transmitted infections are prevalent in South Africa, with young women being at an increased risk. Since imbalances in the FGT microbiome are associated with FGT diseases, it is vital to investigate the factors that influence FGT health. The mycobiome plays an important role in regulating mucosal health, especially when the bacterial component is disturbed. However, we have a limited understanding of the FGT mycobiome since many 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 FGT mycobiome in South African reproductive-age women.

**Results** We examined FGT fungal communities present in 123 women by collecting lateral vaginal wall swabs for liquid chromatography-tandem mass spectrometry. From this, 39 different fungal genera were identified, with *Candida* dominating the mycobiome (53.2% relative abundance). We observed changes in relative abundance at the protein, genus, and functional (gene ontology biological processes) level between BV states. In women with BV, *Malassezia* and *Conidiobolus* proteins were more abundant, while *Candida* proteins were less abundant compared to BV-negative women. Correspondingly, Nugent scores were negatively associated with total fungal protein abundance. The clinical variables, Nugent score, pro-inflammatory cytokines, chemokines, vaginal pH, *Chlamydia trachomatis*, and the presence of clue cells were associated with fungal community composition.

**Conclusions** The results of this study revealed the diversity of FGT fungal communities, setting the groundwork for understanding the FGT mycobiome.

**Keywords** Mycobiome, Metaproteomics, Female genital tract, Bacterial vaginosis, Fungi

<sup>†</sup>Nicola Mulder and Lindi Masson are co-senior authors in this study.

\*Correspondence:

Lindi Masson

Lindi.Masson@burnet.edu.au

Full list of author information is available at the end of the article



## Background

The female genital tract (FGT) microbiome comprises a community of microorganisms (bacteria, archaea, fungi, and viruses), some of which play a major role in maintaining or disrupting FGT health [1]. Although many previous studies have characterized the diversity of the bacterial component of the FGT microbiome and its relationship with adverse health outcomes using next-generation sequencing and metaproteomics approaches [2–4], information about the mycobiome (otherwise known as the fungome) is relatively sparse [5–7]. The FGT mycobiome refers to the yeasts (unicellular) and filamentous fungi that make up the genital fungal community [8]. Human-colonizing fungi have been shown to play a critical role in health outcomes and are capable of altering host function and behavior, metabolic function, energy acquisition, vitamin-cofactor availability, and immune system development and functioning [9]. However, much less is known about the role of fungi in the FGT, and the majority of studies have focused on the genus *Candida*, which includes a group of species that cause vulvovaginal candidiasis (VVC). VVC is the second most reported type of microbial vaginitis and approximately 7% of women globally have recurrent VVC [10, 11]. VVC is associated with a 2.2-fold increased risk of human immunodeficiency virus (HIV) infection [12–17] and may also increase HIV shedding and hence transmission to sexual partners [18]. The majority of human-colonizing fungi are opportunistic pathogens that result in disease when the immune system is compromised, and immunosuppressed individuals with HIV have a greater risk of fungal infection [19, 20]. This is particularly concerning in settings such as South Africa, where VVC affects approximately 1 million women each year and HIV is also prevalent [10, 11, 21].

Few studies have characterized the FGT mycobiome, and none has been conducted using metaproteomics in South Africa, to our knowledge. One of the few next-generation sequencing studies that has been conducted on the FGT mycobiome of Estonian women used internal transcribed spacer 2 (ITS2) sequencing. This study found that, as expected, *Candida albicans* was the most abundant fungal species in the study population (34%) [22]. However, a significant proportion of mycobiome sequences remained unclassified [23–25], highlighting how premature this field is. ITS is the commonly used target for fungi; however, this sequencing method is limited as the length of the ITS1-2 region is extremely variable among different fungal genera and species [26]. The uneven lengths of ITS fragments may promote preferential PCR amplification of shorter ITS sequences that could lead to a biased quantification of relative abundances of fungal taxa [27]. Therefore,

primer choice causes significant bias in PCR-dependent methods [27]. In the past, FGT studies were reliant on culture-based techniques, which restricted the fungal species researchers could access [28] and targeted analyses using PCR were generally limited to known human pathogens (e.g., *C. albicans*, *Candida glabrata*, *Cryptococcus neoformans*, and *Aspergillus fumigatus*) [20]. Consequently, fungi have not been extensively classified in the FGT. Metaproteomics may be a suitable method as research using this approach has 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 [29–31].

Another common vaginal condition is bacterial vaginosis (BV), which occurs when the optimal *Lactobacillus* species are replaced by diverse (facultative) anaerobic bacteria. Like VVC, BV is also associated with an increased risk of HIV, and other sexually transmitted infections (STIs) [32–36]. VVC is a common side effect of antibiotics used to treat BV [37], indicating that FGT bacterial communities might influence the colonization of vaginal yeast (such as *Candida* spp.).

Clinically, *Candida* and BV-associated bacteria commonly co-occur: approximately 20–30% of BV patients have both conditions. When there is a co-occurrence of BV and VVC, microbiome analyses have shown an abundance of *Lactobacillus*, contrary to its depletion in individuals with only BV [38].

Recent research has suggested that the bacteriome and mycobiome directly impact each other [39–41]. Thus, shifts in one community modulate the community structure of the other. Even though BV and VVC may co-occur [38, 42, 43], other studies have shown an inverse association between BV and vaginal yeast, implying that yeast colonization occurs more frequently in an optimal vaginal microbiome [22, 44–48]. However, there is minimal data available on fungal interactions and possible coexistence patterns with bacterial communities [22] and emerging studies have proposed that changes in commensal fungi may be relevant in diseases that are not primarily fungal [49]. Of the research conducted, the fungal genus best studied for its association with bacteria is *Candida*. However, there is conflicting evidence describing the relationship between *Lactobacillus* and *C. albicans* colonization. While *Candida* is highly tolerant of changes in vaginal pH and can co-colonize the FGT with *Lactobacillus* [50–53], some studies suggest that *Lactobacillus* spp. compete with *C. albicans* for adhesion sites and secrete substances, such as lactic acid, that inhibit fungal attachment to control *C. albicans* invasion and disease [54–58].

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 analyzed bacteria and fungi from the same sample. The aim of this study was to characterize the FGT mycobiomes of young women residing in South Africa using a metaproteomics approach and to evaluate the relationship between the mycobiome and BV, as well as other clinical variables.

## Methods

### Study participants

A total of 148 South African women were recruited as part of the Women's Initiative in Sexual Health (WISH) study from a resource-limited, population-dense community in Cape Town [59]. The adolescent population in this community was found to have a high incidence of HIV and other STIs [59]. This sub-study included 123 of the women who participated in the parent study between the ages of 16 and 22 years who had vaginal swabs available for metaproteomics analysis. Seventy-four of these women provided another sample at a second-time point 9 weeks later.

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. Women were enrolled in the main study if they were HIV-negative, in general, in good health, not pregnant or menstruating at the time of sampling, and had not had unprotected sex or douched in the last 48 h. Additional exclusion criteria included the use of antibiotics in the previous 2 weeks. Patient-related metadata was collected and recorded [59]. Metadata included information about age, body mass index (BMI), hormonal contraception (HC), cytokine measurements, STIs, and BV status [60].

### Sample collection and processing

As previously described in [59], one lateral vaginal wall swab was rolled onto a glass slide and Gram-stained for microscopy to evaluate BV status using Nugent score. Vaginal pH was measured from swabs using color-fixed indicator strips (Macherey–Nagel, Düren, Germany) [59]. Swab eluants were analyzed by PCR to identify women with STIs (i.e., *Chlamydia trachomatis*, *Neisseria gonorrhoeae*, *Trichomonas vaginalis*, and *Mycoplasma genitalium*). To identify women with herpes simplex virus 2 (HSV-2) infections, serum was collected for serology. Cytokines and chemokines were measured in menstrual cup secretions to identify the inflammation state of each participant using Luminex [60]. *K*-means clustering was used to identify women with low, medium, and high

pro-inflammatory cytokine profiles [interleukin (IL)–1 $\alpha$ , IL-1 $\beta$ , IL-6, IL-12p40, IL-12p70, tumor necrosis factor (TNF)- $\alpha$ , TNF- $\beta$ , TNF-related apoptosis-inducing ligand, interferon- $\gamma$ ] as described in [60].

### Mass spectrometry

For shotgun liquid chromatography-tandem mass spectrometry (LC–MS/MS), lateral vaginal wall swabs were collected, placed in 1 ml phosphate-buffered saline, transported to the laboratory at 4 °C and stored at –80 °C. Samples were computationally randomized for processing and analysis. Each swab sample was clarified by centrifugation. The protein content of each supernatant was determined by the bicinchoninic acid assay (BCA). Equal protein amounts (100  $\mu$ g) 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 clarify samples. The samples were then incubated overnight with trypsin, and the peptides were eluted with HEPES and dried via vacuum centrifugation. Reversed-phase LC 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 as previously described in [61].

### Protein identification, taxonomic assignment, and functional annotation

Raw MS/MS files were processed with MaxQuant version 1.6.3.4 on Linux using a curated pan proteome database. The database was created by manually curating a fungal pan proteome database, consisting of all publicly available sequences for the fungal species that are predicted to form part of the FGT mycobiome; these species and their importance in relation to the FGT are listed in Supplementary Table 1 in Additional file 1. This fungal pan proteome database was concatenated to a Metanovo database that was generated as previously described [61]. In short, Metanovo is a software pipeline that integrates existing tools (DeNovoGUI, DirecTag, PeptideMapper, and X!Tandem) with a custom algorithm to generate targeted protein sequence databases [62]. In our study, Metanovo created the database by mapping de novo sequence tags to a UniProtKB database that was filtered for human and microbial entries. For MaxQuant parameters, methionine oxidation and acetylation of the protein *N*-terminal amino acid were considered variable modifications and carbamidomethyl (C) was a fixed modification. The digestion enzyme selected was trypsin with a maximum of two missed cleavages.

'Match-between-runs' function was selected for raw spectra processing. For a more detailed description of parameters see Supplementary Table 2 in Additional file 1. The above analyses were run on the high-performance computing cluster (<https://www.ilifu.ac.za/>).

To ensure identifications of protein groups were reliable, a protein group was only accepted for downstream analysis if it had  $\geq 2$  peptides assigned to it. Many identified peptides could not unambiguously be assigned to a single protein, as the peptide sequence is part of multiple protein sequences in the database [63]. These peptides were retained for downstream analyses, and consequently, some proteins shared the same peptides. Potential contaminants, reverse peptides, and deleted protein groups from the UniProtKB database were removed from our dataset. Intensity-Based Absolute Quantitation (iBAQ) values were normalized and  $\log_2$ -transformed using the *MSnbase* R package [64]. To have a complete data matrix, missing iBAQ values were imputed using the *k*-nearest neighbor (KNN) method using the *imputeL-CMD* R package to not introduce any negative values into the dataset [65]. Thereafter, protein groups were separated according to their five taxonomic kingdoms: bacterial, mammalian, fungal, archaeal, and viral to calculate the relative abundance of fungi, and then fungal protein groups were extracted for downstream analysis.

Peptide sequences were also 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.

The functional potential of the FGT mycobiome was estimated using a protein-centric approach. The list of protein IDs from the *proteinGroups.txt* file was assigned to Gene Ontology (GO) IDs using UniProt.

#### Data analysis

The imputed  $\log_2$ -transformed iBAQ values were used to identify significantly differentially abundant fungal genera, proteins, and GO biological processes by BV status using the *limma* R package [66] and a non-parametric Wilcoxon rank sum test. To perform analysis at the biological process level, iBAQ intensities were aggregated for protein IDs mapping to the same biological process. *T*-tests accounted for confounding factors, which were determined using an analysis of variance (ANOVA) test for BV category comparisons. Proteins with false discovery rate (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 BV states, unsupervised hierarchical clustering of iBAQ intensities

was performed for all protein and functional term assignments using the *heatmap.2* R package [67]. Manhattan was used as the distance metric to calculate the absolute differences between samples. The effects of BV on the metaproteome were investigated by PCA using the *mixOmics* R package [68].

Relationships between fungal taxa identified were interrogated using the more robust Spearman's rank correlation analysis with  $\log_2$  normalized iBAQ intensities. Fungal-bacterial relationships were also investigated using fungal taxa identified to be differentially abundant and selecting bacterial taxa found to be associated with BV and inflammation in the literature from our metaproteomic dataset [69–73]. In addition, we investigated the correlation between fungal taxa and clinical variables recorded in this study. Correlation plots were generated in R using the *Hmisc* and *Corrplot* packages [74, 75]. Fungal-bacterial relationships and the correlation between fungal taxa and clinical variables were further interrogated using Bayesian networks. Bayesian networks were constructed in R using the *bnlearn* package [76].

To investigate the effects of clinical variables on FGT 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 [77].

To predict fungal proteins/taxa and clinical variables associated with BV, an ENSEMBLE bagging decision tree was used. The *caret* (Classification And REgression Training) package was used to perform ENSEMBLE bagging in R [78]. 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* [79]. For protein data, default parameters were used.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [1] partner repository with the dataset identifier PXD046053.

## Results

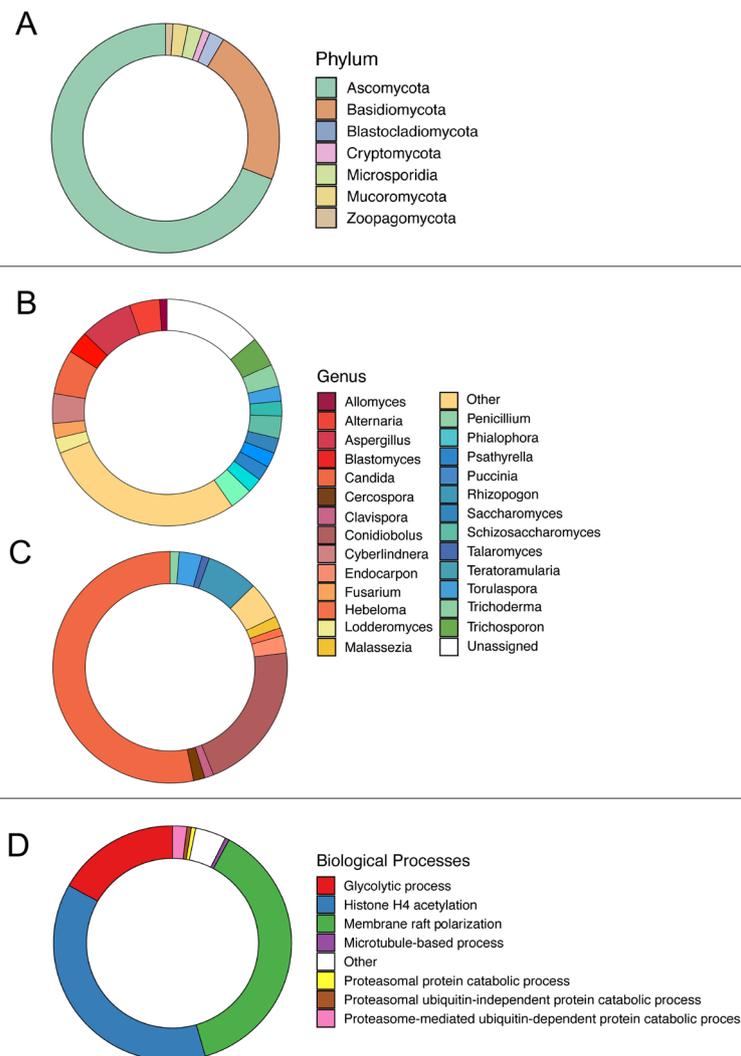
### Vaginal fungal taxa and their co-occurrence

We assessed the FGT mycobiome of young, South African women living in Cape Town, South Africa. As previously reported, STI and BV prevalence were high in this cohort, with 42.5% of the women having at least one STI and 50.4% of the women being BV positive by Nugent scoring at visit 1. Median age was 18.5 years old and 100% of participants were black (Supplementary Table 3 and 4, Additional file 1) [59, 61].

Metaproteomics was used to identify peptides present in lateral vaginal wall swab samples that were

then assigned to proteins. Both peptides and proteins were used for fungal taxonomic analysis. Of the peptides identified post-quality control in lateral vaginal wall swab samples ( $n=17,758$ ), 94 (0.53%) peptides were assigned to a fungal taxonomic lineage as its lowest common ancestor, 7177 (40.42%) to human, 8208 (46.20%) to bacteria, and less than 1% of peptides were assigned to other taxa (i.e., Archaea and viruses). 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. The identified fungal peptide sequences were assigned to seven phyla, with the majority assigned to the phylum Ascomycota (69.15%), followed by Basidiomycota (22.34%) (Fig. 1A). A smaller percentage of peptide sequences were assigned to the phyla, Blastocladiomycota, Microsporidia,



**Fig. 1** Fungal relative abundance was determined using metaproteomics. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) was used to evaluate the fungal metaproteome in lateral vaginal wall swabs from 123 women from Cape Town, South Africa. Raw MS files were processed with MaxQuant using a database that was created by manually curating a fungal pan proteome database and concatenated to a Metanovo database. **A** Taxonomic assignment of peptides. Taxonomic assignment was conducted using Unipept and the distribution of phyla identified is shown as a doughnut chart. The majority of fungal peptide sequences were assigned to the phylum Ascomycota. **B** Distribution of fungal genera using peptide sequences. Genera identified using Unipept, and distribution is depicted as a doughnut chart. The majority of fungal peptide sequences were assigned to the genus *Aspergillus*. **C** Relative abundance of fungal genera pre-normalization using summed total intensity-based absolute quantification (iBAQ) protein values for each genus. *Candida* was the most relatively abundant genus. **D** Relative abundance of fungal protein biological process gene ontology (GO) function. Biological process relative abundance was determined by aggregation of fungal protein iBAQ values assigned to the same GO term;  $n=24$ . The most relatively abundant biological processes were histone H4 acetylation and membrane raft polarization. The relative abundance of biological processes is depicted as a doughnut chart

Mucoromycota, Zoopagomycota, and Cryptomycota (Fig. 1A). At the genus level, 46 distinct fungal genera were identified. Of these genera, 7.4% (7) of sequences were assigned to *Aspergillus*, 6.4% (6) to *Candida*, followed by 4.3% (4) to *Alternaria*, *Cyberlindnera*, and *Trichosporon* and 13.69% of sequences could not be assigned to genus level (Fig. 1B).

The 94 fungal peptides were assigned to 50 protein groups that passed quality control filtering (Supplementary Table 1, Additional file 2). A description of these fungal protein groups and their importance in the FGT can be found in Supplementary Table 1 in Additional file 1. Using iBAQ values, the relative abundance of fungal proteins was 0.4%, human 91.1%, bacteria 8.5%, viruses 0.06%, and archaea 5.46e–05%. A total of 39 fungal genera and 45 fungal species were identified. The most prevalent genera were *Candida* (53.2%), *Conidiobolus* (20.9%), *Rhizopogon* (7.1%), and *Teratoramularia* (3.1%) (Fig. 1C). 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 MaxQuant output files to confirm if these fungal assignments were correct (Supplementary Table 5, Additional file 1). A BLAST search of the protein sequence for unusual species not previously detected in the FGT, such as *Endocarpon pusillum* (a lichen-forming fungus) and *Rhizoclostridium globosum* (fungus usually found in damp and aquatic environments) confirmed that these fungi were correctly annotated during protein and taxonomic assignment procedure.

The majority of fungal protein groups identified were actin and heat shock proteins (HSPs). Of the HSPs, 5 were heat shock protein 70 kDa (HSP-70). The remainder of the prevalent proteins were calmodulin, tubulin beta chain, 78 kDa glucose-regulated proteins (GRP78), and elongation factor 2 proteins (Supplementary Table 1, Additional file 2). According to GO annotation, 48% of the fungal proteins could be assigned to known biological processes. The most relatively abundant pathways were histone H4 acetylation (37.7%), membrane raft polarization (37.7%), and the glycolytic process (16.8%) (Fig. 1D). “Other” included GO biological processes with a relative abundance of less than 0.5%.

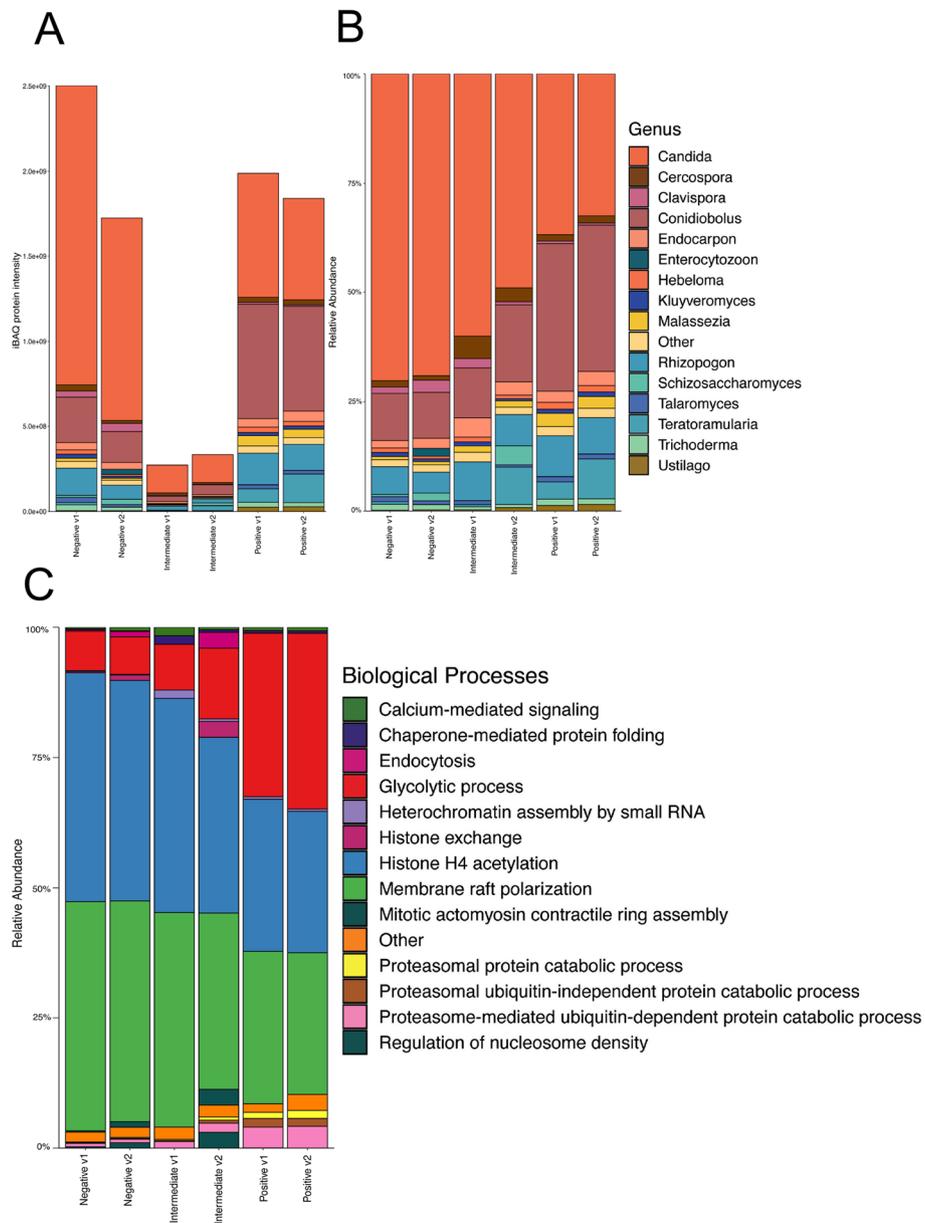
We determined Spearman’s correlation coefficients to explore associations between all fungal species identified using the samples collected only at visit 1 ( $n=113$ ) to investigate which fungal species likely co-occur in the mycobiome (Supplementary Fig. 1, Additional file 2). There were a greater number of positive correlations than negative correlations, many of which were observed between *Candida* spp., including *Candida orthopsilosis*, *Clavispora lusitaniae*, and *Candida thasaenensis* with *C.*

*albicans*, and *Candida orthopsilosis* with *Kluyveromyces marxianus* (*Candida kefir*) and *Candida dubliniensis*, suggesting that these Candidal species are more likely to co-occur (Supplementary Fig. 1, Additional file 2). In contrast, negative correlations were much weaker, with the strongest negative correlation occurring between the commonly found FGT species *C. albicans* with *Alternaria alternata* and *Malassezia sympodialis* suggesting that these species are usually not present at a high abundance at the same time (Supplementary Fig. 1, Additional file 2). Similar correlations were observed between fungal taxa in visit 2 samples ( $n=89$ ); however, the strongest negative correlation was now observed between *M. sympodialis* and *Torulaspota delbrueckii* (*Candida colliculosa*), a yeast species ( $\text{cor} = -0.71$ ;  $p = 3.77e-15$ ; not shown).

#### The relative abundance of fungal genera differed between BV states

We compared the relative abundance of fungal genera, proteins, and GO biological processes across different BV states (BV negative, BV intermediate, BV positive) longitudinally using limma *t*-test and Wilcoxon rank sum test. The genus level was chosen to examine differences between BV states as taxonomic annotation is not very efficient in the differentiation of species [80]. An ANOVA test demonstrated that pro-inflammatory cytokines ( $p\text{-value} = 1.78e-12$ ) and vaginal pH ( $p\text{-value} = 1.04e-08$ ) were significantly associated with BV status. Therefore, differential expression analyses using limma were also adjusted for these variables (Additional file 3).

Protein abundance, as measured by total fungal iBAQ values, was similar between BV-positive and BV-negative women, but was substantially lower in intermediate microbial states (Nugent 4–6), even though samples were normalized to total protein level prior to LC–MS/MS analysis (Fig. 2A). However, only a few individuals were assigned to this state, which may affect the total iBAQ value. As a general trend, *Candida* and *Conidiobolus* were the most relatively abundant fungal genera in each BV state (Fig. 2B; Supplementary Fig. 2, Additional file 2). The BV negative group had the largest relative abundance of *Candida* (69.7% visit 1; 68.1% visit 2) in comparison to other BV groups (Fig. 2B). In the BV intermediate category, the relative abundance of *Candida* was 59.9% at visit 1 and 48.7% at visit 2 (Fig. 2B). 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. 2B). Similarly, the total *Candida* iBAQ values in BV-negative women at both visits were higher compared to the BV-positive women (Fig. 2A), *Candida* was significantly under-abundant in the BV-positive group compared to



**Fig. 2** Fungal relative abundance was determined using metaproteomics for each bacterial vaginosis (BV) group. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) was used to evaluate the fungal metaproteome in lateral vaginal wall swabs from 123 women from Cape Town, South Africa. Raw MS files were processed with MaxQuant using a database that was created by manually curating a fungal pan proteome database and concatenated to a Metanovo database. The iBAQ values for fungal proteins were normalized,  $\log_2$ -transformed, and imputed using the  $k$ -nearest neighbor method. BV groups were defined by the Nugent score. 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. Fungal relative abundance for each BV state was determined by aggregation of fungal proteins for each individual assigned to a specific BV state. **A** The total relative abundance of fungal genus by BV state at each visit was determined by summing total fungal iBAQ values of proteins assigned to each fungal genus. Genus assignments were determined using UniProt on protein groups. *Candida* was the most relatively abundant across all BV states. **B** Fungal community composition at the genus level across BV states at different visits. **C** Relative abundance of biological processes according to Gene Ontology (GO). GO determined using UniProt across BV states at different visits. Biological process relative abundance was determined by aggregation of fungal protein iBAQ values assigned to the same GO term. Histone H4 acetylation and membrane raft polarization were the most abundant biological processes across BV states

the BV-negative group at visit 1 ( $adj. p=2.38e-05$ ), with a similar trend at visit 2, although non-significant ( $adj. p=0.08$ ).

The BV positive group had the highest relative abundance and total iBAQ protein values of *Conidiobolus*, *Rhizopogon*, and *Malassezia* compared to BV negative and BV intermediate women (Fig. 2A, B). These differences were significant for *Conidiobolus* and *Malassezia* at visits 1 and 2 ( $adj. p=0.04$ ;  $adj. p=0.002$  and  $adj. p=2.38e-05$ ;  $adj. p=2.21e-06$ , respectively), suggesting that these genera are likely more abundant during BV. 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.”

### Fungal proteins differed between BV states

We performed a *limma* test and Wilcoxon rank sum test to identify significant differentially abundant ( $adj. p\text{-value}<0.05$ ) proteins between BV groups. The results were found to be consistent between methods (Table 1 and Additional file 3). Based on the *limma* results, the comparison between BV positive and BV negative groups revealed that 17 fungal proteins were significantly different between BV positive compared to negative women at visit 1 (Table 1). However, only two proteins, M5EKD5 (proteasome subunit alpha type, from *M. sympodialis*) and A0A137P973 (fructose-bisphosphate aldolase assigned to *Conidiobolus coranatus*), were more abundant in the BV positive state compared to BV negative groups. At visit 2, six fungal proteins were significantly different (Supplementary Table 2, Additional file 2); five of these fungal proteins were the same as those identified as differentially abundant at visit 1. The comparison between BV negative and BV intermediate groups revealed that 13 fungal proteins were more abundant in the BV negative group in visit 1 (Table 1). Five of these fungal proteins were the same as those identified in the comparison between BV-positive and BV-negative women. The comparison between BV-positive and BV intermediate samples identified one more abundant fungal protein at visit 1 in the BV-positive group (Table 1), C4Y881 (Gp\_dh\_C domain-containing protein, from *C. lusitaniae*). However, some of these associations were no longer significant after adjusting for vaginal cytokine profiles and pH, suggesting that differences in certain fungal taxa and proteins may be influenced by cytokine and pH changes associated with BV.

Hierarchical clustering according to fungal protein iBAQ values confirmed clustering according to BV status (Fig. 3A). 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. 3A).

Four proteins, A0A0C3J1M2 (an uncharacterized protein, from *Pisolithus tinctorius*), A5E2Z2 (an uncharacterized protein, from *Lodderomyces elongisporus*), B6Q757 (translation elongation factor EF-2, from *Talaromyces marneffeii*), and A0A1D8PFR4 (actin, from *C. albicans*), were consistently highly expressed across clusters (Fig. 3A), with A0A1D8PFR4 showing the highest expression. The BV-negative dominant cluster showed the highest expression of the majority of proteins in comparison to the other clusters, suggesting that fungi express more protein during the BV-negative state than the BV-positive state.

Principal component analysis (PCA) using fungal proteins shows that BV-positive and BV-negative individuals tended to cluster together into their respective groups, with a few outliers (Fig. 3B); the clustering was more clearly visible in visit 1 samples (Fig. 3B; ANOSIM  $R=0.2413$ ,  $p=0.001$ ) compared to visit 2 samples (Fig. 3C;  $R=0.1695$ ,  $p=0.001$ ). This suggested that fungal communities differ between BV-positive and BV-negative individuals. Individuals belonging to the BV intermediate group tended to cluster with the BV-negative samples (Fig. 3B, C).

We investigated fungal protein signatures that differed between BV-positive and BV-negative individuals to assess which fungal taxa were likely influenced by the disruption of the vaginal microbiome caused by BV. We used an ENSEMBLE Bagging Decision Tree, which gave the best results compared to other machine learning models when employing a classification approach on protein iBAQ data. For this analysis, we removed the BV intermediate state due to the small sample size ( $n=7$ ). Our training classification model incorporated 80% of the dataset as input, which included 84 samples, 50 predictors (proteins), and two 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% ( $p\text{-value}=0.01$ ). The model had a sensitivity of 100%, a specificity of 85%, the positive predictive value was 78%, and it had a negative predictive value of 100%. Although the moderate positive predictive value of this analysis suggests that some BV-negative women were incorrectly classified as positive, the overall accuracy of 90% upon validation suggests that vaginal fungal and bacterial populations are very closely related. The area under the curve (AUC) was determined to be 0.99. The model found that the most important fungal proteins distinguishing the BV-positive from the BV-negative category were M5EKD5 (assigned to *M. sympodialis*), A0A1D8PFR4 (*C. albicans*), and J7M8M3 (*C. thasaenensis*) (Fig. 4).

**Table 1** Differentially abundant proteins, fungal taxa and biological processes between bacterial vaginosis (BV) state at the first study visit

<b>BV-positive—BV-negative (visit 1)</b>					
<b>Protein</b>	<b>Protein name</b>	<b>Fungal species</b>	<b>logFC</b>	<b>p-value</b>	<b>adj. p-value</b>
J7M8M3	Actin	<i>Candida thasaenensis</i>	−1.31	9.88e−08	1.70e−06
A5E2Z2	Uncharacterized	<i>Lodderomyces elongisporus</i>	−0.64	3.71e−06	4.64e−05
M5EKD5	Proteasome subunit alpha type	<i>Malassezia sympodialis</i>	1.24	1.02e−07	1.70e−06
A0A1D8PFR4	Actin	<i>Candida albicans</i>	−1.76	2.03e−08	1.02e−06
A0A177DGX3	Glutamate dehydrogenase	<i>Alternaria alternata</i>	−1.31	3.00e−05	0.0002
A0A0C3J1M2	Uncharacterized	<i>Pisolithus tinctorius</i>	−0.54	2.11e−05	0.0002
R9APA2	Calmodulin	<i>Wallemia ichthyophaga</i>	−1.14	1.72e−05	0.0002
H8WXM9	Kar2 protein	<i>Candida orthopsilosis</i>	−1.14	3.45e−05	0.0002
C4XYH4	HATPase_c domain-containing protein	<i>Clavispora lusitaniae</i>	−1.06	0.0002	0.001
A0A137P973	Fructose-bisphosphate aldolase	<i>Conidiobolus coronatus</i>	1.85	0.0009	0.004
B6Q757	Translation elongation factor EF-2 subunit, putative	<i>Talaromyces marneffeii</i>	−0.50	0.001	0.005
B3FE91	Calmodulin	<i>Petromyces alliaceus</i>	−0.96	0.001	0.005
A0A1Y1W0Z6	Putative mitochondrial Hsp70 chaperone	<i>Linderina pennispora</i>	−0.84	0.01	0.04
K1W4Z9	Heat shock protein 70	<i>Trichosporon asahii</i>	−0.81	0.004	0.02
A0A1Y2GNC9	Tubulin beta chain	<i>Lobosporangium transversale</i>	−0.85	0.008	0.02
A0A068SI01	Xap5-domain-containing protein	<i>Lichtheimia corymbifera</i>	−0.91	0.006	0.02
S9XAK4	Sporulation protein Spo15	<i>Schizosaccharomyces cryophilus</i>	−1.63	0.004	0.02
		<b>Biological process</b>	<b>logFC</b>	<b>p-value</b>	<b>adj. p-value</b>
		Histone H4 acetylation	−1.77	2.19 e−07	3.28e−06
		Membrane raft polarization	−1.77	2.19 e−07	3.28e−06
		Cell adhesion	−0.61	9.27e−06	3.97e−05
		Cellular response to heat	−0.61	9.27e−06	3.97e−05
		Cellular response to osmotic stress	−0.61	9.27e−06	3.97e−05
		Cellular response to oxidative stress	−0.61	9.27e−06	3.97e−05
		Plasma membrane organization	−0.61	9.27e−06	3.97e−05
		Cellular amino acid metabolic process	−1.24	0.0002	0.0006
		Calcium-mediated signaling	−0.79	7.05e−05	0.0003
		Chromatin organization	−0.87	0.002	0.005
		Adenylate cyclase activating glucose	−0.91	0.002	0.005
		Chaperon-mediated protein folding	−0.91	0.002	0.005
		Heterochromatin assembly by small RNA	−0.91	0.002	0.005
		Glycolytic process	1.49	0.01	0.02
		Protein folding	−0.85	0.01	0.03
		Proteasome-mediated ubiquitin-dependent protein catabolic process	0.47	0.02	0.04
		Microtubule-based process	−0.41	0.02	0.04
<b>BV-negative—BV intermediate</b>					
<b>Protein</b>	<b>Protein name</b>	<b>Fungal species</b>	<b>logFC</b>	<b>p-value</b>	<b>adj. p-value</b>
A0A1V6RS99	Uncharacterized	<i>Penicillium vulpinum</i>	2.53	8.86e−05	0.0004
G8ZUB2	Exocyst complex component SEC15	<i>Torulaspota delbrueckii</i>	2.62	3.98e−05	0.001
A0A0D1CPW1	Proteasome subunit beta	<i>Ustilago maydis</i>	1.77	0.0005	0.009
C4Y881	Gp_dh_C domain-containing protein	<i>Clavispora lusitaniae</i>	1.71	0.0008	0.01
A0A068SI01	Xap5-domain-containing protein	<i>Lichtheimia corymbifera</i>	2.52	0.001	0.01
A5E2Z2	Uncharacterized	<i>Lodderomyces elongisporus</i>	0.97	0.001	0.01
A0A0C3J1M2	Uncharacterized	<i>Pisolithus tinctorius</i>	0.94	0.002	0.01
A0A1Y2CHJ5	H(+)-transporting two-sector ATPase	<i>Rhizoclosmatium globosum</i>	1.70	0.002	0.01
W3VSG0	Elongation factor 2	<i>Pseudozyma aphidis</i>	1.58	0.005	0.02

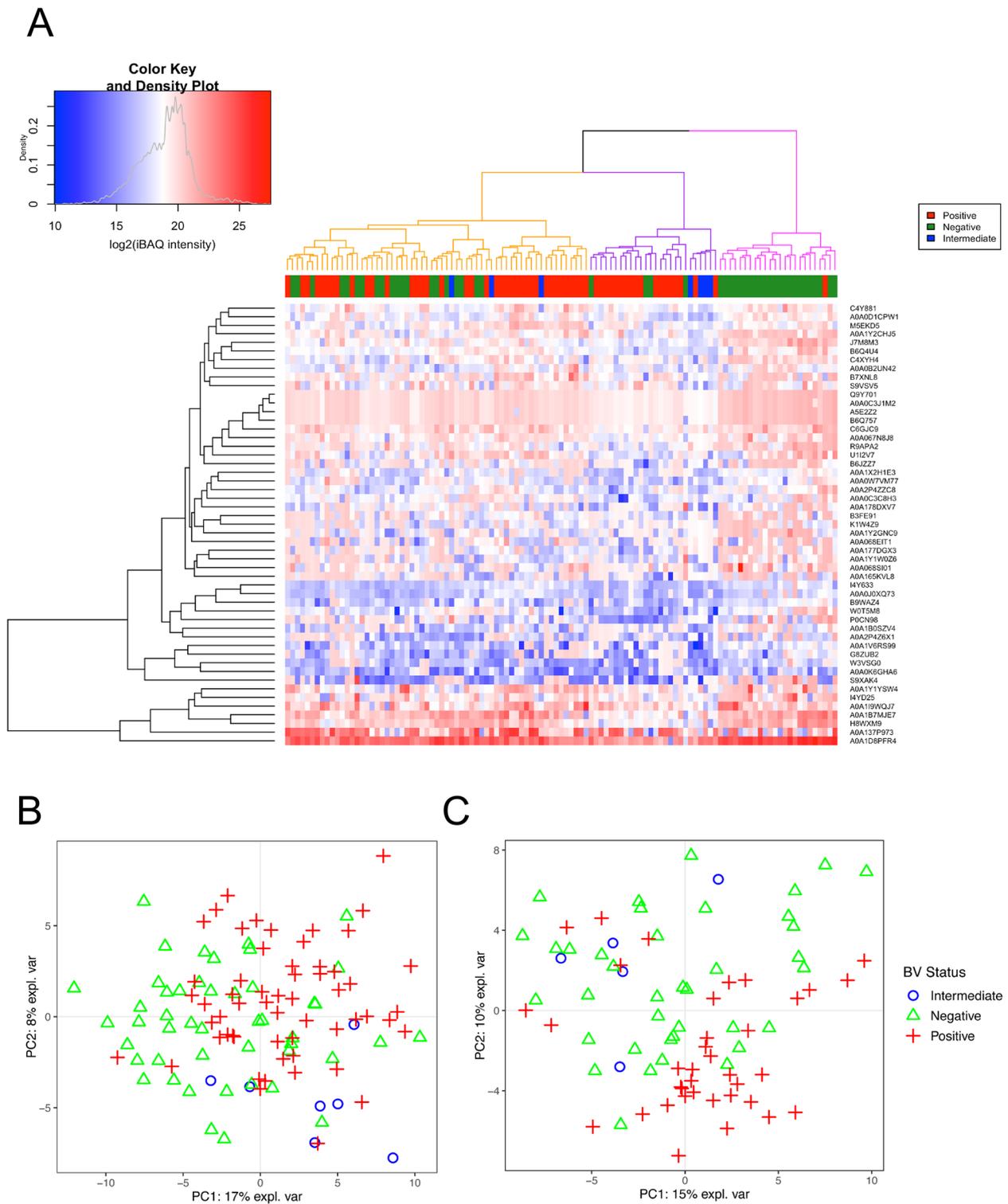
**Table 1** (continued)

<b>BV-positive—BV-negative (visit 1)</b>					
J7M8M3	Actin	<i>Candida thasaenensis</i>	1.34	0.009	0.04
A0A1I9WQJ7	Actin	<i>Teratoramularia persicariae</i>	1.71	0.009	0.04
A0A178DXV7	Elongation factor 2	<i>Pyrenochaeta sp.</i>	1.91	0.01	0.04
A0A1Y2GNC9	Tubulin beta chain	<i>Lobosporangium transversale</i>	1.81	0.01	0.04
		<b>Biological process</b>	<b>logFC</b>	<b>p-value</b>	<b>adj. p-value</b>
		Golgi to plasma membrane transport	2.75	4.38e−06	6.56e−05
		Vesicle docking involved in exocytosis	2.75	4.38e−06	6.56e−05
		Cell adhesion	0.96	0.0005	0.002
		Cellular response to heat	0.96	0.0005	0.002
		Cellular response to osmotic stress	0.96	0.0005	0.002
		Cellular response to oxidative stress	0.96	0.0005	0.002
		Plasma membrane organization	0.96	0.0005	0.002
		Translation	1.74	0.0008	0.002
		Proteasome-mediated ubiquitin independent protein catabolic process	1.72	0.0005	0.002
		Proteasomal protein catabolic process	1.72	0.0005	0.002
		Microtubule-based process	1.17	0.001	0.003
		ATP metabolic process	1.68	0.002	0.006
		Chromatin organisation	1.38	0.02	0.04
<b>BV-positive—BV intermediate</b>					
<b>Protein</b>	<b>Protein name</b>	<b>Fungal species</b>	<b>logFC</b>	<b>p-value</b>	<b>adj. p-value</b>
C4Y881	Gp_dh_C domain-containing protein	<i>Clavispora lusitaniae</i>	1.92	0.0001	0.007
		<b>Biological process</b>	<b>logFC</b>	<b>p-value</b>	<b>adj. p-value</b>
		Proteasome -mediated ubiquitin independent protein catabolic process	1.29	0.004	0.03
		Golgi to plasma membrane transport	2.62	0.006	0.03
		Vesicle docking involved in exocytosis	2.62	0.006	0.03
		Proteasomal protein catabolic process	1.46	0.009	0.03
		Proteasome-mediated ubiquitin-dependent protein catabolic process	1.46	0.009	0.03
		ATP metabolic process	1.27	0.006	0.03
		Endocytosis	1.65	0.01	0.03
		Histone exchange	1.65	0.01	0.03
		Mitotic actomyosin contractile ring assembly	1.65	0.01	0.03
		Regulation of nucleosome density	1.65	0.01	0.03

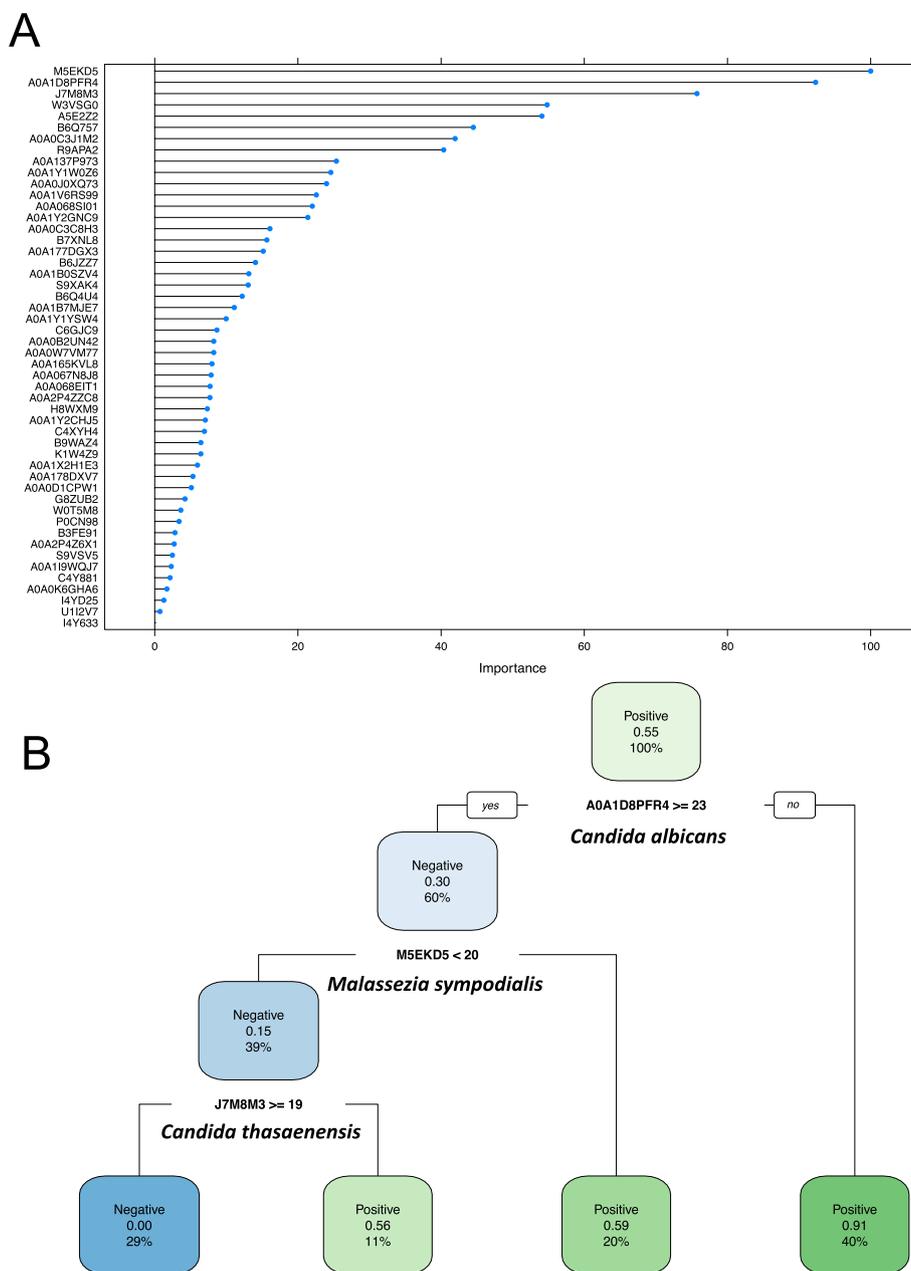
The false discovery rate (FDR) corrected *p*-values were obtained based on the log<sub>2</sub>-transformed intensity-based absolute quantification (iBAQ) values of each protein by applying the moderated *t*-test (limma R package). Proteins with FDR-adjusted *p*-values < 0.05 were considered as differentially abundant proteins

(See figure on next page.)

**Fig. 3** Fungal proteins associated with bacterial vaginosis (BV) status. **A** Unsupervised hierarchical clustering was used to group samples based on the log<sub>2</sub>-transformed imputed intensity-based absolute quantification (iBAQ) values of fungal proteins at visit 1. Red indicates the highest expression, and blue indicates the lowest expression. The darker the red banding pattern, the higher the iBAQ value. The darker the blue banding pattern, the lower the iBAQ value. Samples showed some clustering according to BV status indicated by the top dendrogram. BV status was defined based on Nugent score values. **B** Principal component analysis (PCA) was used to group individuals based on the log<sub>2</sub>-transformed imputed iBAQ intensities of fungal proteins for visit 1. Grouping was based on BV status and each point represented an individual women. BV-positive and BV-negative individuals tended to cluster together into their respective groups, with a few outliers. **C** PCA was used to group individuals based on the log<sub>2</sub>-transformed imputed iBAQ intensities of fungal proteins for visit 2. Grouping was based on BV status and each point represented individual women. BV-positive and BV-negative individuals tended to cluster together into their respective groups, with a few outliers



**Fig. 3** (See legend on previous page.)



**Fig. 4** **A** The importance of identified fungal proteins in discriminating the BV-positive group from the BV-negative group was determined by an ENSEMBLE bagging decision tree. M5EKD5 (*Malassezia sympodialis*) was the most important fungal protein distinguishing the BV-positive from the BV-negative group. **B** An ENSEMBLE bagging decision tree using previously identified important fungal proteins between the BV-positive and the BV-negative state to predict BV status. *Candida albicans* (A0A1D8PFR4) was selected as the root feature of the decision tree. Each fungal protein on the tree has an iBAQ 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

**Fungal protein functional profiles were associated with BV status**

In total, 17 and 12 GO biological processes were significantly differentially abundant between the BV positive and BV-negative groups at visits 1 and visit 2, respectively (Table 1; Supplementary Table 2, Additional

file 2). Both histone H4 acetylation and membrane raft polarization were the most relatively abundant biological processes across BV-negative and BV intermediate samples (Fig. 2C). In the BV-negative group, both biological processes cumulatively made up over 80% of the relative abundance at both visits, similar to the

BV intermediate group at visit 1 (Fig. 2C). Histone H4 acetylation and membrane raft polarization were significantly less abundant in the BV-positive group across both visits compared to the BV-negative groups (Table 1, Supplementary Table 2, Additional file 2).

The glycolytic process had the highest relative abundance in the BV-positive group (31.3% v1; 33.7% v2) (Fig. 2C) and was significantly more abundant in the BV-positive group at both visits in comparison to the BV-negative group (7.6% v1; 7.2% v2; Table 1; Supplementary Table 2, Additional file 2). Proteasome-mediated ubiquitin-dependent protein catabolic processes were also more abundant in BV-positive women (4.0% visit 1; 4.2% visit 2), and this was significant at both visits compared to BV-negative women (0.6% visit 1; 0.7% visit 2) (Table 1; Supplementary Table 2, Additional file 2). Biological processes that constituted less than 1% across each BV state were included under the category “other” for visual purposes.

Thirteen biological processes were more abundant 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, and several cellular responses to stresses (heat, osmotic, and oxidative) (Table 1). Ten biological processes were found to be significantly more abundant in the BV-positive compared to the BV intermediate group.

#### Fungal and bacterial taxa relative abundance correlated

We calculated Spearman’s rank correlation coefficients using iBAQ values to explore the strength of linear relationships between common fungal taxa that were identified to be differentially expressed from the limma *t*-test ( $n=10$ ) and bacteria (identified using our metaproteomics data) found to be associated with BV states in the literature ( $n=11$ ) (Fig. 5). *M. sympodialis* had the strongest positive correlation with the key BV-associated bacterium *Gardnerella vaginalis*, as well as positive associations with other BV-associated bacteria including *Mobiluncus mulieris*, *Prevotella amnii*, *Prevotella* sp. S7-1-8, *Megasphaera genomosp. type\_1*, *Megasphaera* sp. UPII 135-E, and *Sneathia amnii* (Fig. 5). *Candida* spp. (including *C. albicans*, *C. thasaenensis*, *C. lusitaniae*) correlated positively with multiple lactobacilli, including *Lactobacillus crispatus*, *L. iners*, and *L. jensenii* (Fig. 5). The strongest negative correlations were observed between *M. sympodialis* and *L. crispatus*, followed by *G. vaginalis* and *C. albicans* (Fig. 5). Furthermore, *C. albicans*, *C. thasaenensis*, and *C. lusitaniae* had many negative associations with bacteria: *P. amnii*, *Megasphaera genomosp. type\_1*, *Megasphaera* sp. UPII 135-E, and *S. amnii* (Fig. 5). *C. coronatus* also correlated negatively with lactobacilli: *L.*

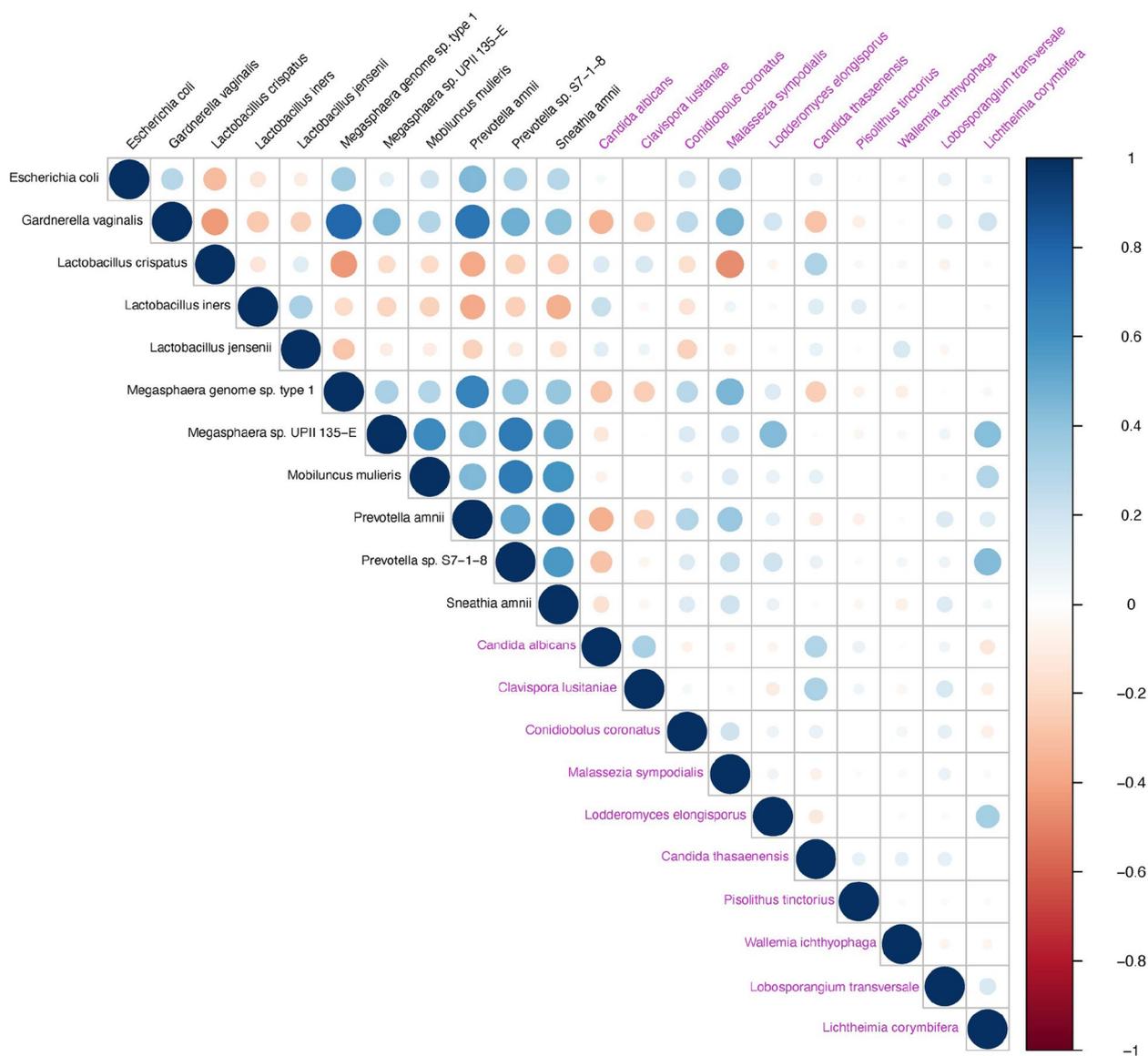
*crispatus*, *L. iners*, and *L. jensenii* (Fig. 5). These correlations reflect that *M. sympodialis* and *C. coronatus* were more prevalent in the BV-positive state than the BV-negative state, whereas *Candida* spp. were more prevalent in BV-negative compared to BV-positive women.

A Bayesian network was used to model the conditional dependencies among fungi and bacterial species, to capture interactions among species [81]. This provides a more accurate representation of the underlying relationships than what is represented by the linear relationships in Spearman’s correlations. This analysis confirmed many of the fungal-bacterial associations identified from the Spearman’s correlation (Fig. 6). The fungus *M. sympodialis* was again strongly associated with *L. crispatus* and *G. vaginalis* (Fig. 6). Likewise, *C. albicans* was strongly associated with *G. vaginalis*, *P. amnii*, and *Prevotella* sp. S77-1-8. *C. thasaenensis* was strongly associated with *Prevotella* sp. S7-1-8, and *L. crispatus*. Regarding the relationships between fungal taxa, the strongest associations were observed between *C. albicans*, *C. lusitaniae*, and *C. thasaenensis* (strength=1) (Fig. 6).

#### Fungal community composition was associated with pH, STIs and inflammation

A redundancy analysis was performed to determine whether any of the clinical variables significantly explained the variation in vaginal fungal community composition (Supplementary Fig. 3, Additional file 2). The variables vaginal pH ( $r^2=0.14$ ;  $p$ -value=0.002), Nugent score ( $r^2=0.28$ ;  $p$ -value=0.001), clue cells ( $r^2=0.14$ ;  $p$ -value=0.001), *C. trachomatis* infection ( $r^2=0.04$ ;  $p$ -value=0.020), pro-inflammatory cytokines ( $r^2=0.05$ ;  $p$ -value=0.044), and chemokines ( $r^2=0.15$ ;  $p$ -value=0.001) were each associated with fungal community composition. We also quantified the associations between fungal taxa ( $n=45$ ) and clinical factors at visit 1 using Spearman’s correlation tests (Supplementary Fig. 4, Additional file 2). Fungal taxa that correlated positively with BV-associated bacteria, including *M. sympodialis* and *A. alternata*, were positively correlated with increased Nugent score, a diagnostic marker of BV. Conversely, other yeast taxa (including *C. albicans*, *C. thasaenensis*, *C. lusitaniae*, *C. orthopsilosis*, and *T. delbrueckii*) were negatively correlated with Nugent score and clue cells (Supplementary Fig. 4, Additional file 2). The taxa *M. sympodialis*, *L. pennisporea*, *Wallemia mellicola*, *A. alternata*, and *R. vinicolor* had positive correlations with vaginal pH (Supplementary Fig. 4, Additional file 2).

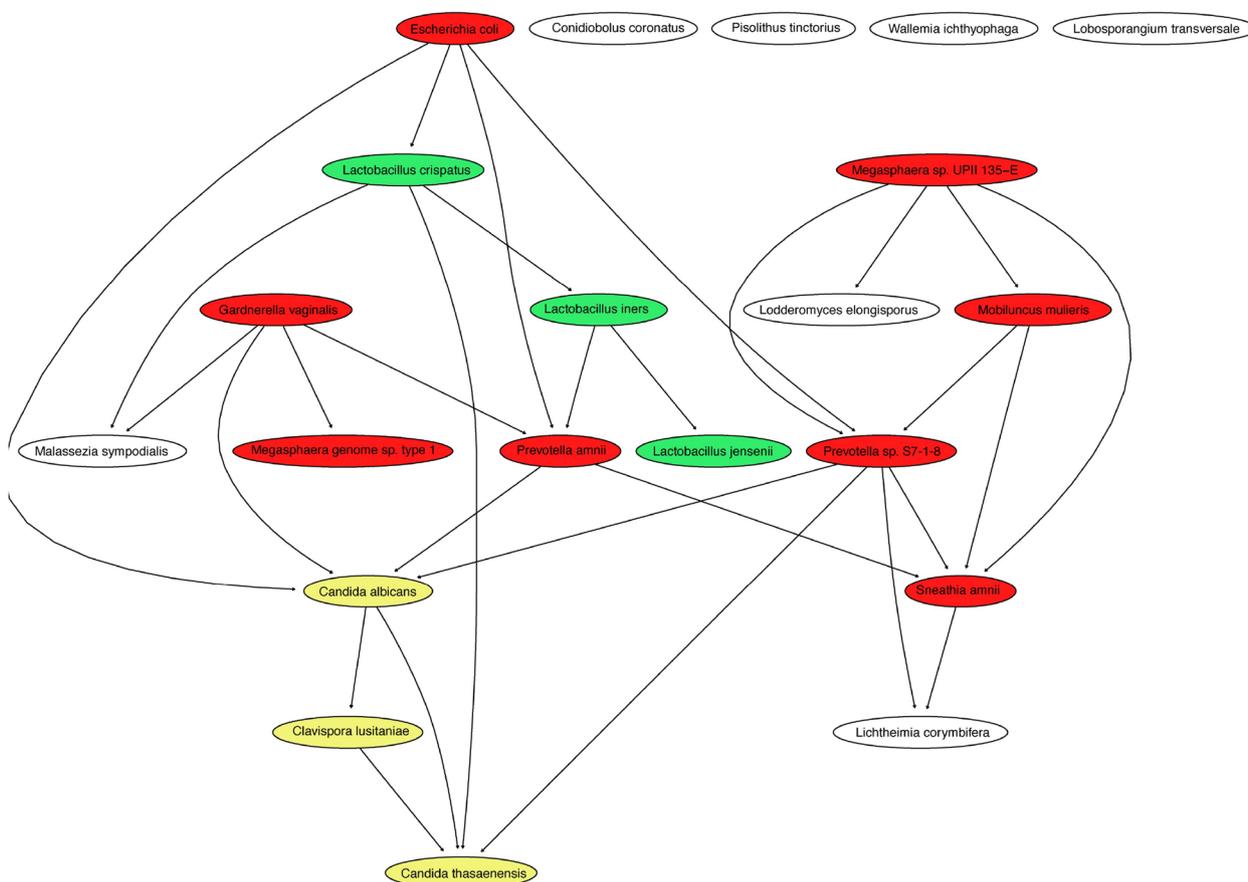
Regarding STI associations, the total iBAQ value for *M. sympodialis* had a negative correlation with the presence of *T. vaginalis*, whereas *R. globosum*, *Schizosaccharomyces japonicus*, *O. colligata*, and *Schizosaccharomyces cryophilus* had a positive correlation with



**Fig. 5** Correlations between differentially expressed common fungal ( $n=10$ ) and bacterial taxa associated with BV and inflammation from visit 1 samples ( $n=113$ ) using a Spearman's correlation. Fungal taxa are highlighted in purple. Red circles represent negative correlations (negative  $r$ -value), and blue circles represent positive correlations (positive  $r$ -value). The size of the circle represents the strength of the correlation ( $r$ -value), darker colors, and circle size represent stronger correlations. *M. sympodialis* had the strongest negative correlation with the bacterium *G. vaginalis*. The strongest positive correlation was observed between *C. albicans* and *G. vaginalis*

the presence of this STI (Supplementary Fig. 4, Additional file 2). The only significantly positively correlated fungal taxon with the presence of *N. gonorrhoeae* was *Cryptococcus neoformans*, an opportunistic fungus that has been previously reported in one study to infect the vagina ([82]; Supplementary Table 1 in Additional file 1). Similarly, the only significantly correlated fungal species with the presence of *M. genitalium* was *C. orthopsilosis*, a commonly found species in the vaginal mycobiome [22, 83] (Supplementary Table 1 in

Additional file 1); however, this correlation was negative (Supplementary Fig. 4, Additional file 2). A number of fungal taxa showed significant positive correlations with the visible presence of yeast cells (determined by Gram staining), including yeast species *C. neoformans*, *C. orthopsilosis*, *Cutaneotrichosporon oleaginosum*, *L. elongisporous*, *T. marneffeii*, *C. dubliniensis*, and *Trichoderma gamsii* (Supplementary Fig. 4, Additional file 2). The presence of visible yeast cells tended to positively correlate with the total fungal iBAQ value, but this



**Fig. 6** Bayesian network depicting the associations between differentially abundant common fungal taxa ( $n = 10$ ) and optimal (green) and non-optimal bacterial (red) taxa linked with BV and inflammation ( $n = 11$ ) at visit 1 ( $n = 113$ ). The selected BV and inflammation-associated bacteria were obtained from the literature. Arrows represent the microbial taxa that have an association with other microbial taxa, with only associations with a strength  $> 0.8$  depicted. *Candida* species are highlighted in yellow and the remainder of fungi are highlighted in white

correlation was not found to be significant (Supplementary Fig. 4, Additional file 2), and *C. trachomatis* showed a positive correlation, although insignificant (Supplementary Fig. 4, Additional file 2). *Candida* species (*C. albicans*, *C. orthopsilosis*, *C. thasaenensis*, *C. lusitanae*, *K. marxianus*, and *C. dubliniensis*) were found to be significantly positively correlated with total fungal iBAQ value, with *C. orthopsilosis* as the most strongly associated (Supplementary Fig. 4, Additional file 2). All correlations observed between fungal taxa and clinical variables were confirmed using samples for visit 2 ( $n = 89$ ).

A Bayesian Network analysis allowed us to confirm most of the fungal-clinical variable relationships identified from Spearman’s correlation analysis (Supplementary Fig. 5, Additional file 2). In this analysis, one of the strongest associations was between *M. sympodialis* with *T. vaginalis* and with Nugent score, which could be

a result of *T. vaginalis* and *M. sympodialis* both associating with BV (Supplementary Fig. 5, Additional file 2). Clue cells had strong associations with *C. thasaenensis*, while *C. albicans* was associated with Nugent score and vaginal pH.

### Discussion

Here, for the first time, we explore the vaginal metaproteome. Fungi constituted 0.4% of the FGT metaproteome and consisted of genera previously detected by molecular techniques in vaginal samples, *Candida*, *Aspergillus*, *Alternaria*, *Clavispora*, *Trichosporon*, and *Malassezia* [22, 83–87]. We also identified genera that have been reported less frequently, including *Cryptococcus*, *Penicillium*, *Torulaspora*, and *Lodderomyces* [84, 86, 88], and several unique genera, such as *Teratoramularia*, *Endocarpon*, and *Rhizoclosmatium* that have not been detected in previous human mycobiome studies, to our

knowledge. *Rhizoclostridium* has characteristically been found in damp and aquatic environments and *Endocarpon* in soil [89] and could be a result of contamination in the environment and swabs. However, many opportunistic pathogens (e.g., *Aspergillus fumigatus* and *C. neoformans*) are naturally found in soil or other environmental niches. Since these taxa contain ubiquitous housekeeping proteins (actin and ATPase), sequences are similar across fungal taxa, peptides can be shared between multiple organisms [90] and clear taxonomic annotation is not always feasible in metaproteomics or could represent unsequenced human fungi. Another hypothesis is that these taxa were present as a result of vaginal product use, hygiene habits, and water purity [91] since surfaces and cavities in the human body are exposed to exogenous environmental fungi that can enter the FGT from various external sources [88]. For example, a South African study by [92] showed that women insert a range of products into their vaginas, such as crushed tobacco, Knorrox stock cubes (a food seasoning), rice water, and water after soaking jellyfish or tree bark. However, more work is required to determine the influence of vaginal insertion and hygiene practices on the microbiome [93]. 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 [94]. Finally, we did not perform proteomics on an environmental swab, and some fungi may have been present in the collection room or on the swabs themselves. The FGT microbiome also differs between different geographic areas; therefore, it is possible these fungal taxa are normal commensals as no other studies have examined the FGT mycobiome in South African women to date [3, 53, 95]. Since many factors are able to drive change and dysbiosis in the microbiome; in the future, it will be beneficial to examine the contribution of environmental factors and geography to the FGT mycobiome.

Some fungal taxa found in the gut are transient and allochthonous [96–98]. Since the anus and the vagina are in close proximity, microorganisms can be translocated externally via wiping and/or during bathing; consequently, members of the gut microbiota may be shared with the FGT microbiota [98]. This may explain the presence of *Ustilago maydis* (a maize fungus) previously found in the gastrointestinal tract [99, 100], and the food-associated fungus *Hebeloma cylindrosporium* (an edible mushroom) in our study, since fungi consumed as part of the diet can be detected in gut microbiome studies [98, 99]. *U. maydis* has been previously found in the FGT [85] and has also been found to be closely related to *Malassezia*, a commonly detected skin fungus which has been found in the vagina [101]. Clinical samples are highly susceptible to environmental contamination, such

as airborne fungi [22]. Controlling all factors in a clinical environment proves difficult during sample collection; however, to determine whether our unique fungal identifications are true FGT colonizers, transient fungi, or merely contaminants, additional controls are required to account for environmental contaminants (e.g., air samples from clinics and laboratories, and skin samples from clinicians and laboratory technicians). Further, whether the proteins are from viable organisms remains to be demonstrated.

Understanding the FGT mycobiome in the context of BV is critical for the development of effective treatment strategies to reduce the prevalence of this condition and the associated adverse outcomes. The many correlations that were observed between fungal taxa and bacteria suggest that interactions between the mycobiome and the bacteriome likely occur and may be explained by biophysical and metabolic interactions, including quorum sensing, biofilm formation, production of secondary metabolites, and cellular signal transduction [102–106]. Negative correlations between *C. albicans* and the fungal taxa that had a greater relative abundance during the BV-positive state (*M. sympodialis*, *C. coronatus*, and *A. alternata*) may be due to differing abilities to survive in different BV states or may be due to quorum-sensing molecules such as farnesol. *C. albicans* secretes farnesol, which is able to control the morphology of other fungi, inhibiting hyphal growth, and early biofilm formation [107, 108], thereby controlling the pathogenesis of fungi. Fungi flourish as biofilms and *Candida*, *Aspergillus*, *Cryptococcus*, *Fusarium*, *Malassezia*, *Trichosporon*, and *Rhizopus* have all been shown to form biofilms [109–115]. In light of this, more work needs to be done to understand the interactions that occur between fungi in the FGT microbiome.

Several *Candida* spp. have been suggested to form part of an optimal microbiome, without causing a symptomatic infection [22, 50, 51, 53, 87]. The majority of *Candida* spp. in our study were prevalent, significantly more abundant in BV-negative women and positively correlated with *Lactobacillus* spp. Similarly, FGT colonization with *Candida* spp. has been shown to be more common in women with a *Lactobacillus*-dominated FGT than in women with dysbiosis [22, 44, 48]. Hong et al. (2016) and Moodley et al. (2002) [4, 116] also showed that women with detectable *C. albicans* were more likely to have a low Nugent score (<4). However, *C. albicans* is also detected and is the most dominant fungal taxon in BV-positive women [88, 117] since *Candida* spp. have the potential to be pathogenic and proliferate due to its dimorphic nature [118]. From previous studies, fungal mutants limited to either the hyphal or yeast form were avirulent, suggesting that the ability to reversibly transition between

morphological forms is essential for virulence [119, 120]. The transition of yeast to hyphae in *C. albicans* is usually triggered by certain environmental cues, such as high temperature (37 °C), pH, nutrition availability, and the transition can be inhibited by quorum-sensing molecules [107, 108]. The hyphal state is required for vaginal infection as it affects numerous properties that contribute to the virulence of this species, including the expression of morphology-dependent cell wall adhesins, secreted hydrolytic enzymes, proteases, and the ability to form biofilms [50, 121, 122]. Membrane raft polarization, a biological process that was inversely associated with BV in this study, has been suggested to contribute to hyphal growth and to the pathogenesis of *C. albicans* and may thus aid in the colonization of BV-negative women [123]. *Candida* can also metabolically adapt to the host micro-environment and the nutrients available [124]. For example, *C. albicans* cells are able to induce glycolytic genes during mucosal invasion [125, 126]. Glucose metabolism by *C. albicans* appears to be essential during infection because immune cells rely on glucose for survival [127]; therefore, glycolysis is thought to contribute to successful host colonization and pathogenesis [128]. This may explain the high relative abundance of the fungal glycolytic biological process during the BV-positive state.

A number of other differences in fungal GO biological processes between BV-positive and BV-negative women were identified. However, these may be a reflection of the differential abundance of the associated fungal taxa, rather than an indication of specific functional changes, as many biological processes identified as being differentially abundant have housekeeping roles. However, differences in processes such as histone H4 acetylation, chromatin organization, and heterochromatin assembly by small RNA may indicate alterations in fungal transcriptional activity between BV-negative and BV-positive women [129]. Additionally, the increased relative abundance of various stress response biological processes in BV-negative women may relate to environmental differences associated with BV status, such as vaginal pH.

*Malassezia* has been found in molecular studies to be a constituent of an optimal FGT microbiome [88], however *Malassezia* and *Lactobacillus* spp. have also been shown to co-occur infrequently [85]. There are two hypotheses that could explain this observation: either reduced colonization with *Lactobacillus* may give *Malassezia* an opportunity to colonize the FGT microbiome from the skin or specific fungal inhabitants could create an environment that inhibits *Lactobacillus* growth. *M. sympodialis* is a lipid-dependent yeast and has the ability to hydrolyze lipids in its surrounding environment by using lipolytic enzymes and integrate the fatty acids into the fungal cells, which are essential for their survival and

possibly play an essential role in the pathogenesis of this species [130]. Fatty acids have been demonstrated to have an inhibitory effect on fungal growth, which may explain the negative correlation between *M. sympodialis* and *C. albicans* (usually highly abundant in BV-negative individuals) [131]. Conversely, pathogens like *C. trachomatis* are obligate intracellular bacteria and auxotrophic for various metabolites, such as nucleotides, ATP, and amino acids [132]. They depend on host cells to provide these compounds, allowing them to replicate within vaginal epithelial cells. For instance, yeasts secrete an array of metabolites, mainly amino acids, and may benefit the feeding of species such as *C. trachomatis* [133]. The upregulation of various amino acid biosynthesis processes in our study may be attributed to this nutrient cross-feeding.

*Malassezia* was also found to be prevalent in the FGT of women with high-risk human papillomavirus types and has been shown to be associated with a high Nugent score ( $\geq 8$ ) [4]. Therefore, the overabundance of *M. sympodialis* during BV is plausible, as *Malassezia* appears to be associated with non-optimal conditions. The relative abundance of the fungal species *C. coronatus* was also greater during BV; however, only one study has shown the presence of this taxon in the FGT and more work would be required to confirm this species' involvement in the FGT [134].

This study had some limitations since mycobiome research is relatively new in the metaproteomics field, and this method requires a well-annotated protein sequence database that sufficiently represents the diversity of proteins present in samples. Therefore, many fungal taxa cannot be classified due to the scarcity of fungi in sequence databases. To overcome this limitation, we suggest the use of shotgun metagenomics to further examine more fungal taxa present in samples, in addition to the bacteriome. It should be noted that databases only allow us to uncover fungi that have already been cultured, and characterized, which remains difficult with ~90% of fungal species remaining unknown. Cultivation of many fungi remains challenging due to their diverse and specialized requirements, but there is significant motivation to advance fungal cultivation techniques to provide medical treatments and ecological insights. The depth of data collected could be improved by developing fractionation methods to increase the representation of lower abundance proteins 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. Future work should incorporate environmental swabs and fungal culture to confirm which aspects of the FGT mycobiome are true inhabitants of that space.

## Conclusion

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. Since we are only beginning to understand the potential role fungal communities play in the FGT microbiome, more work is needed to improve our understanding of the relationship between the FGT mycobiome and FGT diseases.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-025-02066-1>.

Additional file 1: Table S1. Characteristics of where the fungal species identified in this study using protein groups are typically found and their importance. Table S2. Details of parameters used to search raw mass spectrometry (MS) files generated from lateral vaginal wall samples from 113 women from Cape Town, South Africa with MaxQuant version 1.6.3.4. Table S3. Clinical characteristics of patients from which vaginal swabs were collected that formed part of the WISH dataset. Table S4. Prevalence of diseased states in the sample population from which vaginal samples were collected at visit 1 and visit 2. Table S5. Peptide and protein characteristics of fungal species assignments not found in the human body.

Additional file 2: Table S1. Fungal protein groups identified in this study and the percentage of these protein groups in the total fungal dataset. Figure S1. Correlations between fungal taxa were identified in this study at visit 1 ( $n = 113$ ) using a Spearman's correlation. 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 taxa were observed in comparison to negative correlations. The strongest positive correlation occurred between *Ordozpora colligata* and *Rhizoclostridium globosum*. The strongest negative correlation was between *Malassezia sympodialis* and *Torulapora delbrueckii*.

Additional file 2: Figure S2. Fungal relative abundance of each individual sampled was determined using metaproteomics. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) was used to evaluate the fungal metaproteome in lateral vaginal wall swabs from 123 women from Cape Town, South Africa. Raw MS files were processed with MaxQuant using a database that was created by manually curating a fungal pan proteome database and concatenated to a Metanovo database. The intensity-based absolute quantification (iBAQ) values for fungal proteins were normalised,  $\log_2$ -transformed, and imputed using the  $k$ -nearest neighbor method. Bacterial vaginosis (BV) groups were defined by Nugent score. 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. From left to right, BV negative, BV intermediate to BV positive. Table S2. List of differentially abundant proteins, fungal taxa, and biological processes between different bacterial vaginosis (BV) states (negative, positive, and intermediate) at visit 2. Figure S2. Fungal relative abundance of each individual sampled was determined using metaproteomics. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) was used to evaluate the fungal metaproteome in lateral vaginal wall swabs from 123 women from Cape Town, South Africa. Raw MS files were processed with MaxQuant using a database that was created by manually curating a fungal pan proteome database and concatenated to a Metanovo database. The intensity-based absolute quantification (iBAQ) values for fungal proteins were normalised,  $\log_2$ -transformed, and imputed using the  $k$ -nearest neighbor method. Bacterial vaginosis (BV) groups were defined by Nugent score. 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. From left to right, BV negative, BV intermediate to BV positive. Figure S3. Redundancy analysis (RDA) showing the relationships between fungal community composition and clinical factors at visit 1 ( $n = 113$ ). The length of arrows represents 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. Six variables (vaginal pH, Nugent score, clue cells, *C. trachomatis* infection, pro-inflammatory cytokines, and chemokines) were associated with fungal community composition. Nugent score associated the most with fungal community composition. Figure S4. Correlations between fungal taxa identified ( $n = 45$ ) and clinical variables at visit 1 ( $n = 113$ ) using a Spearman's correlation analysis. Fungal taxa are highlighted in purple. Red circles represent negative correlations (negative  $r$ -value), and blue circles represent positive correlations (positive  $r$ -value). The size of the circle represents the strength of the correlation ( $r$ -value), the darker colours and increased circle size represent stronger correlations. Yeast species negatively correlated with Nugent score and clue cells. *Malassezia sympodialis*, *Linderina pennisporea*, *Walleimia mellicola*, *Endocarpon pusillum*, *Hebeloma cylindrosporum*, *Alternaria alternata*, *Trichosporon asahii*, and *Rhizopogon vinicolor* had significant negative correlations with Nugent score and clue cells. *Candida* species were significantly positively correlated with total fungal iBAQ value. Figure S5. Bayesian Network depicting the associations between common differentially abundant fungal taxa ( $n = 14$ ) and clinical variables (blue) visit 1 ( $n = 113$ ). Arrows represent associations between fungal taxa and clinical variables, with only associations with strength values  $> 0.8$  depicted. Thickness of arrow represents strength of the association. The strongest associations were between *Malassezia sympodialis*, *L. transversale* and *Trichomonas vaginalis*. *M. sympodialis* and *C. albicans* showed an association with Nugent score. Clue cells had strong associations with *Candida thasaenensis*. Clinical variables are highlighted in blue.

Additional file 3: Table S1. Differentially abundant proteins, fungal taxa, and biological processes between bacterial vaginosis (BV) states using a non-parametric Wilcoxon rank sum test and adjusted limma test at visit 1. Table S2. Differentially abundant proteins, fungal taxa, and biological processes between bacterial vaginosis (BV) states using a non-parametric rank sum test and adjusted limma test at visit 2.

## Acknowledgements

We would like to acknowledge the participants of the Women's Initiative in Sexual Health Study.

## Authors' contributions

TKG analyzed the data and wrote the manuscript; AA, MP, AI, IA, and DLT analyzed the data and contributed to manuscript preparation; LB, EW, ZM, SD and SB conducted some of the laboratory experiments and contributed to manuscript preparation; LGB managed the clinical site for the WISH study, collected some of the clinical data and contributed to manuscript preparation; JMB and HBJ supervised the data analysis and contributed to manuscript preparation; JSP was Principal Investigator of the WISH cohort, supervised the collection of some of the data and contributed to manuscript preparation; NM was the primary supervisor of the data analysis and contributed to manuscript preparation; LM conceptualized the study, supervised the data collection, supervised the data analysis and contributed to manuscript preparation.

## Funding

The laboratory work, data analysis, and writing were supported by the Carnegie Corporation of New York, the South African National Research Foundation (NRF), the Poliomyelitis Research Foundation, the South African Medical Research Council, and the Australian National Health and Medical Research Council (PI: L. Masson). The WISH cohort was supported by the European and Developing Countries Clinical Trials Partnership Strategic Primer grant (SP.2011.41304.038; PI J.S. Passmore) and the South African Department of Science and Technology (DST/CON 0260/2012; PI J.S. Passmore). J.M.B. thanks the NRF for a SARChI Chair. D.L.T. was supported by the South African Tuberculosis Bioinformatics Initiative (SATBBI), a Strategic Health Innovation Partnership grant from the South African Medical Research Council, and the South African Department of Science and Technology.

## Declarations

### Ethics approval and consent to participate

The University of Cape Town Faculty of Health Sciences Human Research Ethics Committee (HREC REF: 267/2013) approved this study. Women  $\geq$  18 years provided written informed consent, those < 18 years provided informed assent, and informed consent was obtained from parents/guardians.

### Consent for publication

Not applicable.

### Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [1] partner repository with the dataset identifier PXD046053.

### Competing interests

The authors declare no competing interests.

### Author details

<sup>1</sup>Department of Integrative Biomedical Sciences, Computational Biology Division, University of Cape Town, Cape Town 7925, South Africa. <sup>2</sup>Department of Soil and Environment, Swedish University of Agricultural Sciences, 750 07 Uppsala, Sweden. <sup>3</sup>Division of Medical Virology, Department of Pathology, University of Cape Town, Cape Town 7925, South Africa. <sup>4</sup>Department of Integrative Biomedical Sciences, Division of Chemical and Systems Biology, University of Cape Town, Cape Town 7925, South Africa. <sup>5</sup>Centre for Proteomic and Genomic Research, Cape Town 7925, South Africa. <sup>6</sup>Laboratory of Human Pathologies Biology, Department of Biology and Genomic Center of Human Pathologies, Mohammed V University in Rabat, Rabat, Morocco. <sup>7</sup>Seattle Children's Research Institute, University of Washington, Seattle, WA 98101, USA. <sup>8</sup>Institute of Infectious Disease and Molecular Medicine (IDM), University of Cape Town, Cape Town 7925, South Africa. <sup>9</sup>Bioinformatics Unit, South African Tuberculosis Bioinformatics Initiative, Stellenbosch University, Stellenbosch 7602, South Africa. <sup>10</sup>DST-NRF Centre of Excellence for Biomedical Tuberculosis Research, Stellenbosch University, Stellenbosch 7602, South Africa. <sup>11</sup>Desmond Tutu HIV Centre, Cape Town, University of Cape Town, Cape Town 7925, South Africa. <sup>12</sup>Division of Immunology, Department of Pathology, University of Cape Town, Cape Town 7925, South Africa. <sup>13</sup>Centre for the AIDS Programme of Research in South Africa (CAPRISA), Durban 4013, South Africa. <sup>14</sup>National Health Laboratory Service, Cape Town 7925, South Africa. <sup>15</sup>Centre for Infectious Diseases Research (CIDRI) in Africa Wellcome Trust Centre, University of Cape Town, Cape Town 7925, South Africa. <sup>16</sup>Women's, Children's and Adolescents' Health and Disease Elimination Programs, Life Sciences Discipline, Burnet Institute, Melbourne 3004, Australia. <sup>17</sup>Central Clinical School, Monash University, Melbourne 3004, Australia.

Received: 4 September 2023 Accepted: 11 February 2025

Published online: 19 March 2025

## References

- Cheng K, Ning Z, Zhang X, Mayne J, Figeys D. Separation and characterization of human microbiomes by metaproteomics. *TrAC Trends Anal Chem.* 2018;108:221–30.
- Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. *Nature.* 2011;473:174–80.
- Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, et al. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A.* 2011;108 Suppl 1 Suppl 1:4680–7.
- Hong KH, Hong SK, Cho SI, Ra E, Han KH, Kang SB, et al. 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;36:441–9.
- Iliev ID, Leonardi I. Fungal dysbiosis: immunity and interactions at mucosal barriers. *Nat Rev Immunol.* 2017;17:635–46.
- Nash AK, Auchtung TA, Wong MC, Smith DP, Gesell JR, Ross MC, et al. The gut mycobiome of the Human Microbiome Project healthy cohort. *Microbiome.* 2017;5:153.
- Happel A-U, Gasper M, Balle C, Konstantinus I, Gamielien H, Dabee S, et al. Persistent, asymptomatic colonization with *Candida* is associated with elevated frequencies of highly activated cervical Th17-like cells and related cytokines in the reproductive tract of South African adolescents. *Microbiol Spectr.* 2022;10:e01626-e1721.
- Bradford LL, Ravel J. The vaginal mycobiome: a contemporary perspective on fungi in women's health and diseases. *Virulence.* 2017;8:342–51.
- Dworecka-Kaszak B, Dąbrowska J, Kaszak I. The mycobiome – a friendly cross-talk between fungal colonizers and their host. *Ann Parasitol.* 2016;62(3):175–84.
- Denning DW, Kneale M, Sobel JD, Rautemaa-Richardson R. Global burden of recurrent vulvovaginal candidiasis: a systematic review. *Lancet Infect Dis.* 2018;18:e339–47.
- Schwartz I, Boyles T, Kenyon C, Hoving J, Brown G, Denning D. The estimated burden of fungal disease in South Africa. *S Afr Med J.* 2019;109:885.
- Martin HLJ, Nyange PM, Richardson BA, Lavreys L, Mandaliya K, Jackson DJ, et al. Hormonal contraception, sexually transmitted diseases, and risk of heterosexual transmission of human immunodeficiency virus type 1. *J Infect Dis.* 1998;178:1053–9.
- Hester RA, Kennedy SB. Candida infection as a risk factor for HIV transmission. *J Womens Health.* 2002;2003(12):487–94.
- Röttingen JA, Cameron DW, Garnett GP. A systematic review of the epidemiologic interactions between classic sexually transmitted diseases and HIV: how much really is known? *Sex Transm Dis.* 2001;28:579–97.
- Kapiga SH, Lyamuya EF, Lwihula GK, Hunter DJ. The incidence of HIV infection among women using family planning methods in Dar es Salaam. *Tanzania AIDS Lond Engl.* 1998;12:75–84.
- Taha TE, Hoover DR, Dallabetta GA, Kumwenda NJ, Mtimavalye LA, Yang LP, et al. Bacterial vaginosis and disturbances of vaginal flora: association with increased acquisition of HIV. *AIDS Lond Engl.* 1998;12:1699–706.
- Kilmarx PH, Limpakarnjanarat K, Mastro TD, Saisorn S, Kaewkungwal J, Korattana S, et al. HIV-1 seroconversion in a prospective study of female sex workers in northern Thailand: continued high incidence among brothel-based women. *AIDS Lond Engl.* 1998;12:1889–98.
- Mostad SB, Overbaugh J, DeVange DM, Welch MJ, Chohan B, Mandaliya K, et al. Hormonal contraception, vitamin A deficiency, and other risk factors for shedding of HIV-1 infected cells from the cervix and vagina. *Lancet Lond Engl.* 1997;350:922–7.
- Ackerman AL, Underhill DM. The mycobiome of the human urinary tract: potential roles for fungi in urology. *Ann Transl Med.* 2017;5:31.
- Perfect JR, Casadevall A. Fungal molecular pathogenesis: what can it do and why do we need it? In: Heitman J, Filler SG, Edwards Jr. JE, Mitchell AP, editors. *Molecular Principles of Fungal Pathogenesis.* ASM Press; 2006. p. 1–11.
- Zanoni BC, Archary M, Buchan S, Katz IT, Haberer JE. Systematic review and meta-analysis of the adolescent HIV continuum of care in South Africa: the Cresting Wave. *BMJ Glob Health.* 2016;1:e000004.
- Drell T, Lillsaar T, Tummeleht L, Simm J, Aaspõllu A, Väin E, et al. Characterization of the vaginal micro- and mycobiome in asymptomatic reproductive-age Estonian women. *PLoS ONE.* 2013;8:e54379.
- Cui L, Morris A, Ghedin E. The human mycobiome in health and disease. *Genome Med.* 2013;5:63.
- Rizzetto L, De Filippo C, Cavalieri D. Richness and diversity of mammalian fungal communities shape innate and adaptive immunity in health and disease. *Eur J Immunol.* 2014;44:3166–81.
- Seed PC. The human mycobiome. *Cold Spring Harb Perspect Med.* 2014;5:a019810.
- Esteve-Zarzoso B, Belloch C, Uruburu F, Querol A. Identification of yeasts by RFLP analysis of the 5.8S rRNA gene and the two ribosomal internal transcribed spacers. *Int J Syst Bacteriol.* 1999;49 Pt 1:329–37.
- De Filippis F, Laiola M, Blaiotta G, Ercolini D. Different amplicon targets for sequencing-based studies of fungal diversity. *Appl Environ Microbiol.* 2017;83:e00905-e917.
- Forbes JD, Bernstein CN, Tremlett H, Van Domselaar G, Knox NC. A fungal world: could the gut mycobiome be involved in neurological disease? *Front Microbiol.* 2018;9:3249.
- Brooks B, Mueller RS, Young JC, Morowitz MJ, Hettich RL, Banfield JF. Strain-resolved microbial community proteomics reveals simultaneous aerobic and anaerobic function during gastrointestinal tract colonization of a preterm infant. *Front Microbiol.* 2015;6:654.

30. Greenblum S, Carr R, Borenstein E. Extensive strain-level copy-number variation across human gut microbiome species. *Cell*. 2015;160:583–94.
31. Kolmeder CA, Salojarvi J, Ritari J, de Been M, Raes J, Falony G, et al. Faecal metaproteomic analysis reveals a personalized and stable functional microbiome and limited effects of a probiotic intervention in adults. *PLoS ONE*. 2016;11:e0153294.
32. Gosmann C, Anahtar MN, Handley SA, Farcasanu M, Abu-Ali G, Bowman BA, et al. Lactobacillus-deficient cervicovaginal bacterial communities are associated with increased HIV acquisition in young South African women. *Immunity*. 2017;46:29–37.
33. Jones ML, Martoni CJ, Ganopolsky JG, Labbé A, Prakash S. The human microbiome and bile acid metabolism: dysbiosis, dysmetabolism, disease and intervention. *Expert Opin Biol Ther*. 2014;14:467–82.
34. Low N, Chersich MF, Schmidlin K, Egger M, Francis SC, van de Wijgert JHHM, et al. Intravaginal practices, bacterial vaginosis, and HIV infection in women: individual participant data meta-analysis. *PLoS Med*. 2011;8:e1000416.
35. Schwebke JR. Gynecologic consequences of bacterial vaginosis. *Obstet Gynecol Clin North Am*. 2003;30:685–94.
36. Soper DE, Bump RC, Hurt WG. Bacterial vaginosis and trichomoniasis vaginitis are risk factors for cuff cellulitis after abdominal hysterectomy. *Am J Obstet Gynecol*. 1990;163:1016–21; discussion 1021–1023.
37. Pirota MV, Gunn JM, Chondros P. "Not thrush again!" Women's experience of post-antibiotic vulvovaginitis. *Med J Aust*. 2003;179:43–6.
38. Liu M-B, Xu S-R, He Y, Deng G-H, Sheng H-F, Huang X-M, et al. Diverse vaginal microbiomes in reproductive-age women with vulvovaginal candidiasis. *PLoS ONE*. 2013;8:e79812.
39. Oever JT, Netea MG. The bacteriome-mycobiome interaction and antifungal host defense. *Eur J Immunol*. 2014;44:3182–91.
40. Krüger W, Vielreicher S, Kapitan M, Jacobsen ID, Niemiec MJ. Fungal-bacterial interactions in health and disease. *Pathog Basel Switz*. 2019;8:70.
41. Changying Z, Ying Li, Bin C, Yue Kaile Su, Jing ZX, et al. Mycobiome study reveals different pathogens of vulvovaginal candidiasis shape characteristic vaginal bacteriome. *Microbiol Spectr*. 2023;11:e03152-e3222.
42. Donders G, Bellen G, Ausma J, Verguts L, Vaneldere J, Hinoul P, et al. 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 Off Publ Eur Soc Clin Microbiol*. 2011;30:59–63.
43. Wei Q, Fu B, Liu J, Zhang Z, Zhao T. *Candida albicans* and bacterial vaginosis can coexist on Pap smears. *Acta Cytol*. 2012;56:515–9.
44. Biagi E, Vitali B, Pugliese C, Candela M, Donders GGG, Brigidi P. Quantitative variations in the vaginal bacterial population associated with asymptomatic infections: a real-time polymerase chain reaction study. *Eur J Clin Microbiol Infect Dis Off Publ Eur Soc Clin Microbiol*. 2009;28:281–5.
45. Brown SE, Schwartz JA, Robinson CK, O'Hanlon DE, Bradford LL, He X, et al. The vaginal microbiota and behavioral factors associated with genital candida albicans detection in reproductive-age women. *Sex Transm Dis*. 2019;46:753–8.
46. Eastment MC, Balkus JE, Richardson BA, Srinivasan S, Kimani J, Anzala O, et al. Association between vaginal bacterial microbiota and vaginal yeast colonization. *J Infect Dis*. 2021;223:914–23.
47. van de Wijgert JHHM, Borgdorff H, Verhelst R, Crucitti T, Francis S, Verstraelen H, et al. The vaginal microbiota: what have we learned after a decade of molecular characterization? *PLoS ONE*. 2014;9:e105998.
48. Zhou X, Westman R, Hickey R, Hansmann MA, Kennedy C, Osborn TW, et al. Vaginal microbiota of women with frequent vulvovaginal candidiasis. *Infect Immun*. 2009;77:4130–5.
49. Underhill DM, Iliev ID. The mycobiota: interactions between commensal fungi and the host immune system. *Nat Rev Immunol*. 2014;14:405–16.
50. Beigi RH, Meyn LA, Moore DM, Krohn MA, Hillier SL. Vaginal yeast colonization in nonpregnant women: a longitudinal study. *Obstet Gynecol*. 2004;104(5 Pt 1):926–30.
51. Cotch MF, Hillier SL, Gibbs RS, Eschenbach DA. Epidemiology and outcomes associated with moderate to heavy *Candida* colonization during pregnancy. Vaginal Infections and Prematurity Study Group. *Am J Obstet Gynecol*. 1998;178:374–80.
52. Förster TM, Mogavero S, Dräger A, Graf K, Polke M, Jacobsen ID, et al. Enemies and brothers in arms: *Candida albicans* and gram-positive bacteria. *Cell Microbiol*. 2016;18:1709–15.
53. McClelland RS, Richardson BA, Hassan WM, Graham SM, Kiarie J, Baeten JM, et al. Prospective study of vaginal bacterial flora and other risk factors for vulvovaginal candidiasis. *J Infect Dis*. 2009;199:1883–90.
54. Jang SJ, Lee K, Kwon B, You HJ, Ko G. Vaginal lactobacilli inhibit growth and hyphae formation of *Candida albicans*. *Sci Rep*. 2019;9:8121.
55. Li T, Liu Z, Zhang X, Chen X, Wang S. Local probiotic lactobacillus crispatus and lactobacillus delbrueckii exhibit strong antifungal effects against Vulvovaginal Candidiasis in a rat model. *Front Microbiol*. 2019;10:1033.
56. Parolin C, Marangoni A, Laghi L, Foschi C, Nahui Palomino RA, Calonghi N, et al. Isolation of vaginal lactobacilli and characterization of anti-candida activity. *PLoS ONE*. 2015;10:e0131220.
57. Donati L, Di Vico A, Nucci M, Quagliozi L, Spagnuolo T, Labianca A, et al. Vaginal microbial flora and outcome of pregnancy. *Arch Gynecol Obstet*. 2010;281:589–600.
58. O'Hanlon DE, Moench TR, Cone RA. Vaginal pH and microbicidal lactic acid when lactobacilli dominate the microbiota. *PLoS ONE*. 2013;8:e80074.
59. Barnabas SL, Dabee S, Passmore J-AS, Jaspan HB, Lewis DA, Jaumdally SZ, et al. Converging epidemics of sexually transmitted infections and bacterial vaginosis in southern African female adolescents at risk of HIV. *Int J STD Aids*. 2018;29:531–9.
60. Dabee S, Barnabas SL, Lennard KS, Jaumdally SZ, Gamielien H, Balle C, et al. Defining characteristics of genital health in South African adolescent girls and young women at high risk for HIV infection. *PLoS ONE*. 2019;14:e0213975.
61. Alisoltani A, Manhanzva MT, Potgieter M, Balle C, Bell L, Ross E, et al. Microbial function and genital inflammation in young South African women at high risk of HIV infection. *Microbiome*. 2020;8:1–21.
62. Potgieter MG, Nel AJM, Fortuin S, Garnett S, Wendoh JM, Tabb DL, et al. MetaNovo: An open-source pipeline for probabilistic peptide discovery in complex metaproteomic datasets. *PLoS Comput Biol*. 2023;19:e1011163.
63. Huang T, Wang J, Yu W, He Z. Protein inference: a review. *Brief Bioinform*. 2012;13:586–614.
64. Gatto L, Lilley KS. MSnbase—an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinforma Oxf Engl*. 2012;28:288–9.
65. Lazar C, Gatto L, Ferro M, Bruley C, Burger T. Accounting for the multiple natures of missing values in label-free quantitative proteomics data sets to compare imputation strategies. *J Proteome Res*. 2016;15:1116–25.
66. Smyth GK. *limma: Linear Models for Microarray Data*. In: Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S, editors. *bioinformatics and computational biology solutions using R and bioconductor*. New York, NY: Springer New York; 2005. p. 397–420.
67. Warnes MG, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A. Package 'gplots': Various R programming tools for plotting data. 2016;30:112–9.
68. Rohart F, Gautier B, Singh A, Lê Cao K-A. mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol*. 2017;13:e1005752.
69. Gardner HL, Dukes CD. New etiologic agent in nonspecific bacterial vaginitis. *Science*. 1954;120:853.
70. Schwebke JR, Lawing LF. Prevalence of *Mobiluncus* spp among women with and without bacterial vaginosis as detected by polymerase chain reaction. *Sex Transm Dis*. 2001;28:195–9.
71. Fethers K, Twin J, Fairley CK, Fowkes FJL, Garland SM, Fehler G, et al. Bacterial vaginosis (BV) candidate bacteria: associations with BV and behavioural practices in sexually-experienced and inexperienced women. *PLoS ONE*. 2012;7:e30633.
72. McMillan A, Rulisa S, Sumarah M, Macklaim JM, Renaud J, Bisanz JE, et al. A multi-platform metabolomics approach identifies highly specific biomarkers of bacterial diversity in the vagina of pregnant and non-pregnant women. *Sci Rep*. 2015;5:14174.
73. Ceccarani C, Foschi C, Parolin C, D'Antuono A, Gaspari V, Consolandi C, et al. Diversity of vaginal microbiome and metabolome during genital infections. *Sci Rep*. 2019;9:14095.
74. Harrell Jr FE, Harrell Jr MF. Package 'hmisc'. CRAN2018. 2019;235–6.
75. Wei T, Wei MT. Package 'corrplot'. Statistician. 2016;56:316–24.
76. Scutari M, Cheng K, Ning Z, Zhang X, Mayne J, Figeys D. Learning Bayesian networks with the bnlearn R package. *J Stat Softw*. 2018;35:1–22.

77. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin P, O'Hara B, et al. *Vegan: community ecology package*. R Package Version. 2015;22-1(2):1-2.
78. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw*. 2008;28:1-26.
79. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. *Scikit-learn: Machine learning in Python*. *J Mach Learn Res*. 2011;12:2825-30.
80. Armengaud J. Metaproteomics to understand how microbiota function: the crystal ball predicts a promising future. *Environ Microbiol*. 2023;25:115-25.
81. McGeachie MJ, Sordillo JE, Gibson T, Weinstock GM, Liu Y-Y, Gold DR, et al. Longitudinal prediction of the infant gut microbiome with dynamic Bayesian networks. *Sci Rep*. 2016;6:20359.
82. Chen CK, Chang DY, Chang SC, Lee EF, Huang SC, Chow SN. Cryptococcal infection of the vagina. *Obstet Gynecol*. 1993;81 5 (Pt 2):867-9.
83. Guo R, Zheng N, Lu H, Yin H, Yao J, Chen Y. Increased diversity of fungal flora in the vagina of patients with recurrent vaginal candidiasis and allergic rhinitis. *Microb Ecol*. 2012;64:918-27.
84. Fornari G, Vicente VA, Gomes RR, Muro MD, Pinheiro RL, Ferrari C, et al. Susceptibility and molecular characterization of *Candida* species from patients with vulvovaginitis. *Braz J Microbiol Publ Braz Soc Microbiol*. 2016;47:373-80.
85. Godoy-Vitorino F, Romaguera J, Zhao C, Vargas-Robles D, Ortiz-Morales G, Vázquez-Sánchez F, et al. Cervicovaginal fungi and bacteria associated with cervical intraepithelial neoplasia and high-risk human papillomavirus infections in a hispanic population. *Front Microbiol*. 2018;9:2533.
86. Abbasi Nejat Z, Farahay S, Falahati M, Ashrafi Khozani M, Hosseini AF, Faizy A, et al. Molecular identification and antifungal susceptibility pattern of non-albicans *Candida* species isolated from Vulvovaginal Candidiasis. *Iran Biomed J*. 2017;22:33-41.
87. Zheng N-N, Guo X-C, Lv W, Chen X-X, Feng G-F. Characterization of the vaginal fungal flora in pregnant diabetic women by 18S rRNA sequencing. *Eur J Clin Microbiol Infect Dis Off Publ Eur Soc Clin Microbiol*. 2013;32:1031-40.
88. Lehtoranta L, Hibberd AA, Yeung N, Laitila A, Maukonen J, Ouwehand AC. Characterization of vaginal fungal communities in healthy women and women with bacterial vaginosis (BV); a pilot study. *Microb Pathog*. 2021;161 Pt A:105055.
89. Belnap J, Büdel B, Lange OL. Biological soil crusts: characteristics and distribution. In: Belnap J, Lange OL, editors. *Biological Soil Crusts: Structure, Function, and Management*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2003:3-30.
90. Kleiner M, Thorson E, Sharp CE, Dong X, Liu D, Li C, et al. Assessing species biomass contributions in microbial communities via metaproteomics. *Nat Commun*. 2017;8:1558.
91. Humphries H, Mehoul-Loko C, Phakathi S, Mdladla M, Fynn L, Knight L, et al. "You'll always stay right": understanding vaginal products and the motivations for use among adolescent and young women in rural KZN. *Cult Health Sex*. 2019;21:95-107.
92. Gafos M, Mzimela M, Sukazi S, Pool R, Montgomery C, Eford J. Intravaginal insertion in KwaZulu-Natal: sexual practices and preferences in the context of microbicide gel use. *Cult Health Sex*. 2010;12:929-42.
93. Noyes N, Cho K-C, Ravel J, Forney LJ, Abdo Z. Associations between sexual habits, menstrual hygiene practices, demographics and the vaginal microbiome as revealed by Bayesian network analysis. *PLoS ONE*. 2018;13:e0191625.
94. Ma B, Forney LJ, Ravel J. Vaginal microbiome: rethinking health and disease. *Annu Rev Microbiol*. 2012;66:371-89.
95. Srinivasan S, Hoffman NG, Morgan MT, Matsen FA, Fiedler TL, Hall RW, et al. Bacterial communities in women with bacterial vaginosis: high resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. *PLoS ONE*. 2012;7:e37818.
96. El Aila NA, Tency I, Claeys G, Verstraelen H, Saerens B, Santiago GLDS, et al. Identification and genotyping of bacteria from paired vaginal and rectal samples from pregnant women indicates similarity between vaginal and rectal microflora. *BMC Infect Dis*. 2009;9:167.
97. El Aila NA, Tency I, Saerens B, De Backer E, Cools P, dos Santos Santiago GL, et al. Strong correspondence in bacterial loads between the vagina and rectum of pregnant women. *Res Microbiol*. 2011;162:506-13.
98. Suhr MJ, Banjará N, Hallen-Adams HE. Sequence-based methods for detecting and evaluating the human gut mycobiome. *Lett Appl Microbiol*. 2016;62:209-15.
99. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*. 2014;505:559-63.
100. Ott SJ, Kühbacher T, Musfeldt M, Rosenstiel P, Hellmig S, Rehman A, et al. Fungi and inflammatory bowel diseases: alterations of composition and diversity. *Scand J Gastroenterol*. 2008;43:831-41.
101. Xu J, Saunders CW, Hu P, Grant RA, Boekhout T, Kuramae EE, et al. Dandruff-associated *Malassezia* genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. *Proc Natl Acad Sci*. 2007;104:18730-5.
102. O'Toole GA. Classic spotlight: quorum sensing and the multicellular life of unicellular organisms. *J Bacteriol*. 2016;198:601.
103. Polke M, Jacobsen ID. Quorum sensing by farnesol revisited. *Curr Genet*. 2017;63:791-7.
104. Papenfort K, Bassler BL. Quorum sensing signal-response systems in Gram-negative bacteria. *Nat Rev Microbiol*. 2016;14:576-88.
105. Rio RVM. Don't bite the hand that feeds you. *Cell Host Microbe*. 2017;21:552-4.
106. Whiteley M, Diggle SP, Greenberg EP. Progress in and promise of bacterial quorum sensing research. *Nature*. 2017;551:313-20.
107. Lorek J, Pöggeler S, Weide MR, Breves R, Bockmühl DP. Influence of farnesol on the morphogenesis of *Aspergillus niger*. *J Basic Microbiol*. 2008;48:99-103.
108. Semighini CP, Murray N, Harris SD. Inhibition of *Fusarium graminearum* growth and development by farnesol. *FEMS Microbiol Lett*. 2008;279:259-64.
109. Beauvais A, Schmidt C, Guadagnini S, Roux P, Perret E, Henry C, et al. An extracellular matrix glues together the aerial-grown hyphae of *Aspergillus fumigatus*. *Cell Microbiol*. 2007;9:1588-600.
110. Cannizzo FT, Eraso E, Ezkurra PA, Villar-Vidal M, Bollo E, Castellá G, et al. Biofilm development by clinical isolates of *Malassezia pachydermatis*. *Med Mycol*. 2007;45:357-61.
111. Di Bonaventura G, Pompilio A, Picciani C, Iezzi M, D'Antonio D, Piccolomini R. Biofilm formation by the emerging fungal pathogen *Trichosporon asahii*: development, architecture, and antifungal resistance. *Antimicrob Agents Chemother*. 2006;50:3269-76.
112. Harriott MM, Lilly EA, Rodriguez TE, Fidel PL, Noverr MC. *Candida albicans* forms biofilms on the vaginal mucosa. *Microbiol Read Engl*. 2010;156(Pt 12):3635-44.
113. Imamura Y, Chandra J, Mukherjee PK, Lattif AA, Szczotka-Flynn LB, Pearlman E, et al. *Fusarium* and *Candida albicans* biofilms on soft contact lenses: model development, influence of lens type, and susceptibility to lens care solutions. *Antimicrob Agents Chemother*. 2008;52:171-82.
114. Martinez LR, Casadevall A. Specific antibody can prevent fungal biofilm formation and this effect correlates with protective efficacy. *Infect Immun*. 2005;73:6350-62.
115. Nobile CJ, Johnson AD. *Candida albicans* biofilms and human disease. *Annu Rev Microbiol*. 2015;69:71-92.
116. Moodley P, Connolly C, Sturm AW. Interrelationships among human immunodeficiency virus type 1 infection, bacterial vaginosis, trichomoniasis, and the presence of yeasts. *J Infect Dis*. 2002;185:69-73.
117. Pramanick R, Nathani N, Warke H, Mayadeo N, Aranha C. Vaginal dysbiotic microbiome in women with no symptoms of genital infections. *Front Cell Infect Microbiol*. 2021;11:760459.
118. Mayer FL, Wilson D, Hube B. *Candida albicans* pathogenicity mechanisms. *Virulence*. 2013;4:119-28.
119. Lo H-J, Köhler JR, DiDomenico B, Loebenberg D, Cacciapuoti A, Fink GR. Nonfilamentous *C. albicans* mutants are avirulent. *Cell*. 1997;90:939-49.
120. Schweizer A, Rupp S, Taylor BN, Röllinghoff M, Schröppel K. The TEA/ATTS transcription factor CaTec1p regulates hyphal development and virulence in *Candida albicans*. *Mol Microbiol*. 2000;38:435-45.
121. Moyes DL, Wilson D, Richardson JP, Mogavero S, Tang SX, Wernecke J, et al. Candidalysin is a fungal peptide toxin critical for mucosal infection. *Nature*. 2016;532:64-8.
122. Deorukhkar SC. Virulence traits contributing to pathogenicity of *Candida* species. *J Microbiol Exp*. 2017;5:00140.
123. Martin SW, Konopka JB. Lipid raft polarization contributes to hyphal growth in *Candida albicans*. *Eukaryot Cell*. 2004;3:675-84.

124. Brown AJP, Brown GD, Netea MG, Gow NAR. Metabolism impacts upon *Candida* immunogenicity and pathogenicity at multiple levels. *Trends Microbiol.* 2014;22:614–22.
125. Wilson D, Thewes S, Zakikhany K, Fradin C, Albrecht A, Almeida R, et al. Identifying infection-associated genes of *Candida albicans* in the post-genomic era. *FEMS Yeast Res.* 2009;9:688–700.
126. Zakikhany K, Naglik JR, Schmidt-Westhausen A, Holland G, Schaller M, Hube B. In vivo transcript profiling of *Candida albicans* identifies a gene essential for interepithelial dissemination. *Cell Microbiol.* 2007;9:2938–54.
127. Tucey TM, Verma J, Harrison PF, Snelgrove SL, Lo TL, Scherer AK, et al. Glucose homeostasis is important for immune cell viability during *Candida* challenge and host survival of systemic fungal infection. *Cell Metab.* 2018;27:988–1006.e7.
128. Brock M. Fungal metabolism in host niches. *Curr Opin Microbiol.* 2009;12:371–6.
129. Lai Y, Wang L, Zheng W, Wang S. Regulatory roles of histone modifications in filamentous fungal pathogens. *J Fungi.* 2022;8:565.
130. Juntachai W, Oura T, Murayama SY, Kajiwara S. The lipolytic enzymes activities of *Malassezia* species. *Med Mycol.* 2009;47:477–84.
131. Guimarães A, Venâncio A. The potential of fatty acids and their derivatives as antifungal agents: a review. *Toxins.* 2022;14:188.
132. Iliffe-Lee ER, McClarty G. Regulation of carbon metabolism in *Chlamydia trachomatis*. *Mol Microbiol.* 2000;38:20–30.
133. Gabrielli N, Maga-Nteve C, Kafka E, Rettel M, Loeffler J, Kamrad S, et al. Unravelling metabolic cross-feeding in a yeast–bacteria community using <sup>13</sup>C-based proteomics. *Mol Syst Biol.* 2023;19:e11501.
134. Subramanian C, Sobel JD. A case of *Conidiobolus coronatus* in the vagina. *Med Mycol.* 2011;49:427–9.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.