Metataxonomic Methods Assessment and Investigation of the Gut Mycobiota in Pediatric-Onset Multiple Sclerosis

By
Nelson Yuk Hay Mok

A thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
In partial fulfillment of the requirements of the degree of

Master of Science

Department of Medical Microbiology and Infectious Diseases
University of Manitoba
Winnipeg, Manitoba

ABSTRACT

Multiple sclerosis (MS) is a chronic, progressive, and debilitating disease of the central nervous system. This disease is thought to have multiple predisposing factors, including genetic risk factors, vitamin D deficiency, and infectious agents. Although a fungal presence in MS has long been suspected, only recently did studies associate the fungal gut microbiome—the mycobiome—with MS in adult individuals. However, investigations of the mycobiome's association with disease have been hampered by a lack of standardized methodologies. Both the methods for sequencing and bioinformatic characterization of the mycobiome require systematic assessment in order to standardize such methodologies. Furthermore, although alterations in the mycobiome have been associated with MS in adults, no studies to date have examined if the same case can be made for pediatric-onset MS (POMS) individuals. We set out to evaluate the performance of different internal transcribed spacer (ITS) amplicon sequencing and bioinformatics approaches by varying the PhiX concentration and assessing the abilities of different analytical pipelines to characterize a predefined fungal community. We also used the most optimal methods resulting from our evaluation to examine the gut mycobiome in persons with and without POMS. Our methodology evaluation finds that a 50% PhiX spike-in increased the sequence quality, with 15% more bases at a Phred score of 30 or greater (≥ 99.9% accuracy) but decreased the overall number of reads by more than half when compared to a 25% PhiX spike-in. Of the three fungal metabarcoding bioinformatics pipelines we examined, LotuS classified the highest number of expected species. Although we did not find a significant difference in the fungal community between POMS and unaffected control individuals, we found that the genus *Agaricus* was highly abundant in POMS individuals; we also found an association of the species *Candida albicans* with POMS.

ACKNOWLEDGEMENTS

This achievement is made possible thanks to my supervisors, Drs. Natalie Knox and Gary Van Domselaar, are the two pillars instrumental in providing me with the opportunity and environment to develop my computational skill set, critical thinking, language, and overall character traits. The guidance, encouragement, and nurture you have both provided me over the years of my study will undoubtedly be key to my future endeavours.

I would also like to thank my committee members, Drs. Aleeza Gerstein and Paul McLaren, for providing support throughout my project with an extended shout-out to Dr. Gerstein. My graduate study experience may not have materialized had I not met you during the last year of my undergraduate studies. It is thanks to your career counseling that I took the initiative to approach my supervisors, even though I initially had zero experience with the computer science aspect of this project.

To our project collaborators, Drs. Helen Tremlett, Charles N. Bernstein, Jessica Forbes, Ali Mirza, as well as Feng Zhu, thank you for your insights. It was a pleasure to work as a team. The knowledge you provided on multiple sclerosis and bacterial gut microbiome, our cohort, and data and statistical methodology was invaluable throughout this project. Additionally, I am grateful for the supportive environment fostered by the members of the genomics and bioinformatics sections here at the National Microbiology Laboratory in Winnipeg. Special thanks to Dr. Chih-Yu Chen (Julie Chen) for being understanding of my family situation during my moment of crisis.

And to the Visual and Automated Disease Analytics graduate training program and Mindel and Tom Olenick family, thank you for funding my graduate studies. Last but not least, I would like to thank my family for the monumental reinforcement they have provided.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANOVA	Analysis of variance
ASVs	Amplicon sequence variants
BBB	Blood brain barrier
BLAST+	Basic Local Alignment Search Tool+
bp	Base pair
CIS	Clinically isolated syndrome
CLR	Central log-ratio
CNS	Central nervous system
CSF	Cerebrospinal fluid
DMF	Dimethyl fumarate
DNA	Deoxyribose nucleic acid
EAE	Experimental autoimmune encephalomyelitis
EBV	Epstein-Barr virus
EDSS	Expanded disability status scale
ELISA	Enzyme-linked immunosorbent assay
FLASH	Fast Length Adjustment of SHort Reads
FS	Functional systems
gDNA	genomic Deoxyribose nucleic acid
HCAR2	Hydroxyl carboxylic acid receptor 2
HLA	Human leukocyte antigen
IHMS	International Human Microbiota Standards
IL-2	Interleukin-2
IL-6	Interleukin-6

IL-17	Interleukin-17
INSDC	International Nucleotide Sequence Database Collaboration
ITS1	Internal transcribed spacer 1
ITS2	Internal transcribed spacer 2
LEfSe	Linear discriminant analysis effect size
LotuS	Less Operational Taxonomic Unit Scripts
MMF	Monomethyl fumarate
MRI	Magnetic resonance imaging
MS	Multiple sclerosis
monoADS	Monophasic acquired demyelinating syndrome
N	Normality
NCBI	National Center for Biotechnology Information
NGS	Next-generation sequencing
NMDS	Non-metric multidimensional scaling
Nrf2	Nuclear factor erythroid-derived 2-related factor
OTU	Operational taxonomic unit
PERMANOVA	Permutational multivariate analysis of variance
PCA	principal component analysis
PCR	Polymerase chain reaction
PhiX	PhiX174 bacteriophage
pM	Picomolar
POMS	Pediatric-onset multiple sclerosis
PPMS	Primary-progressive multiple sclerosis

RefSeq	NCBI Reference Sequence Database
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
RRMS	Relapsing-remitting multiple sclerosis
SPMS	Secondary-progressive multiple sclerosis
TNF	Tumour necrosis factor
µl	Microlitre(s)
UCHIME	Unoise CHIMEras
USEARCH	Unified Search
µM	Micromolar
qPCR	Quantitative polymerase chain reaction
Q-Q	Quantile-Quantile

AUTHOR CONTRIBUTIONS

The thesis writing, data curation, bioinformatics, and statistical analyses, visualization of figures and tables were performed by Nelson Y. H. Mok with support and data validation from Natalie Knox, Gary Van Domselaar, Feng Zhu, and Helen Tremlett. Natalie Knox, Gary Van Domselaar, Feng Zhu, and Helen Tremlett, Aleeza Gerstein, and Paul McLaren provided input on statistical analyses. Jessica Forbes, Gary Van Domselaar, Morag Graham, Christine Bonner, Julia Hart, Emmanuelle Waubant, and Helen Tremlett wrote the grant which provided funding for this work. Julia Hart facilitated the training and study of coordinators in the collection of stool samples. Charles N. Bernstein oversaw the biobanking. Morag Graham, Natalie Knox, and Christine Bonner oversaw the wet lab methodology, including sequencing. Helen Tremlett contributed to the writing of the sample collection and cohort inclusion criteria. The fungal mock community sample was provided by Matthew Bakker (U.S. Department of Agriculture). Christine Bonner contributed to writing the extraction and sequencing methodology in chapter 2.3.

CHAPTER 1: INTRODUCTION

1.0 Multiple sclerosis

1.0.1 Characterization of multiple sclerosis

The brain and spinal cord form the central nervous system (CNS), which is responsible for cognition, muscle contraction, vision, and other essential life functions. The CNS comprises neuronal cells that conduct electrical signals that are fundamental in cell-cell signalling. Neuronal axons are lined in a myelin sheath composed of fat and protein that insulates the neuronal cell, guiding the electron flow to ensure the signal from one end arrives at the other. Certain brain disorders can cause the deterioration of myelin, leading to aberrant electrical conductance and improper cell-cell signalling.

Multiple sclerosis (MS) is one such neurodegenerative disorder where chronic inflammation and myelin degradation occurs.[1] This demyelinating effect is marked by ovular-shaped plaques in the grey and white matter of the CNS, which can be visualized using magnetic resonance imaging (MRI) and causes effects ranging from muscle weakness involving bowel and bladder dysfunction, visual and cerebral impairments in severe cases, and in extreme cases, cognitive impairments leading to physical disabilities.[1,2] Disease severity of MS can be semi-quantified by a physician using an evaluated score from zero to ten on the Expanded Disability Status Scale (EDSS), where higher values are linked to MS-induced mortality (**Figure 1.0**).[3]

0	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
Normal, no FS[a]	No disability, signs of 1 FS	Minimal disability, 1 FS	Moderate disability (1 FS) or mild disability in 3–4 FS	Severe disability but self-sufficient	Severe disability impairing daily activities (able to walk ~200 m)	Requires walking aid & able to walk ~100 m	Unable to walk even with aid, wheelchair-bound	Restricted to bed/chair or requires wheelchair pushed	Restricted to bed but is still able to communicate and eat	Death by MS

Figure 1.0. Summarized version of the Expanded Disability Status Scale.

[a] FS: functional systems, including pyramidal functions, cerebellar, brain stem, sensory, bowel and bladder, visual, cerebral, and other.[3]

1.0.2 Types of multiple sclerosis

Although the severity of MS can be semi-quantitatively determined by the EDSS, clinical courses leading to disability are variable due to spontaneous or continuous periods of demyelination and remyelination.[2] The definitions proposed by Lublin et al. (1996) are commonly accepted in modern diagnoses. The most common include Relapsing-remitting (RRMS), which occurs when acute demyelinating periods are separated by no disease progression during remyelination; Secondary-progressive (SPMS), which stems from RRMS, where individuals eventually lose the ability to reform myelin by oligodendrocytes, thereby allowing continuous neurodegeneration; and Primary-progressive (PPMS), which is marked by constant demyelination from disease onset due to a lack of or very minimal periods of remyelination.[2,4] Though other definitions exist for non-standard patterns of demyelination, whether these can be characterized as MS or another neurodegenerative disorder is not clear; accordingly, Lublin et al. (2014) added clinically isolated syndrome (CIS) for uncertain diagnoses pertaining to inflammatory demyelination.[5] Additionally, the diagnosis of MS is commonly performed using the McDonald Criteria, which depends on the dissemination of plaques over time and location. These criteria are often enhanced by MRI imaging and gel electrophoresis of oligoclonal bands from cerebrospinal fluid (CSF).[6,7]

1.0.3 Adult-onset multiple sclerosis versus pediatric-onset multiple sclerosis

Multiple sclerosis affects females at least twice as much as males.[8] On average, the global age of MS diagnosis is around 37 years, but symptom onset occurs at around 20–40 years of age.[2,8,9] The most common type of MS diagnosis is RRMS in adult-onset individuals; this diagnosis also extends to pediatric-onset cases, where MS onset typically occurs before the ages

of 16–18.[9,10] Although cases of pediatric-onset MS (POMS) are rare (~5% of total MS cases according to one study involving a Swedish cohort[11]), the reporting of POMS is on the rise, but studies focused on only POMS is rather limited, often due to the low number of individuals available to participate in a study cohort.[8,10,11]

In studies examining the differences between adult-onset and POMS, differences in the disease have been noted.[10] Individuals diagnosed with POMS experience more aggressive inflammation, leading to axonal damage of the myelin and higher relapse rates. The high relapse rate in POMS individuals is offset by quicker recovery, often attributed to remyelination by oligodendrocytes, high neuronal plasticity, or both. Significant cognitive impairment is associated with POMS if unmanaged by early intervention.[2,10] Compared to adult-onset MS, POMS individuals also take longer to reach EDSS disability milestones, but typically these milestones are met at a younger age.[11]

1.0.4 Epidemiology of multiple sclerosis

There are approximately 2.8 million cases of MS globally with a reported increase of approximately 500,000 global cases between 2013 and 2020.[8] Of these cases, 30,000 are POMS, though pediatric-onset cases is are reported by only 47 out of the 115 countries surveyed. POMS is difficult to diagnose due to the similarities in clinical presentation and symptoms compared to monophasic acute demyelinating syndromes (monoADS).[8,12,13]

The most common type of MS, RRMS, affects 85% of adult cases and is diagnosed almost exclusively in pediatric individuals.[8,12] Although twice as many females are diagnosed with MS compared to males, this difference in prevalence is due to an unknown factor potentially influenced by genetics, hormones, or lifestyle.[2,8] Additionally, the incidence of MS is

increased in high income continents, such as North America, Europe, and Australia. The high prevalence in these continents is partially attributed to increased surveillance compared to middle and low income continents.[8] Studies have shown that individuals of European descent are disproportionately affected by MS due to genetic variants associated to antigen presenting cells, such as the human leukocyte antigen (HLA)-DRB1 and CD58 gene, which occurs more frequently in this population.[2,14]

1.0.5 Pathology of multiple sclerosis

The pathology of inflammation during MS involves the blood-brain barrier (BBB), which is composed of tight junctions that selectively allow for the entry of molecules from blood capillaries into the brain.[15,16] Demyelinating plaques are often located near capillaries, suggesting that the breakdown of the BBB is a feature of MS pathology. Pro-inflammatory cytokines, such as tumour necrosis factor (TNF) and interleukin-6 (IL-6), increase leakage of the BBB, thereby allowing the entry of pathological molecules that further exacerbate the immune response, leading to the destruction of oligodendrocytes and self-inflicted demyelination by T-cells, possibly due to molecular mimicry during pathogen infection that influences the progression of MS severity. However, the actual etiology of MS leading to chronic inflammation is still under active investigation.

1.0.6 Risk factors

Certain risk factors, including sex, environment, genetics, and pathological elements such as infection by Epstein-Barr virus (EBV), are suspected to individually or jointly play a role in MS progression. For example, MS is more commonly diagnosed in females than males, especially in individuals of European descent, which can be up to three times more prevalent in some regions.[2] In a study of 199 pregnant women diagnosed with MS, Mirzaei et al. (2011)

found a lower risk of hereditary MS if the mother had a daily intake of ≥ 2 and ≥ 400 IU vitamin D before giving birth.[17] In terms of genetics, an early concordance study by Ebers et al. (1986) revealed that monozygotic twins are more likely to be diagnosed with MS, with a higher ratio in female twins than males (1.5:1), as opposed to dizygotic non-identical twins.[18] Modern studies on genetic associations note that the HLA, namely HLA-DRB1 subtypes commonly found in individuals of European descent, may lead to a higher risk of MS.[19]

A study by Endriz et al. (2017) involving 259 individuals suggests that the EBV may play a role in MS. They demonstrated a correlation of the presence of this virus (using the Epstein-Barr nuclear antigen 1 [EBNA1]) with an increased risk of MS.[20] The study also cites estrogen as a risk factor, as other studies have found that females going through puberty are more likely to develop MS.

The most compelling evidence for EBV as a causal factor of MS involves a longitudinal study of US military service personnel conducted by Bjornevik et al. (2022).[21] This study analyzed blood serum samples from 801 MS cases compared to 1566 matched controls without MS over 20 years. Using enzyme-linked immunosorbent assay (ELISA) and Western blotting, the authors found that the vast majority (800 of 801 individuals) who developed MS during their service were seropositive for EBV, whereas the same could not be said for controls. At the initiation of the study, only 33 MS cases were EBV-negative compared to 90 controls. Although 32 of these MS cases were EBV-positive at the end of the study, only 51 of the controls seroconverted to EBV positivity. Over half of the EBV seropositive controls did not develop MS, and only one of 33 MS cases did not seroconvert to EBV positive at the end of the sampling period. Although the initial cohort enrollment was high, the authors limited their investigation to 33 MS individuals who were initially EBV seronegative before MS onset, which limits the

confidence of the interpretations of the study. The study suggests that EBV may explain MS onset, but the authors also acknowledge alternative factors like the HLA-DR15 antibody may act in synergism. The HLA-DR15 molecule is also suspected to bind fungal-associated peptides, suggesting perhaps other microorganisms such as bacteria or fungi may play a role in MS.[22]

1.1 Human-fungal connections to MS

The hypothesis that fungi may play a role in MS has gained increasing traction over the past decade. This section reviews the current evidence for a fungal association with MS.

1.1.1 Fungal infection in patients with MS

Ramos et al. (2008) investigated seven individuals diagnosed with MS for the presence of specific fungi.[23] Blood serum samples from these individuals were tested by anti-*Candida sp.* antibodies using immunofluorescence and slot-blot assays, β-1,3 glucan biochemical test, and quantitative PCR (qPCR) targeting the fungal internal transcribed spacer 1 (ITS1) region. Antibodies against several fungal species were tested, including *C. albicans, C. famata, C. parapsilosis, C. glabrata, C. krusei, Saccharomyces cerevisiae,* and *Rhodotorula mucilaginosa*, all of which are human pathogens. Results of an immunofluorescence antibody assay against the five *Candida spp.* detected varying levels of these species in six out of seven individuals, whereas all ten non-MS control individuals tested negative. Similarly, all case individuals tested positive for fungal antibodies when using the more sensitive slot-blot assay but at varying levels compared to the immunofluorescence assay, though control individuals remained negative or had a negligible concentration of fungal antigens. Further testing on blood serum samples by qPCR detected fungi in six of seven of the MS individuals but not the non-MS controls. When the authors tested for the presence of β-1,3 glycan (a component of the fungal cell wall), they

discovered a fungal signature in four of seven MS cases, but no signature was found in the control individuals. This study suggests that fungi in MS are either a consequence of immune dysregulation or possibly a direct etiological agent.

1.1.2 Longitudinal fungal examination of a patient in MS

A similar approach to Ramos et al. (2008) was used by Pisa et al. (2011).[23,24] The authors of this study conducted similar assays using antibodies against the same fungi as Ramos et al. (2008), though they instead analyzed longitudinal samples from a single individual diagnosed with MS. Immunofluorescence assay on blood serum samples detected fungal antigens but with variable response. The fungal species detected differed depending on the month of sample collection, where the anti-fungal antibodies sometimes detected a certain species and other times no fungal organism was detected. These variable results were similarly observed by slot-blot assay; although fungal antigens were detected throughout all longitudinally collected sera, some were detected at low levels, which the authors attribute to the increased sensitivity of this assay. Similarly, nested PCR amplification for the fungal ITS1 region and detection of β-1,3 glucan was also variably positive in some blood serum sample collection months but not others. An additional cross-sectional analysis was performed on a CSF sample of this individual. CSF is a naturally sterile fluid,[25] but when the authors used the immunofluorescence and slot-blot assays, they discovered that the MS individual had antigens and antibodies of fungal origin while none were observed in the control.

1.1.3 Fungi in cerebrospinal fluid of MS patients

Pisa et al. (2013) followed up on their original investigation into the presence of fungi in the CSF of MS individuals.[26] The main focus of this study was to test for the presence of *C. famata* and *C. parapsilosis* in the CSF samples of 12 individuals. Although fungi were detected

in all 12 cases, two cases were detected above the study's detection threshold. The authors also tested the blood serum of these 12 individuals using the slot-blot assay. The results showed positivity for fungi in ten of 12 individuals for *C. famata*, and five of 12 individuals positive for *C. parapsilosis*. The results of this study suggest no correlation between the fungal community in CSF and serum samples. When testing for the presence of fungal antigen by immunofluorescence-tagged antibodies, *C. famata* was identified in one individual, and *C. parapsilosis* was identified in four individuals. Blood serum tests, in contrast, detected both *C. famata* and *C. parapsilosis* in nine MS individuals, which again indicates a difference in fungal composition depending on the sampling location and type. The lower detection rate in the CSF samples is attributed to the presence of the BBB and selective permeation of molecules into the CNS. Though the authors observed an 88% detection rate in their PCR test using CSF from eight individuals, this higher detection is possibly due to increased sensitivity by genetic testing over serological testing.

1.1.4 Metagenomic analysis of MS cerebrospinal fluid

Cerebrospinal fluid from ten individuals with MS was studied by Perlejewski et al. (2016) using next-generation sequencing (NGS).[22] This study used single-primer isothermal amplification followed by RNA sequencing to conduct metagenomic analysis. They identified organisms of bacterial, parasitic, protozoan, and fungal origin. Fungal genera identified include *Ascomycota, Brachyakara, Candida, Cladosporium, Debaryomyces, Exophiala, Falciformispora, Funneliformis, Galactomyces, Gigaspora, Glomus, Knufia, Malassezia, Penicillium, Peniophora, Pleosporales, Rhodotorula, Saccharomycetales, Sclerotium*, and *Sclerotium*. Although the presence of fungi was common in some cases, the identified fungal populations differed among individuals with MS. The genera *Candida, Cladosporium,*

Funneliformis, Galactomyces, and *Glomus* were found in more than half of the individuals, with *Malassezia* being found in four individuals. However, the authors caution that these commonly identified organisms may be contaminants of the environment or the utilized PCR reagents.

1.1.5 Detection of fungi in MS nervous tissue

Alonso et al. (2018) investigated individuals for the presence of fungi and bacteria in the brain tissue of ten individuals with MS.[27] They used nested PCR assays to amplify specific fungal signatures, the internal transcribed spacer regions 1 and 2 (ITS1 and ITS2). The ITS1 region amplified in approximately nine cases, and the ITS2 region in all ten cases. In non-MS control individuals, ITS1 was identified in eight of nine samples, and ITS2 was identified in all nine samples. Next-generation sequencing of the ITS1 and ITS2 regions identified *Trichosporon mucoides* in eight of ten MS cases, while non-MS controls identified no fungal species. Metagenomic analysis identified *Candida deformans, Davidiella tassiana, uncultured Malassezia sp.*, and unknown fungi in high abundance. Fungi were also detected in non-MS controls. Differential analysis of cases and controls revealed a relative abundance of the genera *Candida, Malassezia,* and *Trichosporon* in MS cases, and the orders Eurotiales, Lecanorales, Pleosporales, and Tremellales were exclusive to MS individuals. When conducting principal coordinate analysis using the Bray-Curtis distance, the authors found no distinct differences between MS cases and non-MS control samples.

1.1.6 First characterization of the gut microbiome in MS

Shah et al. (2021) published the first paper characterizing the gut fungal community in 25 people with MS and 22 non-MS unaffected controls that were matched for age, gender, BMI, and ethnicity.[28] The authors found that the phylum *Ascomycota* was highly abundant in both MS and controls, whereas *Basidiomycota* was lower in MS individuals compared to controls. The genus

Saccharomyces had the highest relative abundance in both MS and controls. The genera *Xylaria, Pencillium, Agaricus*, and *Aspergillus* were identified in both groups but at lower abundances; furthermore, *Saccharomyces* and *Aspergillus* were significantly higher in MS individuals than controls. The authors found the gut fungal community significantly differed in MS individuals compared to unaffected controls. Beta diversity at the genus level significantly differed between MS cases and controls ($p = 0.04$). Alpha diversity by Shannon index found the same result ($p = 0.043$). Longitudinal sampling over six months did not show significant changes in the gut fungal community, although the genus *Saccharomyces* steadily declined over this period. This suggests that the gut fungal populations form stable communities in individuals with MS.

1.1.7 Second characterization of the gut fungal microbiome in MS

Yadav et al. (2022) investigated the composition of the gut mycobiota in 20 RRMS individuals and 33 unaffected controls aged 18–63.[29] Their beta diversity results found a significant difference in diversity when comparing MS cases to controls ($p = 0.011$), in agreement with the findings of Shah et al. (2021). Although their alpha diversity analysis did not indicate significant differences, they noted a trend of higher richness in MS individuals than in controls. Similar to the findings of Shah et al. (2021), the authors of this study also noted phylum *Ascomycota* and *Basidiomycota* at the highest relative abundances in both groups; however, *Basidiomycota* was found to be higher in MS individuals, whereas this phylum was found to be at low abundance in MS individuals in the Shah et al. (2021) study. Across both studies, the genus *Saccharomyces* was found at the highest abundance in both MS individuals and unaffected controls. The genera *Candida, Malassezia, Penicillium*, and *Cladosporium* were also identified in MS cases and controls, though at lower abundances. Interestingly, a majority of these overlapping genera found between both studies were also discovered in a controlled diet

experiment conducted by Auchtung et al. (2018).[30] The authors of the controlled diet study indicated that *Saccharomyces* and *Candida*, in specific *C.albicans*, are commonly found in foods such as bread and in naturally occurring in saliva, which is a cause of dental cavities. Overall, Yadav et al. (2022) concluded there is a difference between the gut mycobiota of MS individuals and unaffected controls.

1.2 Other fungal connections to MS

1.2.1 Experimental autoimmune encephalomyelitis animal model

Experimental autoimmune encephalomyelitis (EAE) is an inflammatory demyelinating disease of the central nervous system that approximates the clinical features of MS, and is the most commonly used animal model for studying MS.[31] Takata et al. (2015) examined the effect of yeasts from fermented foods on the composition of the intestinal microbiota and their ability to modulate the course of EAE.[32] A total of 18 fungal organisms were cultured, heat-killed, and fed to C57BL/6 inbred mice. EAE was later induced using MOG35-55 oligodendrocyte glycoprotein and Freund's adjuvant. Of the 18 administered fungi, only *Candida kefyr* significantly decreased EAE symptoms, and cytometry results indicate low mononuclear cells in the mice CNS with cytokine assays showing a significant decrease of the proinflammatory interleukin-17 (IL-17). Mice fed with *C. kefyr* were euthanized for their cecal contents, including the microbiota, which was then fed to mice on antibiotics and later induced for EAE. The results demonstrate a decrease in EAE symptoms attributed to a decrease in the bacterial genus *Bacteriodes* and increases in the genus *Prevotella* and order *Lactobacillales*, compared to control EAE-induced mice. This study suggests that *C. kefyr*, through its influence on the bacterial gut microbial population, may ameliorate the effects of EAE if ingested before disease induction.

In contrast, a study performed by Fraga-Silva et al. (2015) injected the fungus *C. albicans* into the tail vein of C57BL/6 mice, exploiting the ability of *C. albicans* to disseminate into the CNS through invasive candidiasis.[33] They later induced EAE using the MOG35-55 peptide, Freund's adjuvant, and *Bordetella pertussis* toxin to immunologically incite limb weakness or paralysis. The presence of *C. albicans* in the CNS of mouse brain biopsy tissue was detected by tissue staining and isolating colonies on Sabouraud-dextrose agar. EAE mice infected with *C. albicans* developed the severe disease by aggravating immune infiltration into the CNS, namely by strong $CD3^{+}CD4^{+}$ T-cells attraction by proinflammatory cytokines such as TNF-α, IL-6, and IL-17, showing that a species of *Candida* can exacerbate EAE inflammation and suggesting a link between fungi and MS, with fungal toxins possibly playing a contributory role.

1.2.2 Mycotoxin: Gliotoxin animal studies

Toxin-induced models for studying MS have explored the effects of gliotoxins, such as ethidium bromide, lysolecithin, and 6-aminonicotinamide, which are known to cause demyelination.[31] Willis et al. (2004) showed that astrocytes, critical for maintaining tight junctions of the BBB, can be destroyed by the gliotoxin 3-chloropropanediol in F344 rats and thereby increasing leakage into the CNS.[34] Fraga-Silva et al. (2019) expanded on this by examining gliotoxins produced by the fungus *Gliocladium fimbriatum* in an EAE-mouse model. By injecting gliotoxin into C57BL/6 mice induced with EAE, the authors found a significant increase of IL-17 and interleukin-2 (IL-2) cytokines, which caused inflammation and demyelination in the CNS, thereby increasing EAE severity.[35] The authors report that the BBB permeability was increased in both control and EAE-induced mice once gliotoxin was administered, which allowed the uptake of sodium fluorescein into the CNS. This study shows that mycotoxins may play a role in MS, and various species of fungi are producers of gliotoxins

and other mycotoxins, which allows them to increase invasiveness and decrease the fungal immune response.[35,36]

1.2.3 Dimethyl fumarate and fungi

A less explored, but intriguing potential link between fungi and MS is the drug dimethyl fumarate (DMF) used for MS symptom management. Although the exact mechanism of action of DMF in MS is not clear, it is used to regulate immune response by promoting anti-inflammation and neuroprotection, which can be extended for use in other immune-mediated diseases such as psoriasis.[37,38] It is hypothesized that DMF, when hydrolyzed into monomethyl fumarate (MMF), can directly activate the nuclear factor erythroid-derived 2-related factor (Nrf2) transcription factor pathway, which ultimately leads to an increase of antioxidants such as carbon monoxide and anti-inflammatory cells while decreasing the inflammatory cell population.[38] DMF may also indirectly activate the hydroxyl carboxylic acid receptor 2 (HCAR2), which is shown to be important to anti-inflammatory responses in the EAE mice model of MS. Prior to DMF being approved as a treatment for MS by the US Food and Drug Administration in 2013, it was used as a biocide to prevent fungal growth on leather products.[39] Ma et al. (2017) conducted a study to examine the effects of DMF on mycotoxins that disrupt tight junctions of the gut-blood barrier, leading to dysbiosis associated with immune-mediated diseases.[40] Feeding BALB/c mice with DMF increased the gut bacterial richness and diversity and decreased tight junction permeability. Specific mycotoxins, including aflatoxin, fumonisin, and zearalenone, were found to be decreased in mice administered with DMF. Although the study outcome on mycotoxins *in vivo* does not clearly indicate a DMF target, the authors advance the idea that this drug may decrease toxins released by fungi.

1.3 Fungal genomics and metabarcoding

1.3.1 Next-generation sequencing

The impact of genomic sequencing technologies on microbial disease investigation has been profound.[41] Sequence-based fungal identification was originally performed by Sanger sequencing, which is limited in its ability to analyze multiple samples in parallel.[42,43] Next-generation sequencing (NGS) technologies, with their increased accuracy, decreased cost, and massive throughput, have resulted in a dramatic proliferation of genomic investigations into microbial disease. Such studies may involve the identification of a specific organism or a community of organisms.[42] These analyses have contributed greatly to identifying pathological agents contributing to disease.

The Illumina platform is a short-read sequencing platform and a popular choice for mycobiome investigations due to its relatively high throughput and low cost relative to long-read options such as PacBio or Oxford Nanopore.[42] There are two main approaches for studying the mycobiome using the Illumina platform: the targeted-amplicon metataxonomic barcoding approach and the shotgun metagenomic approach.[42,44,45] The targeted-amplicon approach amplifies a single short molecular signature unique and common to fungi but with sufficient variability to distinguish between fungal taxa. Shotgun metagenomics, in contrast, randomly samples all the genomic content from a population, including fungal and non-fungal constituents. Although, in theory, shotgun metagenomics offers a comprehensive survey of all the fungal genes within a sample, in practice, this method is hard to perform for the mycobiome due to the high relative abundance of non-fungal species in the gut microbiome. Only ~0.1% of the human gut is inhabited by fungi, while the rest is composed of mostly bacteria, viral and human genes.[46] This means a greater sequencing depth is required to fully capture the fungal diversity within a

sample which corresponds to a high cost of shotgun metagenomic sequencing. Additionally, although fungal databases are sparse in general, there exist less comprehensive and well-curated databases to assemble and identify fungal organisms using the shotgun metagenomic approach.[47] For these reasons, targeted-amplicon approaches, which are specific to the fungal population, are more popular. **Figure 1.1** provides an overview of the mycobiome analysis process.

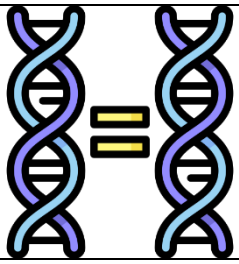
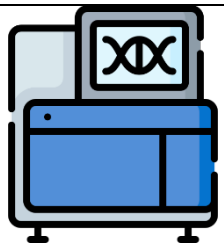
1. Sample Collection	2. DNA Extraction	3. Library Preparation	4. Sequencing	5. Bioinformatics Analysis
				

Figure 1.1. General overview of the gut fungal microbiome analysis process. 1) A sample, typically from stool, is collected. 2) DNA is extracted from the sample. 3) DNA is amplified using a targeted amplicon or shotgun metagenomic approach and ligated with sequencing primers. 4) Sequencing is performed. 5) Bioinformatics analysis is used to identify taxa and conduct downstream analyses.

1.3.2 The challenge of extracting ecological fungal DNA

The identification of organisms in environmental samples by NGS is heavily impacted by the upstream sequence extraction step.[46,48] The DNA extraction step is crucial for capturing the genomic content representing the microbial community of interest in a sample. This challenge is increased for fungi because their chitinous cell wall differs from bacteria, and techniques for extracting DNA from the gut mycobiome are less well-developed than their bacterial counterpart.[42,44–46,48,49] For example, the International Human Microbiota Standards (IHMS) project established standard protocols for extracting and sequencing bacterial DNA from stool, but similar standards for fungal DNA extraction are lacking.[46,50] Although the structure of the fungal cell wall is similar between different fungal phyla, their chemical compositions differ considerably, which poses an additional challenge to DNA extraction.[51] For example, within the *Ascomycota* phylum, the genus *Fusarium* and *Saccharomyces* differ in the ratio of components forming their cell walls, where *Saccharomyces* contains a 1:60:13:8 percent dry weight ratio of chitin:glucans:protein:lipid, whereas *Fusarium* has a 39:29:7:8 ratio. This difference in cell wall composition requires careful consideration of cell lysis techniques since complete lysis methods do not neutralize PCR inhibitors. Techniques that do not address the presence of phenol derivatives in some fungal walls and potential proteases or metal ions can decrease amplification efficiency and may lead to decreased detection during taxonomic assignment.[52,53] In contrast, incomplete lysis can lead to a loss of captured diversity. For example, the cell wall of the genus *Allomyces* contains more chitin as compared to *Saccharomyces* (58% vs. 1% total cell wall dry weight). Such a large range of chitin composition biases nucleic acid extraction towards the fungal population with cell wall structures that are more easily lysed.[51]

To examine the effects of different lysis techniques on the extraction of fungal DNA, Frau et al. (2019) investigated chemical lysis combined with or without mechanical shearing by bead beating on seven positive control fungi and a stool sample.[49] The authors noted that using bead beating before the extraction of stool DNA significantly increased the DNA yield of two of the positive control organisms, *Candida tropicalis,* and *Malassezia furfur*, although reads from *Candida neoformans* decreased compared to not using beads.[49] The authors tested the same extraction techniques on a single human stool sample and found a slight increase in fungal 18S rDNA copies once bead beating was used. This study concluded that mechanical shearing by bead beating is beneficial for mycobioime DNA extraction.

1.3.3 Metabarcoding signatures of the fungal rDNA operon

As previously mentioned, the Illumina platform is often preferred for analyzing fungal communities because of its low cost and high base-calling accuracy.[42] The Illumina protocol involves library preparation and sequencing steps, whereby the amplification of specific fungal barcode signature(s), called metabarcoding, is preferred over whole genome metagenomics for a few reasons: fungi are eukaryotes, and their DNA contains many non-coding introns; the cost of metagenomics is usually higher than metabarcoding; and the databases currently available for assigning taxa to fungal sequences are considerably less well developed than their bacterial counterparts.[42,44,46] Like bacteria, the fungal rDNA operon is taxonomically well conserved. The operon contains the 18S, internal transcribed spacer region 1 (ITS1), 5.8S, ITS2, and 28S subunit regions, which are useful metabarcoding targets for investigating fungal communities (**Figure 1.2**). Although the Illumina platform is commonly used for mycobiome analyses, this technology is limited by the length of nucleotides it can capture in a read, currently 300 bases using the single-end Illumina Miseq approach, expandable to 600 bases by paired-end sequencing. Many

fungal ITS1 and ITS2 regions fall within this sequencing range, making them suitable targets for metabarcoding.

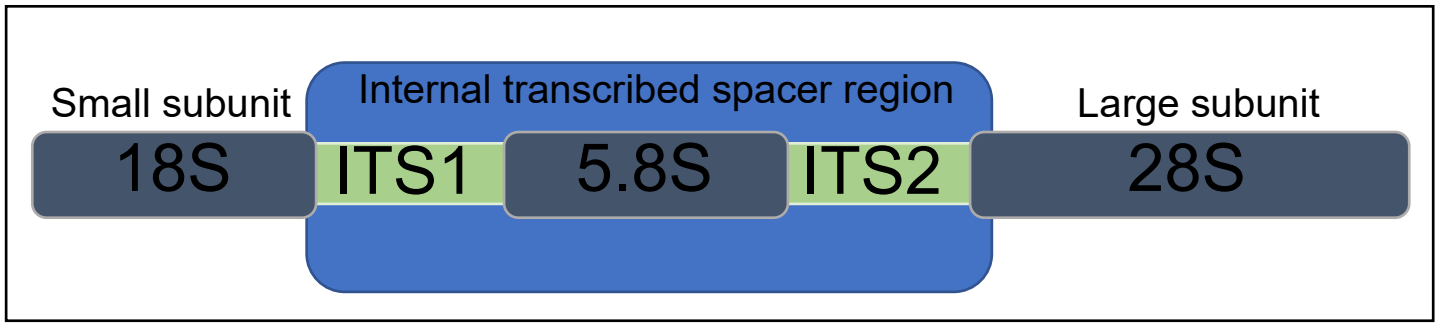


Figure 1.2. The fungal ribosomal DNA. The 18S makes up the small ribosomal subunit, while 5.8S and 28S form the large subunit. Two internal transcribed spacers exist (ITS1 and ITS2); they are non-coding and excised during rRNA processing.[54] The internal transcribed spacer region encompasses the ITS1, 5.8S, and ITS2. The outer regions, 18S and 28S, are each ~1800 bp or greater, while the complete internal transcribed spacer is ~500-700 bp in length.[55]

1.3.3.1 ITS region as a fungal barcode

Due to the length limitation of short-read sequencers, only certain subregions of the rDNA can be captured by the existing sets of degenerate fungal primers, and each primer set possesses a bias toward amplifying specific fungal taxa.[42,44,45,49] As well, rDNA subregions vary in their ability to resolve closely related taxa due to the highly conserved nucleotides, limiting its ability to distinguish between certain taxa. Schoch et al. (2012) investigated multiple fungal DNA candidates as potential barcodes for identifying and differentiating between fungal taxa.[55] The authors found that rDNA genes amplified with a higher success rate than other potential markers, such as RPB1. Namely, the ITS subregion composed of IT1, 5.8S, and ITS2 outperformed all other subregions when used to resolve 226 fungal species in 742 samples. The study also showed an inherent taxonomic bias within each subregion. Specific fungal phyla were differentially captured by different regions; for example, the ITS subregion identified more *Basidiomycota* and fewer *Ascomycota* species, while the reverse was observed for the protein-coding RPB1 gene.

However, Schoch et al. (2012) used a Sanger sequencing approach to characterize the ITS subregion.[55] The complete ITS subregion ranges from approximately 500–700 bp. Both the individual ITS1 and ITS2 spacers of the ITS subregion, range from about 250–400 bp.[42,56] The ITS spacer regions are separated by a ~160 bp 5.8S rDNA region. and outside of the ITS regions flanking the ITS subregion contains the 18S and 28S rDNA, which can be greater than 1800 bp.[57] This is why short-read NGS technologies are best suited to target the fungal ITS1 or ITS2 subregions since their sequence length range can be captured by short-read sequencing technologies.

1.3.3.2 Examining fungal ribosomal barcodes

In addition to testing for DNA extraction methodologies, Frau et al. (2019) also used the Illumina Hiseq platform to sequence individual fungal subregions, including the fungal 18S, ITS1, ITS2, and 28S.[49] The authors pooled the DNA of seven positive control fungi at different compositions and used the UNITE fungal database consisting of full-length fungal rDNA sequences to identify fungal taxa from sequence reads.[49,58] Results from this study indicate that the 18S subregion outperformed the ITS1 and ITS2 amplicons, in that order, for the detection of the seven fungi, although the ITS2 amplicons identified fewer spurious operational taxonomic units (OTUs) and chimeric sequences compared to the ITS1 and 18S subregions. The authors conclude that the fungal 18S subregion more accurately detects fungal taxa but at a high cost of false positive identifications relative to the ITS subregions.

1.3.3.3 ITS1 versus ITS2

Although the study of Frau et al. (2019) found that the 18S marker outperformed the ITS1 and ITS2 markers, in general, the highly conserved 18S subregion typically provides lower resolution, as fungi from different taxonomic orders often cluster into polytomies.[59] Therefore, the ITS1 and ITS2 subregions are used as gold standards to characterize fungi since they contain nucleotides with sufficient diversity to resolve species.[42,55–57] The relative superiority of ITS1 versus ITS2 regions for assigning fungal taxa remains the subject of debate.

The ITS1 subregion is more diverse and evolves faster than ITS2; the ITS2 subregion, in contrast, is less variable in length between primer annealing regions, which has been shown to increase taxonomic resolution compared to ITS1.[42,59] An examination of the variability between the fungal ITS region sequences from the International Nucleotide Sequence Database Collaboration (INSDC), which includes the Genbank, European Molecular Biology Laboratory

(EMBL), and DNA Data Bank of Japan (DDBJ) databases, resulted in a 0.87 correlation between the ITS1 and ITS2 subregions, possibly indicating both do perform similarly for inter- and intra-species identification.[54] In a study investigating the species identification concordance of ITS1 and ITS2, Blaalid et al. (2013) characterized 42 environmental samples using targeted-amplicon sequencing of the ITS1 and ITS2 regions followed by taxon assignment using the UNITE fungal ribosomal database.[60] The ITS1 and ITS2 subregions both identify fungi with high concordance when clustering reads at or above 97% sequence identity. Although there are slight differences in the taxa identified by the ITS1 and ITS2 subregions at the phylum level, these differences are likely influenced by amplification bias, primer choice, and the DNA extraction protocol. In another study comparing fungal marker targets, Hoggard et al. (2018) found the fungal ITS2 subregion resulted in fewer false-negative identifications.[61] The consensus view of existing literature indicates that the fungal ITS2 subregion is the best target for short-read-based identification of fungal communities.

1.4 Role of bioinformatics in fungal metabarcoding

1.4.1 High-throughput sequencing and bioinformatics

Advancements in bioinformatics are important as NGS technologies are now commonplace in many biomedical applications. While NGS technologies are used to determine the genetic composition of organisms, bioinformatics is required to store, manage, process, and interpret the vast amount of sequence data generated by such technologies.[62,63] Bioinformatics is a multidisciplinary field as it not only requires an understanding of biological processes but also computational algorithms, mathematical methods, and statistics to develop and apply bioinformatics methods and tools.

1.4.2 Importance of validating bioinformatic tools

Bioinformatics tools abound, and there often exist multiple tools to carry out a given analysis; this, in turn, affects the interpretation of the results.[63,64] Therefore, it is important to validate bioinformatics tools serving the same purpose, to ensure the data is interpreted correctly. This holds true for the analysis of fungal ITS markers, especially since the field is relatively new, and the available tools to carry out the analysis are not as mature as their bacterial counterparts.[63]

1.4.3 Fungal metabarcoding pipeline workflow

Several pipelines exist to aid in characterizing fungi within an ecological community using metabarcoding to target the ITS subregions of the fungal rDNA. These pipelines consist of multiple bioinformatics tools that work together to convert raw NGS sequence data into taxonomic identities. The steps in the pipeline typically include 1) read processing and quality control (demultiplexing reads, trimming primer and barcode sequences, removing reads containing homopolymeric stretches, and read-quality filtering); 2) merging the overlapping regions of paired-end reads to generate longer, higher quality reads and dereplication to eliminate unnecessary computation; 3) filtering chimeric sequences; 4) ITS subregion extraction; 5) read clustering to generate OTUs or amplicon sequence variants (ASVs); and 6) taxonomic assignment using a specific annotated database of fungal ITS sequences with known identities.[42,63–66]

1.4.3.1 A comparison of fungal metabarcoding pipelines

Many pipelines for metabarcoding analyses exist, but only a handful are tailored for analyzing short-read sequencing of the fungal ITS subregions; popular pipelines include PIPITS, LotuS, PipeCraft, mothur, QIIME, and Galaxy.[42,64,67–69] Anslan et al. (2018) compared the performance of five out of six of these pipelines (mothur was excluded), on samples of arthropod

substrates sequenced for the fungal ITS2 subregion using paired-end Illumina Miseq technology.[63] Some workflow components differed between pipelines; for example, the read-merging step included various joiner tools such as FLASH, DADA2, VSEARCH, or FLASHQ. However, all the evaluated pipelines used the UNITE fungal database for taxonomic assignment. Although the total abundance of reads was similar between each processing step within the pipelines, the number of different OTUs identified (the richness) differed between pipelines. These differences were apparent in the identified taxonomic compositions of the sequenced fungal community. This result indicates that differences exist in the taxonomic profiles generated by the bioinformatics pipelines used to characterize fungal communities.

1.4.4 Choice of annotated databases for fungal metabarcoding

The choice of a database used to taxonomically represent OTUs or ASVs is important since different databases differ in quality and the degree to which they are maintained.[42,44] Several databases exist to aid classifiers in assigning sequences to fungal taxa, each containing annotations specific to subregions of the fungal rDNA. For example, the UNITE database is an archive of annotated sequences spanning the entire fungal ITS subregion (IT1, 5.8S, and ITS2), while the targeted host-associated fungi database contains only fungal ITS1 sequences.[47] Annotated databases also vary in completeness, as many fungal species may not be captured within a database or have yet to be identified. Around 140,000 fungal species have been scientifically described, but it is estimated that about 1.5 million to 3.8 million species are yet to be discovered and sequenced.[51] Because not all fungal marker sequences are represented in all databases, and most databases for assigning sequences to fungal taxa are sparsely annotated, some taxa may be misclassified or absent. This is especially true for rare and undiscovered fungal species.

More comprehensively annotated databases improve overall fungal characterization; however, it is also essential to choose a well-curated database. For example, the INSDC databases contain a large number of annotated fungal taxonomic marker sequences, but these are submitted *ad hoc* from the scientific community, and although there is a basic standard of curation before approval of these sequences, curators of the INSDC databases often lack the domain expertise to perform the specialized curation required for taxonomic assignment from targeted-amplicon sequencing of fungal taxonomic markers.[70] The UNITE database is a specialized database designed specifically for the taxonomic assignment of fungal maker sequences. The UNITE database contains a collection of fungal ITS region sequences and the taxa to which they belong. Each entry is peer-reviewed by mycologists with domain expertise.[58] This makes the UNTIE database attractive for fungal taxonomic assignment of fungal communities from fungal ITS targeted-amplicon sequencing.

The work of Tang et al. (2015) illustrates the impact of database choice on taxonomic assignment.[56] In this study, the authors investigated the fungal community of mouse feces using targeted amplicon sequencing of the ITS1 region. They then compared the taxonomic compositions assigned by the Basic Local Alignment Search Tool (BLAST) classifier using three publicly available annotation databases: UNITE, Findley, and RefSeq Targeted Loci (RTL), with the addition of their custom Targeted Host-associated Fungi (THF) database. Although the THF database was specifically tailored to the experiment, and therefore noted to perform the best, the authors found significant differences in the taxonomic assignments generated from the three other databases tested. They showed that different genera were identified depending on database choice, and if a genus was similarly identified between databases, then the reported relative abundances differed. Notably, the UNITE and Findley databases performed similarly to the

custom THF database in identifying relative abundances of major genera (> 30% of all genera) such as *Candida* and *Saccharomyces*, and minor genera (< 30% of all genera), including *Alternaria*, *Fusarium*, and *Trichosporon*.

1.5 Study objectives

1.5.1 Chapter 3: Assessing pipelines for mycobiome amplicon data analysis - rationale, hypothesis, and aims

Well-established protocols exist to extract, sequence, and identify bacterial communities inhabiting the gut.[46] The same cannot be said for the fungal residents of the microbiome, as there currently exists no standardized methodology for fungal ecological analyses in humans, although some progress has been made.

For fungal sequence extraction, it is recommended to encourage the degradation of the chitinous cell wall by supplementing the standard DNA extraction protocols used for bacteria with a bead-beating step. The Illumina MiSeq platform is an accurate, high-throughput, and cost-efficient technology well-suited for fungal ecology studies. Although the Illumina MiSeq is a short-read sequencer, the majority of ITS1 and ITS2 sequences can be captured by this technology, and it provides a good resolution of the fungal taxa. The UNITE ribosomal database is a good candidate for fungal taxonomic assignment since it is well-curated and designed to assign taxa to fungal ITS1 or ITS2 marker regions.

The choice of bioinformatics pipeline is equally important to the choice of sequencing technology and taxonomic assignment database. Existing pipelines conveniently package together all of the necessary components required to assign taxa from sequence reads. However,

there is a lack of systematic evaluation of these pipelines, which is required to guide the choice of bioinformatics pipeline to adopt for a given study.

To fill this knowledge gap, we used a 19-member mock fungal community as ground truth to assess the suitability of the UNITE fungal database to identify fungal organisms. In addition, we also examined and assessed three bioinformatics pipelines (LotuS, mothur, and PIPITS) for their ability to accurately identify the members of this mock community. We hypothesize that the UNITE database is not comprehensive for fungal identification because many fungal species remain unsequenced.[28,29,71–73] We also hypothesize that the mothur pipeline will perform best in characterizing the mock community because of its demonstrated accuracy in assigning taxa for bacterial communities.

1.5.2 Chapter 4: Investigating the gut mycobiota in pediatric-onset MS - rationale, hypothesis, and aims

The etiology of the autoimmune disease MS has long eluded researchers, though studies have found associations with various factors, such as environment, genetics, microbial infection, hormones, and vitamin D levels. Emerging evidence suggests there is also a fungal link to MS. In a mouse model of MS, the inflammation of EAE is altered by injecting specific fungi and/or their secreted mycotoxins.[33,35] Fungal signatures have been detected in brain tissue from individuals with MS. The CSF, which is normally sterile, was also positive for fungal antibodies and antigens of fungal origin.[25,26] Studies on MS individuals suggest that the human gut may be implicated, including the gut mycobiome.[28,29] However, these studies are largely conducted on adult onset-MS individuals; studies on the association of the gut mycobiome in individuals with POMS are comparatively rare.

Our study, detailed in chapter four, addresses the question of whether a difference exists between the fungal gut microbiome in a cohort of POMS individuals relative to non-MS (unaffected) controls and individuals with monophasic acute demyelinating syndrome (monoADS). We hypothesize that individuals with POMS will have a different gut fungal profile relative to unaffected controls, individuals with monoADS, or both.

CHAPTER 2: MATERIALS AND METHODS

2.1 Mock fungal community

Pre-extracted genomic DNA corresponding to 19 distinct fungi was provided as a pooled mock community at varying ratios of "Staggered A" proportions defined by 18S rRNA qPCR by Dr. Matthew Bakker from the University of Manitoba, formerly with the US Department of Agriculture (**Figure 2.0**).[74] Two members of the pooled community, *Rhizomucor miehei* (5 strains) and *Rhizophagus irregularis* (2 strains), contained additional sequence variant strains for a total of 24 sequences corresponding to 19 unique fungal species.

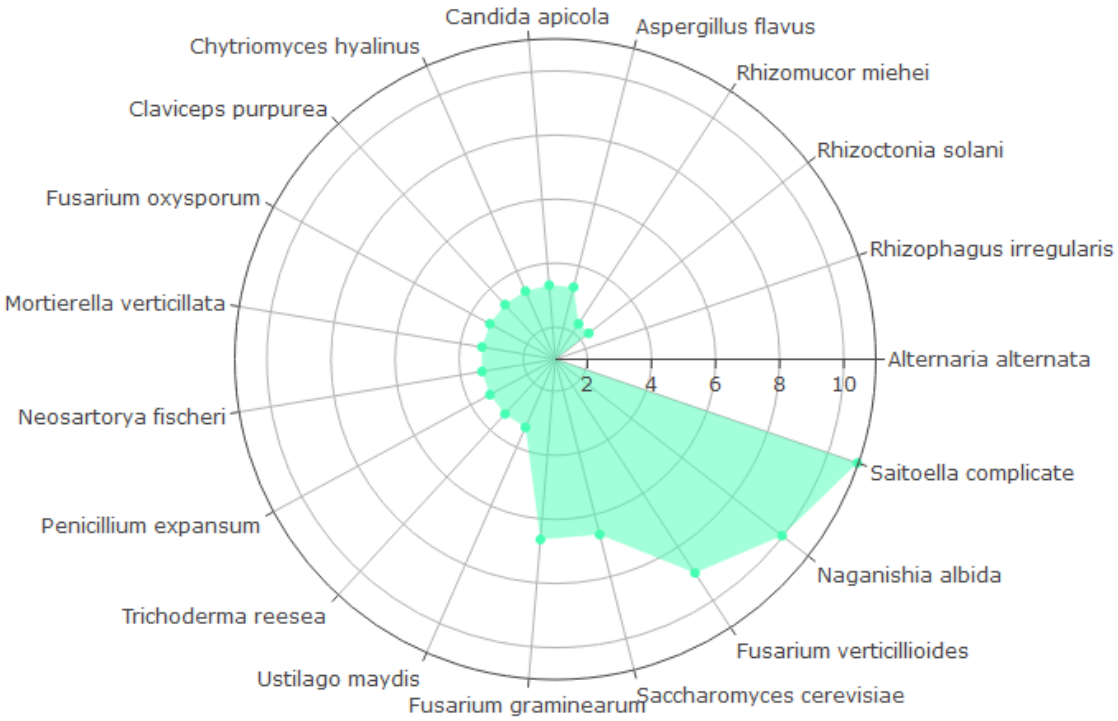


Figure 2.0. Relative ratios of a 19-organism "Staggered A" mock fungal community. The expected ratios are determined by qPCR of 18S rRNA by Bakker (2018) and transformed here on a log2-scale for visualization.[74]

2.2 Study cohort and stool samples

Our study cohort consisted of 119 North American participants, 86 residing in Canada and 33 in the United States of America. All participants were enrolled in a Canadian Paediatric Demyelinating Disease Network study. Canadian participants were excluded from this study if they were diagnosed with relapsing non-MS, which accounted for four participants. An additional Canadian participant was removed from this study due to mislabeling during sequencing. Of the American participants, 23 were excluded from this study because of a lack of cohort metadata. A total of 15 longitudinal samples from 14 Canadians and one American participant were sequenced but not analyzed due to the low sample size. Further selection for pediatric participants included applicants of ≤ 24 years of age at the time of stool sample collection (between 2015–2019). Our cohort was either diagnosed with POMS or monophasic acute demyelinating syndrome (monoADS), all of which had their symptom onset (first clinical attack) at < 18 years of age. Participants not in the POMS or monoADS group are considered unaffected controls, which were recruited through poster and web advertisements.[75] The MS cases fulfilled McDonald's diagnostic criteria. MonoADS was defined as an initial acute clinical episode of symptoms involving the CNS, with evidence of inflammatory demyelination and with no new clinical or MRI findings of recurrent demyelination. Unaffected controls had no known neurological or immune-related conditions (headache/migraine, asthma, and allergies were permissible). Study participants were also excluded from this study if any antibiotics, antivirals, probiotics, or steroids were taken within 30 days of stool collection. Study participants were recruited with the aim of matching age, sex, race, and geographical regions.

The final cohort consisted of 38 Canadian study participants residing in the provinces of British Columbia, Alberta, Manitoba, Ontario, and Quebec, as well as 8 American participants

from Pennsylvania state, for a total of 46 study participants. Of the study cohort, 29 identified as female and 17 as male. Upon stratification by diagnoses, 18 were diagnosed with multiple sclerosis (POMS; 15 females, 3 males), 13 with monoADS (6 females, 7 males), and 15 participants were classified as unaffected controls (8 females, 7 males). The mean age of symptom onset was 15.3 and 5.2 for POMS and monoADS participants, respectively.

Stool samples were shipped on ice and stored at -80 °C. The mean stool sample collection age for all 46 participants was 18.1 for POMS, 12.1 for monoADS, and 15.3 for unaffected control participants. The 46 study participants were assigned to an age group (child or youth) based on their age at the time of sample collection: child ≤ 14 years old (n = 13) and youth between 15 and 24 years old (n = 33).

2.3 Stool DNA extraction from study Participants and sequencing of both participant and pre-extracted mock community DNA

Stool samples were subjected to genomic DNA extraction using the Zymo Quick-DNA™ Fecal/Soil Microbe Miniprep Kit (Zymo Research, Irvine, CA) according to the vendor's protocol with some modifications. Two fecal punches were lysed with 750 µl of BashingBead buffer and were spun in a bead-beater for 10 minutes at 1200 rpm. Subsequent centrifugation occurred for 1 minute at 10,000 x g, and 400 µl supernatant was transferred to a Zymo-Spin III-F filter within a collection tube, then centrifuged for 1 minute at 8,000 x g. The filtrate was then mixed with 467 µl of genomic lysis buffer and 333 µl isopropanol, followed by repeated transferring of ~800 µl into a Zymo-Spin IIC column in a collection tube with centrifugation for 1 minute at 10,000 x g until the mixture was completely collected in the column. The collected material in the column was then transferred into a new collection tube with 200 µl of DNA pre-

wash buffer, followed by centrifugation for 1 minute at 10,000 x g. Column material was washed using 500 µl of gDNA wash buffer and centrifuged for 1 minute at 10,000 x g. The column was then placed into a clean 1.5 ml microcentrifuge tube, and 50 µl of 60 ºC heated DNA elution buffer was added. DNA was eluted after 3 minutes by centrifugation for 30 seconds at 10,000 x g. A secondary elution step was performed by re-adding the eluent back into the column, and the same centrifugation step was repeated after 3 minutes. The remaining extraction steps were carried out as per manufacturer protocol.

Genomic DNA library preparation targeting the fungal ITS2 region of the extracted stool and mock fungal community DNA was PCR amplified using forward primer ITS7-XT (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGARTCATCGAATCTTTG) and reverse ITS4-XT (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCTCCGCTTATTGATATGC), where non-underlined nucleotides correspond to overhang adapter sequences. The targeted amplicon PCR reaction was performed using 12.5 µl of 2x KAPA HiFi HotStart ReadyMix (Roche, Pleasanton, CA), 5 µl of each 1.0 µM forward and reverse amplicon primers, and 2.5 µl of genomic DNA cycled at 95 °C for 3 minutes, followed by 25 cycles at 95 °C for 30 seconds, 55 °C for 30 seconds, and 72 °C for 30 seconds, then a 5-minute final extension at 72 °C and holding at 4 °C. The resulting ITS2 amplicons were then purified with 20 µl AMPure XP (Beckman Counter Canada, LP, Mississauga, ON), and a second amplification was performed to attach multiplexing indices. Second-round PCR used 25 µl of 2x KAPA HiFi HotStart ReadyMix, 5 µl of each Nextera XT index primer (N7XX/N5XX), and 15 µl of purified ITS2 amplicon DNA with the following conditions: 95 °C for 3 minutes, followed by 8 cycles at 95 °C for 30 seconds, 55 °C for 30 seconds, and 72 °C for 30 seconds, then a 5-minute final extension

at 72 °C and holding at 4 °C. Indexed amplicons were purified using 56 µl AMPure XP, then the libraries with positive concentrations were quantitated using PicoGreen and pooled in equimolar amounts. Gating of the pooled libraries was performed using BluePippin 1.5% cassettes (Sage Science Inc., Beverly, MA) to select 300–1000 bp fragments. Size-selected libraries were then purified using 0.6X AMPure XP and assessed for size on an Agilent Tapestation analyzer (Agilent Technologies Canada Inc., Mississauga, ON) and quantitated on a Qubit 2.0 (Thermo Fisher Scientific Inc., Waltham, MA). The libraries were then denatured using 0.5 N NaOH to 11 pM and spiked with 25% PhiX control DNA for stool samples and 25% or 50% PhiX for the fungal mock community. Sequencing was conducted using the Illumina MiSeq V3 (600 cycles; 2 x 300 bp) protocol. A total of 134 stool samples were sequenced, 65 in the first run and 69 in the second. Each MiSeq run was performed with an extraction water blank, no-template control, and positive control (25% PhiX spiked fungal mock community in technical duplicate). An additional run containing no stool samples and only technical duplicates of 50% PhiX-spiked fungal mock positive controls, water blanks, and no-template control was conducted. The resulting FASTQ sequences were demultiplexed and assessed for sequencing quality by FastQC v0.11.9[76] and MultiQC v1.8[77] using default parameters.

2.4 LotuS pipeline analysis of the mock fungal community and study cohort samples

The fungal ITS2 amplicons from the study participants' stool and the 19-member mock community were sequenced using the Illumina MiSeq platform and processed using the Less Operational Taxonomic Unit Scripts (LotuS) v1.62[69] pipeline. The *autoMap.pl* perl script was used to generate a mapping file to match paired-end reads, then primer sequences were appended to each pair of reads for downstream demultiplexing. The *lotus.pl* script was used to trim, quality filter, and dereplicate reads using the following parameters: trim reads to remove primer

sequences; remove sequences longer or shorter than the gated fragment size of 100–1000 bp; post-trimming of reads with a minimum average base quality of 33; remove sequences with ambiguous bases; remove reads with homopolymer length of 15 bp or more; base quality of ≥ 25 within a 50 bp k-mer window; and 50 bp ends of each sequence were trimmed if the sequence quality score fell below 25.

The Unified Search (USEARCH) v11.0.667[78] sequencing clustering tool was used to cluster reads with a ≥ 97% sequence identity. Chimeric and PhiX sequences were removed by alignment against the Unoise CHIMEras (UCHIME) v7.2[79] fungal database. Clustered read pairs overlapping by at least 10 bp with an average quality score of ≥ 33 were merged using Fast Length Adjustment of SHort Reads (FLASH) v1.2.10.[80] The merged reads were filtered for fungal ITS2 sequences by the ITSx v1.0.11.[81] Fungal taxonomic assignment was performed by alignment against the mothur release of the UNITE v8.2[58] fungal ribosomal database using BLAST+ v2.2.29.[82] Non-fungal taxa were removed by specifying parameters *-keepUnclassified* 0 and *-tax_group fungi*.

2.5 Checking for batch effects

Both plates of 25% PhiX spike-in samples were checked for potential batch effects after processing through the LotuS pipeline. Two different methods were utilized, in the first, all samples (including each mock positive control) were visualized using principal component analysis prior to applying any data transformation. The second method filtered for the 46 study participants with available cohort metadata, in which the samples were transformed using center log-ratio transformation and visualized using non-metric multidimensional scaling, specifying the Euclidean distance.

2.6 Mock fungal community assessment

2.6.1 Characterization of pre-defined mock fungal community sequences to determine unite database limitations

To investigate if taxonomic assignment is limited by the reference database, the 24 sequences representing a 19-member fungal mock community were aligned against the mothur v8.2 UNITE fungal ribosomal reference database using the Basic Local Alignment Search Tool+ (BLAST+) v2.9.0 software with parameters *blastn -query x -db y -out z -outfmt '6seqid sseqid length mismatch evalue pident qcovsq bitscore*', where *x* is the set of mock fungal community sequences, *y* is the UNITE database, *z* is the output file, *6seqid* is the query organism, *sseqid* is the hit organism unique digital object identifier (DOI), *length* is the alignment length, *mismatch* is the number of base mismatches, *evalue* is the expected value, *pident* is the percent identity, *qcovsq* is the query coverage per subject, and *bitscore* is the bit score.

The output data file from BLAST+ was processed using a custom script within the R programming language. BLAST data was cleaned and formatted using tidyverse[83] and seqinr[84] packages, used to read in, format, and process sequencing data. The processed data was then matched to the original UNITE database file using the hit organism DOI as a key, which provided the taxonomic assignments. Identified taxa were filtered to the species level in order to compare with the fungal mock community query. If a direct hit to the query was not found in the resulting table, query–hit matches were manually inspected for hit organisms that may have an alternative nomenclature; if so, they were reassigned with a standardized nomenclature to facilitate comparison. Each of the 19 fungal mock organisms was set as an individual group, and the top three hits matching each group were kept based on their highest value of bit score, percent identity, and expect value. Ties (i.e., hits with identical bit score, percent identity, and

expect value) were resolved by arbitrarily removing organisms with only one hit to decrease inaccurate identities and to prioritize hits pertaining to the expected mock member. Each group was further inspected for hits to the query sequence: if the known taxon was identified but was not within the top three hits, they were added to the group for assessment. A final table was generated to indicate if the query organisms were identified in the search, accompanied by the percent sequence identity of the top matching hits.

2.6.2 PIPITS pipeline analysis of the mock fungal community

Taxonomic characterization of fungal ITS2 amplicons from mock community sequencing was performed using the PIPITS v2.7[68] pipeline. A mapping file was created using *pispino_createreadpairslist* to link forward and reverse reads, which were then merged with VSEARCH v2.14.1[85] and quality filtered using the command *pispino_seqprep* with the following parameters: 80% of bases must have a quality score ≥ 30 and an average Phred score of 33. The resulting read pairs were converted to FASTA and merged into a single file.

The command *pipits_funits -x ITS2* was used to trim input reads to just the ITS2 marker region. Default parameters were used for fungal ITS2 extraction. The *pipits_process --unite 04.02.2020 --includeuniqueseq*s command was used to cluster sequences with ≥ 97% sequence similarity with and assign taxa against the UNITE v8.2 database. Chimeric and merged sequences under 100 bp were automatically removed. Singletons were retained in the output counts and taxonomic tables.

2.6.3 Mothur pipeline analysis of the mock fungal community

Reads from the 25% and 50% PhiX sequencing runs were processed by the mothur v1.43.0[86] pipeline using the commands *make.file* to generate a mapping file and *make.contigs* to

assemble contiguous sequences by specifying a base quality threshold of 20. Merged reads were filtered using the command *screen.seqs* for contigs containing no ambiguous bases, the minimum sequence length of 100 bp, 25 bp minimum overlap, and a maximum homopolymer length of 15 bp. Amplicon primer sequences were re-trimmed using *pcr.seqs* command, where a maximum of 3 bp difference in the primer sequences was allowed. In order to decrease the computational processing burden, the *unique.seqs* command was used to group identical sequences and select a representative.

The taxonomically informative ITS2 subregion was extracted from the list of filtered contigs using ITSx v1.1b. The argument *allow_reorder* was set to true, and sequences with single domains were included if they had an E-value of $1x10^{-3}$ and a cutoff score of 0 set by the argument *allow_single_domain*. The mothur command *unique.seqs* was used to group identical sequences outputted by ITSx, and counts corresponding to the representative sequences were added to the existing count file.

Taxonomic identifications were assigned using the UNITE v8.2 database with the command *classify.seqs* specifying the Wang approach with a confidence cutoff of 95%. Unclassified sequences were removed using the command *remove.lineage*. Chimeric sequences were identified and removed using *chimeria.vsearch* specifying the UCHIME v7.2 fungal database as the chimeric reference. The remaining sequences were clustered using the commands *pairwise.seqs* and *cluster* by specifying the default opticlust method and a ≥ 97% sequence similarity threshold to produce the counts and corresponding taxonomic tables.

2.6.4 Data wrangling of the mock fungal community

Count matrices and their corresponding taxonomic tables produced by each pipeline (PIPITS, mothur, and LotuS) for mock fungal sequences were formatted into a mergeable format for downstream analysis using R programming language. Each column heading corresponding to the sequenced samples was denoted in the format of the pipeline used for processing sequencing data, followed by the percentage of PhiX spike-in and a subscript (rep1 or rep2) to denote technical replicates (i.e., mothur-25%$_{rep1}$ corresponds to the first technical replicate of a mock sample sequenced using 25% PhiX, while mothur-25%$_{rep2}$ is the second replicate). Output taxonomic tables were standardized by assigning a standard nomenclature across all pipelines at all taxonomic levels. Using the mothur pipeline taxonomic table format as a guide; if no species-level assignments were generated by the LotuS and PIPITS pipelines, then the genus was used to represent the organism. If no genus-level assignments were generated, then the family was used to represent the organism. If neither the genus nor family levels were generated, the assignment was set to "*f_NA*" to represent a fungal organism with an unknown identity at the family, genus, or species taxonomic levels. All organisms were identified to belong to the fungal kingdom.

To associate the OTUs of each mock fungal sample with the different pipelines used to generate them, the OTU identifiers corresponding to each sample were encoded such that they could be traced back to their processing pipeline (i.e. OTU1y is from mothur-25% and OTU1m is from mothur-50%). A unique numerical key value was then given to each OTU table and merged using the tidyverse packages. The phyloseq[87] package was used to merge counts from different pipelines belonging to the same species taxon level. Each sample was filtered to select the expected 19 taxonomically distinct mock fungal taxa. The ggplot2[88] package was used to visualize the comparative analysis of the different pipelines.

2.6.5 Statistical analysis of mock fungal community

Taxonomic abundances were analyzed using raw values without transformation or scaling since the expected relative ratios of the mock fungal members have been previously established. The phyloseq R package was used to determine alpha diversity indices by the number of observed OTUs, Shannon, and Inverse Simpson metrics. The phyloseq package was also used to generate heatmaps by non-metric multidimensional scaling (NMDS), specifying the Morisa-Horn index to determine sample similarity. The values are transformed to a log4 scale. Beta diversity was calculated with the vegan package using NMDS specifying the Morista-Horn index. The vegan[89] and pairwiseAdonis[90] packages were used to perform the PERMANOVA test for significance between samples. Statistical testing was conducted by Kruskal–Wallis ANOVA and Dunn's post hoc and all hypotheses were tested at a level of significance of $\alpha = 0.05$ with adjustment for false discovery rate using the Benjamini–Hochberg method. K-means clustering was performed using the factoextra[91] package.

2.6.6 Assessment of the mock fungal community from raw sequences

To confirm the sequencing of all 19 mock community members, the raw Illumina sequences were investigated in two ways. In the first method, paired-end reads were merged using FLASH v1.2.11 by setting an overlap parameter of 10–250 bp. The reads were then converted from FASTQ format to FASTA using the command *fastq_to_fasta* from the FASTX v0.0.14 toolkit[92]. Dereplication to obtain unique sequences by VSEARCH v2.22.1 was used before alignment against the reference mock sequences[74] using BLAST+ v2.14.0 with the parameters previously mentioned in section 2.6.1.

The second method to affirm the identities of all 19 mock fungal sequences in our raw sequences also used FLASH v1.2.11 to merge reads in the first step. Any remaining primer

sequences were trimmed using trimmomatic v0.39[93] with parameters, *MINLEN:100 LEADING:25 TRAILING:25 SLIDINGWINDOW:50:25* to remove reads with a length below 100 bp, cut bases from the ends if there was a sequence quality below 25, and a 50 bp sliding window trim if the quality was below 25. Reads were then converted to FASTA using the FASTX v0.0.14 toolkit, and dereplicated by VSEARCH v2.22.1 and the ITS2 region extracted by ITSx v1.1.3. Alignment against the reference mock sequences was performed using BLAST+ v2.14.0.

2.7 Study participant sample assessment

2.7.1 Data cleaning, counts transformation, positive control check, and metadata age stratification of study participants

The LULU[94] R package was used to merge OTUs with a high sequence similarity to decrease spurious OTU counts without removing them from downstream statistical analysis. The LULU algorithm takes an OTU table and a file of associated sequences, performs an all-against-all comparison (via BLAST+) of the sequences to generate a match list, and then groups co-occurring and sequence-similar OTUs together into a single aggregate OTU. The BLAST+ parameters were set to default, retaining sequence alignments with ≥ 84% sequence identity and ≥ 80% query coverage per high-scoring sequence pair. A Good's coverage index was used to estimate the acceptable lower bound of OTUs per sample. Samples with an overall count of ≤ 500 OTUs were excluded from the study to obtain a Good's coverage index of ≥ 96.5%. A Good's coverage index of ≥ 96.5% retains samples with ≤ 3.5% of overall reads mapping to unique/rare OTUs. This is to reduce the number of samples with sequences arising due to sequencing errors.

Since OTUs with zero counts cannot undergo log transformation, the table of counts was transposed, and values with an abundance of 0 were given a pseudocount using default

parameters from the *cmultRepl* function of the zCompositions[95] R package. The table of counts was then transformed using the center log-ratio (CLR) transformation via the *clr* function from the compositions[96] R package. The CLR-transformed samples were visually inspected for normality and skew by density distribution and Quantile-Quantile (Q-Q) plot using the ggplot2 R package. The mock community control was not subject to LULU curation or count transformation since the relative ratio and identities of the controls are known. Instead, the mock community data were visually inspected for the presence or absence of expected species using raw counts.

2.7.2 Visualization and statistical analysis

Visual inspection of the top ten genera was performed for samples grouped by POMS, monoADS, or unaffected controls. Raw counts were transformed into proportions to aid in the visual inspection. Data were visualized using the ggplot2 R package. CLR-transformed values were used in the statistical analyses (except for alpha diversity). Statistical comparisons were stratified by POMS, monoADS, or unaffected controls.

Biodiversity was estimated by alpha diversity using the number of observed OTUs, Shannon, and inverse Simpson metrics. Alpha diversity was compared between the POMS, monoADS, and unaffected participants using the Kruskal–Wallis ANOVA and Dunn's post hoc testing with Benjamini–Hochberg correction (via the ggpubr[97] and rstatix[98] R packages). As a complementary approach, alpha diversity was also examined by sex or age at sample collection grouped as child (≤ 14 years) or youth (15–24 years) for all participants. The examination of alpha diversity on sex or age group was performed by Wilcoxon ranked sum test adjusted for false discovery rate using the Benjamini–Hochberg correction method.

Beta diversity analysis was performed by principal component analysis (PCA) and NMDS, using Euclidean distance via the PCAtools[99] and vegan R packages. Significance testing by PERMANOVA on the calculated Euclidean distances was performed by the *adonis2* function in the vegan R package, followed by pairwiseAdonis. The P-values were adjusted for false discovery rate by Benjamini–Hochberg correction using the stats R package.

Differential species abundance testing between POMS, monoADS, and unaffected control participants was performed on OTUs in two ways. The first method was by the ANOVA-Like Differential Expression (ALDEx2)[100] R package on non-transformed raw counts, and the second is linear discriminant analysis effect size (LEfSe)[101] on CLR-transformed counts using the *run_lefse* function of the microbiomeMarker[102] R package for both individual OTU and aggregated taxon counts. Both ALDEx2 and LEfSe are statistical packages used for differential abundance analysis of biological data. The LEfSe algorithm transforms OTU data in preparation for downstream statistical testing steps. These steps include comparing POMS, monoADS, and unaffected control participants using Kruskal–Wallis ANOVA and Wilcoxon rank sum test. Linear discriminant analysis was used to predict significant taxa associated with each group. Effect sizes (generated by LEfSe) were plotted using a modified *plot_ef_bar* function to display data according to taxonomic levels. Differential testing by ALDEx2 differs from LEfSe in the first step, whereby posterior probabilities are generated for each OTU in ALDEx2, and Monte-Carlo sampling from a Dirichlet distribution is used to re-sample the data, which is then transformed before Kruskal–Wallis ANOVA and Wilcoxon rank sum tests are performed. Linear discriminant modelling is also not performed in ALDEx2. In all statistical tests, an alpha level of 0.05 was used to indicate significance.

The BLAST+ sequence alignment software package and the BLAST fungi NCBI Reference Sequence (RefSeq) ITS database[103] (PRJNA177353 - downloaded on date 2022-04-13) were used to reassign taxa above the species level to a more precise taxonomic level. The BLAST+ parameters used were *qcov_hsp_perc* 80 and *perc_identity* 80. BLAST+ results were filtered for hits with the lowest expected value and highest query coverage per high-scoring segment pair, followed by the highest percent identity.

CHAPTER 3: EVALUATION OF THREE METABARCODING PIPELINES USING A FUNGAL MOCK COMMUNITY

3.1 Introduction

Current evidence suggests the fungal kingdom emerged in the Neoproterozoic era at least 890 million years ago.[104,105] Since the discovery of fungi, studies have predicted the existence of well over 1 million different fungal species.[51,106,107] Approximately 120,000 fungal species have been described, and ~300 of these are known to cause disease in animals.[58,108] Fungi have a significant role in the ecology of our planet. They have also had a significant impact on human history. They are decomposers of plants, symbiotic partners to other microorganisms, producers of alcohol and antibiotics, and are imperative to the bioindustrial and pharmaceutical industries.[51] Despite their significance to our ecosystem and their roles in food and medicine, they are greatly understudied compared to other organisms, such as bacteria, especially in how they affect human health.[71,109]

A basic first step in studying fungi is their taxonomic classification; however, fungi has historically been somewhat subjectively classified based on phenotypic features. The classification of fungi is also limited due to the difficulty in their cultivation.[42,51] There is a challenge to identify asexual fungi using culture-based methods because many are unable to grow in laboratory conditions.[110] Advances in molecular identification methods now provide researchers with the ability to better characterize unculturable species by improving taxonomic resolution; indeed, such advances have led to the reclassification of several fungi.[58,108]

Metataxonomic barcoding using marker genes such as 16S rRNA and chaperonin-60 genes is the standard approach for the taxonomic classification of bacterial communities.[41,111] The metataxonomic counterpart for classifying fungal communities is the internal transcribed

spacer (ITS) region of the fungal ribosomal RNA.[54,55,58,112] The ITS regions of the fungal rRNA operon have traditionally been characterized by Sanger sequencing.[42,43] However, Sanger sequencing is a low throughput method, limited by the number of specimens that can be run at once, rendering this technology obsolete for complex ecological diversity analyses of samples with large numbers of nucleic acid signatures.[42] Advances in sequencing technologies, particularly high throughput short-read sequencing, have greatly enhanced our ability to investigate fungal communities by allowing simultaneous analysis of many fungal sequences in parallel.[28,38,63,113]

The "sequencing by synthesis" short-read sequencing technology, which forms the basis of the Illumina line of sequencing platforms, is a popular choice for characterizing fungal communities, but it is not without its limitations.[42,43,63,112] Such approaches generate clonal sequence clusters that generate a fluorescent signal corresponding to the incorporated base on the nascent strand during the sequencing reaction. The deep amplicon sequencing approach used in metaxonomic barcoding potentially results in low overall diversity in the sequence clusters that diminishes the sequencers ability to discriminate the different clusters, resulting in an erroneously low representation of the true sample diversity.[114] One approach to address this issue is to spike in a high concentration (40-50%) of sequencing control bacteriophage (PhiX), which reduces the density of sequencing clusters and increases the ability to discriminate between fluorescent signals of incorporated bases to detect rare species by distinguishing clusters (**Figure 3.0**).[114,115]

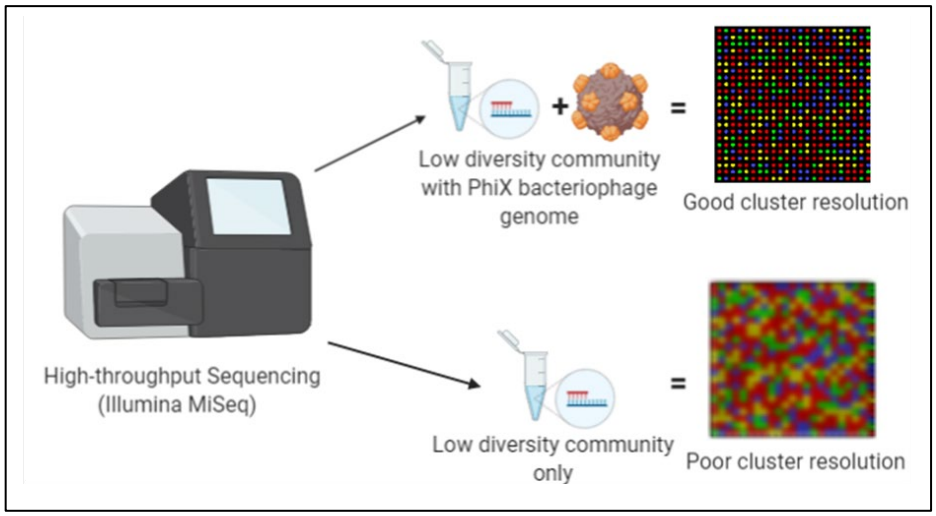


Figure 3.0. Addition of PhiX bacteriophage genome to low-diversity samples improves sequence quality. Image generated with BioRender with licensed publishing permission.

An additional challenge for metataxonomic barcoding of fungal communities lies in the ability to assign taxa to the sequences. Taxonomic assignment of fungal communities requires a database of fungal marker sequences annotated with the taxa to which they belong. The UNITE fungal ribosomal database is a well-curated database widely used in fungal metataxonomic studies targeting the ITS region.[42,58] The process of taxonomic assignment is normally accomplished by bioinformatics pipelines created specifically for this purpose.[51,63] Popular pipelines for taxonomic assignment of fungal communities from metabarcoding sequences include Less Operational Taxonomic Unit Scripts (LotuS), PIPITS, and mothur.[68,69,86] The selection of analytic pipelines and/or software can greatly affect taxonomic identification when characterizing fungal communities.[63]

In order to improve our ability to characterize fungal communities, we set out two objectives: 1) assess the ability of different PhiX spike-ins concentrations to improve fungal metabarcoding sequence quality, and 2) assess the performance of three different bioinformatics pipelines to characterize fungal communities. To achieve these objectives, we used a mock community panel containing 19 known fungal organisms in specific 18S rRNA ratios.[74] The fungal mock community serves as ground truth to assess the effect of varying PhiX spike-ins concentrations and the taxonomic classification abilities of the tested fungal metabarcoding pipelines (**Figure 3.1**).

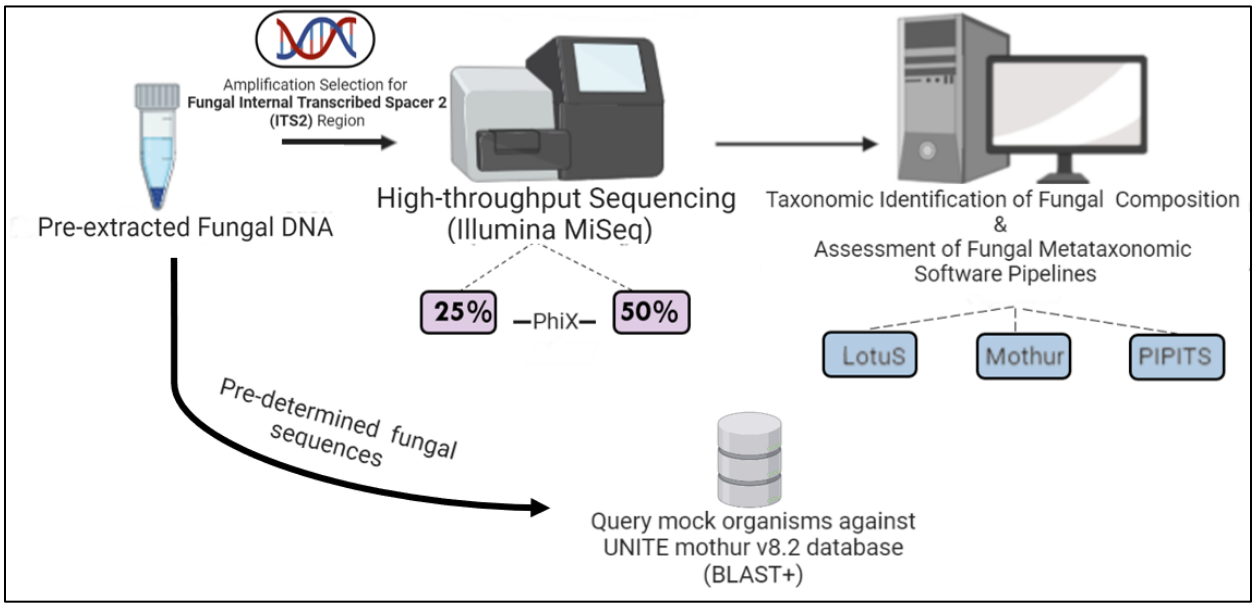


Figure 3.1. Fungal mock community sequencing protocol. Extracted fungal mock community DNA is sequenced using the Illumina platform in technical replicates, with either a 25% or 50% PhiX bacteriophage spike-in. Image generated with BioRender with licensed publishing permission.

3.2 Results

3.2.1 PhiX spike-in concentration affects sequencing quality and the number of sequencing reads

Sequencing quality assessment by FastQC and MultiQC shows that the mean base quality score for mock fungal samples spiked with a 50% PhiX produces a stable Phred score of > 30 over ~200 bp in both paired-end reads and technical replicates (**Figure 3.2a**). In contrast, a 25% PhiX spike-in exhibits a decline in mean sequencing quality after the first 50 bp but still retains a Phred above 30 for the following ~200 bp, then rapidly decreases in sequencing quality thereafter (**Figure 3.2b**). This translates to 86.1% of the bases called by using a 50% PhiX spike-in having a Phred score of 30 or greater (≥ 99.9% accuracy), which is higher compared to a 25% PhiX spike-in where 70.8% of the bases called have a Phred score ≥ 30. However, a 25% PhiX spike-in results in 5.6-fold more reads generated, as an average of 10,168,535 raw reads were produced when accounting for technical replicates, while a 50% PhiX spike-in on average produced 1,821,961 reads. When the raw reads were filtered for fungal annotated identities, following processing by sequence analysis pipelines, a 25% PhiX spike-in yielded an average of 976,441 reads when accounting for all pipelines and technical replicates. This is 4.2-fold more reads on average compared to an average of 235,036 filtered reads produced from 50% PhiX spike-in samples.

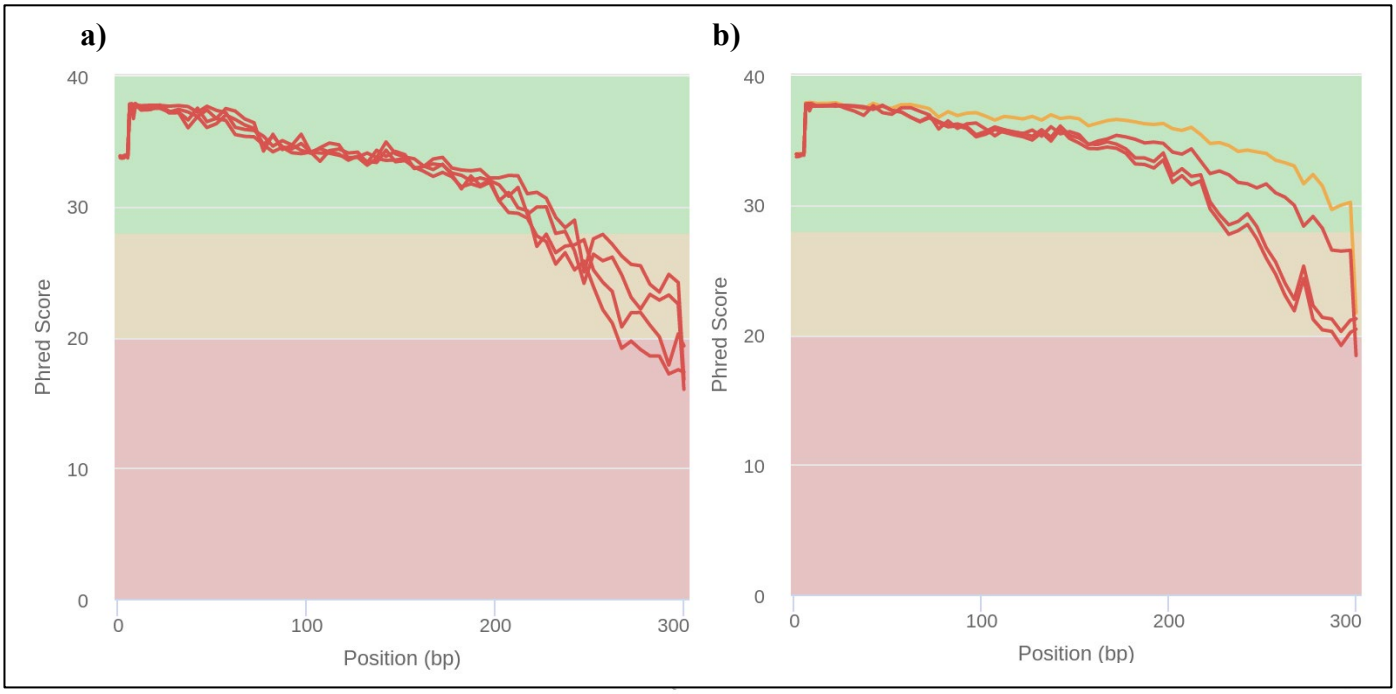


Figure 3.2. Average sequencing quality scores per base position for fungal mock community samples. **a)** paired-end DNA sequenced in duplicate with a 25% PhiX spike-in. **b)** paired-end DNA sequenced in duplicate with a 50% PhiX spike-in. The x-axis represents the base position along the sequence length. A Phred score ≥ 30 (green region) indicates a 99.9% accuracy in the base called, 20–29 (yellow region) is 99.0% base call accuracy, and 0–19 (red region) is 90.0% accuracy.

3.2.2 Some fungal taxa cannot be identified within the UNITE fungal ribosomal database

The mock community was queried against the UNITE fungal ribosomal data using BLAST+. We noticed a nomenclature difference between the queried organisms and the identified best hit for four taxa. The *Neosartorya fischeri* query aligned to *Aspergillus fischeri*; *Fusarium graminearum* aligned to *Gibberella zeae*; *Fusarium verticillioides* aligned to *Gibberella fujikuroi*; and *Rhizoctonia solani* aligned to *Thanatephorus cucumeris*. Upon further inspection, these hit organisms were found to correspond to synonymous or alternative nomenclatures of the query organism, some of which represent teleomorphic or anamorphic forms of the fungi.[116,117] We found that all the expected fungi can be identified within the UNITE fungal ribosomal database with the exception of *Mortierella verticillata*, which did not align to any identical or synonymous hits. (**Table 1**).

Table 3.0. Fungal mock community organisms queried against the UNITE database.

Mock Organism	Match in UNITE Database	Percent ID (%)
Alternaria alternata	Yes	99.4
Aspergillus fischeri (Neosartorya fischeri)[a]	Yes	95.5–100
Aspergillus flavus	Yes	100
Candida apicola	Yes	100
Chytriomyces hyalinus	Yes	98.7–100
Claviceps purpurea	Yes	99
Fusarium graminearum (Gibberella zeae)[a]	Yes	100
Fusarium oxysporum	Yes	99.6
Fusarium verticillioides (Gibberella fujikuroi)[a]	Yes	100
Mortierella verticillata	No	0
Naganishia albida	Yes	99.1
Penicillium expansum	Yes	100
Rhizoctonia solani (Thanatephorus cucumeris)[a]	Yes	96.8
Rhizomucor miehei	Yes	99.5–99.7
Rhizophagus irregularis	Yes	94.6–99.7
Saccharomyces cerevisiae	Yes	99.2
Saitoella complicata	Yes	100
Trichoderma reesea	Yes	100
Ustilago maydis	Yes	99.6–100

[a] Organism name was reassigned using standardized nomenclature.

3.2.3 A 50% PhiX spike-in results in more consistency between technical replicates, but a 25% PhiX spike-in yields more reads, and LotuS retains the most reads

Sequencing data of the mock fungal community and their technical replicates with a 25% or 50% PhiX spike-in were processed through the LotuS, mothur, and PIPITS taxonomic assignment pipelines. The number of assignable reads to the species taxonomic level differed between 25% PhiX spike-in replicates (**Table 3.1**). Samples processed with the LotuS pipeline yielded 1,626,046 reads for the first replicate and 549,078 reads for the second replicate. Similarly, sequence processing by mothur yielded 1,524,364 reads for the first replicate and 516,149 reads for the second. The sequences processed by the PIPITS pipeline produced the least reads: 389,870 for the first replicate and 1,253,141 for the second. In contrast, the number of reads recovered between 50% PhiX spike-in replicates were relatively close: 254,856 vs 220,284 reads for LotuS, 243,010 vs 210,693 reads for mothur, and 253,680 vs 227,690 reads for PIPITS. An experimental error may have occurred during the sequencing of the 25% PhiX spike-in samples, as there is a large difference in the number of reads when comparing against the technical replicate. Due to the fact that these replicates were run in two batches, we checked to see if this difference between technical replicates biases the data by visualizing for batch effect which is discussed in the appendix section *AP1*.Overall, a 50% PhiX spike-in yielded less reads, but generated more consistent results between technical replicates in this study.

We further combined technical replicates for each PhiX spike-in to contrast the total number of reads obtained by different PhiX spike-in concentrations, with the following results: 2,175,124 reads LotuS-25% vs 475,140 reads LotuS-50%; 2,040,513 reads mothur-25% vs 453,703 reads mothur-50%; and 1,643,011 reads PIPITS-25% vs 481,370 reads PIPITS-50%. When combining reads by processing pipelines, the LotuS pipeline notably contained the highest number of reads, followed by mothur, then PIPITS. Taken together, a 25% PhiX spike-in

consistently yields more reads after sequence processing, and the LotuS processing pipeline recovers the highest number of sequencing reads.

Table 3.1. Recovered reads from three different fungal metabarcoding pipelines.

	Samples											
	LotuS-25%$_{rep1}$	LotuS-25%$_{rep2}$	LotuS-50%$_{rep1}$	LotuS-50%$_{rep2}$	Mothur-25%$_{rep1}$	Mothur-25%$_{rep2}$	Mothur-50%$_{rep1}$	Mothur-50%$_{rep2}$	PIPITS-25%$_{rep1}$	PIPITS-25%$_{rep2}$	PIPITS-50%$_{rep1}$	PIPITS-50%$_{rep2}$
Number of Reads	1,626,046	549,078	254,856	220,284	1,524,364	516,149	243,010	210,693	389,870	1,253,141	253,680	227,690
Grouped by PhiX	2,175,124		475,140		2,040,513		453,703		1,643,011		481,370	
Grouped by Pipeline	2,650,264				2,494,216				2,124,381			

3.2.4 Choice of sequence processing pipeline influences community profile

Alpha diversity analysis of replicates grouped by PhiX concentration and analysis pipeline showed differing richness for 25% and 50% PhiX spike-ins (**Figure 3.3a first panel**). The 25% PhiX samples have higher OTU richness than the 50% PhiX samples, although the alpha diversities were similar when accounting for the evenness of the community (**Figure 3.3a second and third panels**). The Shannon and inverse Simpson indices show that the diversity between samples processed with varying methods is generally similar, with the exception of 25% PhiX spike-in samples processed by PIPITS. No significant differences in alpha diversity were found among samples grouped by pipeline (all $p > 0.05$) (**Figure 3.3b**). Although species composition was similar between technical replicates, they differed by processing pipeline (**Figure 3.4**). The mothur and LotuS pipelines identified the species *Saitoella complicata* at the highest proportion. The PIPITS pipeline similarly identified *Saitoella* at the highest proportion, but only to the genus level. The species profiles between mothur and LotuS were most similar in composition, although mothur identified a higher proportion of genus *Naganishia* than LotuS. LotuS was able to classify this OTU to the species level as *Naganishia adeliensis*. The most obvious difference between taxonomic profiles by PhiX spike-in concentration is for the PIPITS pipeline, where the species *Naganishia aleiensis* spiked with 25% PhiX exhibits a much lower proportion when compared to the 50% spike-in.

Inspection of the top-ten most abundant species for each processing pipeline revealed a similar trend (**Figure 3.5**). Taxonomic assignments differed between processing pipelines, but the profiles are nearly identical between technical replicates and PhiX concentrations for samples processed by the LotuS and mothur pipelines. In contrast, samples with different PhiX spike-in concentrations differed when processed by PIPITS. When processed by PIPITS, the 25% PhiX

spike-in resulted in a higher proportion of the genus *Saitoella*, whereas the 50% PhiX spike-in showed a higher proportion of the species *Naganishia adeliensis*.

We performed beta diversity analysis to investigate the inter-sample differences. A preliminary scree plot analysis indicated that three dimensions should account for 99.99% of the variance associated with the mock community profiles; accordingly, the number of dimensions for NMDS was set to three (**Figure 3.6**). The beta diversity plot in **Figure 3.7** shows that the community profiles are similar for samples processed by the same pipeline across PhiX concentrations and between technical replicates. Statistical significance testing by PERMANOVA revealed a significant difference between profiles identified by different processing pipelines ($q = 0.001$). Pairwise testing by post hoc analysis indicated no significance between differing PhiX concentrations for samples processed by the same pipeline, but there are significant differences between pipelines: LotuS vs mothur ($p = 0.035$, $q = 0.029$), LotuS vs PIPITS ($p = 0.025$, $q = 0.028$), and mothur vs PIPITS ($p = 0.020$, $q = 0.031$). Similar grouping trends are found from a silhouette plot analysis used to show how close each point in one cluster is to points in the neighbouring clusters (**Figure 3.8a**). The silhouette plots show that three clusters are appropriate for k-means analysis (**Figure 3.8b**), which distinctly grouped samples by processing pipeline.

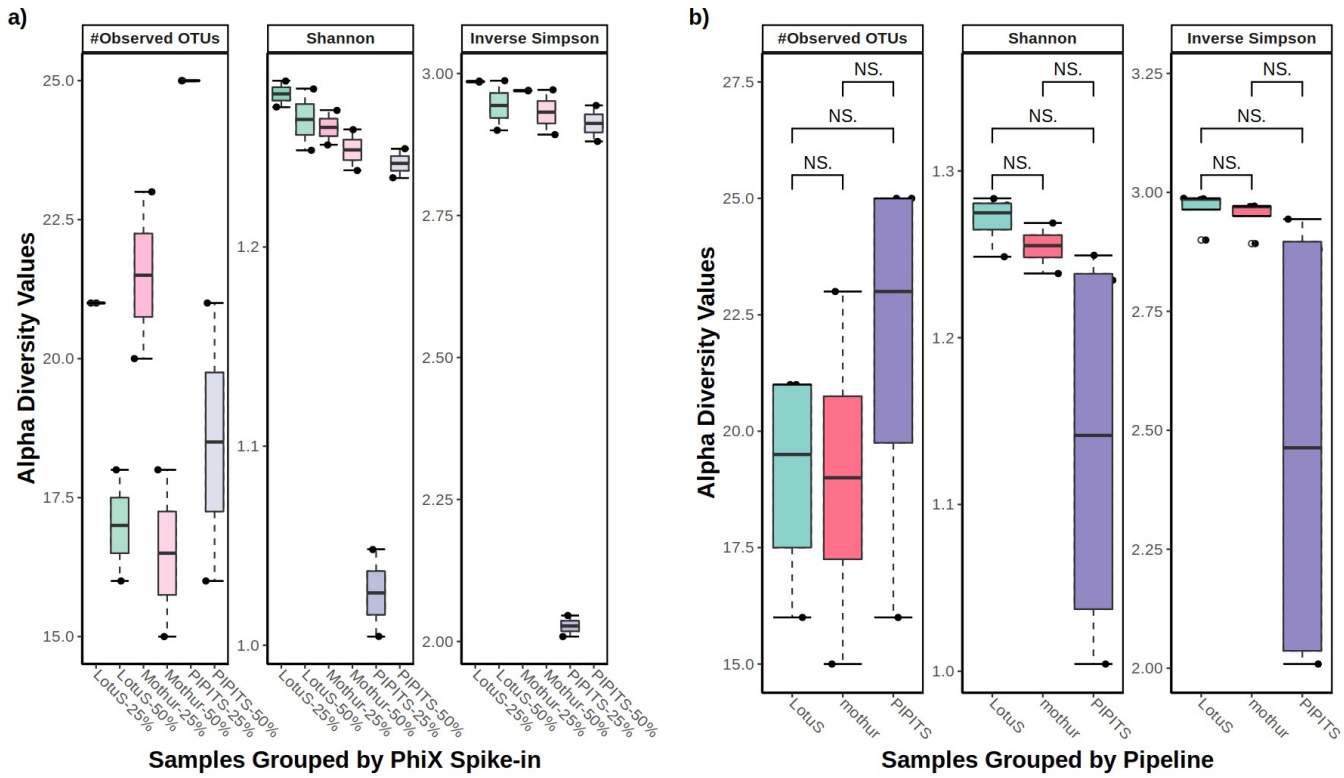


Figure 3.3. Alpha diversity box plots of a fungal mock community processed using LotuS, mothur, and PIPITS pipelines. The technical replicate pairs of the mock community were processed using the LotuS, mothur, and PIPITS pipelines. **a)** Technical replicate pairs grouped PhiX spike-in (n = 2 in each group). **b)** Technical replicate pairs grouped by processing pipeline (n = 4 in each group).

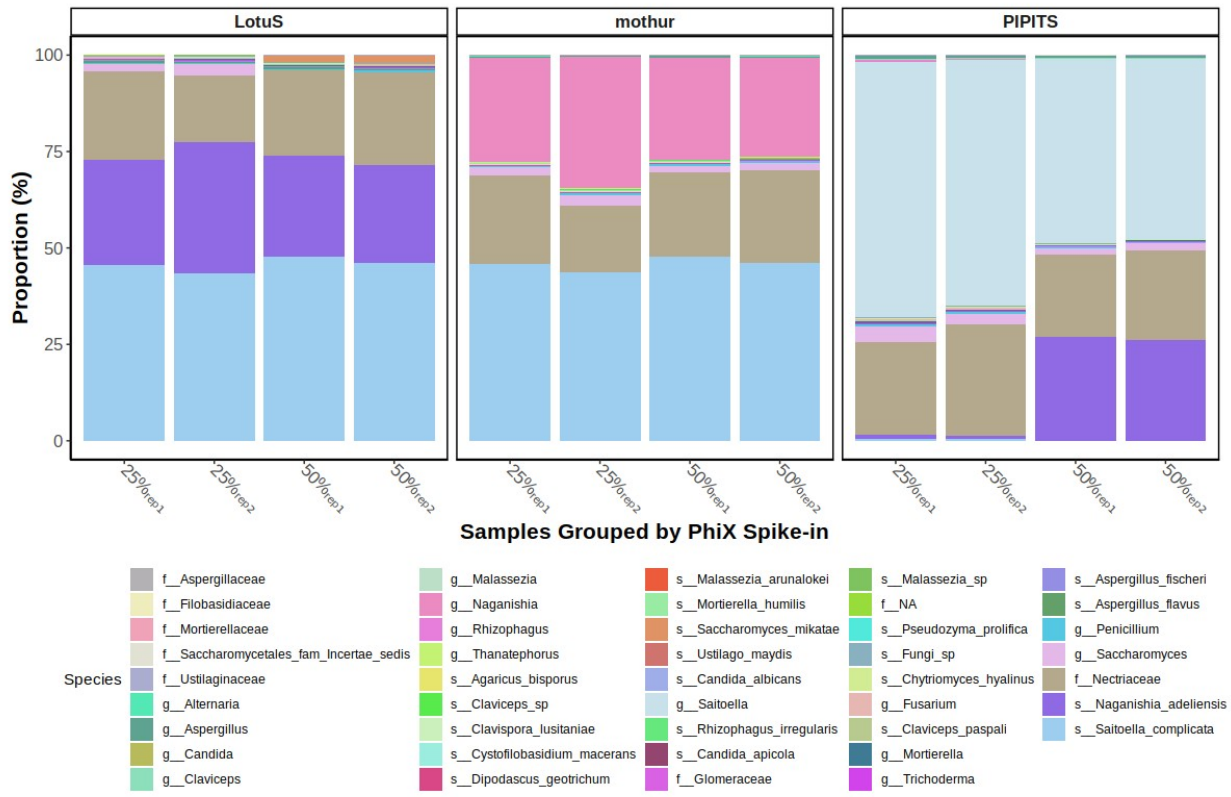


Figure 3.4. Stacked bar plot of complete taxonomic profiles characterized by different fungal metabarcoding pipelines. A 19-member fungal mock community was sequenced using different PhiX concentrations (25% or 50%) in technical replicate pairs (rep1 or rep2) and processed using three different fungal metabarcoding pipelines. Organisms labeled "f_NA" represent no assignment at the family, genus, or species taxonomic levels but belong to the fungal kingdom.

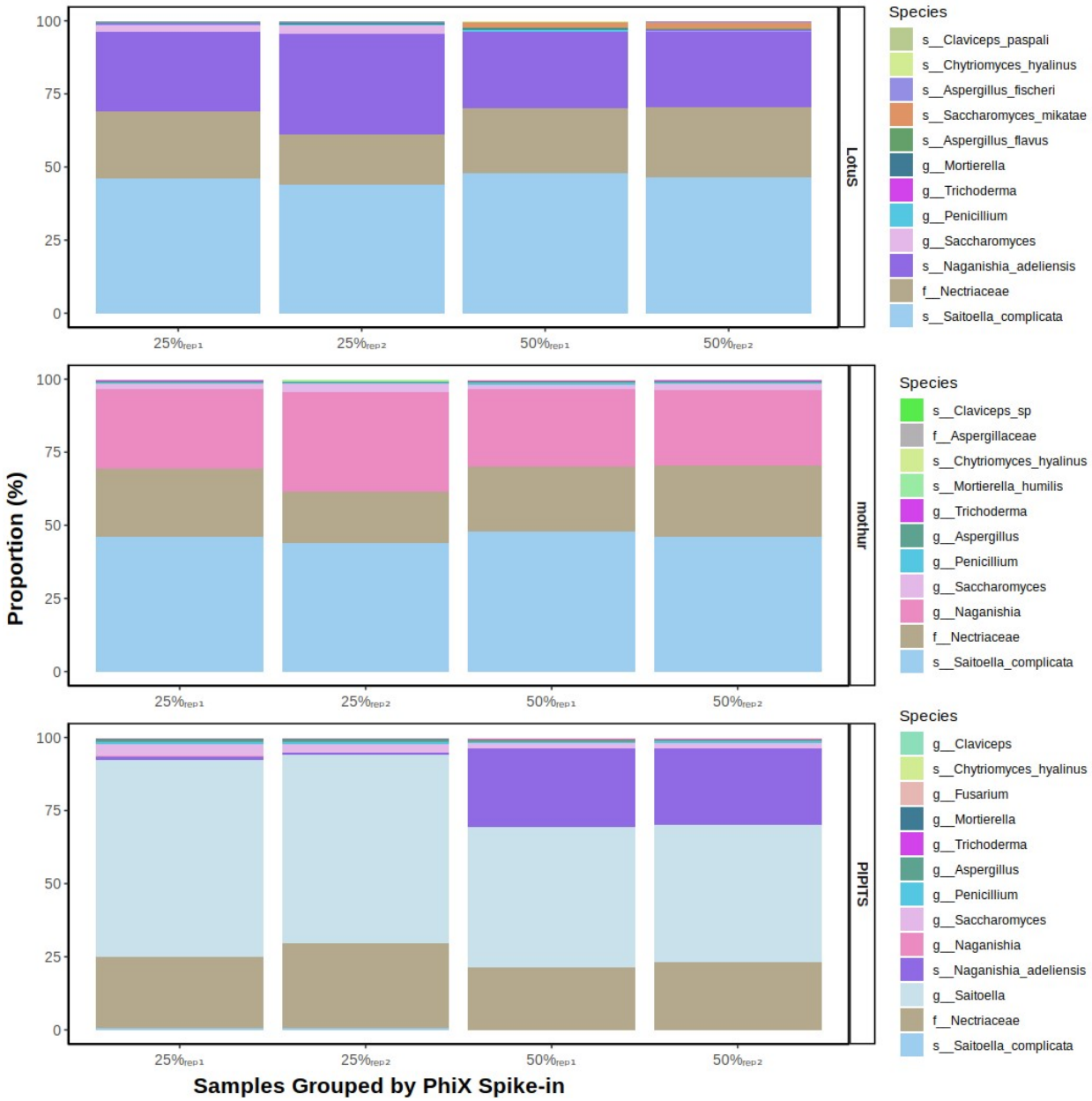


Figure 3.5. Stacked bar plot for the top ten most abundant taxa identified per fungal metabarcoding processing pipeline. A 19-member mock fungal community was sequenced with different spike-in concentrations (25% or 50%) of PhiX in technical replicate pairs (rep1 or rep2) and processed using three different pipelines.

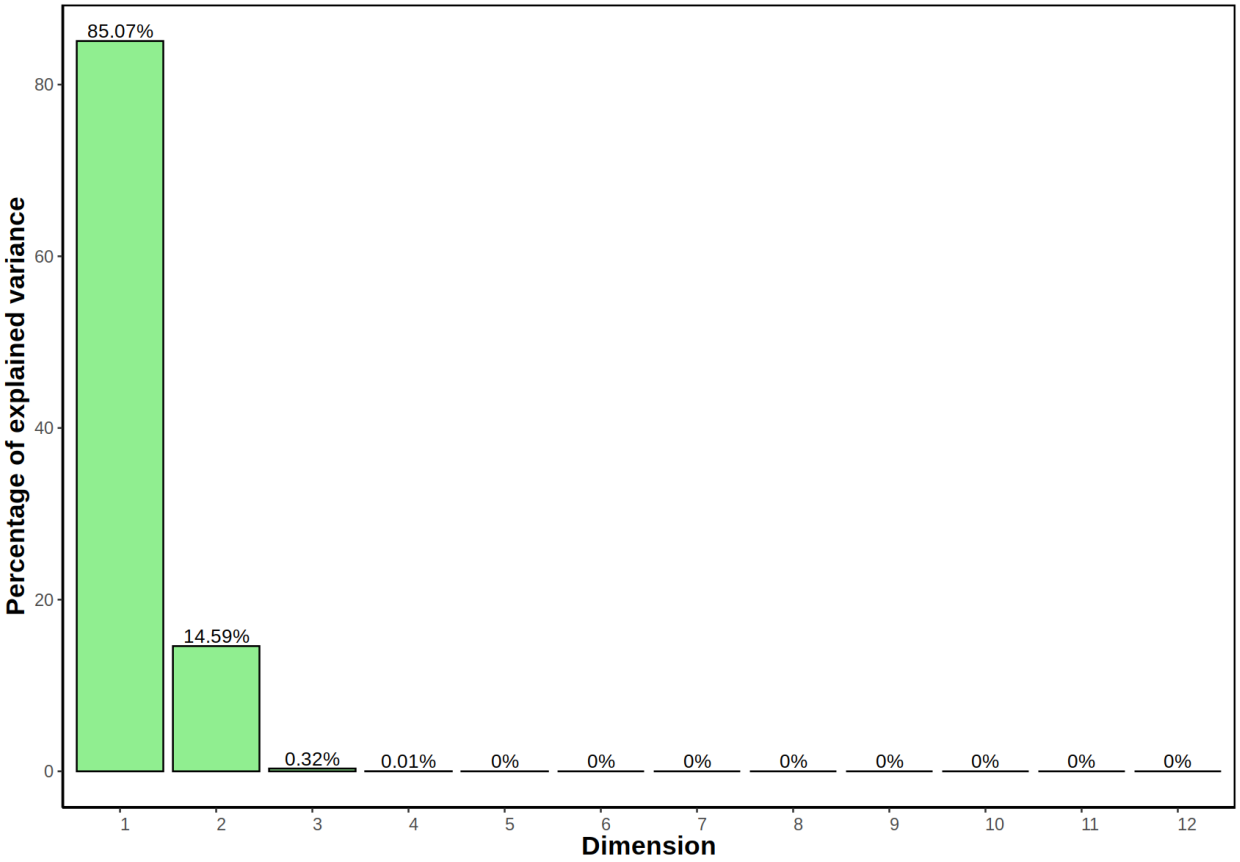


Figure 3.6. Scree plot showing the percent variance accounted for per dimension. The distance measure used is the Morista-Horn index.

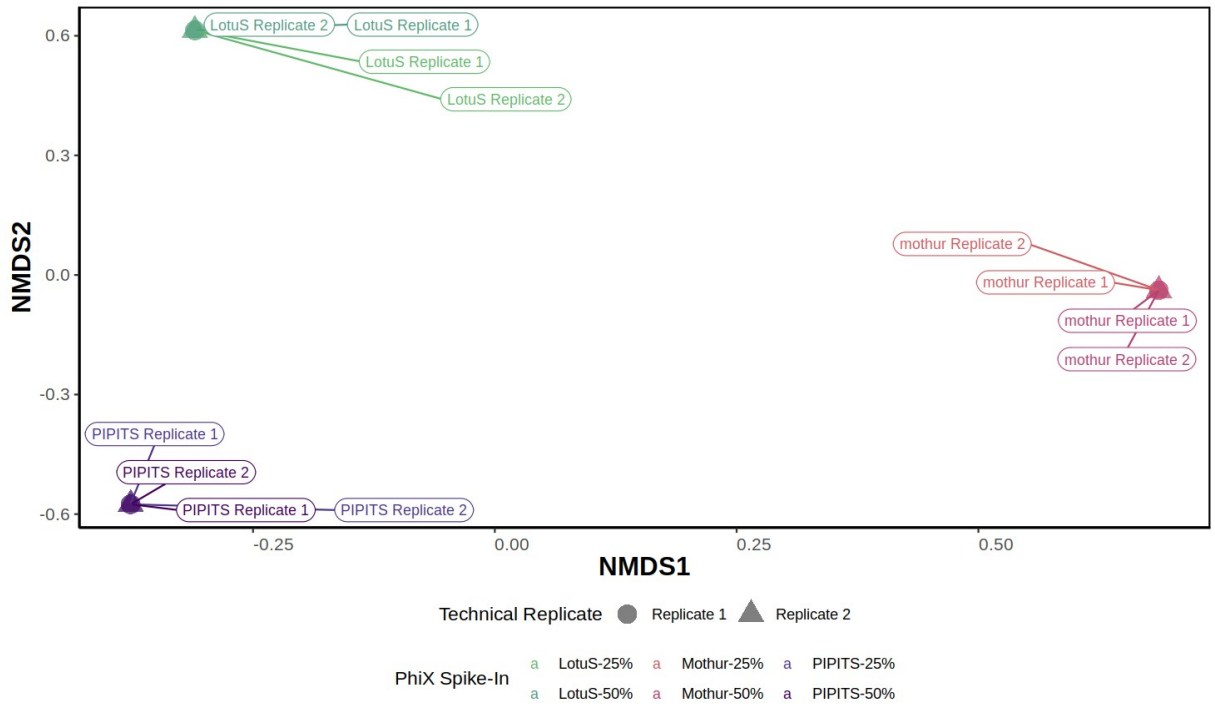


Figure 3.7. NMDS analysis of beta diversity of a fungal mock community sequenced in replicate at 25% and 50% PhiX spike-in concentrations and processed using the LotuS, mothur, and PIPITS pipelines. The distance measure used is the Morista-Horn index. Statistical analysis by PERMANOVA detected significance (q = 0.001). Post-hoc tests reveal differences between pipelines: LotuS vs mothur (p = 0.035, q = 0.029), LotuS vs PIPITS (p = 0.025, q = 0.028), and mothur vs PIPITS (p = 0.020, q = 0.031).

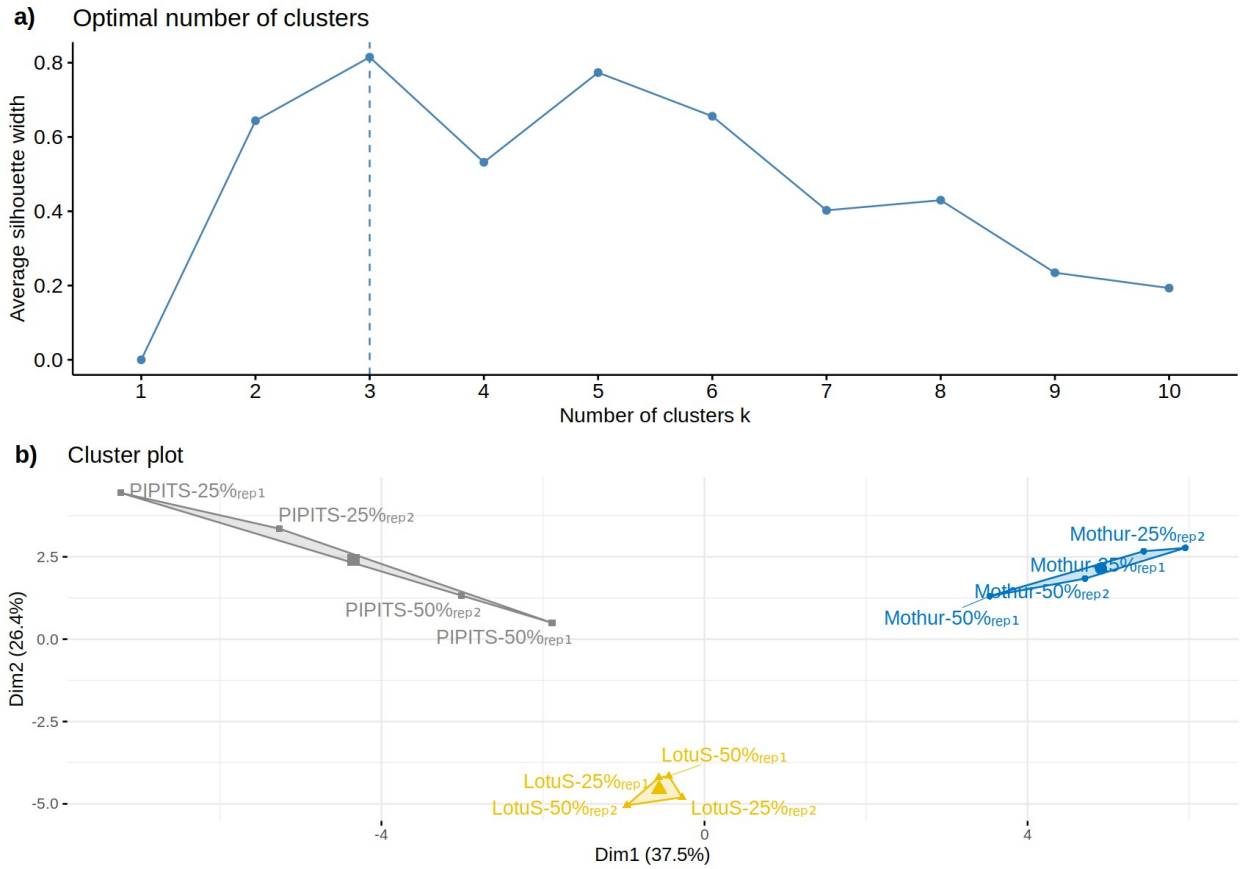


Figure 3.8. **a)** Silhouette plot and **b)** k-means cluster plot of the beta diversity of a fungal mock community sample sequenced in replicate (rep1, rep2) at 25% and 50% PhiX spike-in concentrations and processed by the LotuS, mothur, and PIPITS pipelines. **a**) Silhouette plot to measure the appropriate number of clusters for k-means cluster analysis. **b)** K-means clustering of a 19-member mock fungal community sequenced using different spike-in concentrations of PhiX in technical replicate pairs, processed using three different pipelines.

3.2.5 The LotuS pipeline is performed best to identify members of the fungal mock community

To determine the most optimal pipeline for taxonomic assignment, each of the resulting datasets was filtered for the expected 19 fungal species in the mock community and averaged over technical replicates. The LotuS pipeline performed the best in identifying members of the fungal mock community, followed by mothur, then PIPITS. The LotuS pipeline performed best, correctly assigning an average of 6 organisms at a 25% PhiX spike-in and an average of 5.5 organisms at a 50% PhiX spike-in. The mothur pipeline correctly assigned an average of 5 organisms with a 25% PhiX spike-in and 4.5 at a 50% spike-in. PIPITS performed poorest: only 3 organisms were correctly assigned at a 25% spike-in concentration and 2 at a 50% PhiX spike-in concentration (**Figure 3.9a**). In all processing pipelines, the species *S. complicata* was identified at the highest proportion, matching the expected ratio designed for this mock community.

We investigated the presence of taxa not contained in the fungal mock community (false positive assignment), which may manifest from contamination, artifacts of sequencing, and bioinformatics processing. On average, the LotuS pipeline falsely assigned 13.3 taxa; mothur falsely assigned an average of 14.3 taxa, and PIPITS falsely assigned an average of 19.3 taxa (**Figure 3.9b**).

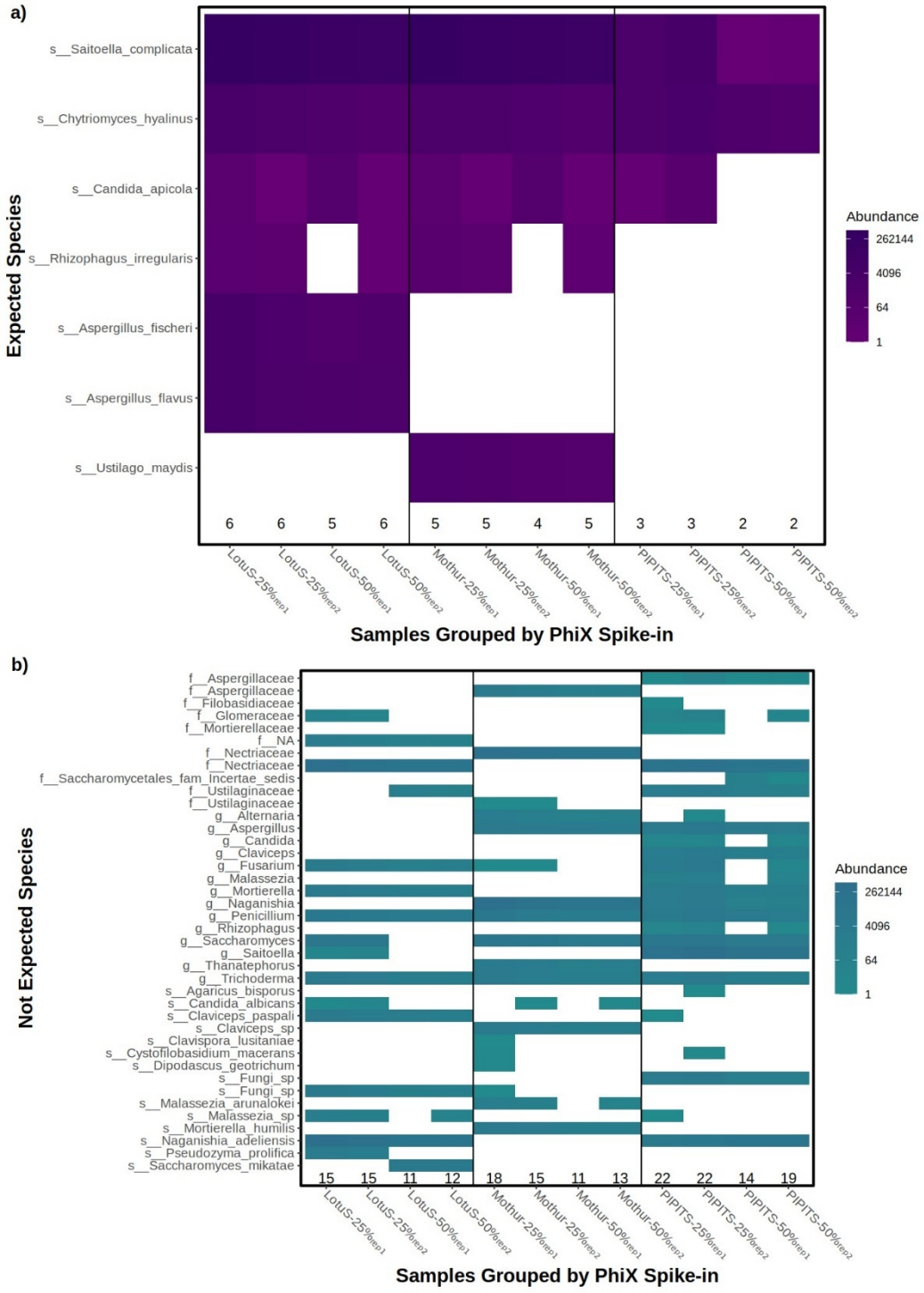


Figure 3.9. Heatmaps showing the abundances of taxa identified from a fungal mock community sample sequenced in replicate at 25% and 50% PhiX spike-in concentrations and processed using the LotuS, mothur, and PIPITS pipelines. **a)** Samples filtered for the expected fungal mock community members. **b)** Samples filtered for taxa not expected in the fungal mock community. The total number of taxa identified appears above its x-axis label. Abundances are transformed on a log4 scale.

3.2.6 Analysis of missing fungal mock community members

Due to all three processing pipelines not capturing all expected mock community members, we examined the raw unprocessed sequences to determine if the mock members were captured after sequencing and perhaps lost during processing by pipelines. Two methods were applied, the first merged paired-end reads and aligned against the expected mock reference sequences, while the second extracted the fungal ITS2 region for alignment. In both methods, all mock community members were identified from the raw sequences, confirming that, in theory, these community members were assignable and were lost during pipeline processing.

3.3 Discussion

Standardized methodologies to sequence and analyze the bacterial gut microbiome were established by the International Human Microbiota Consortium in 2015,[118] and although a similar endeavour has been pursued in fungal research, to date, no such equivalent standard exists for fungal analysis.[42,44,50] This shortcoming is partly attributable to the tough fungal chitinous cell wall that resists disruption by standard bacterial DNA extraction methodologies.[46,48] Furthermore, the gut fungal community occupies a small (~0.1%) fraction of the total intestinal microbiome, which is largely dominated by bacteria. This low proportion limits the number of sequences that can be practically generated to study the human gut mycobiome, which places extra importance on the standardization of sequencing and analysis methods in order to minimize technical variability.[63]

To assist in the effort to standardize the sequencing and analysis of gut fungal communities, we used a well-characterized mock fungal community consisting of 19 fungal species to study the effect of varying PhiX spike-in concentration and taxonomic assignment pipelines on the generated mock fungal community composition. We used PhiX spike-in values of 25% and 50% (recommendations range between 1% and 50%).[114] Two technical replicates were generated for each sample. All samples were analyzed by three popular fungal taxonomic assignment pipelines: LotuS, mothur, and PIPITS.

Once fungal community sequences are generated and quality controlled, they must be clustered for similarity and the clusters assigned to their corresponding taxa. This community profile generated by this process depends strongly on the quality and representativeness of the reference database used for taxonomic assignment.

In our pre-experimental BLAST assessment of the taxonomic assignment accuracy of the popular UNITE fungal ribosomal database, we utilized known sequences of our mock community. The purpose of this initial investigation is to examine if the UNITE database would hinder downstream detection of our sequenced community, such as if all 19 expected mock organisms can be identified within the UNITE database. We find that the UNITE database correctly assigned taxa to the species level for 18 out of the 19 expected fungal organisms in our mock community, with only *Mortierella verticillata* remaining unidentified (**Table 3.0**). We note that the comparison was not straightforward since many fungi possess synonymous scientific names as well as different names corresponding to their growth state; for example, we expected to find *Neosartorya fischeri* but instead identified *Aspergillus fischeri*, the anamorphic form of this organism.[113] Non-standardized nomenclature issues are noted within the mycological research community, and the curators of the UNITE database are actively attempting to steer fungal nomenclature toward a standardized format.[106,114,115]

We found that a sequencing spike-in of 50% PhiX provides little advantage over a 25% spike-in for assigning taxa while highly reducing the number of fungal sequences captured in a sequencing run. The 50% PhiX spike-in samples had a higher percentage of bases with Phred values over 30 for the entire 300 bp read length (86.1% vs. 70.8%). Overall, only a difference of 15.3% exists between 25% and 50% spike-in samples for the total number of bases called at an accuracy of ≥ 99.9% (**Figure 3.2**). However, the 50% PhiX samples yielded an average of 4.2-fold less processed reads relative to a 25% PhiX spike-in (**Table 3.1**). This difference in the quality of bases called over an accuracy of ≥ 99.9% is negligible compared to the higher yield of sequencing reads obtained by a 25% PhiX spike-in, but should nevertheless be a consideration tailored to the experimental design of a study.

We then compared the taxonomic community profiles generated by the various analysis pipelines assessed in this study. The alpha diversity of the communities generated by the pipelines did not significantly differ (**Figure 3.3b**). The effect of PhiX (25% or 50%) on alpha diversity was small (**Figure 3.3a**). We conclude that the PhiX concentrations tested do not affect overall community diversity.

A comparison of the fungal community profiles generated from our tested conditions reveals that the overall species composition and read abundances differed, depending on the taxonomic assignment pipeline used (**Figure 3.4**). The taxonomic profiles generated by a given pipeline were highly similar between technical replicates and PhiX spike-in concentrations across all pipelines. The LotuS and mothur pipelines generated the most similar community compositions, with the exception of *Naganishia adeliensis* identified in higher abundance within the LotuS pipeline, which mothur identifies as genus *Naganishia*.

A simple visual examination of the variability in the top-ten-most-abundant taxa identified from our tested conditions shows a similar pattern of high similarity between technical replicates and PhiX concentrations; variability in the community profiles generated by the different pipelines was more pronounced (**Figure 3.5**). The PIPITS pipeline was found to preferentially classify organisms at the genus level, whereas the LotuS pipeline generated the most species-level classifications, and mothur fell between the other pipelines in terms of genus and species identifications.

To quantify the differences in the fungal profiles generated from our tested conditions, we performed beta diversity analysis by nonmetric multidimensional scaling (NMDS) and significance testing by PERMANOVA (**Figure 3.7**). Significant differences were found in the

community profiles generated by the different taxonomic assignment pipelines: LotuS vs mothur ($q = 0.029$), LotuS vs PIPITS ($q = 0.028$), and PIPITS vs mothur ($q = 0.031$). These findings show that the choice of taxonomic assignment pipeline can significantly impact the results and interpretations generated from metabarcoding studies of fungal communities.

We used our fungal mock community to assess the taxonomic assignment accuracy of the different pipelines tested in this study. The LotuS pipeline accurately assigned the most organisms at a species taxonomic level, regardless of PhiX concentration, followed by mothur and PIPITS (**Figure 3.9a**). Overall, a maximum of six of 18 possible taxonomic assignments were correctly made by the LotuS pipeline.

We also assessed the false taxonomic assignment of the different pipelines tested in this study (**Figure 3.9b**). We found the PIPITS pipeline, on average, produces the most false assignments at the species taxonomic level, followed by mothur and then LotuS. In our pre-experimental BLAST assessment of this study, we also noticed that all queried mock organisms identified false taxonomic assignments when aligned directly against the UNITE database. The sources of these false assignments may arise from multiple factors such as the reagents and pipette tips used during the extraction and sequencing steps, sequencing artifacts, natural microbes in the laboratory environment, or possibly because of misassignment of taxonomic identities by the processing pipelines.[119,120]

Further, we investigated the raw sequences to examine if the uncaptured mock community members were present in the sequence data and lost during processing through the pipelines. We confirmed that the all-uncaptured mock sequences exist within the raw unprocessed data. A possible explanation as to why these taxa were not identified by the

processing pipelines may be due to the stringency of the filtering parameters set in order to reduce the amount of false positive assignments, as we have noticed in our pre-experimental BLAST assessment that without filtering, many false assignments arise.[42,44,45] Our experimental objective is not to optimize each pipeline, but rather to assess their identification abilities when using the default parameters where possible. We also note that these pipelines consist of multiple bioinformatics tools and not all parameters were adjustable.

Nevertheless, our work here using a fungal mock community to examine the effects of varying PhiX concentrations on sequencing, and the impact of three different fungal metataxonomic processing pipelines (LotuS, mothur, and PIPITS) on characterizing these mock identities, suggests that a concentration of 25% PhiX offers balance to retain a good quantity and quality of sequencing reads, and the LotuS pipeline provides adequate resolution to discern fungal identities. The current study offers a guide towards standardizing methodology for sequencing and identifying environmental fungal communities, such as studies seeking to investigate if an association exists between the mycobiota and disease.

CHAPTER 4: INVESTIGATION OF THE GUT FUNGAL COMMUNITY IN A PEDIATRIC-ONSET MULTIPLE SCLEROSIS COHORT

4.1 Introduction

Multiple Sclerosis (MS) is a complex, multifactorial autoimmune disease of the central nervous system (CNS) leading to the degeneration of neuronal axons, whereby demyelination occurs, and disruption of essential cellular functions ensues.[2] An estimated 2.8 million individuals worldwide suffer from this disease, with outcomes ranging from minimal neurological dysfunction to bodily disability and even death.[3,8] This disease also has a severe psychological impact: individuals diagnosed with MS often cope with depression, and caregivers living among MS individuals report a reduced quality of life.[121,122]

Although medical interventions exist to alleviate MS-associated inflammation, no treatment currently exists to eliminate demyelinating episodes, which results in progressive neuronal deterioration.[2] Many investigations have been conducted to examine the factors that contribute to MS, including genetics, vitamin D levels, environmental factors, and viral infections.[2] Emerging evidence suggests that the gut microbiota may play a role in MS inflammation through mechanisms involving gut-brain-axis signalling and disrupting tight junctions of the blood-gut and blood-brain barriers.[72,123] Bacterial gut microbiome studies have associated various bacterial taxa with beneficial or detrimental metabolic changes, which may influence MS pathology.[73,124,125] The fungal component of the gut microbiota is less well-studied, although evidence of a fungal etiology in MS exists.

For example, a study conducted by Ramos et al. (2008) found the presence of fungi in the blood of individuals diagnosed with MS by immunological assay utilizing anti-*Candida sp*.

antibodies, whereas non-MS individuals tested negative with this assay.[23] In a patient-limited longitudinal study by Pisa et al. (2011), the abundance of fungi in the blood of an MS-diagnosed patient was noted to vary temporally.[24] Additional experiments by Pisa et al. (2013) utilizing molecular and immunological assays indicated the presence of fungi in the cerebrospinal fluid of MS individuals.[26] Alonso et al. (2018) found molecular signatures for fungi in the brain tissues of post-mortem MS individuals.[27] In an experimental autoimmune encephalomyelitis study in mice, a commonly used experimental model for MS, an inoculation of *Candida albicans* exacerbated symptoms, and tissue analysis by histopathology of biopsy samples detected fungal invasion of the mouse CNS.[33] There is also evidence of a fungal association with MS by dimethyl fumarate, a disease-modifying drug that was originally used to prevent mold growth on leather products but is now approved for the treatment of psoriasis and MS.[37–40]

Despite the evidence of a fungal role in MS pathogenesis, investigations into the role they may play are lacking; to date, only two studies have reported on the association of the gut mycobiome and MS.[28,29] Both studies focused on the role of the gut mycobiome in adults. No studies to date have investigated the role of the gut mycobiome in pediatric-onset MS (POMS) individuals. Investigating the causes of POMS is important because the disease course differs from adult MS. There is increased inflammation and axonal damage in children compared to adults, which can lead to significant impairment of cognition and/or disability at a young age.[10] A second reason to examine the gut fungi in POMS is that the diversity of the gut fungal community has also been observed to change at different stages of life, suggesting that the fungal community composition of POMS may differ from adult MS and thus may have a different impact on MS.[126]

In this study, we sought to investigate the gut fungal profiles in a cohort of POMS individuals and compare their gut fungal profiles to non-MS pediatric individuals (unaffected controls) and individuals with monophasic acquired demyelinating syndrome (monoADS). Unlike the progressive demyelination of neurons in MS, monoADS individuals present a single demyelinating episode but are not formally classified as MS. Comparing the gut fungal profiles of these similar but distinct diseases may help to reveal common and unique elements of the community profile that may associate with MS.[13] We hypothesize that the gut fungal profile of our MS cohort will differ from that of monoADS or unaffected control individuals. Our investigation used metabarcoding sequencing of the fungal internal transcribed spacer 2 (ITS2) to characterize and compare the gut mycobiome between our study participants and explore whether a fungal association to POMS exists.

4.2 Results

4.2.1 Cohort filtering and determining cohort characteristics

Stool samples from 91 of the 119 individuals collected in this study passed sequencing size selection (300–1000 bp), quality filtering, and contained associated cohort metadata suitable for analysis. Following LULU curation and determination of a Good's coverage, 45 individuals with a Good's coverage index of ≤ 96.5% were removed to control for samples with low sequencing coverage. The final cohort of 46 individuals stratified by diagnoses consisted of 18 POMS, 13 monoADS, and 15 unaffected controls. The mean age of MS onset was 15.3 ± 2.8 years, while monoADS individuals, on average, experienced a demyelination episode at 5.2 ± 3.2 years of age. A total of 14 of the 18 POMS individuals were on a single disease-modifying therapy at the time of stool sample collection: 3 glatiramer acetate, 4 interferon beta-1a, 2 pegylated interferon beta, 1 natalizumab, 1 rituximab, and 2 dimethyl fumarate. A breakdown by cohort characteristics and participant sex is summarized in **Table 4.0**. Although insufficient for statistical analysis, a couple of our study participants used dimethyl fumarate at the time of stool sampling collection; we checked and noticed no visible trends in our beta diversity analysis.

Table 4.0. Characteristics of the POMS, monoADS, and unaffected control participants.

Characteristics	POMS[a] n = 18	MonoADS[b] n = 13	Unaffected Control n = 15
Sex, female: n (%)	15 (83%)	6 (46%)	8 (53%)
Sex, male: n (%)	3 (17%)	7 (54%)	7 (47%)
Age at symptom onset, years: mean (SD)	15.3 (SD: 2.8)	5.2 (SD: 3.2)	\
Age at stool sample collection, years: mean (SD)	18.1 (SD: 2.4)	12.1 (SD: 3.6)	15.3 (SD: 3.4)
Disease-modifying drug exposure status at stool sample collection (ever/naïve): No. (%)	14 (78%) / 4 (22%)	\	\
Ever Avonex - IFNB-1a (SC)	4	\	\
Ever Dimethyl fumarate	2	\	\
Ever Glatiramer Acetate	3	\	\
Ever Natalizumab	1	\	\
Ever PEG-IFNB	2	\	\
Ever Rebif - IFNB-1a (SC)	1	\	\
Ever Rituximab	1	\	\

[a] POMS = pediatric-onset multiple sclerosis, [b] monoADS = monophasic acquired demyelinating syndrome

4.2.2 Fungal ITS2 amplicon sequence data processing

Illumina sequencing generated an average of 30,791,070 paired-end reads across the 46 study individuals, while the mock community yielded an average of 10,168,535 paired-end reads across technical replicates. After sequence processing, quality assessment, and clustering for high-quality fungal reads using the LotuS pipeline, 10,525,102 reads remained across 46 study individuals and 2,175,124 reads across mock community replicates. After LULU curation and Good's analysis filtering, sequences belonging to one OTU were found to be non-fungal associated and removed, leaving 10,500,725 reads remaining for the 46 individuals. Visual inspection of the density distribution for each participant sample showed non-normal distribution, which was confirmed by Q-Q plot (**Figure 4.0A & B**). Following central log-ratio (CLR) transformation, the distribution shape for each participant sample changed (**Figure 4.1A & B**). A check of CLR-transformed values using a Q-Q plot showed that the normality assumption is not met in a majority of the sequenced samples, as the quantile points do not perfectly fit the diagonal line (**Figure 4.2**). Due to the non-normal distribution of the data, nonparametric methods were used for the statistical analysis. A summary of the number of reads reflecting the 46 study individuals is provided in **Table 4.1**.

Out of 19 expected fungal organisms of the mock community, only 6 were identified at the species level: *Saitoella complicata* (also known as *S. complicate*), *Aspergillus fischeri* (also known as *Neosartorya fischeri*), *Aspergillus flavus*, *Chytriomyces hyalinus*, *Rhizophagus irregularis*, and *Candida apicola* (**Figure 4.3A**). The order of our relative assignment of the mock taxa is in agreement with the expected values (**Figure 4.3B)**, with the exception of *C. apicola*, which swapped order with *R. irregularis*.[74] This is due to the second replicate containing an inordinately low (one) count of *C. apicola* (as compared to 32 for the other

replicate). A perfect match to the expected ratios was not anticipated due to the original ratios being determined by sequencing the 18S rRNA, rather than the ITS2 conducted in our present study.

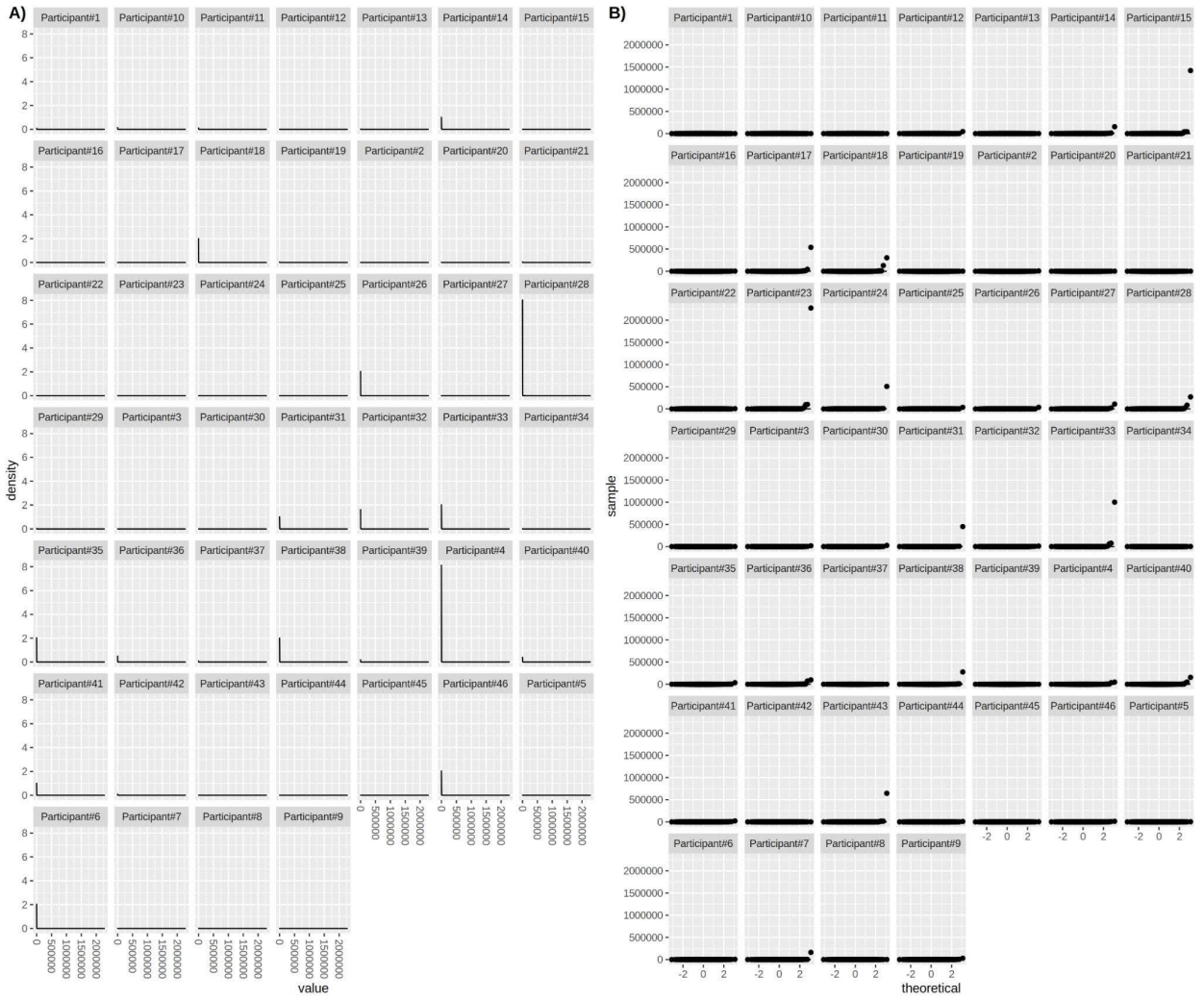


Figure 4.0. Density distributions and quantile-quantile plots of untransformed relative abundances of OTUs within each participant's stool. The sequencing reads from a 46-member cohort composed of pediatric-onset MS (POMS), monoADS, and unaffected control individuals are examined for normality. **A)** Density distributions for each study individual. **B)** Quantile-quantile plots for each study individual.

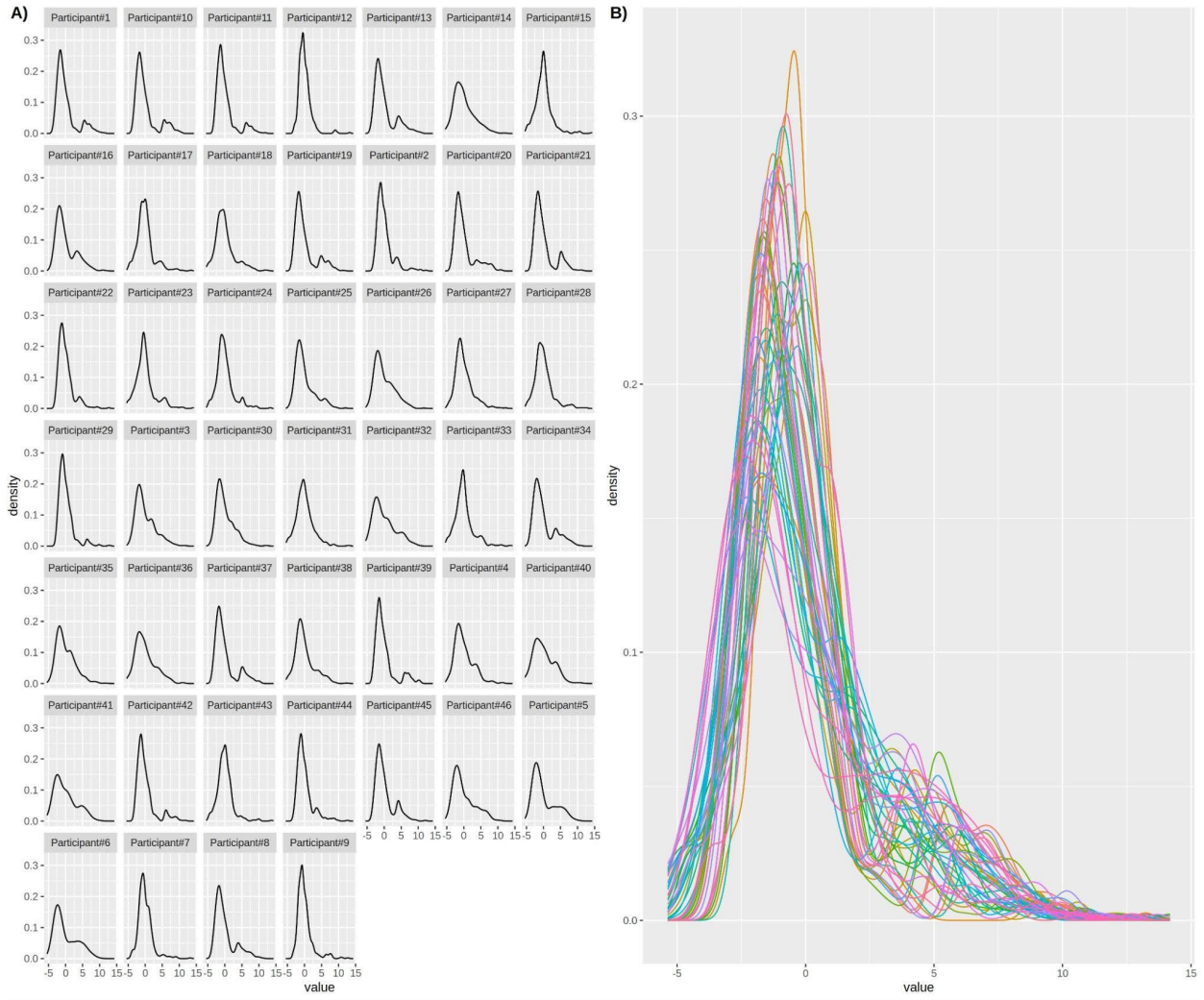


Figure 4.1. Density distributions of CLR-transformed relative abundances of OTUs within each participant's stool. The sequencing reads from a 46-member cohort composed of pediatric-onset MS (POMS), monoADS, and unaffected control individuals were transformed using the CLR method and examined for normality. In the left figure **A)** Density distributions for each study individual; **B)** All 46 density distributions from **A** combined into a single plot.

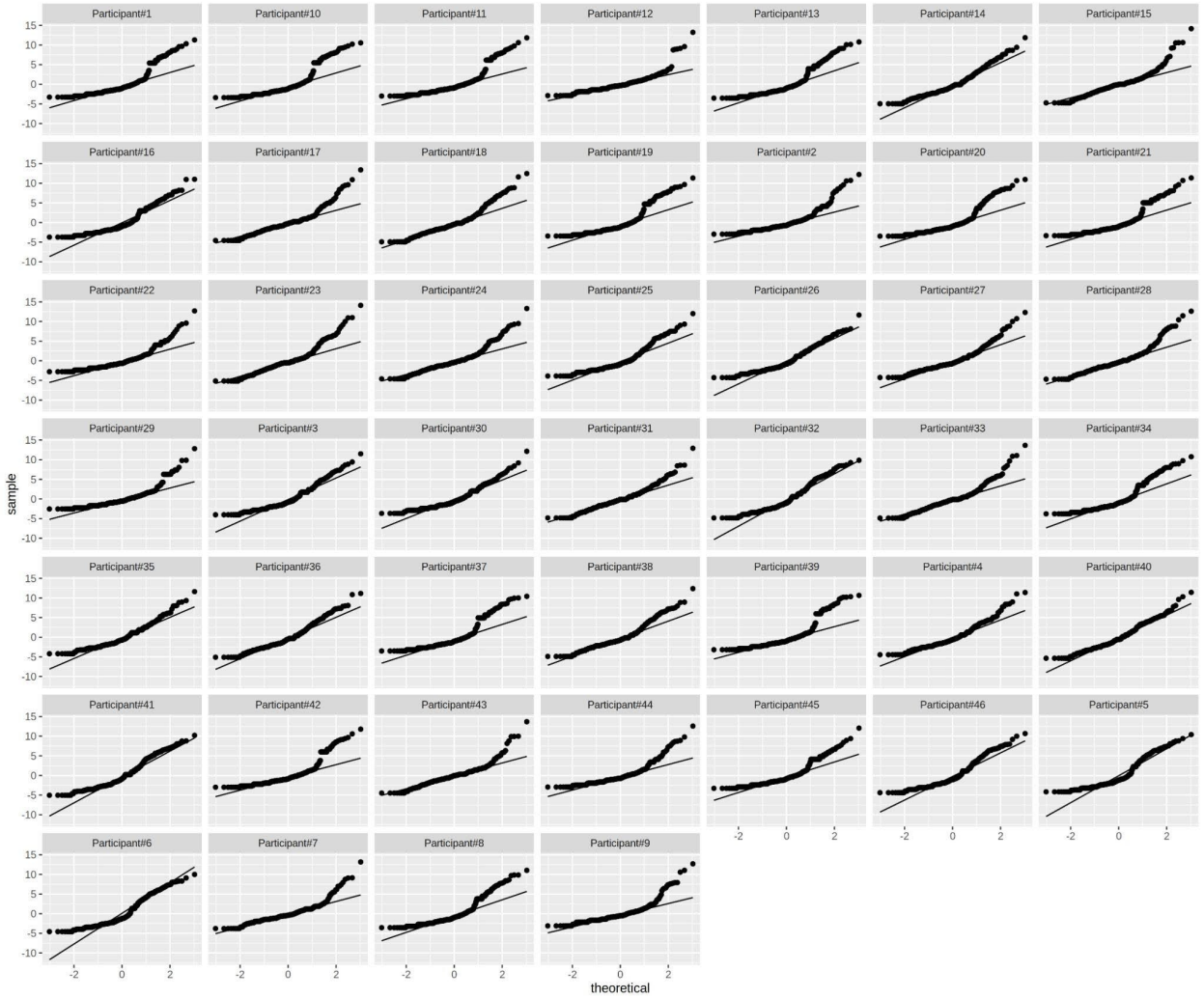


Figure 4.2. Quantile-quantile plots of CLR transformed relative abundances. The sequencing reads from a 46-member cohort composed of pediatric-onset MS (POMS), monoADS, and unaffected control individuals were transformed using the CLR method and examined for normality.

Table 4.1. Number of reads for the study participants (POMS, monoADS, and unaffected controls) following sequencing and bioinformatics processing (n = 46).

	Raw Sequencing (Averaged)	LotuS	LULU[a]
Total reads	30,791,070	10,525,102	10,500,725
Min	1644	563	562[b]
Medium	11,3768	37,945	37,907
Mean	669,371	228,807	228,277
Max	4,854,072	2,566,757	2,560,308

[a] Includes filtering in samples > 500 reads (according to Good's coverage ≥ 96.5%).
[b] The resulting minimum is lower for LULU because a single OTU found in the previous steps was identified to be of non-fungal origin and was removed.

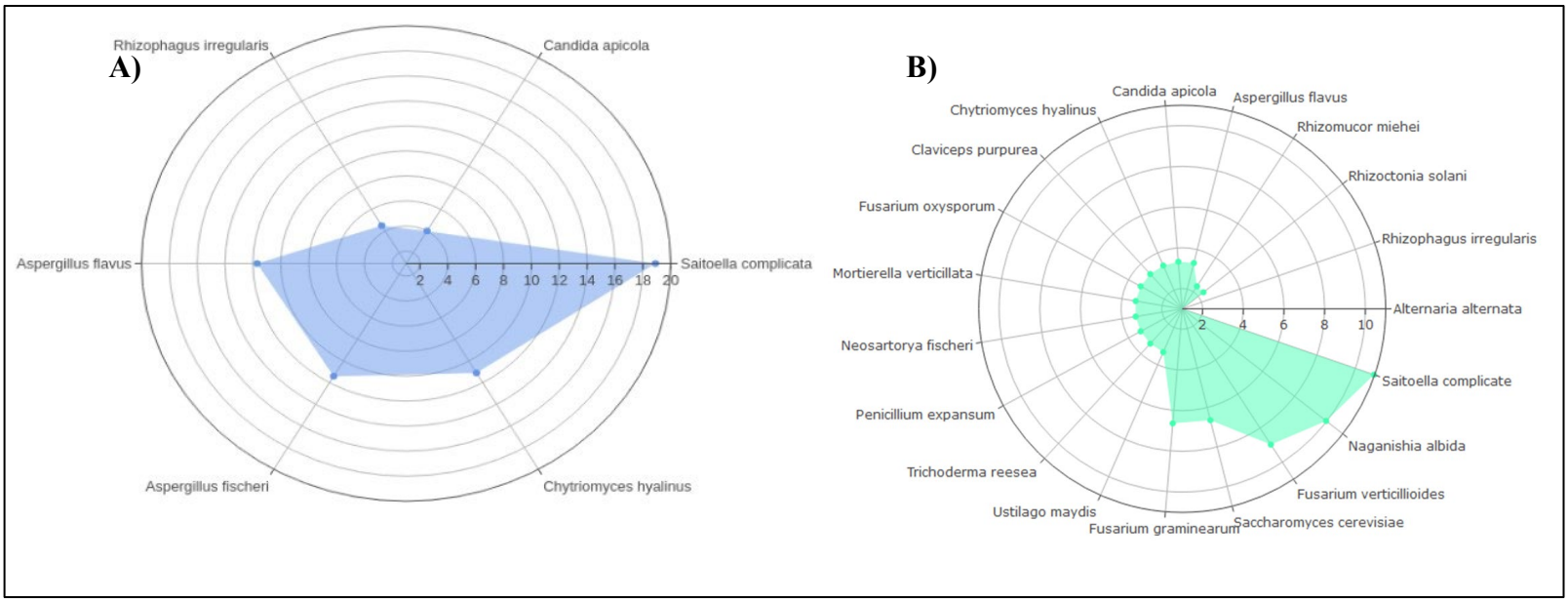


Figure 4.3. Relative ratios of a 19-member mock fungal community. Ratios are presented in a log2 scale. In **A)** fungal ITS2 ratios from fungi identified in this study, averaged between technical replicates of 6 detected mock organisms; **B)** expected 18S rRNA ratios of all 19 organisms.[74]

4.2.3 Abundances of fungal taxa

The *Ascomycota* phylum dominated across all three groups, representing 93.2% of the relative abundance for the POMS individuals, 83.7% for monoADS, and 95.4% for the unaffected controls (**Figure 4.4**). While the *Basidiomycota* phylum appeared higher for the POMS individuals at 6.62% compared to monoADS (0.662%) and the unaffected controls (0.465%). In contrast, *Mucormycota* appeared lower in abundance for POMS individuals (0.0659%) compared to monoADS (1.38%) and unaffected controls (3.80%).

The top ten genera for the POMS, monoADS, and unaffected control groups are presented in **Figure 4.5**. The relative abundances of observed taxa are similar across the three groups, with *Saccharomyces* detected at the highest proportion (48.8% POMS, 56.2% monoADS, 42.0% unaffected control). Another six genera were shared between all three groups: *Aspergillus, Candida, Cladosporium, Mucor, Penicillium*, and an unidentified genus. The *Candida* genus was generally the second-highest in terms of relative abundance (37.96% POMS, 9.25% monoADS, 43.1% unaffected control). Of the top-ten-most abundant genera shared between two or more groups, the genus *Cyberlindnera* was shared between POMS and unaffected individuals, the genus *Malassezia* was shared between POMS and monoADS individuals, and an unclassified genus from the order Eurotiales was shared between monoADS and unaffected individuals. The genera *Aspergillus, Candida, Cladosporium, Mucor, Penicillium, Saccharomyces,* and an unclassified genus were found in all groups. Within the top ten genera, *Agaricus* was found to be uniquely abundant in the POMS group, *Phoma* in monoADS, and *Exophiala* in unaffected controls. We found that 99.9% of Agaricus species in POMS were composed of *Agaricus bisporus*, with the remaining proportion belonging to unclassified *Agaricus spp.* and *Agaricus kriegeri*. Although the genus *Agaricus* was not within

the top-ten-most abundant taxa in monoADS and unaffected controls, the species *A. bisporus* constituted 100.0% of this genus in monoADS and 89.8% in unaffected control participants.

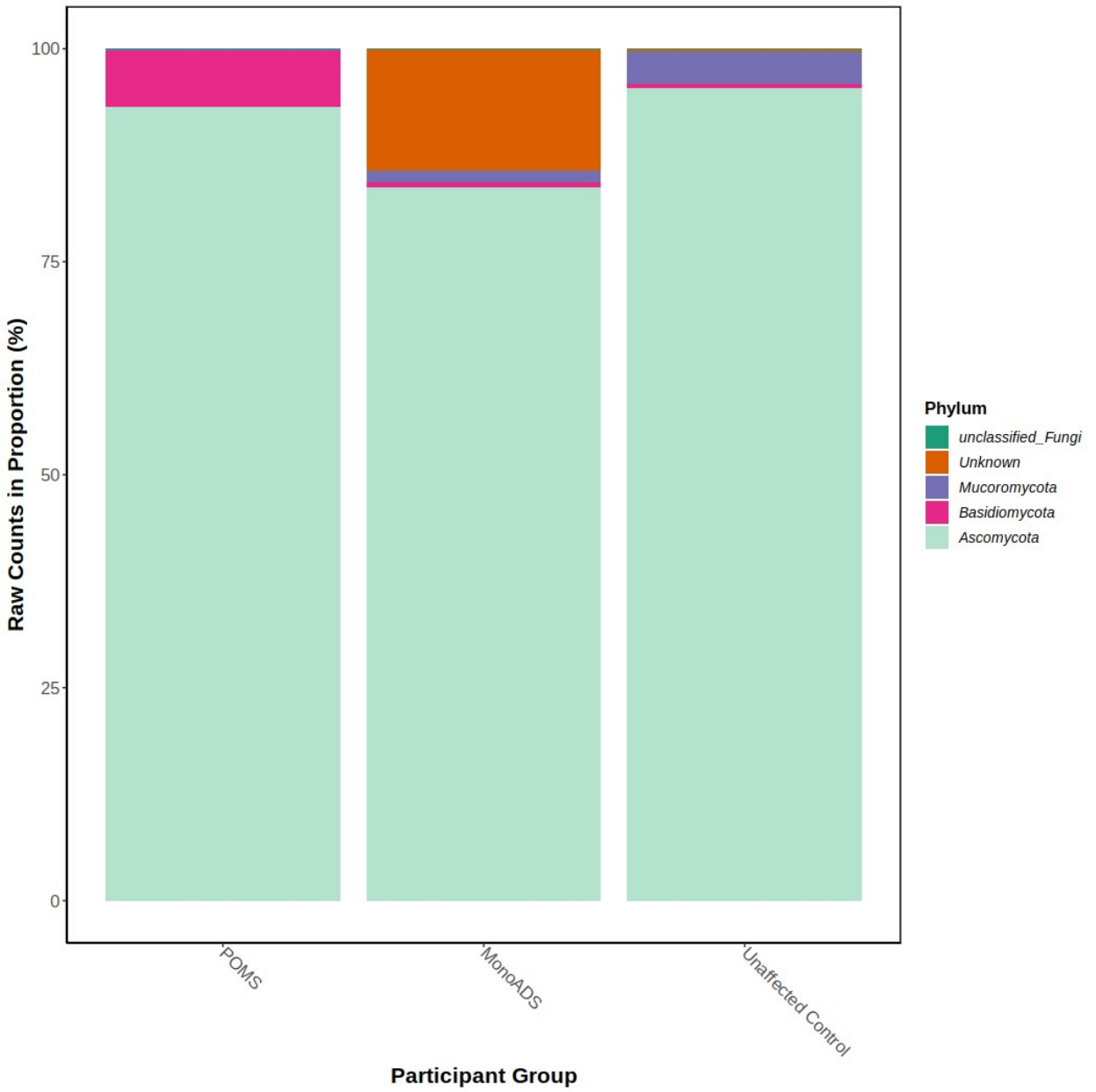


Figure 4.4. Stacked bar plot showing the proportion of fungal phyla found within pediatric onset-MS (POMS), monophasic acquired demyelinating syndrome (monoADS), and unaffected control participants. Phyla are plotted from largest proportion (bottom) to smallest (top) relative to POMS, and the order is maintained for all disease categories.

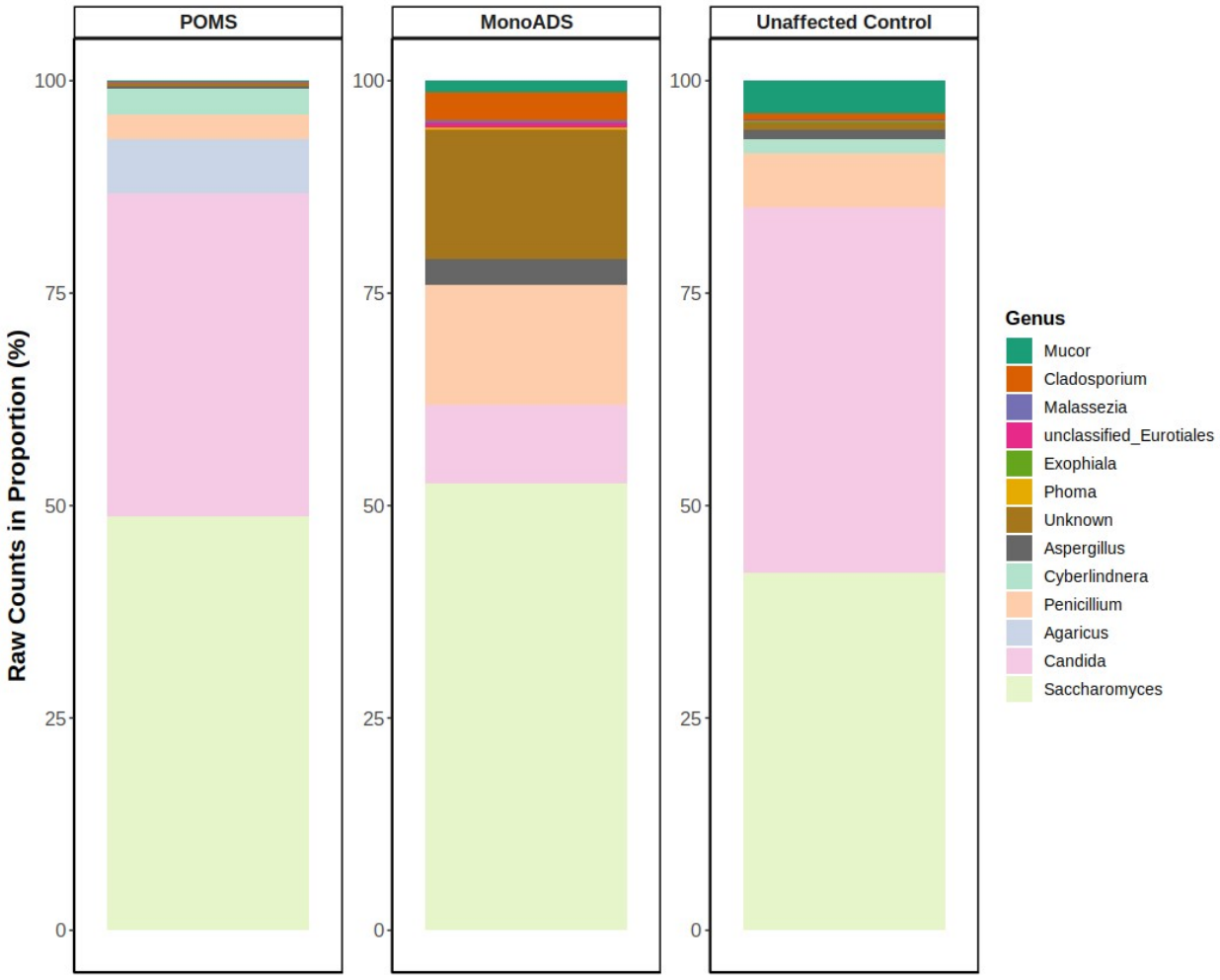


Figure 4.5. Top-ten-most abundant fungal genera for pediatric-onset MS (POMS), monophasic acquired demyelinating syndrome (monoADS), and unaffected control participants. Phyla are plotted from largest proportion (bottom) to smallest (top) relative to POMS.

4.2.4 Alpha and beta diversity of stratified diagnoses

The alpha diversity results appear in **Figure 4.6**. MonoADS participants had the lowest species richness, followed by POMS and unaffected controls (**Figure 4.6A**). However, none of the three groups were significantly different by the Kruskal-Wallis ANOVA test ($p = 0.077$, and adjusted post hoc values are shown in the figure). While there was some suggestion of a difference in diversity between the three groups (**Figure 4.6B & C**), with the POMS participants potentially having a lower alpha diversity versus those with monoADS or unaffected controls (based on Kruskal-Wallis ANOVA of the participants Shannon and inverse Simpson indices, $p = 0.041$ and 0.034, respectively), no comparisons reached significance after post hoc testing ($q = 0.057$ and 0.057, respectively), for POMS vs. monoADS or unaffected in Shannon index, and ($q = 0.055$ and 0.055, respectively, for the same comparisons with the inverse Simpson index).

We also stratified our participants by sex or age group (**Figure 4.7A)**. We found significant differences in overall fungal richness between males and females ($q = 0.035$), as well fungal diversity significantly differed between sexes when measured with the inverse Simpson index ($q = 0.037$). Significant differences in fungal richness were also seen between children and youth ($q = 0.018$ in **Figure 4.7B**), and similarly, the fungal diversity also differed when measured using the inverse Simpson index ($q = 0.033$).

For beta-diversity, while no obvious or distinct clustering patterns could be visually observed between the MS, monoADS, and unaffected participants (**Figure 4.8**), quantification by PERMANOVA showed a significant difference between the three groups ($q = 0.019$). Specifically, pairwise comparisons indicated a significant difference between the monoADS and POMS ($q = 0.005$), but not between the unaffected and POMS ($q = 0.08$) or monoADS ($q = 0.274$).

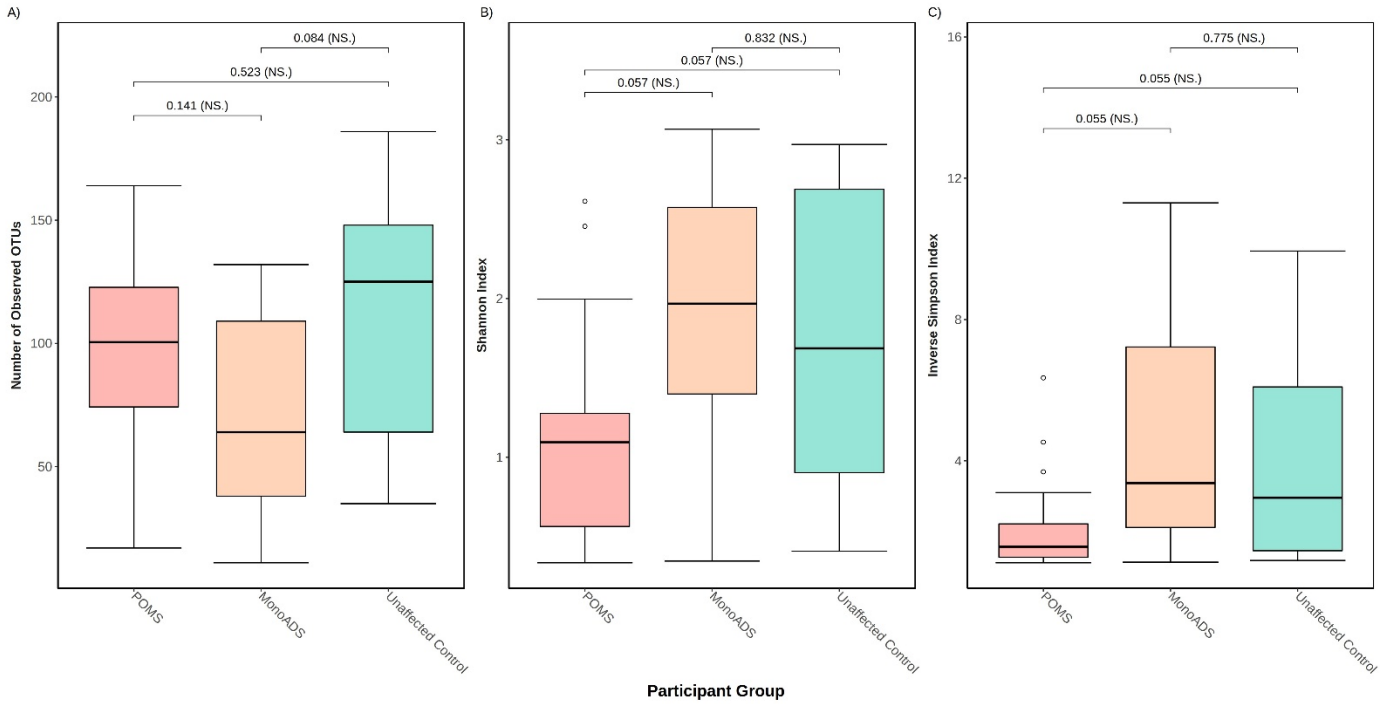


Figure 4.6. Alpha diversity metrics for the pediatric-onset MS (POMS), monophasic acquired demyelinating syndrome (monoADS), and unaffected control participants.

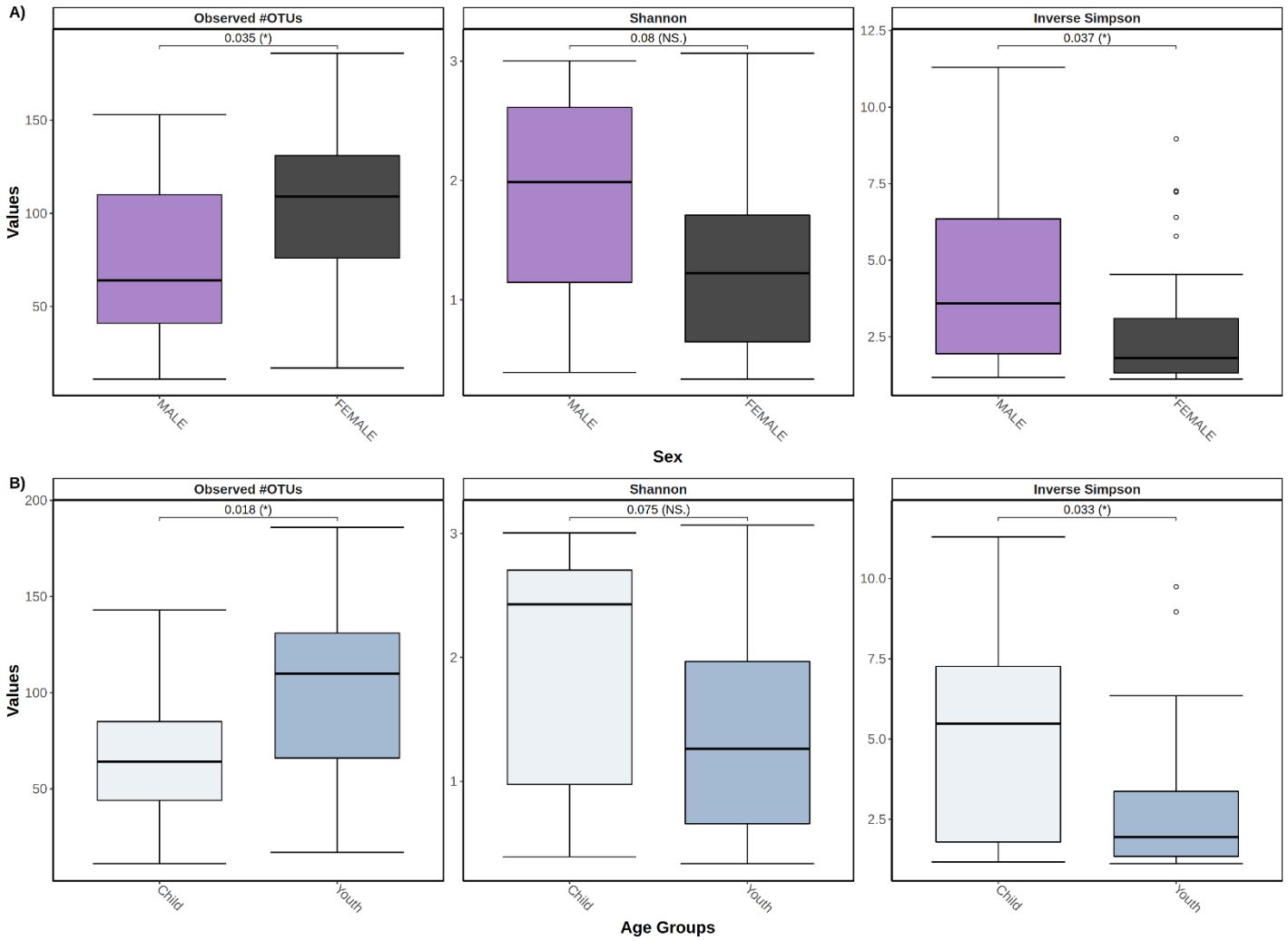


Figure 4.7. Three different alpha diversity metrics were calculated for 46 study participants stratified by either sex or age groups. **A)** Total study cohort stratified by sex (n = 29 females and n = 17 males) and is measured for the number of observed OTUs that represent unique species. **B)** shows a stratification by age groups of child (≤ 14 years old, n = 13) and youth (between 15 and 24 years old, n = 33).

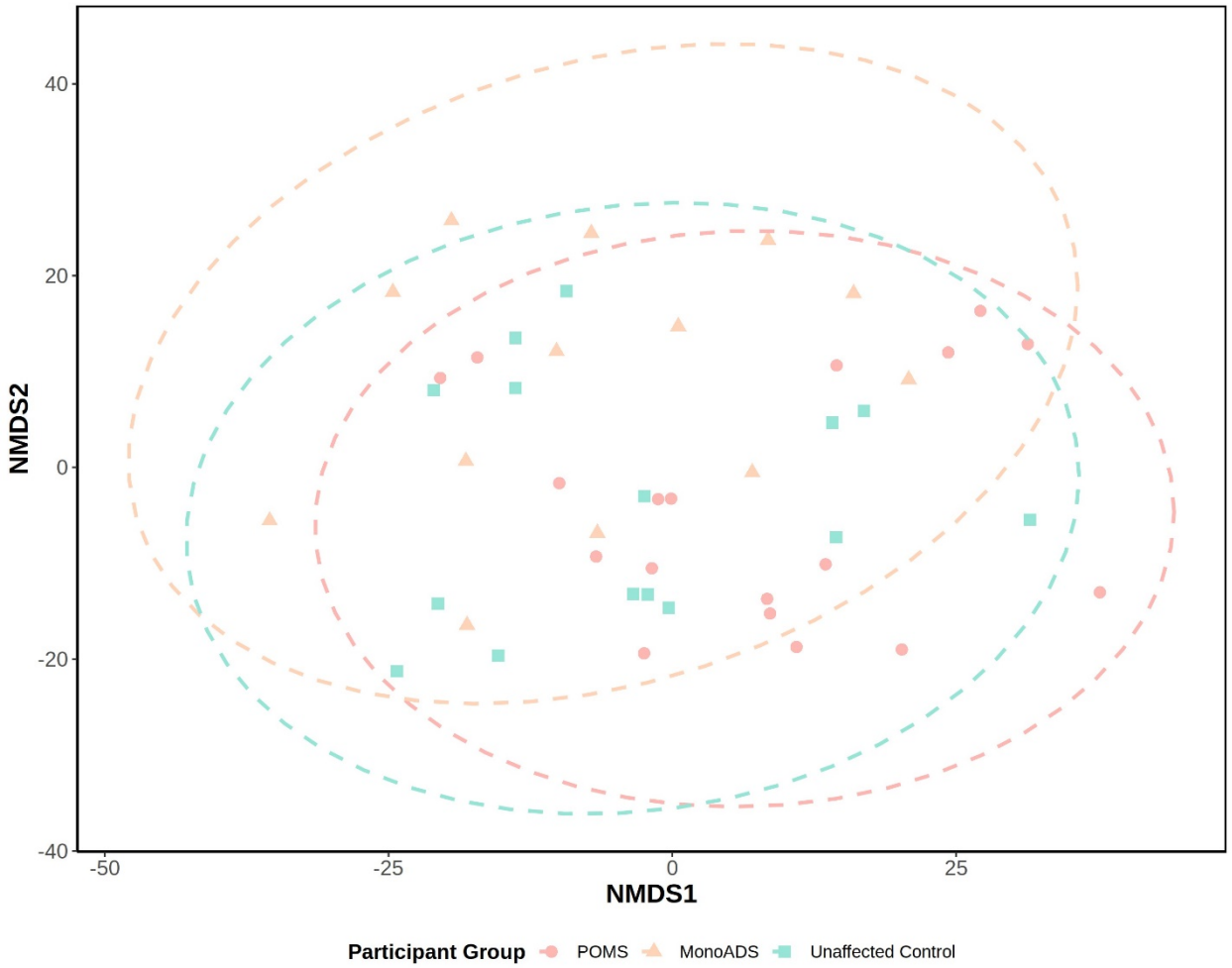


Figure 4.8. Beta diversity analysis by non-metric multidimensional scaling for the pediatric-onset MS (POMS), monophasic acquired demyelinating syndrome (monoADS), and unaffected control participants. Calculations are based upon CLR-transformed counts. Ellipses represent 95% confidence intervals. The Euclidean distance is used for calculations.

4.2.5 Differential abundance analysis

Analysis of taxonomic abundances between the groups yielded no significant differences by ALDEx2 analysis. However, LEfSe analysis predicted marker organisms (**Figure 4.9**). The species *Candida albicans, Cyberlindnera jadinii, and Fusarium poae* were predicted as markers of POMS participants. *Acremonium fusidioides, Cyberlindnera rhodanensis, Exophiala lecanii-corni, Talaromyces islandicus,* and an unclassified fungal species were predicted to be markers of monoADS by LEfSe. In unaffected individuals, *Penicillium aurantiogriseum* and *P. solitum* were predicted as markers. BLAST analysis of species-level taxa identified by LEfSe using the RefSeq ITS database yielded multiple unclassified identities (**Table 4.2**).

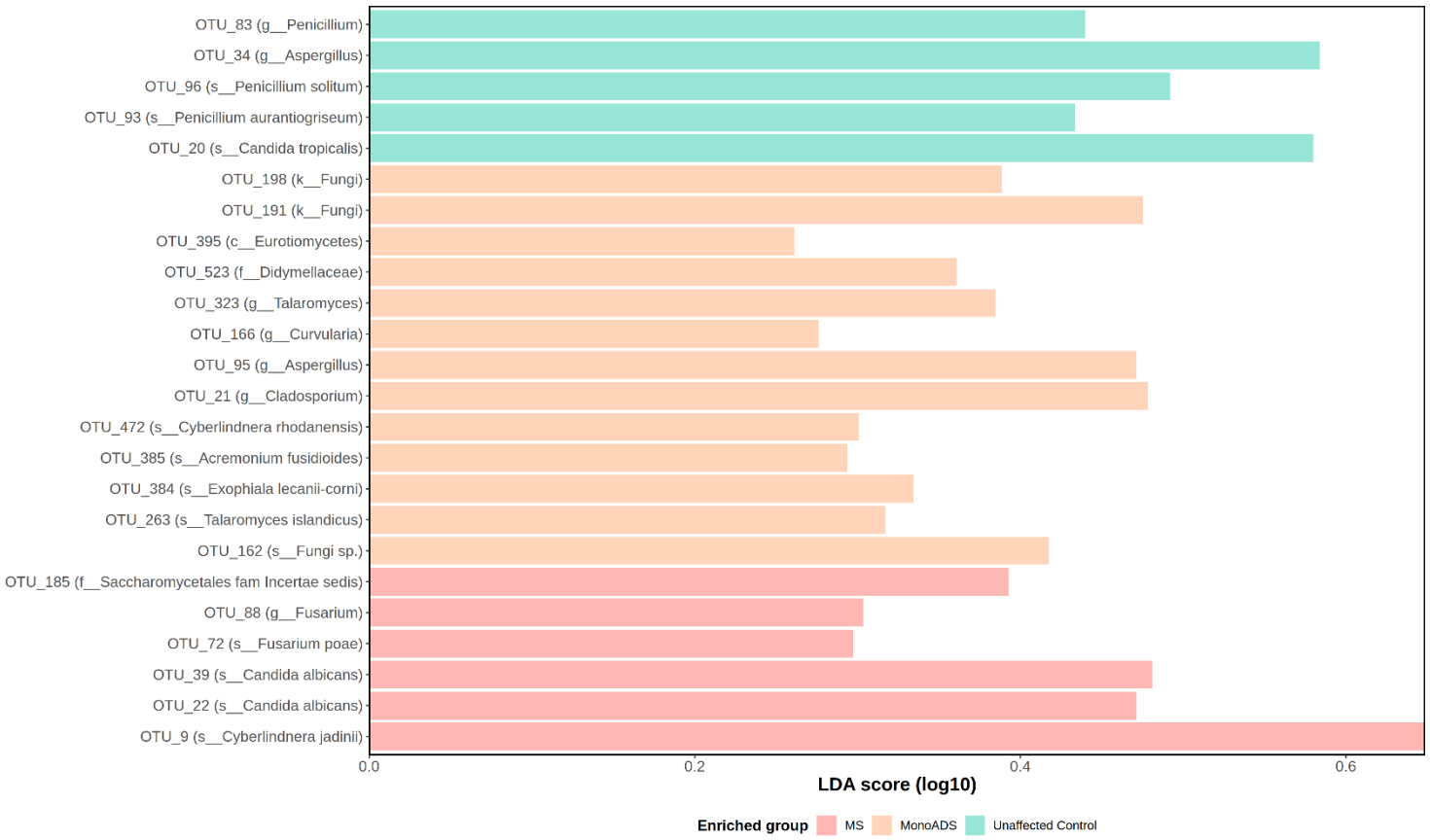


Figure 4.9. Operational Taxonomic Unit (OTU) differential abundance analysis identifying predictive makers by Linear discriminant analysis Effect Size (LEfSe) for pediatric-onset MS (POMS), monophasic acquired demyelinating syndrome (monoADS), and unaffected control participants. Effect sizes are converted into a score and plotted on a log10 scale.

Table 4.2. BLAST identities of organisms predicted above a species taxonomic level from LefSe using the RefSeq ITS database.

OTU	UNITE Taxon ID	BLAST Taxon ID	BLAST Accession	%Identity	QCovHSP[a]	Expected Value
OTU 166	g__Curvularia	*Curvularia alcornii MFLUCC 10-0703*	NR_137091	100	100	1.67E-81
OTU 185	f__Saccharomycetales fam Incertae sedis	*Candida africana*	NR_138276	91.333	100	9.60E-54
OTU 185	f__Saccharomycetales fam Incertae sedis	*Candida albicans*	NR_125332	91.333	100	9.60E-54
OTU 191	k__Fungi	*Cladosporium angustiherbarum*	NR_152286	97.297	100	1.20E-67
OTU 191	k__Fungi	*Cladosporium aggregatocicatricatum*	NR_152300	97.297	100	1.20E-67
OTU 191	k__Fungi	*Cladosporium rhusicola*	NR_152299	97.297	100	1.20E-67
OTU 191	k__Fungi	*Cladosporium puyae*	NR_152298	97.297	100	1.20E-67
OTU 191	k__Fungi	*Cladosporium versiforme*	NR_152297	97.297	100	1.20E-67
OTU 191	k__Fungi	*Cladosporium subcinereum UTHSC DI-13-257*	NR_148193	97.297	100	1.20E-67
OTU 191	k__Fungi	*Cladosporium antarcticum*	NR_121332	97.297	100	1.20E-67
OTU 191	k__Fungi	*Cladosporium phlei*	NR_120013	97.297	100	1.20E-67
OTU 191	k__Fungi	*Cladosporium cucumerinum MUCL 10092*	NR_119841	97.297	100	1.20E-67
OTU 191	k__Fungi	*Cladosporium chubutense*	NR_119728	97.297	100	1.20E-67
OTU 191	k__Fungi	*Cladosporium variabile*	NR_119663	97.297	100	1.20E-67
OTU 191	k__Fungi	*Cladosporium tenellum*	NR_119662	97.297	100	1.20E-67
OTU 191	k__Fungi	*Cladosporium spinulosum*	NR_119660	97.297	100	1.20E-67
OTU 191	k__Fungi	*Cladosporium macrocarpum*	NR_119657	97.297	100	1.20E-67
OTU 191	k__Fungi	*Cladosporium herbarum*	NR_119656	97.297	100	1.20E-67
OTU 191	k__Fungi	*Cladosporium allicinum*	NR_152266	97.297	100	1.20E-67

[a] = QcovHSP stands for query coverage per high scoring pair

Table 4.2. BLAST identities of organisms predicted above a species taxonomic level from LefSe using the RefSeq ITS database (continued #1).

OTU	UNITE Taxon ID	BLAST Taxon ID	BLAST Accession	%Identity	QCovHSP[a]	Expected Value
OTU 191	k__Fungi	*Cladosporium herbaroides*	NR_119655	97.297	100	1.20E-67
OTU 191	k__Fungi	*Cladosporium iridis*	NR_111271	97.297	100	1.20E-67
OTU 198	k__Fungi	*Clavispora lusitaniae*	NR_130677	100	100	5.26E-33
OTU 21	g__Cladosporium	*Cladosporium chasmanthicola CPC 21300*	NR_152307	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium pseudochalastosporoides*	NR_152296	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium welwitschiicola CPC 18648*	NR_152308	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium parapenidielloides*	NR_152293	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium longicatenatum*	NR_152291	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium ipereniae*	NR_152290	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium austroafricanum*	NR_152288	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium angustiterminale*	NR_152287	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium crousii UTHSC DI-13-247*	NR_148192	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium phaenocomae*	NR_119950	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium varians CPC 13658*	NR_119856	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium subuliforme CPC 13735*	NR_119854	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium exile CPC 11828*	NR_111532	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium globisporum*	NR_111534	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium myrtacearum CPC 14567*	NR_119849	100	100	2.56E-74

[a] = QcovHSP stands for query coverage per high scoring pair

Table 4.2. BLAST identities of organisms predicted above a species taxonomic level from LefSe using the RefSeq ITS database (continued #2).

OTU	UNITE Taxon ID	BLAST Taxon ID	BLAST Accession	%Identity	QCovHSP[a]	Expected Value
OTU 21	g__Cladosporium	*Cladosporium angustisporum CPC 12437*	NR_111530	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium iranicum CPC 11554*	NR_111536	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium funiculosum*	NR_119845	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium flabelliforme CPC 14523*	NR_119844	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium exasperatum CPC 14638*	NR_119843	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium montecillanum*	NR_152292	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium gamsianum CPC 11807*	NR_111533	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium australiense CPC 13226*	NR_119837	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium asperulatum CPC 14040*	NR_119836	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium pini-ponderosae*	NR_119730	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium colombiae*	NR_119729	100	100	2.56E-74
OTU 21	g__Cladosporium	*Cladosporium silenes*	NR_111270	100	100	2.56E-74
OTU 323	g__Talaromyces	*Talaromyces wortmannii*	NR_172039	99.419	100	6.44E-86
OTU 34	g__Aspergillus	*Aspergillus tubingensis NRRL 4875*	NR_131293	100	100	2.88E-84
OTU 34	g__Aspergillus	*Aspergillus costaricensis*	NR_103604	100	100	2.88E-84
OTU 34	g__Aspergillus	*Aspergillus luchuensis KACC 46772*	NR_135449	100	100	2.88E-84
OTU 395	c__Eurotiomycetes	*NA*	NA	NA	NA	NA

[a] = QcovHSP stands for query coverage per high scoring pair

Table 4.2. BLAST identities of organisms predicted above a species taxonomic level from LefSe using the RefSeq ITS database (continued #3).

OTU	UNITE Taxon ID	BLAST Taxon ID	BLAST Accession	%Identity	QCovHSP[a]	Expected Value
OTU 523	f__Didymellaceae	*Dimorphoma saxea*	NR_157421	97.368	100	7.38E-70
OTU 83	g__Penicillium	*Penicillium neoechinulatum*	NR_163549	100	100	1.31E-82
OTU 83	g__Penicillium	*Penicillium hirsutum*	NR_163544	100	100	1.31E-82
OTU 83	g__Penicillium	*Penicillium robsamsonii*	NR_144866	100	100	1.31E-82
OTU 83	g__Penicillium	*Penicillium thymicola*	NR_137883	100	100	1.31E-82
OTU 83	g__Penicillium	*Penicillium marinum*	NR_137882	100	100	1.31E-82
OTU 83	g__Penicillium	*Penicillium compactum CGMCC 3.15411*	NR_144844	100	100	1.31E-82
OTU 83	g__Penicillium	*Penicillium concentricum*	NR_138334	100	100	1.31E-82
OTU 83	g__Penicillium	*Penicillium viridicatum FRR 963*	NR_119496	100	100	1.31E-82
OTU 83	g__Penicillium	*Penicillium verrucosum FRR 965*	NR_119495	100	100	1.31E-82
OTU 83	g__Penicillium	*Penicillium expansum ATCC 7861*	NR_077154	100	100	1.31E-82
OTU 83	g__Penicillium	*Penicillium italicum*	NR_163528	100	100	1.31E-82
OTU 83	g__Penicillium	*Penicillium tricolor*	NR_077206	100	100	1.31E-82
OTU 83	g__Penicillium	*Penicillium polonicum*	NR_103687	100	100	1.31E-82
OTU 83	g__Penicillium	*Penicillium albocoremium IBT 10682*	NR_138271	100	100	1.31E-82
OTU 88	g__Fusarium	*Fusarium robustum*	NR_159851	97.987	100	7.21E-70
OTU 95	g__Aspergillus	*Aspergillus heterocaryoticus*	NR_163674	100	100	3.67E-83
OTU 95	g__Aspergillus	*Aspergillus costiformis*	NR_135434	100	100	3.67E-83
OTU 95	g__Aspergillus	*Aspergillus chevalieri NRRL 78*	NR_135340	100	100	3.67E-83

[a] = QcovHSP stands for query coverage per high scoring pair

4.3 Discussion

Despite the growing amount of research on the gut fungal community (the mycobiota), this field is still in the early stages of understanding the role of fungi in human health and disease.[113] The mycobiota has been associated with immune-mediated inflammatory diseases such as Crohn's disease, where Ott et al. (2008) and Li et al. (2014) found an increase in fungal richness and diversity within the mucosa of individuals with Crohn's.[127,128] A study investigating the gut mycobiota of individuals with diabetes by Salamon et al. (2021) also noted an increase in fungal richness and diversity in type 1 diabetes compared to unaffected control participants.[129] Fungal associations with neurological disorders have also been explored in Alzheimer's disease.[130,131] In a dietary intervention study, a small cohort of individuals with mild cognitive impairment were shown to have a lower gut fungal diversity compared to cognitively normal individuals.[130] In a mouse model of Alzheimer's, Wu et al. (2019) showed that *Candida albicans* increase amyloid precursor protein, which is associated with plaque formation in Alzheimer's.[131] In another neurological study, De Pablo-Fernandez et al. (2022) indicated that the fungal mycobiota is more diverse in individuals with Parkinson's compared to unaffected controls.[132] However, the characterization of the gut mycobiota in multiple sclerosis (MS), especially POMS, is limited. Our study aims to investigate whether the gut fungal profiles of POMS individuals differ from that of monophasic acquired demyelinating syndrome (monoADS) or unaffected control individuals without MS or monoADS.

Currently, only two studies have explored the gut mycobiome in MS individuals. The first study was by Shah et al. (2021), who compared the gut mycobiota of 25 adult MS individuals to 22 unaffected controls,[28] and a second, more recent study by Yadav et al. (2022) that included 20 adult MS and 33 unaffected controls.[29] Both studies reported a dominating

presence of *Ascomycota* at the phylum level in both MS and controls, while Yadav et al. (2022) indicated an increased presence of *Basidiomycota* in MS. In our study, we observed the same overwhelming presence of *Ascomycota* in all groups and an increased presence of *Basidiomycota* in only POMS (**Figure 4.4**). In terms of the top genera, Yadav et al. (2022) detected *Saccharomyces, Candida, Cladosporium, Malassezia,* and *Penicillium* within their most abundant genera in both MS and control participants, and Shah et al. (2021) also found *Saccharomyces, Aspergillus, Agaricus,* and *Penicillium* among the most abundant.[28,29] In our study, we also found *Saccharomyces* at the highest abundance in all groups, and we noted that *Candida*, *Cladosporium*, and *Penicillium* can be found in all groups (**Figure 4.5**).

Although most of the top-represented genera overlap between the three groups in our study, they differ in the proportionality of specific genera (**Figure 4.5**). For example, *Saccharomyces* dominates across all three groups (48.8% POMS, 52.6% monoADS, 42.0% unaffected controls), followed by *Candida* (38.0% POMS, 9.3% monoADS, 43.1% unaffected controls) and *Penicillium* (3.0% POMS, 14.2% monoADS, 6.4% unaffected controls). Many of our top-represented genera were also found in other gut mycobiome studies. One study on 215 healthy individuals from the Human Microbiome Project found that *Saccharomyces* and *Candida* were among the highest abundance.[50] The commonality of these two genera may be attributed to extrinsic sources, as a controlled-diet study by Auchtung et al. (2018) noted a decrease in gut colonization by both *Saccharomyces* and *Candida* after administration of an *S. cerevisiae*-free diet and/or increasing dental hygiene.[30] These fungal organisms, therefore, may be poor markers for monoADS or MS. It is possible that rare genera or species may play a larger role in the overall disease schema, which has been hypothesized in bacterial microbiome research, or it may

be that the presence of by-products produced by the mycobiota is more important than the organism itself.[35,133]

Interestingly, we detected a highly abundant genus unique to each of the groups within the top-ten: *Agaricus* in POMS, *Phoma* in monoADS, and *Exophiala* in unaffected participants. Further examination of the data showed that *A. bisporus* composed 99.9% of the *Agaricus* genus in POMS participants. *A. bisporus* is edible and commonly sold in supermarkets as the white button mushroom.[51] Although the *Agaricus* genus is only composed of 0.0198% of the total genera in monoADS and 0.0620% in unaffected controls, *A. bisporus* also constituted a majority of the *Agaricus spp.* (100% monoADS and 89.8% unaffected controls). This supports the hypothesis that the gut mycobiome is largely impacted by diet and transient in nature, as we also detected *Saccharomyces spp.* and *Candida spp.* at the highest proportion in all groups.[30] We also find that the only species within the genus *Phoma* is *P. herbarum.* This organism is known to produce cytochalasins, a mycotoxin that modulates cell division that causes apoptosis, and is thought to have anti-inflammatory effects.[134–136] However, it is not known if *P. herbarum* plays a role in modulating inflammatory responses in monoADS participants or is involved in monoADS pathology.

Both studies examining the gut mycobiome in MS also report increased fungal richness in MS compared to control participants, although only Shah et al. (2021) reported a significantly higher diversity in MS participants.[28] Furthermore, the authors of both studies found a significant difference in beta diversity, with Yadav et al. (2022) reporting a distinct clustering between MS and unaffected controls.[29] In our study, we find that unaffected control participants have the richest fungal community, followed by POMS, and monoADS (**Figure 4.6A**). However, monoADS and unaffected control participants had a higher fungal diversity, whereas POMS had

the lowest diversity according to Shannon and inverse Simpson's indices (**Figure 4.6B & C**). Although trends in alpha diversity indices are notable, no statistical significance was observed between participants with POMS, monoADS, or the unaffected controls, which is consistent with Yadav et al. (2022), who compared adult MS to controls.[29] We also did not observe any distinct clustering between any of the three groups (**Figure 4.8**), Statistical analysis by PERMANOVA did detect a difference between community profiles of monoADS and MS participants, but neither was significant when compared to unaffected participants.

We noted that our POMS cohort contains a less rich and diverse community compared to unaffected controls (**Figure 4.6**), while Shah et al. (2021) and Yadav et al. (2022) found the opposite.[25,26] The difference between our study and those focused on adult MS may be due to the nature of our pediatric cohort. Our observation of a more rich and diverse community within our pediatric controls reflects the hypothesis of eubiosis—intestinal homeostasis is maintained, and a balanced composition of microbes exists—versus dysbiosis—the dysregulation of homeostasis associated with an imbalanced microbial composition—which seemingly occurs in our POMS cohort.[137,138] A study conducted by Strati et al. (2016) on healthy participants supports the idea that the reported gut mycobiota is richer in healthy pediatric individuals such as children (3–10 years) and adolescents (11–17 years), as compared to healthy adults (≥ 18).[141] Furthermore, although the community profiles of our POMS and monoADS cohorts differ in terms of beta diversity, a more thorough investigation is required, as monoADS is a broad category that encompasses multiple diseases, including optic neuritis and acute disseminated encephalomyelitis.[142] Importantly, our results indicate there is no difference between the fungal communities of POMS or monoADS compared to unaffected participants.

Our study also examined if there was a difference in the number of unique fungal species detected between males versus females, and by age groups (child ≤ 14 years and youth 15–24 years). We detected significant differences in the richness of both stratifications by sex and age groups (**Figure 4.7A & B**). When we examined differences in fungal diversity, however, only the inverse Simpson measure indicated a difference between males compared to females and children versus youth. Since both Shannon and inverse Simpson indices measure diversity in similar ways—they both account for the overall richness as well as the relative abundance of each species—the difference detected by the inverse Simpson index can perhaps be attributed to their difference in weight calculations, as inverse Simpson is geared towards higher abundances while the Shannon index favours richness.[139] Our result differs from Strati et al. (2016), who found no significant differences in community richness between males compared to females, though the authors only examined healthy individuals.[141] In the current study, we looked at the alpha diversity of gender using the whole cohort due to the fact that our POMS participants consist of triple the number of females as compared to males (15:3). We find that females exhibit a more rich gut mycobiome but lower overall diversity. This may be because there is nearly double the number of females as compared to males (29:17) when accounting for all participants.

We further explored the idea that rare species play an important role in demyelination within POMS by differential abundance analysis. Our LEfSe predictive results reveal OTUs significantly predicted to be associated with different groups of our cohort (**Figure 4.9**). We focus only on significant OTUs assignable to a known fungal species using the UNITE database because of multiple ambiguous taxonomic classifications when attempting to resolve identities using the RefSeq ITS database (**Table 4.2**).

In particular, our LEfSe analysis predicted *Fusarium poae* as a marker of POMS. This organism has a role in agriculture as a plant pathogen that commonly contaminates animal feed.[140–142] It is known that the genus *Fusarium* has the potential to produce mycotoxins, which have been shown in *vitro* to inhibit sphingosine biosynthesis important for the formation of sphingomyelin lining the neuronal axon.[143] Also predicted with our POMS cohort is *Cyberlindnera jadinii*, which can be identified under the teleomorphic name of *Candida utilis*. Several *Candida* species are known opportunistic pathogens able to cause invasive Candidiasis of the blood.[108] *C. utilis* has been suggested as a possible biomarker of vulvovaginal candidiasis and has been identified as the causative agent in a rare case of candidemia within a participant with Alzheimer's.[144,145]

Perhaps the most interesting LEfSe prediction for POMS is *Candida albicans*. Although *C. albicans* is known to commonly colonize the human microbiome, it is also a known opportunistic human pathogen.[108] Furthermore, *C. albicans* has previously been detected in cerebrospinal fluid and post-mortem brain tissue samples of MS.[23,24,26,27] As well, in the mice model of MS, experimental autoimmune encephalomyelitis, it has been shown that *C. albicans* exacerbates inflammation and demyelination, but a decrease of *C. albicans* will alleviate these effects.[33] On a note related to neurological diseases, *C. albicans* was also found to play a role in the mice model of Alzheimer's disease.[131] There is a compelling case for *C. albicans* as a predictor of POMS.

For our monoADS cohort, LEfSe analysis predicted the species *Acremonium fusidioides*, *Exophiala lecanii-corni,* and *Talaromyces islandicus*. Relatively little information can be found on *A. fusidioides,* but they have been shown to produce compounds able to inhibit the growth of HL-60 human leukemia cells in vitro.[146] In a rare isolated case of phaeohyphomycosis, a disease

documented as having a ~70% mortality rate in CNS infections, *E. lecanii-corni* had been detected and associated as the causative agent.[147,148] Furthermore, studies show that *T. islandicus* is capable of producing the mycotoxin cyclic pentapeptide cyclochlorotine, which is said to have mutagenic and toxic effects on humans.[149] Although these organisms were predicted as significant markers, more research is necessary in order to understand their role in monoADS.

Our LEfSe analysis associated *Penicillium aurantiogriseum* and *P. solitum* with unaffected control participants. *P. aurantiogriseum* has documented use in the biotechnology sector to produce enzymes such as collagenase, which is associated with foods with antioxidant and anti-hypersensitivity activity.[150] *P. solitum* is known to produce many types of metabolites important for sustaining human health, such as compactin, benzodiazepine alkaloids, and meroterpenoids.[151]

In this work, we sought to investigate if a pediatric cohort consisting of POMS and monoADS individuals differs in gut mycobiota composition compared to unaffected controls. We found there is no significant difference in terms of alpha or beta diversity when comparing either neurological disorders to control individuals. The profiles of the top ten fungal genera between conditions are similar, though we note that *A. bisporus* is highly represented in our POMS cohort. Our predictive analysis yielded unique fungal markers representing each of the three groups (POMS, monoADS, and unaffected controls), which are potential candidates requiring further investigation.

CHAPTER 5: DISCUSSION AND CONCLUSIONS

5.1 Conclusions and impacts of chapter 3

In chapter three, we used 19 pre-defined fungal organisms to examine the impact of the PhiX bacteriophage on sequencing quality. We also inspected the limitations of the UNITE ribosomal database and assessed the performance of three bioinformatics pipelines used in fungal metataxonomic classification. In comparison to a 25% PhiX spike-in, sequencing with a 50% PhiX spike-in increases the quality of base calls over the entire length of the ITS2 subregion. Both tested PhiX concentrations sustained ≥ 99.9% accuracy in base calling for a length of ~200 bp, but overall a 50% PhiX spike-in resulted in 15.3% more bases called with a Phred ≥ 30 (**Figure 3.2**). The cost of enhanced base calling by a 50% PhiX spike-in is a ~5.6-fold decrease in the overall sequence abundance. Therefore, our results indicate that a 25% PhiX spike-in is more practical than a 50% spike-in for metataxonomic analysis of the mycobiome.

Through our pre-experimental BLAST analysis, we found 18 of the 19 expected reference fungi in the UNITE ribosomal database (**Table 3.0**), and four of these organisms were found under synonymous/alternative nomenclature. This suggests that while the UNITE database may not be entirely comprehensive in its ability to assign taxa, it does seem to perform quite well.[58,152]

Furthermore, we note that the choice of metataxonomic analysis pipeline is important, as each will yield different fungal characterization profiles (**Figure 3.4, 3.5, 3.7, & 3.8b**). In our analysis, the LotuS pipeline outperformed the other tested fungal metataxonomic analysis pipelines. LotuS was able to identify the most mock community organisms (**Figure 3.5 & 3.9a**),

retained the largest number of sequencing reads (**Table 3.1**) and made the fewest false-positive assignments (**Figure 3.9b**).

We recommend using a 25% PhiX spike-in to maximize the number of usable reads without sacrificing quality. Also, researchers should be mindful that the popular UNITE ribosomal database may not comprehensively identify all fungi in the dataset. As well, researchers should be aware that the fungal nomenclature has yet to be standardized amongst the mycological research community. We find that the LotuS pipeline is a good choice for processing raw sequencing reads and taxonomic assignments and is therefore recommended for use in fungal metataxonomic studies.

5.2 Conclusions and impacts of chapter 4

Chapter four describes a novel investigation of fungi inhabiting the human gut in POMS individuals, as compared to a similar but distinct disease—monoADS—and unaffected control individuals. Although studies have investigated the gut mycobiota in adult-onset MS, our results show similarities and differences compared to POMS. Notable similarities include the overwhelming presence of the phylum *Ascomycota* and the genus *Saccharomyces* (**Figure 4.4**).

In our study, we note that the phylum *Ascomycota* and genus *Saccharomyces* are highly abundant among all three diagnostic groups. In addition to this, the genus *Candida* was also detected within the top ten most abundant genera in the three groups (**Figure 4.5**). In another mycobiota study, Auchtung et al. (2018) concluded that the *Saccharomyces* and *Candida* genera are transient colonizers of the human gut.[30] This is because their abundance can be decreased by consuming diets consisting of low amounts of yeasts and through proper dental hygiene.

Another interesting element of our study is the genus *Agaricus*, which was only detected within the top ten most abundant taxa in POMS individuals. Further inspection of this genus revealed that *Agaricus bisporus* represents 99.9% of the *Agaricus spp*. in POMS individuals, 100.0% in monoADS, and 89.8% within unaffected controls. *A. bisporus* is also known as the edible white button mushroom which is commonly sold in supermarkets, so it is possible that these differential abundances may be due to diet.[51] We also find that the diversity of the fungal profiles of POMS and monoADS individuals is not significantly different when compared to unaffected control individuals (**Figure 4.6 & 4.8**).

Given that *Saccharomyces*, *Candida*, and *A. bisporus* are commonly found within our three cohort groups, and that the fungal profiles between groups are not significantly different compared to unaffected controls, the fungal profiles of our cohort may be largely composed of transient organisms. With this knowledge, we used a predictive technique to infer if less common organisms have a role in the pathophysiology of POMS. Our linear discriminant modelling analysis indicated that *Candida albicans, Cyberlindnera jadinii,* and *Fusarium poae* are predictive biomarkers of POMS in our study (**Figure 4.9**). Of the predictive markers within POMS individuals, *C. albicans* may be of particular interest due to its implications in experimental autoimmune encephalomyelitis.[33]

Our findings provide a guide for future "-omics" based investigations in MS, including analysis of pediatric-onset participants, and potential rare and/or under-detected fungal markers, which may be examined for their functional potential and interaction with the wider gut microbiome community.

5.3 Study limitations

Advancements in the discovery of fungi through refinement of methodology is a work in progress. We acknowledge that metataxonomic analysis of fungi is not ideal as there currently exists no standardized protocols to perform this analysis.[42,46] Methods to extract fungal DNA have been researched, but no general consensus has been made on the technologies and methodologies used to examine fungi in the human gut.[46,48–50]

Steps in the extraction protocol, such as freeze-thaw cycles, the addition of a cryoprotectant and/or other nucleic acid degrading/preserving chemicals, and cell wall lysis strategy all play a role in determining the final DNA extraction yield, quality, and species of fungi captured.[45] For our fungal mock community control used in chapter three, the extraction step is less of a problem because we received pre-extracted fungal DNA of known identities. The limitation of our mock community is not knowing if the buffer storing our fungal mock community contains any additive chemicals, or how many freeze-thaw cycles were performed prior to use. In terms of the stool samples from our study cohort presented in chapter four, the integrity of the sample is hard to confirm. Some of the stool samples had been in storage for four years, and multiple transfers have been made between locations (hospital, university, and lab), meaning the storage conditions of these samples have varied over time. Additionally, many of our stool samples were processed for another study involving 16S bacterial sequencing prior to the start of the mycobiome projects, meaning they have undergone freeze-thaw cycles, and there is limited/no quantity of stool remaining for certain individuals. This hinders us from re-analysis of specific study individuals and from conducting any future analyses on participants with insufficient stool samples.

As previously noted, a consensus for standardized DNA extraction methods for mycobiome studies has not been reached. Given the cell wall composition of eukaryotic fungi differs from bacteria, we added an additional lysis step to the sequencing protocol kit.[56] In our study in chapter four, bead beating was used to lyse the fungal cell wall, but this method is rather harsh on the DNA and may lead to shearing.[45,49] Alternative methodologies for fungal lysis exist, such as enzymatic lysis, sonification, and the traditional phenol/chloroform extraction method. Each extraction method has its pros and cons. For example, the phenol/chloroform protocol is said to outperform kit-based methods, but these chemicals are rather harmful to humans and the environment, making disposal challenging.[45]

There are also limitations in the sequencing of our experimental stool samples. Our choice of the Illumina sequencing platform for examining the fungal communities had previously been validated.[42] The Illumina platform is rather limiting in terms of the maximum length of nucleotides that can be sequenced (~600 bp), and in many cases, this limitation cannot fully capture the fungal ITS2 metabarcoding marker, which is noted to vary in length.[55,74] Certain fungal species may contain IT2 lengths surpassing the current Illumina platform limit of 600 bp, which may result in an inaccurate representation of an organism or prevent the identification of certain fungi.[74]

Furthermore, it is well documented that the choice of metabarcoding target, as well as universal primers used to amplify the fungal subregions, will ultimately bias the amplification and identification of certain fungal phyla.[42,55] We chose to use a recommended set of universal primers to amplify the fungal ITS2, but our choice of primers is only able to capture 98–99% of all currently documented fungi as determined by past ecological studies.[42] We did not assess alternative primer pairs in comparison to the recommended set.

Although our experiment to investigate batch effect did not reveal a significant association of our identified taxa with batch, the notable size difference in sequences generated per run implies we cannot conclusively rule it out.

Bioinformatics software for analyzing the fungal components of the human gut has also not been standardized.[45,63] Although we conducted a study in chapter three to compare the identification abilities of three fungal metabarcoding pipelines (LotuS, mothur, and PIPITS), it should be noted that the bioinformatics software comprising these pipelines are constantly being updated. Notably, the LotuS pipeline used in chapters three and four of this thesis has since released a new version (LotuS2) in late 2022.[153]

Another limitation in our study is that we were not able to identify a majority of the members expected within our mock fungal control, possibly due to the bioinformatics software parameters set or the database used for annotation. The UNITE fungal ribosomal database used in this study was determined to be limiting as we were not able to identify one expected mock community member in our evaluation of three metabarcoding pipelines (chapter three), and a lot of unknown fungal organisms were identified within the cohort of our POMS study (chapter four). Our investigation of the gut mycobiome of our POMS cohort (chapter four) identified many fungi at the species level using the RefSeq ITS database that could not be definitively assigned using the UNITE database. Although the use of the RefSeq ITS database did provide a higher taxonomic resolution for most OTUs resulting from our LEfSe analysis, some OTUs could not be unambiguously classified to a single species level. This is likely because many fungal organisms are yet to be discovered, and the fungal nomenclature system is undergoing active restructuring. Further, it is difficult to distinguish between anamorphic and teleomorphic forms of fungi through sequencing.[58]

We also acknowledge that our study cohort is limited in sample size, which is expected as the prevalence of pediatric-onset multiple sclerosis is a rare disease,[8] but this, in turn, affected our statistical interpretation of the results and led to the exclusion of certain research questions, such as the effect of dimethyl fumarate on the gut mycobiome. We also had to exclude a majority of our American study participants due to the challenges related to obtaining and accessing associated metadata. As well, we are aware of demographic confounders potentially impacting our study results, such as individuals representing different income levels, gender, and potentially the HLA-DRB1 allele and other genetic variants found in higher frequency within individuals of European descent.[2,14] Our study is North American-based, therefore our cohort is limited to represent a high-income continent and gender balance is difficult due to POMS being a female-dominated disease, as well, metadata pertaining to an individual's ancestral background is often ambiguous.[8]

5.4 Future directions

In our analysis of the mock community using three different bioinformatics pipelines (chapter three), our results concluded that the LotuS pipeline best identified six out of the 19 expected mock fungal organisms when a 25% PhiX spike-in was used. Of the 13 mock organisms we could not identify, we noted that 12 of them had annotations within the UNITE fungal ribosomal database. Further BLAST analysis showed that these 12 missing mock organisms can be identified from the raw sequences obtained through Illumina sequencing. This means that the limiting factor here was a step in selecting the processing parameters set in the bioinformatics pipelines. In the future, we recommend users of a pipeline verify its taxonomic assignment accuracy through the use of the mock community, confirm that the pipeline is suitable for the analysis, and fine tune the pipeline parameters where possible.

It has been shown that amplification of the entire ITS subregion (ITS1, 5.8S, and ITS2) is superior to capturing just the ITS2 subregion alone,[55] suggesting that our studies may benefit from using long-read sequencing. Instead of the Illumina platform, we should explore if technologies such as the Oxford Nanopore would result in improved taxonomic classifications using our mock control as a ground truth.

Furthermore, our current study investigating the gut mycobiome within POMS individuals (chapter four) did not examine the functional potential of the identified fungi. In order to dive deeper into the functional potential of identified fungi within POMS, we could employ the use of transcriptomics through long-read sequencing or shotgun metagenomics. Though it's difficult to infer fungal functional potential from shotgun metagenomics studies as reference databases for functional genes of fungi are lacking. However, tools such as FungalTraits can potentially be used to infer functional potential from ITS sequences, which is based upon conserved functional traits observed for limited groups (guilds) of fungi.[47] Another future objective is to include biochemical analysis of certain fungal species associated with POMS individuals identified in chapter four. We can either attempt proteomics analysis or check the biochemical properties of metabolites, and perhaps examine if mycotoxins are produced, possibly with the aid of chromatography techniques. Adding to this, supplementing the current studies using culture-based and microscopy techniques would enhance the confidence of our bioinformatics analysis. Although many fungal organisms currently cannot be cultured, phenotypic inspection using microscopy is still considered the gold standard for fungal identification.[51]

Currently, available methods to study the fungal mycobiome require further development to better study the role of the gut mycobiota within diseases such as POMS and monoADS. Our

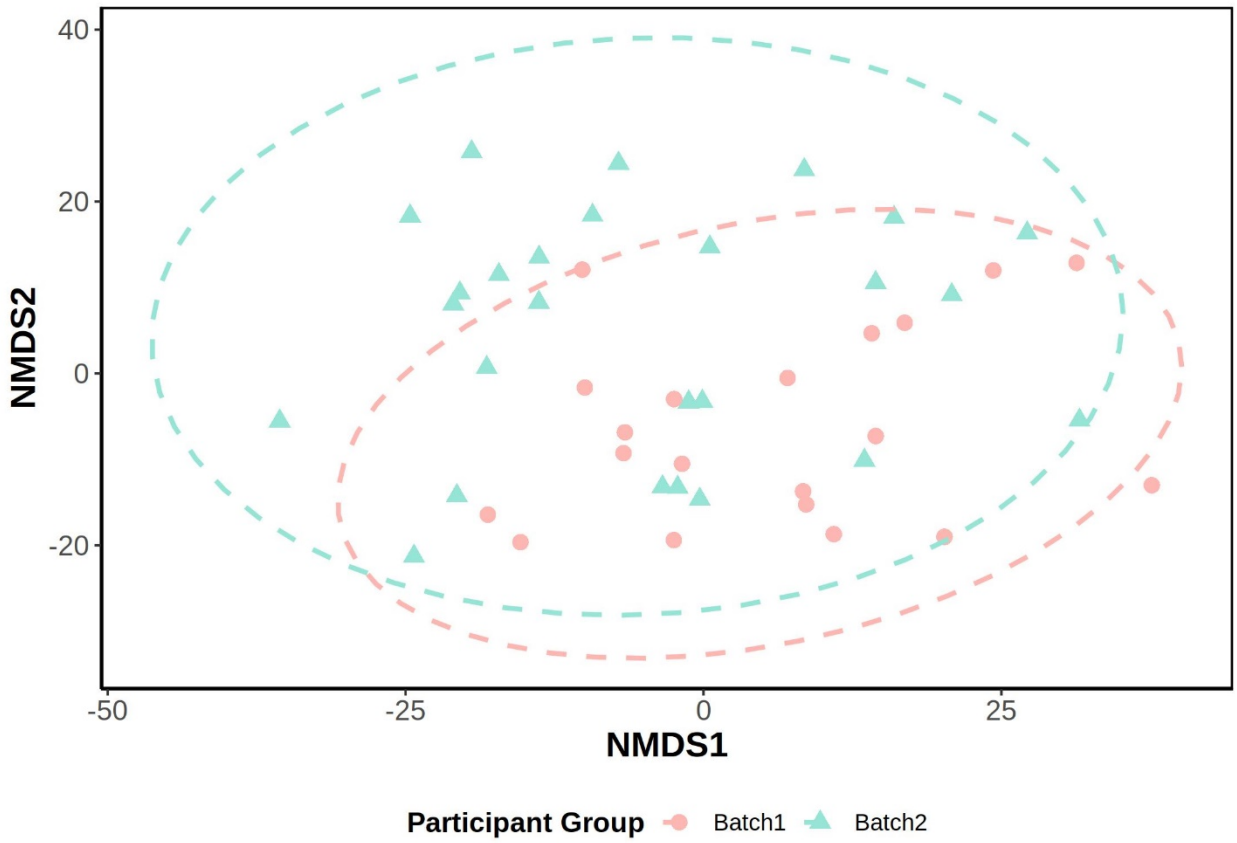
study investigating the mycobiome in chapter four provides a high-level characterization of the fungal gut community within POMS, compared to monoADS, and unaffected control participants, though our limited study size prevented our ability to substratify our cohorts by potential confounders such as disease-modifying drugs, as well, a larger sample size would improve the power of our diversity analysis on sex and age groups. In future analyses, an increase in study participants representing each of the study groups would enhance the ability to explore potential confounders. In addition to the current analysis, our study would benefit from exploring if a controlled diet influences the gut mycobiota in POMS individuals, as well as examining immunological aspects through blood-typing in order to identify subsets of immune cells.

Overall, our methodology assessment of mycobiota analysis techniques (chapter three) guides future research by showing the effect of varying PhiX concentrations (25% vs. 50%) on sequencing quality and quantity, and we show that the choice of fungal metataxonomic pipelines impacts the profiling of a fungal community. We found that the POMS gut mycobiota does not significantly differ relative to unaffected controls (chapter four). We also found that rare or underrepresented species may play a role, as indicated by our predictive analysis. These results provide a guide towards standardizing fungal metataxonomic methodology and offer a glimpse into the gut mycobiota of POMS individuals.

APPENDIX

AP1 No clear indication of batch effects

Analysis of samples from both batches spiked with 25% PhiX and processed through the LotuS pipeline, showed no clear indication of batch effects. The 95% confidence interval ellipses from both batches in the principal component analysis overlapped. This result was also observed when filtering for the 46 participants discussed in the POMS study of chapter 4 (**Supplementary Figure 1**). Although no clear indication of a batch effect is apparent, these results may be limited due to the small sample size of our study, which could potentially be enhanced by adding technical sequencing replicates. This however, was not possible because our study is limited by the quantity of remaining stool samples.



Supplementary Figure 1. NMDS analysis of batch effect for 46 study participants of the POMS study in chapter 4. The distance method used is Euclidean.

REFERENCES

1. Petrova N, Carassiti D, Altmann DR, Baker D, Schmierer K. Axonal loss in the multiple sclerosis spinal cord revisited. Brain Pathol Zurich Switz. 2018 May;28(3):334–48.

2. Filippi M, Bar-Or A, Piehl F, Preziosa P, Solari A, Vukusic S, et al. Multiple sclerosis. Nat Rev Dis Primer. 2018 Nov 8;4(1):1–27.

3. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). Neurology. 1983 Nov 1;33(11):1444–1444.

4. Lublin FD, Reingold SC, Sclerosis* NMSS (USA) AC on CT of NA in M. Defining the clinical course of multiple sclerosis: Results of an international survey. Neurology. 1996 Apr 1;46(4):907–11.

5. Lublin FD, Reingold SC, Cohen JA, Cutter GR, Sørensen PS, Thompson AJ, et al. Defining the clinical course of multiple sclerosis: The 2013 revisions. Neurology. 2014 Jul 15;83(3):278–86.

6. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. Ann Neurol. 2011;69(2):292–302.

7. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol. 2018 Feb 1;17(2):162–73.

8. MSIF MSIF. Atlas of MS 3rd Edition Epidemiology Report [Internet]. 2020 Sep [cited 2022 Feb 7]. Available from: https://www.msif.org/wp-content/uploads/2020/10/Atlas-3rd-Edition-Epidemiology-report-EN-updated-30-9-20.pdf

9. Walton C, King R, Rechtman L, Kaye W, Leray E, Marrie RA, et al. Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition. Mult Scler Houndmills Basingstoke Engl. 2020 Dec;26(14):1816–21.

10. Alroughani R, Boyko A. Pediatric multiple sclerosis: a review. BMC Neurol. 2018 Mar 9;18(1):27.

11. McKay KA, Hillert J, Manouchehrinia A. Long-term disability progression of pediatric-onset multiple sclerosis. Neurology. 2019 Jun 11;92(24):e2764–73.

12. Bigi S, Banwell B. Pediatric Multiple Sclerosis. J Child Neurol. 2012;27(11):1378–83.

13. Aubert-Broche B, Weier K, Longoni G, Fonov VS, Bar-Or A, Marrie RA, et al. Monophasic demyelination reduces brain growth in children. Neurology. 2017 May 2;88(18):1744–50.

14. Nakatsuka N, Patterson N, Patsopoulos NA, Altemose N, Tandon A, Beecham AH, et al. Two genetic variants explain the association of European ancestry with multiple sclerosis risk in African-Americans. Sci Rep. 2020 Oct 9;10(1):1–9.

15. Minagar A, Alexander JS. Blood-brain barrier disruption in multiple sclerosis. Mult Scler J. 2003 Dec 1;9(6):540–9.

16. Ortiz GG, Pacheco-Moisés FP, Macías-Islas MÁ, Flores-Alvarado LJ, Mireles-Ramírez MA, González-Renovato ED, et al. Role of the Blood–Brain Barrier in Multiple Sclerosis. Arch Med Res. 2014 Nov 1;45(8):687–97.

17. Mirzaei F, Michels KB, Munger K, O'Reilly E, Chitnis T, Forman MR, et al. Gestational Vitamin D and the Risk of Multiple Sclerosis in the Offspring. Ann Neurol. 2011 Jul;70(1):30–40.

18. Ebers GC, Bulman DE, Sadovnick AD, Paty DW, Warren S, Hader W, et al. A Population-Based Study of Multiple Sclerosis in Twins. N Engl J Med. 1986 Dec 25;315(26):1638–42.

19. Moutsianas L, Jostins L, Beecham AH, Dilthey AT, Xifara DK, Ban M, et al. Class II HLA interactions modulate genetic risk for multiple sclerosis. Nat Genet. 2015 Oct;47(10):1107–13.

20. Endriz J, Ho PP, Steinman L. Time correlation between mononucleosis and initial symptoms of MS. Neurol - Neuroimmunol Neuroinflammation [Internet]. 2017 May 1 [cited 2021 Sep 6];4(3). Available from: https://nn.neurology.org/content/4/3/e308

21. Bjornevik K, Cortese M, Healy BC, Kuhle J, Mina MJ, Leng Y, et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. Science. 2022 Jan 21;375(6578):296–301.

22. Benito-León J, Laurence M. The Role of Fungi in the Etiology of Multiple Sclerosis. Front Neurol. 2017;8:535.

23. Ramos M, Pisa D, Molina S, Rábano A, Juarranz A, Carrasco L. Fungal Infection in Patients with Multiple Sclerosis. Open Mycol J. 2008 Apr 18;2.

24. Pisa D, Alonso R, Carrasco L. Fungal infection in a patient with multiple sclerosis. Eur J Clin Microbiol Infect Dis Off Publ Eur Soc Clin Microbiol. 2011 Oct;30(10):1173–80.

25. Perlejewski K, Bukowska-Ośko I, Nakamura S, Motooka D, Stokowy T, Płoski R, et al. Metagenomic Analysis of Cerebrospinal Fluid from Patients with Multiple Sclerosis. Adv Exp Med Biol. 2016;935:89–98.

26. Pisa D, Alonso R, Jiménez-Jiménez FJ, Carrasco L. Fungal infection in cerebrospinal fluid from some patients with multiple sclerosis. Eur J Clin Microbiol Infect Dis Off Publ Eur Soc Clin Microbiol. 2013 Jun;32(6):795–801.

27. Alonso R, Fernández-Fernández AM, Pisa D, Carrasco L. Multiple sclerosis and mixed microbial infections. Direct identification of fungi and bacteria in nervous tissue. Neurobiol Dis. 2018 Sep 1;117:42–61.

28. Shah S, Locca A, Dorsett Y, Cantoni C, Ghezzi L, Lin Q, et al. Alterations of the gut mycobiome in patients with MS. EBioMedicine. 2021 Sep;71:103557.

29. Yadav M, Ali S, Shrode RL, Shahi SK, Jensen SN, Hoang J, et al. Multiple sclerosis patients have an altered gut mycobiome and increased fungal to bacterial richness. PLOS ONE. 2022 Apr 26;17(4):e0264556.

30. Auchtung TA, Fofanova TY, Stewart CJ, Nash AK, Wong MC, Gesell JR, et al. Investigating Colonization of the Healthy Adult Gastrointestinal Tract by Fungi. mSphere. 2018 Apr;3(2):e00092-18.

31. Bjelobaba I, Begovic-Kupresanin V, Pekovic S, Lavrnja I. Animal models of multiple sclerosis: Focus on experimental autoimmune encephalomyelitis. J Neurosci Res. 2018;96(6):1021–42.

32. Takata K, Tomita T, Okuno T, Kinoshita M, Koda T, Honorat JA, et al. Dietary Yeasts Reduce Inflammation in Central Nerve System via Microflora. Ann Clin Transl Neurol. 2015 Jan;2(1):56–66.

33. Fraga-Silva TFC, Mimura LAN, Marchetti CM, Chiuso-Minicucci F, França TGD, Zorzella-Pezavento SFG, et al. Experimental Autoimmune Encephalomyelitis Development Is Aggravated by Candida albicans Infection. J Immunol Res. 2015;2015:635052.

34. Willis CL, Leach L, Clarke GJ, Nolan CC, Ray DE. Reversible disruption of tight junction complexes in the rat blood-brain barrier, following transitory focal astrocyte loss. Glia. 2004;48(1):1–13.

35. Fraga-Silva TF de C, Mimura LAN, Leite L de CT, Borim PA, Ishikawa LLW, Venturini J, et al. Gliotoxin Aggravates Experimental Autoimmune Encephalomyelitis by Triggering Neuroinflammation. Toxins. 2019 Jul 26;11(8):443.

36. Purzycki CB, Shain DH. Fungal toxins and multiple sclerosis: A compelling connection. Brain Res Bull. 2010 Apr 29;82(1):4–6.

37. Saidu NEB, Kavian N, Leroy K, Jacob C, Nicco C, Batteux F, et al. Dimethyl fumarate, a two-edged drug: Current status and future directions. Med Res Rev. 2019;39(5):1923–52.

38. Yadav SK, Soin D, Ito K, Dhib-Jalbut S. Insight into the mechanism of action of dimethyl fumarate in multiple sclerosis. J Mol Med. 2019 Apr 1;97(4):463–72.

39. Stefanelli P, Girolimetti S, Santilio A, Dommarco R. Survey of dimethyl fumarate in desiccant products during 2009 in Italy. Bull Environ Contam Toxicol. 2011 Apr;86(4):428–32.

40. Ma N, Wu Y, Xie F, Du K, Wang Y, Shi L, et al. Dimethyl fumarate reduces the risk of mycotoxins via improving intestinal barrier and microbiota. Oncotarget. 2017 May 16;8(27):44625–38.

41. Reuter JA, Spacek D, Snyder MP. High-Throughput Sequencing Technologies. Mol Cell. 2015 May 21;58(4):586–97.

42. Nilsson RH, Anslan S, Bahram M, Wurzbacher C, Baldrian P, Tedersoo L. Mycobiome diversity: high-throughput sequencing and identification of fungi. Nat Rev Microbiol. 2019 Feb;17(2):95–109.

43. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016 Jun;17(6):333–51.

44. Thielemann N, Herz M, Kurzai O, Martin R. Analyzing the human gut mycobiome – A short guide for beginners. Comput Struct Biotechnol J. 2022 Jan 19;20:608–14.

45. Tiew PY, Mac Aogain M, Ali NABM, Thng KX, Goh K, Lau KJX, et al. The Mycobiome in Health and Disease: Emerging Concepts, Methodologies and Challenges. Mycopathologia. 2020;185(2):207–31.

46. Fiedorová K, Radvanský M, Němcová E, Grombiříková H, Bosák J, Černochová M, et al. The Impact of DNA Extraction Methods on Stool Bacterial and Fungal Microbiota Community Recovery. Front Microbiol. 2019;10:821.

47. Zhang F, Aschenbrenner D, Yoo JY, Zuo T. The gut mycobiome in health, disease, and clinical applications in association with the gut bacterial microbiome assembly. Lancet Microbe. 2022 Dec 1;3(12):e969–83.

48. Angebault C, Ghozlane A, Volant S, Botterel F, d'Enfert C, Bougnoux ME. Combined bacterial and fungal intestinal microbiota analyses: Impact of storage conditions and DNA extraction protocols. PLOS ONE. 2018 Aug 3;13(8):e0201174.

49. Frau A, Kenny JG, Lenzi L, Campbell BJ, Ijaz UZ, Duckworth CA, et al. DNA extraction and amplicon production strategies deeply inf luence the outcome of gut mycobiome studies. Sci Rep. 2019 Jun 27;9(1):1–17.

50. 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 Nov 25;5(1):153.

51. Webster J, Weber R. Introduction to Fungi. 3rd ed. New York: Cambridge University Press; 2007.

52. Möhlenhoff P, Müller L, Gorbushina AA, Petersen K. Molecular approach to the characterisation of fungal communities: methods for DNA extraction, PCR amplification and DGGE analysis of painted art objects. FEMS Microbiol Lett. 2001 Feb 1;195(2):169–73.

53. Schrader C, Schielke A, Ellerbroek L, Johne R. PCR inhibitors – occurrence, properties and removal. J Appl Microbiol. 2012;113(5):1014–26.

54. Nilsson RH, Kristiansson E, Ryberg M, Hallenberg N, Larsson KH. Intraspecific ITS Variability in the Kingdom Fungi as Expressed in the International Sequence Databases and Its Implications for Molecular Species Identification. Evol Bioinforma Online. 2008 May 26;4:193–201.

55. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci U S A. 2012 Apr 17;109(16):6241–6.

56. Tang J, Iliev ID, Brown J, Underhill DM, Funari VA. Mycobiome: Approaches to Analysis of Intestinal Fungi. J Immunol Methods. 2015 Jun;421:112–21.

57. Iwen PC, Hinrichs SH, Rupp ME. Utilization of the internal transcribed spacer regions as molecular targets to detect and identify human fungal pathogens. Med Mycol. 2002 Jan 1;40(1):87–109.

58. Nilsson RH, Larsson KH, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, et al. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Res. 2019 Jan 8;47(D1):D259–64.

59. Tedersoo L, Anslan S, Bahram M, Põlme S, Riit T, Liiv I, et al. Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. MycoKeys. 2015 May 13;10:1–43.

60. Blaalid R, Kumar S, Nilsson RH, Abarenkov K, Kirk PM, Kauserud H. ITS1 versus ITS2 as DNA metabarcodes for fungi. Mol Ecol Resour. 2013 Mar;13(2):218–24.

61. Hoggard M, Vesty A, Wong G, Montgomery JM, Fourie C, Douglas RG, et al. Characterizing the Human Mycobiota: A Comparison of Small Subunit rRNA, ITS1, ITS2, and Large Subunit rRNA Genomic Targets. Front Microbiol [Internet]. 2018 [cited 2023 Jan 9];9. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2018.02208

62. Carriço JA, Rossi M, Moran-Gilad J, Domselaar GV, Ramirez M. A primer on microbial bioinformatics for nonbioinformaticians. Clin Microbiol Infect. 2018 Apr 1;24(4):342–9.

63. Anslan S, Nilsson RH, Wurzbacher C, Baldrian P, Leho Tedersoo null, Bahram M. Great differences in performance and outcome of high-throughput sequencing data analysis platforms for fungal metabarcoding. MycoKeys. 2018;(39):29–40.

64. Pauvert C, Buée M, Laval V, Edel-Hermann V, Fauchery L, Gautier A, et al. Bioinformatics matters: The accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. Fungal Ecol. 2019 Oct 1;41:23–33.

65. Caruso V, Song X, Asquith M, Karstens L. Performance of Microbiome Sequence Inference Methods in Environments with Varying Biomass. mSystems. 2019 Feb 19;4(1):e00163-18.

66. Marizzoni M, Gurry T, Provasi S, Greub G, Lopizzo N, Ribaldi F, et al. Comparison of Bioinformatics Pipelines and Operating Systems for the Analyses of 16S rRNA Gene Amplicon Sequences in Human Fecal Samples. Front Microbiol. 2020;11:1262.

67. Anslan S, Bahram M, Hiiesalu I, Tedersoo L. PipeCraft: Flexible open-source toolkit for bioinformatics analysis of custom high-throughput amplicon sequencing data. Mol Ecol Resour. 2017 Nov;17(6):e234–40.

68. Gweon HS, Oliver A, Taylor J, Booth T, Gibbs M, Read DS, et al. PIPITS: an automated pipeline for analyses of fungal internal transcribed spacer sequences from the Illumina sequencing platform. Methods Ecol Evol. 2015;6(8):973–80.

69. Hildebrand F, Tadeo R, Voigt AY, Bork P, Raes J. LotuS: an efficient and user-friendly OTU processing pipeline. Microbiome. 2014 Sep 30;2(1):30.

70. Federhen S. The NCBI Taxonomy database. Nucleic Acids Res. 2012 Jan;40(Database issue):D136-143.

71. 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 [Internet]. 2019 [cited 2022 Apr 11];9. Available from: https://www.frontiersin.org/article/10.3389/fmicb.2018.03249

72. Cantoni C, Lin Q, Dorsett Y, Ghezzi L, Liu Z, Pan Y, et al. Alterations of host-gut microbiome interactions in multiple sclerosis. eBioMedicine [Internet]. 2022 Feb 1 [cited 2022 Feb 7];76. Available from: http://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(21)00592-2/fulltext

73. Tremlett H, Fadrosh DW, Faruqi AA, Zhu F, Hart J, Roalstad S, et al. Gut microbiota in early pediatric multiple sclerosis: a case-control study. Eur J Neurol. 2016 Aug;23(8):1308–21.

74. Bakker MG. A fungal mock community control for amplicon sequencing experiments. Mol Ecol Resour. 2018;18(3):541–56.

75. Tremlett H, Zhu F, Arnold D, Bar-Or A, Bernstein CN, Bonner C, et al. The gut microbiota in pediatric multiple sclerosis and demyelinating syndromes. Ann Clin Transl Neurol. 2021;8(12):2252–69.

76. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Internet]. 2010. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

77. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016 Oct 1;32(19):3047–8.

78. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinforma Oxf Engl. 2010 Oct 1;26(19):2460–1.

79. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011 Aug 15;27(16):2194–200.

80. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011 Nov 1;27(21):2957–63.

81. Bengtsson-Palme J, Ryberg M, Hartmann M, Branco S, Wang Z, Godhe A, et al. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. Methods Ecol Evol. 2013;4(10):914–9.

82. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009 Dec 15;10:421.

83. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Softw. 2019;4(43):1686-.

84. Charif D, Lobry JR. SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. In: Structural Approaches to Sequence Evolution. Berlin, Heidelberg: Springer Berlin Heidelberg; p. 207–32. (Biological and Medical Physics, Biomedical Engineering).

85. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016;4:e2584.

86. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009 Dec;75(23):7537–41.

87. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE. 2013 Apr 22;8(4):e61217.

88. Wickham H. ggplot2: elegant graphics for data analysis. 2nd ed. Cham: Springer International Publishing; 2016. (Use R!).

89. Oksanen J, Blanchet G, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package [Internet]. 2020. Available from: https://CRAN.R-project.org/package=vegan

90. Arbizu M. pairwiseAdonis: Pairwise Multilevel Comparison using Adonis [Internet]. 2017. Available from: https://github.com/pmartinezarbizu/pairwiseAdonis

91. Kassambara A, Mundt F. factoextra: Extract and Visualize the Results of Multivariate Data Analyses [Internet]. 2020. Available from: https://CRAN.R-project.org/package=factoextra

92. FASTX-Toolkit [Internet]. [cited 2023 Jun 6]. Available from: http://hannonlab.cshl.edu/fastx_toolkit/index.html

93. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.

94. Frøslev TG, Kjøller R, Bruun HH, Ejrnæs R, Brunbjerg AK, Pietroni C, et al. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nat Commun. 2017 Oct 30;8(1):1–11.

95. Palarea-Albaladejo J, Martín-Fernández JA. zCompositions — R package for multivariate imputation of left-censored data under a compositional approach. Chemom Intell Lab Syst. 2015 Apr 15;143:85–96.

96. Boogaart G van den, Tolosana-Delgado R, Bren M. Compositions: Compositional Data Analysis [Internet]. 2021. Available from: https://CRAN.R-project.org/package=compositions

97. Kassambara A. ggpubr: 'ggplot2' Based Publication Ready Plots [Internet]. 2020. Available from: https://CRAN.R-project.org/package=ggpubr

98. Kassambara A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests [Internet]. 2021. Available from: https://CRAN.R-project.org/package=rstatix

99. Blighe K, Lun A. PCAtools: PCAtools: Everything Principal Components Analysis [Internet]. 2020. Available from: https://github.com/kevinblighe/PCAtools

100. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. Microbiome. 2014 May 5;2(1):15.

101. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. Genome Biol. 2011 Jun 24;12(6):R60.

102. Cao Y. microbiomeMarker: microbiome biomarker analysis [Internet]. 2021. Available from: https://github.com/yiluheihei/microbiomeMarker

103. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016 Jan 4;44(D1):D733–45.

104. Bonneville S, Delpomdor F, Préat A, Chevalier C, Araki T, Kazemian M, et al. Molecular identification of fungi microfossils in a Neoproterozoic shale rock. Sci Adv. 2020;6(4):eaax7599.

105. Loron CC, François C, Rainbird RH, Turner EC, Borensztajn S, Javaux EJ. Early fungi from the Proterozoic era in Arctic Canada. Nature. 2019 Jun;570(7760):232–5.

106. Hawksworth DL, Lücking R. Fungal Diversity Revisited: 2.2 to 3.8 Million Species. Microbiol Spectr [Internet]. 2017 Jul 28 [cited 2022 Feb 14]; Available from: http://journals.asm.org/doi/abs/10.1128/microbiolspec.FUNK-0052-2016

107. Webster J. Introduction to fungi. 3rd ed. Cambridge, UK ; Cambridge University Press; 2007. xviii+841.

108. Goering R, Dockrell H, Zuckerman M, Chiodini P. Mims' Medical Microbiology and Immunology. 6th ed. Elsevier Health Sciences; 2019.

109. Ziemons S, Koutsantas K, Becker K, Dahlmann T, Kück U. Penicillin production in industrial strain Penicillium chrysogenum P2niaD18 is not dependent on the copy number of biosynthesis genes. BMC Biotechnol. 2017 Feb 16;17(1):16.

110. Gautam AK, Verma RK, Avasthi S, Sushma, Bohra Y, Devadatha B, et al. Current Insight into Traditional and Modern Methods in Fungal Diversity Estimates. J Fungi. 2022 Feb 24;8(3):226.

111. Links MG, Dumonceaux TJ, Hemmingsen SM, Hill JE. The chaperonin-60 universal target is a barcode for bacteria that enables de novo assembly of metagenomic sequence data. PloS One. 2012;7(11):e49755.

112. Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, et al. Towards a unified paradigm for sequence-based identification of fungi. Mol Ecol. 2013 Nov;22(21):5271–7.

113. Galloway-Peña JR, Kontoyiannis DP. The gut mycobiome: The overlooked constituent of clinical outcomes and treatment complications in patients with cancer and other immunosuppressive conditions. PLOS Pathog. 2020 Apr 2;16(4):e1008353.

114. Illumina I. Using a PhiX Control for HiSeq Sequencing Runs. 2013.

115. Illumina I. Optimizing Cluster Density on Illumina Sequencing Systems [Internet]. 2016. Available from: https://www.illumina.com/content/dam/illumina-marketing/documents/products/other/miseq-overclustering-primer-770-2014-038.pdf

116. Fungorum. Index Fungorum [Internet]. www.indexfungorum.org. 2022 [cited 2021 Nov 18]. Available from: http://www.indexfungorum.org/Names/Names.asp

117. Robert V, Vu D, Amor ABH, van de Wiele N, Brouwer C, Jabas B, et al. MycoBank gearing up for new horizons. IMA Fungus. 2013 Dec;4(2):371–9.

118. IHMS [Internet]. [cited 2022 Apr 11]. Available from: https://human-microbiome.org/index.php

119. Borst A, Box ATA, Fluit AC. False-Positive Results and Contamination in Nucleic Acid Amplification Assays: Suggestions for a Prevent and Destroy Strategy. Eur J Clin Microbiol Infect Dis. 2004 Apr 1;23(4):289–99.

120. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 2014 Nov 12;12(1):87.

121. Figved N, Myhr K, Larsen J, Aarsland D. Caregiver burden in multiple sclerosis: the impact of neuropsychiatric symptoms. J Neurol Neurosurg Psychiatry. 2007 Oct;78(10):1097–102.

122. Siegert R, Abernethy D. Depression in multiple sclerosis: a review. J Neurol Neurosurg Psychiatry. 2005 Apr;76(4):469–75.

123. Parodi B, Kerlero de Rosbo N. The Gut-Brain Axis in Multiple Sclerosis. Is Its Dysfunction a Pathological Trigger or a Consequence of the Disease? Front Immunol. 2021 Sep 21;12:718220.

124. Berer K, Gerdes LA, Cekanaviciute E, Jia X, Xiao L, Xia Z, et al. Gut microbiota from multiple sclerosis patients enables spontaneous autoimmune encephalomyelitis in mice. Proc Natl Acad Sci. 2017 Oct 3;114(40):10719–24.

125. Sterlin D, Larsen M, Fadlallah J, Parizot C, Vignes M, Autaa G, et al. Perturbed Microbiota/Immune Homeostasis in Multiple Sclerosis. Neurol - Neuroimmunol Neuroinflammation [Internet]. 2021 Jul 1 [cited 2022 Feb 7];8(4). Available from: https://nn.neurology.org/content/8/4/e997

126. Zhang F, Zuo T, Yeoh YK, Cheng FWT, Liu Q, Tang W, et al. Longitudinal dynamics of gut bacteriome, mycobiome and virome after fecal microbiota transplantation in graft-versus-host disease. Nat Commun. 2021 Jan 4;12(1):1–11.

127. 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(7):831–41.

128. Li Q, Wang C, Tang C, He Q, Li N, Li J. Dysbiosis of Gut Fungal Microbiota is Associated With Mucosal Inflammation in Crohn's Disease. J Clin Gastroenterol. 2014 Jul;48(6):513–23.

129. Salamon D, Sroka-Oleksiak A, Gurgul A, Arent Z, Szopa M, Bulanda M, et al. Analysis of the Gut Mycobiome in Adult Patients with Type 1 and Type 2 Diabetes Using Next-Generation Sequencing (NGS) with Increased Sensitivity—Pilot Study. Nutrients. 2021 Mar 25;13(4):1066.

130. Nagpal R, Neth BJ, Wang S, Mishra SP, Craft S, Yadav H. Gut mycobiome and its interaction with diet, gut bacteria and alzheimer's disease markers in subjects with mild cognitive impairment: A pilot study. EBioMedicine. 2020 Sep;59:102950.

131. Wu Y, Du S, Johnson JL, Tung HY, Landers CT, Liu Y, et al. Microglia and amyloid precursor protein coordinate control of transient Candida cerebritis with memory deficits. Nat Commun. 2019 Jan 4;10(1):1–15.

132. De Pablo-Fernandez E, Gebeyehu GG, Flain L, Slater R, Frau A, Ijaz UZ, et al. The faecal metabolome and mycobiome in Parkinson's disease. Parkinsonism Relat Disord. 2022 Feb;95:65–9.

133. Dai W, Chen J, Xiong J. Concept of microbial gatekeepers: Positive guys? Appl Microbiol Biotechnol. 2019 Jan;103(2):633–41.

134. Kim MY, Kim JH, Cho JY. Cytochalasin B Modulates Macrophage-Mediated Inflammatory Responses. Biomol Ther. 2014 Jul;22(4):295–300.

135. Trendowski M. Using Cytochalasins to Improve Current Chemotherapeutic Approaches. Anticancer Agents Med Chem. 2015 Mar;15(3):327–35.

136. Zohri AA, Saber SM. New sources of cytochalasins A and B from dematiaceous hyphomycetes. Lett Appl Microbiol. 1994 Jul;19(1):37–9.

137. Al-Rashidi HE. Gut microbiota and immunity relevance in eubiosis and dysbiosis. Saudi J Biol Sci. 2022 Mar 1;29(3):1628–43.

138. Iebba V, Totino V, Gagliardi A, Santangelo F, Cacciotti F, Trancassini M, et al. Eubiosis and dysbiosis: the two sides of the microbiota. New Microbiol. 2016 Jan;39(1):1–12.

139. Roswell M, Dushoff J, Winfree R. A conceptual guide to measuring species diversity. Oikos. 2021;130(3):321–38.

140. Ghimire B, Martinez-Espinoza AD, Ghimire B, Harrelson BC, Youmans J, Mergoum M, et al. First Report of Fusarium poae Causing Fusarium Head Blight of Wheat in Georgia, U.S.A. Plant Dis. 2021 Feb;105(2):491–491.

141. Tan J, De Zutter N, De Saeger S, De Boevre M, Tran TM, van der Lee T, et al. Presence of the Weakly Pathogenic Fusarium poae in the Fusarium Head Blight Disease Complex Hampers Biocontrol and Chemical Control of the Virulent Fusarium graminearum Pathogen. Front Plant Sci [Internet]. 2021 [cited 2022 Apr 26];12. Available from: https://www.frontiersin.org/article/10.3389/fpls.2021.641890

142. Yang Y, Zhao H, Barrero RA, Zhang B, Sun G, Wilson IW, et al. Genome sequencing and analysis of the paclitaxel-producing endophytic fungus Penicillium aurantiogriseum NRRL 62431. BMC Genomics. 2014 Jan 25;15(1):69.

143. Norred WP, Wang E, Yoo H, Riley RT, Merrill AH. In vitro toxicology of fumonisins and the mechanistic implications. Mycopathologia. 1992 Feb;117(1–2):73–8.

144. Paulovičová E, Hrubiško M. Humoral immune responses against facultative pathogen Candida utilis in atopic patients with vulvovaginal candidiasis. Candida utilis glucomannan – New serologic biomarker. Immunobiology. 2022 Jan 1;227(1):152154.

145. Bougnoux ME, Gueho E, Potocka AC. Resolutive Candida utilis fungemia in a nonneutropenic patient. J Clin Microbiol. 1993 Jun;31(6):1644–5.

146. An X, Feng BM, Chen G, Chen SF, Wang HF, Pei YH. Isolation and identification of two new compounds from marine-derived fungus Acremonium fusidioides RZ01. Chin J Nat Med. 2016 Dec 1;14(12):934–8.

147. Lee KC, Kim MJ, Chae SY, Lee HS, Jang YH, Lee SJ, et al. A Case of Phaeohyphomycosis Caused by Exophiala lecanii-corni. Ann Dermatol. 2016 Jun;28(3):385–7.

148. Revankar SG, Sutton DA, Rinaldi MG. Primary Central Nervous System Phaeohyphomycosis: A Review of 101 Cases. Clin Infect Dis. 2004 Jan 15;38(2):206–16.

149. Schafhauser T, Wibberg D, Rückert C, Winkler A, Flor L, van Pée KH, et al. Draft genome sequence of Talaromyces islandicus (“Penicillium islandicum”) WF-38-12, a neglected mold with significant biotechnological potential. J Biotechnol. 2015;211:101–2.

150. Lima CA, Campos JF, Filho JLL, Converti A, da Cunha MGC, Porto ALF. Antimicrobial and radical scavenging properties of bovine collagen hydrolysates produced by Penicillium aurantiogriseum URM 4622 collagenase. J Food Sci Technol. 2015 Jul;52(7):4459–66.

151. Boruta T, Przerywacz P, Ryngajllo M, Bizukojc M. Bioprocess-related, morphological and bioinformatic perspectives on the biosynthesis of secondary metabolites produced by Penicillium solitum. Process Biochem 1991. 2018;68:12–21.

152. Hawksworth DL. A new dawn for the naming of fungi: impacts of decisions made in Melbourne in July 2011 on the future publication and regulation of fungal names1. IMA Fungus Glob Mycol J. 2011 Dec;2(2):155–62.

153. Özkurt E, Fritscher J, Soranzo N, Ng DYK, Davey RP, Bahram M, et al. LotuS2: an ultrafast and highly accurate tool for amplicon sequencing analysis. Microbiome. 2022 Oct 19;10(1):176.