

# **PUNDEMIC**

## **Elucidating gut mycobiome and Pundemic disease relation, potential for FMT treatment**

Group Semester Project HS2024

**Marie Becker, Aidan Meara, Hannah Seifried**  
Department of Biosystems Science and Engineering

**Course:** Applied Bioinformatics: Microbiomes  
**Supervisor:** Prof. Dr. Nicholas Bokulich

November 16, 2025

# Abstract

This study investigates the gut mycobiome's role in the "Pundemic" disease, a unique condition characterized by excessive pun-making, and evaluates the efficacy of fecal microbiota transplantation (FMT) as a potential treatment. Using ITS sequencing and QIIME2-based analysis, we examined fungal diversity and composition in samples collected from Pundemic patients, healthy controls, and FMT recipients.

While fungal alpha and beta diversity distinguished healthy individuals from diseased patients, no significant changes were observed in fungal communities throughout the treatment. Differential abundance and functional predictions showed no disease-specific fungal markers or significant shifts after treatment. These findings suggest that the fungal microbiome remains stable in Pundemic patients throughout treatment, with limited evidence supporting the role of FMT in altering fungal diversity or composition. Future studies with larger cohorts and advanced methodologies are needed to explore the fungal microbiome's potential involvement in this peculiar condition and its treatment.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Methods</b>	<b>4</b>
2.1	ITS-1 Fungal Reads from Stool Samples . . . . .	4
2.2	Quality Control and Denoising . . . . .	5
2.3	Taxonomy Classification . . . . .	6
2.3.1	Alternative Taxonomy Classification using Naive Bayes Classifier . . . . .	6
2.4	Diversity Analysis . . . . .	6
2.4.1	Alpha Diversity . . . . .	7
2.4.2	Beta Diversity . . . . .	8
2.5	Differential Abundance Analysis . . . . .	10
2.6	Functional Predictions . . . . .	11
<b>3</b>	<b>Results</b>	<b>12</b>
3.1	Quality Control and Denoising . . . . .	12
3.2	Taxonomy Classification . . . . .	13
3.3	Alpha Diversity . . . . .	14
3.3.1	Rarefaction Analysis . . . . .	14
3.3.2	Comparison of Alpha Diversity Metrics . . . . .	15
3.3.3	Alpha Diversity Correlation with Disease Intensity . . . . .	17
3.4	Beta Diversity . . . . .	19
3.5	Differential Abundance . . . . .	20
3.5.1	Data Exploration and Normality Assessment . . . . .	20
3.5.2	ANCOM-BC Results . . . . .	21
3.6	Functional Predictions using FunFun . . . . .	22
<b>4</b>	<b>Discussion</b>	<b>24</b>
<b>5</b>	<b>Supplementary Material</b>	<b>27</b>
<b>A</b>	<b>Appendix</b>	<b>28</b>

# List of Figures

1.1	Conceptual Overview of QIIME2 . . . . .	2
2.1	Overview of the used methods . . . . .	4
2.2	Visualization of the dataset including the distribution of sexes . . . . .	5
3.1	QC plot sequences . . . . .	13
3.2	Combined Taxonomy Barplots . . . . .	14
3.3	Rarefaction plots . . . . .	15
3.4	Pairwise differences (pre-treatment vs. post-treatment) of Shannon Entropy . . . . .	17
3.5	Alpha diversity correlation with disease intensity . . . . .	18
3.6	2D visualization of PCoA using Bray-Curtis dissimilarity metric . . . . .	19
3.7	2D visualization of PCoA using Jaccard Index similarity metric . . . . .	19
3.8	Violin plots illustrating feature sparsity in the dataset. . . . .	21
3.9	PCA plot of KEGG orthology group predictions per sample . . . . .	22
A.1	Distribution of features. . . . .	28
A.2	Kruskal-Wallis test results for Shannon Entropy. . . . .	29
A.3	Boxplot visualizing Shannon Entropy for different treatment response groups pre- and post-treatment. . . . .	30
A.4	PERMANOVA test results for beta diversity based on the Jaccard Index for disease status	30
A.5	Bar plot showing relative abundances of KEGG orthology groups for fungal sequences averaged over samples . . . . .	34

# List of Tables

3.1	Evaluation of preprocessing. . . . .	12
3.2	Kruskal-Wallis Test Results for Shannon Entropy . . . . .	16
3.3	Correlation between alpha diversity metrics and disease intensity (Puns per Hour) . . . . .	18
3.4	PERMANOVA test results for Bray-Curtis dissimilarity based on OTU sequence counts .	20
3.5	PERMANOVA test results for Jaccard index similarity based on OTU sequence counts .	20
3.6	PERMANOVA test results for Bray-Curtis dissimilarity based on relative abundance of KEGG orthology groups. . . . .	23
3.7	PERMANOVA test results for Jaccard Index similarity based on relative abundance of KEGG orthology groups. . . . .	23
A.1	Five most enriched and five most depleted ASVs across the tested metadata categories .	31
A.2	Five most enriched taxa across the tested metadata categories . . . . .	32
A.3	Five most depleted taxa across the tested metadata categories . . . . .	33

# Chapter 1

## Introduction

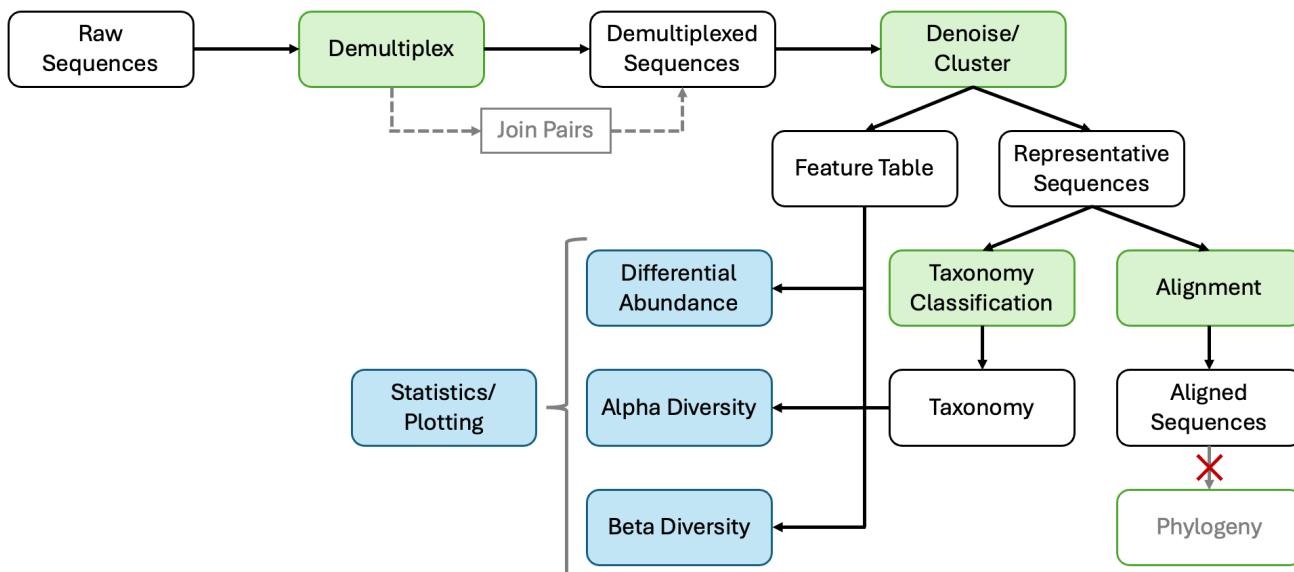
The gut microbiome is the complete community of microorganisms, their theatre of activity and their combined genetic material found within the gastrointestinal tract (GIT). It is composed of various microbes including fungi, bacteria, viruses, and protozoa and influences key physiological processes such as digestion, immunity, and metabolism. The term "gut microbiota" specifically refers to the collection of microorganisms, both beneficial and harmful, living in the GIT [1]. Although fungi have received less attention than bacteria, recent studies suggest they may be also important in maintaining gut health and influencing disease development [2]. The GIT is home to a wide range of fungi, including harmless residents, temporary colonizers, and dangerous pathogens, referred to as "gut mycobiome" [3]. How these fungi interact with their hosts can have a big impact on health. For example, conditions like hepatitis B, obesity, and inflammatory bowel disease (IBD) have all been linked to fungal imbalances [4–8].

Despite these associations, our understanding of how changes in fungal diversity contribute to the progression of these conditions remains limited. Expanding our knowledge of the mycobiome is therefore essential for revealing its possible role in human health and disease. One interesting condition that might be associated with alterations in the gut microbiome, especially the fungal part, is the so-called "pundemic", a peculiar disease where those affected constantly make puns. The pandemic could provide important insights into how imbalances in the gut's fungi might influence behavior.

The goal of this study is to examine the fungal part of the gut microbiome in people affected by the pandemic, particularly in the context of a fecal microbiota transplant (FMT) trial. While FMT has been shown to help recover the balance of gut microbiota in several digestive diseases [9–11], its effects on fungal populations are not yet fully understood. This is why there is growing interest in its potential to affect the fungal microbiome.

In this study, we want to see if FMT can reestablish a healthy balance of fungi in the gut and help reduce the symptoms of the pandemic. To this end, we will compare samples taken before and after treatment from both the FMT and placebo groups. If our study demonstrates that FMT can effectively recover a healthy mycobiome in the context of the pandemic, it could pave the way for further research into fungal-based therapies for other diseases linked to fungal imbalances in the gut.

To analyze the fungal microbiome, we used nuclear ribosomal internal transcribed spacer (ITS) sequences, which are widely recognized for identifying fungi at the species level. The ITS region, particularly its two subregions ITS1 and ITS2, has high variability in both sequence and length, ranging from 200 to 600 base pairs. This variability is due to biological factors, such as high rates of insertions and deletions during evolution, rather than technical issues. These characteristics make ITS regions an



**Figure 1.1:** Conceptual Overview of QIIME2

excellent choice for distinguishing fungal species with precision [12]. For this study, we used forward reads of ITS sequences from fecal samples, alongside metadata, as the foundation for our analysis (chapter 2.1).

The entire analysis was conducted using the QIIME2 bioinformatics pipeline (see Figure 1.1), which allowed us to process and compare fungal data efficiently across different groups, including pandemic patients, healthy donors, and both treatment and placebo groups [13].

We first performed quality control, applying appropriate filtering and denoising techniques. This included trimming the ITS1 primer and filtering out low-quality reads based on quality scores. Denoising was then applied using the DADA2 pipeline, which corrected sequencing errors and removed chimeric sequences, ensuring high-quality data [14, 15] (chapter 2.2). Following this, we performed taxonomy classification (chapter 2.3) to assign fungal sequences to taxonomic groups. This was achieved using both a self-trained and pre-trained classifier, as well as closed-reference clustering against the UNITE database, to ensure an accurate fungal identification [16].

We then analyzed alpha diversity (chapter 2.4.1) to measure fungal diversity within individual samples. Since ITS data does not incorporate phylogenetic information [17], we focused on non-phylogenetic diversity metrics, including Shannon entropy, Pielou evenness, and observed features to capture different aspects of diversity [18, 19]. Pairwise comparisons of diversity were conducted to explore differences, particularly between pre- and post-treatment samples, as well as between treatment and placebo groups. For beta diversity (chapter 2.4.2) we evaluated how the mycobial communities vary across conditions or groups. We visualized these differences using principal coordinate analysis (PCoA) plots based on Jaccard and Bray-Curtis distances and conducted pairwise significance tests to assess the influence of variables such as disease status, treatment response, sex, and ethnicity.

Next, we conducted a differential abundance (chapter 2.5) analysis to identify fungal features that were significantly different between groups. We tested for data normality using the Shapiro-Wilk test

to guide our statistical approach and performed ANCOM (Analysis of Composition of Microbiomes) to pinpoint features that were differentially abundant across conditions, including disease status, pre- and post-treatment, and sex.

To explore the functional capabilities of the fungal microbiome, we used the FunFun algorithm, which evaluates the gene content from fungal ITS data. While ITS regions do not encode functional genes, FunFun infers potential functions of fungal communities [20]. For each fungus it determines the relative abundance of proteins belonging to protein families which represent molecular functions as specified in the Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology (KO) database [21]. This allows us to identify functional differences linked to disease status or treatment effects.

With these goals in mind, we implemented a step-by-step analysis of fungal ITS sequence data, employing bioinformatics tools and statistical approaches to explore fungal diversity and abundance across treatment groups and conditions.

# Chapter 2

## Methods

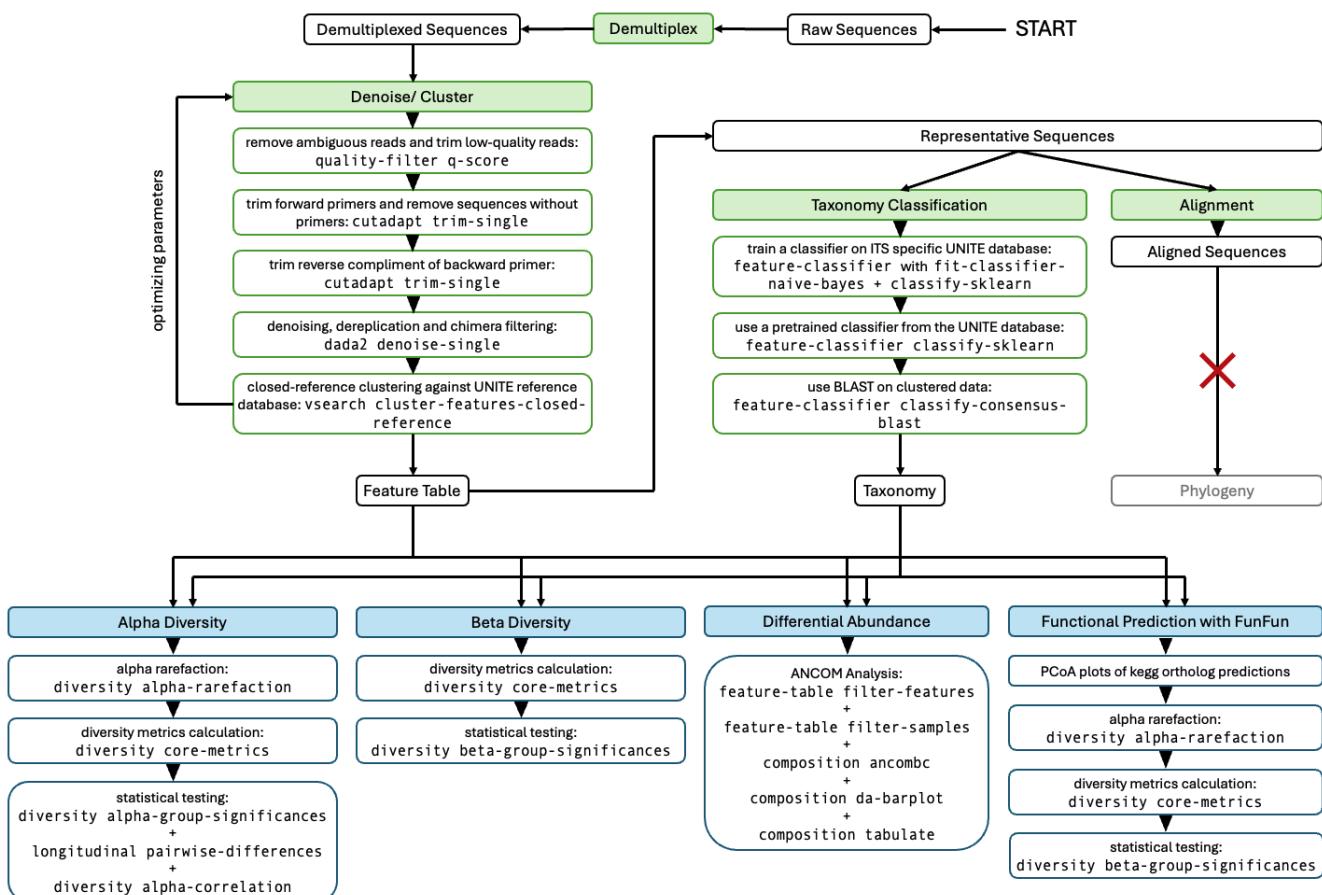
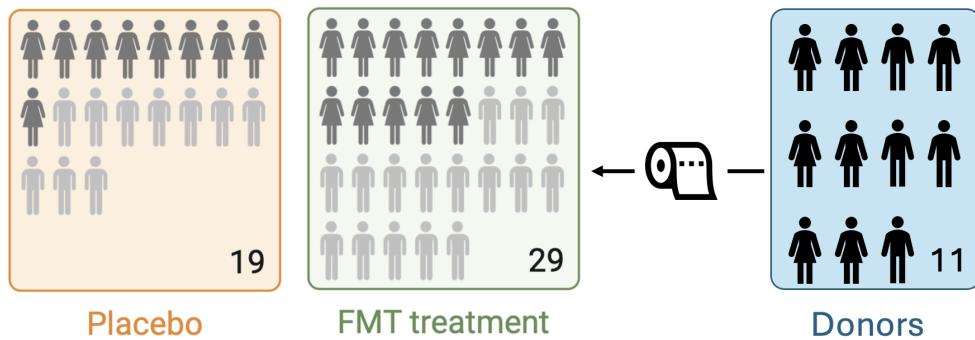


Figure 2.1: Overview of the used methods

### 2.1 ITS-1 Fungal Reads from Stool Samples

To analyze the effectiveness of Fecal Microbiota Transplant (FMT) treatment for the pandemic disease, a clinical trial was conducted. The trial included 48 patients, with 29 receiving FMT treatment and 19 receiving a placebo (Figure 2.2). Fecal samples used for the transplant were sourced from 11 donors at two time points. To evaluate the success of the treatment, patient samples were collected both before and



**Figure 2.2:** Visualization of the dataset including the distribution of sexes

after treatment. The study included patients aged 34 to 53 years, with an approximately equal distribution of males and females (see Figure 2.2). Most participants were Caucasian (70%), with smaller proportions identifying as Mediterranean (12%), Indian Subcontinental (8%), Arabic/Middle Eastern (6%), or Asian (4%). The dataset provided no information on the sex, age, or ethnicity of the donors.

For all samples, ITS-1 amplicons were generated and sequenced on an Illumina MiSeq machine. The sequence files consisted of single-end reads, demultiplexed prior to analysis, and were provided as QIIME2 artifacts (semantic type: SampleData [SequencesWithQuality]).

## 2.2 Quality Control and Denoising

To ensure optimal preprocessing for our single-end Illumina reads, various workflows were tested and evaluated based on their ability to improve the accuracy of subsequent taxonomy classification and minimize unassigned reads. Following this comparison, we implemented the following workflow for all downstream analyses.

Initial quality control was performed using the `quality-filter q-score` plugin in QIIME2, which trims sequences with more than three consecutive bases scoring below a quality score of 15. Reads containing ambiguous base calls were discarded entirely, ensuring cleaner data for further steps.

Following quality filtering, primers were removed from the sequences: CUTADAPT version 4.9 was used to trim the forward primers from the sequences. The `cutadapt trim-single` command was employed with an error rate of 0.2 and an overlap of 8 bases. Sequences that did not contain the primer were filtered out to ensure the removal of any non-target sequences. Additionally, the reverse complement of the backward primer was also trimmed using the same error rate of 0.2, with an overlap of 16 bases.

To distinguish true biological sequence variants from errors introduced during sequencing, the sequences were denoised. Denoising was applied using the `denoise-single` function from DADA2 [22], which performs denoising, dereplication, and chimera filtering simultaneously. The default parameters were applied, and no further length filtering was performed.

Afterwards, sequences were additionally clustered against the UNITE reference database [23] using VSEARCH [24]. VSEARCH implements distance-based greedy clustering around centroids and was run via `vsearch cluster-features-closed-reference` inside the qiime environment with 90% sequence identity. The UNITE database version 10 containing all eukaryotic sequences with 99% identity

was chosen as reference.

During the analysis, different preprocessing approaches were compared. We optimized the error rates for trimming of the primers, the number of consecutive bases with too low quality scores during quality filtering, and sequence trimming during denoising. We also checked whether we are dealing with multi-orientation reads. The best possible parameters were selected and applied as described above.

## 2.3 Taxonomy Classification

With the clustered DADA2 denoised data, we used QIIME2's feature-classifier classify-consensus-blast tool to get the taxonomy classification [25]. The reference taxonomy and reference sequences came from the UNITE database, version 10 with all eukaryotes, 99% identity and no singletons. These were cleaned to remove sh tags corresponding to hypothesized species. The blast used a 90% identity proportion for matching, and a query coverage of 70%. After this classification, the taxonomy was further cleaned by removing non fungal species and unassigned taxonomies, such as plants and metazoa in order to focus on the fungal data. Clustering and blast taxonomy classification were performed on the ETHZ Euler HPC cluster.

### 2.3.1 Alternative Taxonomy Classification using Naive Bayes Classifier

Before applying BLAST consensus classification, classification of taxonomy using a naive Bayes classifier via the SCIKIT-LEARN package was attempted. Initially, a custom classifier was trained using `qiime feature classifier-fit-classifier-naive-bayes`. The same reference taxonomy and sequences from the UNITE database (version 10 with all eukaryotes, 99% identity and no singletons) as in the BLAST classification were obtained and cleaned to remove sh tags. The trained classifier was then evaluated with `qiime rescript evaluate-fit-classifier` and `qiime rescript evaluate-taxonomy`. As comparison, a pre-trained classifier was obtained from *colinbrislawn*'s github as suggested in the qiime forum. The pre-trained classifier was also trained on sequences of all eukaryotes from the UNITE database with 99% identity and no singletons. For additional comparison, we also checked whether a classifier trained with 97% identity or a dynamic classifier could improve the performance.

These classifiers were then applied to our data using `qiime feature-classifier classify-sklearn`. `qiime taxa barplot` was used to evaluate the resulting taxonomy classifications.

## 2.4 Diversity Analysis

A characteristic of the ITS marker gene is its higher sequence variability compared to other commonly used markers. While this facilitates more accurate taxonomic identification, it also leads to highly unreliable multiple sequence alignment of ITS sequences across evolutionarily distant groups of fungi. This, in turn, results in unreliable phylogenetic trees and prohibits the use of phylogenetic distance metrics for analyzing alpha or beta diversity [17]. Additionally, our taxonomy classification showed a high proportion of "unclassified" taxa for a few samples, further contributing to the unassigned branches in the phylogenetic tree generated using MAFFT and FASTTREE. These limitations confirmed the unsuitability of phylogeny-based diversity metrics, such as FAITH's PHYLOGENETIC DIVERSITY and UNIFRAC, for our dataset.

## 2.4.1 Alpha Diversity

Alpha diversity refers to the diversity within individual samples and provides insights into the richness and evenness of fungal communities in the gut. Given the high variability in ITS sequences and the inability to construct reliable phylogenetic trees, we used non-phylogenetic alpha diversity metrics for this analysis.

### Data Preparation and Rarefaction

To ensure comparability across samples with varying sequencing depths, rarefaction analyses for all three metrics were performed using the `qiime diversity alpha-rarefaction` command, with a maximum depth of 10,000 sequences per sample. This approach allowed us to evaluate how sequencing depth influences alpha diversity estimates and to determine an appropriate sampling threshold. The upper resulting rarefaction plot assessed whether diversity metrics stabilized (leveled out) at specific sequencing depths, while the lower plot indicated how many samples were retained at each rarefaction depth. This guided the selection of a sequencing depth that maximized data retention while ensuring reliable diversity estimates. Rarefaction plots were generated for Shannon's Diversity Index and Observed Features, using `disease_subgroup` (FMT, Placebo, Donor) and `time_point` (pre- and post-treatment for patients; t1 and t2 for donors) as metadata categories (see Results Figure 3.3).

### Diversity Metrics Calculation

Alpha diversity metrics were calculated using the `qiime diversity core-metrics` command at the selected rarefaction depth of 3,000 sequences per sample (see Results 3.3.1). The metrics calculated included Shannon's Diversity Index, Pielou's Evenness, and Observed Features:

Shannon's diversity index  $H'$  is a measure of the entropy of a sample. It quantifies community diversity by incorporating both richness, which is the total number of unique features, and evenness, which reflects how evenly the features are distributed within a sample. Mathematically, it is defined as:

$$H' = - \sum_{i=1}^N p_i \ln(p_i) \quad (2.1)$$

where  $p_i$  represents the proportion of sequences belonging to the  $i$ -th feature, and  $N$  is the total number of observed features. Higher values of  $H'$  indicate greater diversity [18].

Pielou's Evenness  $J'$  measures the uniformity of a feature distribution by averaging the entropy  $H'$  (Shannon's Diversity Index) of a sample over the total number of observed features  $N$ . It thereby doesn't take the richness of a sample into account:

$$J' = \frac{H'}{\ln(N)} \quad (2.2)$$

Observed Features is a straightforward measure of community richness, representing the total number of unique features identified in a sample. Unlike other metrics, it does not account for the abundance or distribution of features, focusing only on presence. These metrics were analyzed to identify relationships with categorical metadata, such as time point, sex, and disease status.

## Statistical Testing

To determine whether there were significant differences in alpha diversity across different groups, statistical tests were applied using both non-parametric and longitudinal approaches:

The non-parametric Kruskal-Wallis test was performed with the `qiime diversity alpha-group-significance` command to evaluate diversity differences across categorical groups, such as sex or disease status. This method compares group medians without assuming normal data distribution [26], making it suitable for our data [26].

To analyze changes in diversity over time between paired samples, such as pre- and post-treatment, we used the `qiime longitudinal pairwise-differences` command. For this analysis we used the Wilcoxon signed-rank test, accounting for repeated measures by pairing data within individuals. It evaluates whether the median of the differences between paired observations deviates from zero [27]. Sex-related effects on alpha diversity were addressed by normalizing Shannon diversity values for male and female patients separately. Z-scores were calculated for each individual using the formula:

$$Z = \frac{x_i - \mu}{\sigma} \quad (2.3)$$

where  $x_i$  is the individual's Shannon diversity value,  $\mu$  is the mean for the respective sex group, and  $\sigma$  is the standard deviation for that group. These z-scores were added as a new column (`shannon_sex_bias`) in the metadata to control for any potential bias related to sex differences.

## Alpha Correlation Analysis

To explore the relationship between alpha diversity and disease intensity, we performed correlation analyses between the alpha diversity metrics and the `puns_per_hour` metric (a measure of disease severity). Analyses were conducted separately for pre-treatment and post-treatment time points to identify potential associations between diversity and symptom intensity.

### 2.4.2 Beta Diversity

In contrast to alpha diversity which determines the diversity within a sample, beta diversity determines the diversity in between samples. Because phylogenetic metrics are unreliable for ITS data, only the non-phylogenetic metrics, Bray-Curtis dissimilarity and Jaccard index were considered in this study.

The Jaccard Index  $J$  [28] measures similarity between two samples based on the presence or absence of microbial species. It compares the number of species that are contained in both samples with the number of species that are contained in the union of the two samples. Here  $A$  describes the set of unique species contained in the first sample and  $B$  the set of unique species in the second sample. A Jaccard Index close to 1 indicates high similarity, while a Jaccard Index close to 0 means that the samples differ a lot:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (2.4)$$

Bray-Curtis dissimilarity  $BC$  takes presence as well as abundance of species into account. It sums up the minimal abundance of each species comparing two samples, doubles it and divides it by the total sum of counts of all species. To get a measure of dissimilarity, it calculates 1 minus the obtained fraction [29].

Values close to 1 indicate high dissimilarity while values close to 0 show high similarity. In the following formula,  $A$  and  $B$  are the sets of all species in the two samples,  $a_i$  and  $b_i$  are the abundance of each species in the two samples and  $S$  is the total number of species:

$$BC(A, B) = 1 - \frac{2 \sum_{i=1}^S \min(a_i, b_i)}{\sum_{i=1}^S (a_i + b_i)} \quad (2.5)$$

To assess beta diversity metrics (Bray-Curtis dissimilarity and Jaccard Index), the `qiime diversity core-metrics` command was used with a sampling depth of 3000, consistent with the approach for alpha diversity. This process also generated principal coordinate data, which were visualized as 3D plots using EMPEROR [30] or as 2D principal coordinate analysis (PCoA) plots. To explore potential correlations between sample groups and beta diversity, the first and second principal coordinates were plotted against each other, with each data point colored based on sample categories such as disease status, sex, and ethnicity.

## Statistical Testing

Permutational multivariate analysis of variance (PERMANOVA) [31] is a permutation-based statistical test, which can be used to evaluate whether groups of samples significantly differ from each other based on beta-diversity metrics. It computes variability within a group and between groups and calculates a pseudo F-statistic based on this.

Firstly, sum of squared distances for all samples  $SS_T$  and for all samples within the same group  $SS_W$  are determined.  $N$  defines the total number of samples,  $p$  the number of groups,  $n_k$  the number of samples per group,  $d_{ij}$  the beta diversity between sample  $i$  and sample  $j$  and  $\delta_{ij}$  a delta function which is 1 when the samples  $i$  and  $j$  belong to the same group.

$$SS_T = \frac{1}{N} \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij}^2 \quad (2.6)$$

$$SS_W = \sum_{k=1}^p \frac{1}{n_k} \sum_{i \in G_k} \sum_{\substack{j \in G_k \\ j > i}} d_{ij}^2 \quad (2.7)$$

The between sample variability  $SS_A$  is then determined by calculating  $SS_A = SS_T - SS_W$  and used to calculate the pseudo F-statistics:

$$F = \frac{\text{Between-group variability}}{\text{Within-group variability}} = \frac{\frac{SS_A}{p-1}}{\frac{SS_W}{N-p}} \quad (2.8)$$

To determine statistical significance, the test randomly shuffles group labels and calculates an F-statistic for every draw. This yields a null distribution which can be compared to the actual F-statistic and a p-value can be determined:

$$p = \frac{\text{Number of permutations with } F^\pi \geq F}{\text{Total number of permutations}} \quad (2.9)$$

PERMANOVA tests were conducted using `qiime diversity beta-group-significance`. This assesses significant differences across all groups within a metadata category and performs pairwise comparisons between groups. In this study, beta diversity differences were evaluated based on Bray-Curtis dissimilarity and the Jaccard Index for various metadata categories, including sex, ethnicity, disease status (healthy, healed, diseased), and treatment response subgroups (FMT response, FMT no response, placebo no response).

## 2.5 Differential Abundance Analysis

To identify fungal features that were significantly different between groups, we conducted a differential abundance analysis. First, we examined the feature table, where rows represented samples, columns represented features (ASVs), and the values contained feature frequencies. Using the QIIME2 Python API, we explored the data for sparsity, which refers to the presence of many zero values, and checked whether the data followed a normal distribution. For this, we applied the SHAPIRO-WILK TEST [32] to the three most abundant features and ten randomly selected ASVs, recording whether each sample's features showed normality. The Shapiro-Wilk test computes a test statistic  $W$ , where values close to 1 indicate that the data are normally distributed. The formula for the Shapiro-Wilk test is:

$$W = \frac{(\sum_{i=1}^n a_i x_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.10)$$

Where  $x_i$  are the ordered samples and  $\bar{x}$  their mean,  $a_i$  are constants determined by the expected values of the order statistics of the normal distribution, and  $n$  is the sample size. A significance threshold  $\alpha$  of 0.05 was used to determine if the data significantly deviated from normality. If the p-value was less than  $\alpha$ , the null hypothesis of normality was rejected, indicating that the data were not normally distributed.

To reduce false discoveries and improve resolution, features were filtered to retain those present at a minimum frequency of 25 and in at least six samples. This was achieved using the `qiime feature-table filter-features` command.

Differential abundance testing was conducted using the ANCOM-BC (Analysis of Composition of Microbiomes using Bias Correction) tool. It is a statistical method that compares the relative abundance of features (taxa) between multiple groups while accounting for compositionality of the data [33].

Comparisons were conducted between specific groups, including healed versus healthy samples, healthy versus puns samples, and pre- versus post-treatment groups. For the pre- and post-treatment comparisons, we further separated the data into FMT and placebo groups and distinguished between responders and non-responders. We again evaluated potential sex batch effects to ensure the robustness of the results. The feature table was filtered using the `qiime feature-table filter-samples` command, specifying which samples were included via the `--p-where` parameter, which was customized to specific metadata conditions.

The ANCOM-BC tool was used to determine which features were differentially abundant across groups. The reference group for each comparison was explicitly defined to enable accurate interpretation of results. This was performed using the `qiime composition ancombc`. To visualize the findings, we generated bar plots showing features that were enriched or depleted in each group using the `qiime composition da-barplot` command. Additionally, we created detailed tables listing the exact values for all tested features using the `qiime composition tabulate` command. The results were reviewed to identify fungal taxa that were significantly associated with disease status, treatment response, or other relevant factors.

## 2.6 Functional Predictions

Natural fungal communities exhibit diverse biochemical pathways due to varying gene content, influencing their functional capabilities [34]. These molecular functions can be related to diseases and are thus of interest for our study on the pandemic. Since whole genome sequencing of fungi is costly, technically challenging, and wasn't conducted in this study, gene content profiles can't be inferred directly. Unlike bacterial 16S-sequencing data, the ITS sequencing data obtained in this study can't be used as input for complex predictive models like PICRUST2 [35] because the number of precisely annotated fungi is not high enough.

That is why FUNFUN was used, a modified K-nearest neighbors (KNN) model, which exploits the assumption that functional capabilities of fungi are in strong correlation with their genetic-related organisms. FUNFUN uses fungal genomes with annotated gene content as reference and calculates k-mer relative abundances vectors from their ITS regions. To predict the gene content profile of a target sequence, a cosine distance to every reference sequence is determined. The KNNs within an  $\epsilon$ -neighborhood are used to calculate the relative abundance of genes belonging to different KEGG orthology groups for every target fungus [20].

To obtain a matrix containing relative abundances of KEGG orthology groups for every fungal OTU, the command `funfun -its dna-sequences.fasta -type its1` from the FUNFUN package version 0.1.15 was applied. Since FUNFUN provides fungus-specific abundance estimates, its results were aggregated to obtain sample-specific insights by calculating weighted averages across all sequences present in a sample, using their respective counts as weights. Subsequently, the data was re-normalized to ensure comparability across samples.

To explore potential differences in functional abundances between sample categories, principal component analysis (PCA) was performed using the SCIKIT-LEARN package. PCA dimensionality-reduced KEGG orthology group abundances were visualized and colored by sex, ethnicity, subgroup response, puns per hour (pre- and post-treatment), and disease status.

To statistically test for functional differences between sample groups, beta diversity metrics (Jaccard Index and Bray-Curtis dissimilarity) were determined based on relative abundances of KEGG orthology groups. Subsequent PERMANOVA tests were applied to look for significant differences. This was done by repeating the beta-diversity analysis workflow via QIIME based on functional abundances instead of sequence counts.

# Chapter 3

## Results

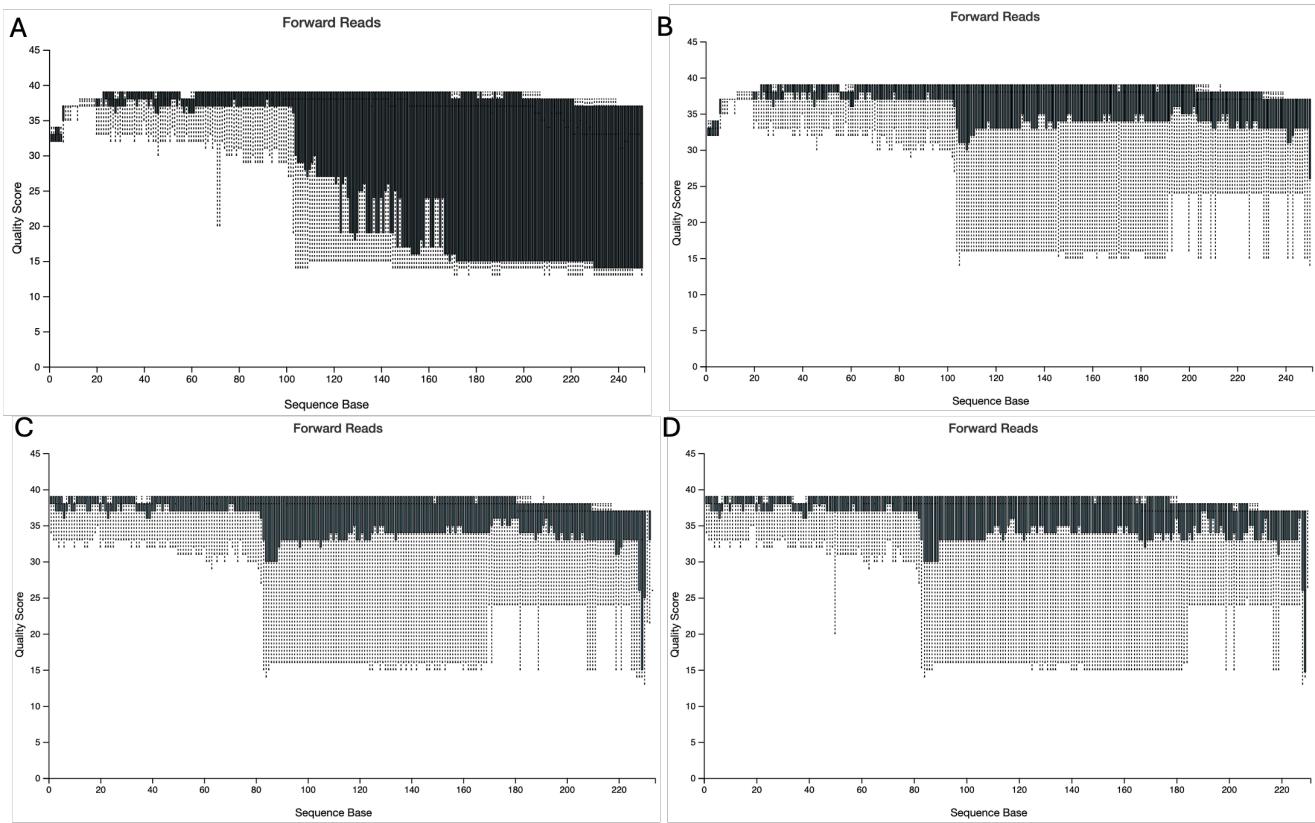
### 3.1 Quality Control and Denoising

A closer inspection of the demultiplexed data (used as the starting point of our study) revealed a total of 7,977,313 reads across all samples, with an approximately tenfold variation in reads per sample and a median read count of about 44,000. Each read had a fixed length of 250 base pairs, with a 25th percentile quality score of 15 and a median quality score of 35. This indicates that subsequent quality filtering and preprocessing was necessary. After quality filtering and primer trimming, we obtained a 127% improvement of the 25 percentile quality score by losing 2.3% of our reads (table 3.1). Figure 3.1 visualizes the increase in base quality.

A minimum of 70.5% of amplicon sequence variants (ASVs) per sample were retained after denoising and chimera removal. However, the number of ASVs across samples and their feature frequencies varied considerably (Figure A.1).

**Table 3.1: Evaluation of preprocessing.** Preceding our analysis, different preprocessing steps were applied to the sequencing reads. The following table compares total number of reads, minimum and median number of reads (ASVs after denoising) per sample, 25th percentile and median read length and the 25th percentile and median base quality score after each preprocessing step. We received demultiplexed reads as the starting point of this study. Afterwards, quality filtering was applied and forward primers were cut off at the start of the sequences, reverse complements of backward primers were cut off at the end of the sequences and denoising, dereplication and chimera-removal were applied using DADA2.

	total reads	number of features		read length		quality score	
		minimum	median	25 %	median	25 %	median
Demultiplexed sequences	7 977 313	11 337	44 076	250	250	15	35
Quality filtered	7 898 931	11 337	41 945	174	250	34	38
Forward primer trimmed	7 875 193	11 309	41 858	129	149	34	37
Backward primer trimmed	7 792 119	11 309	41 858	129	149	33	38
Dada2 denoise-single	7 113 576	9 851	37 077	75	141	-	-
Closed-reference clustering	4 464 876	475	14 925	197	228	-	-
Filtering out non-fungal taxa	4 011 422	462	12 135	194	228	-	-

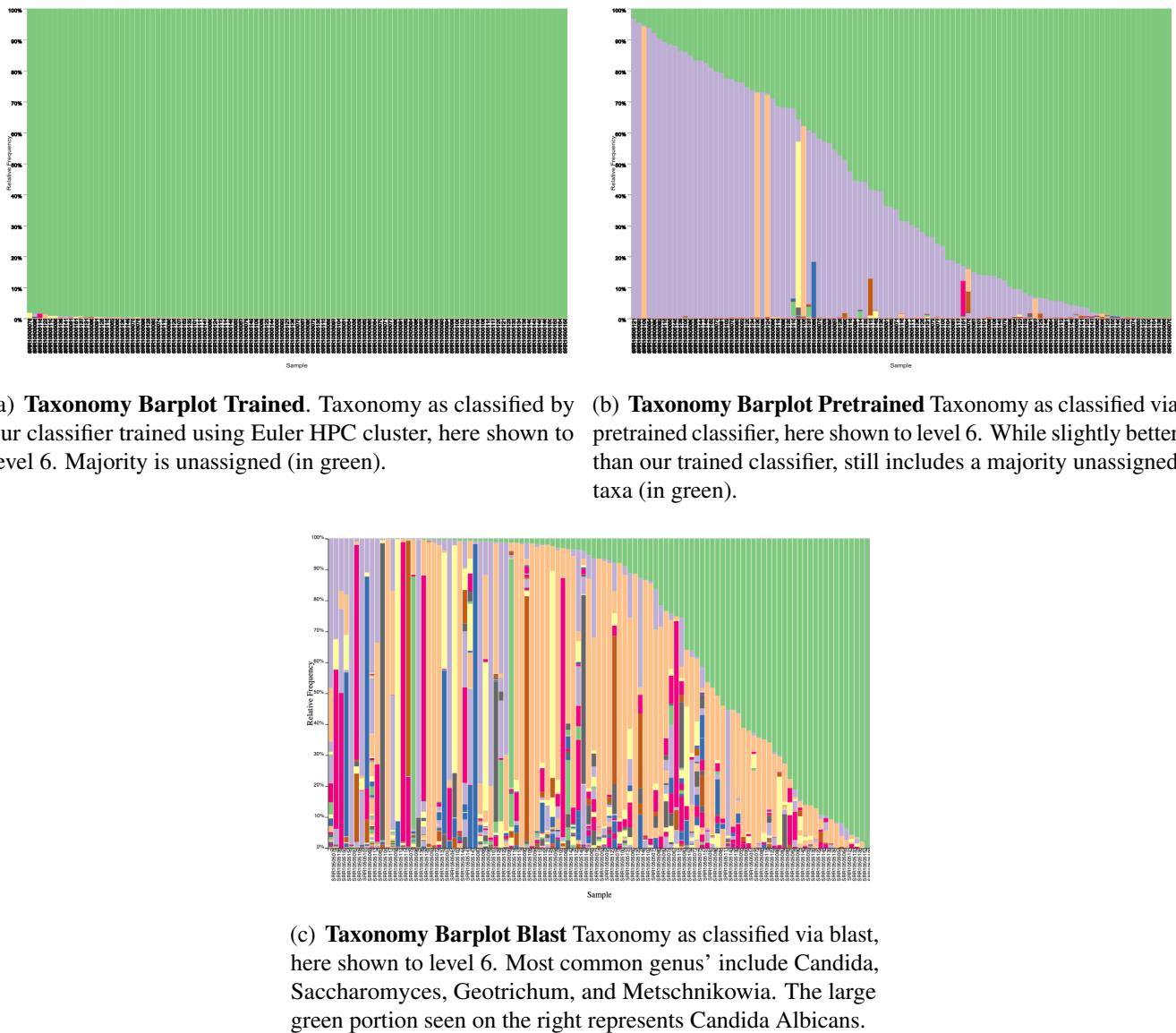


**Figure 3.1: Boxplots visualizing base quality scores over different read positions.** Quality scores are shown for the demultiplexed reads (A), after quality filtering (B), after trimming of forward primers (C) and after trimming of the reverse complement of backward primers (D).

Subsequent closed-reference clustering further reduced the number of sequences, with the minimum number of features per sample dropping to 475. This process, however, increased the median read length to 228 base pairs. Additional filtering of non-fungal and unassigned reads caused only a minor data loss. In total, 50.3% of the demultiplexed reads were retained after preprocessing (Table 3.1).

## 3.2 Taxonomy Classification

First, the classifier we trained on the UNITE database produced a majority unassigned taxonomies as seen in figure 3.2(a). Next, the pre-trained classifier produced slightly better results, as visible in figure 3.2(b), but still left the majority unassigned. Finally, taxonomy classification was completed via Blast using Qiime's feature-classifier classify-consensus-blast. Initial Blast results are similar to figure 3.2(c), with a few additional non fungi species that were later filtered out. Interestingly, a few samples had higher levels (around 15%, with one sample that had a majority classification) of *Blastocystis* sp., an enteric anaerobic parasite which often reaches levels of over 5% of the population in developed countries, and often affects immunocompromised patients [36]. Another non fungal target was the family of Solanaceae, commonly known as nightshades, which may have been due to patient diets. The most common fungal species were *Candida albicans*, a common member of human gut, and *Geotrichum candidum*, another mold like yeast common to the human gut microbiome [37].



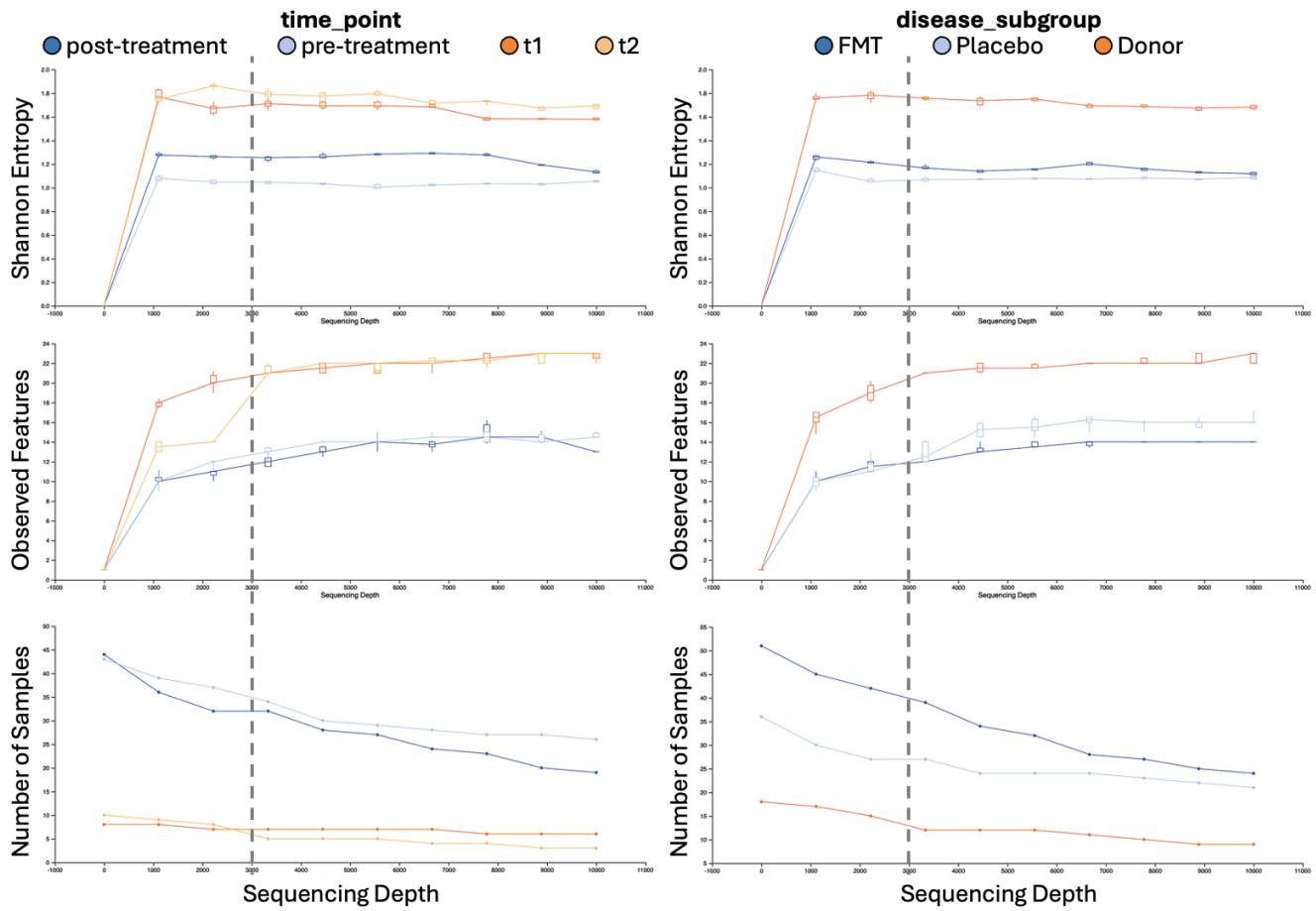
**Figure 3.2:** Comparison of Taxonomy Barplots: (a) Classifier trained on Euler HPC cluster, (b) Pretrained classifier, and (c) BLAST classification. All barplots are shown to level 6.

## 3.3 Alpha Diversity

### 3.3.1 Rarefaction Analysis

The rarefaction analysis demonstrated that Shannon Entropy values stabilized at approximately 3,000 reads per sample. At this depth, diversity metrics leveled off, indicating that sequencing depth was sufficient to provide reliable estimates of community richness and evenness. Observed Features, on the other hand, required a greater sequencing depth of approximately 6,000 reads per sample to fully stabilize. However, choosing a sequencing depth of 6,000 reads would have resulted in a substantial loss of samples (as shown in the lower rarefaction plot of Figure 3.3). Given the trade-off between sequencing depth and sample retention, 3,000 reads per sample was selected as the threshold for subsequent analyses. This decision

prioritized Shannon Entropy, a metric more critical for understanding community richness and evenness in this context.



**Figure 3.3: Rarefaction plots** for Shannon Entropy, Observed Features, and the number of samples retained. Stabilization of Shannon Entropy around 3,000 reads per sample guided the selection of this sequencing depth, balancing diversity estimation accuracy and sample retention. Plots are stratified by *time\_point* (pre-treatment, post-treatment, t1, and t2) and *disease\_status* (FMT, Placebo, and Donor).

### 3.3.2 Comparison of Alpha Diversity Metrics

Statistical tests were then applied to assess differences in alpha diversity across key metadata categories, including *sex* and *disease status*.

#### Initial Statistical Testing

Initial Kruskal-Wallis tests on Shannon entropy revealed significant differences in diversity across categories of *sex* (male, female, and unknown). Pairwise comparisons confirmed significant differences, specifically between the "unknown" category and both males and females (Table 3.2). However, as all donors were classified under "unknown" sex while patients had known sexes, this effect is likely unrelated to biological sex differences. To investigate further, samples with "unknown" sex were excluded, and no significant differences were observed ( $p = 0.168$ ). Moreover, sex differences for Pielou's Evenness and Observed Features were not significant ( $p = 0.126$  and  $p = 0.513$ , respectively).

For the *disease status* variable, significant differences in diversity were observed between healthy controls and patients with the disease (table 3.2). However, the existing categories did not differentiate between healed and unaffected patients post-FMT treatment, limiting interpretability. To address this, a new metadata category was introduced, distinguishing between healthy, healed, and diseased patients. Reanalysis using the updated metadata confirmed a significant association between Shannon entropy and disease status ( $p = 0.001$ ). Pairwise comparisons revealed significant differences only between healthy controls and diseased patients, even though healed patients appeared more similar to healthy individuals than to diseased patients (table 3.2).

Similar trends were observed for Pielou's Evenness and Observed Features. Disease status was significantly associated with these metrics across groups ( $p = 0.027$  and  $p = 0.012$ , respectively). However, pairwise comparisons reached significance only between healthy controls and diseased patients for Observed Features ( $p = 0.012$ ). These findings indicate that fungal diversity is more strongly associated with disease status than with sex, particularly in the context of the Pandemic.

**Table 3.2:** Kruskal-Wallis Test Results for Shannon Entropy (summarized from Figure A.2)

Metadata	Condition	Kruskal-Wallis (all groups)		Kruskal-Wallis (pairwise)				
		H	p-value	group 1	group 2	H	p-value	q-value
sex	unknown, female, male	11.68303	0.003	unknown	female	5.33597	0.021	0.031
				unknown	male	11.91740	0.001	0.002
	female, male	1.90311	0.168	female	male	1.90311	0.168	0.168
group	Healthy, Puns	9.69851	0.002	Healthy	Puns	9.69851	0.002	0.002
disease_status	Healthy, Puns, Healed	13.33923	0.001	Healed	Healthy	1.33929	0.247	0.247
				Healed	Puns	4.04387	0.044	0.066
				Healthy	Puns	10.68009	0.001	0.003

One limitation of the analyses above is the inclusion of multiple time points per subject, which violates the Kruskal-Wallis test assumption of independent samples. This lack of independence could mask meaningful patterns specific to individual time points. To mitigate this issue, pairwise difference tests and longitudinal analyses were performed to account for temporal dependencies and provide a more accurate interpretation of the data.

### Pairwise Differences Across Time Points to evaluate Treatment Response

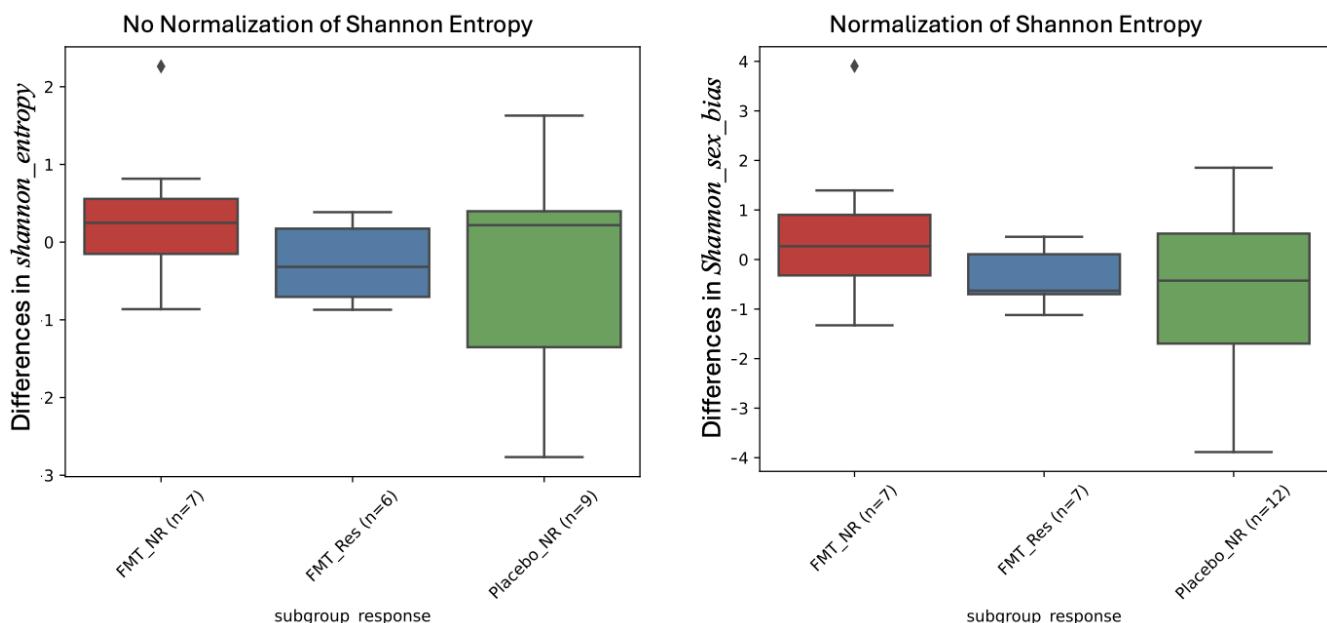
Pairwise difference tests were conducted to assess the potential influence of alpha diversity on treatment outcomes throughout the treatment period (pre- and post-treatment). Notably, 31% (15 out of 48) of patients were excluded from the analysis because only one sample, either pre- or post-treatment, was available.

While differences in alpha diversity between sexes were no longer statistically significant after excluding samples with unknown sex, a visible trend persisted in the data. To minimize potential batch effects related to sex, Shannon diversity values were normalized by sex.

To enhance the evaluation of treatment response, a new metadata category, *subgroup\_response*, was introduced, which combined treatment group (placebo or FMT) with clinical response (response or no

response). However, rarefaction filtering resulted in the exclusion of the post-treatment sample from the only placebo patient who responded to treatment, necessitating its removal from subsequent analyses.

Pairwise difference tests comparing pre- and post-treatment samples revealed no significant differences in alpha diversity for all subgroup response groups and for Shannon entropy, Pielou's Evenness, and Observed Features. Similarly, Kruskal-Wallis tests examining changes in alpha diversity throughout treatment across all subgroup response groups showed no significant associations for Shannon entropy ( $p = 0.508$ ), Pielou's Evenness ( $p = 0.171$ ), or Observed Features ( $p = 0.790$ ). Pairwise group comparisons across all metrics also failed to reach significance. These results suggest that alpha diversity does not play a role in determining the outcome of FMT treatment. Importantly, normalization for sex differences did not alter these findings, confirming that the observed diversity patterns were not influenced by sex-related biases (figure 3.4).



**Figure 3.4:** Pairwise differences (pre-treatment vs. post-treatment) of Shannon Entropy before and after normalization for sex differences

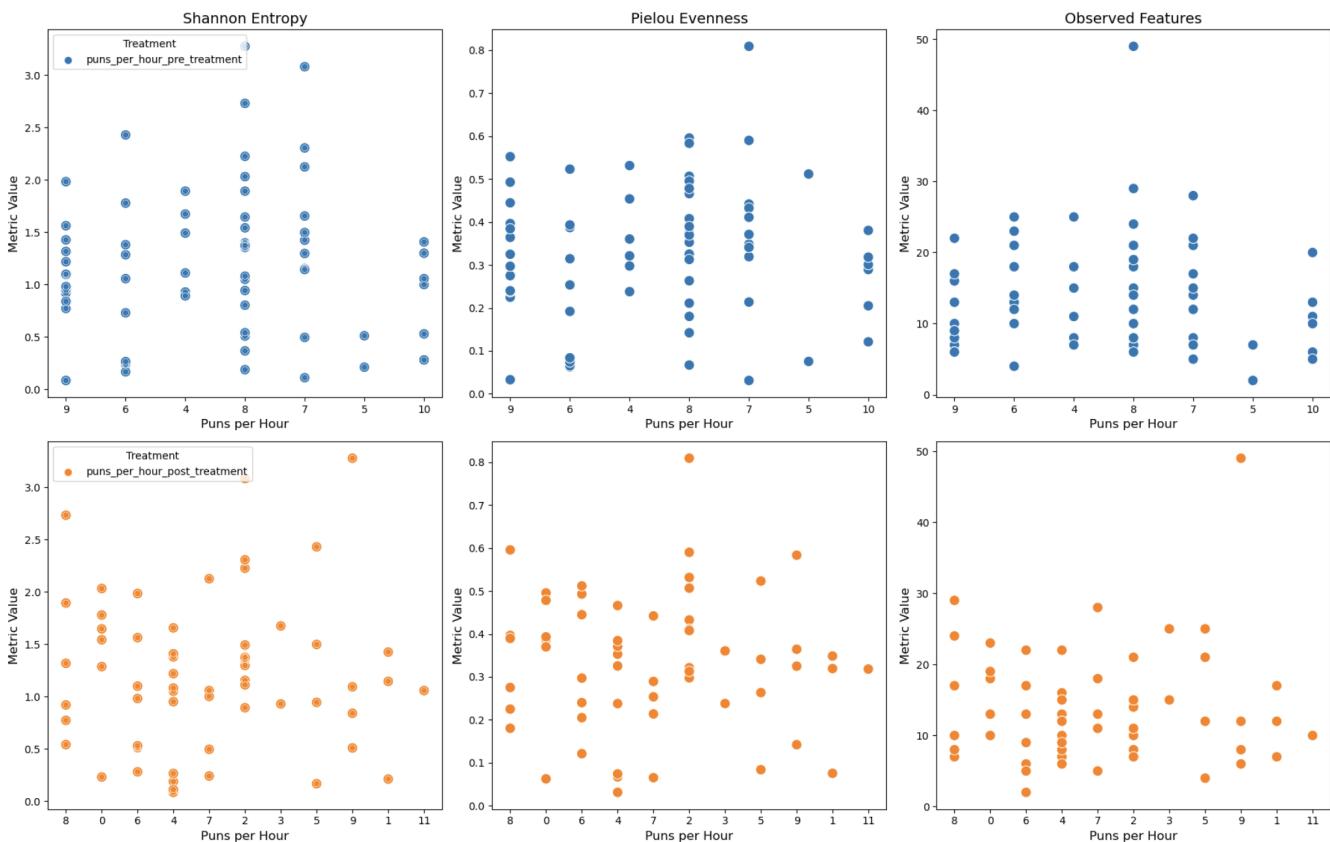
Even though none of the results reached statistical significance, a subtle difference in changes in alpha diversity between time points is evident in Figure 3.4. Specifically, the placebo no-response and FMT no-response groups appear more similar to each other than to the FMT response group. Interestingly, patients who responded to treatment showed a reduction in alpha diversity after treatment, whereas patients without a response exhibited an increase in alpha diversity. However, figure A.3 shows that patients responding to treatment showed a higher alpha diversity even before treatment compared to the other groups. These patterns, while not statistically significant, suggest potential trends that merit further investigation.

### 3.3.3 Alpha Diversity Correlation with Disease Intensity

To investigate the relationship between alpha diversity and disease intensity, correlation analyses were performed between the alpha diversity metrics and the *puns\_per\_hour* metric, which measures disease

severity. Analyses were conducted separately for pre-treatment and post-treatment time points (Figure 3.5).

The results demonstrated no significant correlations between any of the alpha diversity metrics and disease intensity pre-treatment ( $p(\text{shannon}) = 0.808$ ,  $p(\text{evenness}) = 0.789$ ,  $p(\text{observed features}) = 0.227$ ). Also for the post-treatment time point, no statistically significant values were observed ( $p(\text{shannon}) = 0.115$ ,  $p(\text{evenness}) = 0.218$ ,  $p(\text{observed features}) = 0.581$ ). A summary of the correlation results, including test statistics and p-values, is provided in Table 3.3.



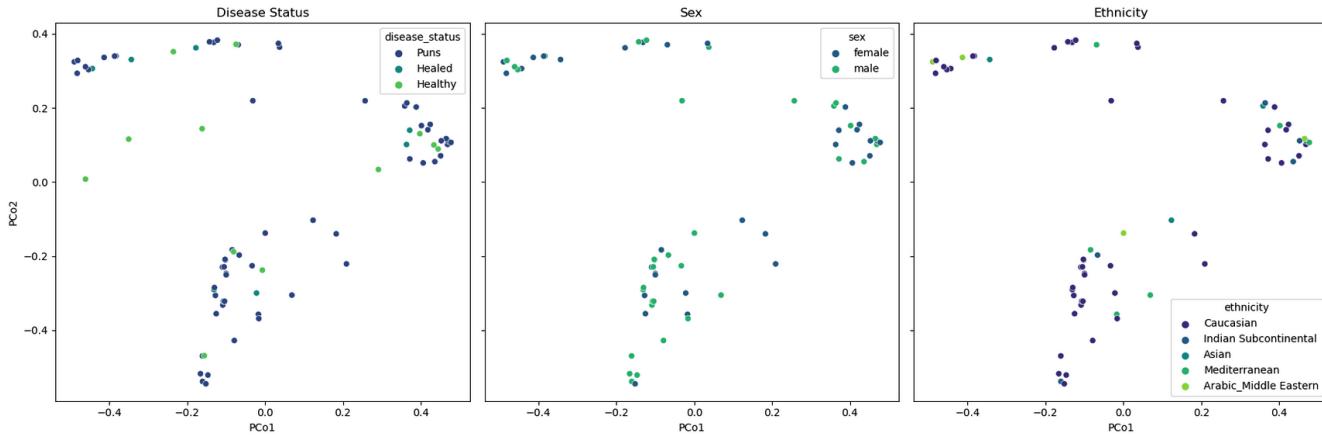
**Figure 3.5:** Alpha diversity correlation with disease intensity

**Table 3.3:** Correlation between alpha diversity metrics and disease intensity (Puns per Hour)

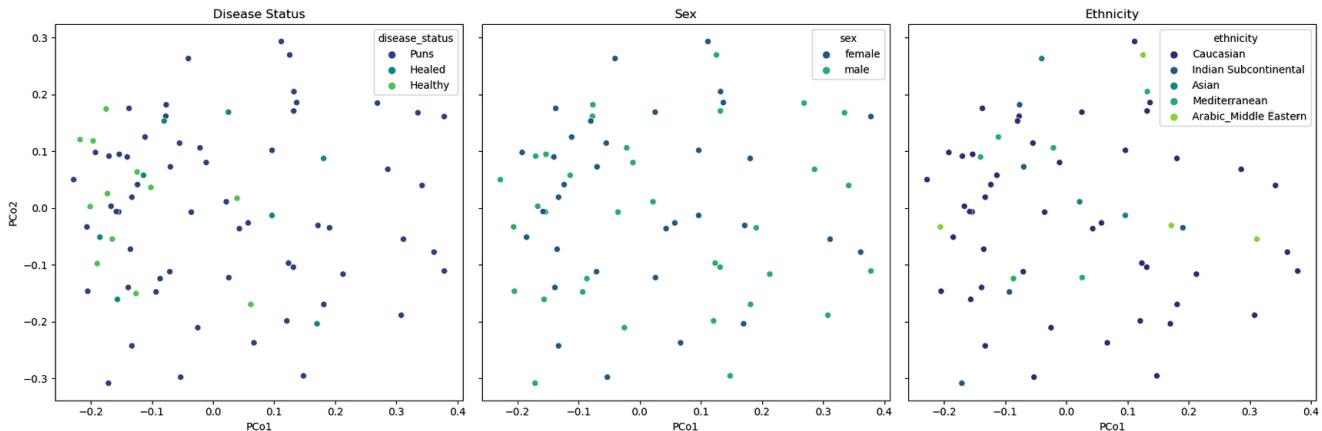
Metric	Pre-treatment		Post-treatment	
	Test Statistic	p-value	Test Statistic	p-value
Shannon Entropy	-0.0303	0.8078	-0.2038	0.1152
Pielou's Evenness	-0.0333	0.7888	-0.1601	0.2176
Observed Features	-0.1497	0.2266	-0.0721	0.5808

## 3.4 Beta Diversity

The analysis of beta diversity metrics aimed to identify potential patterns related to sample categories such as disease status, sex, and ethnicity. By looking at the PCoA plots for Bray-Curtis dissimilarity (Figure 3.6) and the Jaccard Index (Figure 3.7), no clear correlation was observed for either diversity metric. This finding suggests that there are no batch effects related to sex or ethnicity in the dataset while also indicating no significant differences in beta diversity between healthy, healed, and diseased patients.



**Figure 3.6: 2D visualization of PCoA using Bray-Curtis dissimilarity metric.** The plots are colored after disease status, sex and ethnicity. No correlation between beta-diversity and any of the sample categories is visible.



**Figure 3.7: 2D visualization of PCoA using Jaccard Index similarity metric.** The plots are colored after disease status, sex and ethnicity. No correlation between beta-diversity and any of the sample categories is visible.

To statistically evaluate these findings, the PERMANOVA tests were conducted (table 3.4, table 3.5). No significant differences in beta diversity were found for sex or ethnicity in either metric, ruling out potential batch effects. Additionally, no significant differences were observed between treatment response groups, indicating that FMT treatment and placebo responses did not differ in terms of beta diversity. While Bray-Curtis dissimilarity revealed no significant differences in disease status, the Jaccard Index showed a significant beta diversity difference between healthy and diseased patients (table 3.5, figure A.4).

Interestingly, healed patients exhibited greater dissimilarity from healthy individuals than from diseased individuals, suggesting that their microbiomes may not have fully recovered despite the lack of statistical significance.

**Table 3.4:** PERMANOVA test results for Bray-Curtis dissimilarity based on OTU sequence counts. Differences in beta diversity between sexes, ethnicities, disease statuses and subgroup responses post-treatment.

Metadata	Condition	PERMANOVA (all groups)		PERMANOVA (pairwise)			
		pseudo-F	p-value	group_1	group_2	pseudo-F	p-value
sex	female, male	1.236	0.266	female	male	1.236	0.283
ethnicity	Arabic_Middle Eastern, Asian, Caucasian, Indian Subcontinental, Mediterranean	0.806	0.771	Arabic_Middle Eastern	Asian	0.506	0.711
				Arabic_Middle Eastern	Caucasian	0.780	0.590
				Arabic_Middle Eastern	Indian Subcontinental	1.387	0.241
				Arabic_Middle Eastern	Mediterranean	1.159	0.348
				Asian	Caucasian	0.539	0.868
				Asian	Indian Subcontinental	0.785	0.609
				Asian	Mediterranean	0.571	0.803
				Caucasian	Indian Subcontinental	0.986	0.437
				Caucasian	Mediterranean	0.908	0.494
				Indian Subcontinental	Mediterranean	0.434	0.949
disease status	healthy, healed, puns	0.598	0.883	healthy	healed	0.548	0.794
subgroup response	FMT_NR, FMT_Res, Placebo_NR	0.386	0.991	healthy	puns	0.607	0.797
				healed	puns	0.602	0.772
				FMT_NR	FMT_Res	0.233	0.997
				FMT_NR	Placebo_NR	0.243	0.995
				FMT_Res	Placebo_NR	0.656	0.738
							0.997

**Table 3.5:** PERMANOVA test results for Jaccard index similarity based on OTU sequence counts. Differences in beta diversity between sexes, ethnicities, disease statuses and subgroup responses post-treatment.

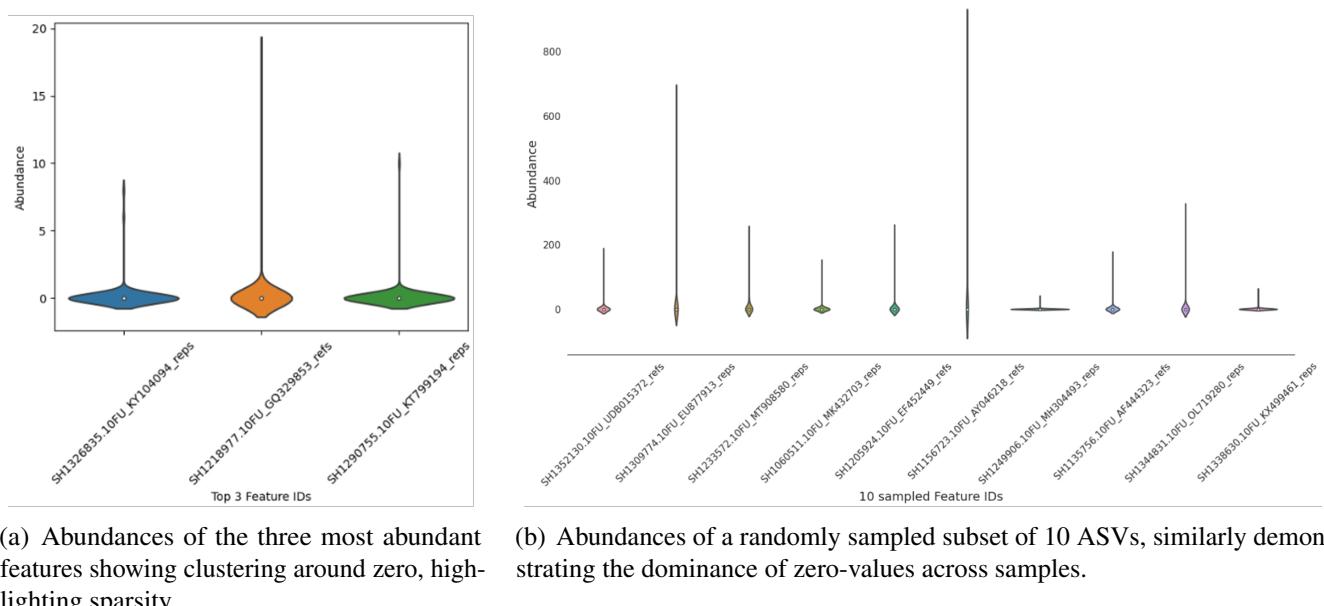
Metadata	Condition	PERMANOVA (all groups)		PERMANOVA (pairwise)			
		pseudo-F	p-value	group_1	group_2	pseudo-F	p-value
sex	female, male	0.876	0.701	female	male	0.876	0.700
ethnicity	Arabic_Middle Eastern, Asian, Caucasian, Indian Subcontinental, Mediterranean	0.920	0.767	Arabic_Middle Eastern	Asian	1.128	0.186
				Arabic_Middle Eastern	Caucasian	0.845	0.766
				Arabic_Middle Eastern	Indian Subcontinental	0.949	0.532
				Arabic_Middle Eastern	Mediterranean	0.847	0.855
				Asian	Caucasian	1.092	0.285
				Asian	Indian Subcontinental	1.147	0.269
				Asian	Mediterranean	0.866	0.742
				Caucasian	Indian Subcontinental	0.820	0.835
				Caucasian	Mediterranean	0.909	0.645
				Indian Subcontinental	Mediterranean	0.810	0.836
disease status	healthy, healed, puns	1.360	0.023	healthy	healed	1.251	0.077
subgroup response	FMT_NR, FMT_Res, Placebo_NR	0.887	0.804	healthy	puns	1.754	0.004
				healed	puns	0.976	0.501
				FMT_NR	FMT_Res	0.832	0.827
				FMT_NR	Placebo_NR	0.816	0.828
				FMT_Res	Placebo_NR	0.999	0.466
							0.828

## 3.5 Differential Abundance

### 3.5.1 Data Exploration and Normality Assessment

Initial exploration of the feature table revealed significant sparsity, with many values equal to zero, indicating limited representation of certain fungal features across samples. Violin plots of the top three abundant features and a random subset of ASVs confirmed that most feature abundances were concentrated around zero (figure 3.8). The Shapiro-Wilk test for normality ( $\alpha = 0.05$ ) revealed that none of the ASVs

exhibited normal distribution, with all results indicating "False" for normality. These findings confirm the dataset's sparse and non-normal nature. That is why we need statistical approaches, such as ANCOM, that are robust to such data characteristics.



**Figure 3.8:** Violin plots illustrating feature sparsity in the dataset.

### 3.5.2 ANCOM-BC Results

All the data discussed here are presented in Tables A.1, A.2, and A.3.

Comparing healthy individuals with healed patients, as well as healthy individuals with diseased patients, revealed no significantly abundant features at either the ASV or taxa level. This suggests a lack of clear fungal biomarkers associated with the disease.

For FMT responders, fungal communities were compared before and after treatment. However, no significant changes were detected, with all q-values equal to 1. As a control, fungal communities were also compared pre- and post-treatment for FMT non-responders and placebo non-responders. Similarly, no significantly abundant features were identified for FMT non-responders. For placebo non-responders *Trichosporonaceae* and *Bulleribasidiaceae* were significantly depleted after treatment (table A.3). The analysis for placebo responders could not be performed due to insufficient sample size, as only two samples were available. This small sample size rendered the model too complex, preventing the completion of the analysis.

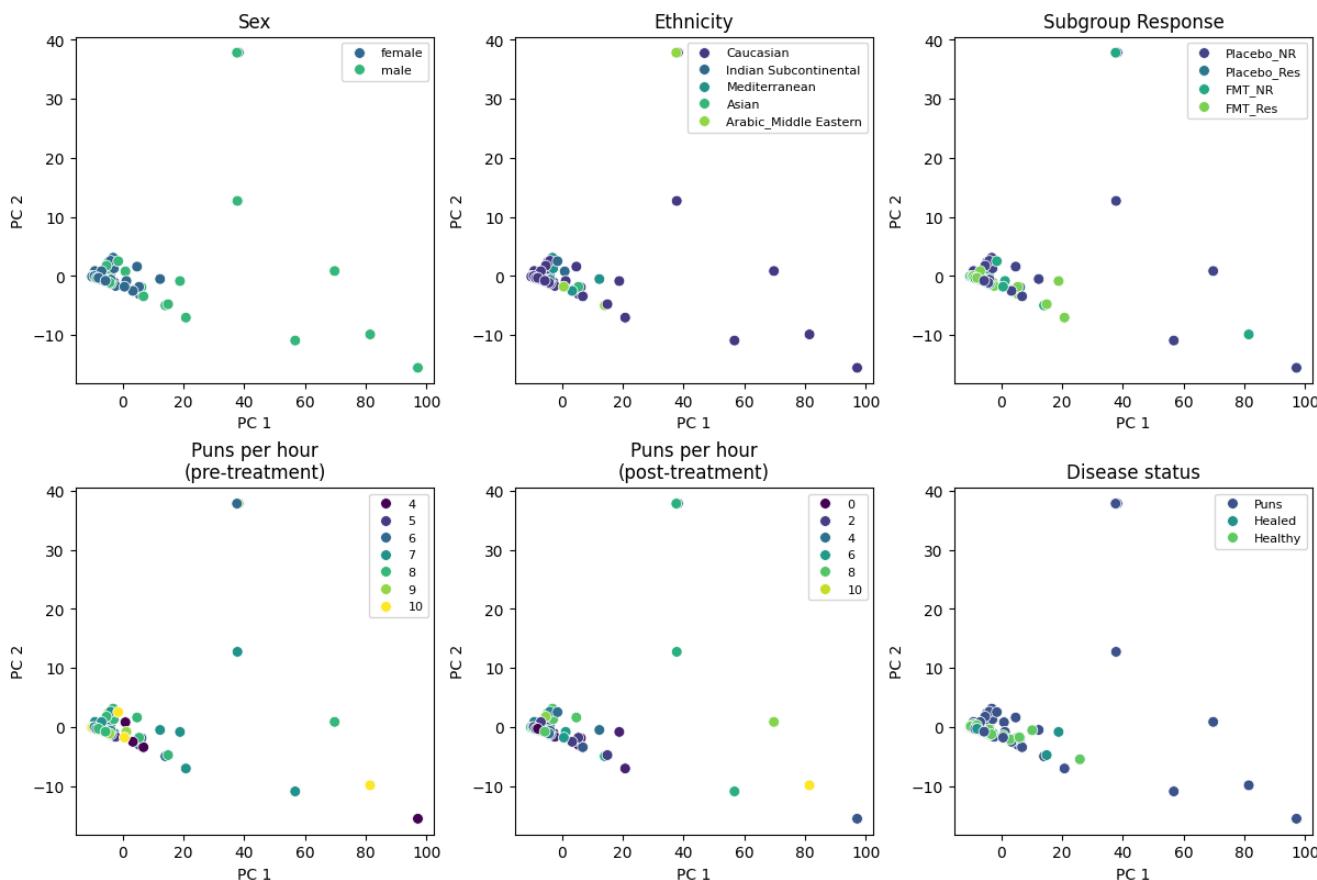
Finally, an investigation of potential sex-related effects on fungal composition showed no significant differences, with all q-values equal to 1.

In summary, our findings indicate that fungal communities remain largely stable across different conditions. While this stability rules out sex-related biases, it also provides no evidence of specific fungal taxa as potential contributors to the disease.

## 3.6 Functional Predictions using FunFun

Applying FUNFUN 78.99% percent of OTU sequences could be annotated. Figure A.5 gives an overview of the relative abundance of the different KEGG orthology groups. Some variation can be seen when comparing different samples. However, after grouping samples by disease status (healthy, healed or diseased patients), this variation averages out, and no clear differences are visible.

As shown in figure 3.9, no correlation with functional abundances is visible after applying PCA across any of the tested sample categories, including sex, ethnicity, subgroup response, puns per hour, and disease status. This finding rules out potential batch effects related to sex and ethnicity while indicating no differences between healthy and diseased patients and no observable effects of treatment outcomes. In accordance with the PCA plots (figure 3.9, none of the PERMANOVA tests showed significant results, neither for all groups nor for pairwise tests in between groups. The following tables (3.6, 3.7) show the obtained pseudo-F-statistics and p-values using Bray-curtis dissimilarity and Jaccard Index metrics.



**Figure 3.9: PCA plot of KEGG orthology group predictions per sample.** Samples were colored after sex, ethnicity, subgroup response, puns per hour (pre- and post-treatment) and disease status. No clear correlation between sample categories and functional predictions is visible.

**Table 3.6:** PERMANOVA test results for Bray-Curtis dissimilarity based on relative abundance of KEGG orthology groups.

Metadata	Condition	PERMANOVA (all groups)		PERMANOVA (pairwise)				
		pseudo-F	p-value	group_1	group_2	pseudo-F	p-value	q-value
sex	female, male	0.898	0.648	female	male	0.898	0.648	0.648
disease status	healthy, healed, puns	0.779	0.92	healthy	healed	0.977	0.56	0.88
				healthy	puns	0.654	0.98	0.98
				healed	puns	0.871	0.59	0.88
subgroup response	FMT_NR, FMT_Res, Placebo_NR	0.891	0.745	FMT_NR	FMT_Res	1.061	0.331	0.877
				FMT_NR	Placebo_NR	0.760	0.877	0.877
				FMT_Res	Placebo_NR	0.927	0.603	0.877

**Table 3.7:** PERMANOVA test results for Jaccard Index similarity based on relative abundance of KEGG orthology groups.

Metadata	Condition	PERMANOVA (all groups)		PERMANOVA (pairwise)				
		pseudo-F	p-value	group_1	group_2	pseudo-F	p-value	q-value
sex	female, male	0.906	0.528	female	male	0.960	0.528	0.528
disease status	healthy, healed, puns	0.712	0.949	healthy	healed	0.708	0.905	0.905
				healthy	puns	0.744	0.812	0.905
				healed	puns	0.679	0.868	0.905
subgroup response	FMT_NR, FMT_Res, Placebo_NR	0.707	0.96	FMT_NR	FMT_Res	0.700	0.883	0.953
				FMT_NR	Placebo_NR	0.592	0.953	0.953
				FMT_Res	Placebo_NR	0.839	0.744	0.953

# Chapter 4

## Discussion

This study aimed to explore the role of the fungal microbiome in the context of the Pandemic disease by analyzing data from a clinical trial evaluating fecal microbiota transplantation (FMT) as a potential treatment approach. Its first focus was to identify significant differences in fungal microbiome composition among healthy, recovered, and diseased patients. The second focus was to assess changes in the fungal microbiome during treatment, comparing FMT and placebo groups, as well as patients with and without treatment success. These changes were examined through analyses of taxonomic composition, alpha diversity, beta diversity, differential abundance of taxa, and functional predictions for fungal species present in the samples.

Before further analyses could begin, the data underwent quality control and preprocessing. This process was successfully completed, resulting in approximately 50% of the reads being retained, with acceptable read lengths and satisfactory quality scores. However, taxonomy classification proved to be more challenging. Classification using a custom-trained classifier resulted in nearly all reads being unclassified. Switching to a pre-trained classifier improved the results slightly, but approximately 60% of reads remained unclassified.

We systematically ruled out potential causes, such as issues with multiorientation reads or the pre-trained classifier being trained without trimming primer sequences. Ultimately, we found that classification could be performed effectively using BLAST, with nearly all sequences successfully classified. However, the reason why the naive Bayes classifier failed to classify the reads, despite successful classification via BLAST, remains unclear. This discrepancy should be noted and could serve as an interesting research question for future studies.

A closer examination of the BLAST classification results revealed that some samples exhibited a high abundance of *Blastocystis sp.* or *Candida albicans*. Both are known parasites frequently associated with immunocompromised patients [36, 37], suggesting a potential link between immune deficiencies and the Pandemic disease.

Firstly, we investigated a potential causal relationship between disease status and the composition of a patient's gut microbiome, categorizing individuals as healthy, diseased (patients making puns), or healed (patients with treatment response).

Significant differences in alpha and beta diversity were observed between healthy and diseased patients, with healed patients being in between healthy individuals and diseased patients. These differences were significant for Shannon's entropy, observed features, and Jaccard Index metrics, indicating that the primary driver of these changes is a reduction in richness rather than evenness within the microbial composition.

Alpha diversity for healed patients was more similar to that of healthy patients than to diseased patients, indicating a recovery in microbiome diversity. In contrast, beta diversity revealed that healed patients were more closely related to diseased patients. Despite this discrepancy, both metrics reflect aspects of the recovery process, with alpha diversity suggesting restored richness and beta diversity highlighting a transitional stage in microbial community composition.

Since higher alpha diversity and a more stable microbiome are generally associated with better host health [38], this finding suggests that the Pundemic either negatively impacts the microbiome of affected patients or that individuals with a less healthy microbiome are more susceptible to infection. Importantly, FMT treatment mitigates this effect in patients who respond to the therapy.

However, differential abundance analysis and functional predictions did not yield significant results. Consequently, we lack evidence pointing to specific microbial species or functions that could be responsible for the disease. This highlights the need for further research to uncover the microbial mechanisms underlying the Pundemic.

After examining changes in the fungal microbiome associated with the Pundemic, we extended our analysis to evaluate the impact of FMT treatment on microbial composition by comparing pre- and post-treatment samples. However, no significant differences were detected in beta diversity, differentially abundant features, or functional predictions, except for two depleted species identified in the placebo no response group, where no active treatment was administered. Similarly, pairwise tests for alpha diversity revealed no significant differences between time points or in changes throughout treatment across response groups (FMT no response, FMT response, and placebo no response).

These findings suggest that, while some patients exhibited significant improvements in the number of puns after treatment, indicating recovery, these changes were not reflected in the fungal microbiome. A potential explanation could be the small cohort size, which may lack the statistical power to detect significant microbial differences. Larger studies could help identify specific fungi or fungal functions associated with disease progression or recovery.

Despite the lack of significant findings, subtle patterns emerged. Patients in the no-response groups (FMT and placebo) exhibited similar trends, whereas those in the response group showed distinct behavior. Interestingly, patients who responded to FMT treatment experienced a reduction in alpha diversity after treatment, while non-responders demonstrated an increase in alpha diversity. Given that donors had a more diverse microbiome, the opposite trend might have been expected.

This unexpected effect could be explained by the observation that patients who responded to FMT treatment had higher alpha diversity before treatment compared to non-responders. This suggests that recovery may only be achievable for patients starting with a more diverse microbiome. Considering that all other tests were non-significant, a potential explanation is that the improvement in disease over time may not be directly attributable to the FMT treatment itself but instead to a natural recovery process that occurs over time.

Generally, fungal communities in the gut stayed mostly stable across different groups, even after Fecal Microbiotal Transplant. This stability could be due to the presence of a “core” fungal community that is resistant to changes. Fungi like *Saccharomyces* and *Candida* are commonly found in the gut and may serve important functions, helping them remain stable even with changes in diet, health, or treatments such as FMT [39–41]. However, this apparent stability may also be influenced by the low relative abundance of fungi compared to bacteria in the gut microbiome, which complicates the detection of less abundant taxa. Methodological limitations, including the use of ITS sequencing, which offers limited phylogenetic resolution, may further obscure subtle community shifts [42, 43].

Despite employing advanced statistical methods, this study faced several limitations. The fungal dataset was sparse, meaning that many of the fungal taxa detected were in very low abundance or absent in many samples. Additionally, the dataset did not follow a normal distribution, which made statistical analysis challenging. We used methods like PERMANOVA (Permutational Multivariate Analysis of Variance) and ANCOM-BC (Analysis of Composition of Microbiomes with Bias Correction) to address these issues.

PERMANOVA is particularly useful for analyzing differences in community composition in non-normal data and is robust against sparse datasets [44]. However, the high sparsity of the data limited its power to detect significant differences, especially when many taxa were present in very few samples [45].

Similarly, while ANCOM-BC is designed to handle compositional data and adjust for biases introduced by differences in sequencing depth, the small sample sizes still posed challenges in detecting meaningful shifts in the fungal community [46]. Additionally, approximately one-third of the patients lacked matching pre- and post-treatment samples, being excluded from the analysis and further reducing the dataset size. A notable issue was the analysis of placebo responders, who were excluded from the final differential abundance tests due to insufficient sample size. This limitation reduced the overall interpretability of our results and highlights the importance of having larger sample sizes in microbiome studies.

Another limitation of this study was the reliance on ITS sequencing to analyze fungi. While ITS sequencing is effective for identifying fungal diversity, it provides limited information about the evolutionary relationships and ecological roles of specific fungi in the gut. This constraint made it challenging to delve deeper into the functional and phylogenetic aspects of fungal communities.

Future studies could address these limitations by including a larger number of participants, improving data collection strategies, and employing advanced sequencing techniques, such as whole-genome shotgun sequencing, to gain more comprehensive insights into fungal activity in the gut. Additionally, refining analysis methods to better handle fungal data and updating bioinformatics pipelines to account for the unique complexities of fungal communities would further enhance the reliability and interpretability of results [47].

Despite these challenges, our study demonstrated a connection between the Pundemic and the fungal microbiome. While we could not establish the effectiveness of FMT treatment against the Pundemic in terms of inducing significant changes in the gut microbiome, this limitation may stem from the design of the clinical trial rather than the treatment itself. Future research may uncover the exact causes of the Pundemic and potentially identify effective treatment strategies.

# **Chapter 5**

## **Supplementary Material**

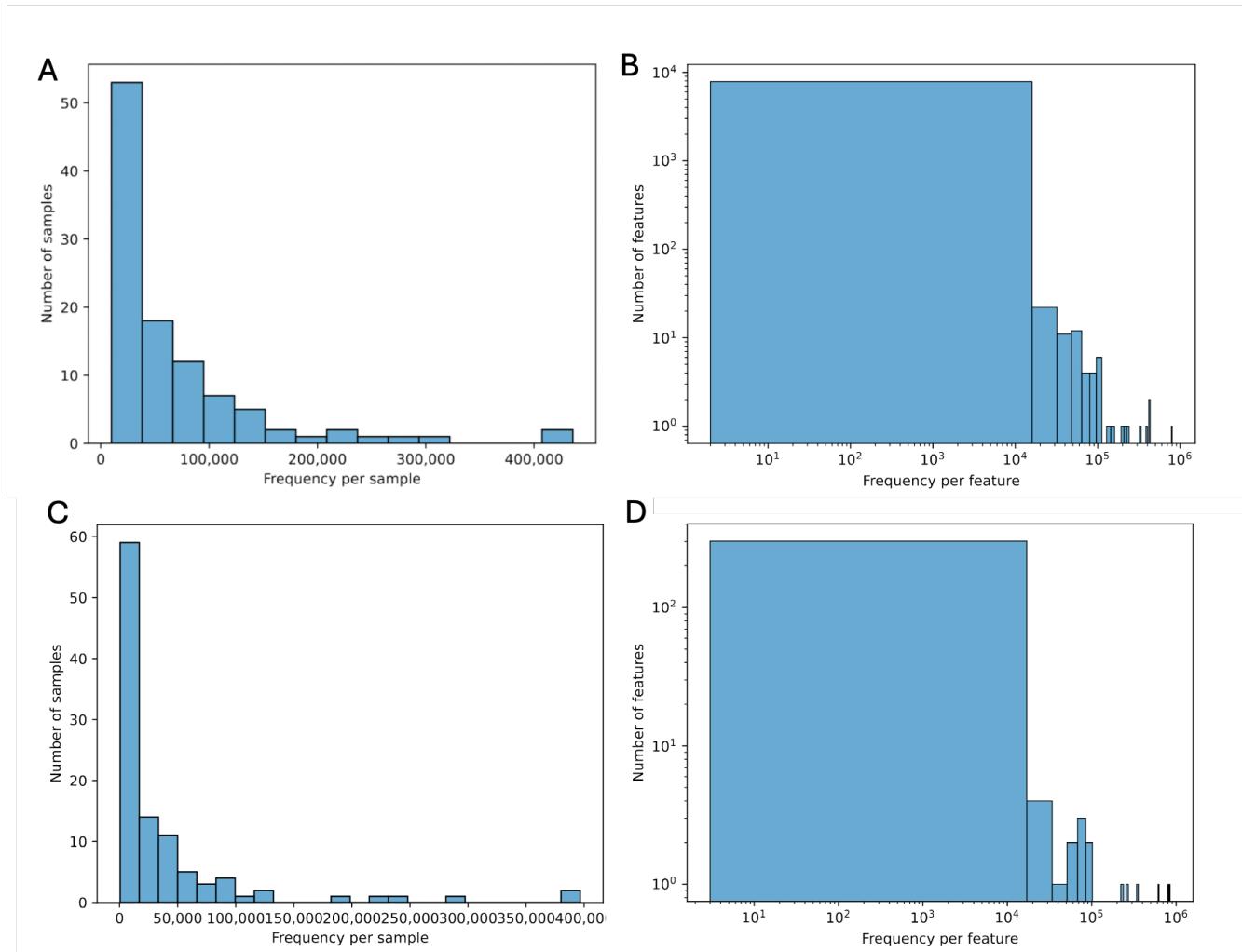
### **Data and Code Availability**

The raw and processed datasets and materials as well as our Jupyter Notebooks used for processing and analyzing the data is available in this GitHub repository: *pundemic*.

The UNITE database v10 (eukaryotes, 99% identity, no singletons) is available at: <https://unite.ut.ee/repository.php>.

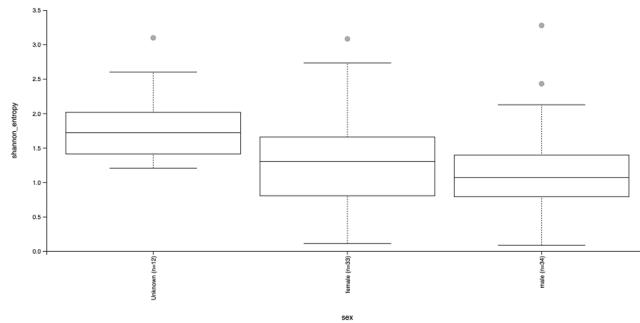
# Appendix A

## Appendix



**Figure A.1: Histogram showing the distribution of the number of features per samples and the distribution of features.** The number of ASVs after denoising (A) and of OTUs after closed-reference clustering (C) differs a lot between samples. The feature abundance of ASVs (B) and OTUs (D) also varies.

## APPENDIX A. APPENDIX



[Download raw data as TSV](#)

### Kruskal-Wallis (all groups)

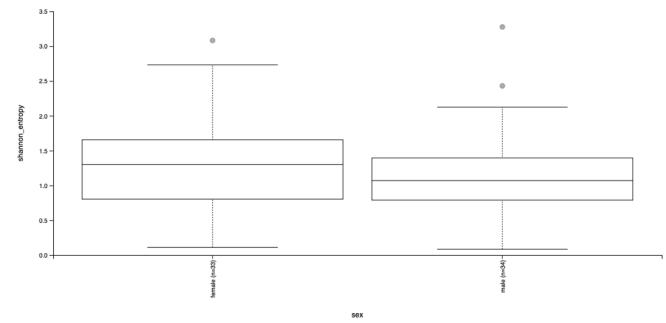
#### Result

H	11.6830298517566
p-value	0.0029044392810554398

### Kruskal-Wallis (pairwise)

[Download CSV](#)

Group 1	Group 2	H	p-value	q-value
Unknown (n=12)	female (n=33)	5.335968	0.020890	0.031335
	male (n=34)	11.917397	0.000556	0.001668
female (n=33)	male (n=34)	1.903114	0.167730	0.167730



[Download raw data as TSV](#)

### Kruskal-Wallis (all groups)

#### Result

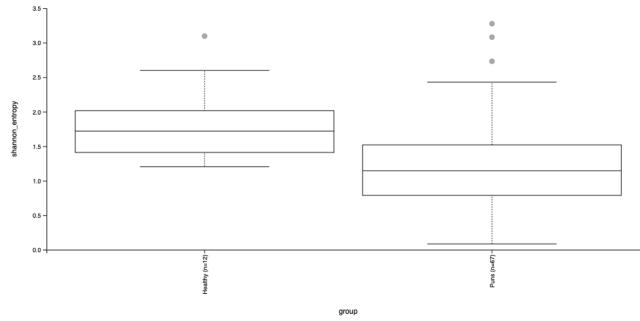
H	1.9031141868511838
p-value	0.1677301562899127

### Kruskal-Wallis (pairwise)

[Download CSV](#)

Group 1	Group 2	H	p-value	q-value
female (n=33)	male (n=34)	1.903114	0.16773	0.16773

(a) Kruskal-Wallis results for the metadata column *sex* with and without the "unknown" condition. The plot compares the test results when the "unknown" condition is included and excluded.



[Download raw data as TSV](#)

### Kruskal-Wallis (all groups)

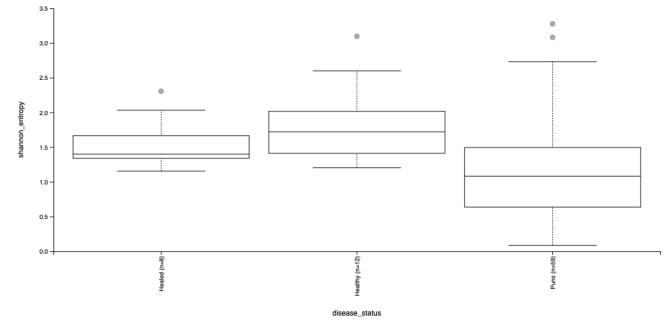
#### Result

H	9.698507462686536
p-value	0.001844177386860712

### Kruskal-Wallis (pairwise)

[Download CSV](#)

Group 1	Group 2	H	p-value	q-value
Healthy (n=12)	Puns (n=67)	9.698507	0.001844	0.001844



[Download raw data as TSV](#)

### Kruskal-Wallis (all groups)

#### Result

H	13.339231119931327
p-value	0.001268864680095917

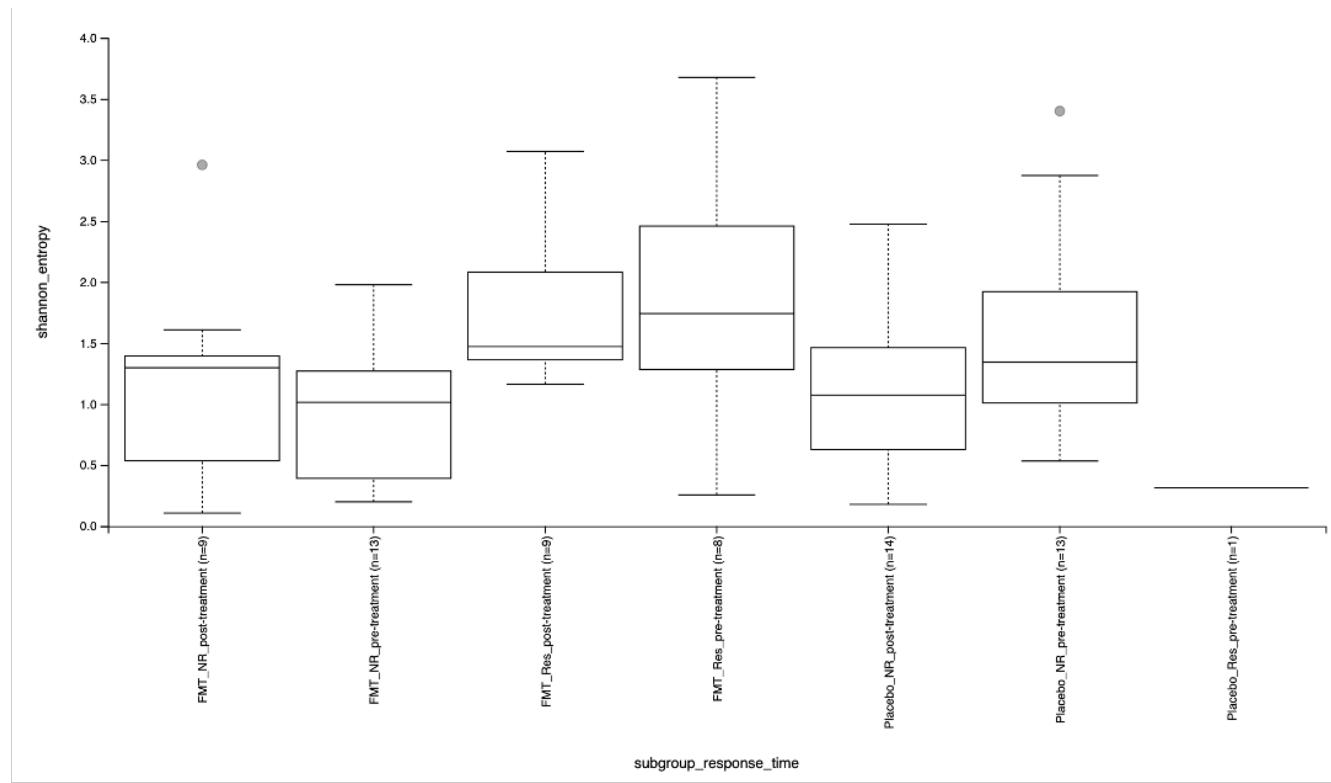
### Kruskal-Wallis (pairwise)

[Download CSV](#)

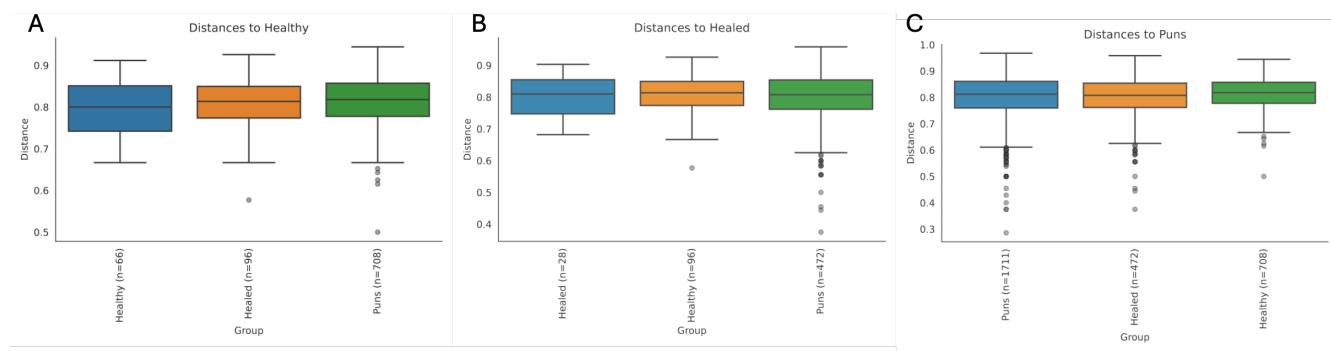
Group 1	Group 2	H	p-value	q-value
Healed (n=8)	Healthy (n=12)	1.339286	0.247160	0.247160
	Puns (n=59)	4.043668	0.044332	0.066498
Healthy (n=12)	Puns (n=59)	10.680085	0.001083	0.003249

(b) Kruskal-Wallis results for the metadata columns *group* and the added column *disease\_status*, with and without the "Healed" condition. The plot shows the effect of adding the "Healed" condition and compares results with its exclusion.

**Figure A.2:** Kruskal-Wallis test results for Shannon Entropy.



**Figure A.3: Boxplot visualizing Shannon Entropy for different treatment response groups pre- and post-treatment.** It can be seen that alpha diversity is higher for patients with successful FMT treatment pre- and post-treatment than for patients not responding to FMT treatment and patients receiving placebo treatment.



**Figure A.4: PERMANOVA test results for beta diversity based on the Jaccard Index for disease status.** The box plots display beta diversity distances among healthy, healed, and diseased (puns) patients relative to healthy (A), healed (B), and diseased (C) reference groups. In all three plots, healed and diseased patients appear to have more similar beta diversity profiles compared to their respective distances from healthy patients. This pattern suggests closer microbial community composition between diseased and healed individuals than between either group and healthy patients.

**Table A.1:** Five most enriched and five most depleted ASVs across the tested metadata categories

		5 most enriched			5 most depleted		
		sample_id	lfc	q_val	sample_id	lfc	q_val
disease_status	Healthy vs. Healed	SH1206929.10FU_JX332067_reps	1.33927	1.000	SH1392968.10FU_KJ956016_reps	-1.69761	1.000
		SH1319629.10FU_EU594608_reps	1.05732	1.000	SH1049518.10FU_KM231814_refs	-1.64055	0.058
		SH1239183.10FU_KY103663_refs	0.84811	1.000	SH1218574.10FU_KX859709_reps	-1.57913	1.000
		SH1269478.10FU_KY307852_refs	0.82280	1.000	SH1419285.10FU_KJ180599_refs	-1.34358	1.000
		SH1239211.10FU_KY101920_refs	0.79280	1.000	SH1338631.10FU_FJ178185_refs	-1.23427	1.000
time_point	Healthy vs. Puns	SH1076932.10FU_AF444417_refs	1.07360	0.187	SH1239211.10FU_KY101920_refs	-1.48866	1.000
		SH1419285.10FU_KJ180599_refs	0.94937	1.000	SH1393052.10FU_KT758093_refs	-1.22004	1.000
		SH1242810.10FU_JN942657_refs	0.66858	1.000	SH1239183.10FU_KY103663_refs	-1.17177	0.677
		SH1049518.10FU_KM231814_refs	0.61729	1.000	SH1319629.10FU_EU594608_refs	-0.87732	1.000
		SH1382014.10FU_AY743636_refs	0.52889	1.000	SH1206929.10FU_JX332067_reps	-0.79495	1.000
FMT Response pre vs. post	FMT Response pre vs. post	SH1049518.10FU_KM231814_refs	0.98940	1.000	SH1205313.10FU_KJ194370_reps	-1.19375	1.000
		SH1393052.10FU_KT758093_refs	0.89427	1.000	SH1303804.10FU_AF444469_refs	-1.15545	1.000
		SH1382505.10FU_EF567995_refs	0.70935	1.000	SH1242810.10FU_JN942657_refs	-1.13929	1.000
		SH1239211.10FU_KY101920_refs	0.55206	1.000	SH1338507.10FU_MT974688_refs	-1.12269	1.000
		SH1085902.10FU_AB018043_refs	0.53493	1.000	SH1242869.10FU_EF568003_refs	-1.01869	1.000
Placebo Response pre vs. post	FMT No Response pre vs. post	SH1049518.10FU_KM231814_refs	0.94206	1.000	SH1085902.10FU_AB018043_refs	-1.95664	0.498
		SH1393052.10FU_KT758093_refs	0.89578	1.000	SH1338507.10FU_MT974688_refs	-1.83630	1.000
		SH1392968.10FU_KJ956016_refs	0.87129	1.000	SH1245248.10FU_AY939798_refs	-1.17955	1.000
		SH1242669.10FU_AY542871_refs	0.66574	1.000	SH1382505.10FU_EF567995_refs	-0.85837	1.000
		SH1233572.10FU_MT908580_refs	0.64408	1.000	SH1382014.10FU_AY743636_refs	-0.85391	1.000
Placebo No Response pre vs. post	Placebo No Response pre vs. post	-	-	-	-	-	-
		SH1392968.10FU_KJ956016_reps	2.55367	0.291	SH1382014.10FU_AY743636_refs	-1.63810	0.143
		SH1393052.10FU_KT758093_reps	1.82041	1.000	SH1205797.10FU_AF444415_refs	-1.39457	0.460
		SH1357311.10FU_AJ549822_refs	1.60117	1.000	SH1303804.10FU_AF444469_refs	-1.26542	0.419
		SH1277523.10FU_AJ229065_refs	0.69155	1.000	SH1339598.10FU_KJ507660_reps	-0.88867	1.000
sex	male vs. female	SH1338507.10FU_MT974688_reps	0.67833	1.000	SH1301142.10FU_KX516054_reps	-0.75737	1.000
		SH1357311.10FU_AJ549822_refs	0.84058	1.000	SH1206929.10FU_JX332067_reps	-0.59018	1.000
		SH1076932.10FU_AF444417_refs	0.54228	1.000	SH1339315.10FU_EF566016_refs	-0.51782	1.000
		SH1242810.10FU_JN942657_refs	0.50897	1.000	SH1393052.10FU_KT758093_reps	-0.50747	1.000
		SH1242669.10FU_AY542871_refs	0.49439	1.000	SH1382014.10FU_AY743636_refs	-0.43807	1.000
		SH1233572.10FU_MT908580_reps	0.46588	1.000	SH1215953.10FU_EF566003_refs	-0.43362	1.000

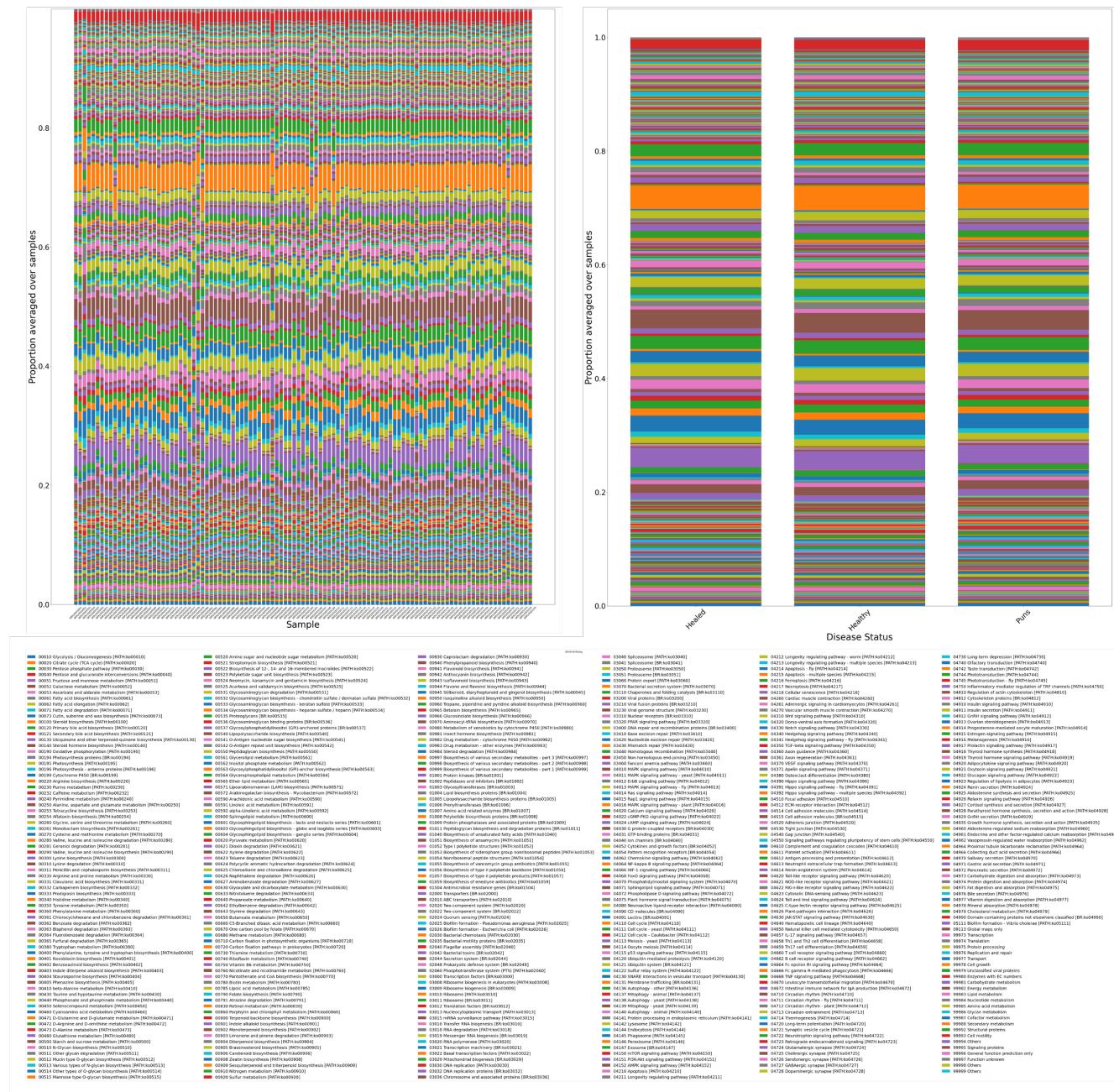
**Table A.2:** Five most enriched taxa across the tested metadata categories

	Kingdom	Phylum	Class	Order	Family	Ifc	q_val
Healthy vs. Healed disease-status	Fungi	Ascomycota	Dothideomycetes	Pleosporales	Pleosporaceae	1.49699	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomycpsidaceae	0.98052	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomyctales_fam_Incertae_sedis	0.95051	1.000
	Fungi	Ascomycota	Eurotiomycetes	Eurotiales	Aspergillaceae	0.92368	1.000
	Fungi	Basidiomycota	Tremellomycetes	Filobasidiales	Filobasidiaceae	0.83686	1.000
Healthy vs. Puns	Fungi	Basidiomycota	Tremellomycetes	Cystofilobasidiales	Mrakiaceae	0.75880	1.000
	Fungi	Basidiomycota	Microbotryomycetes	Sporidiobolales	Sporidiobolaceae	0.63457	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Pichiaceae	0.37772	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Debaryomycetaceae	0.35378	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Debaryomycetaceae	0.31716	1.000
FMT Response pre vs. post	Fungi	Ascomycota	Sordariomycetes	Hypocreales	Nectriaceae	1.01155	1.000
	Fungi	Ascomycota	Ascomycota.cls_Incertae_sedis	Ascomycota_ord_Incertae_sedis	Ascomycota_fam_Incertae_sedis	0.91642	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomyctales_fam_Incertae_sedis	0.57421	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomyctaceae	0.55708	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Metschnikowiaeae	0.54072	1.000
time-point	Fungi	Ascomycota	Sordariomycetes	Hypocreales	Nectriaceae	0.91615	1.000
	Fungi	Ascomycota	Ascomycota.cls_Incertae_sedis	Ascomycota_ord_Incertae_sedis	Ascomycota_fam_Incertae_sedis	0.86987	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Dipodascaceae	0.84538	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Debaryomycetaceae	0.63982	1.000
	Fungi	Basidiomycota	Agaricomycetes	Agaricales	Physalacriaceae	0.61817	1.000
Placebo Response pre vs. post	-	-	-	-	-	-	-
Placebo No Response pre vs. post	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Dipodascaceae	2.20924	0.519
	Fungi	Ascomycota	Ascomycota.cls_Incertae_sedis	Ascomycota_ord_Incertae_sedis	Ascomycota_fam_Incertae_sedis	1.47597	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomyctales_fam_Incertae_sedis	1.25674	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomyctaceae	0.89431	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Metschnikowiaeae	0.32552	1.000
sex	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomyctales_fam_Incertae_sedis	0.85287	1.000
	Fungi	Basidiomycota	Tremellomycetes	Cystofilobasidiales	Mrakiaceae	0.74144	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Dipodascaceae	0.44055	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Metschnikowiaeae	0.38136	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Metschnikowiaeae	0.33189	1.000

**Table A.3:** Five most depleted taxa across the tested metadata categories

	Kingdom	Phylum	Class	Order	Family	Ifc	q_val
disease-status	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Dipodascaceae	-1.53989	1.000
	Fungi	Ascomycota	Sordariomycetes	Hypocreales	Nectriaceae	-1.48284	0.102
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomycetales_fam_Incertae_sedis	-1.42141	1.000
	Fungi	Basidiomycota	Microbotryomycetes	Sporidiobolatales	Sporidiobolataceae	-1.18586	1.000
	Fungi	Fungi_phl_Incertae_sedis	Fungi_cls_Incertae_sedis	Fungi_ord_Incertae_sedis	Fungi_fam_Incertae_sedis	-1.00304	0.699
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomycetales_fam_Incertae_sedis	-1.80346	0.293
Healthy vs. Healed	Fungi	Ascomycota	Ascomycota_cls_Incertae_sedis	Ascomycota_ord_Incertae_sedis	Ascomycota_fam_Incertae_sedis	-1.53485	0.356
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomycetaceae	-1.45156	0.069
	Fungi	Ascomycota	Eurotiomycetes	Eurotiales	Aspergillaceae	-1.16940	0.643
	Fungi	Ascomycota	Dothideomycetes	Pleosporales	Pleosporaceae	-1.10975	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Debaromyctetaceae	-1.33994	1.000
	Fungi	Basidiomycota	Tremellomycetes	Tremellales	Bulleribasidiaceae	-1.13330	1.000
FMT Response pre vs. post	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Debaromyctetaceae	-1.11714	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	-	-0.98023	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Pichiaceae	-0.66062	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomycetaceae	-1.98255	0.320
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	-	-1.87280	1.000
	Fungi	Basidiomycota	Malasseziomycetes	Malasseziales	Malasseziaceae	-1.80406	0.305
time-point	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomycetales_fam_Incertae_sedis	-0.99106	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Debaromyctetaceae	-0.86656	1.000
	Placebo Response pre vs. post	-	-	-	-	-	-
	Fungi	Basidiomycota	Tremellomycetes	Tremellales	Trichosporonaceae	-1.73901	0.049
	Fungi	Basidiomycota	Tremellomycetes	Malasseziales	Bulleribasidiaceae	-1.60985	0.047
	Fungi	Basidiomycota	Malasseziomycetes	Saccharomycetales	Malasseziaceae	-1.41533	0.664
sex	Fungi	Ascomycota	Saccharomycetes	Tremellomycetes	Saccharomycodaceae	-1.23310	1.000
	Fungi	Basidiomycota	Tremellomycetes	Filibasidiales	Filibasidiaceae	-1.10180	0.525
	Fungi	Ascomycota	Dothideomycetes	Pleosporales	Pleosporaceae	-0.57788	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomycetaceae	-0.56429	1.000
	Fungi	Basidiomycota	Malasseziomycetes	Malasseziales	Malasseziaceae	-0.50150	1.000
	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomycodaceae	-0.43330	1.000
male vs. female	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Debaromyctetaceae	-0.34324	1.000

## APPENDIX A. APPENDIX



**Figure A.5: Bar plot showing relative abundances of KEGG orthology groups for fungal sequences averaged over samples.** On the left, the functional predictions for all samples are visualized showing small differences between samples. On the right, the samples were grouped by disease status and averaged by group (healthy, healed, diseased). No clear differences are visible.

# Bibliography

- [1] Gail AM Cresci and Kristin Izzo. Gut microbiome. In *Adult short bowel syndrome*, pages 45–54. Elsevier, 2019.
- [2] Harry Sokol, Valentin Leducq, Hugues Aschard, Hang-Phuong Pham, Sarah Jegou, Cecilia Landman, David Cohen, Giuseppina Liguori, Anne Bourrier, Isabelle Nion-Larmurier, et al. Fungal microbiota dysbiosis in ibd. *Gut*, 66(6):1039–1048, 2017.
- [3] Fen Zhang, Dominik Aschenbrenner, Ji Youn Yoo, and Tao Zuo. The gut mycobiome in health, disease, and clinical applications in association with the gut bacterial microbiome assembly. *The Lancet Microbe*, 3(12):e969–e983, 2022.
- [4] Chloe E Huseyin, Raul Cabrera Rubio, Orla O’Sullivan, Paul D Cotter, and Pauline D Scanlan. The fungal frontier: a comparative analysis of methods used in the study of the human gut mycobiome. *Frontiers in Microbiology*, 8:1432, 2017.
- [5] Nina Gouba, Didier Raoult, and Michel Drancourt. Plant and fungal diversity in gut microbiota as revealed by molecular and culture investigations. *PloS one*, 8(3):e59474, 2013.
- [6] N Gouba and M Drancourt. Digestive tract mycobiota: a source of infection. *Medecine et maladies infectieuses*, 45(1-2):9–16, 2015.
- [7] Giuseppina Liguori, Bruno Lamas, Mathias L Richard, Giovanni Brandi, Gregory Da Costa, Thomas W Hoffmann, Massimo Pierluigi Di Simone, Carlo Calabrese, Gilberto Poggioli, Philippe Langella, et al. Fungal dysbiosis in mucosa-associated microbiota of crohn’s disease patients. *Journal of Crohn’s and Colitis*, 10(3):296–305, 2016.
- [8] M Mar Rodríguez, Daniel Pérez, Felipe Javier Chaves, Eduardo Esteve, Pablo Marin-Garcia, Gemma Xifra, Joan Vendrell, Mariona Jové, Reinald Pamplona, Wifredo Ricart, et al. Obesity changes the human gut mycobiome. *Scientific reports*, 5(1):14600, 2015.
- [9] Karen MJ Waller, Rupert W Leong, and Sudarshan Paramsothy. An update on fecal microbiota transplantation for the treatment of gastrointestinal diseases. *Journal of gastroenterology and hepatology*, 37(2):246–255, 2022.
- [10] Samuel P Costello, Patrick A Hughes, Oliver Waters, Robert V Bryant, Andrew D Vincent, Paul Blatchford, Rosa Katsikeros, Jesica Makanya, Melissa A Campaniello, Chris Mavrellos, et al. Effect of fecal microbiota transplantation on 8-week remission in patients with ulcerative colitis: a randomized clinical trial. *Jama*, 321(2):156–164, 2019.
- [11] Elaine W Yu, Liu Gao, Petr Stastka, Michael C Cheney, Jasmin Mahabamunuge, Mariam Torres Soto, Christopher B Ford, Jessica A Bryant, Matthew R Henn, and Elizabeth L Hohmann. Fecal microbiota

- transplantation for the improvement of metabolism in obesity: The fmt-trim double-blind placebo-controlled pilot trial. *PLoS medicine*, 17(3):e1003051, 2020.
- [12] R Henrik Nilsson, Gunilla Bok, Martin Ryberg, Erik Kristiansson, and Nils Hallenberg. A software pipeline for processing and identification of fungal its sequences. *Source code for biology and medicine*, 4:1–6, 2009.
- [13] Mehrbod Estaki, Lingjing Jiang, Nicholas A Bokulich, Daniel McDonald, Antonio González, Tomasz Kosciolek, Cameron Martino, Qiyun Zhu, Amanda Birmingham, Yoshiki Vázquez-Baeza, et al. Qiime 2 enables comprehensive end-to-end analysis of diverse microbiome data and comparative studies with publicly available data. *Current protocols in bioinformatics*, 70(1):e100, 2020.
- [14] Nicholas A Bokulich, Sathish Subramanian, Jeremiah J Faith, Dirk Gevers, Jeffrey I Gordon, Rob Knight, David A Mills, and J Gregory Caporaso. Quality-filtering vastly improves diversity estimates from illumina amplicon sequencing. *Nature methods*, 10(1):57–59, 2013.
- [15] Benjamin J Callahan, Paul J McMurdie, Michael J Rosen, Andrew W Han, Amy Jo A Johnson, and Susan P Holmes. Dada2: High-resolution sample inference from illumina amplicon data. *Nature methods*, 13(7):581–583, 2016.
- [16] Kessy Abarenkov, R Henrik Nilsson, Karl-Henrik Larsson, Andy FS Taylor, Tom W May, Tobias Guldberg Frøslev, Julia Pawlowska, Björn Lindahl, Kadri Pöldmaa, Camille Truong, et al. The unite database for molecular identification and taxonomic communication of fungi and other eukaryotes: sequences, taxa and classifications reconsidered. *Nucleic Acids Research*, 52(D1):D791–D797, 2024.
- [17] Jennifer Fouquier, Jai Ram Rideout, Evan Bolyen, John Chase, Arron Shiffer, Daniel McDonald, Rob Knight, J Gregory Caporaso, and Scott T Kelley. Ghost-tree: creating hybrid-gene phylogenetic trees for diversity analyses. *Microbiome*, 4:1–10, 2016.
- [18] Warren Weaver. *The mathematical theory of communication*. University of Illinois Press, 1963.
- [19] Evelyn C Pielou. The measurement of diversity in different types of biological collections. *Journal of theoretical biology*, 13:131–144, 1966.
- [20] Danil V Krivonos, Dmitry N Konanov, and Elena N Ilina. Funfun: Its-based functional annotator of fungal communities. *Ecology and Evolution*, 13(3):e9874, 2023.
- [21] Minoru Kanehisa, Yoko Sato, and Kanae Morishima. Blastkoala and ghostkoala: Kegg tools for functional characterization of genome and metagenome sequences. *Journal of molecular biology*, 428(4):726–731, 2016.
- [22] Benjamin J Callahan, Paul J McMurdie, Michael J Rosen, Andrew W Han, Amy Jo A Johnson, and Susan P Holmes. Dada2: High-resolution sample inference from illumina amplicon data. *Nature methods*, 13(7):581–583, 2016.
- [23] Kessy Abarenkov, R Henrik Nilsson, Karl-Henrik Larsson, Andy FS Taylor, Tom W May, Tobias Guldberg Frøslev, Julia Pawlowska, Björn Lindahl, Kadri Pöldmaa, Camille Truong, et al. The unite database for molecular identification and taxonomic communication of fungi and other eukaryotes: sequences, taxa and classifications reconsidered. *Nucleic Acids Research*, 52(D1):D791–D797, 2024.

- [24] Torbjørn Rognes, Tomáš Flouri, Ben Nichols, Christopher Quince, and Frédéric Mahé. Vsearch: a versatile open source tool for metagenomics. *PeerJ*, 4:e2584, 2016.
- [25] D Loos, L Zhang, C Beemelmanns, O Kurzai, and G Panagiotou. Daniel: a user-friendly web server for fungal its amplicon sequencing data. *front microbiol*. 2021; 12: 720513, 2021.
- [26] Patrick E McKnight and Julius Najab. Kruskal-wallis test. *The corsini encyclopedia of psychology*, pages 1–1, 2010.
- [27] SM Taheri and Gholamreza Hesamian. A generalization of the wilcoxon signed-rank test and its applications. *Statistical Papers*, 54:457–470, 2013.
- [28] John Hancock. *Jaccard Distance (Jaccard Index, Jaccard Similarity Coefficient)*. 10 2004.
- [29] C. Ricotta and Janos Podani. On some properties of the bray-curtis dissimilarity and their ecological meaning. *Ecological Complexity*, 31:201–205, 09 2017.
- [30] Yoshiki Vázquez-Baeza, Meg Pirrung, Antonio Gonzalez, and Rob Knight. Emperor: a tool for visualizing high-throughput microbial community data. *Gigascience*, 2(1):2047–217X, 2013.
- [31] Marti J. Anderson. *Permutational Multivariate Analysis of Variance (PERMANOVA)*, pages 1–15. John Wiley Sons, Ltd, 2017.
- [32] Samuel Sanford Shapiro and Martin B Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611, 1965.
- [33] Huang Lin and Shyamal Das Peddada. Analysis of compositions of microbiomes with bias correction. *Nature communications*, 11(1):3514, 2020.
- [34] Sara Branco. Fungal diversity from communities to genes. *Fungal Biology Reviews*, 33(3-4):225–237, 2019.
- [35] Gavin M Douglas, Vincent J Maffei, Jesse R Zaneveld, Svetlana N Yurgel, James R Brown, Christopher M Taylor, Curtis Huttenhower, and Morgan GI Langille. Picrust2 for prediction of metagenome functions. *Nature biotechnology*, 38(6):685–688, 2020.
- [36] Ivan Wawrzyniak, Philippe Poirier, Eric Viscogliosi, Meloni Dionigia, Catherine Texier, Frédéric Delbac, and Hicham El Alaoui. Blastocystis, an unrecognized parasite: an overview of pathogenesis and diagnosis. *Therapeutic Advances in Infectious Disease*, 1(5):167–178, 2013.
- [37] Judith Berman. Candida albicans. *Current biology*, 22(16):R620–R622, 2012.
- [38] Claire E Williams, Tobin J Hammer, and Candace L Williams. Diversity alone does not reliably indicate the healthiness of an animal microbiome. *The ISME Journal*, 18(1):wrae133, 2024.
- [39] E. Van Syoc, M. P. Nixon, J. D. Silverman, Y. Luo, F. J. Gonzalez, I. Elbere, J. Klovins, A. D. Patterson, C. J. Rogers, and E. Ganda. Changes in the type 2 diabetes gut mycobiome associate with metformin treatment across populations. *mBio*, 15, 2024.
- [40] L. Wu, T. Zeng, M. Deligios, L. Milanesi, M. G. I. Langille, A. Zinelli, S. Rubino, C. Carru, and D. J. Kelvin. Age-related variation of bacterial and fungal communities in different body habitats across the young, elderly, and centenarians in sardinia. *mSphere*, 5, 2020.

## BIBLIOGRAPHY

---

- [41] A. K. Nash, T. A. Auchtung, M. C. Wong, D. P. Smith, J. R. Gesell, M. C. Ross, C. J. Stewart, G. Metcalf, D. M. Muzny, R. A. Gibbs, N. J. Ajami, and J. F. Petrosino. The gut mycobiome of the human microbiome project healthy cohort. *Microbiome*, 5, 2017.
- [42] Z. Ling, M. Zhu, X. Liu, L. Shao, Y. Cheng, X. Yan, R. Jiang, and S. Wu. Fecal fungal dysbiosis in chinese patients with alzheimer's disease. *Frontiers in Cell and Developmental Biology*, 8, 2021.
- [43] K. Fiedorová, M. Radvanský, E. Němcová, H. Grombiříková, J. Bosák, M. Černochová, M. Lexa, D. Šmajš, and T. Freiberger. The impact of dna extraction methods on stool bacterial and fungal microbiota community recovery. *Frontiers in Microbiology*, 10, 2019.
- [44] E. Rahkola, S. Rautava, H. Hiltunen, C. Ross, L. Lahti, and E. Isolauri. The preterm gut microbiota and administration routes of different probiotics: a randomized controlled trial. *Pediatric Research*, 94:1480–1487, 2023.
- [45] J. D. Silverman, L. Shenhav, E. Halperin, S. Mukherjee, and L. A. David. Statistical considerations in the design and analysis of longitudinal microbiome studies. 2018.
- [46] H. Lin and S. Peddada. Analysis of compositions of microbiomes with bias correction. *Nature Communications*, 11, 2020.
- [47] K. Chen, H. Geng, J. Liu, and C. Ye. Alteration in gut mycobiota of patients with polycystic ovary syndrome. *Microbiology Spectrum*, 11, 2023.