A short data exploration of the Fungi Kingdom in BugSigDB

YC

August 27, 2025

Overview

The fungi kingdom is an underrepresented taxon in microbiome studies (Perez and Johnson, 2013). Technical difficulties in sampling and identification, as well as a lack of quality controlled reference databases (Pérez, 2021) are some of the obstacles in including fungi in microbiome studies. Most studies have focused on fungal outgrowth when the host is compromised, with little known about the dynamics of the mycobiome during health (Huffnagle & Noverr, 2013).

This exploratory analysis seeks to assess how fungi is represented in the BugSigDB database. Reporting on the most common taxa, distribution across body sites, increasing and decreasing signatures, and diversity based on presence/absence data across experiments.

Setup and Data Import

Call libraries and import or load the BugSigDB database.

```r
library(bugsigdbr)
library(BugSigDBStats)
library(tidyverse)
library(RColorBrewer)
# Import the full BugSigDB database
#bsdb <- importBugSigDB(cache = FALSE) # from BugSigDB
bsdb <- readRDS("bsdb.rds") # load from saved file
```

Identifying Studies and Experiments with Fungi Taxa

To identify the experiments that presented fungi taxa, values belonging to the Fungi kingdom were extracted from the `MetaPhlAn taxon names` column of the "bsdb" dataframe.

```r
# Get experiments that have fungi
fungi <- grepl("k__Fungi", bsdb[["MetaPhlAn taxon names"]])
# Subset
fungi <- bsdb[fungi, ]
```

How many studies in BugSigDB have included fungi taxa in their analysis?
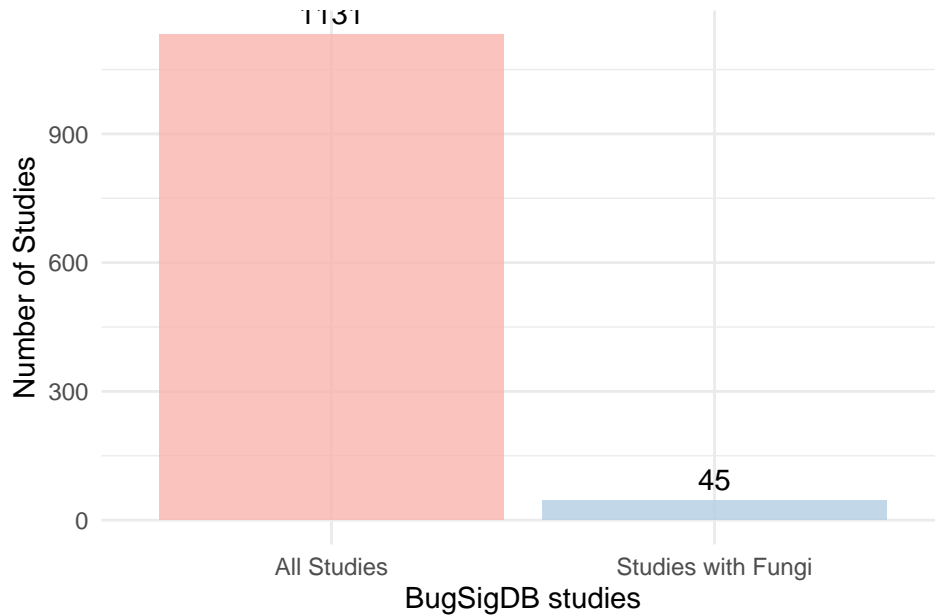

```
## The official BugSigDB release has a total of 1131 studies
## with 5520 experiments. 45 of these studies, with a total
## of 111 experiments included the Fungi kingdom in their analysis.
```

What sequencing type and method was used?

Fungi do not have the 16Sr RNA gene, instead a specific DNA region is targeted.

```
##          PMID    Platform Region Type
## 1   27007700     RT-qPCR     23  16S
## 2   30508628     RT-qPCR    345  16S
## 3   29031597     Illumina      4  16S
## 4   29783199     Roche454     NA  16S
## 5   30405584     Illumina     12  16S
## 6   31834952     Illumina     34  16S
## 7   33806027     Illumina     NA  16S
## 8   32934117     Illumina     34  16S
## 9   22949626     Roche454      4  16S
## 10  30576008     Illumina     34  16S
## 11  29276170     Illumina      4  16S
## 12  31247001     Illumina     NA  16S
## 13  32024712     Illumina     34  16S
## 14        NA     Illumina     NA  16S
## 15  31672155     Illumina      4  16S
## 16  30631088     Roche454    567  16S
## 17  33565055     Illumina     45  16S
## 18  38589368     Illumina     34  16S
## 19  37158877     Illumina     34  16S
## 20  34376634  Ion Torrent     34  16S
```

There are 20 studies that should be reviewed since 16S is listed as part of the sequencing protocol. In many cases there was more than one type of library used and the results for both were grouped together. BugSigDB does not allow for more than one type to be inputted and this might be the issue.

What taxa do we find in the experiments that have fungi?

Identify the taxonomic signatures in the experiments that contain fungi.

```
##  [1] "Malassezia"                        "Malassezia globosa"
##  [3] "Malassezia"                        "Malassezia globosa"
##  [5] "Malassezia globosa"                "Saccharomyces sp."
##  [7] "[Candida] sp. (uncertain placement)" "Saccharomyces sp."
##  [9] "[Candida] sp. (uncertain placement)" "Candida albicans"
```

What percentage of taxa in BugSigDB belong to the fungi kingdom?

Number of Taxa by Category
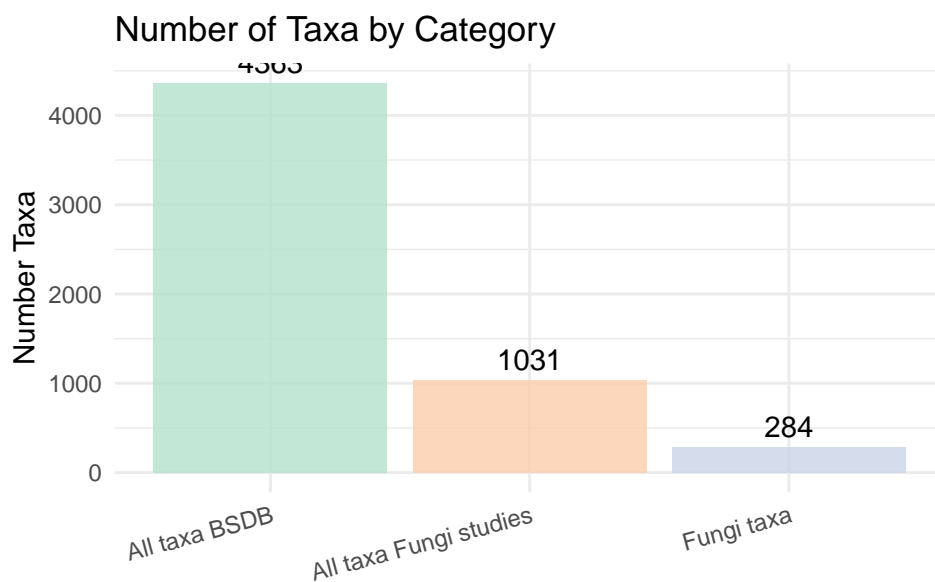

```
## In the 4363 signatures curated in BugSigDB, 4079 taxa are classified as Bacteria, Archaebacteria,
##  Eukaryota or Virus. Out of these, 284 ( 6.5 %) belong to the Fungi kingdom.
```

Heattree for Bacteria, Archea, Fungi and Viruses found in fungi experiments
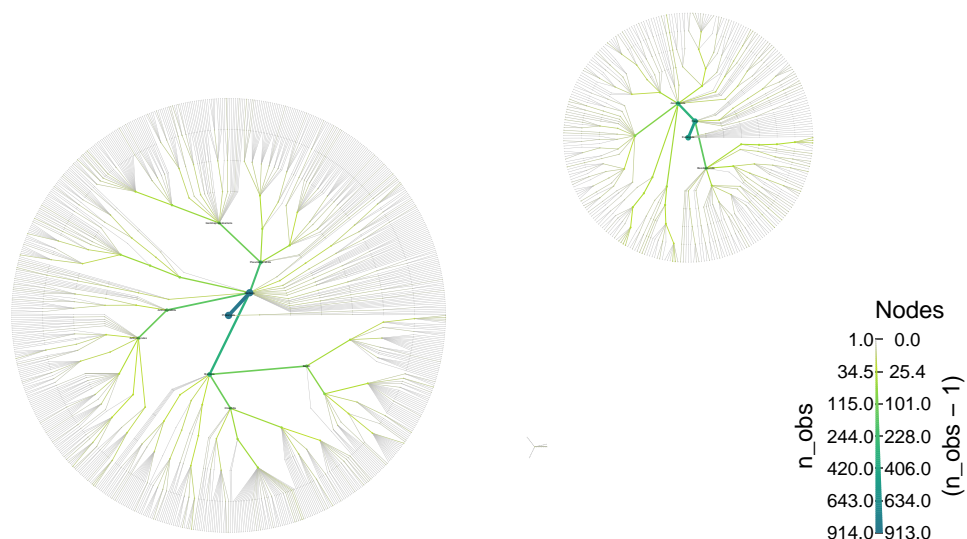
Unexpected taxa found

```
##     superkingdom        kingdom              phyla          class
## 1      Eukaryota        Metazoa           Chordata     Actinopteri
## 2      Eukaryota        Metazoa         Arthropoda         Insecta
## 3      Eukaryota         Evosea      Mastigamoebida     Entamoebidae
## 4      Eukaryota        Metazoa         Arthropoda       Collembola
## 5      Eukaryota      Euglenozoa       Kinetoplastea Trypanosomatida
## 6      Eukaryota        Metazoa           Nematoda      Chromadorea
## 7      Eukaryota     Apicomplexa             <NA>            <NA>
## 8      Eukaryota     Apicomplexa        Aconoidasida     Haemosporida
## 9      Eukaryota        Cercozoa Cercozoa sp. B134           <NA>
## 10     Eukaryota Viridiplantae        Streptophyta    Magnoliopsida
## 11     Eukaryota        Cercozoa      Cercomonadida   Cercomonadidae
## 12     Eukaryota Synurophyceae      Ochromonadales  Ochromonadaceae
## 13     Eukaryota        Metazoa         Arthropoda        Arachnida
##              order                 family            genus
## 1      Gerreiformes              Gerreidae           Gerres
## 2         Blattodea       Rhinotermitidae   Coptotermitinae
## 3         Entamoeba                  <NA>             <NA>
## 4       Poduromorpha       Hypogastruridae     Hypogastrura
## 5   Trypanosomatidae         Leishmaniinae        Leishmania
## 6         Rhabditida        Hoplolaimidae  Rotylenchulinae
## 7              <NA>                  <NA>             <NA>
## 8       Plasmodiidae            Plasmodium  Plasmodium ovale
## 9              <NA>                  <NA>             <NA>
## 10      Malpighiales              Linaceae            Linum
## 11     Paracercomonas      Paracercomonas sp.          <NA>
## 12 Poterioochromonas Poterioochromonas malhamensis        <NA>
## 13         Ricinulei             Ricinoididae      Cryptocellus
```

```
##                             species
## 1             Gerres decacanthus
## 2                    Coptotermes
## 3                           <NA>
## 4   Hypogastrura sp. BOLD:AAI2333
## 5                           <NA>
## 6                    Rotylenchus
## 7                           <NA>
## 8                           <NA>
## 9                           <NA>
## 10           Linum usitatissimum
## 11                          <NA>
## 12                          <NA>
## 13             Cryptocellus magnus
```
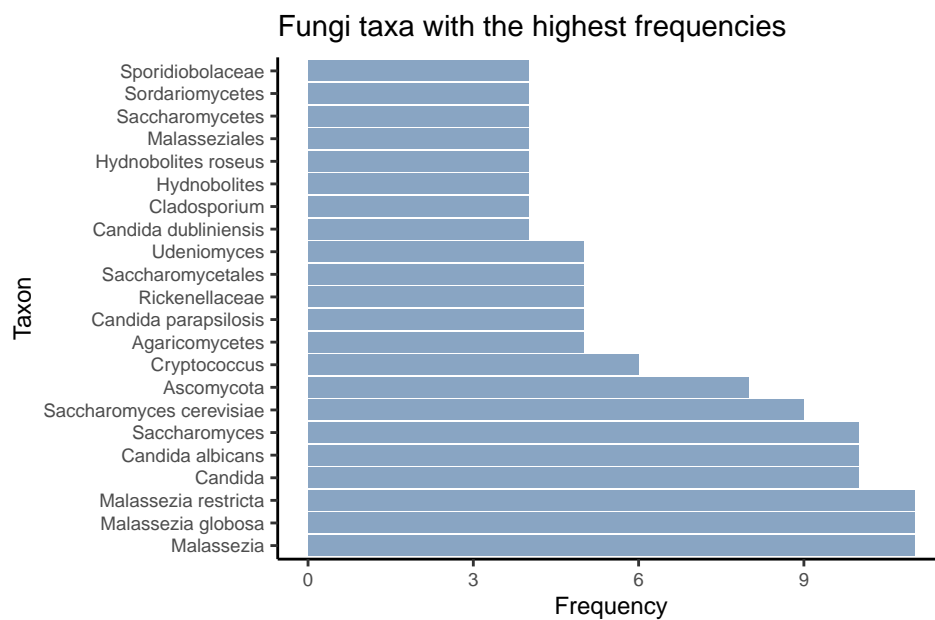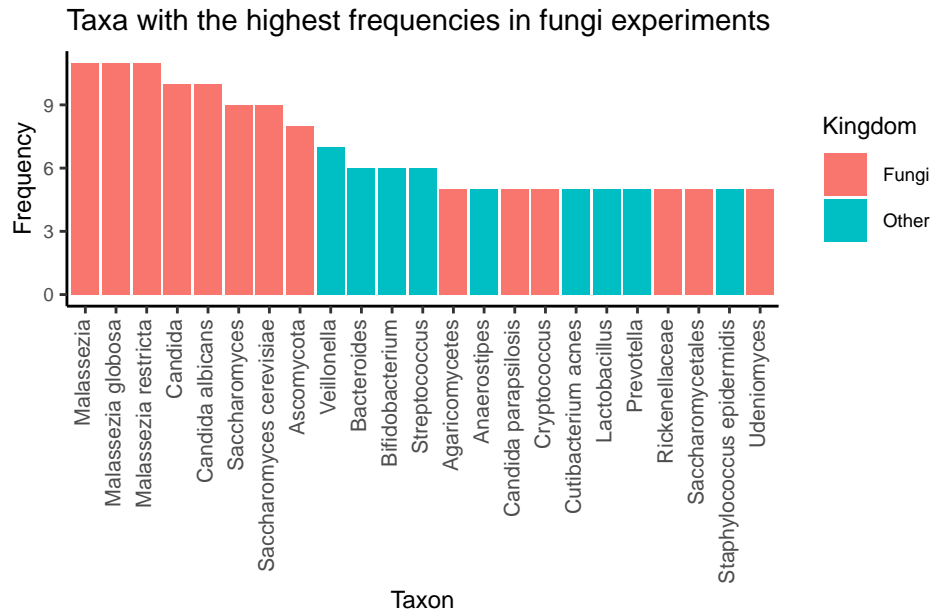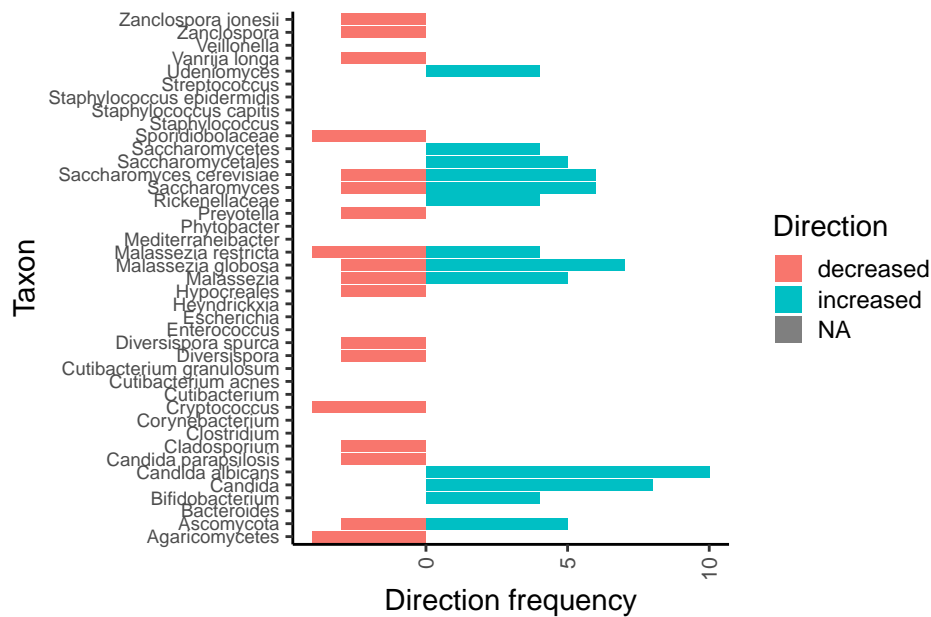
Exploring Frequencies

Which are the most frequent taxa in BugSigDB fungi experiments?

Fungi taxa with the highest frequencies

Taxa with the highest frequencies in fungi experiments
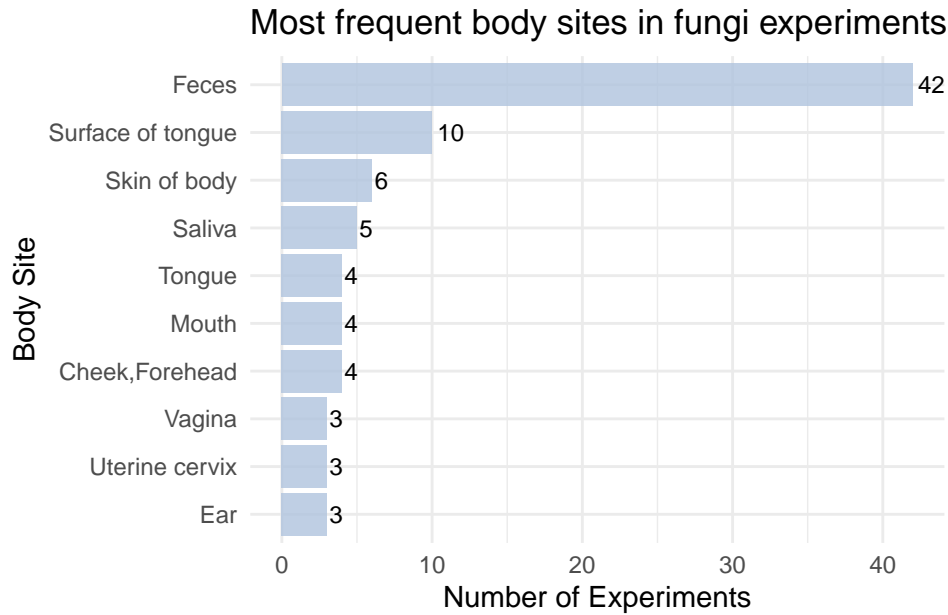
Increasing vs Deacresing signatures

What taxa usually increases or decreases for the condition being tested?

Body Site Analysis

What body sites were tested for fungi? ### Body sites for the experiments that countain fungal taxa
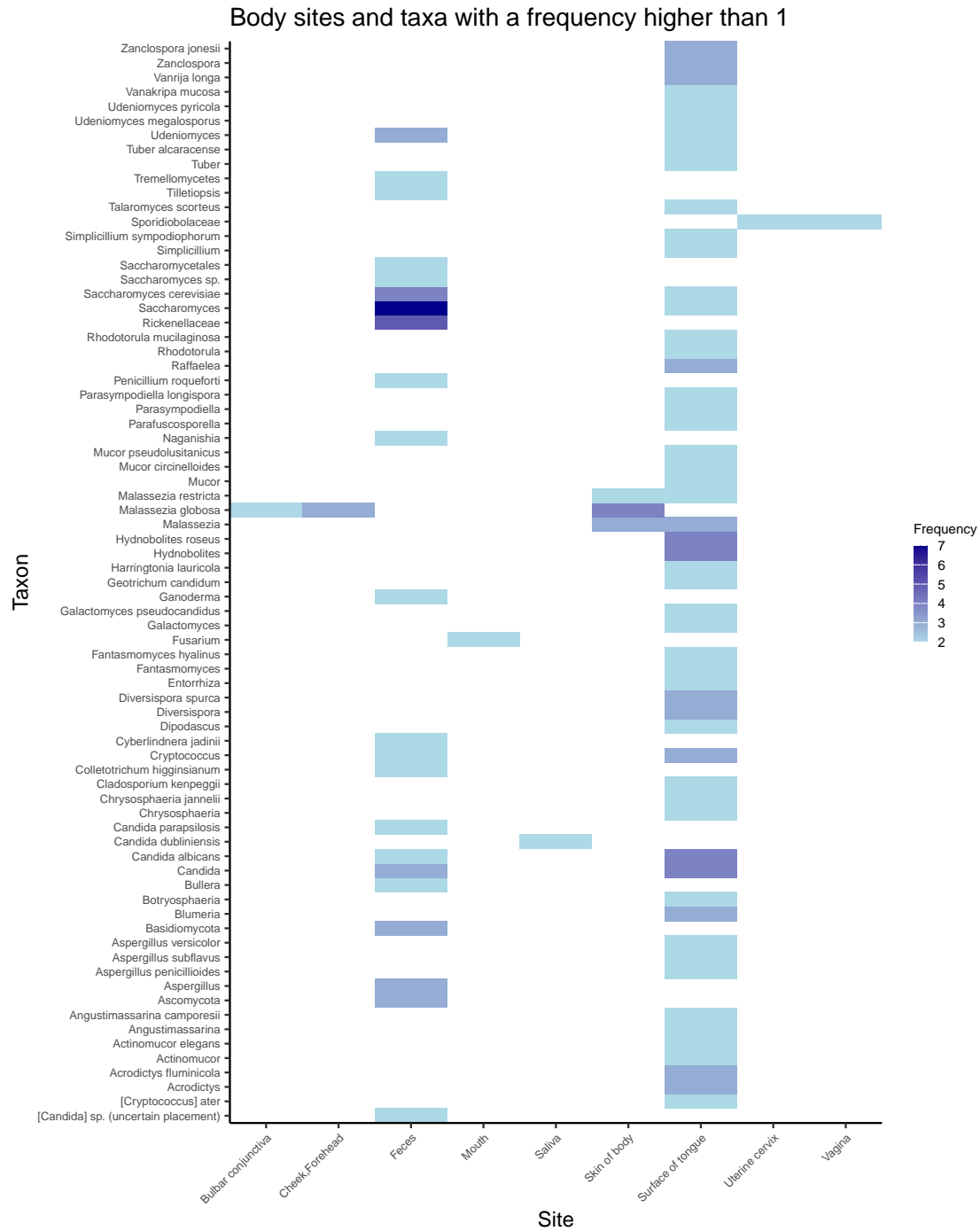
```
## [1] "25  body sites were sampled in the fungi experiments."
```
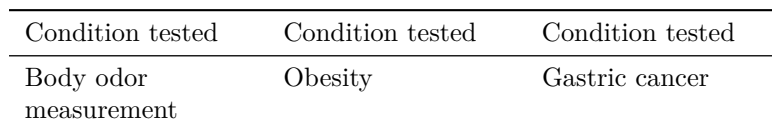
Most frequent body sites in fungi experiments

Body Site	Number of Experiments
Feces	42
Surface of tongue	10
Skin of body	6
Saliva	5
Tongue	4
Mouth	4
Cheek,Forehead	4
Vagina	3
Uterine cervix	3
Ear	3

Significant fungal species often found in the human intestine (From Pérez, 2021).

Name	Observations/Comments	In BSDB
Candida albicans	Most prevalent fungus in the gastrointestinal tract of human adults. Major inducers of Th17 cells.	Is among top 20 frequent fungi species, not frequently repeated in body sites.
Candida glabrata	Opportunistic pathogen. Unknown role(s) in the gut.	Listed in NCBI as Nakaseomyces glabratus. Not found in BSDB
Candida parapsilosis	Opportunistic pathogen. Unknown role(s) in the gut.	Is among top 20 frequent fungi species, not frequently repeated in body sites.
Rhodotorula mucilaginosa	Higher abundance in neonates associated with childhood atopy (Fujimura et al., 2016).	Found in 1 study.
Issatchenkia orientalis (Candida krusei or Pichia kudriavzevii)	Increased abundance in infants associated with atopic wheeze (Arrieta et al., 2018).	Pichia kudriavzevii in NCBI. Not found in BSDB
Malassezia restricta	While often found in stool samples, this skin fungus may not be a true symbiont of the human gut (Fiers et al., 2019).	Frequent in BSDB, found only in feces in decreasing signatures.
Saccharomyces spp.	While they are often found in stool samples (likely because they are components of many beverages and foods), these yeasts are not considered natural residents of the human gut.	Frequent in BSDB, found in feces.

Body sites and taxa with a frequency higher than 1

All taxa frequency by body site

Body sites and taxa with a frequency higher than 1

What conditions were tested for in the experiments that include fungi?

Condition tested	Condition tested	Condition tested
Body odor measurement	Obesity	Gastric cancer

Asthma	Acute myeloid leukemia	Antimicrobial agent
Air pollution	Human papilloma virus infection,Cervical glandular intraepithelial neoplasia	Cervical glandular intraepithelial neoplasia,Human papilloma virus infection
Human papilloma virus infection	Type II diabetes mellitus	Type I diabetes mellitus
Diabetes mellitus	COVID-19	Parkinson's disease
High fat diet	Health study participation	HIV infection
Ethnic group	Age	Crohn's disease
Colorectal cancer	Root	Antimicrobial agent,Amoxicillin
Biological sex	Dry eye syndrome	Acne
Urinary tract infection	Allergic asthma	Behcet's syndrome
Gut microbiome measurement	Microbiome measurement	Polycystic ovary syndrome
Otitis media,Otitis externa	Otitis externa,Otitis media	Hypertension
Response to risperidone	Acute pancreatitis	Lower respiratory tract disease

# What can presence/absence tell us about diveristy?

Diversity in Fungi taxa vs Other taxa

```r
# reorganize for iNext
incidence <- freqKingdom %>%
  add_count(Kingdom, name = "Samples") %>%
  dplyr::select(Kingdom, Samples, Frequency, Taxon, everything())

incidence <- reshape2::dcast(incidence, formula =
            Taxon~Kingdom,      value.var = "Frequency", fill = 0)
## convert to list
incidence_list<-as.list(incidence)
# Remove zeros from each element of the list
incidence <- lapply(incidence_list, function(x) {
  x[x > 0]})
incidence<- incidence[-1]
incidence$Fungi<-c(264, incidence$Fungi)
incidence$Other<-c(264, incidence$Other)

library(iNEXT)
out.inc <- iNEXT(incidence, q=0, datatype="incidence_freq")
```
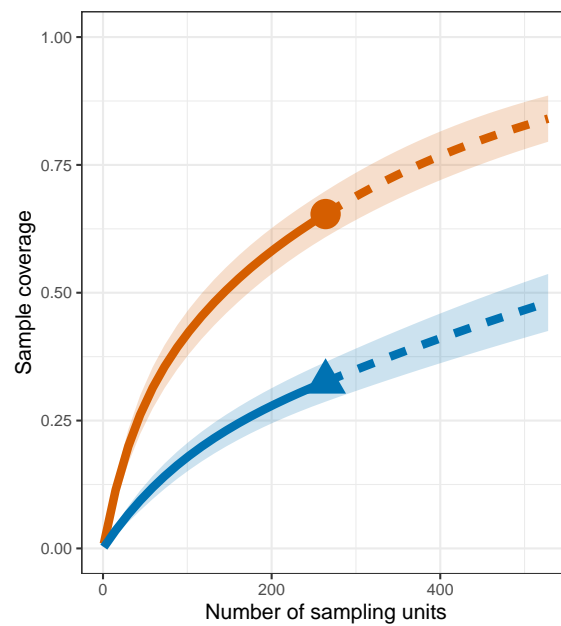
```
ggiNEXT(out.inc, type=1, color.var="Assemblage") +
theme_bw(base_size = 8) +
theme(legend.position="None")
```
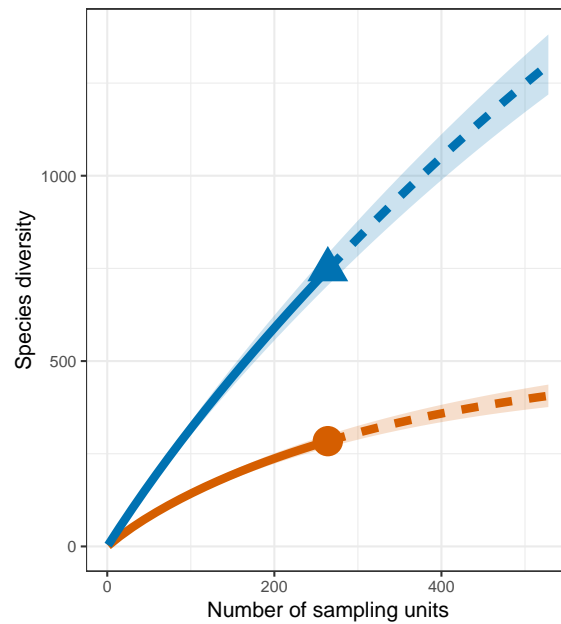


```
ggiNEXT(out.inc, type=2, color.var="Assemblage") +
ylim(c(0,1)) +
theme_bw(base_size = 8) +
theme(legend.position="None")
```

```
ggiNEXT(out.inc, type=3, color.var ="Assemblage") +
xlim(c(0,1)) +
theme_bw(base_size = 18) +
theme(legend.position="bottom",
legend.title=element_blank(),
text=element_text(size=18),
legend.box = "vertical")
```
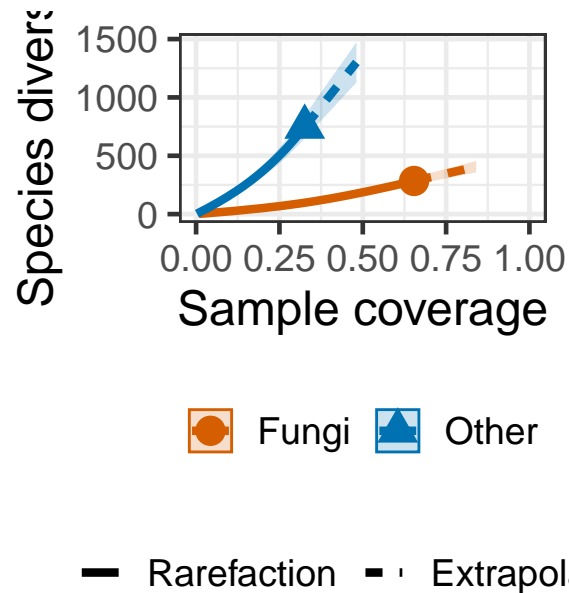

Conclusions

This short exploratory analysis suggests there is an under-representation of fungi compared to bacteria in BugSigDB. This could just be the reflection of an existing bias in microbiome research toward bacterial studies.

At least 20 studies require a second revision to ensure the sequencing method is correctly input.

A taxonomic exploration revealed some unexpected taxa, like the insecta class and Malpighiales order.

The most frequent fungal taxa belong to the Malassezia genus, and include the phylum Ascomycota and genus Saccharomyces. The most frequent bacteria is Staphylococcus epidermis.

The increasing/decreasing abundance patterns may suggest taxa that don't coexist or are antagonistic to each other. This analysis also highlighted there are studies that are missing this input.

The studies took samples from several body sites and tested for different conditions. The frequencies for certain taxa seem to point these may be more prevelant in some body part than in others.

The diversity analysis suggests that fungal diversity is poorly captured with current sampling efforts. However this analysis should be carried out with the same taxa level and not with a mix of taxa levels.