# Appendix S9: results after Qiime Closed ref clustering

Adrien Taudiere*

*CEFE - Centre d'Ecologie Fonctionnelle et Evolutive, Montpellier: France*

July 25, 2016

**To set the filter parameter, see directly section 'Choice of filter parameters' 2.1.**
**Don't forgot to set working directory.**

*adrien.taudiere@cefe.cnrs.fr

# Contents

# 1 Introduction

This supplementary material presents the ecological analysis of endophytic fungal communities in *Pinus nigra* subsp. *laricio*, an endemic species of Corsica. The dataset analyse here was computed using Qiime Closed ref clustering (see article for more details).

## 1.1 R requirements

First we need to install packages.

```
install.packages(c('ape', 'biom', 'optparse', 'RColorBrewer', 'randomForest',  'vegan',
                   'VennDiagram', 'venneuler', 'xtable', 'schoRsch', 'ape',
                  'ips', 'adegenet', 'mvabund', 'rCharts', 'networkD3', 'data.tree'))

# Upgrade Bioconductor to the latest version available for this version of R
source("http://bioconductor.org/biocLite.R")
biocLite(c("multtest", "DECIPHER", "edgeR", "phyloseq", 'DESeq2', 'metagenomeSeq'))

require(devtools)
install_github('ramnathv/rCharts')
install_github("timelyportfolio/d3treeR")
```

```
#Load the packages.
lapply(list("ggplot2", "phyloseq", "cluster", "plyr", "VennDiagram",
            "circlize", "xtable", "schoRsch", "DESeq2", "mvabund",
            "edgeR", "phangorn", "DECIPHER", "ips", "adegenet", "multtest",
            "networkD3", "treemap", "data.tree", "d3treeR", "venneuler",
            "gridExtra"), library,
       character.only = TRUE)
library(vegan)
```

## 1.2 System and session informations

This document was created with R version 3.3.1 (2016-06-21) on Windows the 2016-07-25 10:02:41. See below for more information.

```
sessionInfo()

## R version 3.3.1 (2016-06-21)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 8.1 x64 (build 9600)
##
## locale:
## [1] LC_COLLATE=French_France.1252  LC_CTYPE=French_France.1252
## [3] LC_MONETARY=French_France.1252 LC_NUMERIC=C
## [5] LC_TIME=French_France.1252
##
## attached base packages:
##  [1] parallel  stats4    grid      stats     graphics  grDevices utils
```

```
##  [8] datasets  methods   base
##
## other attached packages:
##  [1] vegan_2.4-0                lattice_0.20-33
##  [3] permute_0.9-0              gridExtra_2.2.1
##  [5] venneuler_1.1-0            rJava_0.9-8
##  [7] d3treeR_0.1                data.tree_0.3.5
##  [9] treemap_2.4-1              networkD3_0.2.11
## [11] multtest_2.28.0            adegenet_2.0.1
## [13] ade4_1.7-4                 ips_0.0-7
## [15] XML_3.98-1.4               colorspace_1.2-6
## [17] DECIPHER_2.0.2             RSQLite_1.0.0
## [19] DBI_0.4-1                  Biostrings_2.40.2
## [21] XVector_0.12.0             phangorn_2.0.4
## [23] ape_3.5                    edgeR_3.14.0
## [25] limma_3.28.12              mvabund_3.11.9
## [27] DESeq2_1.12.3              SummarizedExperiment_1.2.3
## [29] Biobase_2.32.0             GenomicRanges_1.24.2
## [31] GenomeInfoDb_1.8.2         IRanges_2.6.1
## [33] S4Vectors_0.10.1           BiocGenerics_0.18.0
## [35] schoRsch_1.2               xtable_1.8-2
## [37] circlize_0.3.7             VennDiagram_1.6.17
## [39] futile.logger_1.4.1        plyr_1.8.4
## [41] cluster_2.0.4              phyloseq_1.16.2
## [43] ggplot2_2.1.0              knitr_1.13
##
## loaded via a namespace (and not attached):
##  [1] seqinr_3.1-5         deldir_0.1-12        GlobalOptions_0.0.10
##  [4] rstudioapi_0.6       AnnotationDbi_1.34.3 codetools_0.2-14
##  [7] splines_3.3.1        geneplotter_1.50.0   Formula_1.2-1
## [10] jsonlite_0.9.22      gridBase_0.4-7       annotate_1.50.0
## [13] shiny_0.13.2         DiagrammeR_0.8.2     assertthat_0.1
## [16] Matrix_1.2-6         formatR_1.4          visNetwork_1.0.1
## [19] acepack_1.3-3.3      htmltools_0.3.5      tools_3.3.1
## [22] igraph_1.0.1         coda_0.18-1          gtable_0.2.0
## [25] reshape2_1.4.1       dplyr_0.5.0          gmodels_2.16.2
## [28] fastmatch_1.0-4      Rcpp_0.12.5          RJSONIO_1.3-0
## [31] spdep_0.6-5          gdata_2.17.0         nlme_3.1-128
## [34] iterators_1.0.8      stringr_1.0.0        mime_0.4
## [37] gtools_3.5.0         statmod_1.4.24       LearnBayes_2.15
## [40] zlibbioc_1.18.0      MASS_7.3-45          scales_0.4.0
## [43] biomformat_0.99.4    rhdf5_2.16.0         lambda.r_1.1.7
## [46] RColorBrewer_1.1-2   rpart_4.1-10         latticeExtra_0.6-28
## [49] stringi_1.1.1        highr_0.6            genefilter_1.54.2
## [52] gridSVG_1.5-0        foreach_1.4.3        boot_1.3-18
## [55] BiocParallel_1.6.2   shape_1.4.2          chron_2.3-47
## [58] evaluate_0.9         htmlwidgets_0.6      magrittr_1.5
## [61] R6_2.1.2             nnls_1.4             Hmisc_3.17-4
## [64] foreign_0.8-66       mgcv_1.8-12          survival_2.39-5
## [67] sp_1.2-3             nnet_7.3-12          tibble_1.0
## [70] futile.options_1.0.0 locfit_1.5-9.1       data.table_1.9.6
```

```
## [73] digest_0.6.9          httpuv_1.3.3          munsell_0.4.3
## [76] tweedie_2.2.1         quadprog_1.5-5
```

### 1.3 Some usefull functions

The function `as.binaryOtuTable` convert a phyloseq object into a phyloseq object with binary (*i.e.* 0/1) OTU table. It allow to suppress effect due to number of sequences wich may be the result of a lot of molecular artefact (Lindhal et al., 2013).

`funky.color` and `transpa` allow to create nice color palette.

`accu_plot` allow to plot accumulation curves in fonction of a factor in samples data (`@sam_data` of phyloseq object).

`otu_circle` use the package `circlize` to plot circle of OTUs/sequences distributions in samples. `sankey_phyloseq` is an alternative using Sankey plot.

`phyloseq_to_edgeR`, wrote by Paul J. McMurdie, convert phyloseq OTU count data into DGEList for edgeR package.

`plot_deseq2_phyloseq` and `plot_edgeR_phyloseq` plot the result of differential analysis of count data (either using package DESeq2 or edgeR).

```
setwd("~/Documents/GitHub/FEF_paper/")
source(file = "functions_for_phyloseq.R")
```

## 2 Data

### 2.1 Choice of filter parameters

```
#Choose the dataset folder
data_folder <- "Closed_ref"

#Choose the minimum number of sequences by sample.
N_sam_min <- 20000

#Choose the minimum number of samples by OTU.
N_otu_sam_min <- 1

#Choose the minimum number of sequences by OTU.
N_seq_otu_min <- 5
```

### 2.2 Load and convert loading

#### 2.2.1 Otu table

```
#Import biom data
dataBiom   <- import_biom(paste("data/", data_folder, "/otu_table.biom", sep=""))
```

### 2.2.2 Taxonomy

```
#Import taxonomy data
taxRDP_brut <- readLines(paste("data/", data_folder, "/tax_assignments.txt", sep=""))
taxRDP_brut <- gsub(";", "\t", taxRDP_brut)
taxRDP_brut <- gsub(")", "", taxRDP_brut)
taxRDP_brut <- gsub("\\(", "\t", taxRDP_brut)
taxRDP_brut <- gsub("*__", "\t", taxRDP_brut)
taxRDP_brut <- read.table(textConnection(taxRDP_brut), sep = "\t", fill = TRUE)
```

```
# Format taxonomy for phyloseq
taxRDP <- taxRDP_brut[match(taxa_names(dataBiom), taxRDP_brut[, 1]),
                      c(3, 5, 7, 9, 11, 13, 15)]
taxRDP <- tax_table(as.matrix(taxRDP))
taxa_names(taxRDP) <- taxa_names(dataBiom)
colnames(taxRDP) <- c("Domain", "Phylum", "Class", "Order", "Family",
                      "Genus", "Species")
```

### 2.2.3 Add FUNguild information to taxonomy Table

```
taxRDP2 <- as.data.frame(taxRDP)
funguild <- read.delim(paste("data/", data_folder, "/FUNGUILD.guilds.txt", sep=""))
funguild <- funguild[!is.na(match(funguild$OTU_ID, rownames(taxRDP2))),]

match_interm <- match(funguild$OTU_ID, rownames(taxRDP2))

taxRDP2$Trophic_Mode <- NA
taxRDP2$Trophic_Mode[match_interm] <- as.character(funguild$Trophic.Mode)
taxRDP2$Guild <- NA
taxRDP2$Guild[match_interm] <- as.character(funguild$Guild)
taxRDP2$Confidence_Ranking <- NA
taxRDP2$Confidence_Ranking[match_interm] <- as.character(funguild$Confidence.Ranking)
taxRDP2$Growth_Morphology <- NA
taxRDP2$Growth_Morphology[match_interm] <- as.character(funguild$Growth.Morphology)
taxRDP2$Trait<-NA
taxRDP2$Trait[match_interm] <- as.character(funguild$Trait)


taxRDP2 <- tax_table(as.matrix(taxRDP2))
taxa_names(taxRDP2) <- taxa_names(dataBiom)
colnames(taxRDP2) <- c("Domain", "Phylum", "Class", "Order", "Family", "Genus", "Species",
                       "Trophic_Mode", "Guild", "Confidence_Ranking", "Growth_Morphology",
                       "Trait")
```

### 2.2.4 Representative sequences

```
map_endo <-
  import_qiime(map = "data/map_qiimedata.txt")

## Processing map file...

map_endo <- map_endo[order(rownames(map_endo)),]
```

### 2.2.5 Samples information

```
repset <- import_qiime(refseqfilename = paste("data/", data_folder, "/seq.fasta", sep=""))

## Processing Reference Sequences...

taxa_names(repset) <- unlist(strsplit(taxa_names(repset), split = " "))[seq(1, 2*length(taxa_names(r
```

### 2.2.6 Create the phyloseq object

```
data_all <- merge_phyloseq(dataBiom, repset, taxRDP2)
sample_data(data_all) <- map_endo
data_all@tax_table[data_all@tax_table == ""] <- NA
```

### 2.2.7 Caracteristics of the phyloseq data

```
data_all

## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 256 taxa and 80 samples ]
## sample_data() Sample Data:       [ 80 samples by 6 sample variables ]
## tax_table()   Taxonomy Table:    [ 256 taxa by 12 taxonomic ranks ]
## refseq()      DNAStringSet:      [ 256 reference sequences ]
```

The data are made of $6.473782 \times 10^6$ sequences representing 256 OTUs allocate to 80 samples.

## 2.3 Filter sample by number of sequences

```
N_sam_min

## [1] 20000
```

If we discard samples with less than $2 \times 10^4$ sequences, we keep 72 on the 80 samples (90%).

```
barplot(sort(sample_sums(data_all)))
abline(h = N_sam_min)
data.f1 <- prune_samples(sample_sums(data_all) > N_sam_min, data_all)
data.f1 <- prune_taxa(taxa_sums(data.f1) >=  1, data.f1)
```

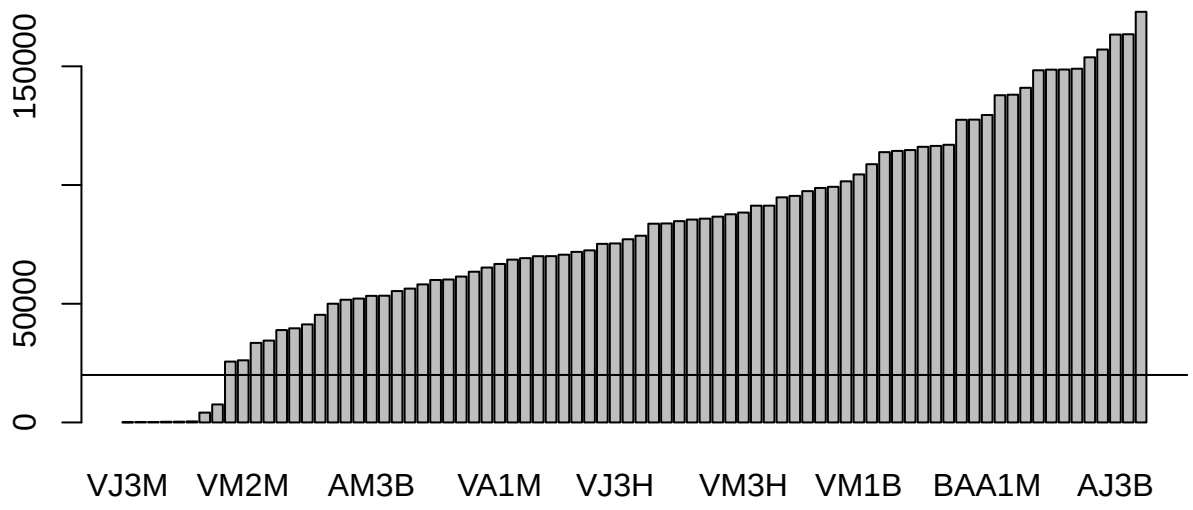


**Figure 2.1:** Number of sequences by sample

## 2.4 Filter OTUs by number of samples

First, we can visualize the number of OTU present in a given number of samples (Figure 2.2).

```
df_nbOtu_sample <- data.frame("Nb of OTUs" = table(rowSums(as.binaryOtuTable(
  data.f1)@otu_table))[table(rowSums(as.binaryOtuTable(data.f1)@otu_table)) > 1],
  "Nb samples" = as.numeric(names(table(rowSums(as.binaryOtuTable(data.f1)@otu_table))
                            [table(rowSums(as.binaryOtuTable(data.f1)@otu_table)) > 1])))

g <- ggplot(df_nbOtu_sample, aes(y = Nb.of.OTUs.Freq, x = Nb.samples))
g + geom_point(size = 4, col = rgb(0.1, 0.1, 0.1, 0.5)) +
  scale_y_continuous(trans = 'log10') +
  geom_smooth(size = 2, col = rgb(0.1, 0.1, 0.1, 0.5))

summary(df_nbOtu_sample$Nb.samples)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    7.75   16.00   25.93   40.50   72.00
```

```
N_otu_sam_min

## [1] 1
```

```
data.f2 <- prune_taxa(rowSums(as.binaryOtuTable(data.f1)@otu_table) >=
                        N_otu_sam_min, data.f1)
```

If we discard OTUs present in less than 1 sample, we keep 252 on the 252 OTUs (100%).

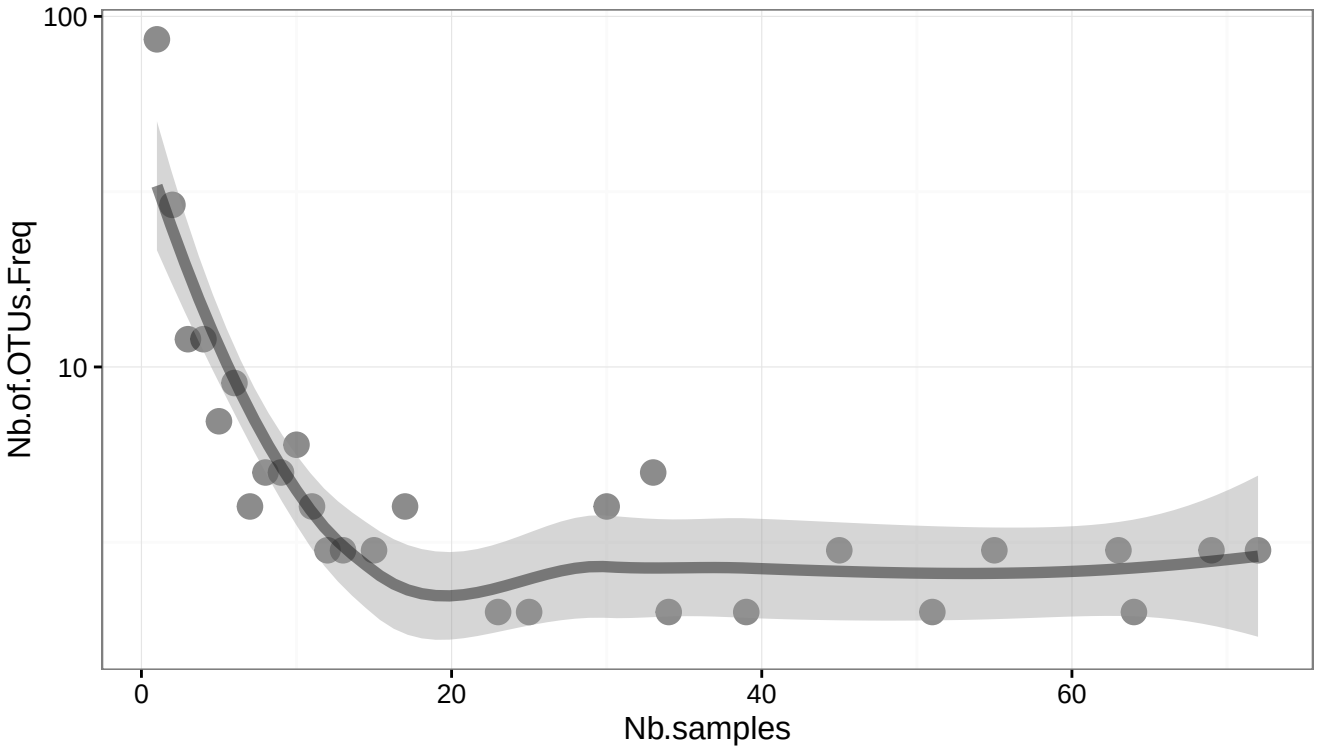


**Figure 2.2:** Number of OTU present in a given number of samples. Vertical bar illustrate the filtering parameter.

## 2.5 Filter OTUs by number of sequences

First, we can visualize the number of sequences by OTU (Figure 2.3).

```
df_nbseq_Otu <- data.frame("Nb of sequences by OTUs" = rowSums(data.f2@otu_table))
g <- ggplot(df_nbseq_Otu, aes(x = Nb.of.sequences.by.OTUs))
g + geom_histogram(size = 2, col = rgb(0.8, 0.8, 0.8, 0.3)) +
  scale_x_continuous(trans = 'log10') +
  geom_vline(xintercept= N_seq_otu_min)

## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.

summary(df_nbseq_Otu[, 1])

##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##       1.0      32.8     203.5   25640.0    1856.0 2292000.0
```

```
N_seq_otu_min

## [1] 5
```

If we discard OTUs with less than 1 sequences, we keep 233 on the 256 OTUs (91.02%).

```
data.f3 <- prune_taxa(rowSums(data.f2@otu_table) >= N_seq_otu_min, data.f2)
```

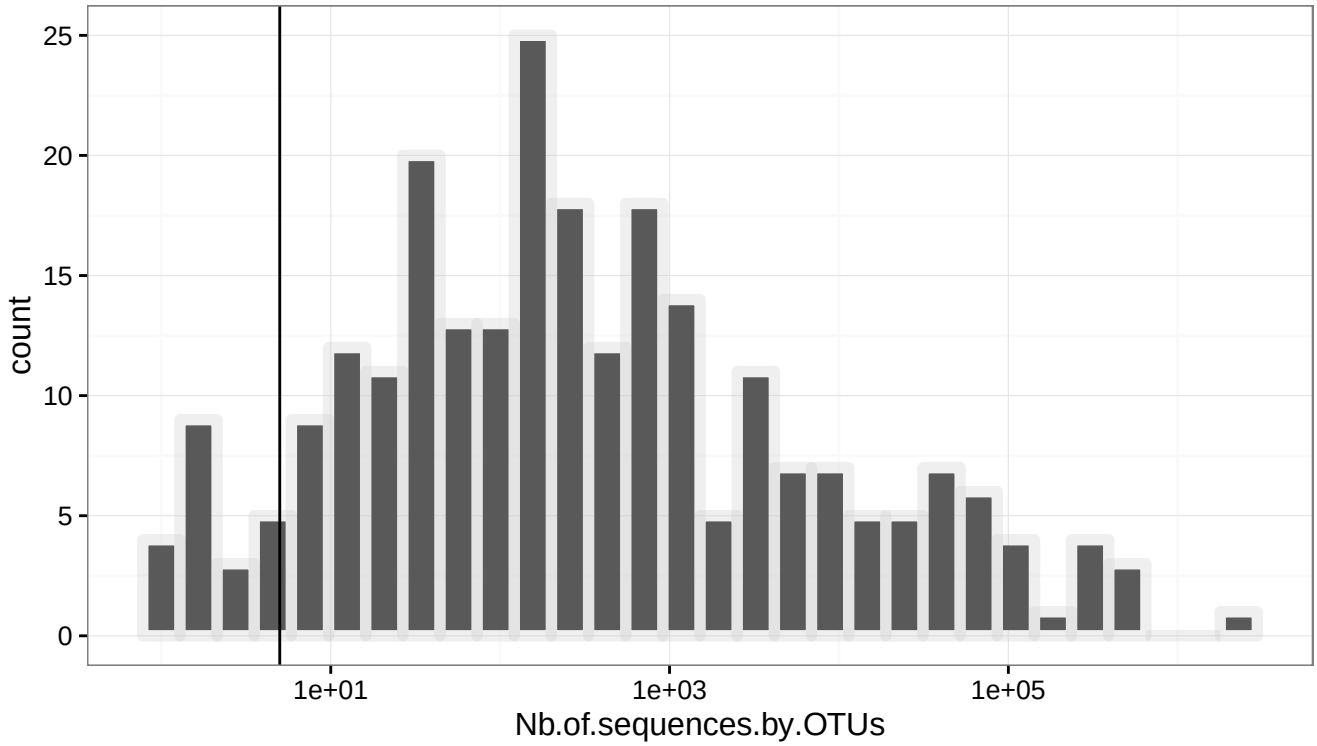


**Figure 2.3:** Number of sequences by OTU (log10 transformed). Horizontal bar illustrate the filtering parameter.

## 2.6 Summary of filtration workflow

The filtered data are made of $6.460489 \times 10^6$ sequences representing 233 OTUs allocate to 72 samples.

| | Nb.of.OTUs | Nb.of.samples | Nb.of.sequences |
|---|---|---|---|
| No filter | 256 | 80 | 6473782.00 |
| Nb of sequences by sample >= 20000 | 252 | 72 | 6460532.00 |
| Nb of sample by OTUs >= 1 | 252 | 72 | 6460532.00 |
| Nb of sequences by OTUs >= 5 | 233 | 72 | 6460489.00 |

**Table 1:** Number of OTUs, samples and sequences after filtering

# 3 Simple description of the dataset

## 3.1 Number of sequences and OTUs by samples

```
df_nbseq_nbotu <- data.frame("Nb of sequences by samples" = colSums(data.f3@otu_table),
                             "Nb of OTUs by samples" =
                               colSums(as.binaryOtuTable(data.f3)@otu_table))

g <- ggplot(df_nbseq_nbotu, aes(x = Nb.of.OTUs.by.samples,
                                y = Nb.of.sequences.by.samples))
g + geom_point(size = 3, col = rgb(0.1, 0.1, 0.1, 0.5)) +
  scale_y_continuous(trans = 'log10') +
```

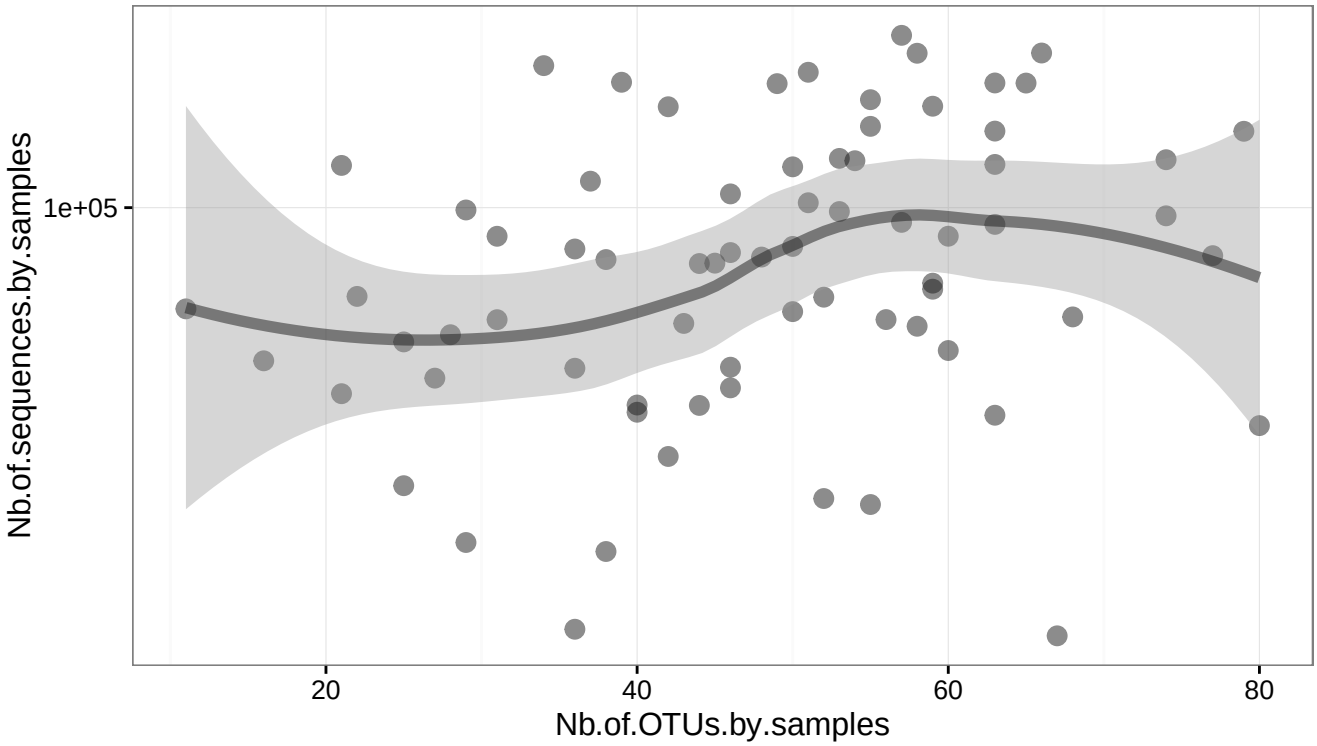


**Figure 3.1:** Number of OTUs by samples in fonction the number of sequences by samples (log10 axe). The tendency is represented by the line obtain from loess (Local Polynomial Regression Fitting).

```
  geom_smooth(size = 2, col = rgb(0.1, 0.1, 0.1, 0.5))
```

```
ggplot(as.data.frame(data.f3@refseq@ranges), aes(x = width)) + geom_density() +
  ylab("Reference sequences length")
```

## 3.2 Number of sequences and samples for each OTUs

```
df_nbseq_nbsam <- data.frame("Nb of sequences by OTUs" = rowSums(data.f3@otu_table)
                             [rowSums(data.f3@otu_table) > 0],
                             "Nb of samples by OTUs" =
                               rowSums(as.binaryOtuTable(data.f3)@otu_table)
                             [rowSums(data.f3@otu_table) > 0])

g <- ggplot(df_nbseq_nbsam, aes(y = Nb.of.samples.by.OTUs,
                                x = Nb.of.sequences.by.OTUs))
g + geom_point(size = 3, col = rgb(0.1, 0.1, 0.1, 0.5)) +
  scale_x_continuous(trans = 'log10') +
  geom_smooth(size = 2, col = rgb(0.1, 0.1, 0.1, 0.5), method = "gam",
              formula  = y ~ s(x, bs = "cs"))
```

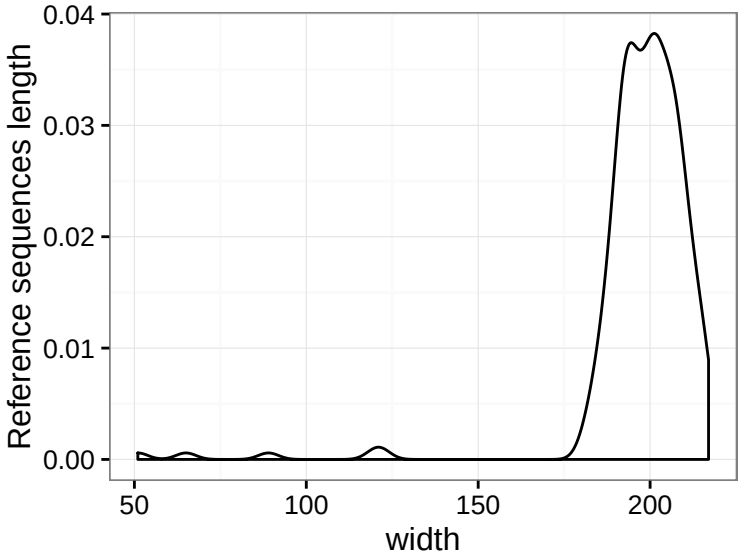


**Figure 3.2:** Distribution of reference sequences length.

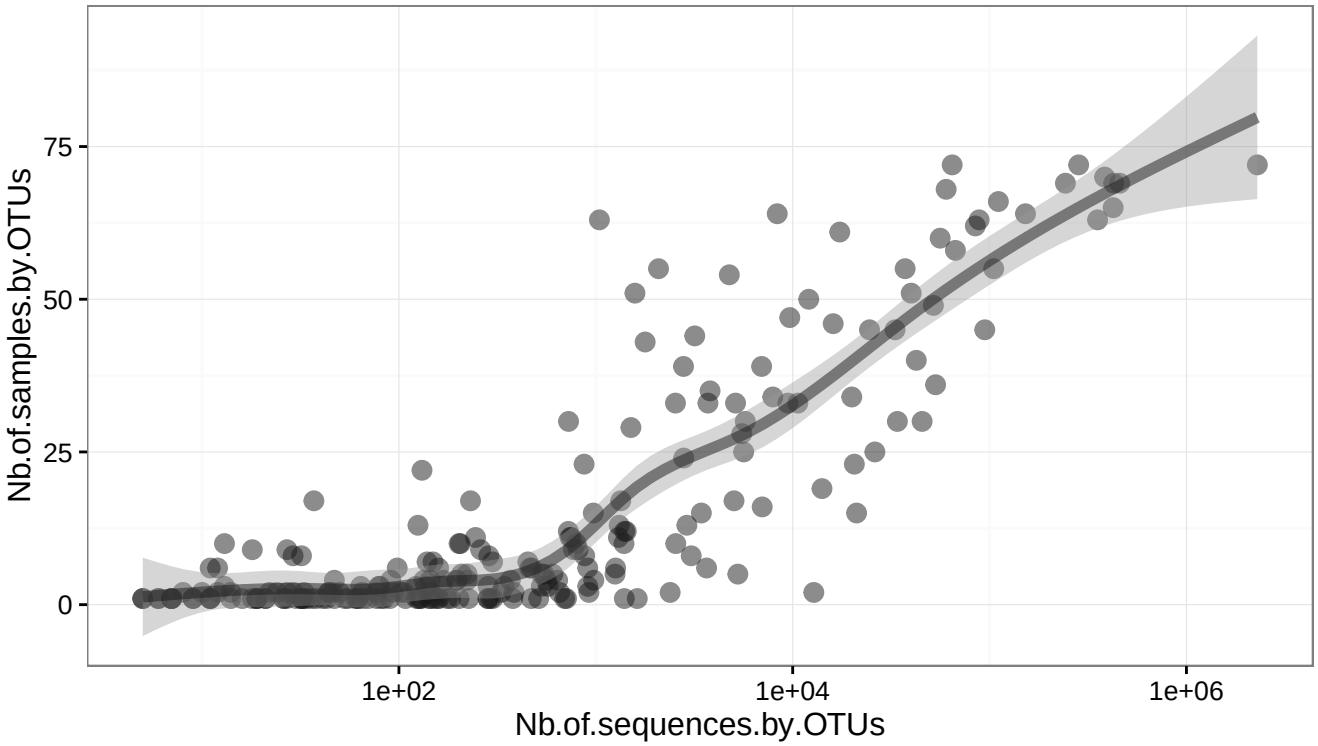


**Figure 3.3:** Number of sequences by OTUs (log10 axe) in fonction the number of samples where OTUs were found. The tendency is represented by the line obtain from gam (Generalized additive models with integrated smoothness estimation).

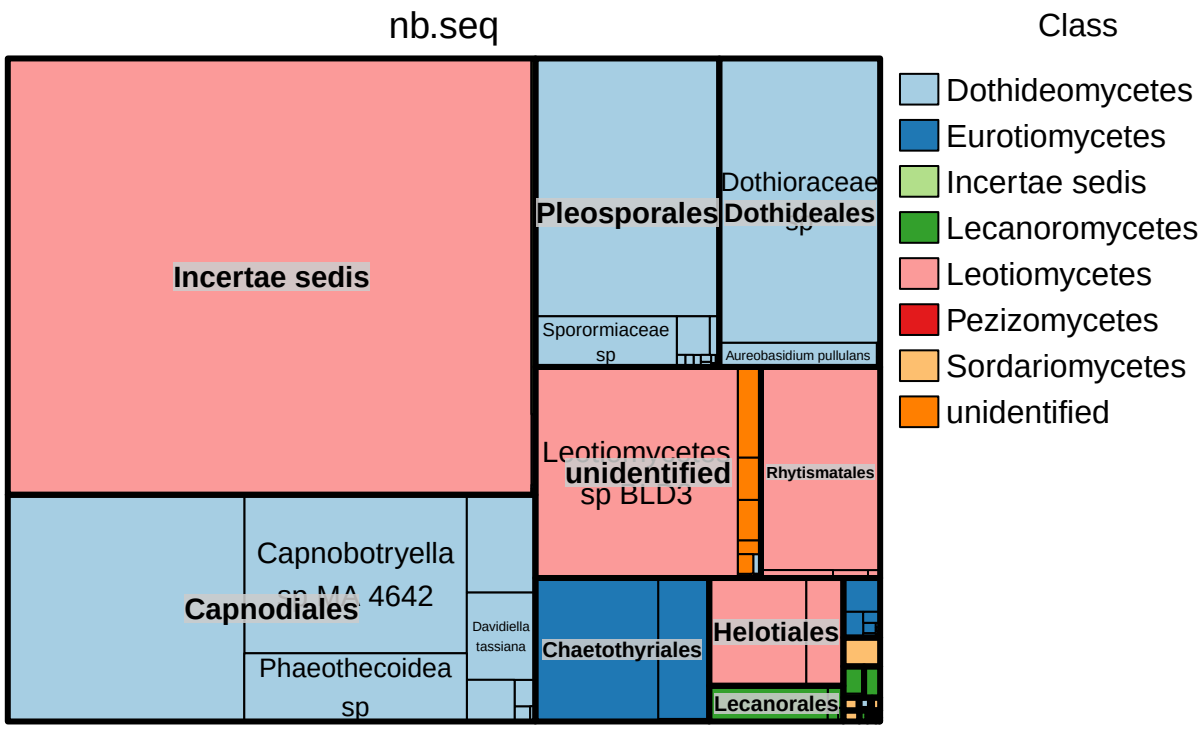


**Figure 3.4:** Distribution of the number of sequences in the taxonomy. Color represent Class, bold lines delimit Order and thick line delimit species.

## 3.3 Distribution of sequences in the taxonomy

```
df3 <- data.frame(as.data.frame(data.f3@tax_table), nb.seq = rowSums(data.f3@otu_table))
tm <- treemap(df3, index = c("Order", "Species"), vSize = "nb.seq", vColor = "Class",
        type = "categorical", palette = "Paired")
# For an interactive version in html
# d3tree(tm)
```

```
data.f3_MINSEQ1000 <- subset_taxa(data.f3, rowSums(data.f3@otu_table)>999)
sankey_phyloseq(data.f3_MINSEQ1000, tax2remove =
                  c("Incertae sedis", "unidentified", "Xylariales", "NA"),
                nbSeq = TRUE, taxa = c(1:6))
```

```
sankey_phyloseq(data.f3, tax2remove = c("Incertae sedis", "unidentified", "Xylariales"),
                nbSeq = FALSE, taxa = c(1:5), min.prop.tax = 0.01)
```

## 3.4 Focus on the 30 more abundant OTUs (number of sequences)

```
the30mostfrequents <- sort(decreasing = T, rowSums(data.f3@otu_table))[1:30]
barplot(the30mostfrequents, horiz = T, cex.names = 0.4, las = 2)
```

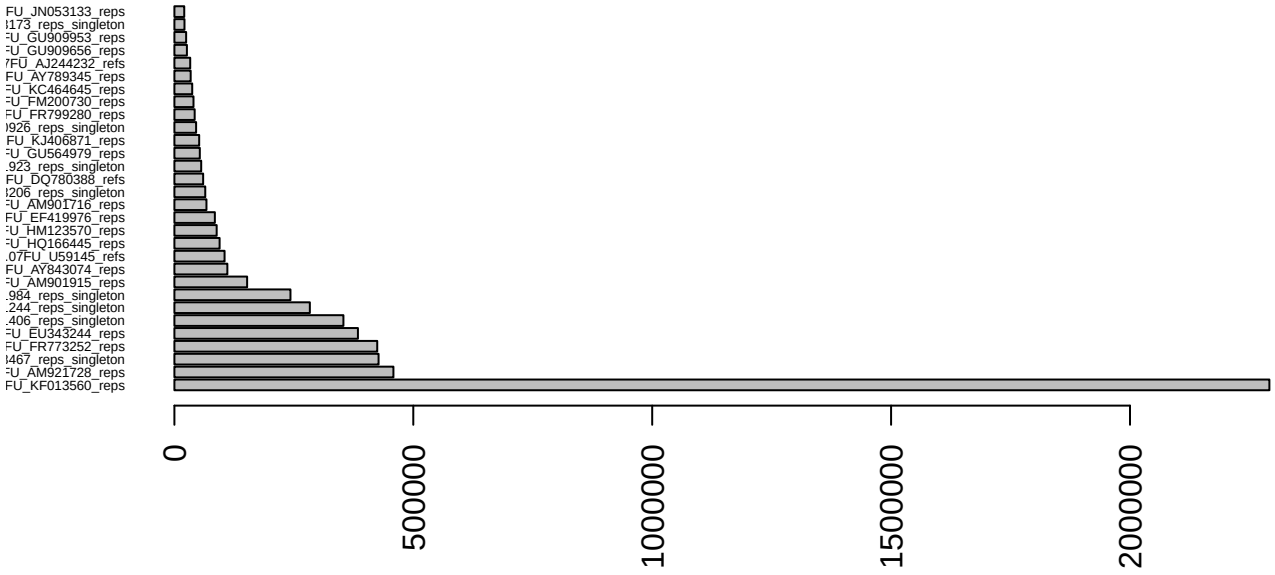


**Figure 3.5:** Number of sequences of the 30 more abundant OTUs (number of sequences).

```
print(xtable(df_the30mostfrequents[, c(1:8, 12)], auto = T,
             caption = "Taxonomie of the 30 more
             frequent OTUs (number of sequences)"),
      size = "\\tiny", include.rownames = FALSE)
```

| Phylum | Class | Order | Family | Genus | Species | Trophic_Mode | Guild | Nb.sequences |
|---|---|---|---|---|---|---|---|---|
| Ascomycota | Leotiomycetes | Incertae sedis | Incertae sedis | Cyclaneusma | Cyclaneusma minus | - | - | 2291731 |
| Ascomycota | Dothideomycetes | Dothideales | Dothioraceae | unidentified | Dothioraceae sp | - | - | 458380 |
| Ascomycota | Leotiomycetes | unidentified | unidentified | unidentified | Leotiomycetes sp BLD3 | - | - | 427300 |
| Ascomycota | Dothideomycetes | Pleosporales | | | | - | - | 424485 |
| Ascomycota | Dothideomycetes | Capnodiales | | | | - | - | 384108 |
| Ascomycota | Dothideomycetes | Capnodiales | Incertae sedis | Capnobotryella | Capnobotryella sp MA 4642 | Saprotroph | Undefined Saprotroph | 353758 |
| Ascomycota | | | | | | - | - | 283446 |
| Ascomycota | Leotiomycetes | Rhytismatales | Rhytismataceae | Lophodermium | | Pathotroph | Plant Pathogen | 242877 |
| Ascomycota | Dothideomycetes | Capnodiales | Mycosphaerellaceae | Phaeothecoidea | Phaeothecoidea sp | Saprotroph | Undefined Saprotroph | 152206 |
| Ascomycota | Dothideomycetes | Capnodiales | | | | - | - | 110862 |
| Ascomycota | Leotiomycetes | Helotiales | Hyaloscyphaceae | Lachnellula | Lachnellula calyciformis | Saprotroph | Undefined Saprotroph | 104983 |
| | | | | | | - | - | 94575 |
| Ascomycota | Eurotiomycetes | Chaetothyriales | Herpotrichiellaceae | Phaeomoniella | Phaeomoniella sp | Saprotroph | Undefined Saprotroph | 88487 |
| Ascomycota | Eurotiomycetes | Chaetothyriales | Herpotrichiellaceae | Phaeomoniella | Phaeomoniella sp | Saprotroph | Undefined Saprotroph | 84672 |
| Ascomycota | Dothideomycetes | Capnodiales | Incertae sedis | Phaeotheca | Phaeotheca sp | - | - | 67092 |
| Ascomycota | | | | | | - | - | 64582 |
| Ascomycota | Dothideomycetes | Capnodiales | Davidiellaceae | Davidiella | Davidiella tassiana | Saprotroph | Undefined Saprotroph | 60185 |
| Ascomycota | | | | | | - | - | 56116 |
| Ascomycota | | | | | | - | - | 53196 |
| Ascomycota | Eurotiomycetes | Chaetothyriales | Herpotrichiellaceae | | | - | - | 51862 |
| Ascomycota | Dothideomycetes | Pleosporales | Sporormiaceae | unidentified | Sporormiaceae sp | - | - | 45446 |
| Ascomycota | Lecanoromycetes | Lecanorales | Parmeliaceae | Pseudevernia | Pseudevernia furfuracea | Symbiotroph | Lichenized | 42439 |
| | | | | | | - | - | 39986 |
| Ascomycota | Leotiomycetes | Helotiales | Hyaloscyphaceae | Lachnellula | | Saprotroph | Undefined Saprotroph | 37259 |
| Ascomycota | Leotiomycetes | | | | | - | - | 34018 |
| Ascomycota | Dothideomycetes | Dothideales | Dothioraceae | Aureobasidium | Aureobasidium pullulans | Saprotroph | Undefined Saprotroph | 33113 |
| Ascomycota | Dothideomycetes | Pleosporales | Sporormiaceae | unidentified | Sporormiaceae sp | - | - | 26156 |
| Ascomycota | Dothideomycetes | Pleosporales | Sporormiaceae | unidentified | | - | - | 24563 |
| unidentified | unidentified | unidentified | unidentified | unidentified | fungal sp agrD488 | - | - | 21114 |

**Table 2:** Taxonomie of the 30 more frequent OTUs (number of sequences)

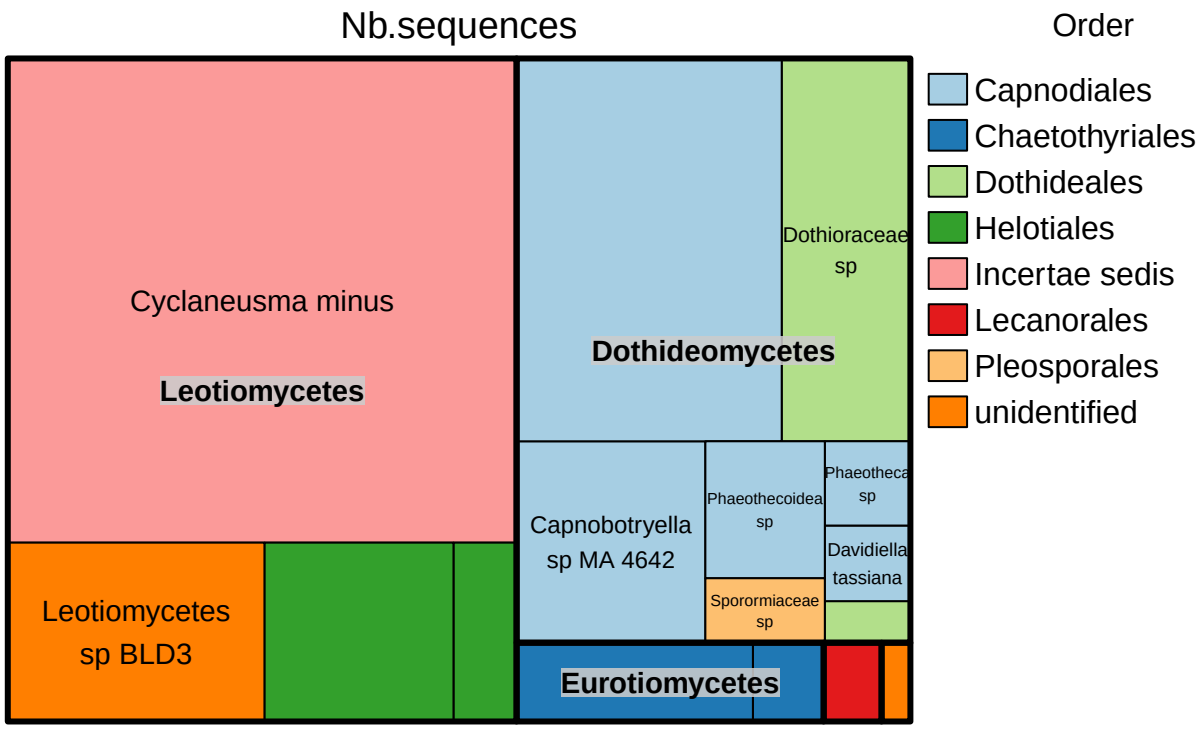


**Figure 3.6:** Number of sequences of the 30 most abundant OTUs (number of sequences). Colors indicate Order, bold lines delimit Class and thick lines delimit species.

```
treemap(df_the30mostfrequents, index = c("Class", "Species"), vSize = "Nb.sequences",
        vColor = "Order", type = "categorical", palette = "Paired")
```

## 3.5 Focus on the 30 more frequent OTUs (number of samples)

```
the30mostfrequents_samp <- sort(decreasing = T,
                                rowSums(as.binaryOtuTable(data.f3)@otu_table))[1:30]
barplot(the30mostfrequents_samp, horiz = T, cex.names = 0.4, las = 2)
```

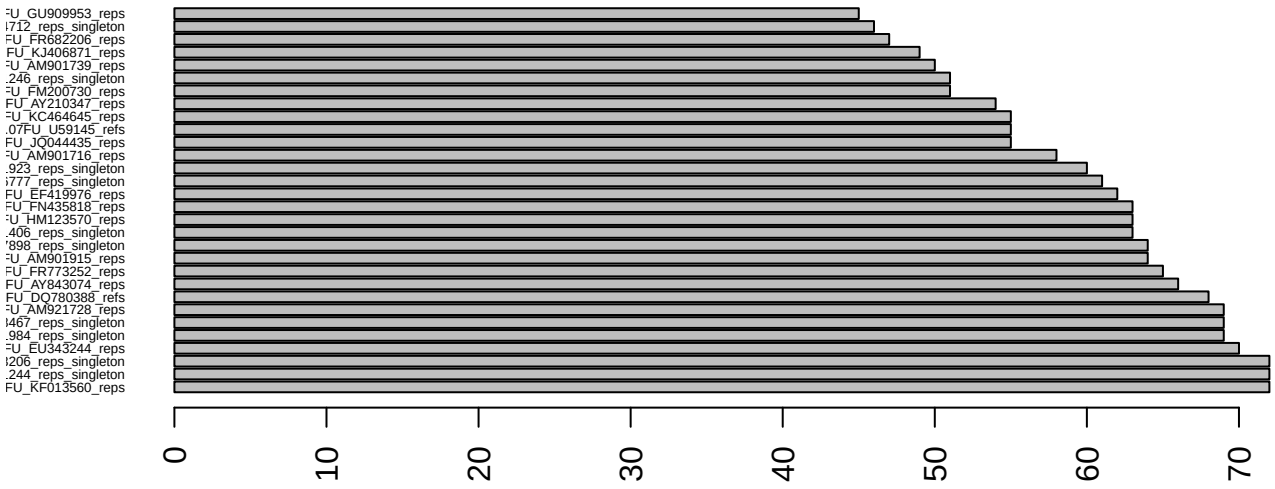


**Figure 3.7:** Number of samples of the 30 more frequent OTUs (number of samples).

```
print(xtable(df_the30mostfrequents_samp[, c(1:8, 12)], auto = T,
      caption = "Taxonomie of the 30 more frequent OTUs (number of samples)"),
      size = "\\tiny", include.rownames = FALSE)
```

| Phylum | Class | Order | Family | Genus | Species | Trophic_Mode | Guild | Nb.samples |
|---|---|---|---|---|---|---|---|---|
| Ascomycota | Leotiomycetes | Incertae sedis | Incertae sedis | Cyclaneusma | Cyclaneusma minus | - | - | 72 |
| Ascomycota | | | | | | - | - | 72 |
| Ascomycota | | | | | | - | - | 72 |
| Ascomycota | Dothideomycetes | Capnodiales | | | | - | - | 70 |
| Ascomycota | Leotiomycetes | Rhytismatales | Rhytismataceae | Lophodermium | | Pathotroph | Plant Pathogen | 69 |
| Ascomycota | Leotiomycetes | unidentified | unidentified | unidentified | Leotiomycetes sp BLD3 | - | - | 69 |
| Ascomycota | Dothideomycetes | Dothideales | Dothioraceae | unidentified | Dothioraceae sp | - | - | 69 |
| Ascomycota | Dothideomycetes | Capnodiales | Davidiellaceae | Davidiella | Davidiella tassiana | Saprotroph | Undefined Saprotroph | 68 |
| Ascomycota | Dothideomycetes | Capnodiales | | | | - | - | 66 |
| Ascomycota | Dothideomycetes | Pleosporales | | | | - | - | 65 |
| Ascomycota | Dothideomycetes | Capnodiales | Mycosphaerellaceae | Phaeothecoidea | Phaeothecoidea sp | Saprotroph | Undefined Saprotroph | 64 |
| Ascomycota | Dothideomycetes | Capnodiales | Mycosphaerellaceae | Phaeothecoidea | | Saprotroph | Undefined Saprotroph | 64 |
| Ascomycota | Dothideomycetes | Capnodiales | Incertae sedis | Capnobotryella | Capnobotryella sp MA 4642 | Saprotroph | Undefined Saprotroph | 63 |
| Ascomycota | Eurotiomycetes | Chaetothyriales | Herpotrichiellaceae | Phaeomoniella | Phaeomoniella sp | Saprotroph | Undefined Saprotroph | 63 |
| Ascomycota | Leotiomycetes | unidentified | unidentified | unidentified | Leotiomycetes sp BLD3 | - | - | 63 |
| Ascomycota | Eurotiomycetes | Chaetothyriales | Herpotrichiellaceae | Phaeomoniella | Phaeomoniella sp | Saprotroph | Undefined Saprotroph | 62 |
| Ascomycota | Dothideomycetes | Capnodiales | | | | - | - | 61 |
| Ascomycota | | | | | | - | - | 60 |
| Ascomycota | Dothideomycetes | Capnodiales | Incertae sedis | Phaeotheca | Phaeotheca sp | - | - | 58 |
| Ascomycota | Eurotiomycetes | Chaetothyriales | Herpotrichiellaceae | Phaeomoniella | Phaeomoniella sp | Saprotroph | Undefined Saprotroph | 55 |
| Ascomycota | Leotiomycetes | Helotiales | Hyaloscyphaceae | Lachnellula | Lachnellula calyciformis | Saprotroph | Undefined Saprotroph | 55 |
| Ascomycota | Leotiomycetes | Helotiales | Hyaloscyphaceae | Lachnellula | | Saprotroph | Undefined Saprotroph | 55 |
| Ascomycota | Dothideomycetes | Pleosporales | | | | - | - | 54 |
| | | | | | | - | - | 51 |
| Ascomycota | Leotiomycetes | Helotiales | Hyaloscyphaceae | | | - | - | 51 |
| Ascomycota | Eurotiomycetes | Chaetothyriales | | | | - | - | 50 |
| Ascomycota | Eurotiomycetes | Chaetothyriales | Herpotrichiellaceae | | | - | - | 49 |
| unidentified | unidentified | unidentified | unidentified | unidentified | fungal sp TRN256 | - | - | 47 |
| Ascomycota | Dothideomycetes | Pleosporales | Pleosporaceae | Alternaria | | Pathotroph | Plant Pathogen | 46 |

**Table 3:** Taxonomie of the 30 more frequent OTUs (number of samples)

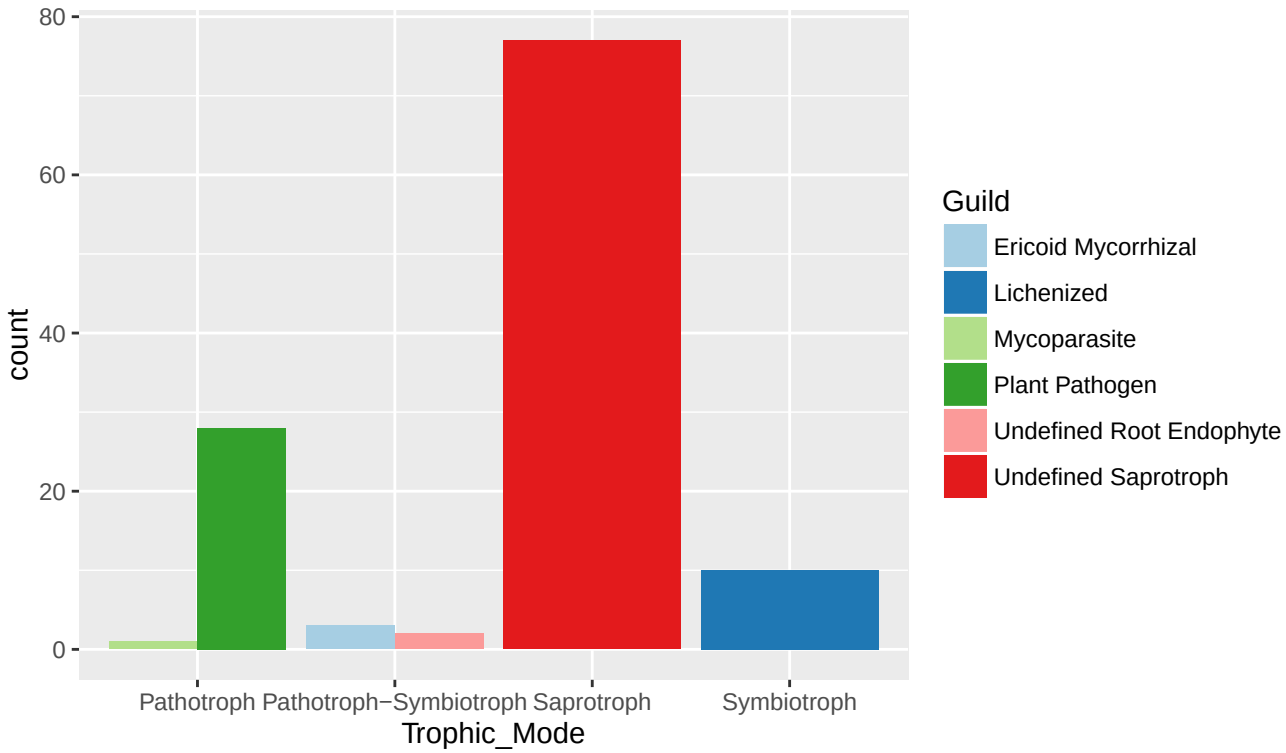


**Figure 4.1:** Distribution of OTUs into functional Guild.

# 4 Number of sequences and OTUs in function of putative ecology (using FUNGuild software; Nguyen et al, 2015)

```
tabPutativeEcology <- apply(data.f3@tax_table, 2, function(x) table(x))
tabPutativeEcology_percent <- apply(data.f3@tax_table, 2, function(x)
  round(table(x)/dim(data.f3@tax_table)[1]*100, 3))
sum(data.f3@otu_table[data.f3@tax_table[,"Trophic_Mode"]=="-"]) /
  sum(data.f3@otu_table)*100

## [1] 79.43185

tmdata <- as.data.frame(data.f3@tax_table[data.f3@tax_table[,"Trophic_Mode"]!="-"])
tmdata$Nb.sequences <- rowSums(data.f3@otu_table[data.f3@tax_table[,"Trophic_Mode"]!="-"])
tmdata$Nb.OTU <- rep(1, length(tmdata$Nb.sequences))
```

```
ggplot(tmdata) + geom_bar(aes(x= Trophic_Mode, fill=Guild), position = "dodge") +
  scale_fill_discrete("Paired")+ theme_grey()
```

```
ggplot(tmdata, stat="identity") +
  geom_bar(aes(x= Trophic_Mode, weight = Nb.sequences,  fill=Guild), position = "dodge") +
  scale_fill_discrete("Paired") + scale_y_log10() + theme_grey()
```

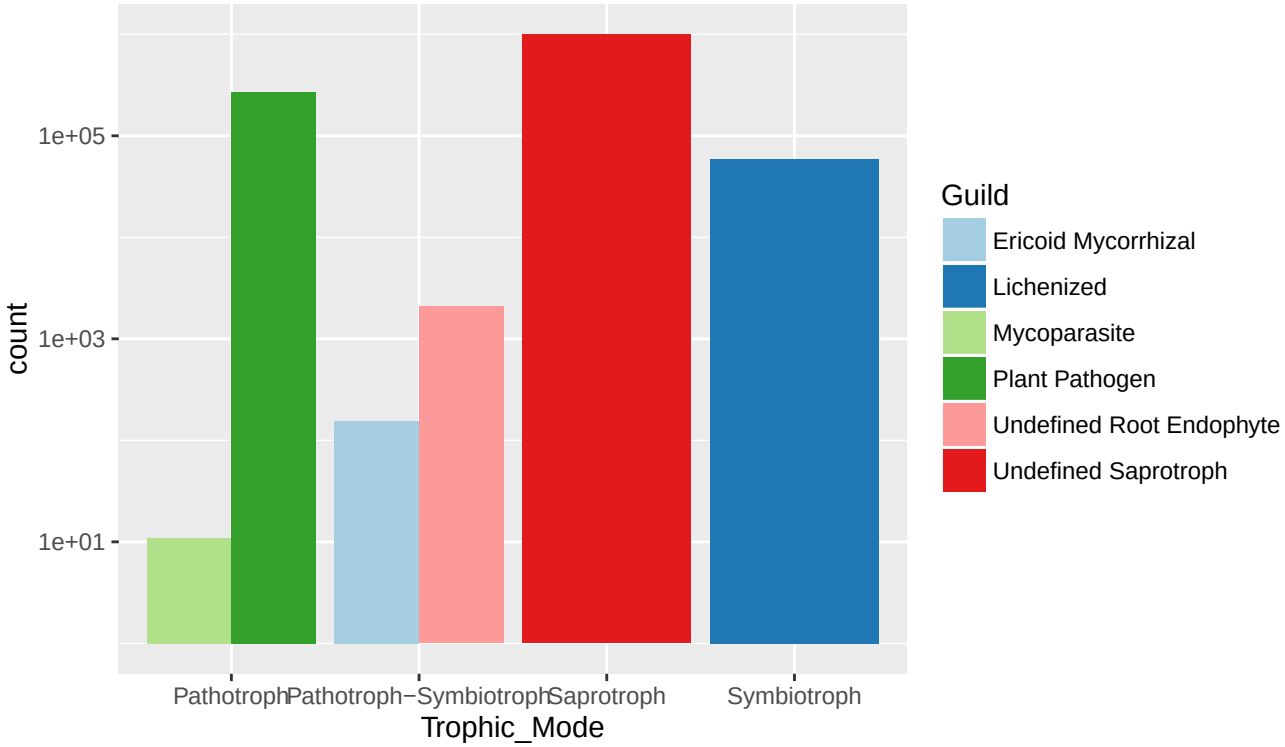


**Figure 4.2:** Distribution of sequences (log10 transformed) into functional Guild.

# 5 Distribution of fungal endophytic alpha-biodiversity

## 5.1 Local diversity = Diversity by sites

```
accu_plot(data.f3, "Sites", nbSeq = FALSE)
```

```
accu_plot(data.f3, "Sites", step = 5000)
```

```
measures_index <- c("Observed", "Chao1", "Shannon", "Simpson")
p <- plot_richness(data.f3, x = "Sites", color = "Sites", measures = measures_index)
p + geom_boxplot(data = p$data, alpha = 0.5)
```

## 5.2 Diversity by age of tree

```
accu_plot(data.f3, "Age", nbSeq = FALSE)
```

```
measures_index <- c("Observed", "Chao1", "Shannon", "Simpson")
p <- plot_richness(data.f3, x = "Age", color = "Sites", measures = measures_index)
p + geom_boxplot(data = p$data, aes(x = p$data$Age, y = value, color = NULL),
                 alpha = 0.5)
```

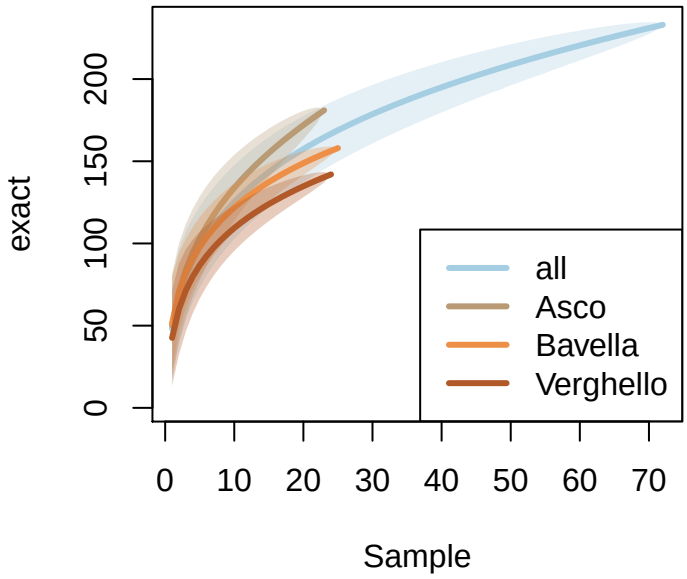


**Figure 5.1:** Rarefaction curves for each sites. Notes that if singletons were removed, these curves are biaised.

## 5.3 Diversity by elevation of the sample

```
accu_plot(data.f3, "Elevation", nbSeq = FALSE)
```

```
measures_index <- c("Observed", "Chao1", "Shannon", "Simpson")
p <- plot_richness(data.f3, x = "Elevation", color = "Sites", measures = measures_index)
p + geom_boxplot(data = p$data, aes(x = p$data$Elevation, y = value, color = NULL),
                 alpha = 0.5)
```

## 5.4 Which factor affect diversity?

```
## Uneven sequencing depth may have an impact
readNumbers = apply(t(data.f3@otu_table), 1, sum)

otuHill <- renyi(t(data.f3@otu_table), scale = c(0, 1, 2), hill = T)

hill.1 = otuHill$"0"
hill.2 = otuHill$"1"
hill.3 = otuHill$"2"

hill.1.m1 = lm(hill.1 ~ sqrt(readNumbers) + data.f3@sam_data$Sites +
                 data.f3@sam_data$Age + data.f3@sam_data$Elevation)
```

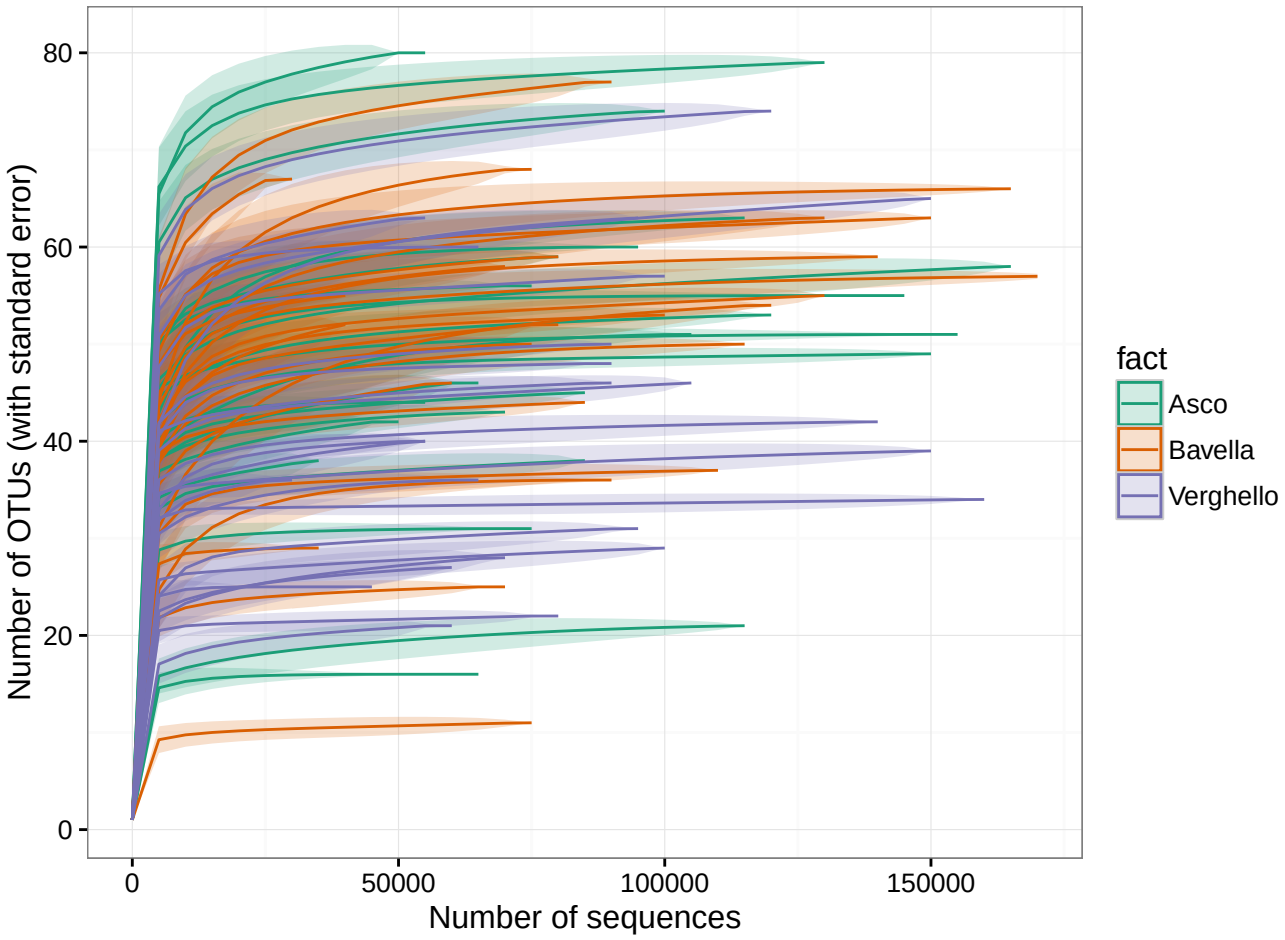


**Figure 5.2:** Rarefaction curves for each samples using sequences number on x-axes. Notes that if singletons were removed, these curves are biaised.

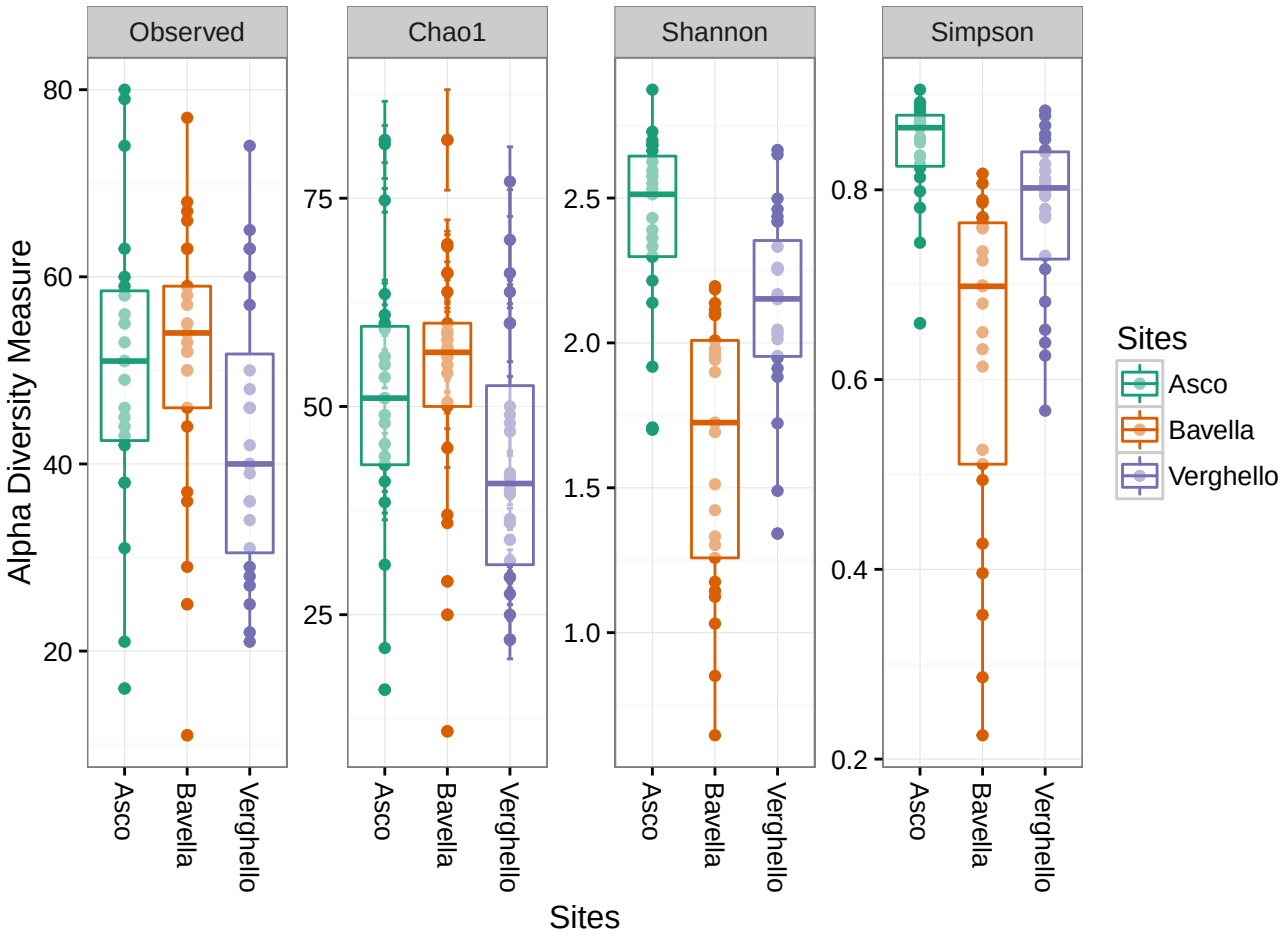


**Figure 5.3:** Diversity of each sites

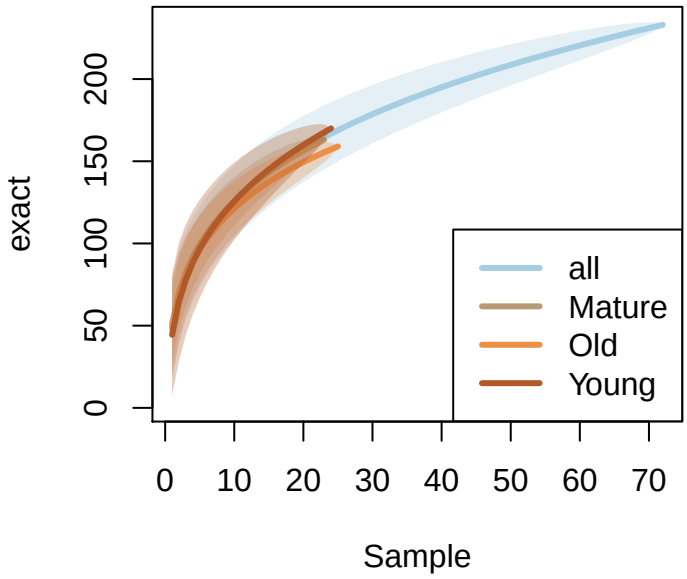


**Figure 5.4:** Rarefaction curves for each tree age modalities. Notes that if singletons were removed, these curves are biaised.

```
hill.2.m1 = lm(hill.2 ~ sqrt(readNumbers) + data.f3@sam_data$Sites +
                 data.f3@sam_data$Age + data.f3@sam_data$Elevation)
hill.3.m1 = lm(hill.3 ~ sqrt(readNumbers) + data.f3@sam_data$Sites +
                 data.f3@sam_data$Age + data.f3@sam_data$Elevation)
```

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 27.0377659 | 9.4442299 | 2.8628873 | 0.0056684 |
| sqrt(readNumbers) | 0.0755160 | 0.0279419 | 2.7026053 | 0.0087990 |
| data.f3@sam_data$SitesBavella | 1.0370913 | 4.2018371 | 0.2468185 | 0.8058387 |
| data.f3@sam_data$SitesVerghello | -6.8254256 | 4.2448276 | -1.6079394 | 0.1127711 |
| data.f3@sam_data$AgeOld | -1.1132151 | 4.2063030 | -0.2646540 | 0.7921265 |
| data.f3@sam_data$AgeYoung | -7.4022901 | 4.3206519 | -1.7132346 | 0.0915101 |
| data.f3@sam_data$ElevationMiddle | 6.6296207 | 4.2620963 | 1.5554835 | 0.1247625 |
| data.f3@sam_data$ElevationTop | 4.1963758 | 4.2136880 | 0.9958914 | 0.3230534 |

**Table 4:** Summary of the linear model of species richness (Hill number 1 (q = 0)

Post-hoc Tukey tests among the three experimental treatments with partial residuals, after accounting for differential sequencing success.

```
tuk1 <- TukeyHSD(aov(lm(hill.1 ~ sqrt(readNumbers))$residuals ~ data.f3@sam_data$Site))
tuk2 <- TukeyHSD(aov(lm(hill.2 ~ sqrt(readNumbers))$residuals ~ data.f3@sam_data$Site))
tuk3 <- TukeyHSD(aov(lm(hill.3 ~ sqrt(readNumbers))$residuals ~ data.f3@sam_data$Site))
```

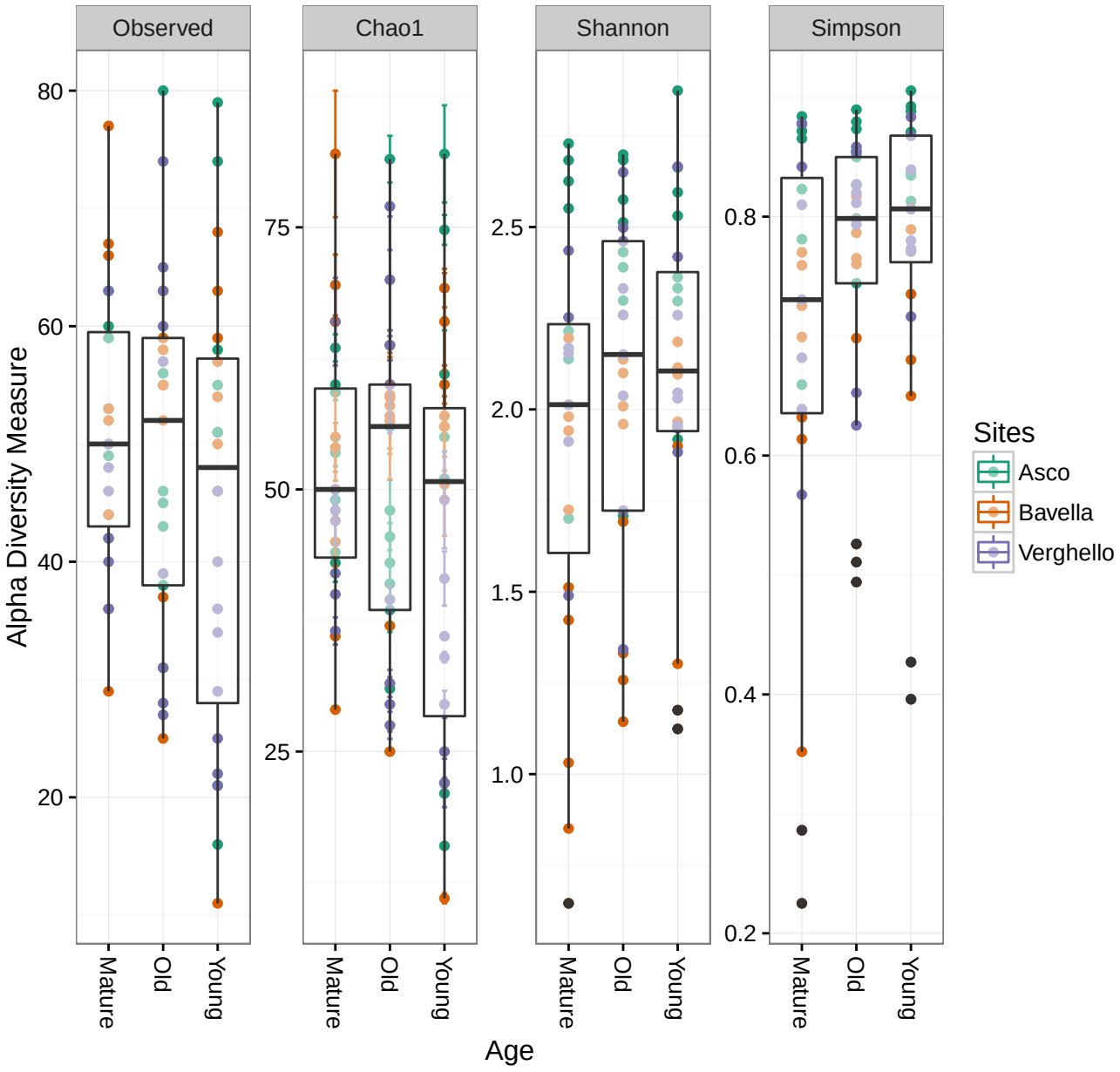


**Figure 5.5:** Diversity in function of tree age. Color represent sites.

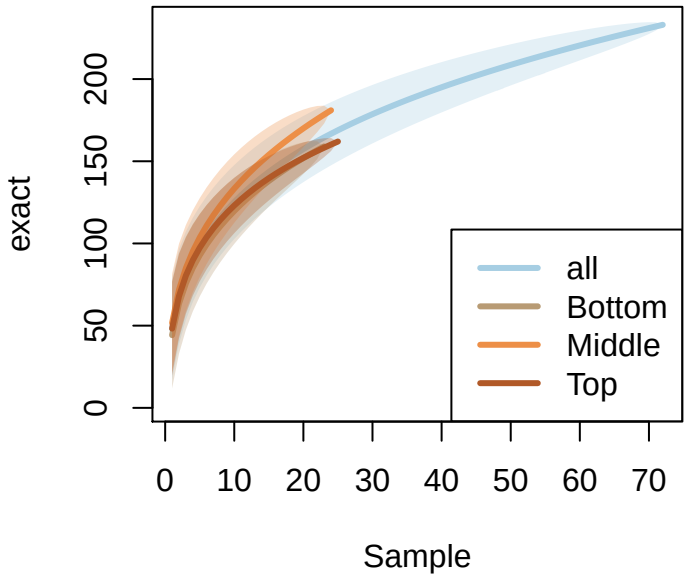


**Figure 5.6:** Rarefaction curves for each elevation. Notes that if singletons were removed, these curves are biaised.

```
ggplot(data = df) + geom_linerange(aes(ymax = xSup, ymin = xInf, x = y), size = 2) +
  geom_point(aes(x=y, y=x), size=4, shape=21, fill="white") +
  coord_flip() + theme_gray() + geom_hline(yintercept = 0) +
  ylab("Differences in mean levels") + xlab("")
```

# 6 Effect of site, age and elevation on fungal endophytic beta-diversity

## 6.1 Venn diagramm

```
venn_phyloseq(data.f3, "Sites", printValues = F)
```

```
venn_phyloseq(data.f3, "Age", printValues = F)
```

```
venn_phyloseq(data.f3, "Elevation", printValues = F)
```

## 6.2 Ordination

Ordination of the OTUs table using NMDS (Non-metric MultiDimenstional Scaling).

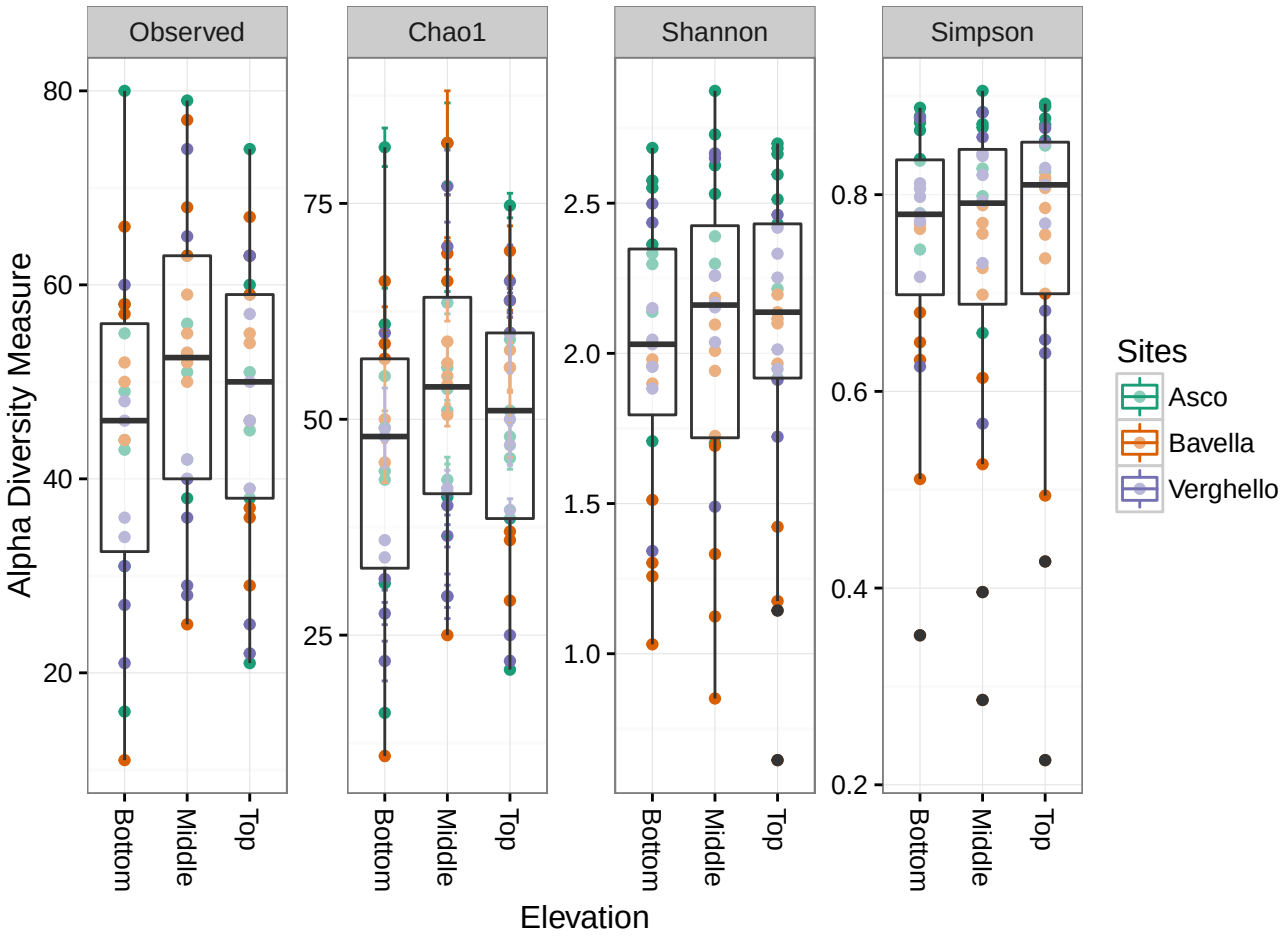


**Figure 5.7:** Diversity in function of elevation. Color represent sites.

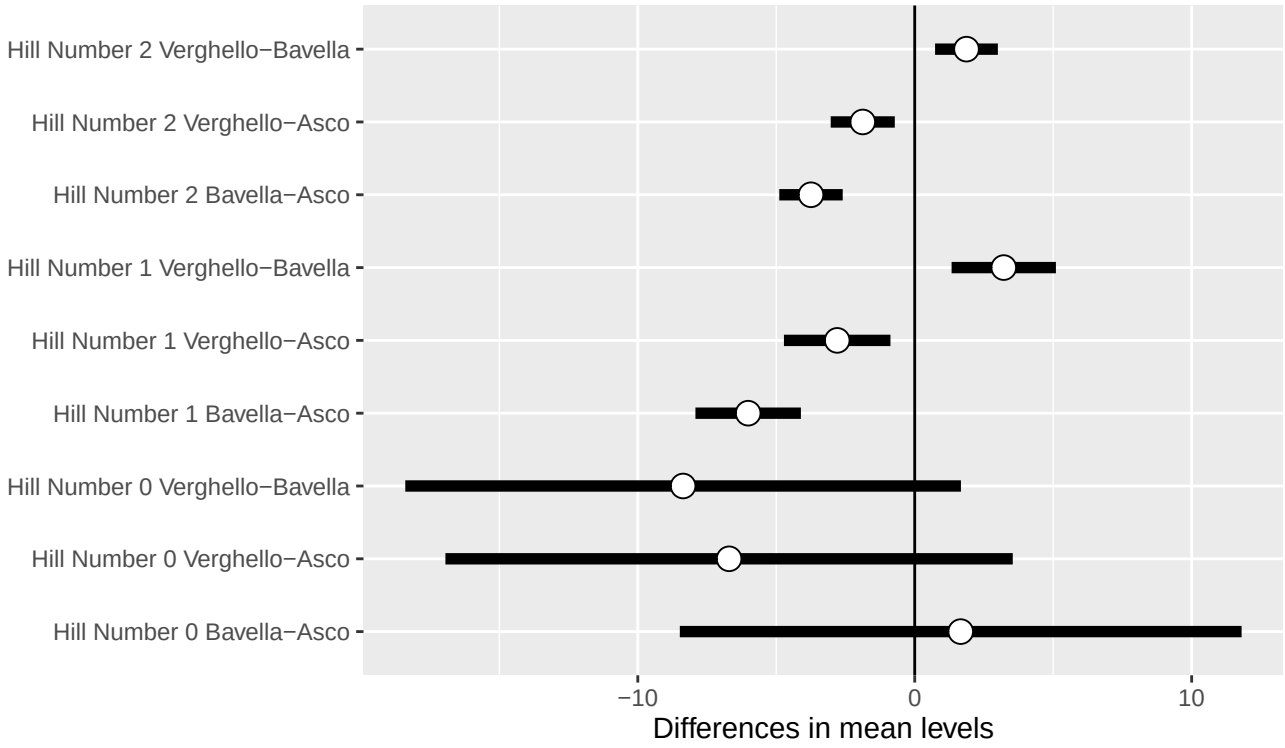


**Figure 5.8:** Results of the Tuckey HSD testing for differences in mean Hill numbers among pairs of modalities

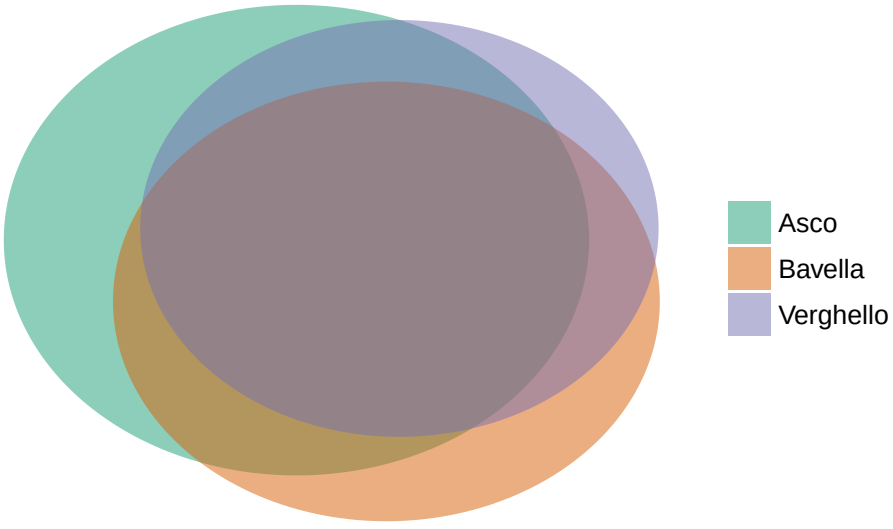


**Figure 6.1:** Venn diagramm ef the distribution of OTUs among Sites

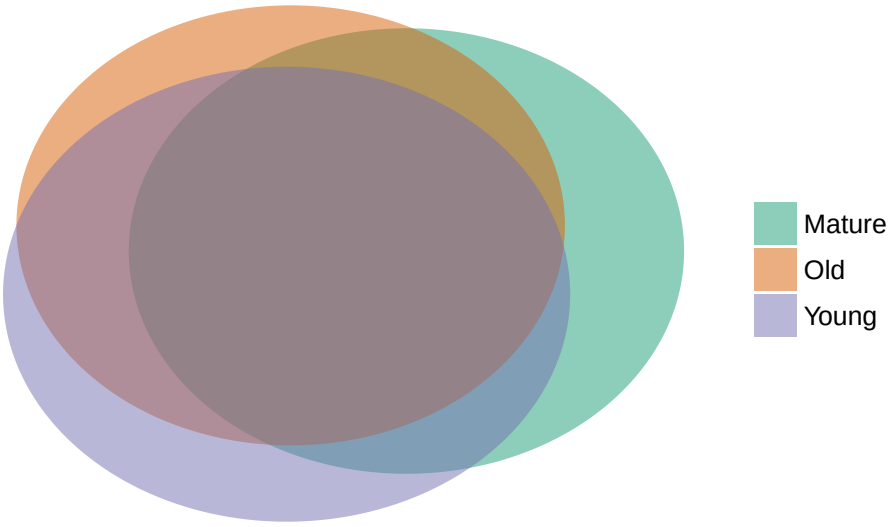


**Figure 6.2:** Venn diagramm ef the distribution of OTUs among host age

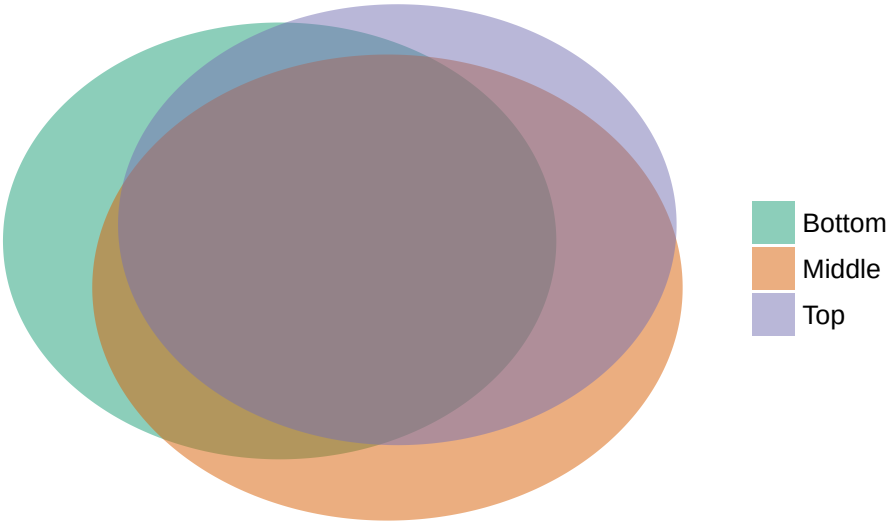


**Figure 6.3:** Venn diagramm ef the distribution of OTUs among elevation of samples

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 9.7457312 | 1.8007852 | 5.4119343 | 0.0000010 |
| sqrt(readNumbers) | 0.0016818 | 0.0053278 | 0.3156609 | 0.7532870 |
| data.f3@sam_data$SitesBavella | -6.0579667 | 0.8011882 | -7.5612276 | 0.0000000 |
| data.f3@sam_data$SitesVerghello | -2.8490706 | 0.8093855 | -3.5200416 | 0.0008006 |
| data.f3@sam_data$AgeOld | 0.8624918 | 0.8020398 | 1.0753729 | 0.2862452 |
| data.f3@sam_data$AgeYoung | 0.9578824 | 0.8238433 | 1.1626997 | 0.2492685 |
| data.f3@sam_data$ElevationMiddle | 1.3526575 | 0.8126782 | 1.6644442 | 0.1009127 |
| data.f3@sam_data$ElevationTop | 1.0510900 | 0.8034479 | 1.3082242 | 0.1954764 |

**Table 5:** Summary of the linear model of the exponential of Shannons entropy index (Hill number 2 (q = 1)

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 6.3011288 | 1.0754044 | 5.8593111 | 0.0000002 |
| sqrt(readNumbers) | -0.0006218 | 0.0031817 | -0.1954387 | 0.8456687 |
| data.f3@sam_data$SitesBavella | -3.7412953 | 0.4784587 | -7.8194739 | 0.0000000 |
| data.f3@sam_data$SitesVerghello | -1.9015445 | 0.4833540 | -3.9340618 | 0.0002082 |
| data.f3@sam_data$AgeOld | 0.5320685 | 0.4789672 | 1.1108662 | 0.2707820 |
| data.f3@sam_data$AgeYoung | 1.0402492 | 0.4919880 | 2.1143792 | 0.0383839 |
| data.f3@sam_data$ElevationMiddle | 0.4753999 | 0.4853204 | 0.9795590 | 0.3309923 |
| data.f3@sam_data$ElevationTop | 0.4678373 | 0.4798081 | 0.9750507 | 0.3332062 |

**Table 6:** Summary of the linear model of inverse of Simpsons concentration index (Hill number 3 (q = 2)

```
my.ord.nmds <- ordinate(data.f3, method = "NMDS")
my.ord.nmds$stress
```

```
stressplot(my.ord.nmds)
```

```
p <- plot_ordination(data.f3, my.ord.nmds, color = "Sites", shape = "Age")
p + geom_point(size = 0) +
  geom_text(aes(x = p$data$NMDS1, y = p$data$NMDS2,
                label = as.character(as.vector(data.f3@sam_data[, "CODE"]$CODE))))
```

```
my.ord.nmds_gower <- ordinate(data.f3, distance = "gower",  method = "NMDS")

## Square root transformation
## Wisconsin double standardization
## Run 0 stress 0.2451014
## Run 1 stress 0.4083446
## Run 2 stress 0.2452108
## ... Procrustes: rmse 0.02660568  max resid 0.1488458
## Run 3 stress 0.2449015
## ... New best solution
## ... Procrustes: rmse 0.0524592  max resid 0.2614784
```

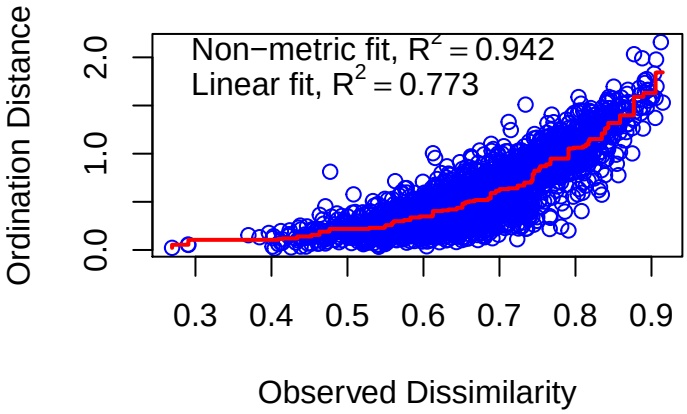


**Figure 6.4:** Stress plot of the NMDS

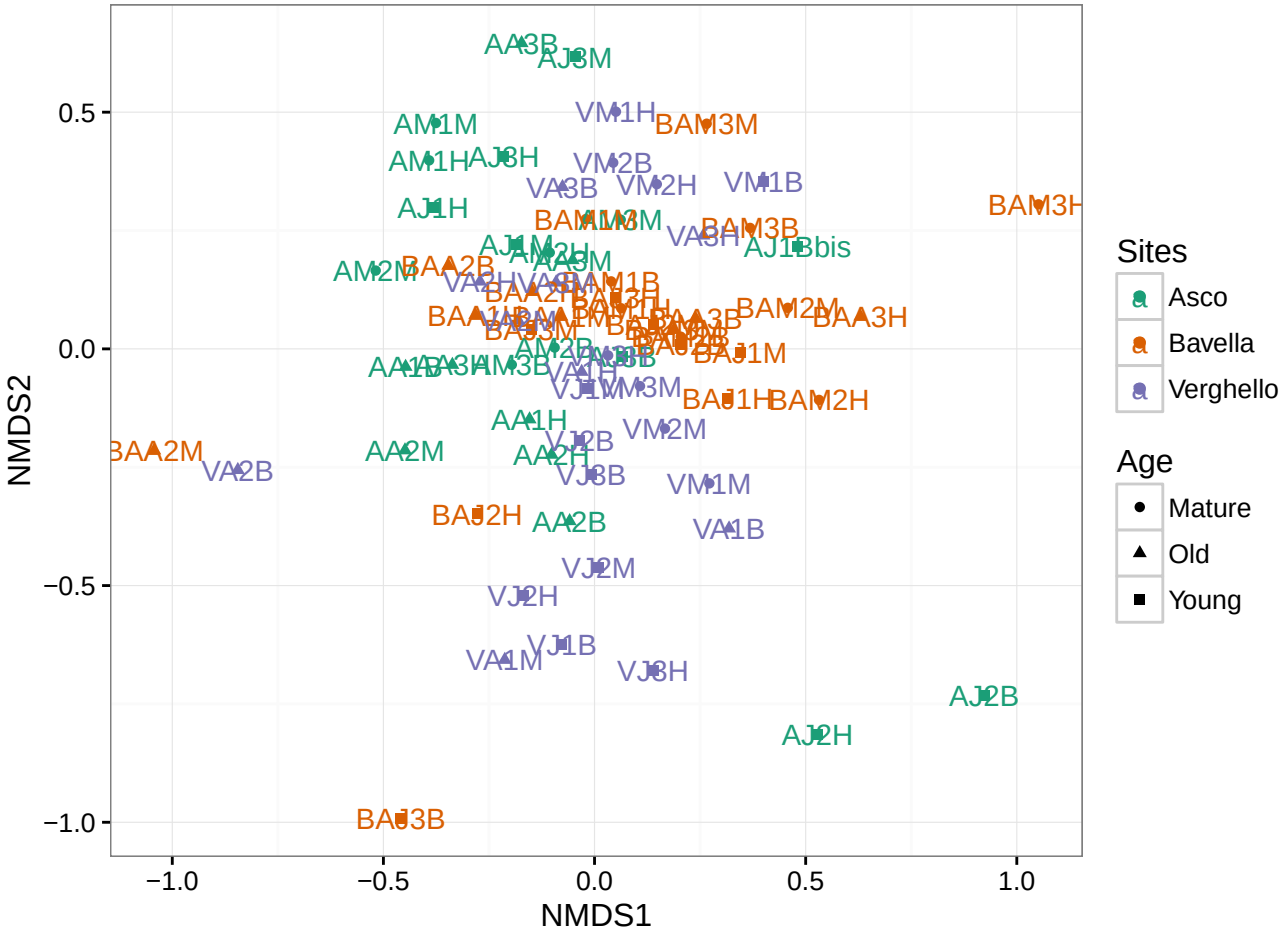


**Figure 6.5:** NMDS of OTU table. Colors represent sites and shape the age of tree.

```
## Run 4 stress 0.2646983
## Run 5 stress 0.2427006
## ... New best solution
## ... Procrustes: rmse 0.04659547  max resid 0.2705798
## Run 6 stress 0.2451439
## Run 7 stress 0.2449139
## Run 8 stress 0.2537435
## Run 9 stress 0.2537289
## Run 10 stress 0.2489884
## Run 11 stress 0.260967
## Run 12 stress 0.2442576
## Run 13 stress 0.2534864
## Run 14 stress 0.2442781
## Run 15 stress 0.2594249
## Run 16 stress 0.260253
## Run 17 stress 0.2429693
## ... Procrustes: rmse 0.01077865  max resid 0.06500734
## Run 18 stress 0.2433502
## Run 19 stress 0.2438756
## Run 20 stress 0.2552936
## *** No convergence -- monoMDS stopping criteria:
##      3: no. of iterations >= maxit
##     17: stress ratio > sratmax

my.ord.PCoA <- ordinate(data.f3, method = "PCoA")
my.ord.PCoA_gower <- ordinate(data.f3, distance = "gower", method = "PCoA")
my.ord.DCA <- ordinate(data.f3, method = "DCA")
my.ord.DCA_gower <- ordinate(data.f3, distance = "gower", method = "DCA")

p_NMDS_BRAY <- plot_ordination(data.f3, my.ord.nmds, color = "Sites",
                               shape = "Age", label = "CODE") + geom_point(size = 5)
p_NMDS_GOWER <- plot_ordination(data.f3, my.ord.nmds_gower, color = "Sites",
                                shape = "Age", label = "CODE") + geom_point(size = 5)
p_PCoA_BRAY <- plot_ordination(data.f3, my.ord.PCoA, color = "Sites",
                               shape = "Age", label = "CODE") + geom_point(size = 5)
p_PCoA_GOWER <- plot_ordination(data.f3, my.ord.PCoA_gower, color = "Sites",
                                shape = "Age", label = "CODE") + geom_point(size = 5)
p_DCA_BRAY <- plot_ordination(data.f3, my.ord.DCA, color = "Sites",
                              shape = "Age", label = "CODE") + geom_point(size = 5)
p_DCA_GOWER <- plot_ordination(data.f3, my.ord.DCA_gower, color = "Sites",
                               shape = "Age", label = "CODE") + geom_point(size = 5)
```

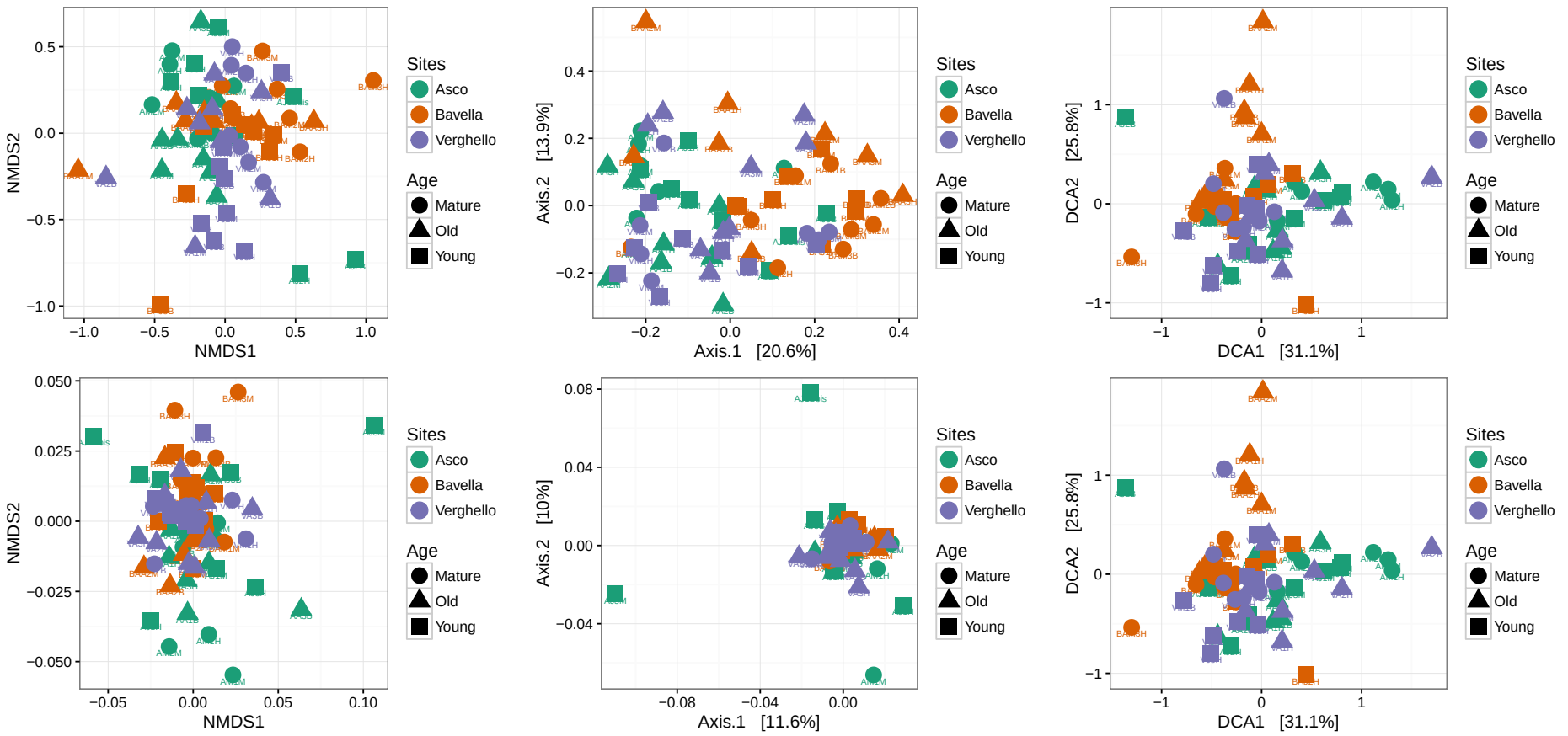


**Figure 6.6:** Comparison of different distances (bray (up) and gower (bottom)) and ordination methods (NMDS (left), PCoA (center) and DCA (right)).

```
multiplot(p_NMDS_BRAY, p_NMDS_GOWER, p_PCoA_BRAY, p_PCoA_GOWER, p_DCA_BRAY, p_DCA_GOWER,
          cols = 3)
```

## 6.3 Permanova on sites, host ages and elevation

```
sam_data <- as.data.frame(unclass(data.f3@sam_data))
sam_data$IndividualTree <- paste(sam_data$Sites, sam_data$Age, sam_data$Tree)
res.ado <- adonis(t(data.f3@otu_table) ~  Sites * Age * Elevation, sam_data,
                  permutation = 9999)
```

If we only keep the 50 OTUs present in more than 30 sample, the Permanova results is the following:

```
res.ado_sampMin30 <- adonis(t(data.f3@otu_table[rowSums(data.f3@otu_table>0)>=30,]) ~
                            Sites * Age * Elevation, sam_data, permutation = 9999)
```

```
xtable(res.ado$aov.tab, caption = "Result of the permanova on abundances
       (number of sequence).")
```

| | Df | SumsOfSqs | MeanSqs | F.Model | R2 | Pr(>F) |
|---|---|---|---|---|---|---|
| Sites | 2 | 1.61 | 0.81 | 5.30 | 0.13 | 0.0001 |
| Age | 2 | 0.55 | 0.28 | 1.82 | 0.04 | 0.0176 |
| Elevation | 2 | 0.43 | 0.21 | 1.41 | 0.03 | 0.1006 |
| Sites:Age | 4 | 1.31 | 0.33 | 2.15 | 0.10 | 0.0003 |
| Sites:Elevation | 4 | 0.52 | 0.13 | 0.85 | 0.04 | 0.7382 |
| Age:Elevation | 4 | 0.53 | 0.13 | 0.87 | 0.04 | 0.7018 |
| Sites:Age:Elevation | 8 | 1.06 | 0.13 | 0.87 | 0.08 | 0.7889 |
| Residuals | 45 | 6.84 | 0.15 | | 0.53 | |
| Total | 71 | 12.85 | | | 1.00 | |

**Table 7:** Result of the permanova on abundances (number of sequence).

```
xtable(res.ado_sampMin30$aov.tab, caption = "Result of the permanova on abundances
       (number of sequence) using only OTUs present in more than 30 samples")
```

| | Df | SumsOfSqs | MeanSqs | F.Model | R2 | Pr(>F) |
|---|---|---|---|---|---|---|
| Sites | 2 | 1.59 | 0.79 | 5.42 | 0.13 | 0.0001 |
| Age | 2 | 0.54 | 0.27 | 1.84 | 0.04 | 0.0186 |
| Elevation | 2 | 0.42 | 0.21 | 1.43 | 0.03 | 0.1015 |
| Sites:Age | 4 | 1.28 | 0.32 | 2.18 | 0.10 | 0.0004 |
| Sites:Elevation | 4 | 0.49 | 0.12 | 0.84 | 0.04 | 0.7375 |
| Age:Elevation | 4 | 0.51 | 0.13 | 0.86 | 0.04 | 0.7051 |
| Sites:Age:Elevation | 8 | 1.01 | 0.13 | 0.86 | 0.08 | 0.7915 |
| Residuals | 45 | 6.59 | 0.15 | | 0.53 | |
| Total | 71 | 12.41 | | | 1.00 | |

**Table 8:** Result of the permanova on abundances (number of sequence) using only OTUs present in more than 30 samples

```
res.ado_bin <- adonis(t(as.binaryOtuTable(data.f3)@otu_table) ~  Sites * Age *
                        Elevation, sam_data, permutation = 9999)
```

```
xtable(res.ado_bin$aov.tab, caption = "Result of the permanova on OTUs
       (each OTU is representing by one sequence)).")
```

| | Df | SumsOfSqs | MeanSqs | F.Model | R2 | Pr(>F) |
|---|---|---|---|---|---|---|
| Sites | 2 | 0.52 | 0.26 | 2.84 | 0.08 | 0.0002 |
| Age | 2 | 0.31 | 0.16 | 1.72 | 0.05 | 0.0167 |
| Elevation | 2 | 0.18 | 0.09 | 0.96 | 0.03 | 0.4961 |
| Sites:Age | 4 | 0.45 | 0.11 | 1.22 | 0.07 | 0.1275 |
| Sites:Elevation | 4 | 0.29 | 0.07 | 0.80 | 0.04 | 0.8671 |
| Age:Elevation | 4 | 0.35 | 0.09 | 0.95 | 0.05 | 0.5666 |
| Sites:Age:Elevation | 8 | 0.61 | 0.08 | 0.84 | 0.09 | 0.8747 |
| Residuals | 45 | 4.10 | 0.09 | | 0.60 | |
| Total | 71 | 6.81 | | | 1.00 | |

**Table 9:** Result of the permanova on OTUs (each OTU is representing by one sequence)).

## 6.4 Permanova on sites, host ages and individual trees

```
res.ado_Tree <- adonis(t(data.f3@otu_table) ~   Sites*Age + Sites:Age:IndividualTree ,
                       sam_data, permutation = 9999)
res.ado_Tree_sampMin30 <- adonis(t(data.f3@otu_table[rowSums(data.f3@otu_table>0)>=30,]) ~
                                   Sites*Age + Sites:Age:IndividualTree , sam_data,
                                 permutation = 9999)
res.ado_Tree_bin <- adonis(t(as.binaryOtuTable(data.f3)@otu_table) ~
                             Sites*Age + Sites:Age:IndividualTree , sam_data,
                           permutation = 9999)
```

```
xtable(res.ado_Tree$aov.tab, caption = "Result of the permanova on abundances
       (number of sequence).")
```

| | Df | SumsOfSqs | MeanSqs | F.Model | R2 | Pr(>F) |
|---|---|---|---|---|---|---|
| Sites | 2 | 1.61 | 0.81 | 6.50 | 0.13 | 0.0001 |
| Age | 2 | 0.55 | 0.28 | 2.23 | 0.04 | 0.0026 |
| Sites:Age | 4 | 1.32 | 0.33 | 2.67 | 0.10 | 0.0001 |
| Sites:Age:IndividualTree | 18 | 3.78 | 0.21 | 1.69 | 0.29 | 0.0001 |
| Residuals | 45 | 5.59 | 0.12 | | 0.43 | |
| Total | 71 | 12.85 | | | 1.00 | |

**Table 10:** Result of the permanova on abundances (number of sequence).

```
xtable(res.ado_Tree_sampMin30$aov.tab, caption = "Result of the permanova on abundances
       (number of sequence) using only OTUs present in more than 30 samples")
```

| | Df | SumsOfSqs | MeanSqs | F.Model | R2 | Pr(>F) |
|---|---|---|---|---|---|---|
| Sites | 2 | 1.59 | 0.79 | 6.65 | 0.13 | 0.0001 |
| Age | 2 | 0.54 | 0.27 | 2.26 | 0.04 | 0.0018 |
| Sites:Age | 4 | 1.29 | 0.32 | 2.71 | 0.10 | 0.0001 |
| Sites:Age:IndividualTree | 18 | 3.63 | 0.20 | 1.69 | 0.29 | 0.0001 |
| Residuals | 45 | 5.37 | 0.12 | | 0.43 | |
| Total | 71 | 12.41 | | | 1.00 | |

**Table 11:** Result of the permanova on abundances (number of sequence) using only OTUs present in more than 30 samples

```
xtable(res.ado_Tree_bin$aov.tab, caption = "Result of the permanova on OTUs
       (each OTU is representing by one sequence)).")
```

| | Df | SumsOfSqs | MeanSqs | F.Model | R2 | Pr(>F) |
|---|---|---|---|---|---|---|
| Sites | 2 | 0.52 | 0.26 | 3.40 | 0.08 | 0.0002 |
| Age | 2 | 0.31 | 0.16 | 2.06 | 0.05 | 0.0023 |
| Sites:Age | 4 | 0.46 | 0.12 | 1.51 | 0.07 | 0.0135 |
| Sites:Age:IndividualTree | 18 | 2.08 | 0.12 | 1.52 | 0.31 | 0.0002 |
| Residuals | 45 | 3.43 | 0.08 | | 0.50 | |
| Total | 71 | 6.81 | | | 1.00 | |

**Table 12:** Result of the permanova on OTUs (each OTU is representing by one sequence)).

## 6.5 Differences in abundances and OTUs number by Order.

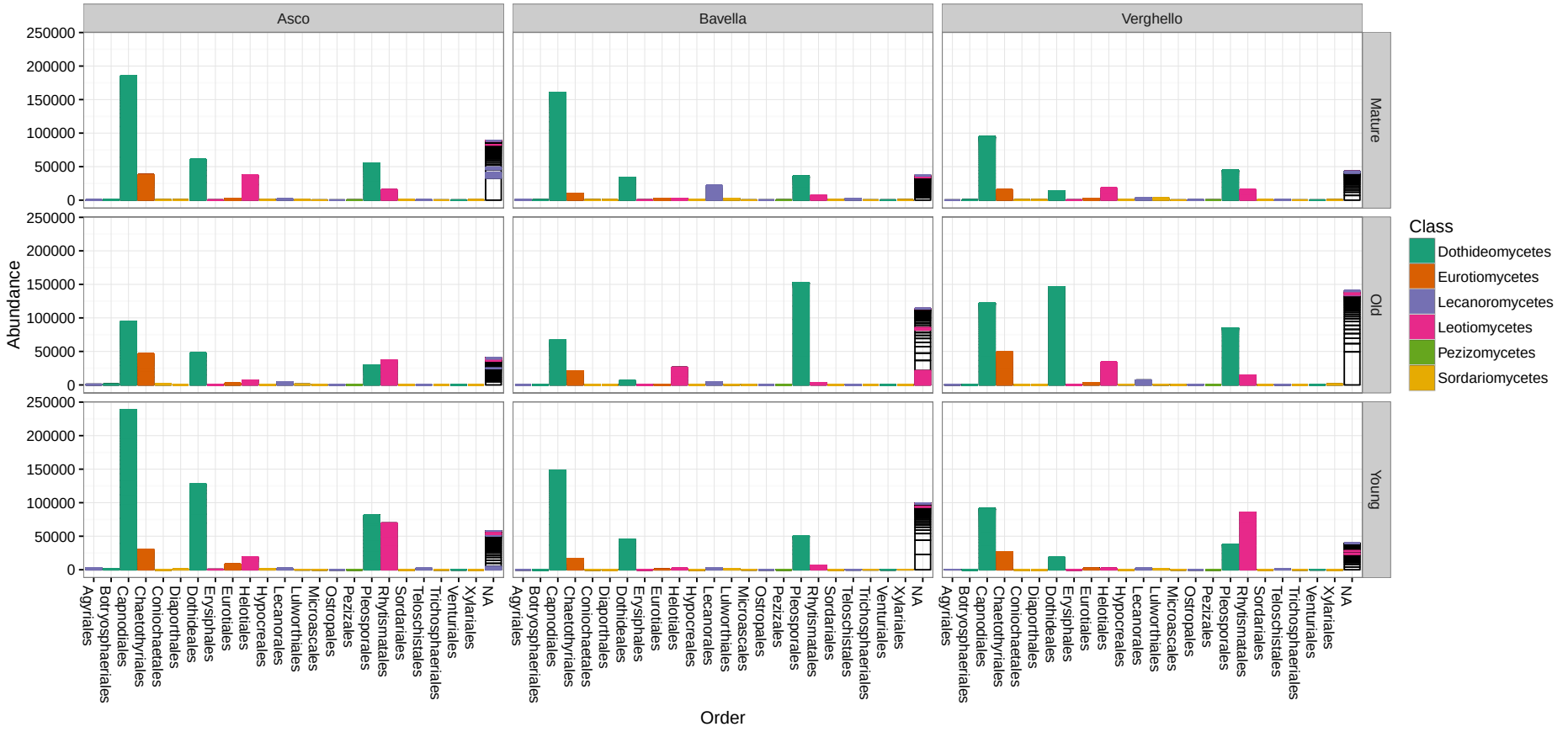


**Figure 6.7:** Taxonomic distribution of sequences in the different site * age combinaison.

```
data.f3_taxo_known <- prune_taxa(taxa_names(data.f3@tax_table)
          [!(data.f3@tax_table[, 4] %in% c("unidentified", "Incertae sedis"))], data.f3)
p <- plot_bar(data.f3_taxo_known, "Order", fill = "Class", facet_grid = Age ~ Sites)
p + geom_bar(aes(color = Class, fill = Class), stat = "identity", position = "stack")
```

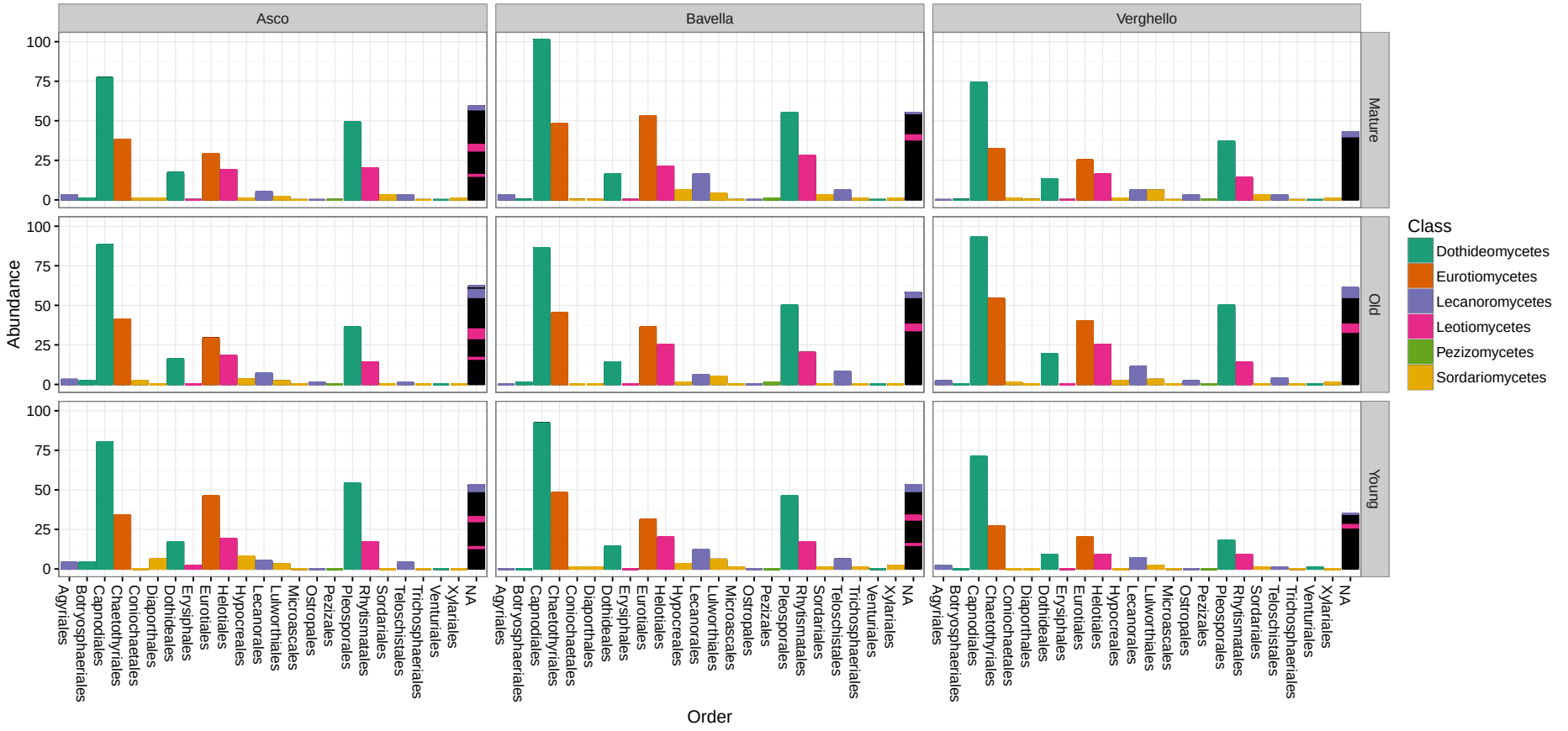


**Figure 6.8:** Taxonomic distribution of OTUs in the different site * age combinaison.

```
p <- plot_bar(as.binaryOtuTable(data.f3_taxo_known), "Order", fill = "Class",
              facet_grid = Age ~ Sites)
p + geom_bar(aes(color = Class, fill = Class), stat = "identity", position = "stack")
```


**Figure 6.9:** OTUs significantly different in terms of abundances between Verghello (positive values) and Asco (negative values)

## 6.6 Differences in abundances for each OTUs

### 6.6.1 Pairwise comparison of the OTUs composition by sites

```
library("DESeq2")
packageVersion("DESeq2")

## [1] '1.12.3'

data.f3_deseq2 <- phyloseq_to_deseq2(data.f3, ~ Sites)
data.f3_deseq2 <- DESeq(data.f3_deseq2, test = "Wald", fitType = "parametric")
res.f3_deseq2 <- results(data.f3_deseq2)
```

```
res_VA <- plot_deseq2_phyloseq(data.f3_deseq2, tax_table = data.f3@tax_table,
                               contrast = c("Sites", "Verghello", "Asco"),
                               taxa = "Species", color_tax = "Class")
res_VA
```

```
res_VB <- plot_deseq2_phyloseq(data.f3_deseq2, tax_table = data.f3@tax_table,
                               contrast = c("Sites", "Verghello", "Bavella"),
                               taxa = "Species", color_tax = "Class")
res_VB
```

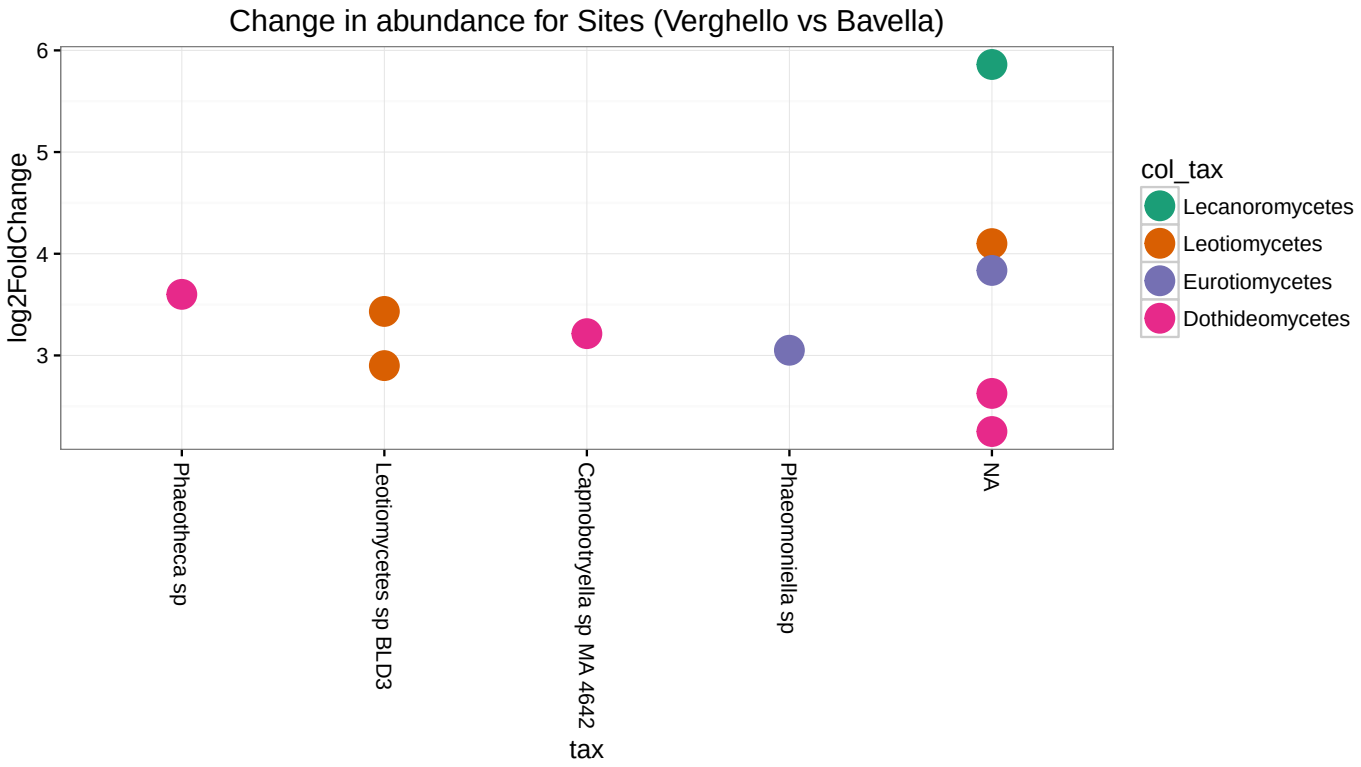


**Figure 6.10:** OTUs significantly different in terms of abundances between Verghello (positive values) and Bavella (negative values)

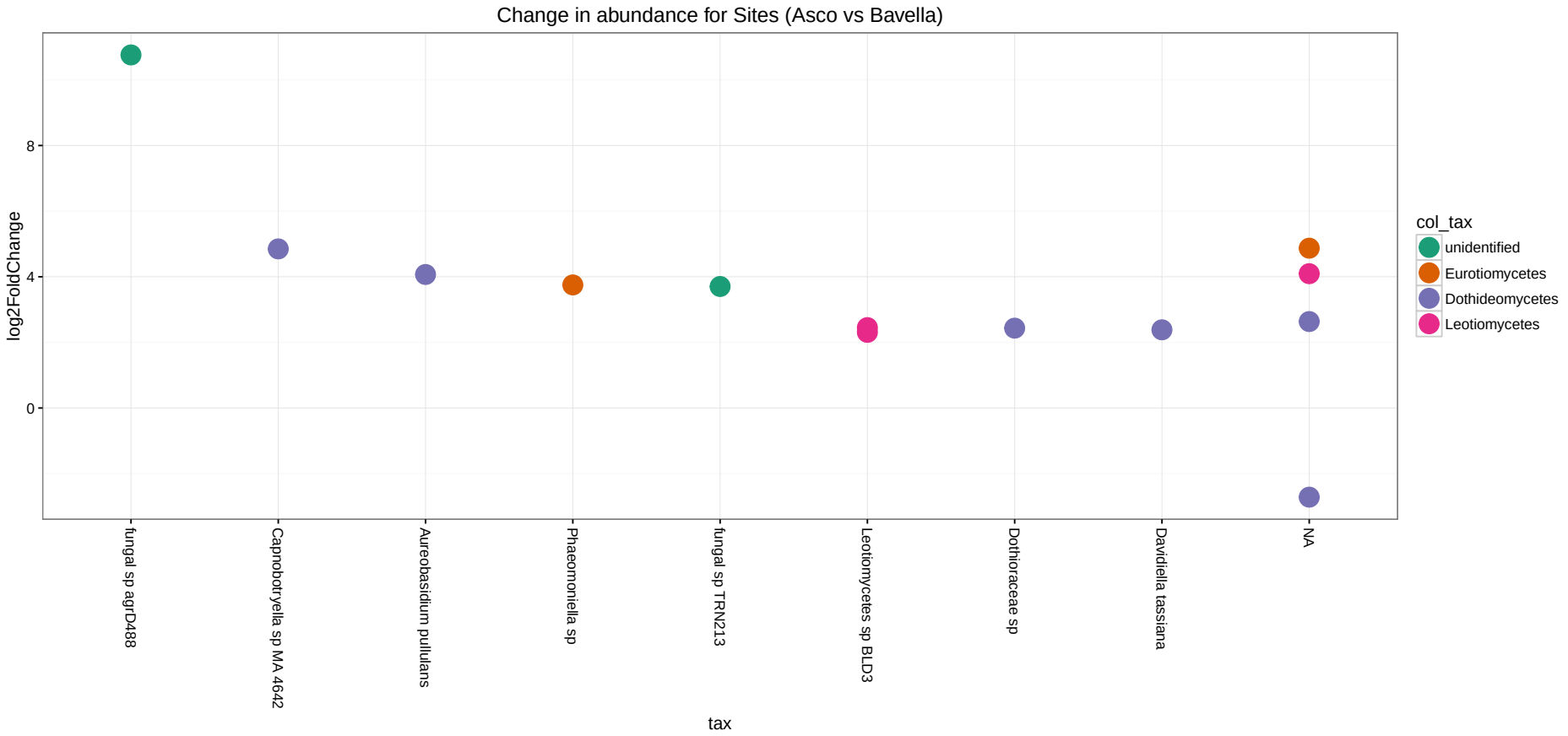


**Figure 6.11:** OTUs significantly different in terms of abundances between Asco (positive values) and Bavella (negative values)

```
res_AB <- plot_deseq2_phyloseq(data.f3_deseq2, tax_table = data.f3@tax_table,
                               contrast = c("Sites", "Asco", "Bavella"),
                               taxa = "Species", color_tax = "Class")
res_AB
```

| | Comparison | Species | Class | log2FoldChange (negative = more on second levels) |
|---|---|---|---|---|
| 1 | Verghello vs Asco | fungal sp agrD488 | unidentified | -7.91852173454766 |
| 2 | Verghello vs Bavella | Phaeomoniella sp | Eurotiomycetes | 3.05108123150561 |
| 3 | Verghello vs Bavella | | Dothideomycetes | 2.25172858097608 |
| 4 | Verghello vs Bavella | | Leotiomycetes | 4.09979744691054 |
| 5 | Verghello vs Bavella | Capnobotryella sp MA 4642 | Dothideomycetes | 3.21422973803097 |
| 6 | Verghello vs Bavella | Leotiomycetes sp BLD3 | Leotiomycetes | 3.43262525351344 |
| 7 | Verghello vs Bavella | | | 2.86541653606385 |
| 8 | Verghello vs Bavella | Phaeotheca sp | Dothideomycetes | 3.60064496739776 |
| 9 | Verghello vs Bavella | | Dothideomycetes | 2.62644677295921 |
| 10 | Verghello vs Bavella | | Eurotiomycetes | 3.83584829614036 |
| 11 | Verghello vs Bavella | Leotiomycetes sp BLD3 | Leotiomycetes | 2.89991498941365 |
| 12 | Verghello vs Bavella | | Lecanoromycetes | 5.86119968105531 |
| 13 | Asco vs Bavella | Phaeomoniella sp | Eurotiomycetes | 3.75100890773472 |
| 14 | Asco vs Bavella | | Dothideomycetes | 2.63479749902359 |
| 15 | Asco vs Bavella | | Leotiomycetes | 4.09351716256786 |
| 16 | Asco vs Bavella | | | 3.91942914222669 |
| 17 | Asco vs Bavella | Capnobotryella sp MA 4642 | Dothideomycetes | 4.84844225899154 |
| 18 | Asco vs Bavella | Leotiomycetes sp BLD3 | Leotiomycetes | 2.44958904773538 |
| 19 | Asco vs Bavella | | | 3.0359708877309 |
| 20 | Asco vs Bavella | Dothioraceae sp | Dothideomycetes | 2.43340283569607 |
| 21 | Asco vs Bavella | | Eurotiomycetes | 4.87040817171346 |
| 22 | Asco vs Bavella | Davidiella tassiana | Dothideomycetes | 2.38299410673209 |
| 23 | Asco vs Bavella | fungal sp TRN213 | unidentified | 3.70296099806453 |
| 24 | Asco vs Bavella | Leotiomycetes sp BLD3 | Leotiomycetes | 2.30562454059822 |
| 25 | Asco vs Bavella | | Dothideomycetes | -2.71760492205897 |
| 26 | Asco vs Bavella | Aureobasidium pullulans | Dothideomycetes | 4.06844436964516 |
| 27 | Asco vs Bavella | fungal sp agrD488 | unidentified | 10.760574211121 |

**Table 13:** OTUs showing differential abundances in the different sites.

### 6.6.2 Pairwise comparison of Order composition by sites

```
res_VA_o <- plot_deseq2_phyloseq(data.f3, contrast = c("Sites", "Verghello", "Asco"),
                                 taxDepth = "Order", color_tax = "Class")
res_VB_o <- plot_deseq2_phyloseq(data.f3, contrast = c("Sites", "Verghello", "Bavella"),
                                 taxDepth = "Order", color_tax = "Class")
res_AB_o <- plot_deseq2_phyloseq(data.f3, contrast = c("Sites", "Asco", "Bavella"),
                                 taxDepth = "Order", color_tax = "Class")
```

| | Comparison | Order | Class | log2FoldChange (negative = more on second levels) |
|---|---|---|---|---|
| 1 | Verghello vs Asco | None | None | None |
| 2 | Verghello vs Bavella | Dothideales | Dothideomycetes | 1.88407036655979 |
| 3 | Verghello vs Bavella | Rhytismatales | Leotiomycetes | 3.04476524722504 |
| 4 | Verghello vs Bavella | unidentified | Leotiomycetes | 2.28029302510296 |
| 5 | Asco vs Bavella | Capnodiales | Dothideomycetes | 0.706545841889738 |
| 6 | Asco vs Bavella | Chaetothyriales | Eurotiomycetes | 1.73298392435311 |
| 7 | Asco vs Bavella | Dothideales | Dothideomycetes | 2.01524198984585 |
| 8 | Asco vs Bavella | Eurotiales | Eurotiomycetes | 2.22677286037957 |
| 9 | Asco vs Bavella | Incertae sedis | Leotiomycetes | -0.81652237136615 |
| 10 | Asco vs Bavella | Rhytismatales | Leotiomycetes | 3.22335653047601 |
| 11 | Asco vs Bavella | unidentified | Leotiomycetes | 1.93274285148088 |

**Table 14:** Order showing differential abundances in the different sites.

# List of Figures



# List of Tables