Appendix S3: results after UPARSE clustering discarding unique sequences (Usearch function sortbysize with argument -minsize 2). Supplementary Materials of "Finding fungi in a needle stack: high alpha and low beta-diversity of foliar endophytic Ascomycetes revealed by metabarcoding in Corsican pine forests".

Adrien Taudiere*

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

November 9, 2017

Abstract

Plant leaves host highly diverse communities of foliar endophytic fungi (FEF). Compared to the other compartments of the plant microbiome, FEF diversity is poorly known. We here document the communities of FEF associated with the endemic Corsican black pine *Pinus nigra* subsp. *laricio* at three sites across its natural range and examine the effect of tree age and light exposure on FEF composition. Metabarcoding using next-generation sequencing provided 8243608 Ascomycota ITS2 sequences clustered into 642 FEF operational taxonomic units (OTUs). Site is the main determinant to explain the diversity and composition of FEF communities. Tree age somewhat affects FEF community composition, whereas needle location (shade vs canopy) has no effect. Results are robust against the various options of the bioinformatic pipeline specifically developed. This study provides the first picture of FEF diversity in a Mediterranean island and underlines the complementarity of forest massifs for fungal conservation.

Key words: foliar endophyte; fungi; community ecology; metabarcoding; Cyclaneusma minus, Pinus nigra subsp. laricio, Mediterranean, endemism, environmental sequencing

To set the filter parameter, see directly section 'Choice of filter parameters' 2.1.
To read a summary of this appendix, see directly section 'Summary' 7.

*adrien.taudiere@zaclys.net

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 analysed here was computed using UPARSE clustering discarding unique sequences (see main article and Sup. Mat. 1 for more details).

1.1 R requirements

First, set the working directory. In this directory, there is data folder and a R script "functions_for_phyloseq.R".

```
setwd("~/Nextcloud/GitHub/FEF_paper/")
```

Then, we may 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")
```

```
## May be needed under windows
Sys.setenv(JAVA_HOME = "C:\\Program Files\\Java\\jdk1.8.0_73")

#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 3.4.2 (2017-09-28) on Linux the 2017-11-09 10:52:47. See below for more information.

```
sessionInfo()
```

```
## R version 3.4.2 (2017-09-28)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 16.04.3 LTS
##
## Matrix products: default
## BLAS: /usr/lib/libblas/libblas.so.3.6.0
## LAPACK: /usr/lib/lapack/liblapack.so.3.6.0
##
```

```
## locale:
##  [1] LC_CTYPE=fr_FR.UTF-8          LC_NUMERIC=C
##  [3] LC_TIME=fr_FR.UTF-8           LC_COLLATE=fr_FR.UTF-8
##  [5] LC_MONETARY=fr_FR.UTF-8       LC_MESSAGES=fr_FR.UTF-8
##  [7] LC_PAPER=fr_FR.UTF-8          LC_NAME=fr_FR.UTF-8
##  [9] LC_ADDRESS=fr_FR.UTF-8        LC_TELEPHONE=fr_FR.UTF-8
## [11] LC_MEASUREMENT=fr_FR.UTF-8    LC_IDENTIFICATION=fr_FR.UTF-8
##
## attached base packages:
##  [1] parallel  stats4    grid      stats     graphics  grDevices utils
##  [8] datasets  methods   base
##
## other attached packages:
##  [1] vegan_2.4-4                lattice_0.20-35
##  [3] permute_0.9-4              gridExtra_2.2.1
##  [5] venneuler_1.1-0            rJava_0.9-8
##  [7] d3treeR_0.1                data.tree_0.7.0
##  [9] treemap_2.4-2              networkD3_0.4
## [11] multtest_2.32.0            adegenet_2.1.0
## [13] ade4_1.7-8                 ips_0.0-7
## [15] XML_3.98-1.9               colorspace_1.3-2
## [17] DECIPHER_2.4.0             RSQLite_2.0
## [19] Biostrings_2.44.2          XVector_0.16.0
## [21] phangorn_2.2.0             ape_4.1
## [23] edgeR_3.18.1               limma_3.32.5
## [25] mvabund_3.12.3             DESeq2_1.16.1
## [27] SummarizedExperiment_1.6.3 DelayedArray_0.2.7
## [29] matrixStats_0.52.2         Biobase_2.36.2
## [31] GenomicRanges_1.28.4       GenomeInfoDb_1.12.2
## [33] IRanges_2.10.3             S4Vectors_0.14.3
## [35] BiocGenerics_0.22.0        schoRsch_1.4
## [37] xtable_1.8-2               circlize_0.4.1
## [39] VennDiagram_1.6.17         futile.logger_1.4.3
## [41] plyr_1.8.4                 cluster_2.0.6
## [43] phyloseq_1.20.0            ggplot2_2.2.1
## [45] knitr_1.17
##
## loaded via a namespace (and not attached):
##   [1] backports_1.1.0         Hmisc_4.0-3
##   [3] fastmatch_1.1-0         igraph_1.1.2
##   [5] lazyeval_0.2.0          sp_1.2-5
##   [7] splines_3.4.2           BiocParallel_1.10.1
##   [9] gridBase_0.4-7          digest_0.6.12
##  [11] foreach_1.4.3           htmltools_0.3.6
##  [13] viridis_0.4.0           gdata_2.18.0
##  [15] magrittr_1.5            checkmate_1.8.3
##  [17] memoise_1.1.0           readr_1.1.1
##  [19] annotate_1.54.0         gmodels_2.16.2
##  [21] blob_1.1.0              dplyr_0.7.2
##  [23] RCurl_1.95-4.8          jsonlite_1.5
##  [25] genefilter_1.58.1       bindr_0.1
##  [27] brew_1.0-6              survival_2.41-3
##  [29] iterators_1.0.8         glue_1.1.1
##  [31] gtable_0.2.0            zlibbioc_1.22.0
##  [33] seqinr_3.4-5            Rook_1.1-1
##  [35] shape_1.4.3             scales_0.5.0
##  [37] futile.options_1.0.0    DBI_0.7
##  [39] Rcpp_0.12.12            viridisLite_0.2.0
##  [41] htmlTable_1.9           foreign_0.8-69
##  [43] bit_1.1-12              spdep_0.6-15
##  [45] Formula_1.2-2           tweedie_2.2.5
##  [47] htmlwidgets_0.9         DiagrammeR_0.9.1
##  [49] RColorBrewer_1.1-2      acepack_1.4.1
##  [51] pkgconfig_2.0.1         nnet_7.3-12
##  [53] deldir_0.1-14           locfit_1.5-9.1
##  [55] rlang_0.1.2             reshape2_1.4.2
##  [57] AnnotationDbi_1.38.2    visNetwork_2.0.1
##  [59] munsell_0.4.3           tools_3.4.2
##  [61] downloader_0.4          evaluate_0.10.1
##  [63] biomformat_1.4.0        stringr_1.2.0
##  [65] bit64_0.9-7             purrr_0.2.3
##  [67] bindrcpp_0.2            nlme_3.1-131
##  [69] mime_0.5                rstudioapi_0.6
##  [71] compiler_3.4.2          rgexf_0.15.3
##  [73] tibble_1.3.4            statmod_1.4.30
##  [75] geneplotter_1.54.0      stringi_1.1.5
##  [77] highr_0.6               Matrix_1.2-11
##  [79] LearnBayes_2.15         GlobalOptions_0.0.12
##  [81] data.table_1.10.4       bitops_1.0-6
##  [83] httpuv_1.3.5            R6_2.2.2
##  [85] latticeExtra_0.6-28     gridSVG_1.5-1
##  [87] codetools_0.2-15        lambda.r_1.1.9
##  [89] boot_1.3-20             MASS_7.3-47
##  [91] gtools_3.5.0            assertthat_0.2.0
##  [93] rhdf5_2.20.0            GenomeInfoDbData_0.99.0
##  [95] mgcv_1.8-22             expm_0.999-2
##  [97] hms_0.3                 influenceR_0.1.0
##  [99] quadprog_1.5-5          rpart_4.1-11
## [101] tidyr_0.7.1             coda_0.19-1
## [103] shiny_1.0.5             base64enc_0.1-3
```

1.3 Some usefull functions

The function `as.binaryOtuTable` converts a phyloseq object into a phyloseq object with binary (*i.e.* 0/1) OTU table. It allows to suppress effect due to the 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` allows to plot accumulation curves in fonction of a factor in samples data (`@sam_data` of phyloseq object).

`otu_circle` uses 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, converts 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 (using either the package DESeq2 or edgeR).

```
source(file = "functions_for_phyloseq.R")
```

2 Data

2.1 Choice of filter parameters

```
#Choose the dataset folder
data_folder <- "Uparse_min2"

#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)
```

```
# Sort taxonomy
sort_taxRDP_brut <- unlist(strsplit(unlist(strsplit(rownames(dataBiom), split = ";"))
                                    [seq(1, length(rownames(dataBiom))*2, by = 2)],
                                    split =  "_"))[seq(2, length(rownames(dataBiom))*2,
                                                       by = 2)]
taxRDP_brut <- taxRDP_brut[1:dim(taxRDP_brut)[1] %in% sort_taxRDP_brut,]

# Format taxonomy for phyloseq
taxRDP <- taxRDP_brut[match(taxa_names(dataBiom),
                            paste(taxRDP_brut[, 1], taxRDP_brut[, 2], "", sep = ";")),
                      c(5, 7, 9, 11, 13, 15, 17)]
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 = ""))

match_interm <- match(paste(funguild$OTU_ID,";", sep = ""), gsub(";size=", "_",
                                                                 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...
```

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
taxa_names(data_all) <-
  unlist(strsplit(taxa_names(data_all) ,
                  split = ";"))[seq(1, 2*length(taxa_names(data_all)), by=2)]
```

2.2.7 Caracteristics of the phyloseq data

```
data_all

## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 662 taxa and 80 samples ]
## sample_data() Sample Data:       [ 80 samples by 6 sample variables ]
## tax_table()   Taxonomy Table:    [ 662 taxa by 12 taxonomic ranks ]
## refseq()      DNAStringSet:      [ 662 reference sequences ]
```

The data are made of 8.265594×10^6 sequences representing 662 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 2×10^4 sequences, we keep 72 on the 80 samples (90%).

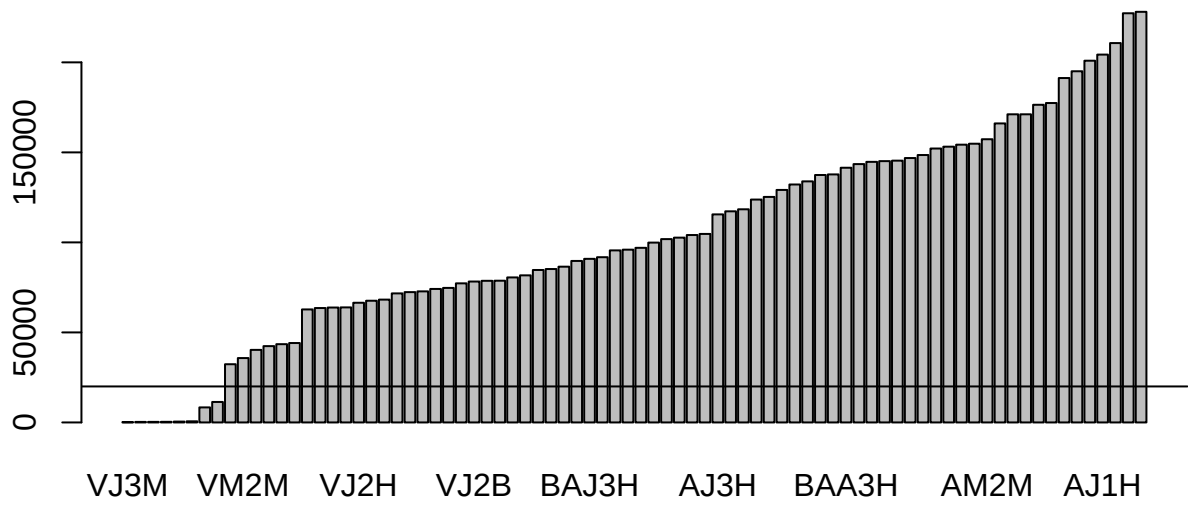


Figure 2.1: Number of sequences by sample. Horizontal line indicates the filtering parameter.

```
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)
```

2.4 Filter OTUs by number of samples

First, we can visualize the number of OTUs 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)) +
  geom_vline(xintercept= N_otu_sam_min)

## `geom_smooth()` using method = 'loess'

summary(df_nbOtu_sample$Nb.samples)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00   14.00   30.00   31.96   49.00   70.00
```

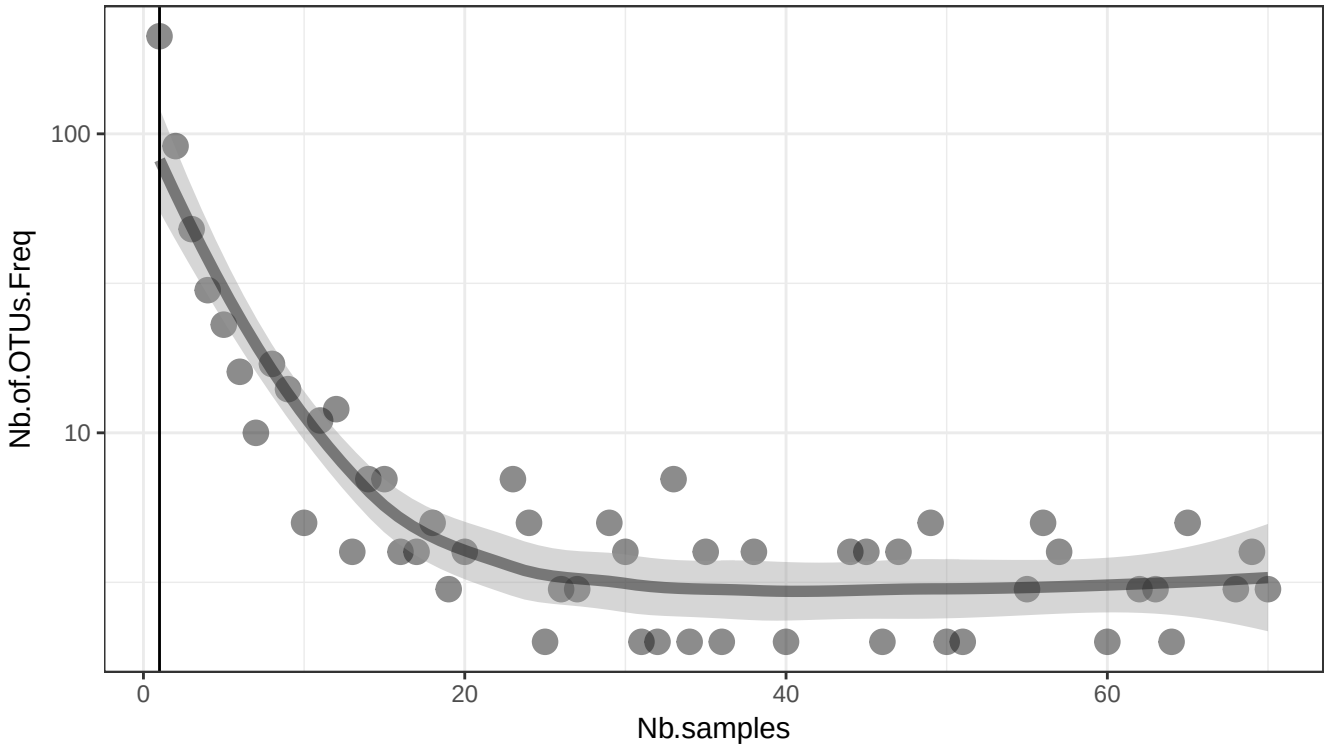


Figure 2.2: Number of OTU present in a given number of samples. Vertical bar illustrates the filtering parameter.

```
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 654 on the 654 OTUs (100%).

2.5 Filter OTUs by number of sequences

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.
##       2.0      46.0     218.5   12605.0    1360.8 2226714.0
```

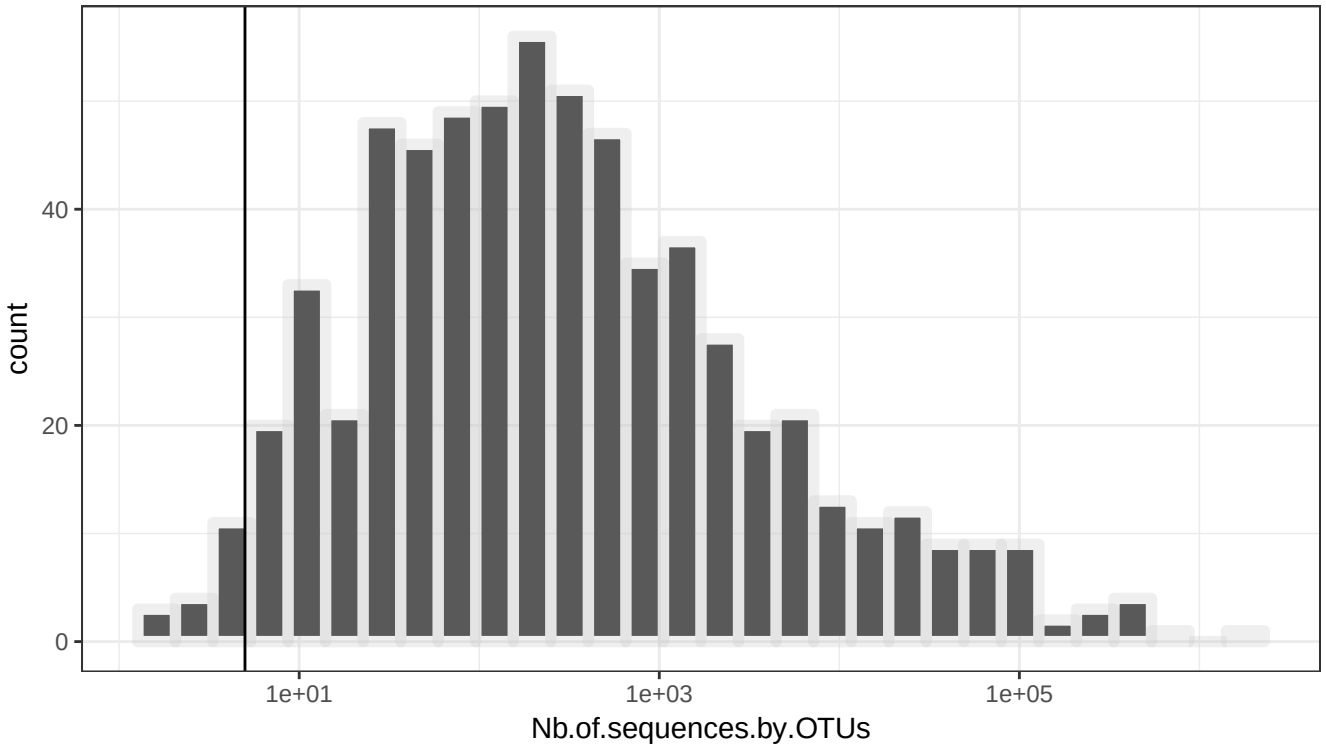


Figure 2.3: Number of sequences by OTU (log10 transformed). Horizontal bar illustrates the filtering parameter.

```
N_seq_otu_min

## [1] 5
```

If we discard OTUs with less than 1 sequences, we keep 642 on the 662 OTUs (96.98%).

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

2.6 Summary of filtration workflow

The filtered data are made of 8.243608×10^6 sequences representing 642 OTUs allocate to 72 samples.

	Nb.of.OTUs	Nb.of.samples	Nb.of.sequences
No filter	662	80	8265594.00
Nb of sequences by sample >= 20000	654	72	8243646.00
Nb of sample by OTUs >= 1	654	72	8243646.00
Nb of sequences by OTUs >= 5	642	72	8243608.00

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

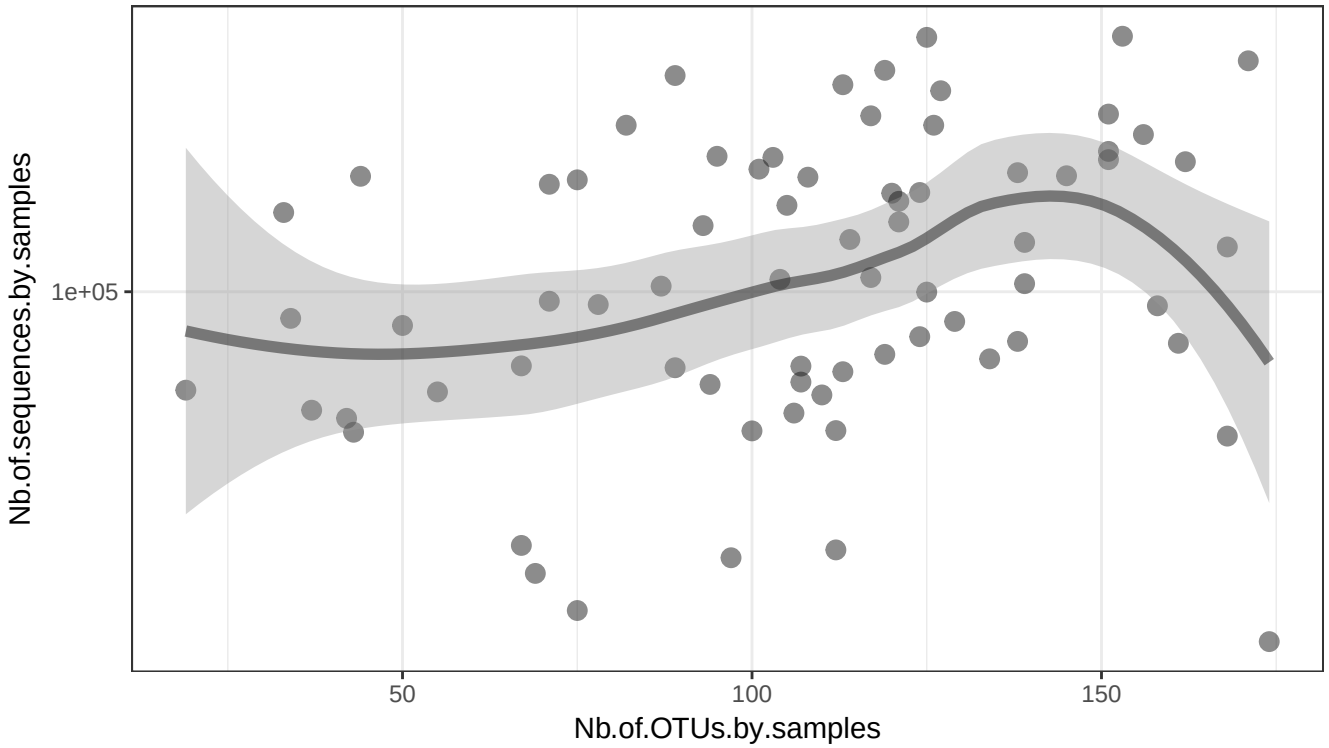


Figure 3.1: Number of OTUs by sample in fonction of the number of sequences by sample (log10 axe). The tendency is represented by the line obtained from loess (Local Polynomial Regression Fitting).

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

## `geom_smooth()` using method = 'loess'
```

```
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])
```

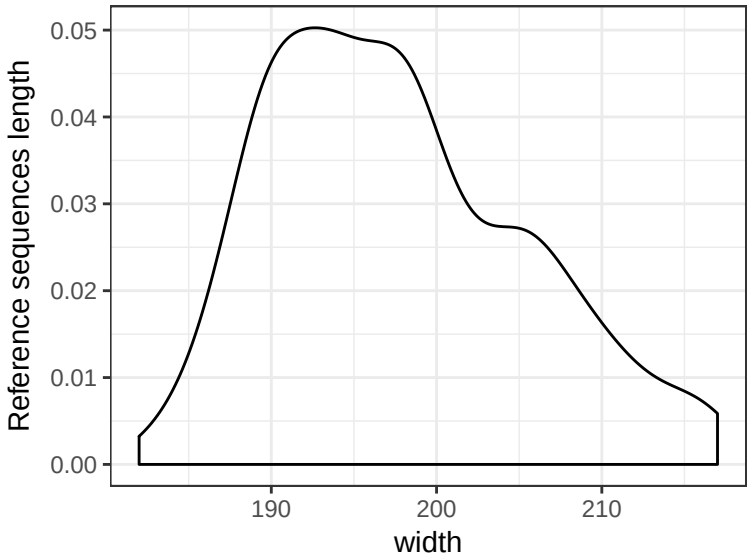


Figure 3.2: Distribution of reference sequences length.

```
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"))
```

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)
```

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)
```

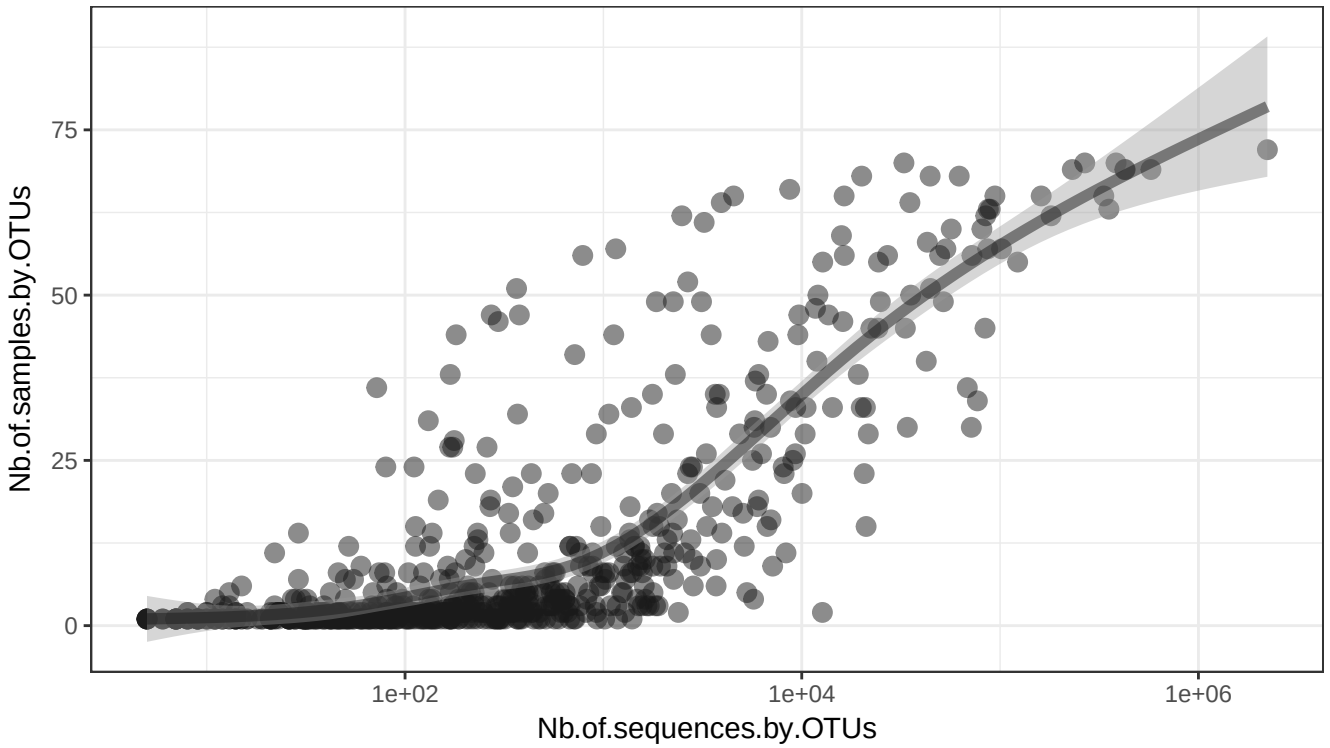


Figure 3.3: Number of sequences by OTUs (log10 axe) in fonction of the number of samples where OTUs were found. The tendency is represented by the line obtain from gam (Generalized additive models with integrated smoothness estimation).

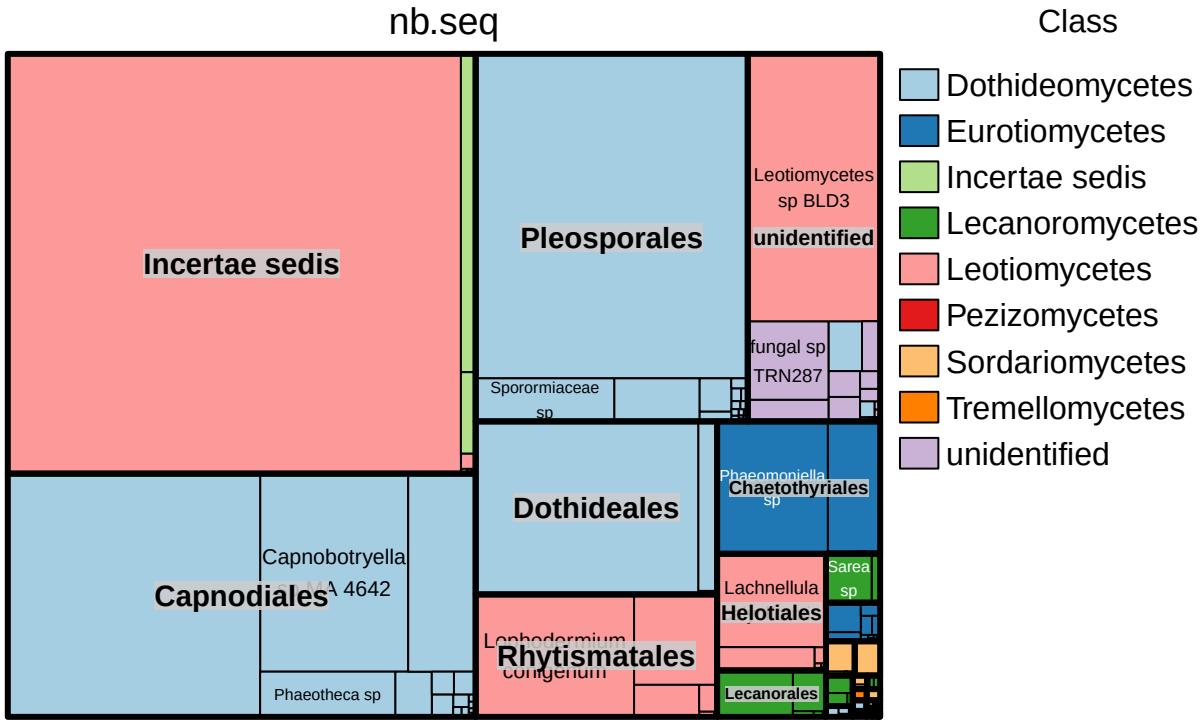


Figure 3.4: Distribution of the number of sequences in the Ascomycota taxonomy. Colors represent Class, bold lines delimit Order and thick line delimit species.

```
print(xtable(df_the30mostfrequents[, c(1:8, 12)], auto = T,
             caption = "Taxonomie of the 30 more
             abundant 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	-	-	2226714
Ascomycota	Dothideomycetes	Pleosporales				-	-	576435
Ascomycota	Leotiomycetes	unidentified	unidentified	unidentified	Leotiomycetes sp BLD3	-	-	427178
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp	-	-	425340
Ascomycota	Dothideomycetes	Capnodiales				-	-	384261
Ascomycota	Dothideomycetes	Capnodiales	Incertae sedis	Capnobotryella	Capnobotryella sp MA 4642	Saprotroph	Undefined Saprotroph	353748
Ascomycota	Dothideomycetes	Pleosporales	Incertae sedis			-	-	333382
						-	-	267278
Ascomycota	Leotiomycetes	Rhytismatales	Rhytismataceae	Lophodermium	Lophodermium conigenum	Pathotroph	Plant Pathogen	231199
						-	-	180724
Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Phaeothecoidea	Phaeothecoidea sp	Saprotroph	Undefined Saprotroph	161122
Ascomycota	Leotiomycetes	Helotiales	Hyaloscyphaceae	Lachnellula	Lachnellula calyciformis	Saprotroph	Undefined Saprotroph	122624
						-	-	101702
Ascomycota	Dothideomycetes	Pleosporales				-	-	94186
Ascomycota	Dothideomycetes	Capnodiales				-	-	89179
Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Phaeomoniella	Phaeomoniella sp	Saprotroph	Undefined Saprotroph	87249
						-	-	86659
Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Phaeomoniella	Phaeomoniella sp	Saprotroph	Undefined Saprotroph	84706
						-	-	83965
Ascomycota	Leotiomycetes	Rhytismatales	Rhytismataceae	Lophodermium	Lophodermium seditiosum	Pathotroph	Plant Pathogen	80980
Ascomycota						-	-	76823
unidentified	unidentified	unidentified	unidentified	unidentified	fungal sp TRN287	-	-	72163
Ascomycota	Dothideomycetes	Pleosporales	Sporormiaceae	unidentified	Sporormiaceae sp	-	-	71602
Ascomycota						-	-	68360
Ascomycota	Dothideomycetes	Capnodiales	Davidiellaceae	Cladosporium		-	-	62215
Ascomycota	Incertae sedis	Incertae sedis	Incertae sedis	Knufia		-	-	56720
Ascomycota	Dothideomycetes	Capnodiales				-	-	53305
Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae			-	-	51765
Ascomycota	Dothideomycetes	Capnodiales	Incertae sedis	Phaeotheca	Phaeotheca sp	-	-	49592

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

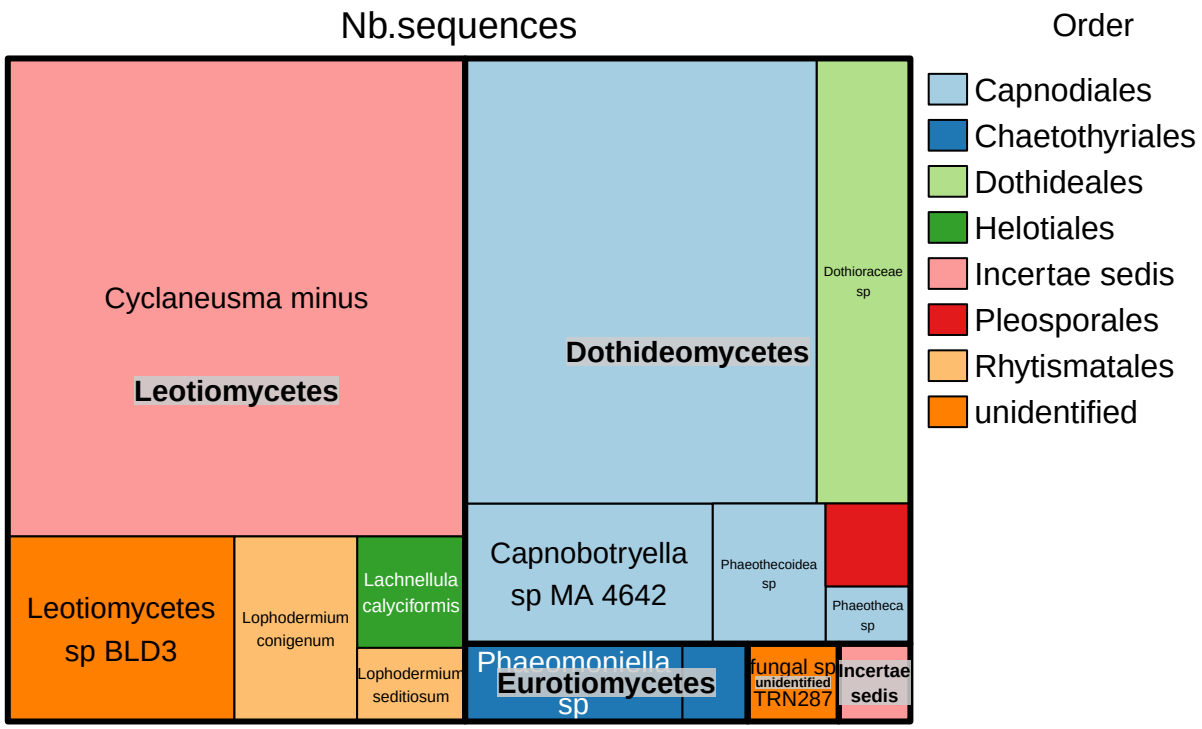


Figure 3.5: Number of sequences of the 30 most abundant OTUs (number of sequences). Colors indicate Order, bold lines delimit Class and thick lines delimit species.

```
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)
```

```
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	Dothideomycetes	Capnodiales				-	-	70
						-	-	70
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	70
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp	-	-	69
Ascomycota	Leotiomycetes	Rhytismatales	Rhytismataceae	Lophodermium	Lophodermium conigenum	Pathotroph	Plant Pathogen	69
Ascomycota	Dothideomycetes	Pleosporales				-	-	69
Ascomycota	Leotiomycetes	unidentified	unidentified	unidentified	Leotiomycetes sp BLD3	-	-	69
Ascomycota	Dothideomycetes	Pleosporales	Incertae sedis	Ochrocladosporium	Ochrocladosporium sp	Saprotroph	Undefined Saprotroph	68
Ascomycota	Dothideomycetes	Capnodiales	Davidiellaceae	Cladosporium		-	-	68
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	68
Ascomycota	Dothideomycetes	Capnodiales				-	-	66
Ascomycota	Dothideomycetes	Capnodiales				-	-	65
Ascomycota	Dothideomycetes	Pleosporales	Incertae sedis			-	-	65
Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Phaeothecoidea	Phaeothecoidea sp	Saprotroph	Undefined Saprotroph	65
Ascomycota	Dothideomycetes	Pleosporales				-	-	65
Ascomycota	Leotiomycetes	Rhytismatales	Rhytismataceae	Lophodermium		Pathotroph	Plant Pathogen	65
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp	-	-	64
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp	-	-	64
Ascomycota	Dothideomycetes	Capnodiales				-	-	63
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	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Phaeomoniella	Phaeomoniella sp	Saprotroph	Undefined Saprotroph	62
						-	-	62
Ascomycota	Leotiomycetes	Rhytismatales	Rhytismataceae	Lophodermium		Pathotroph	Plant Pathogen	62
Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Phaeomoniella	Phaeomoniella sp	Saprotroph	Undefined Saprotroph	61
Ascomycota	Incertae sedis	Incertae sedis	Incertae sedis	Knufia		-	-	60
Ascomycota	Leotiomycetes	Rhytismatales	Rhytismataceae	Lophodermium	Lophodermium seditiosum	Pathotroph	Plant Pathogen	60
Ascomycota	Dothideomycetes	unidentified	unidentified	unidentified	Dothideomycetes sp 11147	-	-	59

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

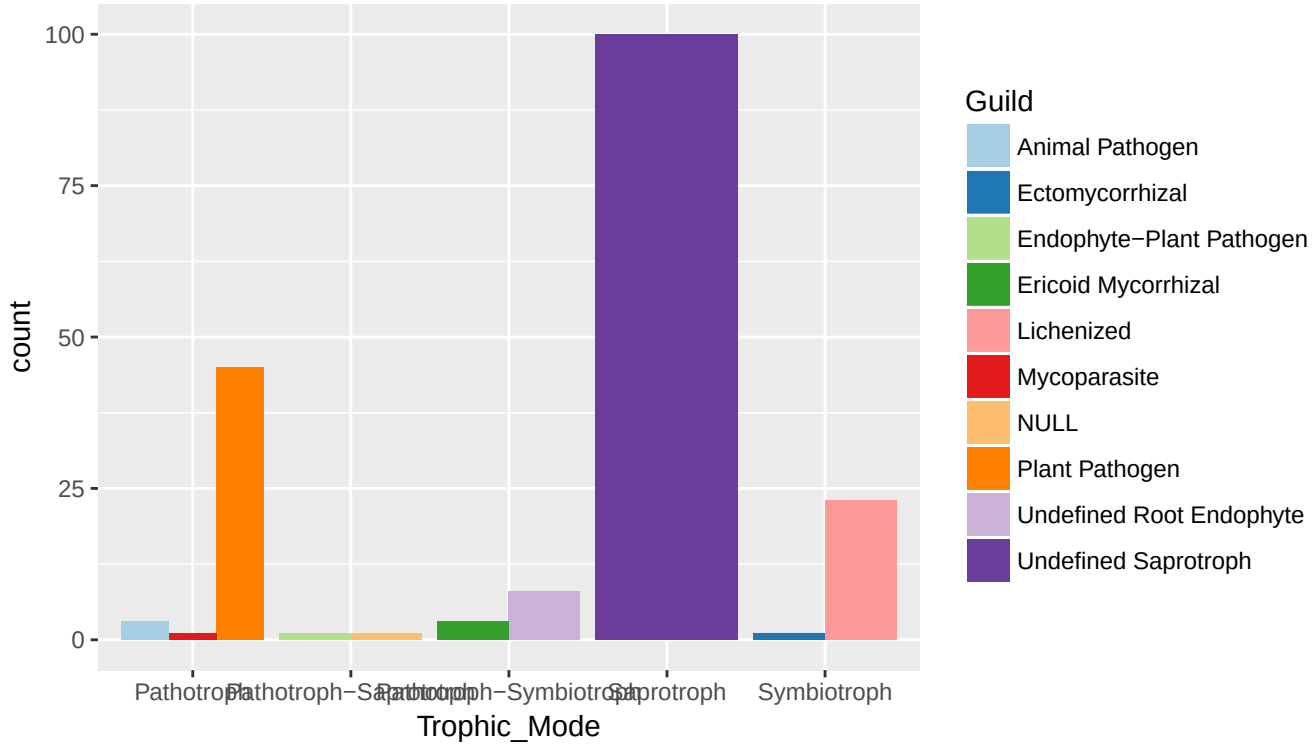


Figure 4.1: Distribution of OTUs into functional Guild.

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] 82.10002

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()
```

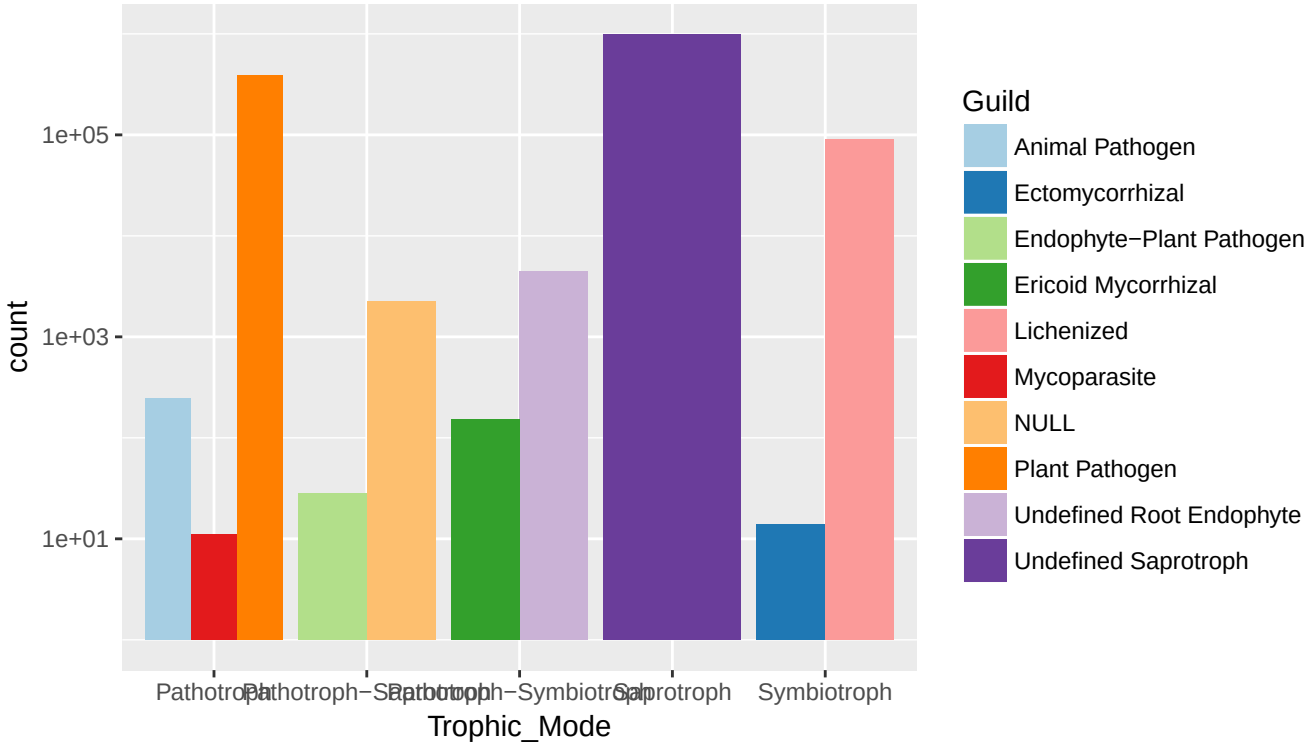


Figure 4.2: Distribution of sequences (log10 transformed) into functional Guild.

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)
```

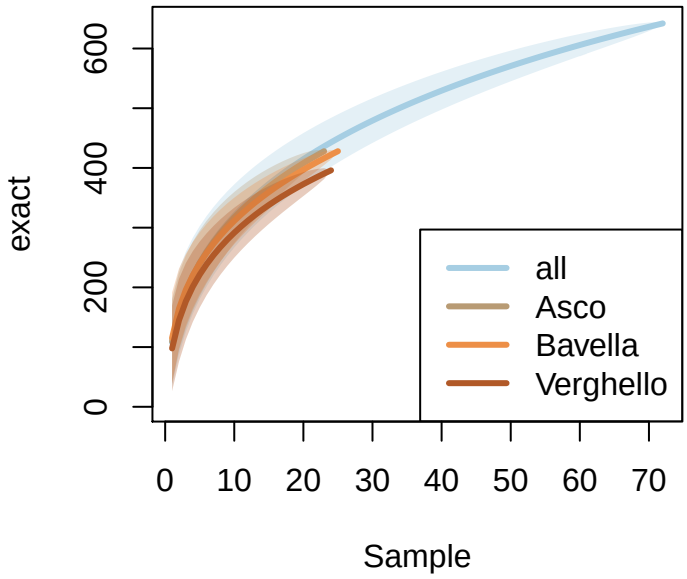


Figure 5.1: Rarefaction curves for each site. Note that if singletons were removed, these curves are biaised.

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)
```

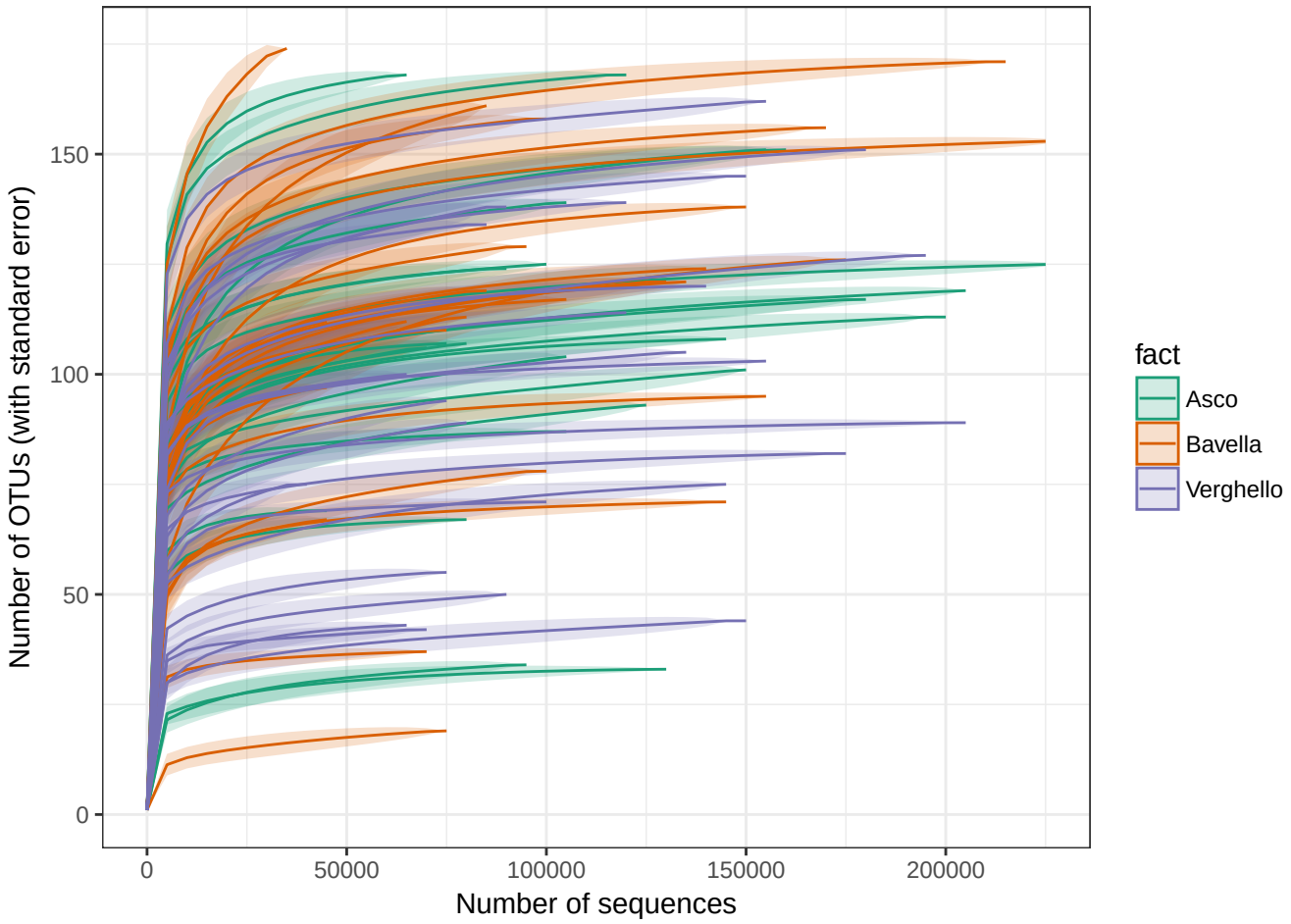


Figure 5.2: Rarefaction curves for each sample using sequences number on x-axes. Note that if singletons were removed, these curves are biaised.

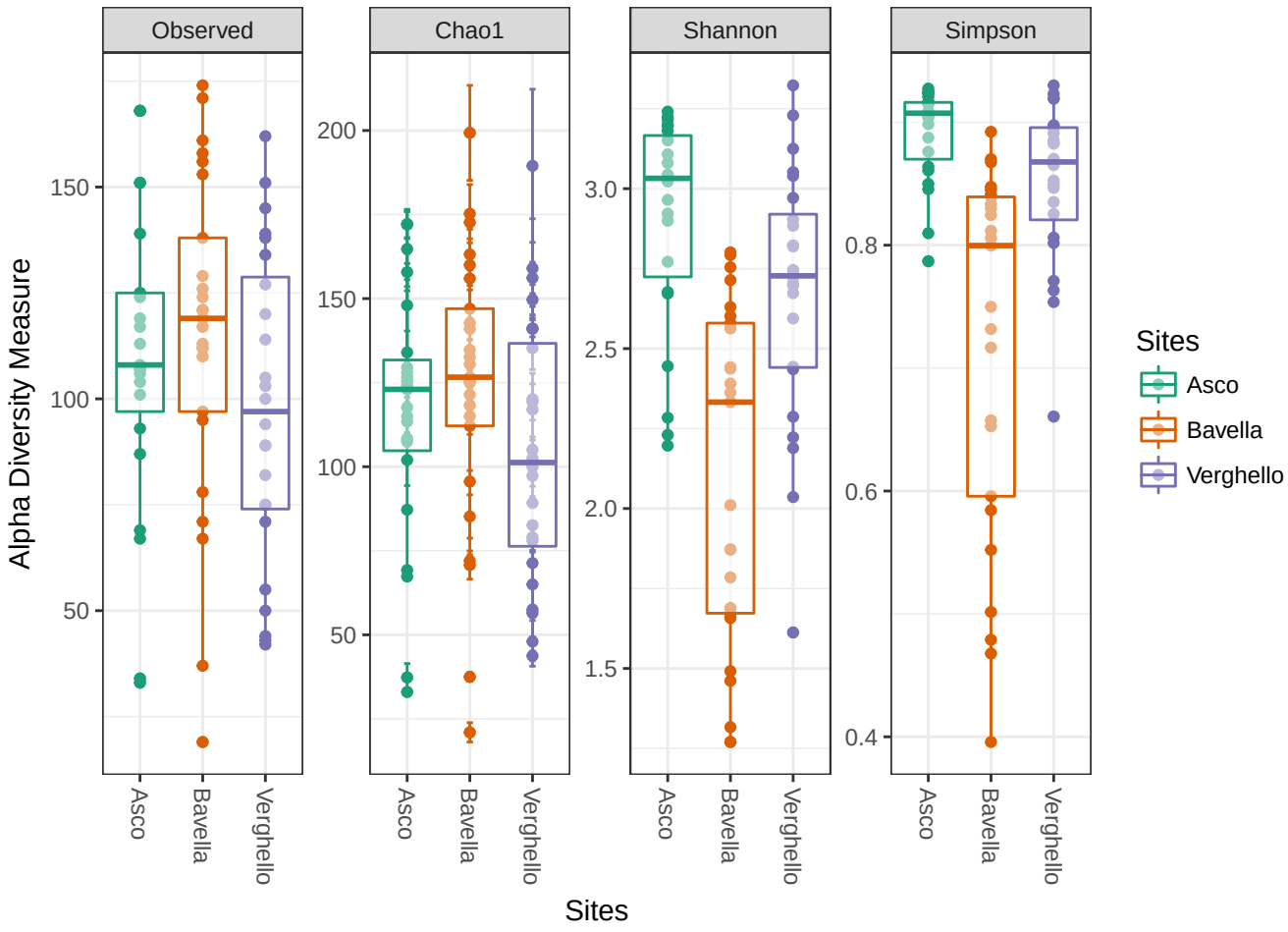


Figure 5.3: Diversity of each sites

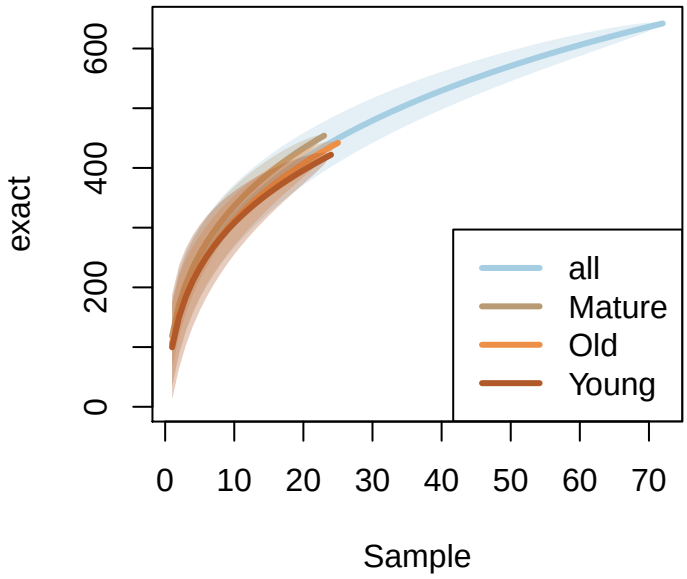


Figure 5.4: Rarefaction curves for each host age. Note that if singletons were removed, these curves are biaised.

```
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)	51.6106359	22.8040019	2.2632271	0.0270214
sqrt(readNumbers)	0.1846663	0.0584640	3.1586321	0.0024196
data.f3@sam_data$SitesBavella	8.0713418	10.1446192	0.7956279	0.4291895
data.f3@sam_data$SitesVerghello	-10.9441994	10.1525433	-1.0779761	0.2850907
data.f3@sam_data$AgeOld	-9.6814931	10.0984648	-0.9587094	0.3413130
data.f3@sam_data$AgeYoung	-21.8600818	10.3378053	-2.1145767	0.0383664
data.f3@sam_data$ElevationMiddle	13.2802087	10.2254910	1.2987356	0.1986947
data.f3@sam_data$ElevationTop	5.6168606	10.0869046	0.5568468	0.5795742

Table 4: Summary of the linear model of species richness (Hill number with 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))
```

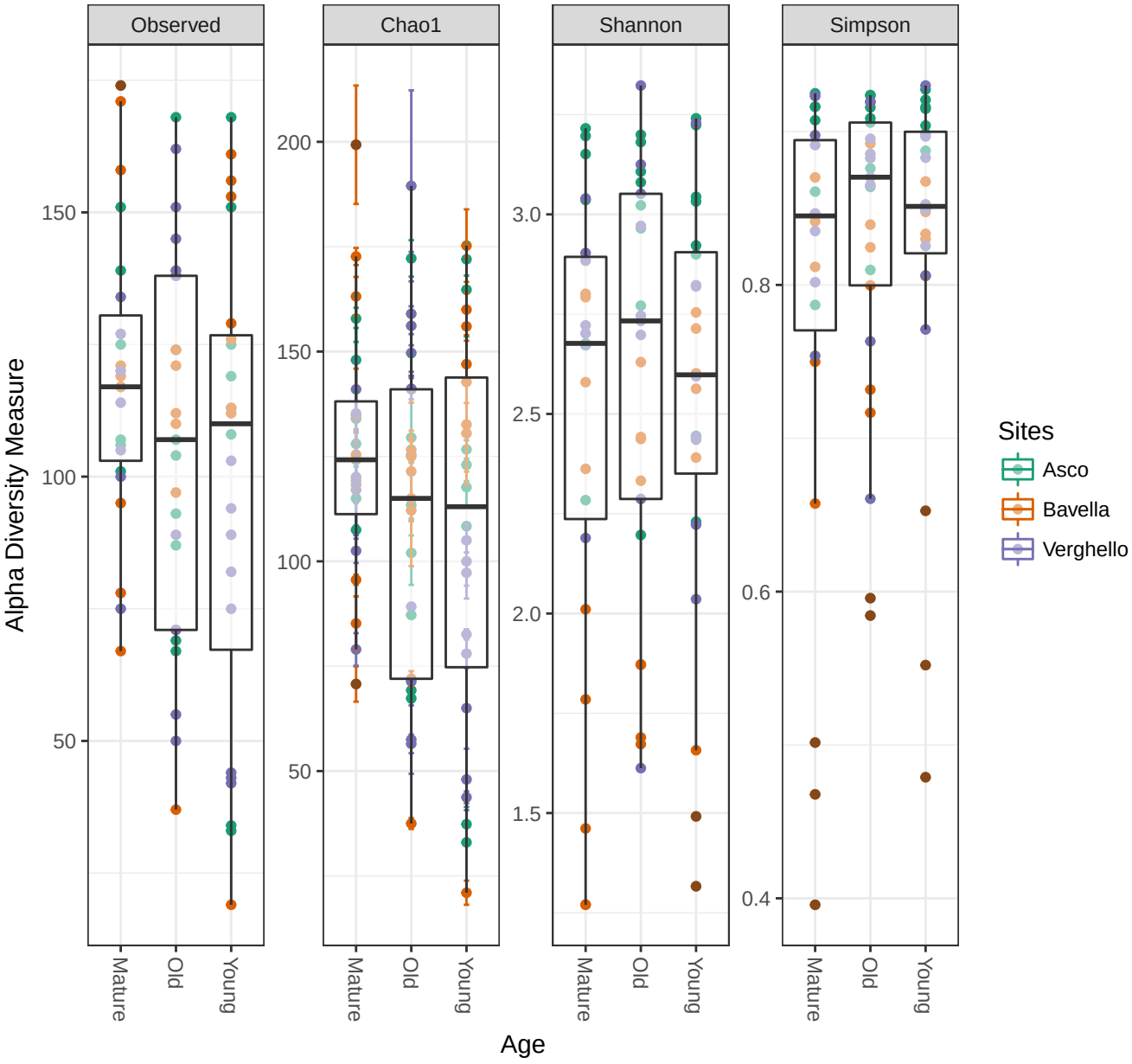


Figure 5.5: Diversity in function of tree age. Color represent sites.

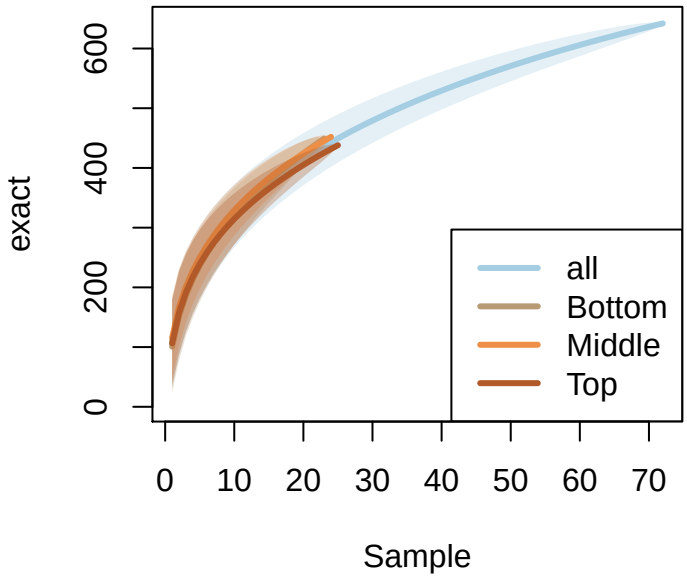


Figure 5.6: Rarefaction curves for each elevation. Notes that if singletons were removed, these curves are biaised.

```
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("")
```

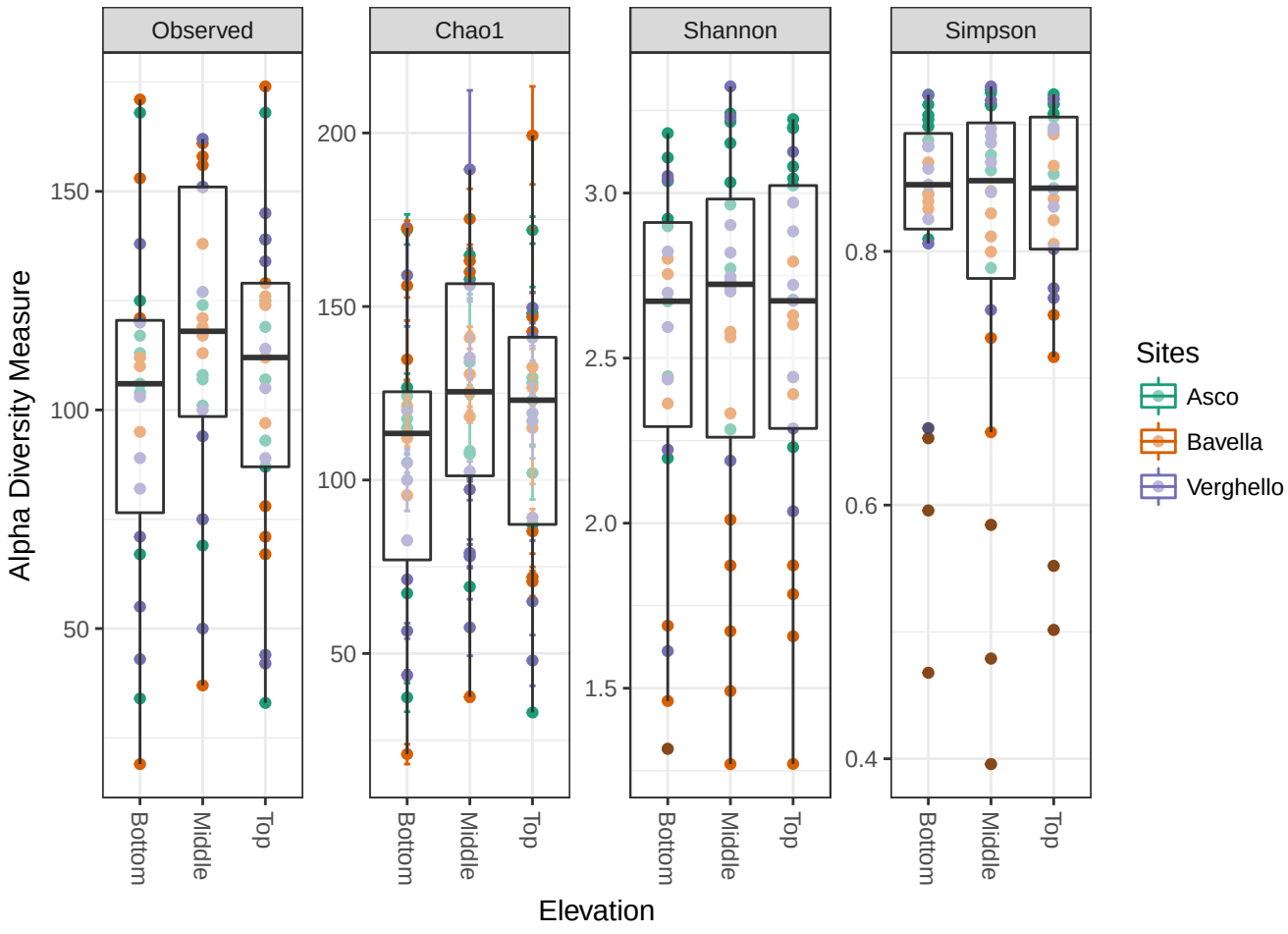


Figure 5.7: Diversity in function of elevation. Color represent sites.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	12.2399544	3.3645138	3.6379564	0.0005501
sqrt(readNumbers)	0.0167792	0.0086258	1.9452279	0.0561432
data.f3@sam_data$SitesBavella	-9.5245180	1.4967422	-6.3634995	0.0000000
data.f3@sam_data$SitesVerghello	-3.5032554	1.4979113	-2.3387603	0.0224830
data.f3@sam_data$AgeOld	0.8726120	1.4899325	0.5856722	0.5601555
data.f3@sam_data$AgeYoung	-0.5162360	1.5252449	-0.3384610	0.7361236
data.f3@sam_data$ElevationMiddle	1.9175411	1.5086740	1.2710109	0.2083251
data.f3@sam_data$ElevationTop	1.3815149	1.4882269	0.9282959	0.3567420

Table 5: Summary of the linear model of the exponential of Shannon's entropy index (Hill number with q = 1)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	7.0395839	1.7684611	3.9806268	0.0001781
sqrt(readNumbers)	0.0067710	0.0045339	1.4934165	0.1402421
data.f3@sam_data$SitesBavella	-5.3837680	0.7867200	-6.8433087	0.0000000
data.f3@sam_data$SitesVerghello	-2.2278060	0.7873345	-2.8295545	0.0062192
data.f3@sam_data$AgeOld	0.6502993	0.7831407	0.8303735	0.4094146
data.f3@sam_data$AgeYoung	0.4607476	0.8017017	0.5747120	0.5675009
data.f3@sam_data$ElevationMiddle	0.6016373	0.7929917	0.7586930	0.4508213
data.f3@sam_data$ElevationTop	0.5727127	0.7822442	0.7321406	0.4667555

Table 6: Summary of the linear model of inverse of Simpson's concentration index (Hill number with q = 2)

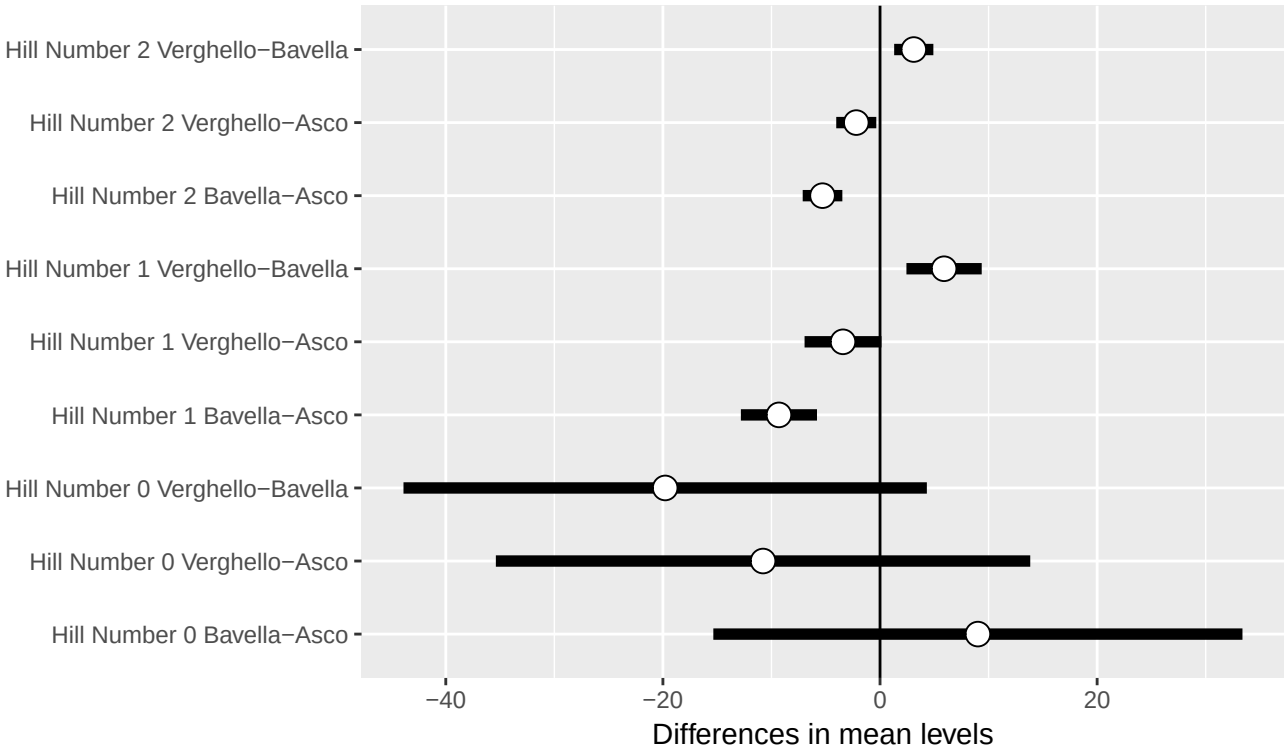


Figure 5.8: Results of the Tuckey HSD testing for differences in mean Hill numbers among pairs of modalities

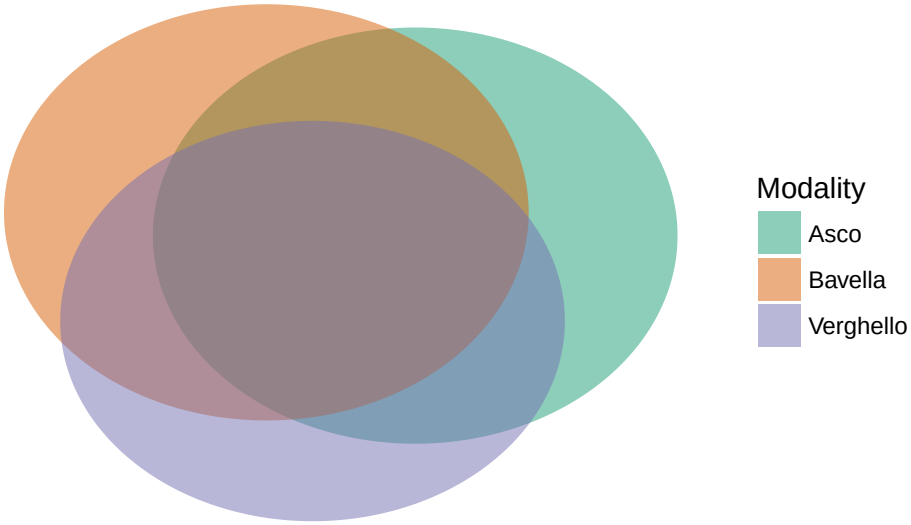


Figure 6.1: Venn diagramm of the distribution of OTUs among Sites

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 Venn diagramm for OTUs present in at least 3 samples

```
data.f3_3samp <- subset_taxa(data.f3, rowSums(data.f3@otu_table>0)>2)
venn_phyloseq(data.f3_3samp, "Sites", printValues = F)
```

```
venn_phyloseq(data.f3_3samp, "Age", printValues = F)
```

```
venn_phyloseq(data.f3_3samp, "Elevation", printValues = F)
```

6.3 Ordination

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

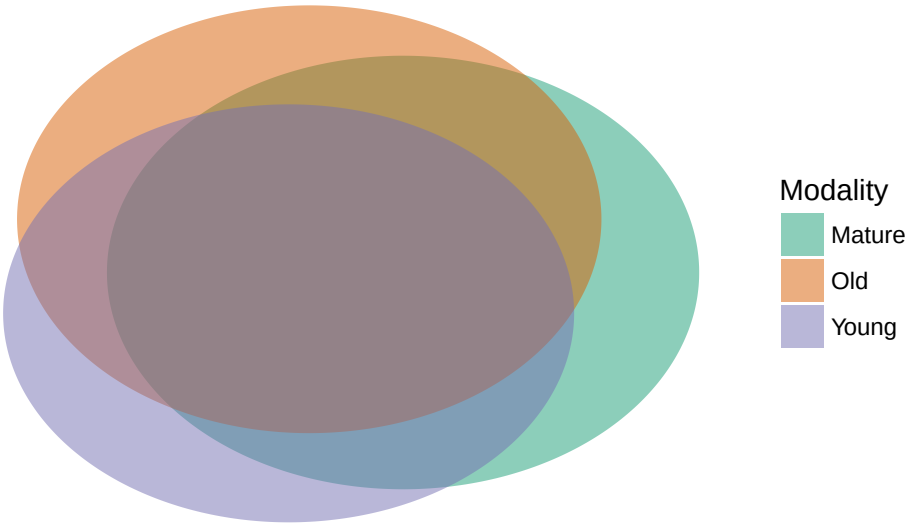


Figure 6.2: Venn diagramm of the distribution of OTUs among host age

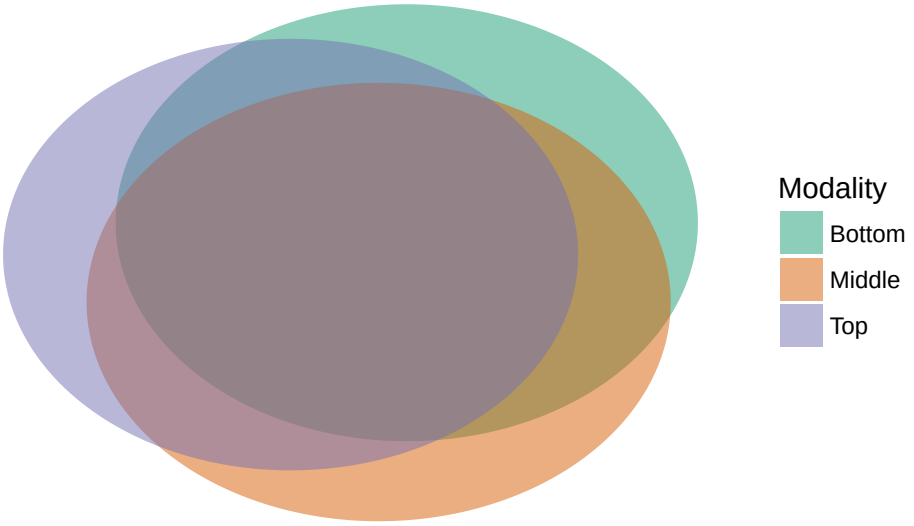


Figure 6.3: Venn diagramm of the distribution of OTUs among elevation of samples

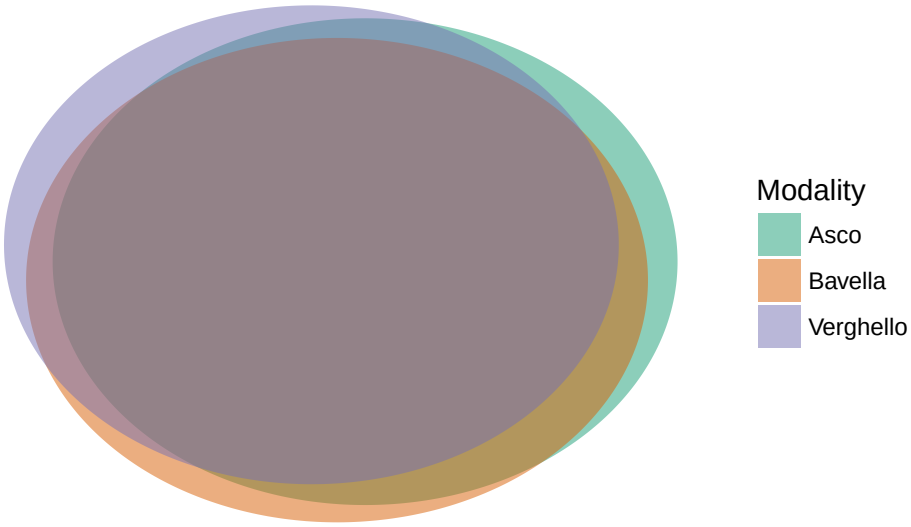


Figure 6.4: Venn diagramm of the distribution of OTUs among Sites

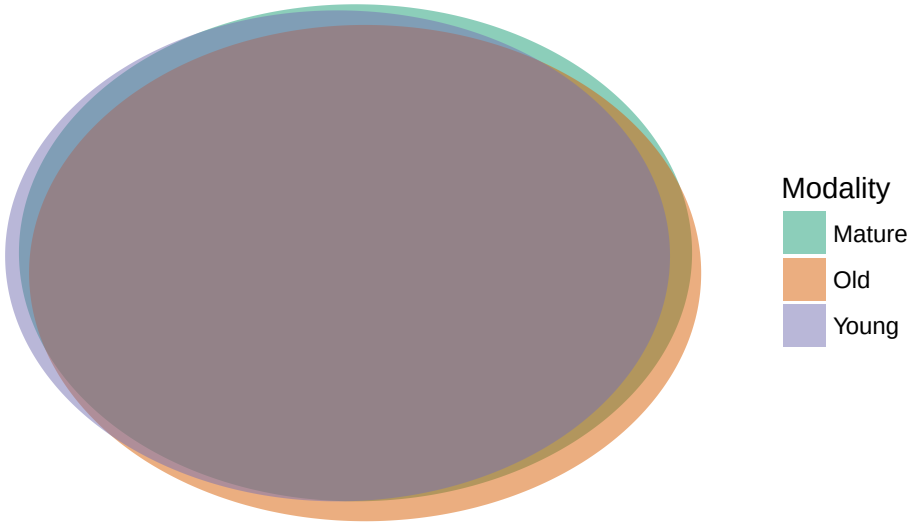


Figure 6.5: Venn diagramm of the distribution of OTUs among host age

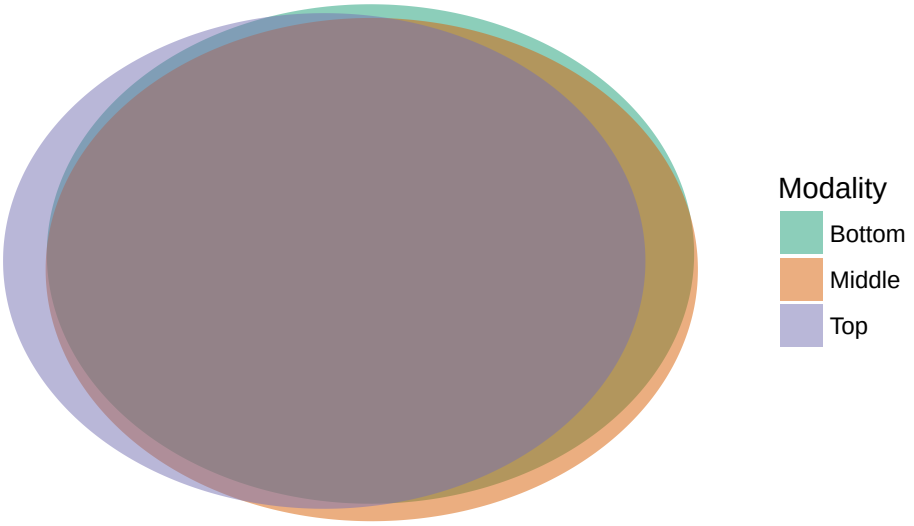


Figure 6.6: Venn diagramm of the distribution of OTUs among elevation of samples whitin the tree

```
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 0.2351585
## Run 1 stress 0.2475769
## Run 2 stress 0.2489581
## Run 3 stress 0.2434148
## Run 4 stress 0.2394274
## Run 5 stress 0.238669
## Run 6 stress 0.2422414
## Run 7 stress 0.2425637
## Run 8 stress 0.2444472
## Run 9 stress 0.2406118
## Run 10 stress 0.2419227
## Run 11 stress 0.2381512
## Run 12 stress 0.2442435
## Run 13 stress 0.2446742
## Run 14 stress 0.2387914
```

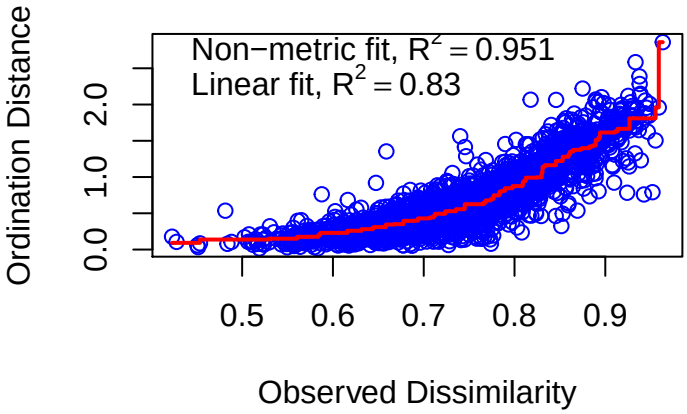


Figure 6.7: Stress plot of the NMDS

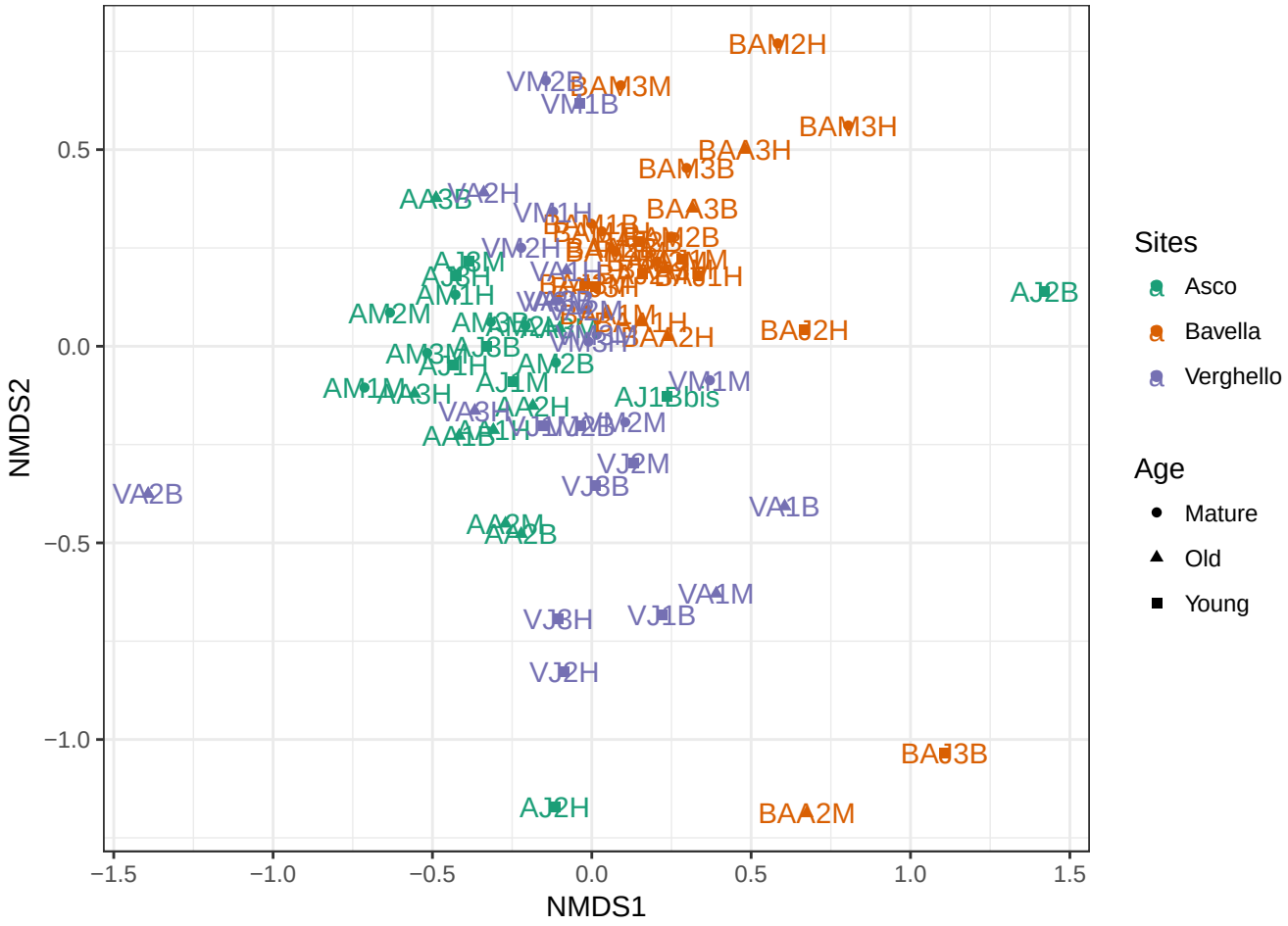


Figure 6.8: NMDS of OTU table. Colors represent sites and shape the age of tree.

```
## Run 15 stress 0.2464117
## Run 16 stress 0.2427213
## Run 17 stress 0.2447614
## Run 18 stress 0.2449986
## Run 19 stress 0.234431
## ... New best solution
## ... Procrustes: rmse 0.03818597  max resid 0.2230861
## Run 20 stress 0.2353794
## *** No convergence -- monoMDS stopping criteria:
##      1: no. of iterations >= maxit
##     19: 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)
```

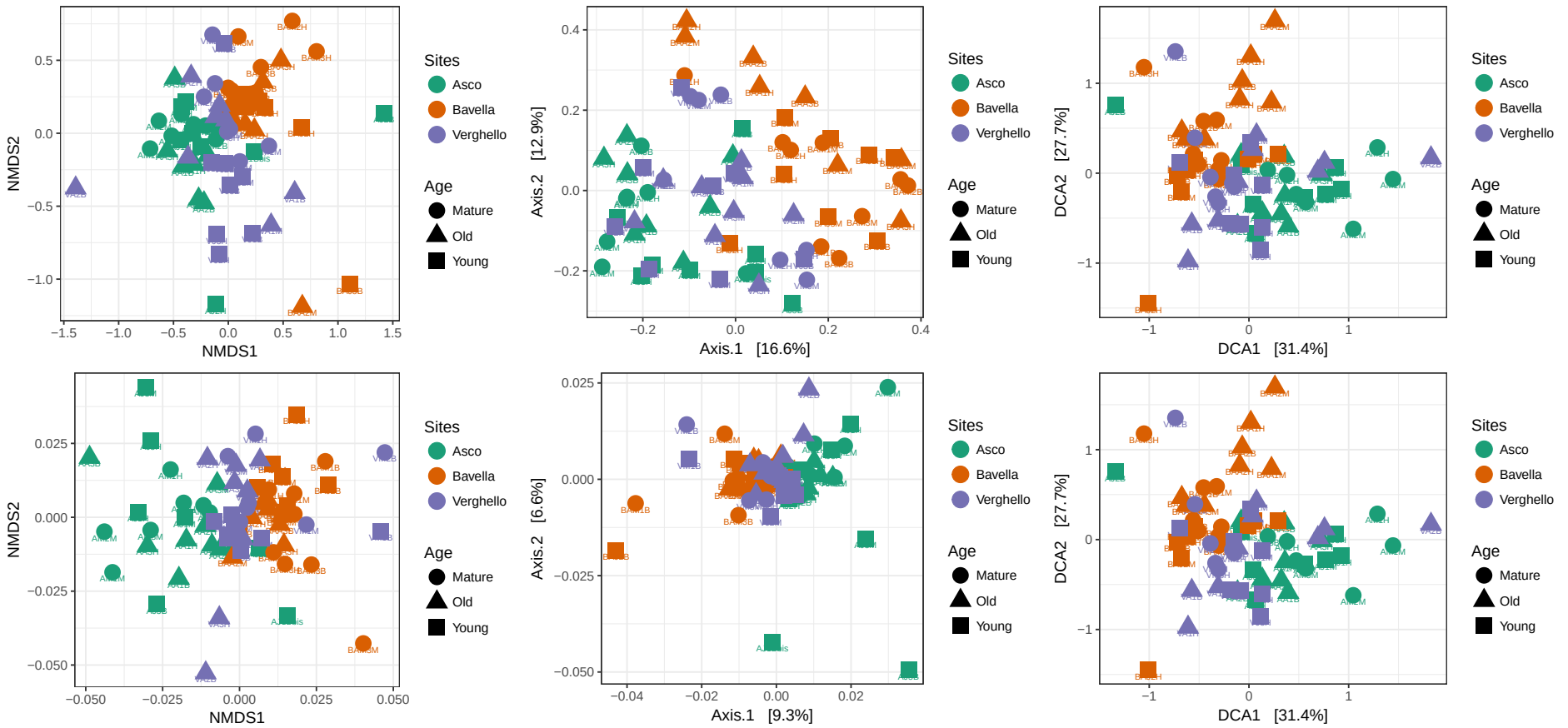


Figure 6.9: Comparison of different distances (bray (up) and gower (bottom)) and ordination methods (NMDS (left), PCoA (center) and DCA (right)).

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

6.4 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 99 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)
```

```
data.f3_without_C_minus <- subset_taxa(data.f3, taxa_names(data.f3)!="OTU_1")
res.ado_without_C_minus <- adonis(t(data.f3_without_C_minus@otu_table) ~
                            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	1.99	1.00	5.78	0.14	0.0001
Age	2	0.58	0.29	1.69	0.04	0.0165
Elevation	2	0.49	0.25	1.43	0.03	0.0682
Sites:Age	4	1.34	0.33	1.94	0.09	0.0007
Sites:Elevation	4	0.66	0.16	0.95	0.04	0.5715
Age:Elevation	4	0.64	0.16	0.93	0.04	0.6294
Sites:Age:Elevation	8	1.20	0.15	0.87	0.08	0.8209
Residuals	45	7.76	0.17		0.53	
Total	71	14.65			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")
```

```
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)).")
```

6.5 Permanova on sites, host ages and individual trees

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Sites	2	1.94	0.97	5.96	0.14	0.0001
Age	2	0.56	0.28	1.71	0.04	0.0158
Elevation	2	0.48	0.24	1.47	0.03	0.0676
Sites:Age	4	1.29	0.32	1.98	0.09	0.0003
Sites:Elevation	4	0.62	0.15	0.95	0.04	0.5697
Age:Elevation	4	0.60	0.15	0.92	0.04	0.6372
Sites:Age:Elevation	8	1.10	0.14	0.84	0.08	0.8632
Residuals	45	7.33	0.16		0.53	
Total	71	13.90			1.00	

Table 8: 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	0.81	0.40	3.47	0.09	0.0001
Age	2	0.37	0.18	1.57	0.04	0.0158
Elevation	2	0.24	0.12	1.01	0.03	0.4241
Sites:Age	4	0.67	0.17	1.43	0.07	0.0101
Sites:Elevation	4	0.43	0.11	0.93	0.05	0.6709
Age:Elevation	4	0.51	0.13	1.09	0.06	0.2641
Sites:Age:Elevation	8	0.84	0.10	0.90	0.09	0.8018
Residuals	45	5.24	0.12		0.58	
Total	71	9.10			1.00	

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

```
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).")
```

```
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")
```

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

6.6 Differences in abundances and OTUs number by Order.

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Sites	2	1.99	1.00	6.88	0.14	0.0001
Age	2	0.58	0.29	2.01	0.04	0.0019
Sites:Age	4	1.33	0.33	2.30	0.09	0.0001
Sites:Age:IndividualTree	18	4.23	0.23	1.62	0.29	0.0001
Residuals	45	6.52	0.14		0.44	
Total	71	14.65			1.00	

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

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Sites	2	1.94	0.97	7.18	0.14	0.0001
Age	2	0.56	0.28	2.07	0.04	0.0030
Sites:Age	4	1.29	0.32	2.39	0.09	0.0002
Sites:Age:IndividualTree	18	4.04	0.22	1.66	0.29	0.0001
Residuals	45	6.08	0.14		0.44	
Total	71	13.90			1.00	

Table 11: 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	0.81	0.40	3.83	0.09	0.0001
Age	2	0.37	0.18	1.74	0.04	0.0048
Sites:Age	4	0.68	0.17	1.60	0.07	0.0014
Sites:Age:IndividualTree	18	2.49	0.14	1.31	0.27	0.0016
Residuals	45	4.75	0.11		0.52	
Total	71	9.10			1.00	

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

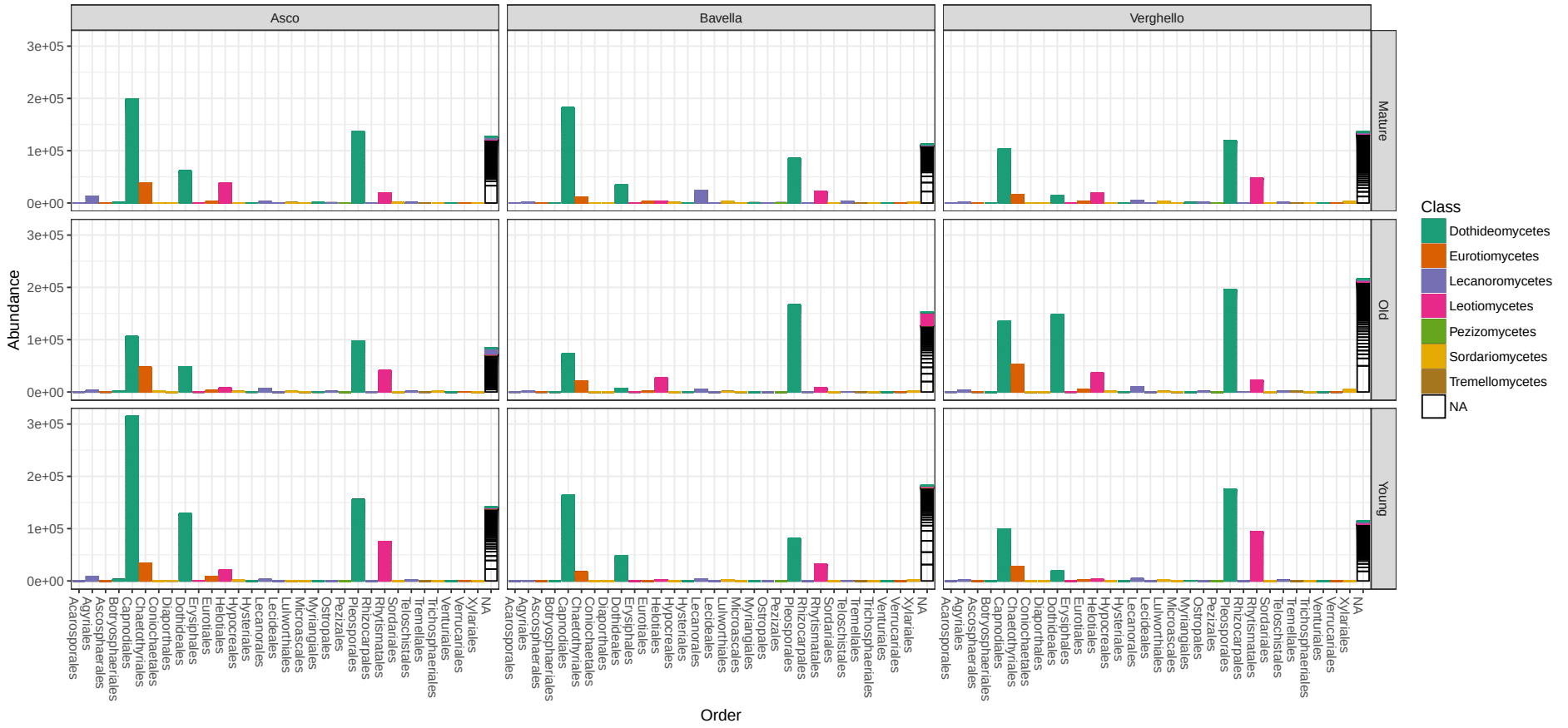


Figure 6.10: Taxonomic distribution of sequences in the different site * age combinaison.

```
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")
```

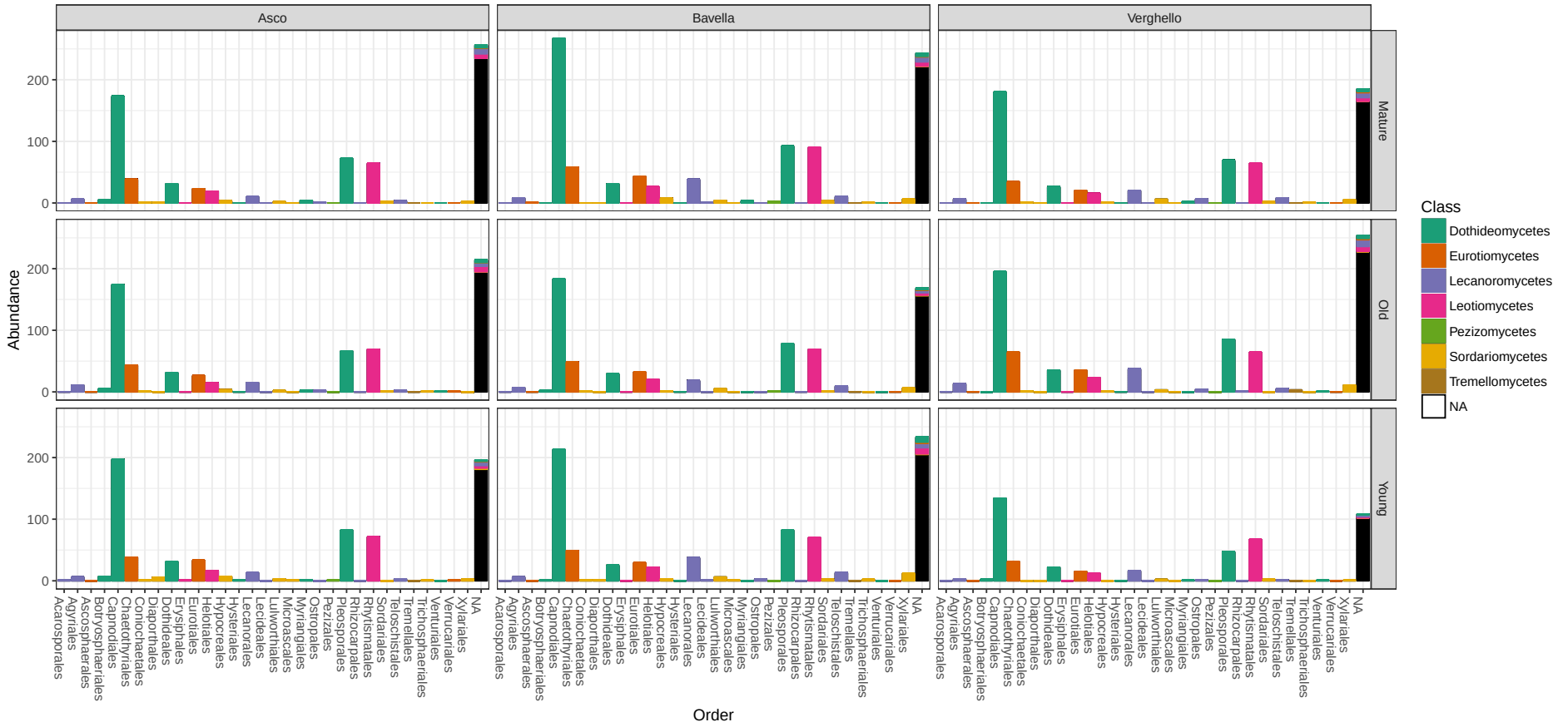


Figure 6.11: Taxonomic distribution of OTUs in the different site * age combinaison.

```
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")
```

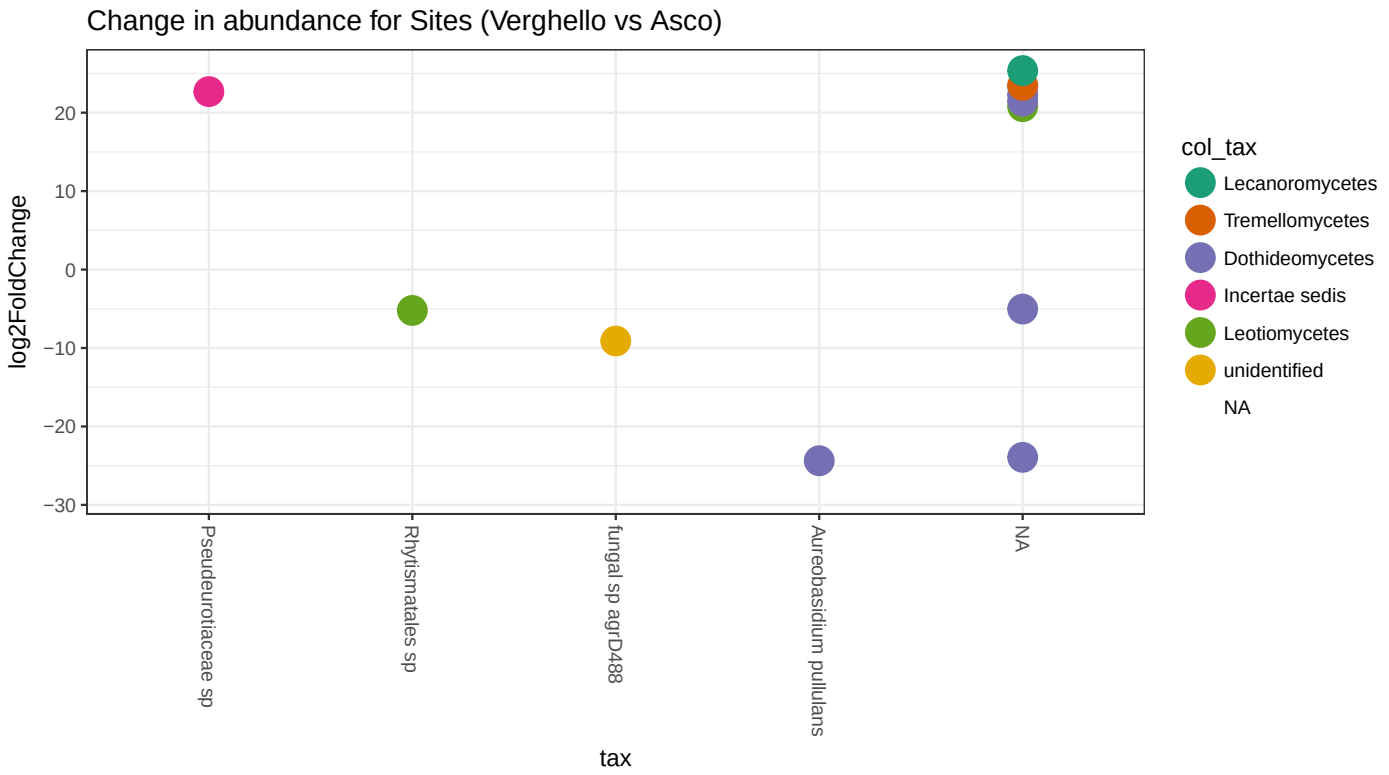


Figure 6.12: OTUs significantly different in terms of abundances between Verghello (positive values) and Asco (negative values)

6.7 Differences in abundances for each OTUs

6.7.1 Pairwise comparison of the OTUs composition by sites

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

## [1] '1.16.1'

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
```

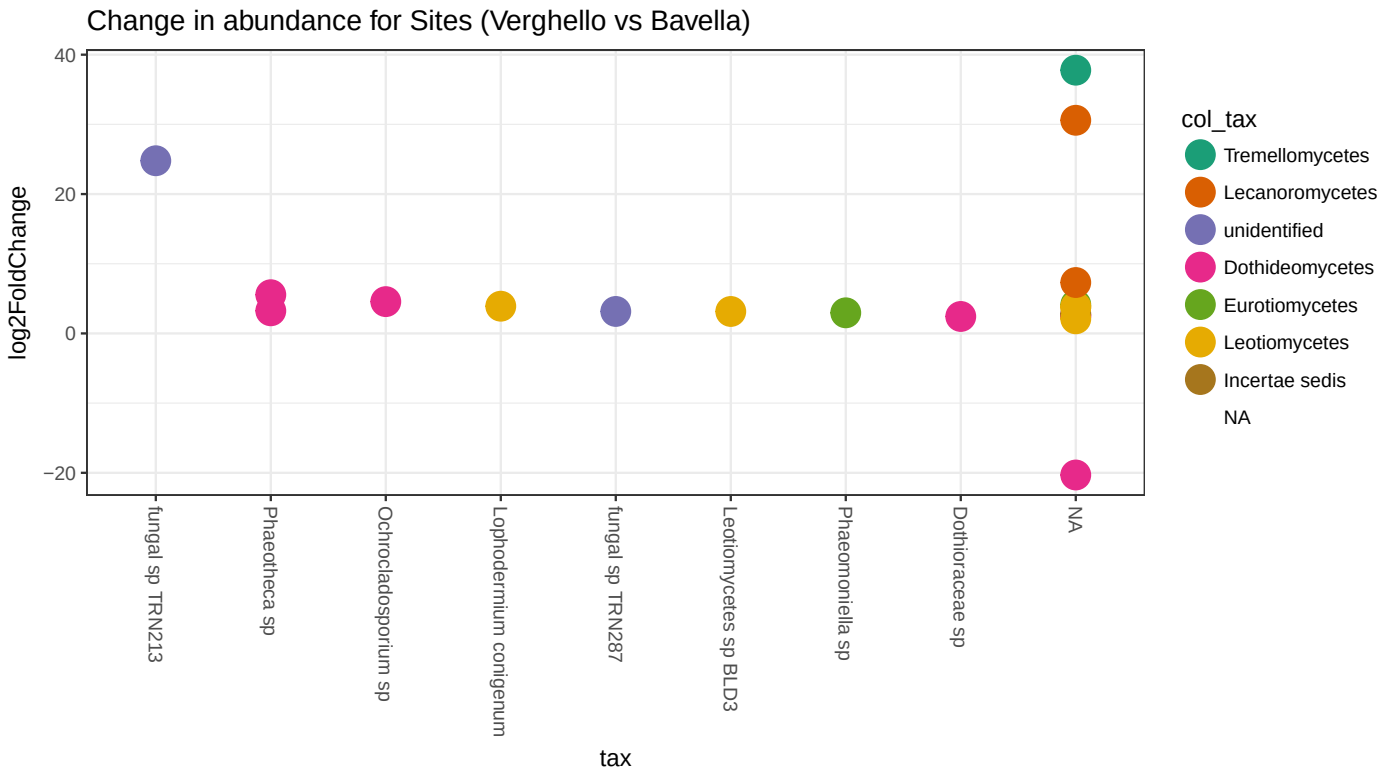


Figure 6.13: OTUs significantly different in terms of abundances between Verghello (positive values) and Bavella (negative values)

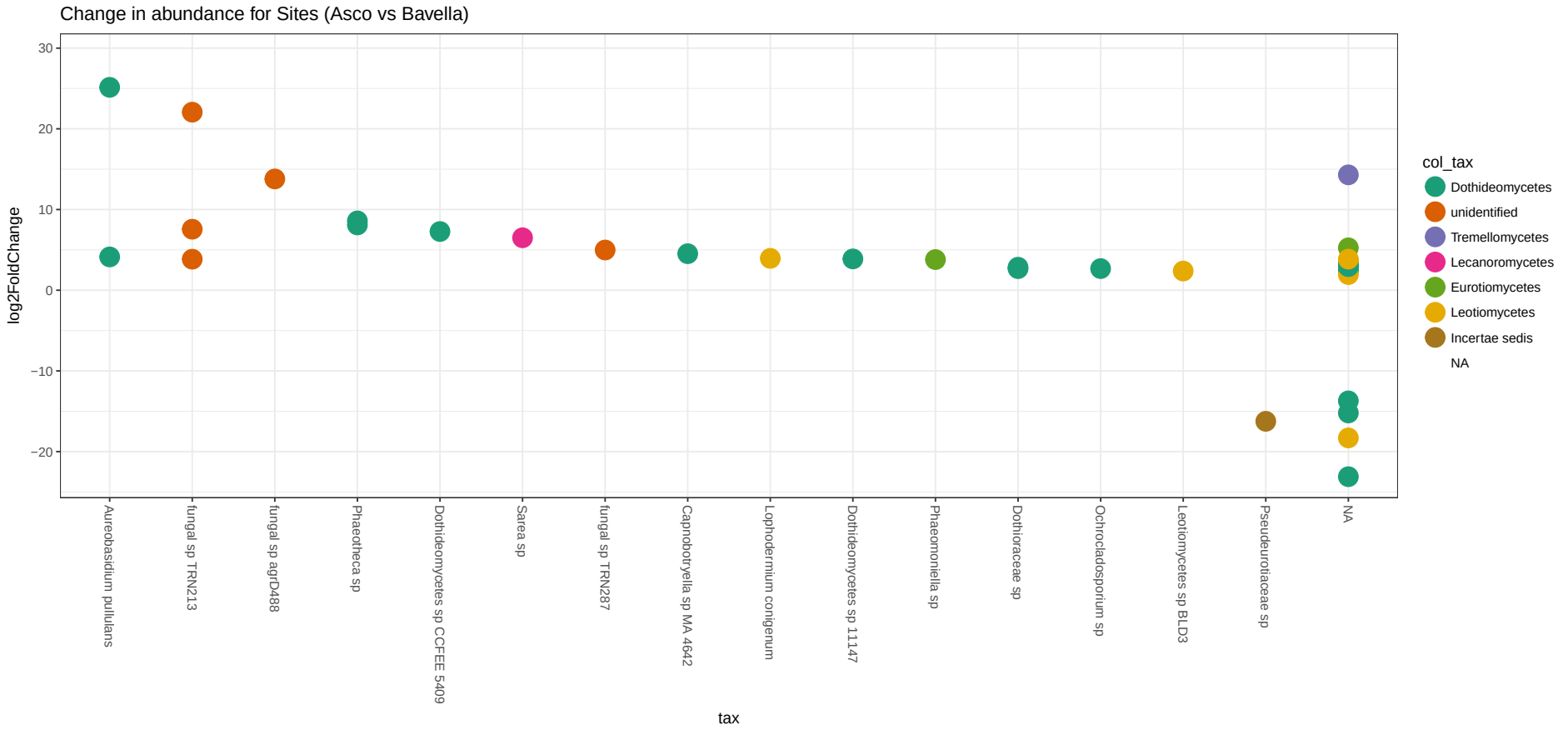


Figure 6.14: OTUs significantly different in terms of abundances between Asco (positive values) and Bavella (negative values)

```
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
```

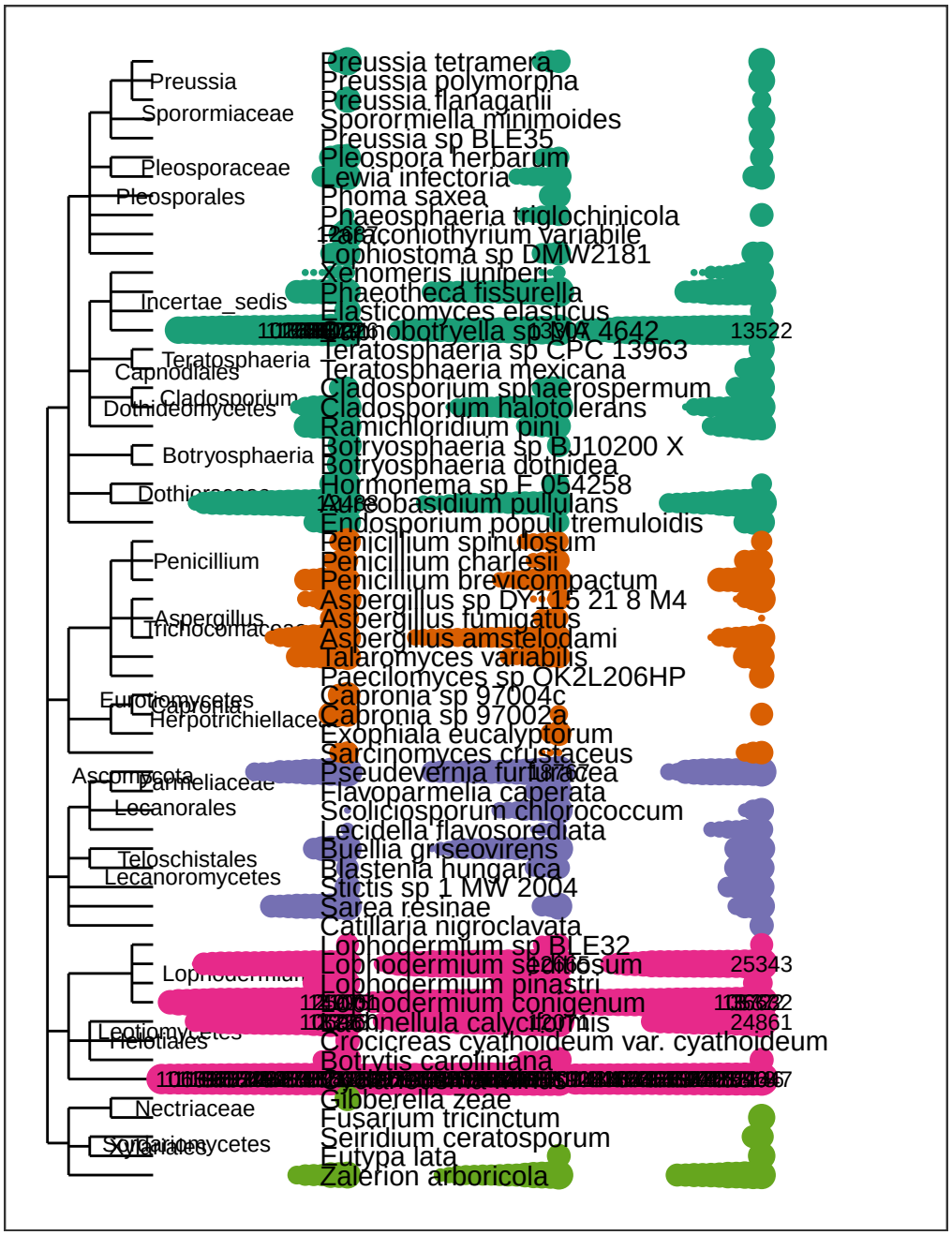
6.7.2 Pairwise comparison of Order composition by sites

```
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")
```

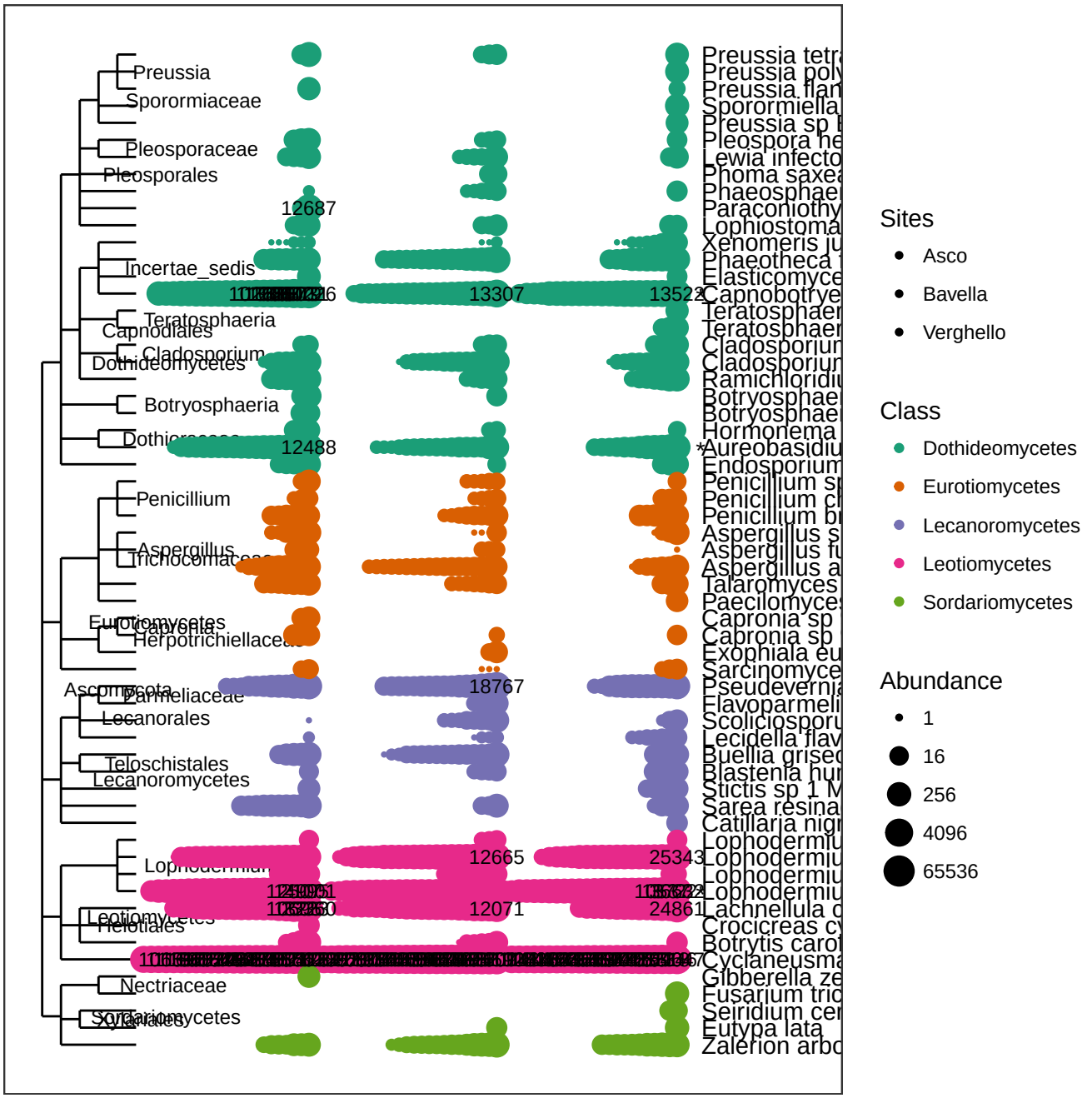
6.8 Distribution of OTUs abundance in the fungal phylogeny

```
library("cluster")
library("phytools")

## Loading required package:  maps
##
## Attaching package:  'maps'
## The following object is masked from 'package:plyr':
##
##    ozone
## The following object is masked from 'package:cluster':
##
##    votes.repub

data.f3_interm <- data.f3
data.f3_interm@otu_table <- otu_table(apply(data.f3@otu_table, 2, function(x) tapply(x, as.factor(da
data.f3_interm@tax_table <- tax_table(apply(data.f3@tax_table, 2, function(x) tapply(x, as.factor(da
data.f3_interm@refseq <- NULL

data.f3_interm <- subset_taxa(data.f3_interm, !grepl("uncultured", data.f3_interm@tax_table[,"Specie
data.f3_interm <- subset_taxa(data.f3_interm, !grepl("sp$", data.f3_interm@tax_table[,"Species"]))
data.f3_interm <- subset_taxa(data.f3_interm, !grepl("unidentified", data.f3_interm@tax_table[,"Fami
data.f3_interm <- subset_taxa(data.f3_interm, !grepl("unidentified", data.f3_interm@tax_table[,"Orde
data.f3_interm <- subset_taxa(data.f3_interm, !grepl("unidentified", data.f3_interm@tax_table[,"Clas
data.f3_interm <- subset_taxa(data.f3_interm, !grepl("Myxotrichaceae", data.f3_interm@tax_table[,"Sp
data.f3_interm <- subset_taxa(data.f3_interm, rowSums(data.f3_interm@otu_table)>100)


tree_tax_interm <- as.data.frame(unclass(data.f3_interm@tax_table))
tree_tax_interm$OTUs <- rownames(tree_tax_interm)

tree_tax_interm <- as.data.frame(replace(as.matrix(tree_tax_interm), which(is.na(tree_tax_interm)),

data.f3_interm@tax_table <- tax_table(as.matrix(tree_tax_interm))

tree_tax_interm$pathString <- paste("Fungi",
                            tree_tax_interm$Phylum,
                            tree_tax_interm$Class,
                            tree_tax_interm$Order,
```

```
                            tree_tax_interm$Family,
                            tree_tax_interm$Genus,
                            tree_tax_interm$OTUs,
                            sep = "/")

write(ToNewick(as.Node(tree_tax_interm, na.rm = TRUE)), file="tree.txt")
tree <- phytools::read.newick(file="tree.txt")
tree <- ape::collapse.singles(tree)

data.f3_interm@phy_tree <- tree
taxa_names(data.f3_interm@phy_tree) <- gsub("_", " ", taxa_names(data.f3_interm@phy_tree))
taxa_names(data.f3_interm@otu_table) <- gsub("_", " ", taxa_names(data.f3_interm@otu_table))
taxa_names(data.f3_interm@tax_table) <- gsub("_", " ", taxa_names(data.f3_interm@tax_table))
taxa_names(data.f3_interm@phy_tree) <- gsub(",", "", taxa_names(data.f3_interm@phy_tree))
taxa_names(data.f3_interm@otu_table) <- gsub(",", "", taxa_names(data.f3_interm@otu_table))
taxa_names(data.f3_interm@tax_table) <- gsub(",", "", taxa_names(data.f3_interm@tax_table))

ptree <- plot_tree(data.f3_interm, color = "Class", shape = "Sites", ladderize = "left", justify = '

cond <- gsub(",", "", rownames(data.f3_interm@otu_table)[rowSums(data.f3_interm@otu_table) >= 1])
df_cond <- as.data.frame(ptree$data)[ptree$data$OTU %in% cond,]
df_cond$Species <- data.f3_interm@tax_table[taxa_names(data.f3_interm) %in% cond,"OTUs"]


cond_Deseq <- levels(df$Species)
df_cond_Deseq <- as.data.frame(ptree$data)[ptree$data$OTU %in% cond_Deseq,]
df_cond_Deseq$Species <- data.f3_interm@tax_table[gsub("_", " ", taxa_names(data.f3_interm)) %in% co

ptree + geom_text(data = df_cond, aes(x = 215, y = y, label = OTU), hjust = "left") + scale_shape_ma

sum(!is.na(match(gsub("_", " ", data.f3@tax_table[,"Species"]), gsub("_", " ", tree$tip.label))))

## [1] 82

sum(rowSums(data.f3@otu_table)[gsub("_", " ", data.f3@tax_table[,"Species"]) %in% gsub("_", " ", tre

## [1] 39.78694
```




```
ptree + geom_text(data = df_cond, aes(x = 585, y = y, label = OTU), hjust = "left") + scale_shape_ma
ggsave("phylo_map.pdf",  width = 20, height = 15)
```



7 Summary

7.1 Filtering summary

The raw data are made of 8.265594×10^6 sequences representing 662 OTUs allocated to 80 samples.

After filtering, the dataset includes 8.243608×10^6 sequences representing 642 OTUs allocated to 72 samples.

7.2 Alpha diversity

Host age and elevation within tree do not impact any aspect of fungal local diversity. Despite similar OTUs richness, Asco is a site more diverse than Verghello and Bavella.

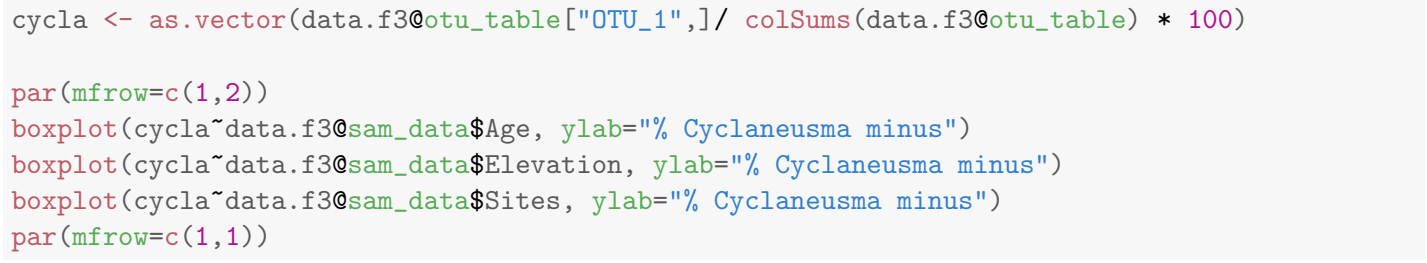
7.3 Beta diversity

Site (R2 = 0.136), age (R2 = 0.04) and interaction age*site (R2 = 0.091) statistically structured the fungal endophytic beta-diversity.

7.4 Special case of *Cyclaneusma minus*

Cyclaneusma minus account for 27.01% of total sequences.

```
cycla <- as.vector(data.f3@otu_table["OTU_1",]/ colSums(data.f3@otu_table) * 100)

par(mfrow=c(1,2))
boxplot(cycla~data.f3@sam_data$Age, ylab="% Cyclaneusma minus")
boxplot(cycla~data.f3@sam_data$Elevation, ylab="% Cyclaneusma minus")
boxplot(cycla~data.f3@sam_data$Sites, ylab="% Cyclaneusma minus")
par(mfrow=c(1,1))
```



```
tapply(cycla, data.f3@sam_data$Age, mean)

##   Mature      Old    Young
## 32.53361 26.34175 27.54212

tapply(cycla, data.f3@sam_data$Age, sd)

##   Mature      Old    Young
## 20.60563 15.48185 16.91920

tapply(cycla, data.f3@sam_data$Elevation, mean)

##   Bottom   Middle      Top
## 28.02981 32.06605 26.14226

tapply(cycla, data.f3@sam_data$Elevation, sd)

##   Bottom   Middle      Top
## 16.80456 19.75448 16.52367

tapply(cycla, data.f3@sam_data$Sites, mean)

##      Asco   Bavella Verghello
##  18.43440  42.90536  23.80010

tapply(cycla, data.f3@sam_data$Sites, sd)

##      Asco   Bavella Verghello
##  7.527581 20.169205 11.669665

wilcox.test(cycla[data.f3@sam_data$Sites=="Asco"], cycla[data.f3@sam_data$Sites=="Bavella"])

##
##  Wilcoxon rank sum test
##
## data:  cycla[data.f3@sam_data$Sites == "Asco"] and cycla[data.f3@sam_data$Sites == "Bavella"]
## W = 73, p-value = 2.292e-06
## alternative hypothesis: true location shift is not equal to 0
```

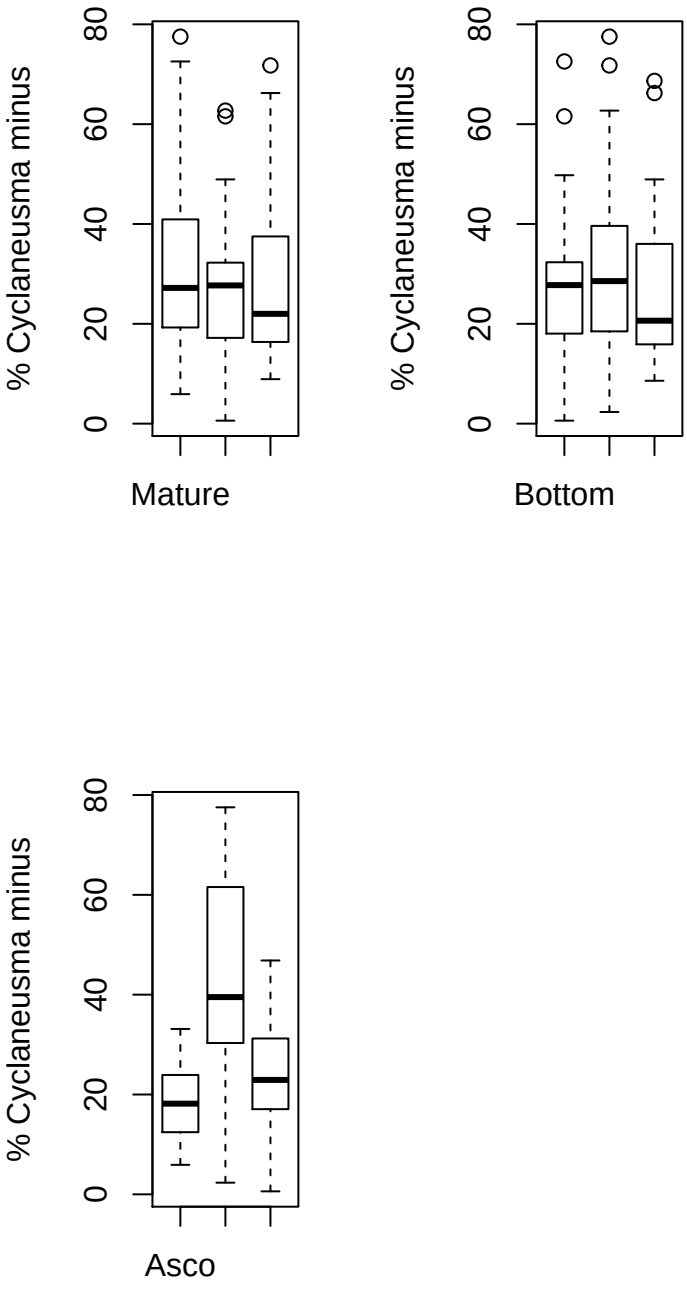


Figure 7.1: Number of sequences assigned to extitCyclaneusma minus across host age, elevation whitin tree and sites

```r
wilcox.test(cycla[data.f3@sam_data$Sites=="Verghello"], cycla[data.f3@sam_data$Sites=="Bavella"])

##
##  Wilcoxon rank sum test
##
## data:  cycla[data.f3@sam_data$Sites == "Verghello"] and cycla[data.f3@sam_data$Sites == "Bavella"
## W = 124, p-value = 0.0002846
## alternative hypothesis: true location shift is not equal to 0

wilcox.test(cycla[data.f3@sam_data$Sites=="Asco"], cycla[data.f3@sam_data$Sites=="Verghello"])

##
##  Wilcoxon rank sum test
##
## data:  cycla[data.f3@sam_data$Sites == "Asco"] and cycla[data.f3@sam_data$Sites == "Verghello"]
## W = 194, p-value = 0.08283
## alternative hypothesis: true location shift is not equal to 0
```

	Comparison	OTU_names	Species	Class	log2FoldChange (negative = more on second levels)
1	Verghello vs Asco	OTU_263			3.45292228957523
2	Verghello vs Asco	OTU_35			3.67157932437273
3	Verghello vs Asco	OTU_18		Dothideomycetes	-5.03682119702355
4	Verghello vs Asco	OTU_72	Rhytismatales sp	Leotiomycetes	-5.2132830208298
5	Verghello vs Asco	OTU_66			-24.986623457262
6	Verghello vs Asco	OTU_33	Aureobasidium pullulans	Dothideomycetes	-24.3713044357013
7	Verghello vs Asco	OTU_252		Dothideomycetes	-23.93410929467
8	Verghello vs Asco	OTU_30	fungal sp agrD488	unidentified	-9.10853333945915
9	Verghello vs Asco	OTU_42			-28.4820305751737
10	Verghello vs Asco	OTU_634			-23.9583367798625
11	Verghello vs Asco	OTU_188			-24.7300982035579
12	Verghello vs Asco	OTU_137		Leotiomycetes	20.7642479218886
13	Verghello vs Asco	OTU_260	Pseudeurotiaceae sp	Incertae sedis	22.684399409629
14	Verghello vs Asco	OTU_110		Dothideomycetes	21.4575349987156
15	Verghello vs Asco	OTU_419		Dothideomycetes	22.2879506251018
16	Verghello vs Asco	OTU_111		Dothideomycetes	23.342726115979
17	Verghello vs Asco	OTU_74		Tremellomycetes	23.4785654447647
18	Verghello vs Asco	OTU_105		Lecanoromycetes	25.3600815820556
19	Verghello vs Bavella	OTU_6	Dothioraceae sp	Dothideomycetes	2.43436510638349
20	Verghello vs Bavella	OTU_22		Dothideomycetes	2.50843902068639
21	Verghello vs Bavella	OTU_9	Lophodermium conigenum	Leotiomycetes	3.90899802973164
22	Verghello vs Bavella	OTU_23	fungal sp TRN287	unidentified	3.15506013630955
23	Verghello vs Bavella	OTU_13	Phaeomoniella sp	Eurotiomycetes	2.95955869810274
24	Verghello vs Bavella	OTU_46		Dothideomycetes	2.57668151165601
25	Verghello vs Bavella	OTU_39	Leotiomycetes sp BLD3	Leotiomycetes	3.1527692438903
26	Verghello vs Bavella	OTU_17		Incertae sedis	2.80911250488933
27	Verghello vs Bavella	OTU_189	Phaeotheca sp	Dothideomycetes	5.555732305606
28	Verghello vs Bavella	OTU_11			5.45724396452158
29	Verghello vs Bavella	OTU_20		Eurotiomycetes	4.14024539940136
30	Verghello vs Bavella	OTU_299	Ochrocladosporium sp	Dothideomycetes	4.56521436577709
31	Verghello vs Bavella	OTU_263			4.88652195288756
32	Verghello vs Bavella	OTU_227		Leotiomycetes	2.09712798527887
33	Verghello vs Bavella	OTU_35			7.46159411852082
34	Verghello vs Bavella	OTU_503		Leotiomycetes	3.77997102666681
35	Verghello vs Bavella	OTU_457	fungal sp TRN213	unidentified	24.7816410765751
36	Verghello vs Bavella	OTU_25	Phaeotheca sp	Dothideomycetes	3.24844039890926
37	Verghello vs Bavella	OTU_616		Leotiomycetes	3.80094139025534
38	Verghello vs Bavella	OTU_34			6.08840587096621
39	Verghello vs Bavella	OTU_66			-19.3731903051315
40	Verghello vs Bavella	OTU_252		Dothideomycetes	-20.3123823547936
41	Verghello vs Bavella	OTU_91		Lecanoromycetes	7.3064755898133
42	Verghello vs Bavella	OTU_634			-18.1098301812661
43	Verghello vs Bavella	OTU_89			22.3975058686059
44	Verghello vs Bavella	OTU_37			7.27182326168398
45	Verghello vs Bavella	OTU_74		Tremellomycetes	37.7775555062092
46	Verghello vs Bavella	OTU_105		Lecanoromycetes	30.5880791727876
47	Asco vs Bavella	OTU_6	Dothioraceae sp	Dothideomycetes	2.84850212169648
48	Asco vs Bavella	OTU_22		Dothideomycetes	2.99639401122662
49	Asco vs Bavella	OTU_9	Lophodermium conigenum	Leotiomycetes	3.94725546572457
50	Asco vs Bavella	OTU_23	fungal sp TRN287	unidentified	4.98215684102213
51	Asco vs Bavella	OTU_13	Phaeomoniella sp	Eurotiomycetes	3.7970632519075
52	Asco vs Bavella	OTU_63			4.57089540611465
53	Asco vs Bavella	OTU_3	Capnobotryella sp MA 4642	Dothideomycetes	4.5260359419882
54	Asco vs Bavella	OTU_46		Dothideomycetes	2.33214625459389
55	Asco vs Bavella	OTU_39	Leotiomycetes sp BLD3	Leotiomycetes	2.37642667202418
56	Asco vs Bavella	OTU_17		Incertae sedis	3.25658820351033
57	Asco vs Bavella	OTU_189	Phaeotheca sp	Dothideomycetes	8.59019650937849
58	Asco vs Bavella	OTU_11			7.25646667517131
59	Asco vs Bavella	OTU_20		Eurotiomycetes	5.25324352998105
60	Asco vs Bavella	OTU_299	Ochrocladosporium sp	Dothideomycetes	2.67956567352929
61	Asco vs Bavella	OTU_186		Dothideomycetes	2.51497201245116
62	Asco vs Bavella	OTU_380	fungal sp TRN213	unidentified	7.5578902365681
63	Asco vs Bavella	OTU_343			6.23573431660615
64	Asco vs Bavella	OTU_198	fungal sp TRN213	unidentified	3.84569827016386
65	Asco vs Bavella	OTU_60			6.36136947697016
66	Asco vs Bavella	OTU_36		Dothideomycetes	3.50517598773726
67	Asco vs Bavella	OTU_227		Leotiomycetes	1.93649677430791
68	Asco vs Bavella	OTU_35			3.7900147941481
69	Asco vs Bavella	OTU_67	Dothideomycetes sp 11147	Dothideomycetes	3.87834808875795
70	Asco vs Bavella	OTU_484	Dothioraceae sp	Dothideomycetes	2.67444483897171
71	Asco vs Bavella	OTU_503		Leotiomycetes	3.37019176795725
72	Asco vs Bavella	OTU_633	Phaeotheca sp	Dothideomycetes	8.08905161068396
73	Asco vs Bavella	OTU_570		Dothideomycetes	2.91884253804365
74	Asco vs Bavella	OTU_457	fungal sp TRN213	unidentified	22.0647898849091
75	Asco vs Bavella	OTU_640		Leotiomycetes	3.65504402154229
76	Asco vs Bavella	OTU_18		Dothideomycetes	3.1887203580742
77	Asco vs Bavella	OTU_616		Leotiomycetes	3.85167352567851
78	Asco vs Bavella	OTU_64			4.5988128981545
79	Asco vs Bavella	OTU_33	Aureobasidium pullulans	Dothideomycetes	25.1427779779013
80	Asco vs Bavella	OTU_16	Aureobasidium pullulans	Dothideomycetes	4.12488534210475
81	Asco vs Bavella	OTU_30	fungal sp agrD488	unidentified	13.7892943991293
82	Asco vs Bavella	OTU_73	Dothideomycetes sp CCFEE 5409	Dothideomycetes	7.27386361978797
83	Asco vs Bavella	OTU_379			7.32992029936617
84	Asco vs Bavella	OTU_42			29.1857906770534
85	Asco vs Bavella	OTU_65			11.057073409853
86	Asco vs Bavella	OTU_104	Sarea sp	Lecanoromycetes	6.498453306476
87	Asco vs Bavella	OTU_89			26.0991540009176
88	Asco vs Bavella	OTU_102			6.84778357772273
89	Asco vs Bavella	OTU_37			9.8277399967367
90	Asco vs Bavella	OTU_101			7.97860444533908
91	Asco vs Bavella	OTU_188			25.6204181467846
92	Asco vs Bavella	OTU_137		Leotiomycetes	-18.3044357308405
93	Asco vs Bavella	OTU_260	Pseudeurotiaceae sp	Incertae sedis	-16.2452623442606
94	Asco vs Bavella	OTU_110		Dothideomycetes	-15.2103922196156
95	Asco vs Bavella	OTU_419		Dothideomycetes	-13.71336281922
96	Asco vs Bavella	OTU_111		Dothideomycetes	-23.1078137625913
97	Asco vs Bavella	OTU_74		Tremellomycetes	14.2989900614445

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

	Comparison	Order	Class	log2FoldChange (negative = more on second levels)
1	Verghello vs Asco	Xylariales	Sordariomycetes	5.04685635990588
2	Verghello vs Bavella	Incertae sedis	Leotiomycetes	-1.24945441899881
3	Verghello vs Bavella	unidentified	unidentified	1.54763743320594
4	Asco vs Bavella	Botryosphaeriales	Dothideomycetes	7.32494363316305
5	Asco vs Bavella	Eurotiales	Eurotiomycetes	1.81014696013106
6	Asco vs Bavella	Incertae sedis	Leotiomycetes	-1.64446153540851
7	Asco vs Bavella	unidentified	unidentified	1.48905182023372
8	Asco vs Bavella	Xylariales	Sordariomycetes	-4.8057409530055

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

	Nb.of.OTUs	Nb.of.samples	Nb.of.sequences
No filter	662	80	8265594.00
Nb of sequences by sample >= 20000	654	72	8243646.00
Nb of sample by OTUs >= 1	654	72	8243646.00
Nb of sequences by OTUs >= 5	642	72	8243608.00

Table 15: Number of OTUs, samples and sequences after filtering

List of Figures

List of Tables