Appendix S8: results after Qiime Closed reference clustering. 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 Qiime Closed reference clustering (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 15:01:39. 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 <- "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(1, 3, 5, 7, 9, 11, 13, 15)]
taxRDP <- tax_table(as.matrix(taxRDP))
taxa_names(taxRDP) <- taxa_names(dataBiom)
colnames(taxRDP) <- c("Species Hypothesis", "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(rownames(taxRDP2), funguild$OTU_ID)

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


taxRDP2 <- tax_table(as.matrix(taxRDP2))
taxa_names(taxRDP2) <- taxa_names(dataBiom)
colnames(taxRDP2) <- c("Species Hypothesis", "Domain", "Phylum", "Class", "Order", "Family", "Genus"
                       "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(repset), by = 2)]
```

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 13 taxonomic ranks ]
## refseq()      DNAStringSet:      [ 256 reference sequences ]
```

The data are made of 6.473782×10^6 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 2×10^4 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)
```

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).

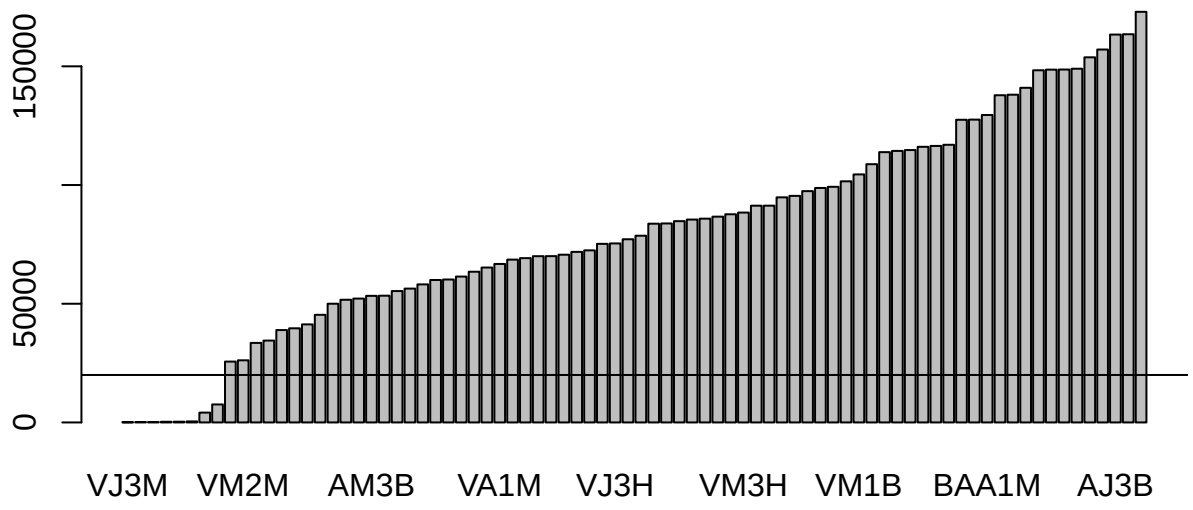


Figure 2.1: Number of sequences by sample. Horizontal line indicates the filtering parameter.

```
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    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%).

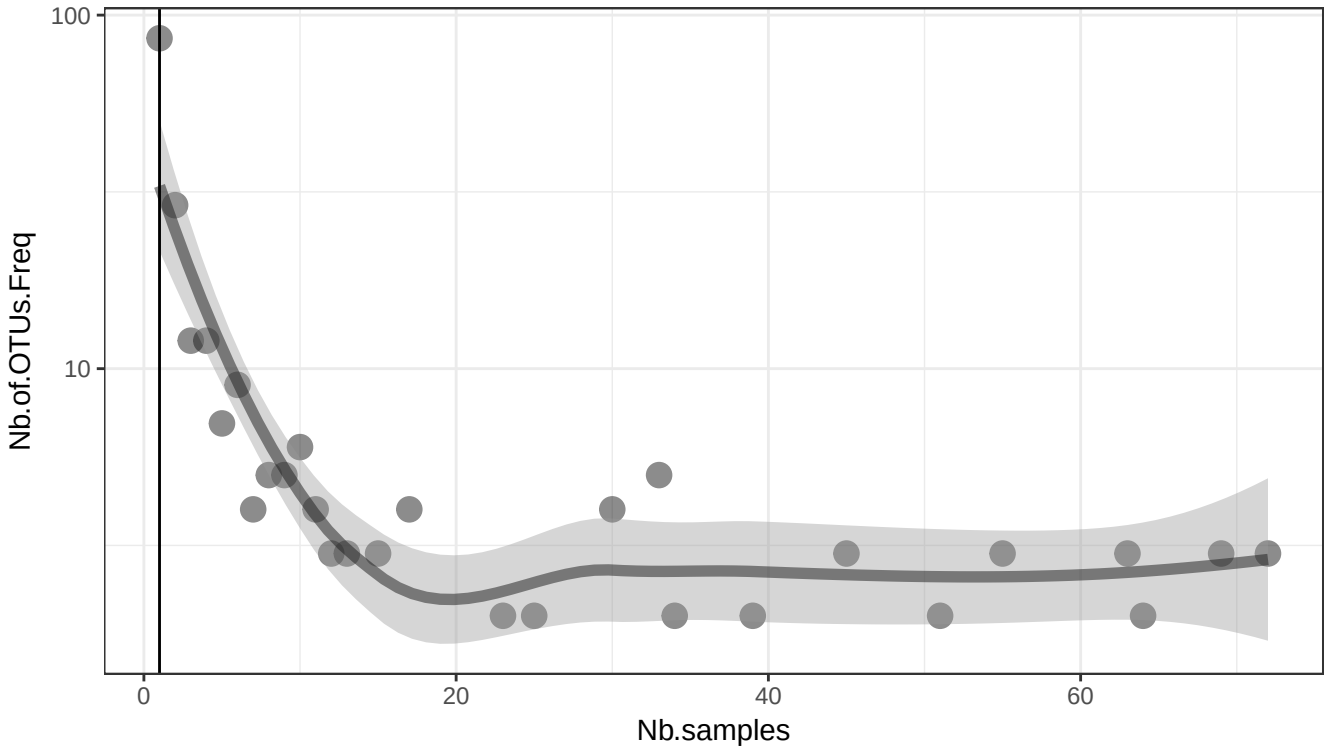


Figure 2.2: Number of OTU present in a given number of samples. Vertical bar illustrates the filtering parameter.

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.
##       1.0      32.8     203.5   25637.0    1856.5 2291731.0
```

```
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)
```

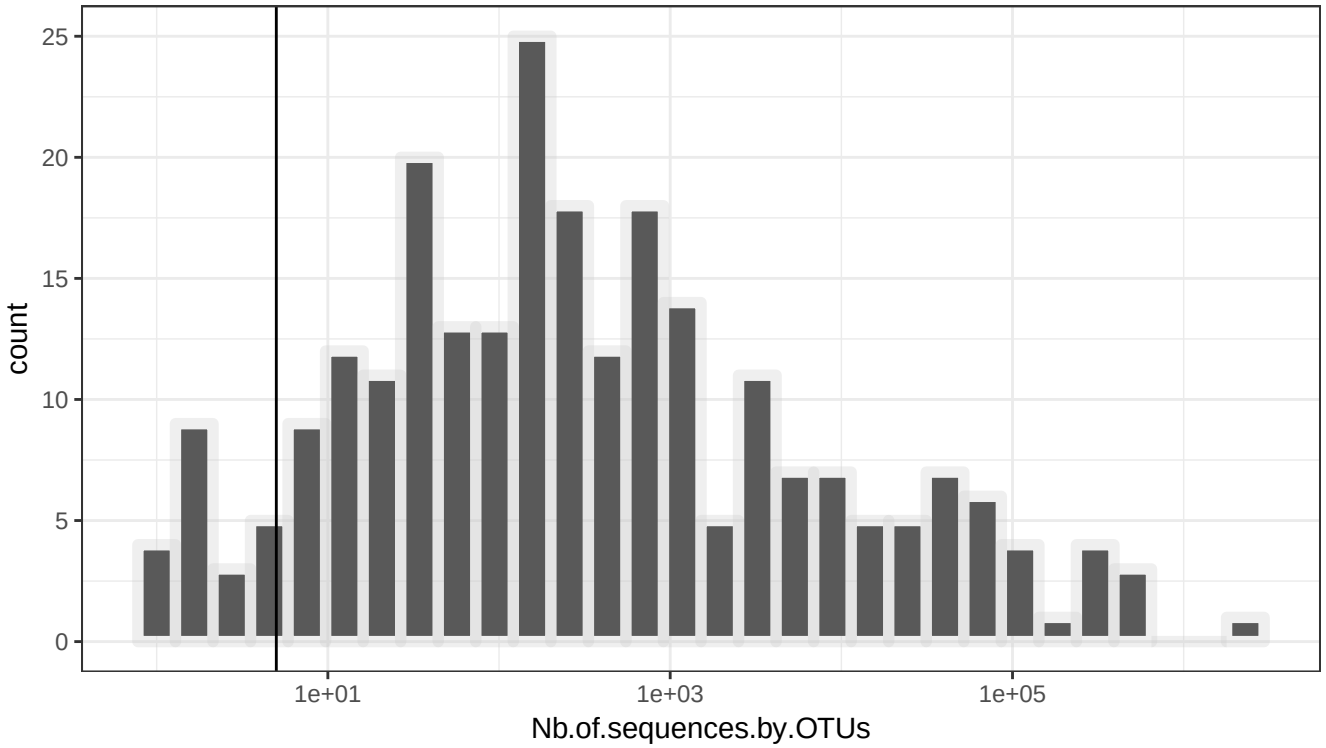


Figure 2.3: Number of sequences by OTU (log10 transformed). Horizontal bar illustrates the filtering parameter.

2.6 Summary of filtration workflow

The filtered data are made of 6.460489×10^6 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') +
```

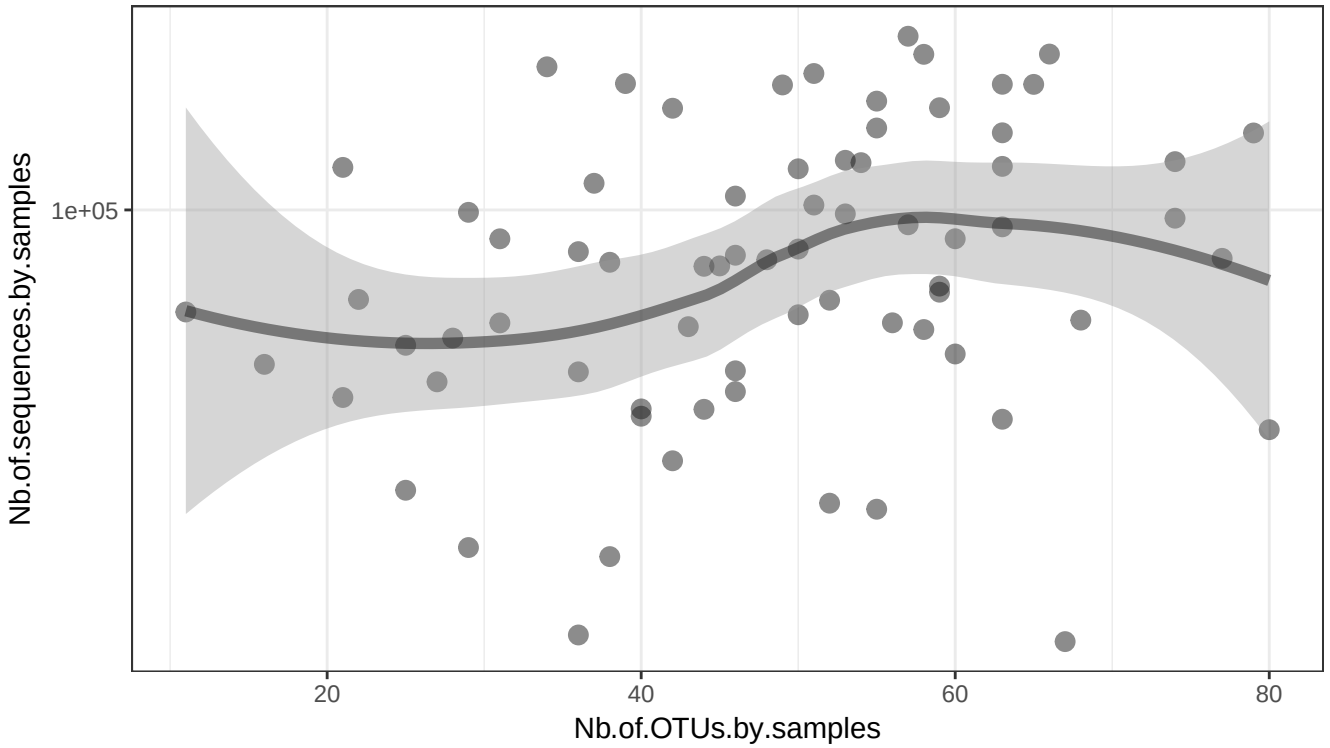


Figure 3.1: Number of OTUs by sample in fonction of the number of sequences by sample (log10 axe). The tendency is represented by the line obtained from loess (Local Polynomial Regression Fitting).

```
  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])

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"))
```

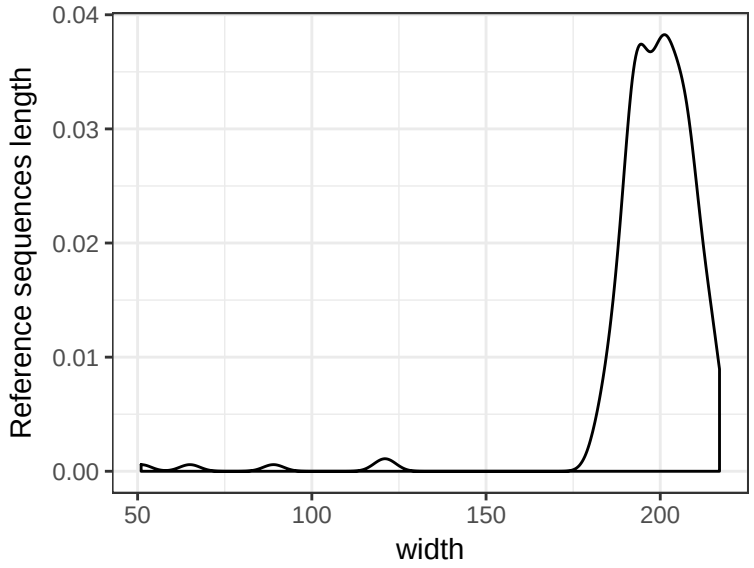


Figure 3.2: Distribution of reference sequences length.

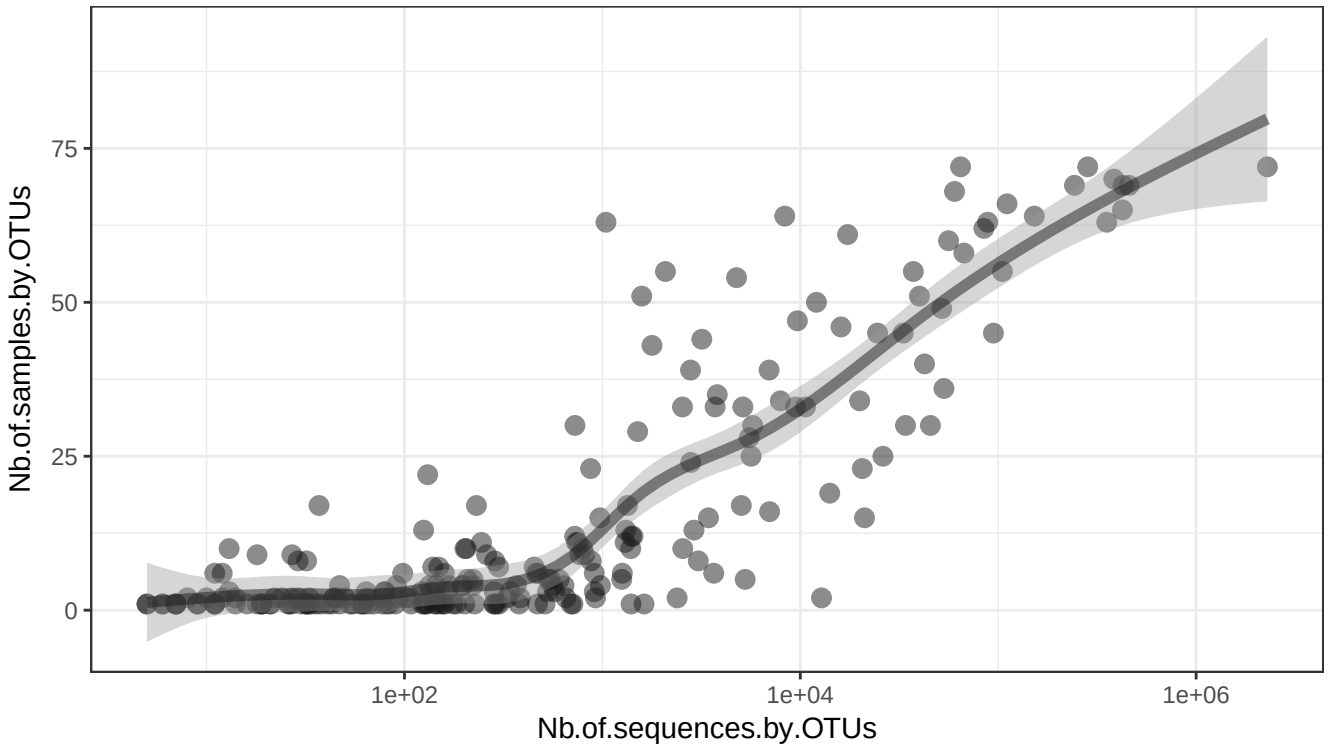


Figure 3.3: Number of sequences by OTUs (log10 axe) in fonction of the number of samples where OTUs were found. The tendency is represented by the line obtain from gam (Generalized additive models with integrated smoothness estimation).

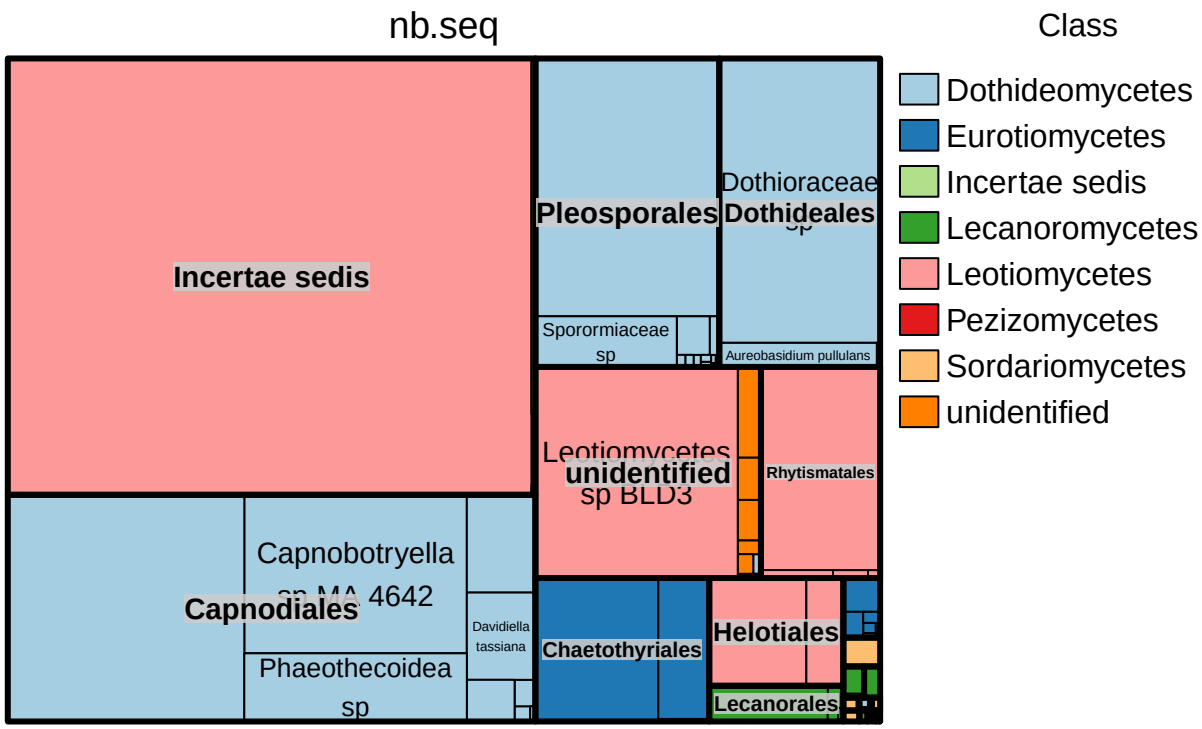


Figure 3.4: Distribution of the number of sequences in the Ascomycota taxonomy. Colors represent Class, bold lines delimit Order and thick line delimit species.

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)
```

```
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)
```

Domain	Phylum	Class	Order	Family	Genus	Species	Trophic_Mode	Trait
Fungi	Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-
Fungi	Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp	-	-
Fungi	Ascomycota	Leotiomycetes	unidentified	unidentified	unidentified	Leotiomycetes sp BLD3	-	-
Fungi	Ascomycota	Dothideomycetes	Pleosporales				-	-
Fungi	Ascomycota	Dothideomycetes	Capnodiales				-	-
Fungi	Ascomycota	Dothideomycetes	Capnodiales	Incertae sedis	Capnobotryella	Capnobotryella sp MA 4642	Saprotroph	NULL
Fungi	Ascomycota						-	-
Fungi	Ascomycota	Leotiomycetes	Rhytismatales	Rhytismataceae	Lophodermium		Pathotroph	NULL
Fungi	Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Phaeothecoidea	Phaeothecoidea sp	Saprotroph	NULL
Fungi	Ascomycota	Dothideomycetes	Capnodiales				-	-
Fungi	Ascomycota	Leotiomycetes	Helotiales	Hyaloscyphaceae	Lachnellula	Lachnellula calyciformis	Saprotroph	NULL
Fungi							-	-
Fungi	Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Phaeomoniella	Phaeomoniella sp	Saprotroph	NULL
Fungi	Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Phaeomoniella	Phaeomoniella sp	Saprotroph	NULL
Fungi	Ascomycota	Dothideomycetes	Capnodiales	Incertae sedis	Phaeotheca	Phaeotheca sp	-	-
Fungi	Ascomycota						-	-
Fungi	Ascomycota	Dothideomycetes	Capnodiales	Davidiellaceae	Davidiella	Davidiella tassiana	Saprotroph	NULL
Fungi	Ascomycota						-	-
Fungi	Ascomycota						-	-
Fungi	Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae			-	-
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Sporormiaceae	unidentified	Sporormiaceae sp	-	-
Fungi	Ascomycota	Lecanoromycetes	Lecanorales	Parmeliaceae	Pseudevernia	Pseudevernia furfuracea	Symbiotroph	NULL
Fungi							-	-
Fungi	Ascomycota	Leotiomycetes	Helotiales	Hyaloscyphaceae	Lachnellula		Saprotroph	NULL
Fungi	Ascomycota	Leotiomycetes					-	-
Fungi	Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	Aureobasidium	Aureobasidium pullulans	Saprotroph	NULL
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Sporormiaceae	unidentified	Sporormiaceae sp	-	-
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Sporormiaceae	unidentified		-	-
Fungi	unidentified	unidentified	unidentified	unidentified	unidentified	fungal sp agrD488	-	-

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

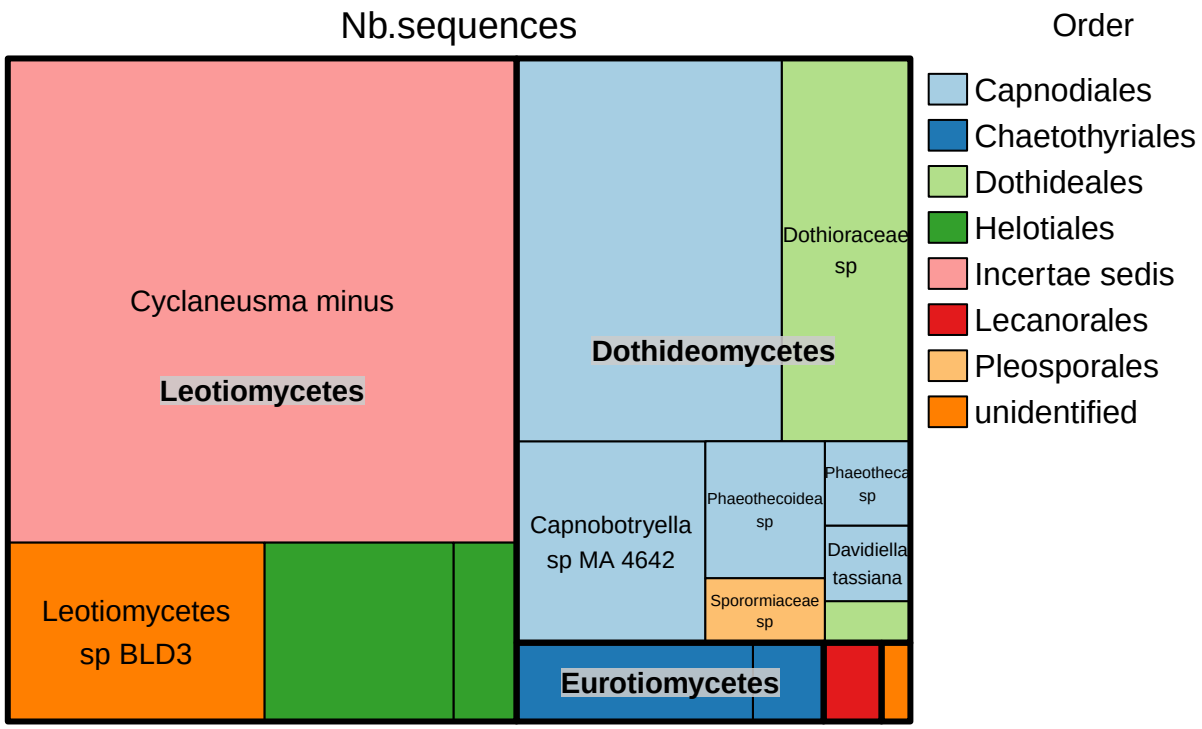


Figure 3.5: Number of sequences of the 30 most abundant OTUs (number of sequences). Colors indicate Order, bold lines delimit Class and thick lines delimit species.

```
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)
```

Domain	Phylum	Class	Order	Family	Genus	Species	Trophic_Mode	Trait
Fungi	Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-
Fungi	Ascomycota						-	-
Fungi	Ascomycota						-	-
Fungi	Ascomycota	Dothideomycetes	Capnodiales				-	-
Fungi	Ascomycota	Leotiomycetes	Rhytismatales	Rhytismataceae	Lophodermium		Pathotroph	NULL
Fungi	Ascomycota	Leotiomycetes	unidentified	unidentified	unidentified	Leotiomycetes sp BLD3	-	-
Fungi	Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp	-	-
Fungi	Ascomycota	Dothideomycetes	Capnodiales	Davidiellaceae	Davidiella	Davidiella tassiana	Saprotroph	NULL
Fungi	Ascomycota	Dothideomycetes	Capnodiales				-	-
Fungi	Ascomycota	Dothideomycetes	Pleosporales				-	-
Fungi	Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Phaeothecoidea	Phaeothecoidea sp	Saprotroph	NULL
Fungi	Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Phaeothecoidea		Saprotroph	NULL
Fungi	Ascomycota	Dothideomycetes	Capnodiales	Incertae sedis	Capnobotryella	Capnobotryella sp MA 4642	Saprotroph	NULL
Fungi	Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Phaeomoniella	Phaeomoniella sp	Saprotroph	NULL
Fungi	Ascomycota	Leotiomycetes	unidentified	unidentified	unidentified	Leotiomycetes sp BLD3	-	-
Fungi	Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Phaeomoniella	Phaeomoniella sp	Saprotroph	NULL
Fungi	Ascomycota	Dothideomycetes	Capnodiales				-	-
Fungi	Ascomycota						-	-
Fungi	Ascomycota	Dothideomycetes	Capnodiales	Incertae sedis	Phaeotheca	Phaeotheca sp	-	-
Fungi	Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Phaeomoniella	Phaeomoniella sp	Saprotroph	NULL
Fungi	Ascomycota	Leotiomycetes	Helotiales	Hyaloscyphaceae	Lachnellula	Lachnellula calyciformis	Saprotroph	NULL
Fungi	Ascomycota	Leotiomycetes	Helotiales	Hyaloscyphaceae	Lachnellula		Saprotroph	NULL
Fungi	Ascomycota	Dothideomycetes	Pleosporales				-	-
Fungi							-	-
Fungi	Ascomycota	Leotiomycetes	Helotiales	Hyaloscyphaceae			-	-
Fungi	Ascomycota	Eurotiomycetes	Chaetothyriales				-	-
Fungi	Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae			-	-
Fungi	unidentified	unidentified	unidentified	unidentified	unidentified	fungal sp TRN256	-	-
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Pleosporaceae	Alternaria		Pathotroph	NULL

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

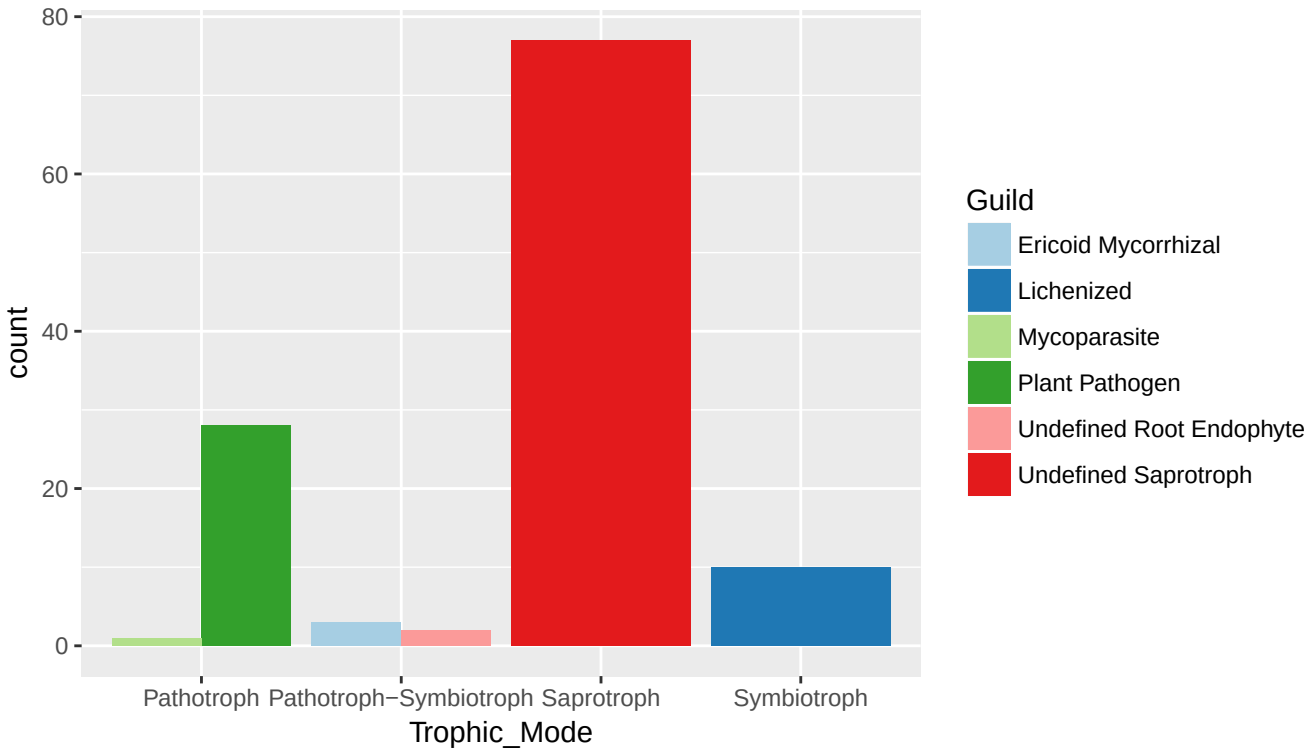


Figure 4.1: Distribution of OTUs into functional Guild.

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()
```

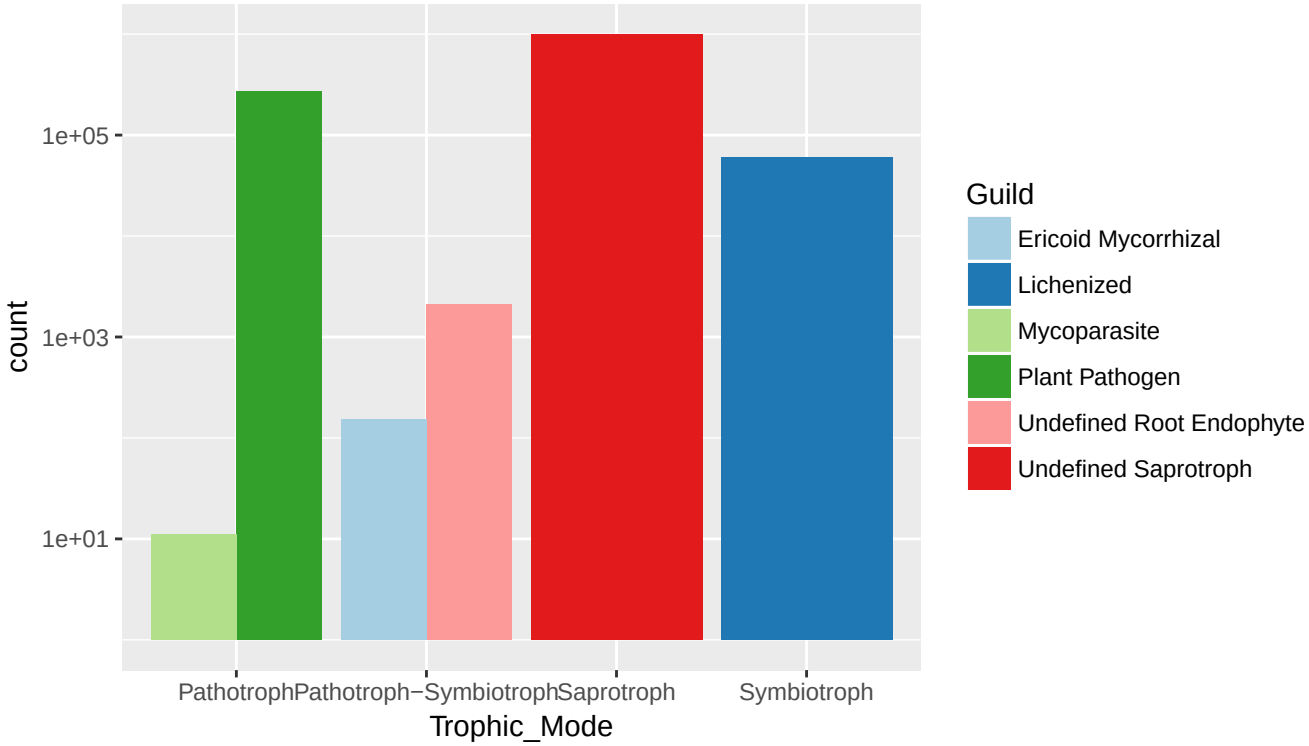


Figure 4.2: Distribution of sequences (log10 transformed) into functional Guild.

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)
```

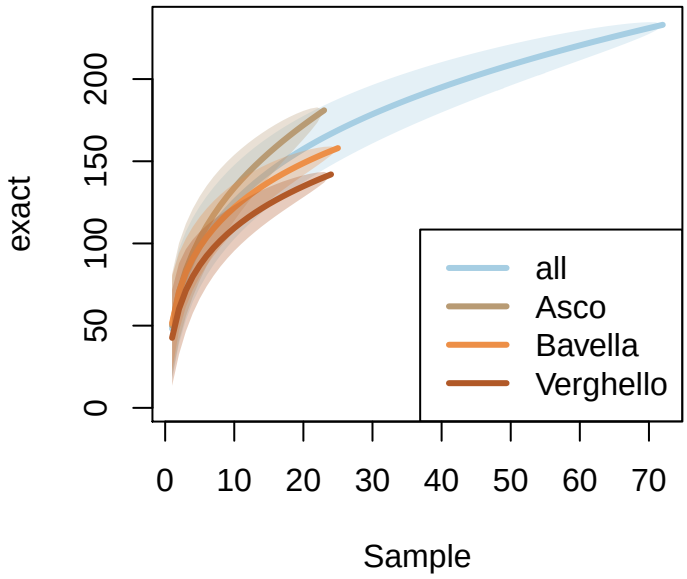


Figure 5.1: Rarefaction curves for each site. Note that if singletons were removed, these curves are biaised.

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)
```

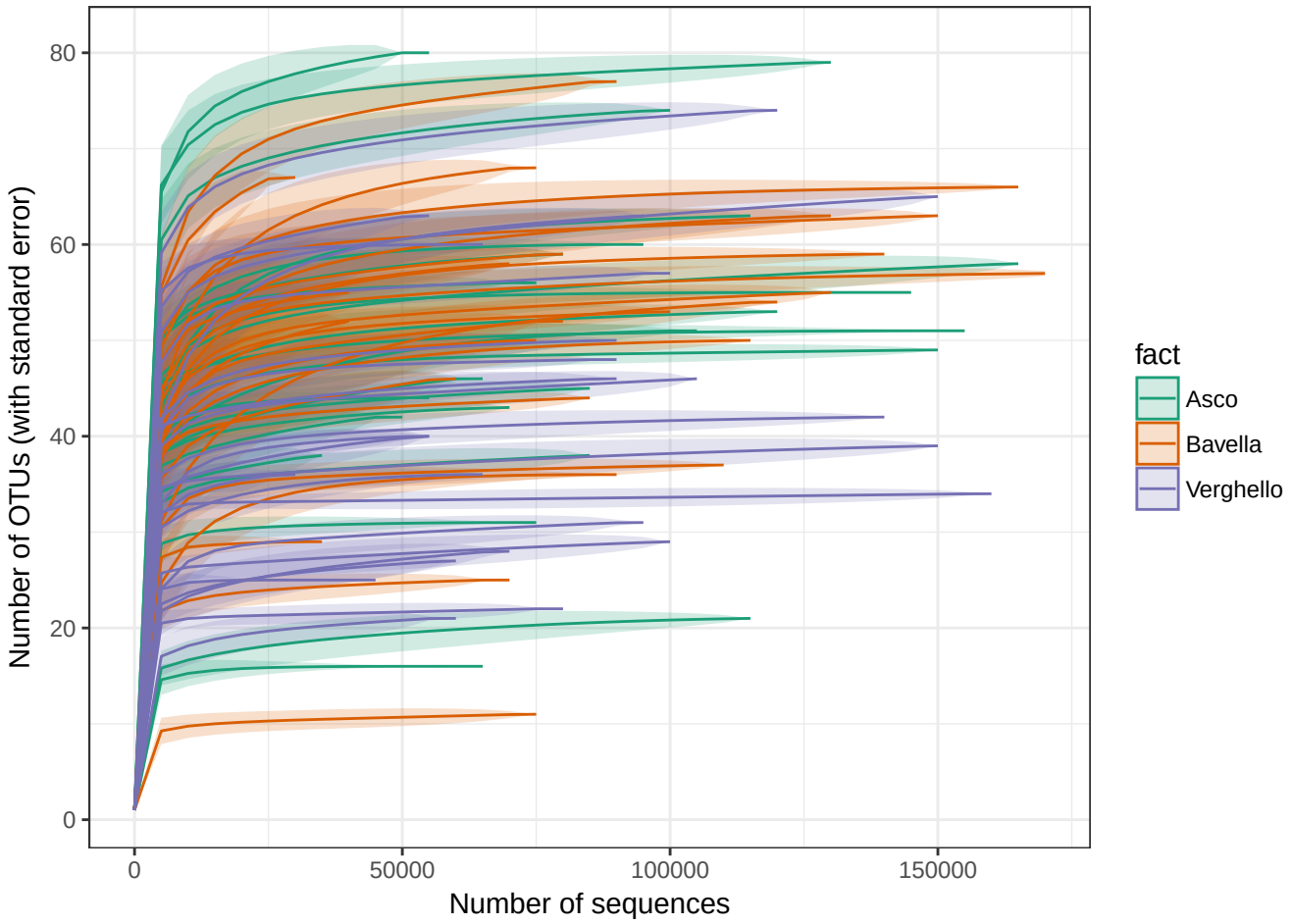


Figure 5.2: Rarefaction curves for each sample using sequences number on x-axes. Note that if singletons were removed, these curves are biaised.

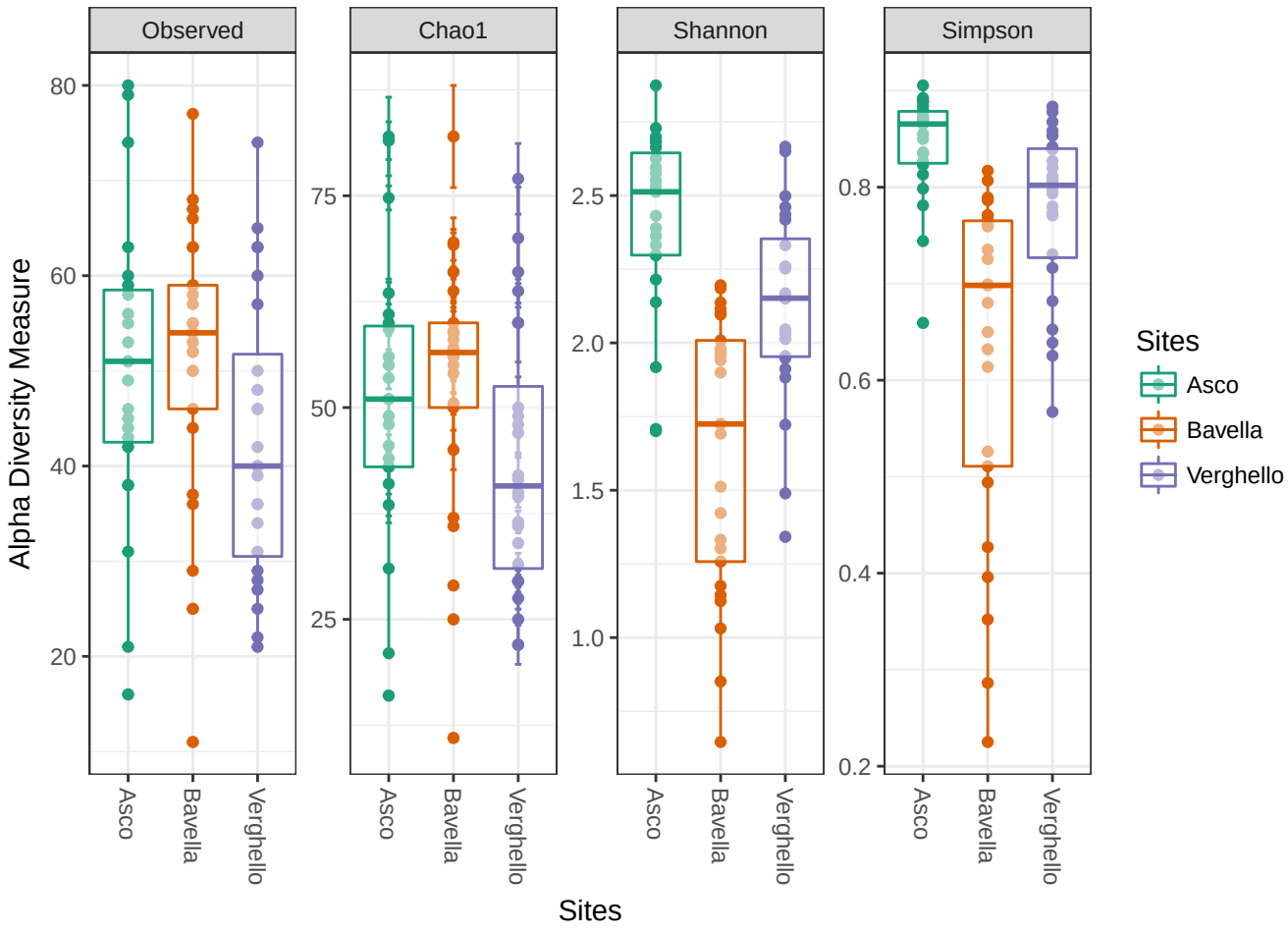


Figure 5.3: Diversity of each sites

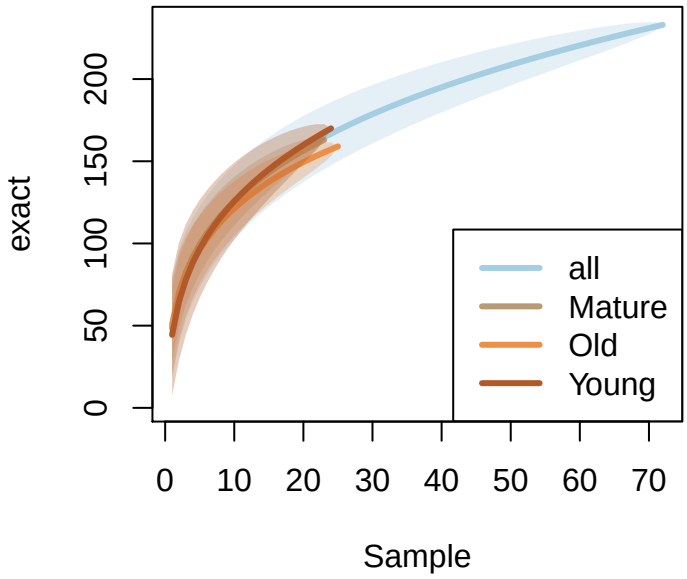


Figure 5.4: Rarefaction curves for each host age. Note that if singletons were removed, these curves are biaised.

```
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)	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 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))
```

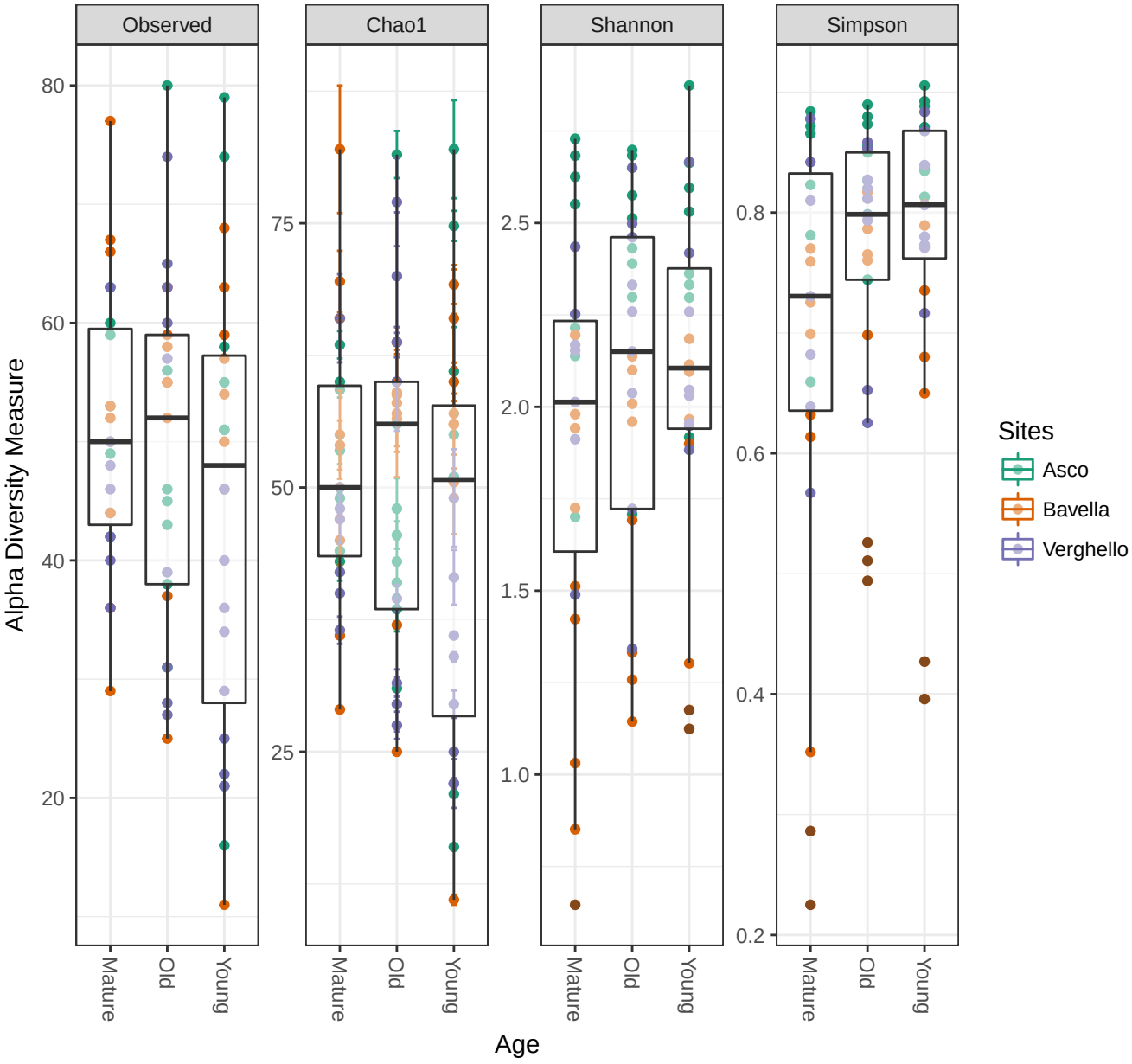


Figure 5.5: Diversity in function of tree age. Color represent sites.

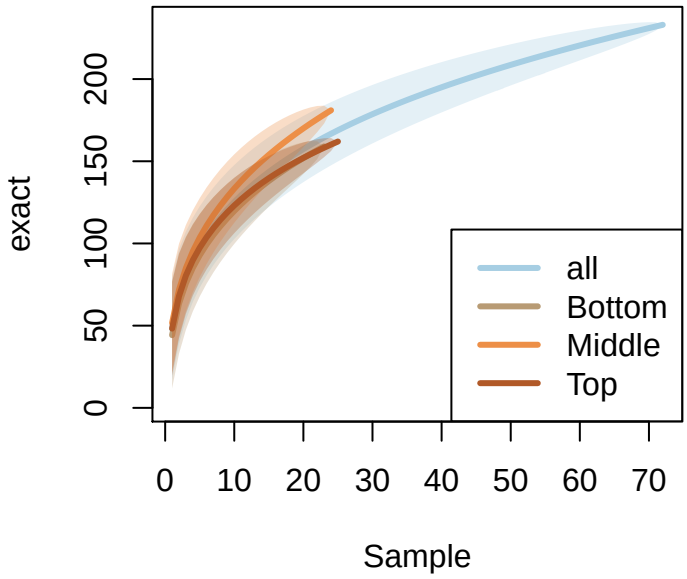


Figure 5.6: Rarefaction curves for each elevation. Notes that if singletons were removed, these curves are biaised.

```
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("")
```

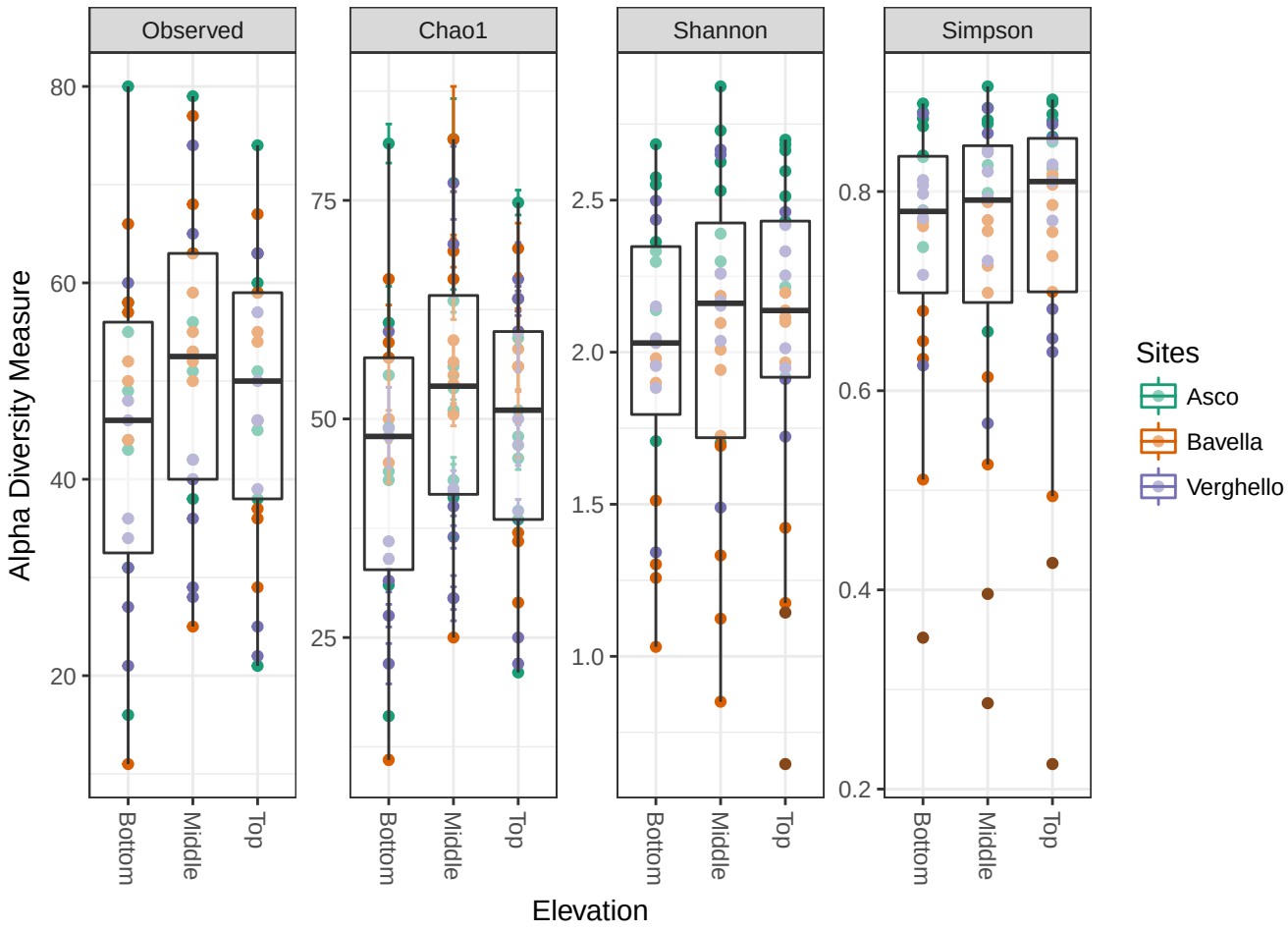


Figure 5.7: Diversity in function of elevation. Color represent sites.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	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 Shannon's entropy index (Hill number with 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 Simpson's concentration index (Hill number with q = 2)

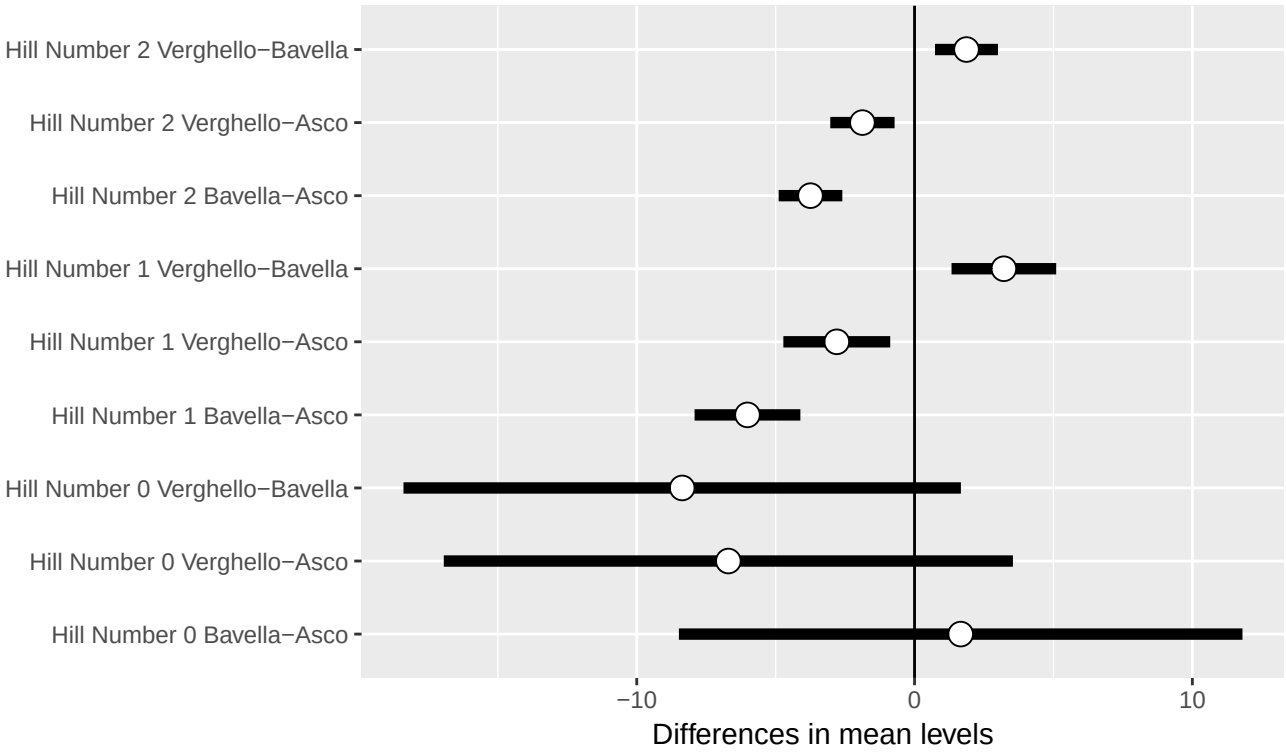


Figure 5.8: Results of the Tuckey HSD testing for differences in mean Hill numbers among pairs of modalities

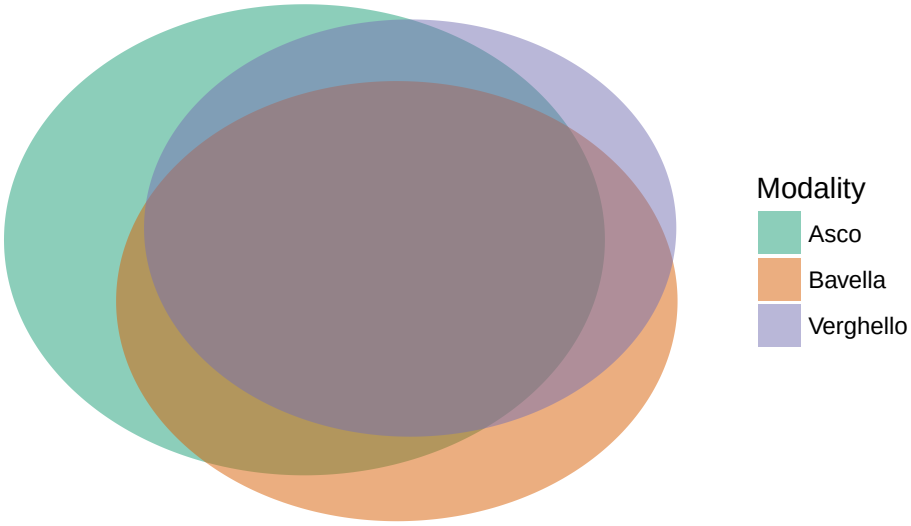


Figure 6.1: Venn diagramm of the distribution of OTUs among Sites

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).

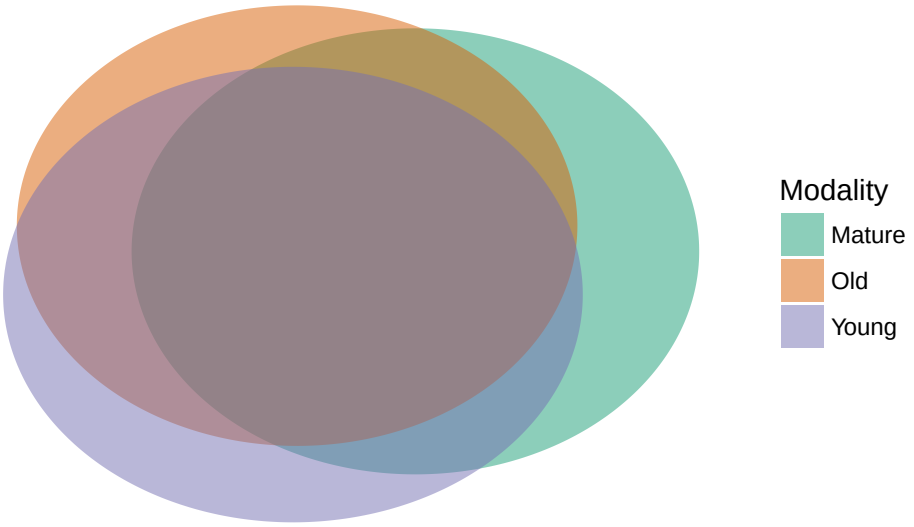


Figure 6.2: Venn diagramm of the distribution of OTUs among host age

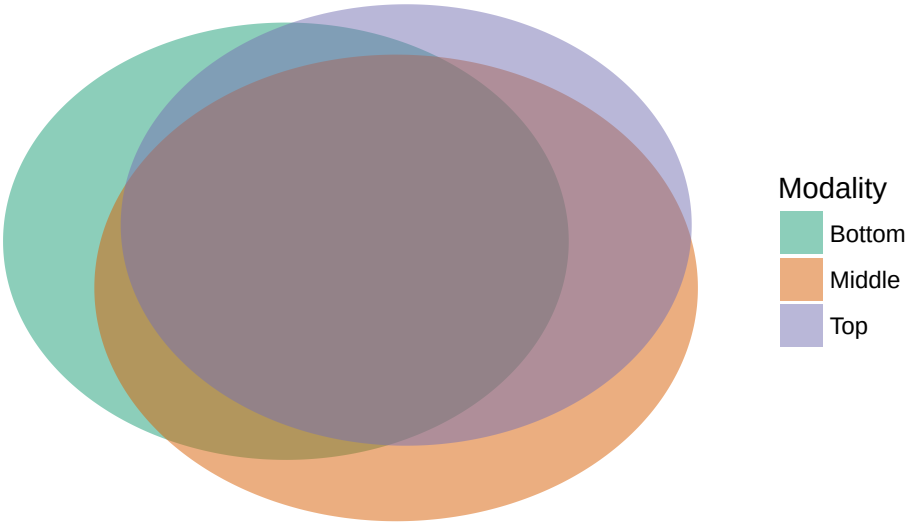


Figure 6.3: Venn diagramm of the distribution of OTUs among elevation of samples

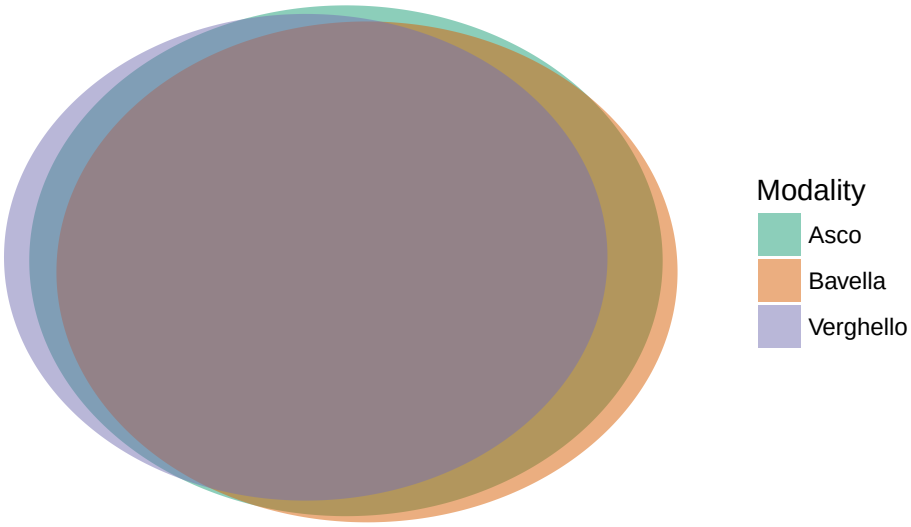


Figure 6.4: Venn diagramm of the distribution of OTUs among Sites

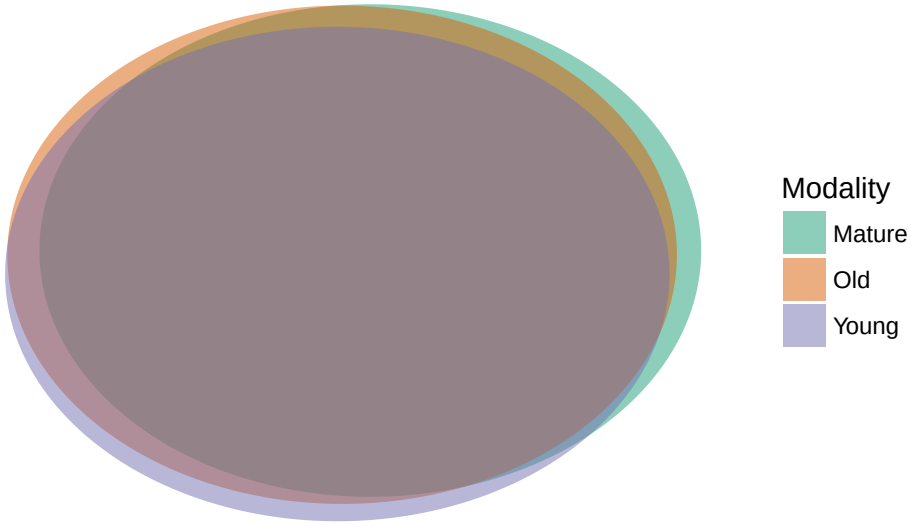


Figure 6.5: Venn diagramm of the distribution of OTUs among host age

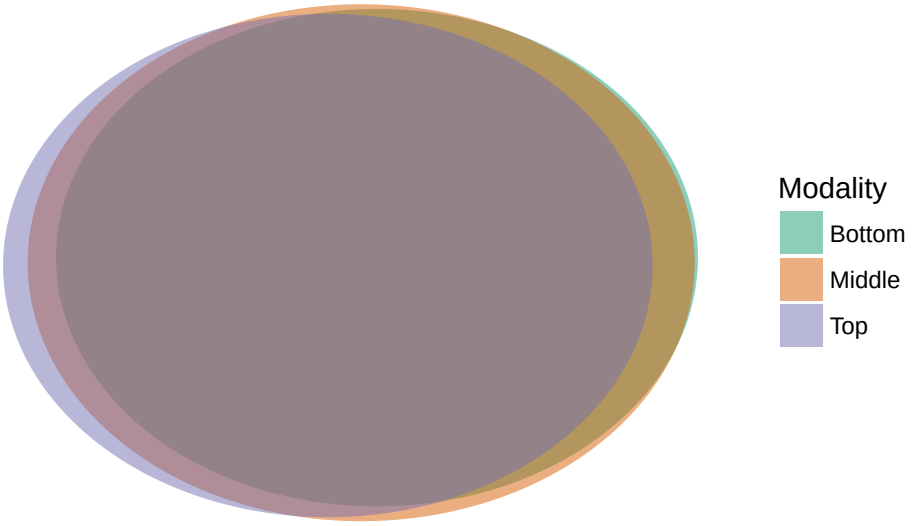


Figure 6.6: Venn diagramm of the distribution of OTUs among elevation of samples whitin the tree

```
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 0.2451014
## Run 1 stress 0.2483763
## Run 2 stress 0.2443706
## ... New best solution
## ... Procrustes: rmse 0.04031939  max resid 0.1755517
## Run 3 stress 0.2429748
## ... New best solution
## ... Procrustes: rmse 0.03382549  max resid 0.1397093
## Run 4 stress 0.2518202
## Run 5 stress 0.2604871
## Run 6 stress 0.2548659
## Run 7 stress 0.2464947
## Run 8 stress 0.2439988
## Run 9 stress 0.2560798
## Run 10 stress 0.2466527
```

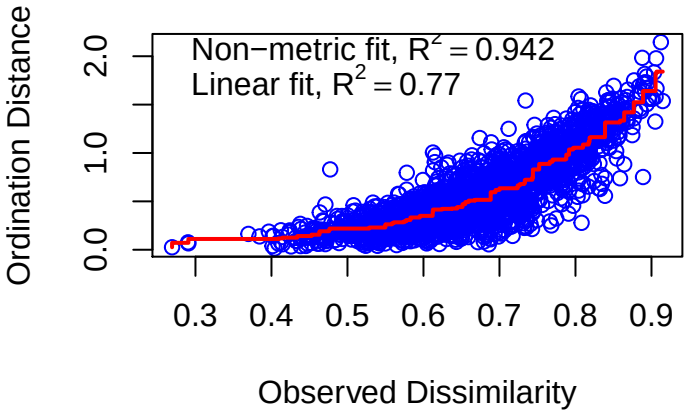


Figure 6.7: Stress plot of the NMDS

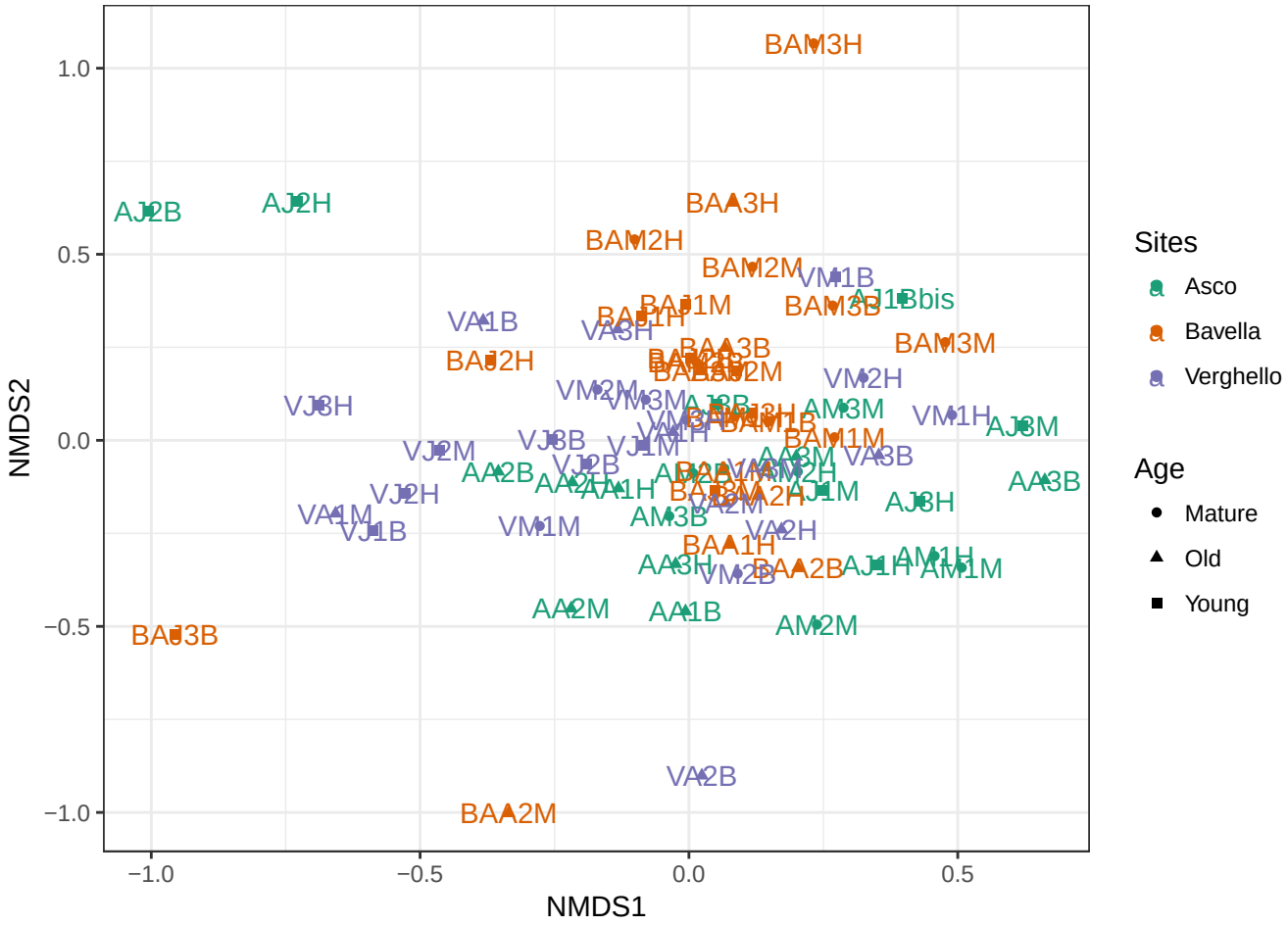


Figure 6.8: NMDS of OTU table. Colors represent sites and shape the age of tree.

```
## Run 11 stress 0.2514242
## Run 12 stress 0.2565368
## Run 13 stress 0.260919
## Run 14 stress 0.2453975
## Run 15 stress 0.2592331
## Run 16 stress 0.244006
## Run 17 stress 0.246079
## Run 18 stress 0.2635747
## Run 19 stress 0.2442591
## Run 20 stress 0.2448261
## *** 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)
```

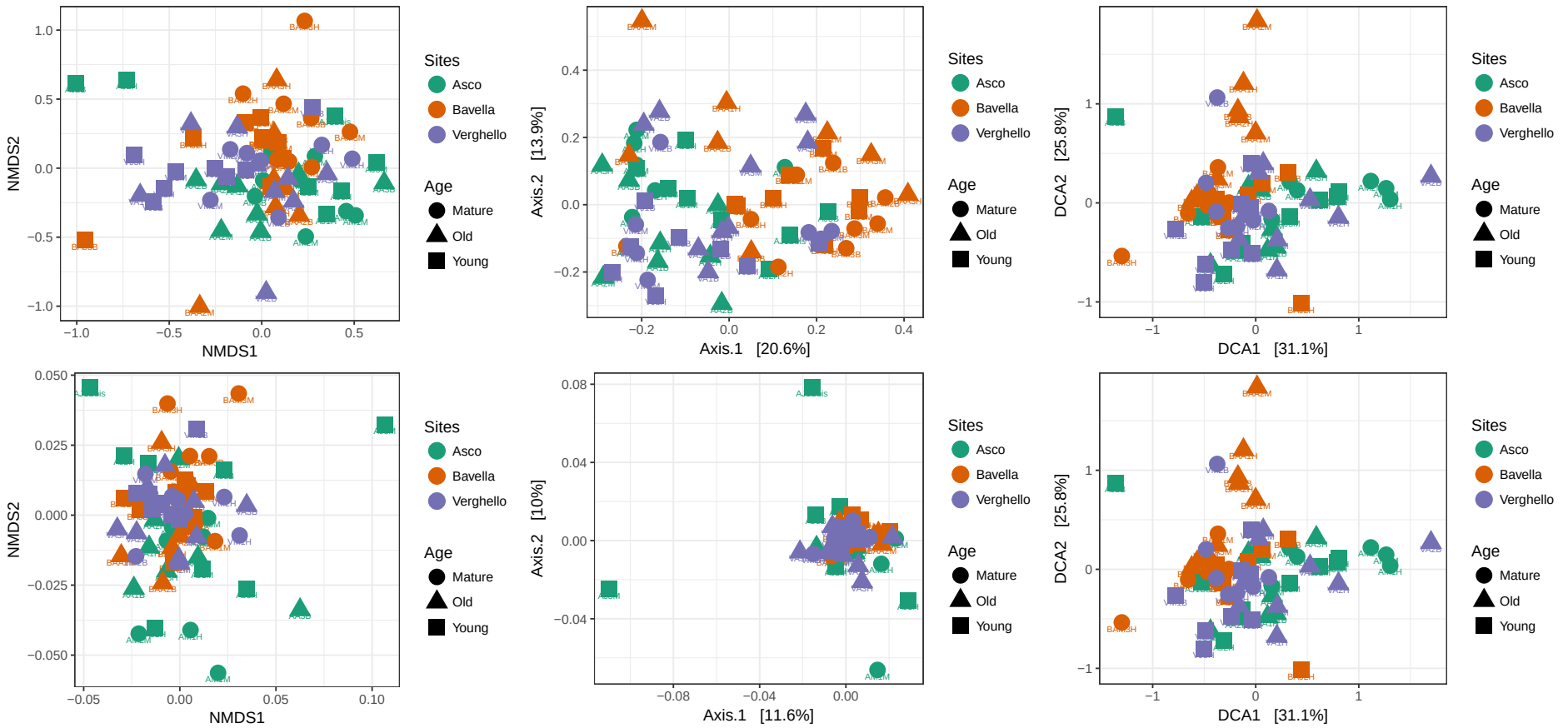


Figure 6.9: Comparison of different distances (bray (up) and gower (bottom)) and ordination methods (NMDS (left), PCoA (center) and DCA (right)).

```
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 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)
```

```
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	1.61	0.81	5.30	0.13	0.0001
Age	2	0.55	0.28	1.82	0.04	0.0141
Elevation	2	0.43	0.21	1.41	0.03	0.1014
Sites:Age	4	1.31	0.33	2.15	0.10	0.0002
Sites:Elevation	4	0.52	0.13	0.85	0.04	0.7380
Age:Elevation	4	0.53	0.13	0.87	0.04	0.7054
Sites:Age:Elevation	8	1.06	0.13	0.87	0.08	0.7873
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")
```

```
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	1.59	0.79	5.42	0.13	0.0001
Age	2	0.54	0.27	1.84	0.04	0.0170
Elevation	2	0.42	0.21	1.43	0.03	0.1085
Sites:Age	4	1.28	0.32	2.18	0.10	0.0004
Sites:Elevation	4	0.49	0.12	0.84	0.04	0.7482
Age:Elevation	4	0.51	0.13	0.86	0.04	0.6972
Sites:Age:Elevation	8	1.01	0.13	0.86	0.08	0.7926
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

	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	0.0183
Elevation	2	0.18	0.09	0.96	0.03	0.5061
Sites:Age	4	0.45	0.11	1.22	0.07	0.1353
Sites:Elevation	4	0.29	0.07	0.80	0.04	0.8563
Age:Elevation	4	0.35	0.09	0.95	0.05	0.5578
Sites:Age:Elevation	8	0.61	0.08	0.84	0.09	0.8720
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)).

```
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	1.61	0.81	6.50	0.13	0.0001
Age	2	0.55	0.28	2.23	0.04	0.0032
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).

	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	0.0022
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

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Sites	2	0.52	0.26	3.40	0.08	0.0001
Age	2	0.31	0.16	2.06	0.05	0.0038
Sites:Age	4	0.46	0.12	1.51	0.07	0.0153
Sites:Age:IndividualTree	18	2.08	0.12	1.52	0.31	0.0001
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)).

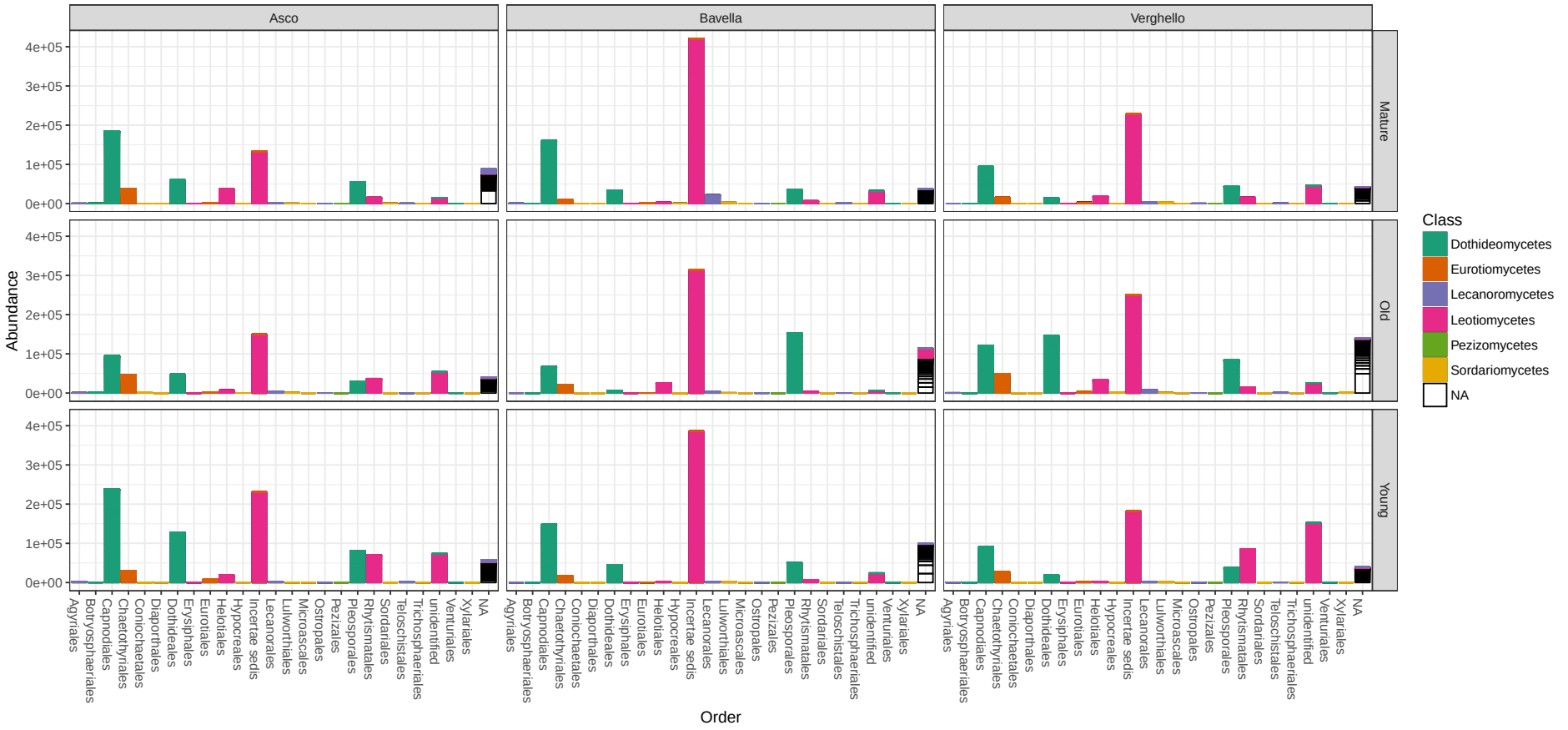


Figure 6.10: Taxonomic distribution of sequences in the different site * age combinaison.

```
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")
```

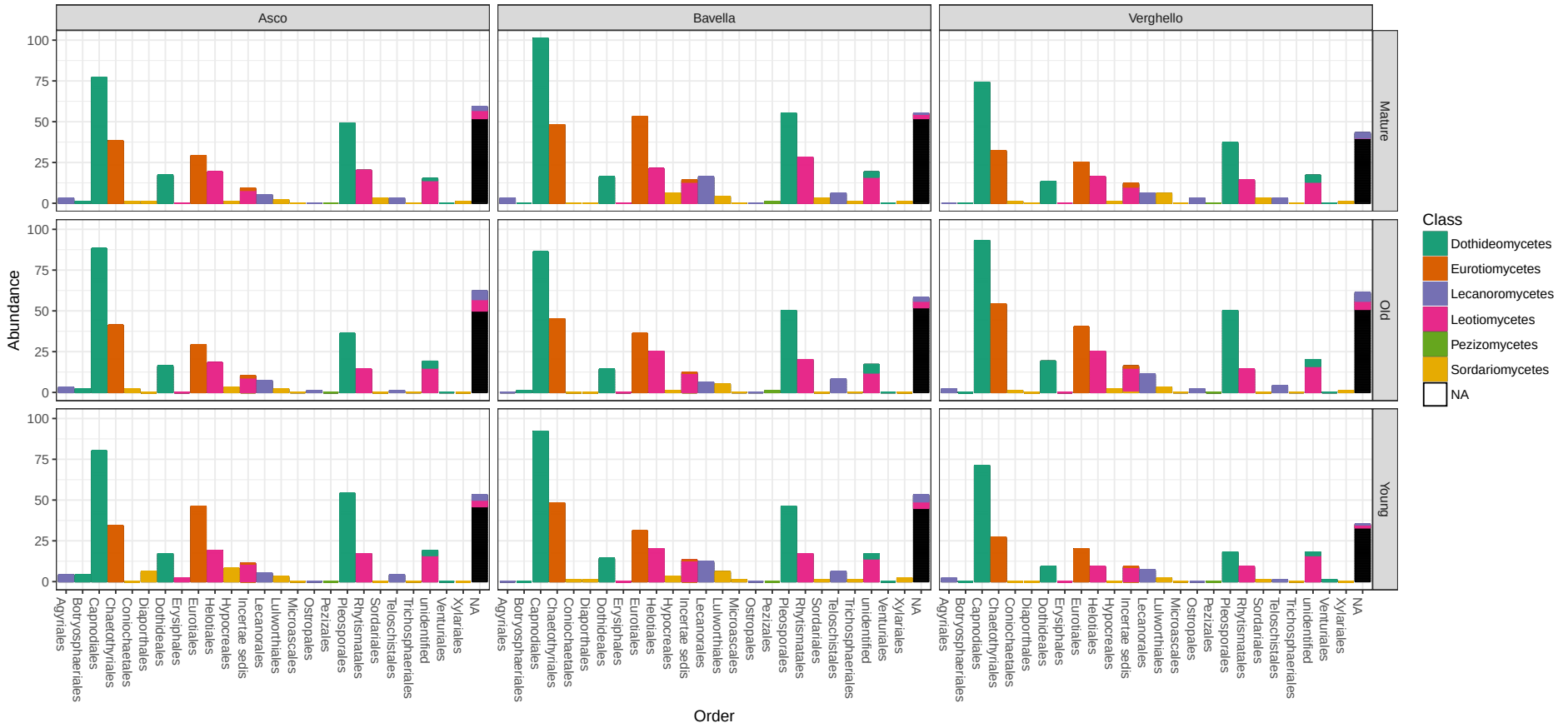


Figure 6.11: Taxonomic distribution of OTUs in the different site * age combinaison.

```
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")
```

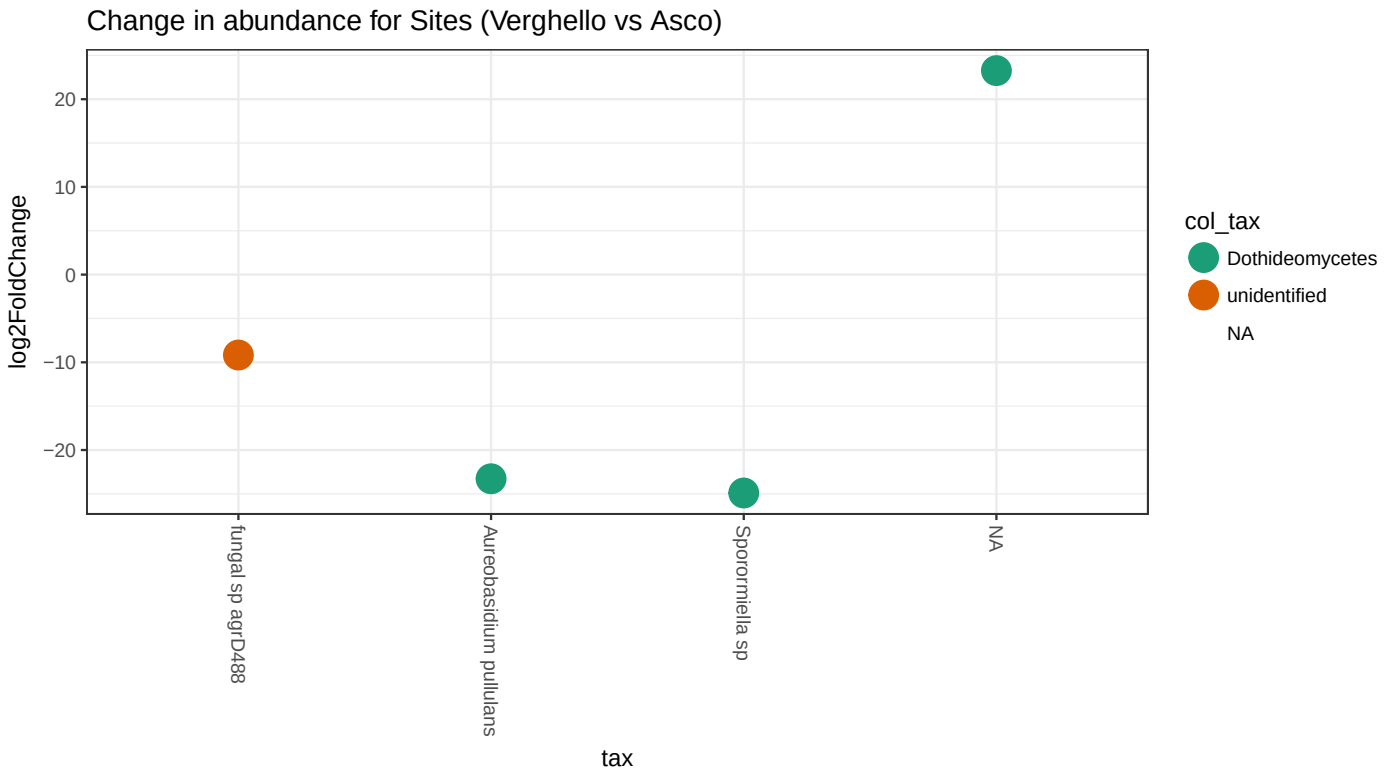


Figure 6.12: OTUs significantly different in terms of abundances between Verghello (positive values) and Asco (negative values)

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
```

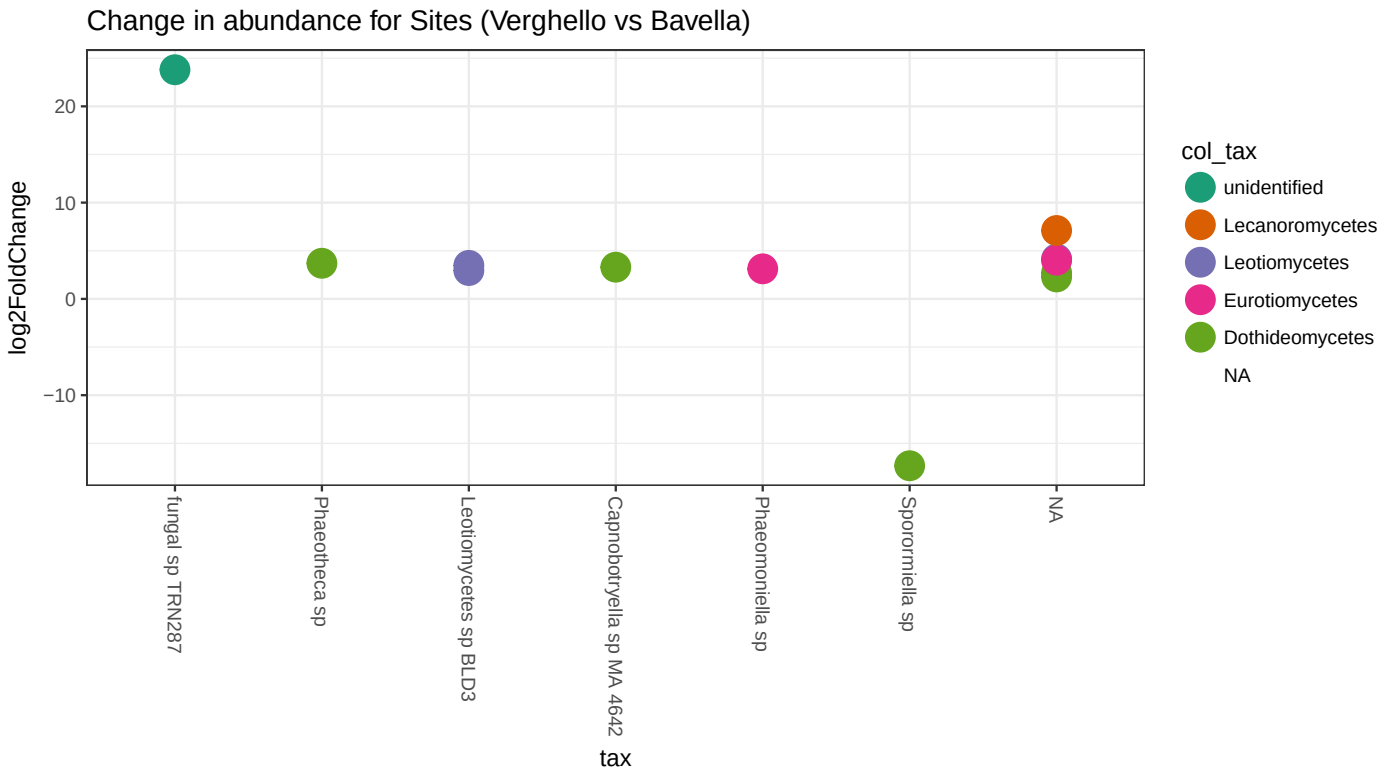


Figure 6.13: OTUs significantly different in terms of abundances between Verghello (positive values) and Bavella (negative values)

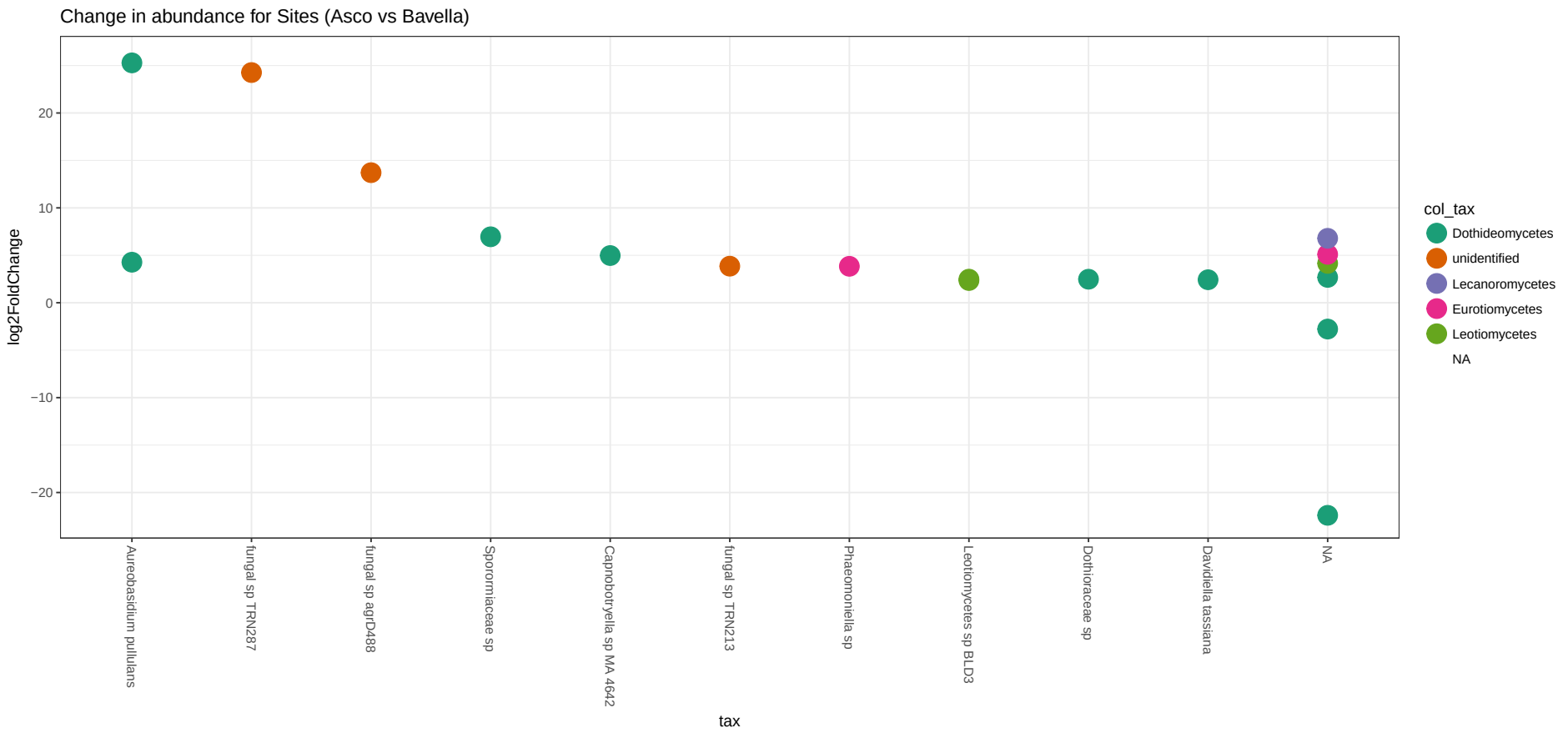


Figure 6.14: OTUs significantly different in terms of abundances between Asco (positive values) and Bavella (negative values)

```
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
```

	Comparison	OTU_names	Species	Class	log2FoldChange (negative = more on second lev
1	Verghello vs Asco	SH020264.07FU_KF800472_reps_singleton	Aureobasidium pullulans	Dothideomycetes	-23.2796299708717
2	Verghello vs Asco	SH018491.07FU_JN053173_reps_singleton	fungal sp agrD488	unidentified	-9.16956539526295
3	Verghello vs Asco	SH184189.07FU_FJ475668_reps	Sporormiella sp	Dothideomycetes	-24.9011554391882
4	Verghello vs Asco	SH205426.07FU_AY843155_reps			-24.5843452735373
5	Verghello vs Asco	SH202297.07FU_KF675366_reps		Dothideomycetes	23.2611842441584
6	Verghello vs Bavella	SH197740.07FU_EF419976_reps	Phaeomoniella sp	Eurotiomycetes	3.1234231651877
7	Verghello vs Bavella	SH214165.07FU_AY843074_reps		Dothideomycetes	2.28415036525221
8	Verghello vs Bavella	SH000664.07FU_FJ861984_reps_singleton		Leotiomycetes	4.1527936984384
9	Verghello vs Bavella	SH017706.07FU_AJ971406_reps_singleton	Capnobotryella sp MA 4642	Dothideomycetes	3.29734697730253
10	Verghello vs Bavella	SH006502.07FU_FN868467_reps_singleton	Leotiomycetes sp BLD3	Leotiomycetes	3.4828699802412
11	Verghello vs Bavella	SH004201.07FU_AM901923_reps_singleton			2.92901288737759
12	Verghello vs Bavella	SH189181.07FU_AM901716_reps	Phaeotheca sp	Dothideomycetes	3.70092841777665
13	Verghello vs Bavella	SH017712.07FU_KJ406777_reps_singleton		Dothideomycetes	2.70668773083002
14	Verghello vs Bavella	SH211751.07FU_KJ406871_reps		Eurotiomycetes	4.01529691348447
15	Verghello vs Bavella	SH017608.07FU_AY843125_reps_singleton	fungal sp TRN287	unidentified	23.8011581994588
16	Verghello vs Bavella	SH184038.07FU_FN435818_reps	Leotiomycetes sp BLD3	Leotiomycetes	2.94834147549805
17	Verghello vs Bavella	SH013575.07FU_HQ605938_reps_singleton		Lecanoromycetes	7.1024418420342
18	Verghello vs Bavella	SH184189.07FU_FJ475668_reps	Sporormiella sp	Dothideomycetes	-17.3266859306514
19	Asco vs Bavella	SH197740.07FU_EF419976_reps	Phaeomoniella sp	Eurotiomycetes	3.84143220242642
20	Asco vs Bavella	SH214165.07FU_AY843074_reps		Dothideomycetes	2.67345053947652
21	Asco vs Bavella	SH000664.07FU_FJ861984_reps_singleton		Leotiomycetes	4.14719888912293
22	Asco vs Bavella	SH200057.07FU_AY843076_reps			4.13735870298141
23	Asco vs Bavella	SH017706.07FU_AJ971406_reps_singleton	Capnobotryella sp MA 4642	Dothideomycetes	4.97577936963069
24	Asco vs Bavella	SH006502.07FU_FN868467_reps_singleton	Leotiomycetes sp BLD3	Leotiomycetes	2.48560370010328
25	Asco vs Bavella	SH004201.07FU_AM901923_reps_singleton			3.10438809472506
26	Asco vs Bavella	SH206392.07FU_AM921728_reps	Dothioraceae sp	Dothideomycetes	2.47894056064314
27	Asco vs Bavella	SH211751.07FU_KJ406871_reps		Eurotiomycetes	5.0996489147093
28	Asco vs Bavella	SH127907.07FU_DQ780388_refs	Davidiella tassiana	Dothideomycetes	2.42586179159098
29	Asco vs Bavella	SH023168.07FU_AY843079_reps_singleton	fungal sp TRN213	unidentified	3.85464235927135
30	Asco vs Bavella	SH017608.07FU_AY843125_reps_singleton	fungal sp TRN287	unidentified	24.2630375758631
31	Asco vs Bavella	SH184038.07FU_FN435818_reps	Leotiomycetes sp BLD3	Leotiomycetes	2.34473164793211
32	Asco vs Bavella	SH027337.07FU_EU707898_reps_singleton		Dothideomycetes	-2.77325759863912
33	Asco vs Bavella	SH020264.07FU_KF800472_reps_singleton	Aureobasidium pullulans	Dothideomycetes	25.2823841356578
34	Asco vs Bavella	SH195774.07FU_AJ244232_refs	Aureobasidium pullulans	Dothideomycetes	4.27262265532015
35	Asco vs Bavella	SH018491.07FU_JN053173_reps_singleton	fungal sp agrD488	unidentified	13.7101337513165
36	Asco vs Bavella	SH209326.07FU_JN053133_reps		Lecanoromycetes	6.79055875149354
37	Asco vs Bavella	SH184176.07FU_GU909656_reps	Sporormiaceae sp	Dothideomycetes	6.95375964063089
38	Asco vs Bavella	SH205426.07FU_AY843155_reps			25.6970707902463
39	Asco vs Bavella	SH202297.07FU_KF675366_reps		Dothideomycetes	-22.4062762093678

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

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")
```

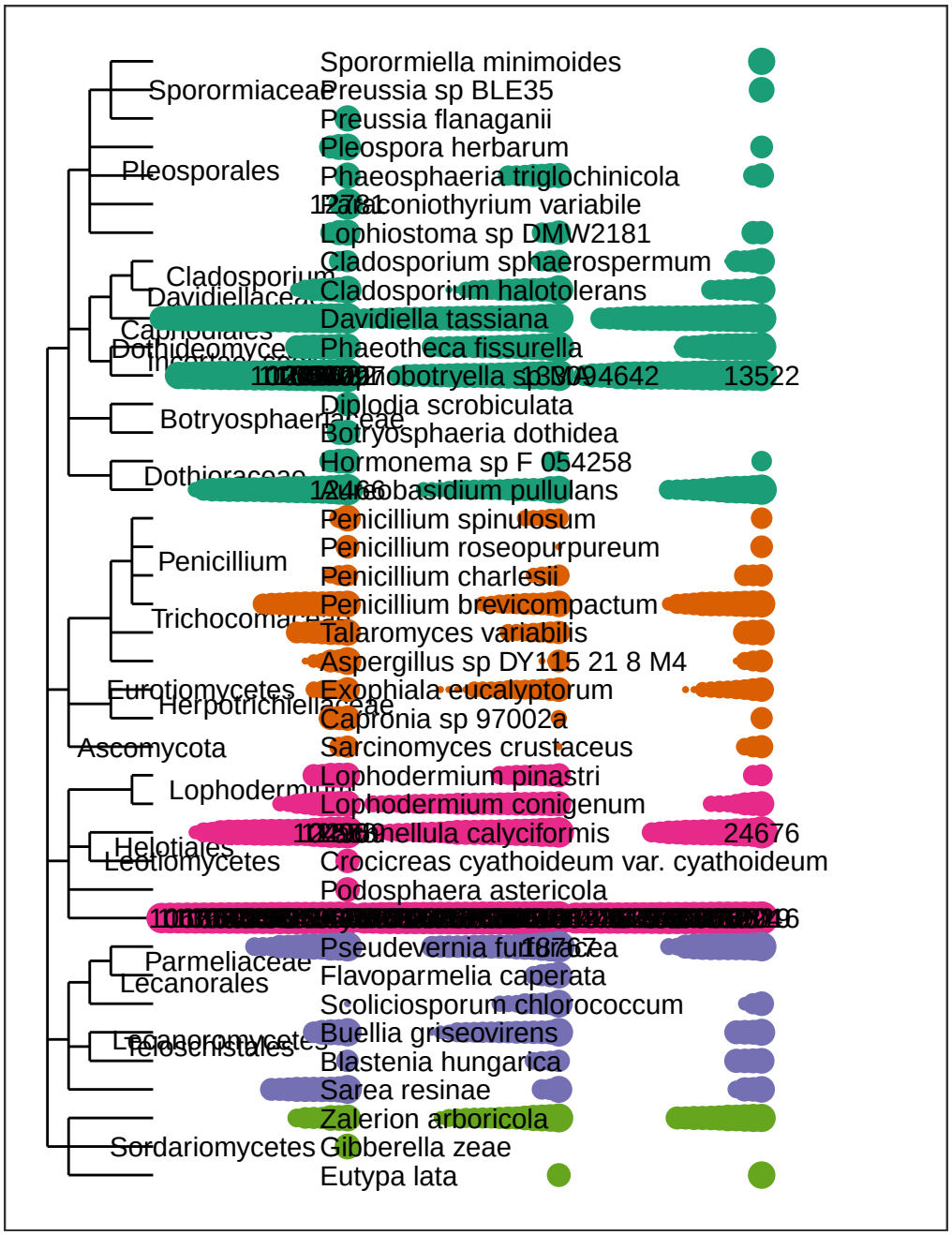
	Comparison	Order	Class	log2FoldChange (negative = more on second levels)
1	Verghello vs Asco	Botryosphaeriales	Dothideomycetes	-24.5249819855987
2	Verghello vs Asco	Ostropales	Lecanoromycetes	23.4855131491163
3	Verghello vs Bavella	Botryosphaeriales	Dothideomycetes	-16.6345809184226
4	Verghello vs Bavella	Dothideales	Dothideomycetes	2.01225332737348
5	Verghello vs Bavella	Eurotiales	Eurotiomycetes	1.68439599616236
6	Verghello vs Bavella	Incertae sedis	Leotiomycetes	-0.567752191873621
7	Verghello vs Bavella	Ostropales	Lecanoromycetes	29.0241120523582
8	Verghello vs Bavella	Rhytismatales	Leotiomycetes	3.19485840143062
9	Verghello vs Bavella	unidentified	Leotiomycetes	2.3727066682193
10	Asco vs Bavella	Agyriales	Lecanoromycetes	7.0973636413091
11	Asco vs Bavella	Capnodiales	Dothideomycetes	0.712165754006809
12	Asco vs Bavella	Chaetothyriales	Eurotiomycetes	1.81475008606722
13	Asco vs Bavella	Dothideales	Dothideomycetes	2.15424108078424
14	Asco vs Bavella	Eurotiales	Eurotiomycetes	2.36148954531956
15	Asco vs Bavella	Incertae sedis	Leotiomycetes	-0.822561360482083
16	Asco vs Bavella	Rhytismatales	Leotiomycetes	3.38436652075133
17	Asco vs Bavella	unidentified	Leotiomycetes	2.01203541569241

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

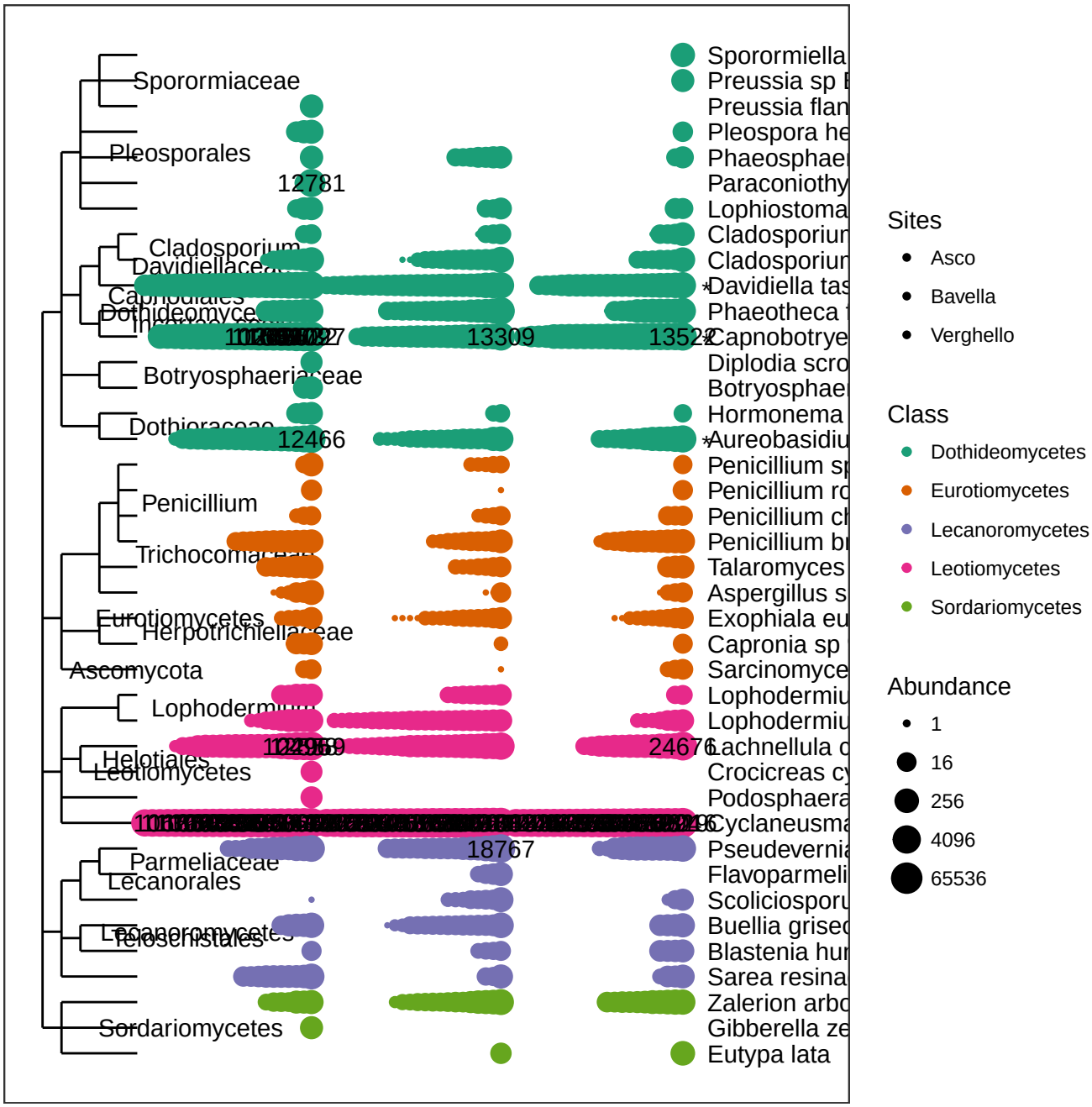
6.8 Distribution of OTUs abundance in the fungal phylogeny

```
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] 44

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

## [1] 46.05385
```




```
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 6.473782×10^6 sequences representing 256 OTUs allocated to 80 samples.

After filtering, the dataset includes 6.460489×10^6 sequences representing 233 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.

	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 15: Number of OTUs, samples and sequences after filtering

7.3 Beta diversity

Site (R2 = 0.126), age (R2 = 0.043) and interaction age*site (R2 = 0.102) statistically structured the fungal endophytic beta-diversity.

List of Figures

List of Tables