Appendix S9: results after UPARSE clustering allowing unique sequences

Adrien Taudiere*

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

July 19, 2016

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

*adrien.taudiere@cefe.cnrs.fr

Contents

1 Introduction **3**
- 1.1 R requirements . . . 3
- 1.2 System and session informations . . . 3
- 1.3 Some usefull functions . . . 5

2 Data **5**
- 2.1 Choice of filter parameters . . . 5
- 2.2 Load and convert loading . . . 5
  - 2.2.1 Otu table . . . 5
  - 2.2.2 Taxonomy . . . 6
  - 2.2.3 Add FUNguild information to taxonomy Table . . . 6
  - 2.2.4 Representative sequences . . . 7
  - 2.2.5 Samples information . . . 7
  - 2.2.6 Create the phyloseq object . . . 7
  - 2.2.7 Caracteristics of the phyloseq data . . . 7
- 2.3 Filter sample by number of sequences . . . 7
- 2.4 Filter OTUs by number of samples . . . 8
- 2.5 Filter OTUs by number of sequences . . . 9
- 2.6 Summary of filtration workflow . . . 10

3 Simple description of the dataset **11**
- 3.1 Number of sequences and OTUs by samples . . . 11
- 3.2 Number of sequences and samples for each OTUs . . . 11
- 3.3 Distribution of sequences in the taxonomy . . . 12
- 3.4 Focus on the 30 more abundant OTUs (number of sequences) . . . 14
- 3.5 Focus on the 30 more frequent OTUs (number of samples) . . . 16

4 Number of sequences and OTUs in function of putative ecology (using FUNGuild software; Nguyen et al, 2015) **19**

5 Distribution of fungal endophytic alpha-biodiversity **20**
- 5.1 Local diversity = Diversity by sites . . . 20
- 5.2 Diversity by age of tree . . . 20
- 5.3 Diversity by elevation of the sample . . . 21
- 5.4 Which factor affect diversity? . . . 21

6 Effect of site, age and elevation on fungal endophytic beta-diversity **26**
- 6.1 Venn diagramm . . . 26
- 6.2 Ordination . . . 26
- 6.3 Permanova on sites, host ages and elevation . . . 34
- 6.4 Permanova on sites, host ages and individual trees . . . 35
- 6.5 Differences in abundances and OTUs number by Order. . . . 36
- 6.6 Differences in abundances for each OTUs . . . 39
  - 6.6.1 Pairwise comparison of the OTUs composition by sites . . . 39
  - 6.6.2 Pairwise comparison of Order composition by sites . . . 42

1 Introduction

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

1.1 R requirements

First we need to install packages.

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

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

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

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

1.2 System and session informations

This document was created with R version 3.3.1 (2016-06-21) on Windows the 2016-07-19 16:59:06. See below for more information.

```
sessionInfo()

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

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

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

1.3 Some usefull functions

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

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

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

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

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

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

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

2 Data

2.1 Choice of filter parameters

```
#Choose the dataset folder
data_folder <- "Uparse_min1"

#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, "/UPARSE_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, "/UPARSE.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:         [ 1667 taxa and 80 samples ]
## sample_data() Sample Data:       [ 80 samples by 6 sample variables ]
## tax_table()   Taxonomy Table:    [ 1667 taxa by 12 taxonomic ranks ]
## refseq()      DNAStringSet:      [ 1667 reference sequences ]
```

The data are made of 8.335341×10^6 sequences representing 1667 OTUs allocate to 80 samples.

2.3 Filter sample by number of sequences

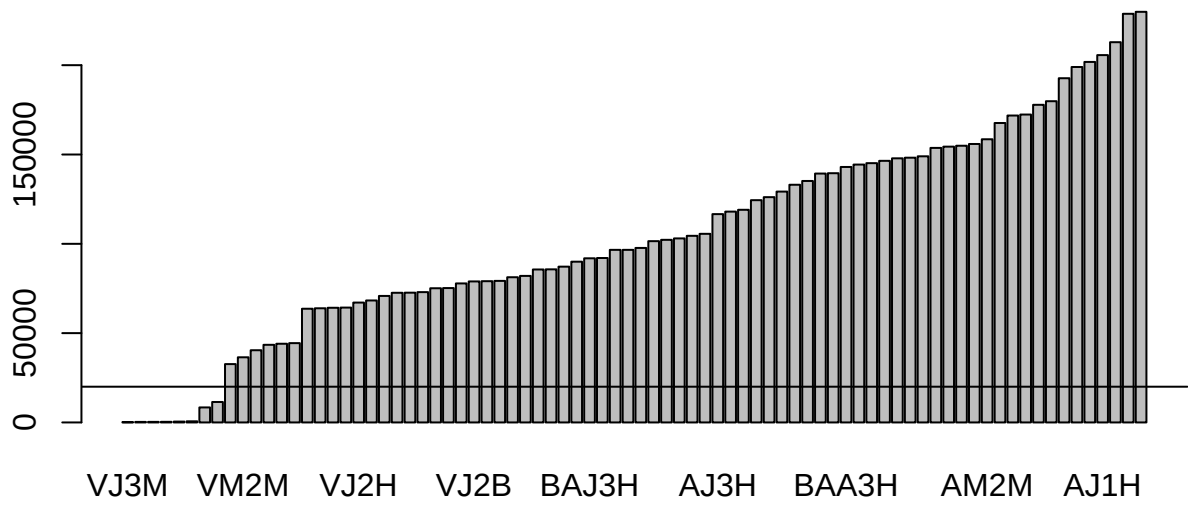

Figure 2.1: Number of sequences by sample

```
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 OTU present in a given number of samples (Figure 2.2).

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

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

summary(df_nbOtu_sample$Nb.samples)
```

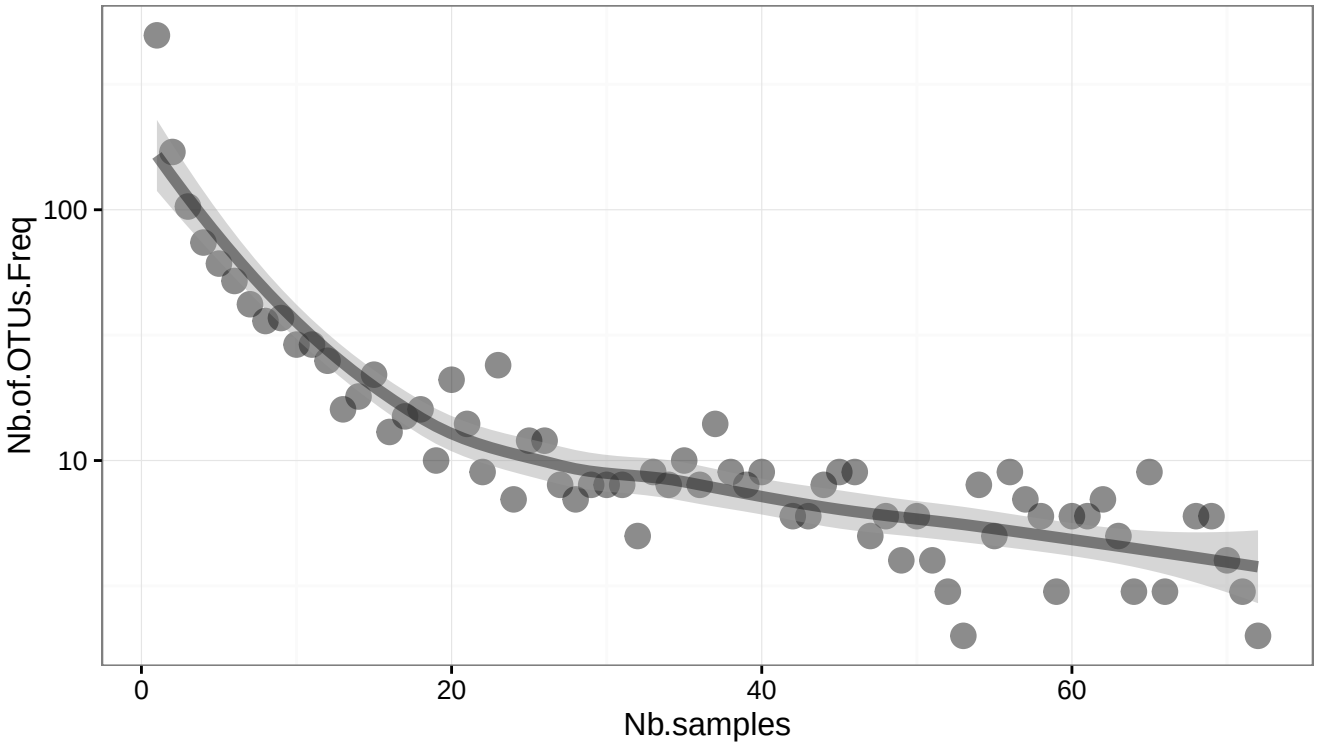

Figure 2.2: Number of OTU present in a given number of samples. Vertical bar illustrate the filtering parameter.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00   18.25   35.50   36.00   53.75   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 1650 on the 1650 OTUs (100%).

2.5 Filter OTUs by number of sequences

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

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

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

summary(df_nbseq_Otu[, 1])
```

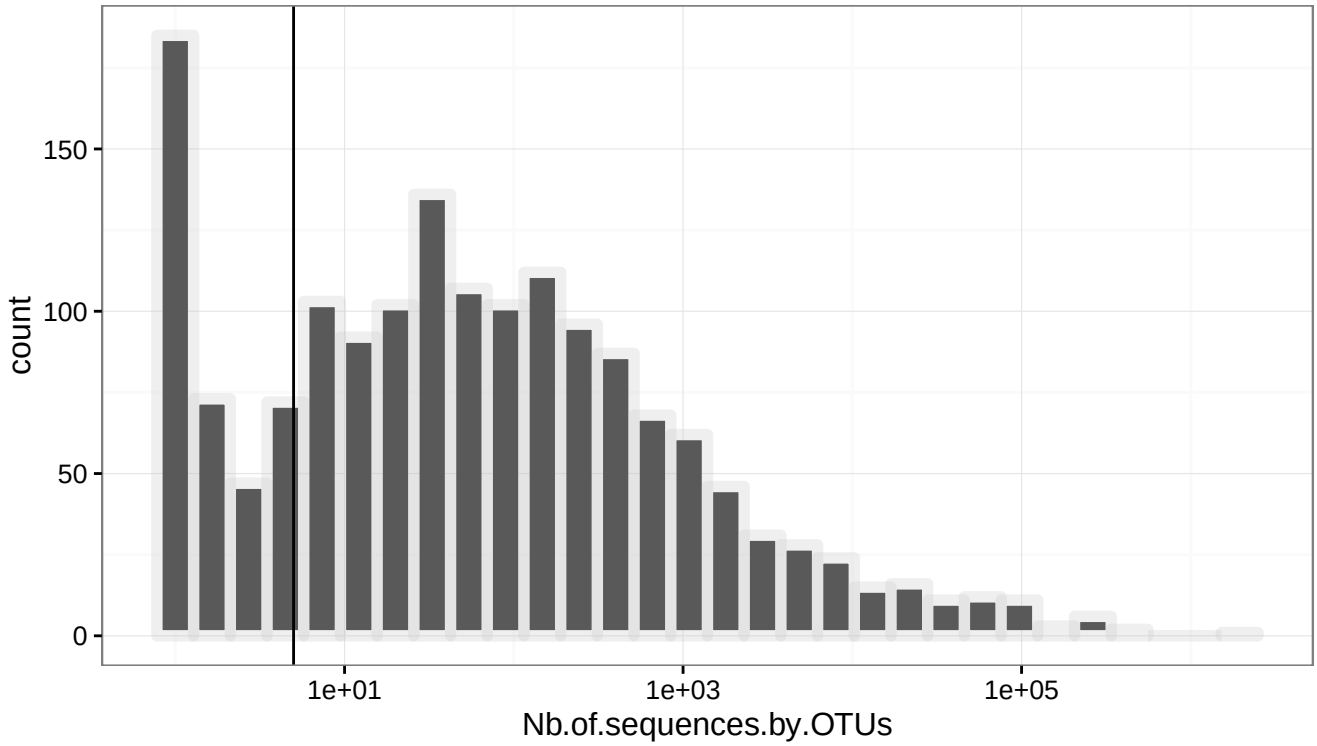

Figure 2.3: Number of sequences by OTU (log10 transformed). Horizontal bar illustrate the filtering parameter.

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##       1.0       7.0      46.0    5038.0     337.8 1943000.0
```

```
N_seq_otu_min

## [1] 5
```

If we discard OTUs with less than 1 sequences, we keep 1302 on the 1667 OTUs (78.1%).

```
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 8.312594×10^6 sequences representing 1302 OTUs allocate to 72 samples.

	Nb.of.OTUs	Nb.of.samples	Nb.of.sequences
No filter	1667	80	8335341.00
Nb of sequences by sample >= 20000	1650	72	8313238.00
Nb of sample by OTUs >= 1	1650	72	8313238.00
Nb of sequences by OTUs >= 5	1302	72	8312594.00

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

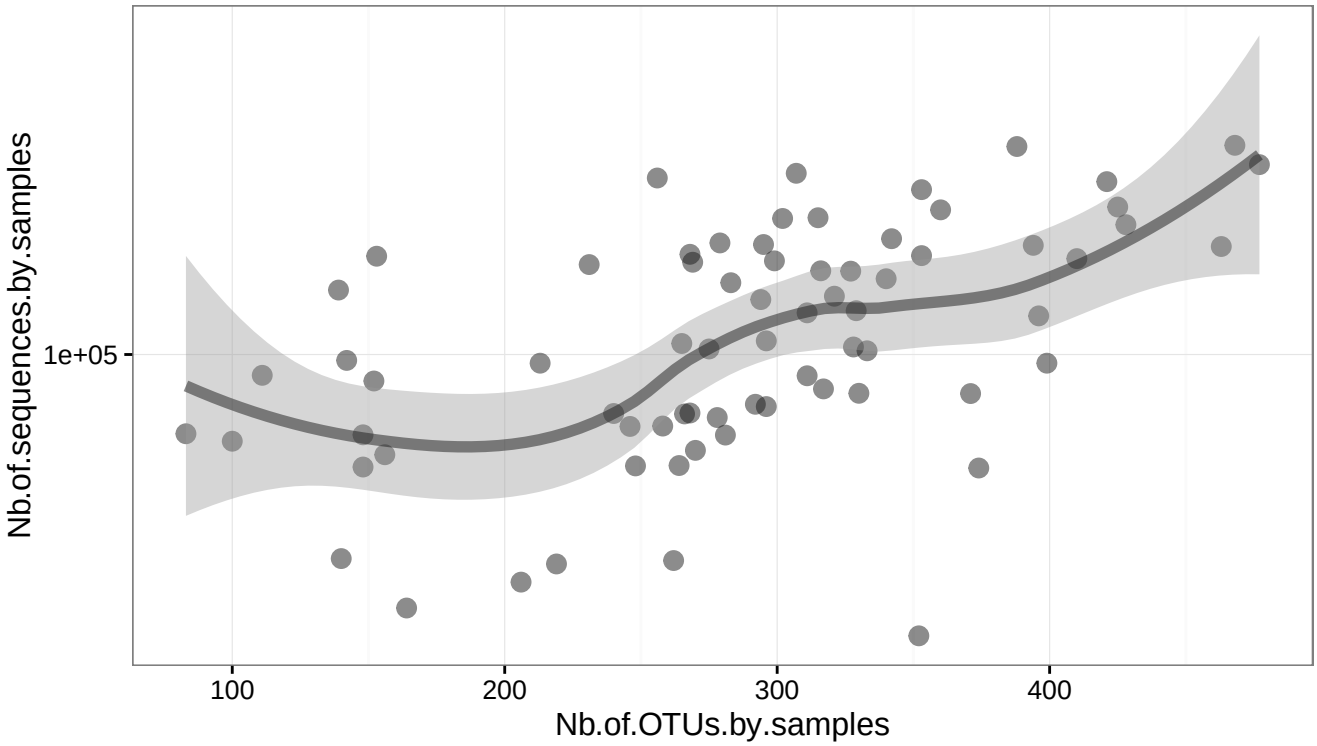

Figure 3.1: Number of OTUs by samples in fonction the number of sequences by samples (log10 axe). The tendency is represented by the line obtain from loess (Local Polynomial Regression Fitting).

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

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

3.2 Number of sequences and samples for each OTUs

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

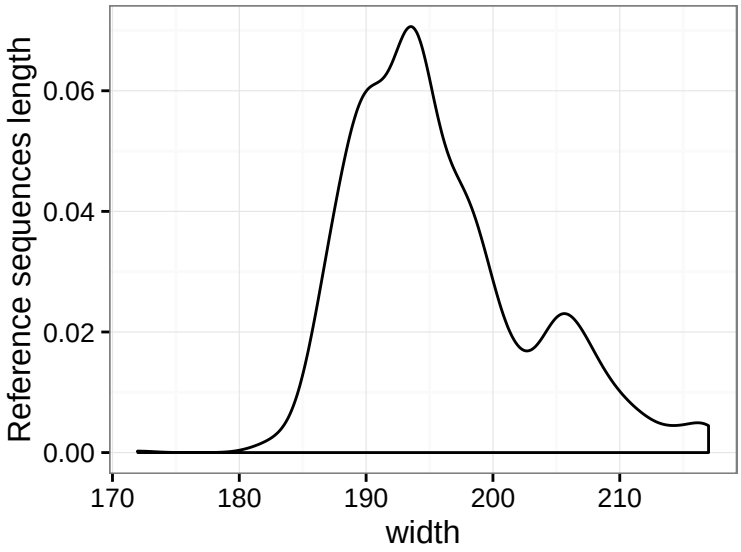

Figure 3.2: Distribution of reference sequences length.

```
                                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"))
```

3.3 Distribution of sequences in the taxonomy

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

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

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

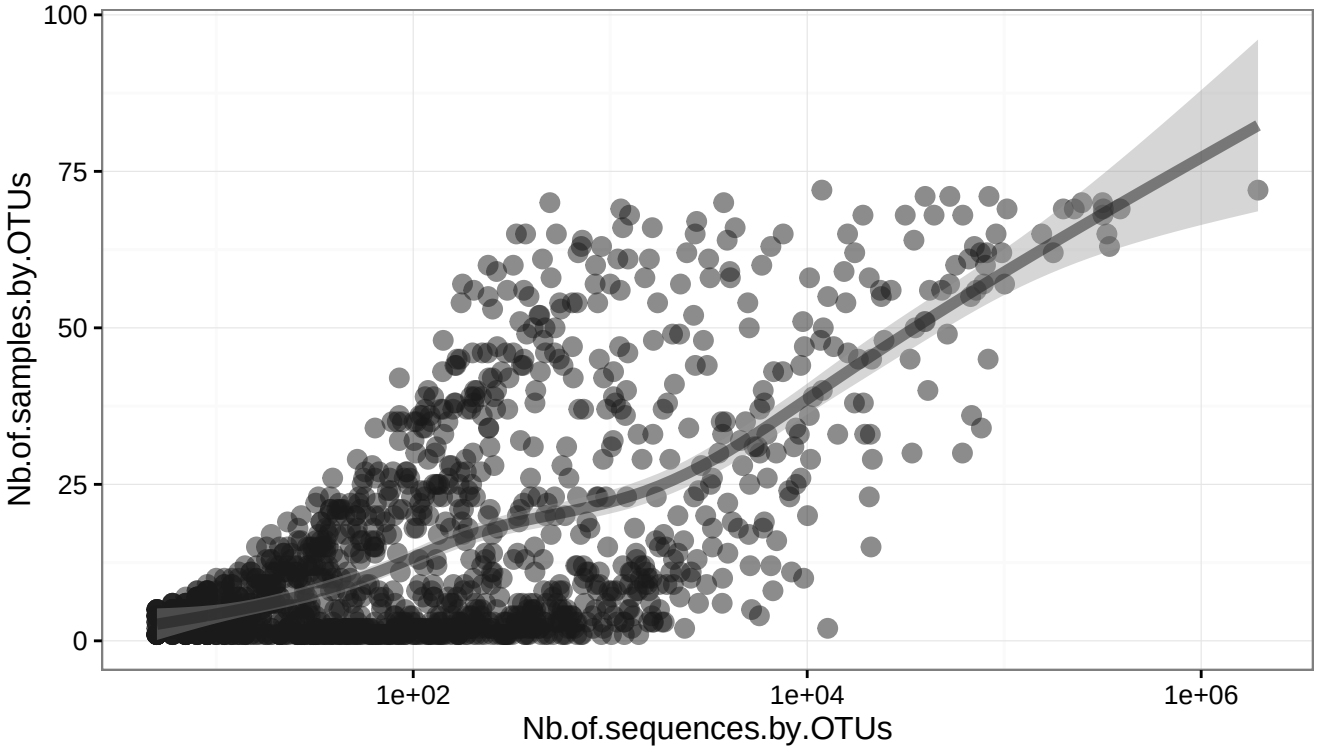

Figure 3.3: Number of sequences by OTUs (log10 axe) in fonction the number of samples where OTUs were found. The tendency is represented by the line obtain from gam (Generalized additive models with integrated smoothness estimation).

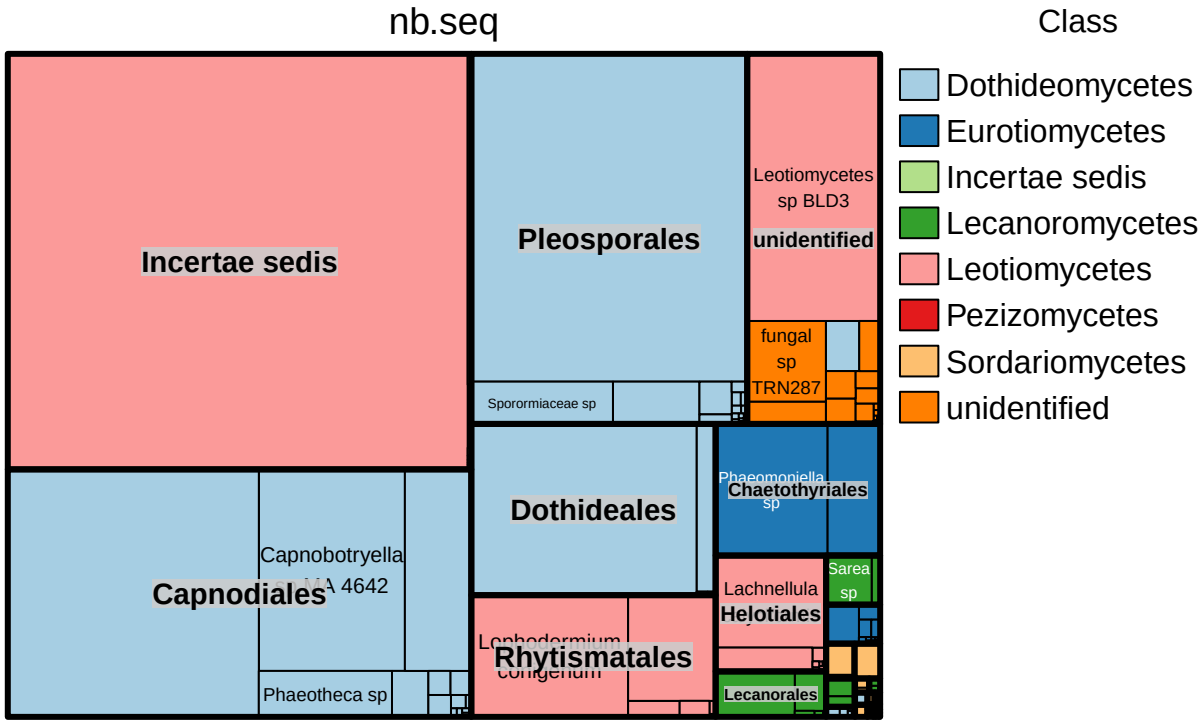

Figure 3.4: Distribution of the number of sequences in the taxonomy. Color represent Class, bold lines delimit Order and thick line delimit species.

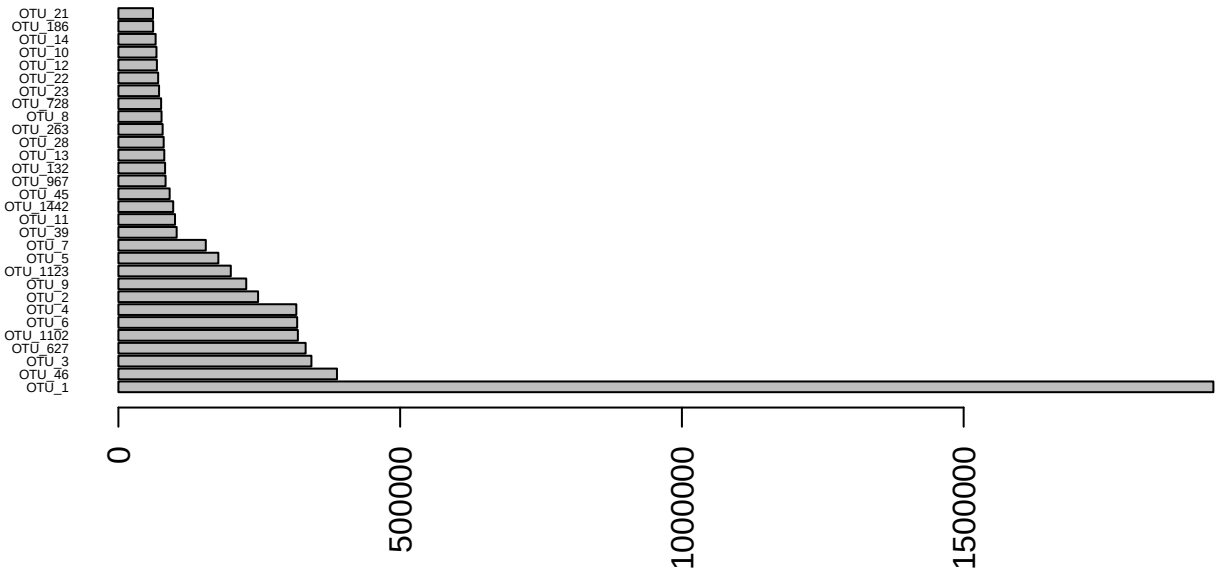

Figure 3.5: Number of sequences of the 30 more abundant OTUs (number of sequences).

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
             frequent OTUs (number of sequences)"),
      size = "\\tiny", include.rownames = FALSE)
```

Phylum	Class	Order	Family	Genus	Species	Trophic_Mode	Guild	Nb.sequences
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	1943171
Ascomycota	Dothideomycetes	Pleosporales				-	-	387885
Ascomycota	Dothideomycetes	Capnodiales	Incertae sedis	Capnobotryella	Capnobotryella sp MA 4642	Saprotroph	Undefined Saprotroph	342517
Ascomycota	Dothideomycetes	Pleosporales	Incertae sedis			-	-	332263
Ascomycota	Leotiomycetes	unidentified	unidentified	unidentified	Leotiomycetes sp BLD3	-	-	318495
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp	-	-	317246
Ascomycota	Dothideomycetes	Capnodiales				-	-	315723
						-	-	247896
Ascomycota	Leotiomycetes	Rhytismatales	Rhytismataceae	Lophodermium	Lophodermium conigenum	Pathotroph	Plant Pathogen	226872
Ascomycota	Dothideomycetes	Pleosporales				-	-	199465
						-	-	177383
Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Phaeothecoidea	Phaeothecoidea sp	Saprotroph	Undefined Saprotroph	155089
Ascomycota	Leotiomycetes	unidentified	unidentified	unidentified	Leotiomycetes sp BLD3	-	-	103401
						-	-	100459
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp	-	-	97232
Ascomycota	Dothideomycetes	Pleosporales				-	-	90937
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	83720
						-	-	82876
Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Phaeomoniella	Phaeomoniella sp	Saprotroph	Undefined Saprotroph	81451
Ascomycota	Leotiomycetes	Rhytismatales	Rhytismataceae	Lophodermium		Pathotroph	Plant Pathogen	80357
						-	-	78492
Ascomycota						-	-	76573
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	75793
unidentified	unidentified	unidentified	unidentified	unidentified	fungal sp TRN287	-	-	72163
Ascomycota	Dothideomycetes	Capnodiales				-	-	70557
Ascomycota						-	-	68326
Ascomycota	Leotiomycetes	Helotiales	Hyaloscyphaceae	Lachnellula	Lachnellula calyciformis	Saprotroph	Undefined Saprotroph	67576
Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Phaeomoniella	Phaeomoniella sp	Saprotroph	Undefined Saprotroph	66072
Ascomycota	Dothideomycetes	Capnodiales	Davidiellaceae	Cladosporium		-	-	61649

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

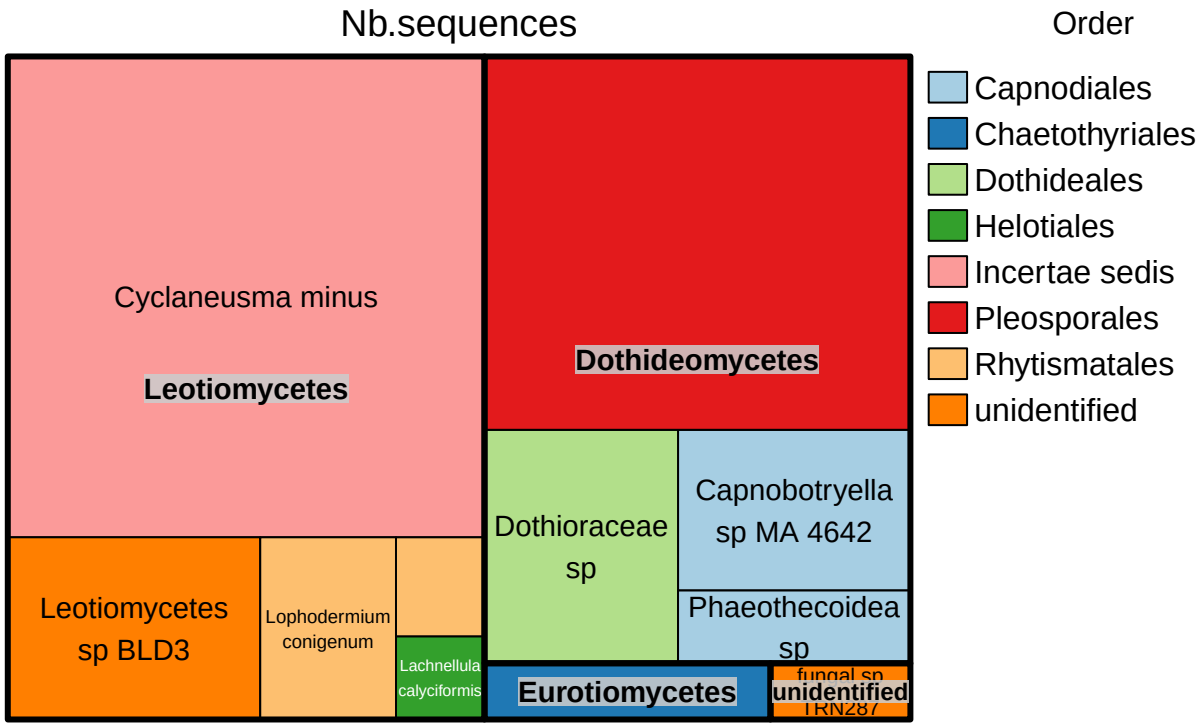

Figure 3.6: Number of sequences of the 30 most abundant OTUs (number of sequences). Colors indicate Order, bold lines delimit Class and thick lines delimit species.

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

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

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

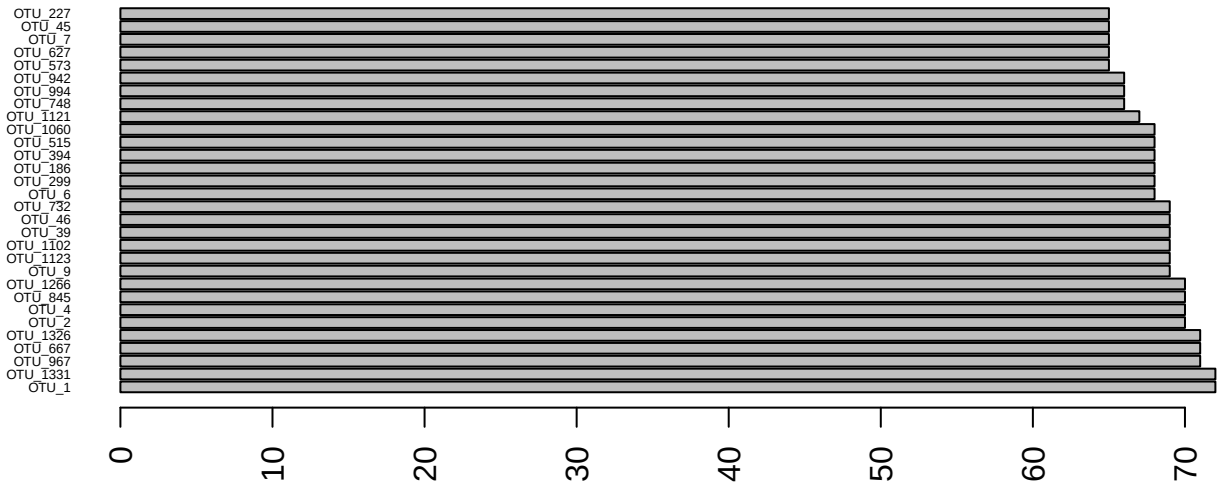

Figure 3.7: Number of samples of the 30 more frequent OTUs (number of samples).

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

Phylum	Class	Order	Family	Genus	Species	Trophic_Mode	Guild	Nb.samples
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	72
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	72
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	71
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	71
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	71
						-	-	70
Ascomycota	Dothideomycetes	Capnodiales				-	-	70
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	70
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	70
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	Leotiomycetes	unidentified	unidentified	unidentified	Leotiomycetes sp BLD3	-	-	69
Ascomycota	Dothideomycetes	Pleosporales				-	-	69
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	69
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp	-	-	68
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	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	68
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	68
Ascomycota	Leotiomycetes	unidentified	unidentified	unidentified	Leotiomycetes sp BLD3	-	-	67
Ascomycota	Dothideomycetes	Capnodiales				-	-	66
						-	-	66
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	66

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

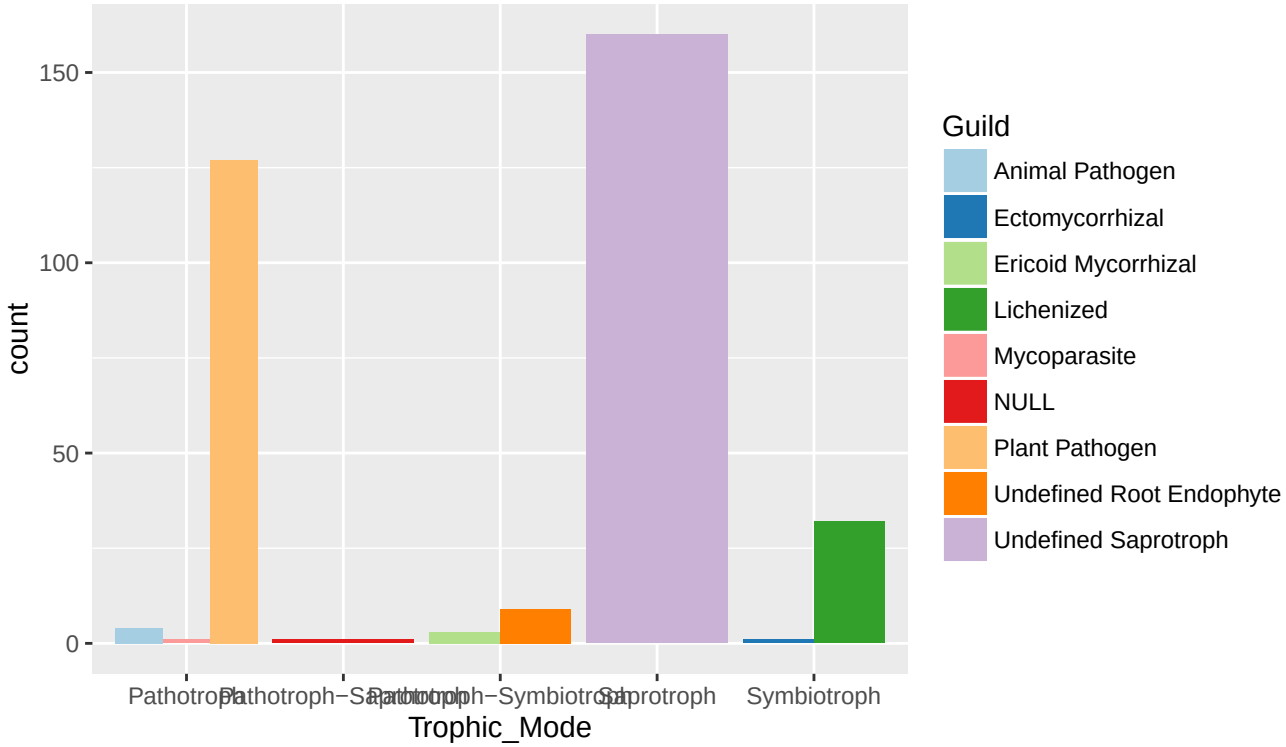

Figure 4.1: Distribution of OTUs into functional Guild.

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] 82.18797

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

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

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

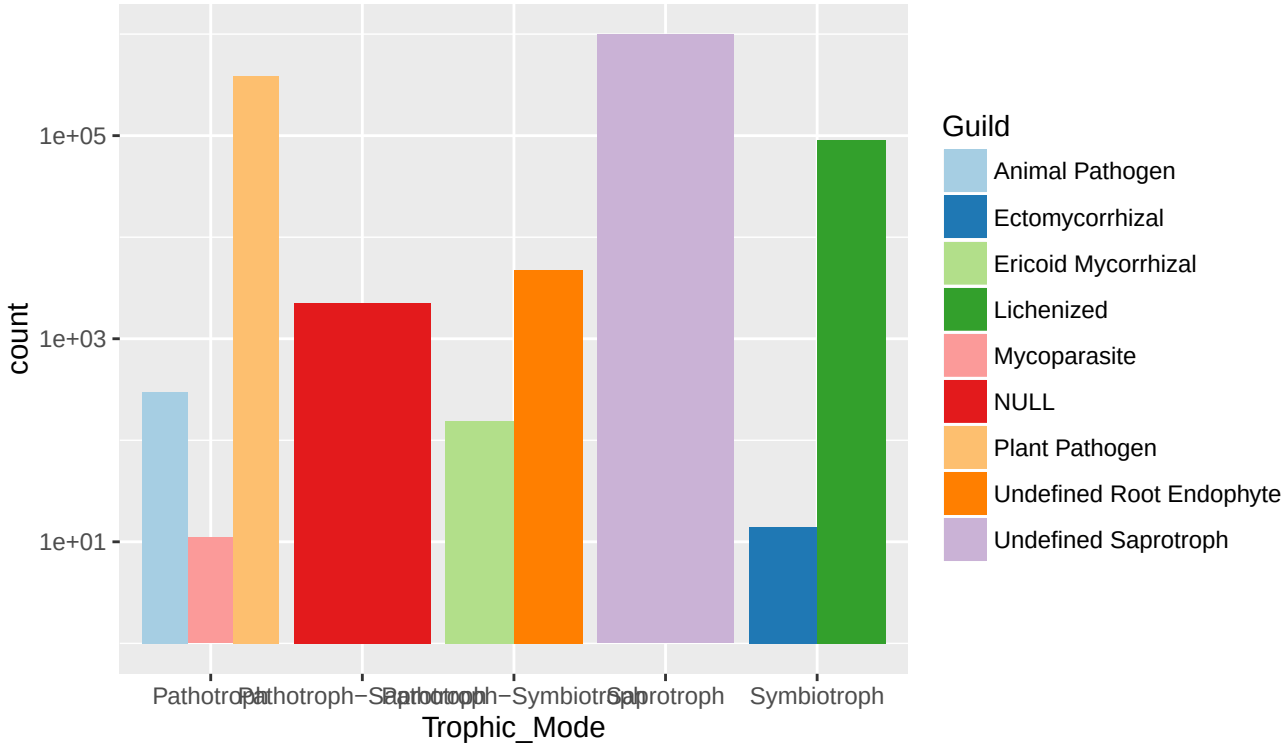

Figure 4.2: Distribution of sequences (log10 transformed) into functional Guild.

5 Distribution of fungal endophytic alpha-biodiversity

5.1 Local diversity = Diversity by sites

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

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

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

5.2 Diversity by age of tree

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

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

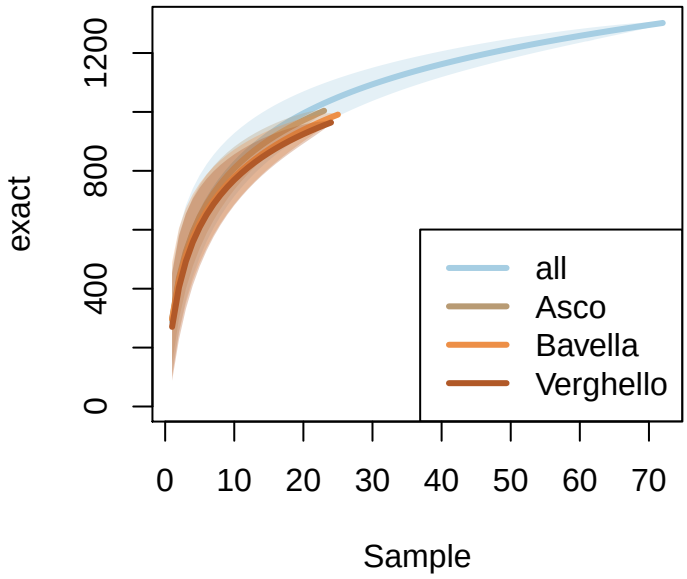

Figure 5.1: Rarefaction curves for each sites. Notes that if singletons were removed, these curves are biaised.

5.3 Diversity by elevation of the sample

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

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

5.4 Which factor affect diversity?

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

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

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

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

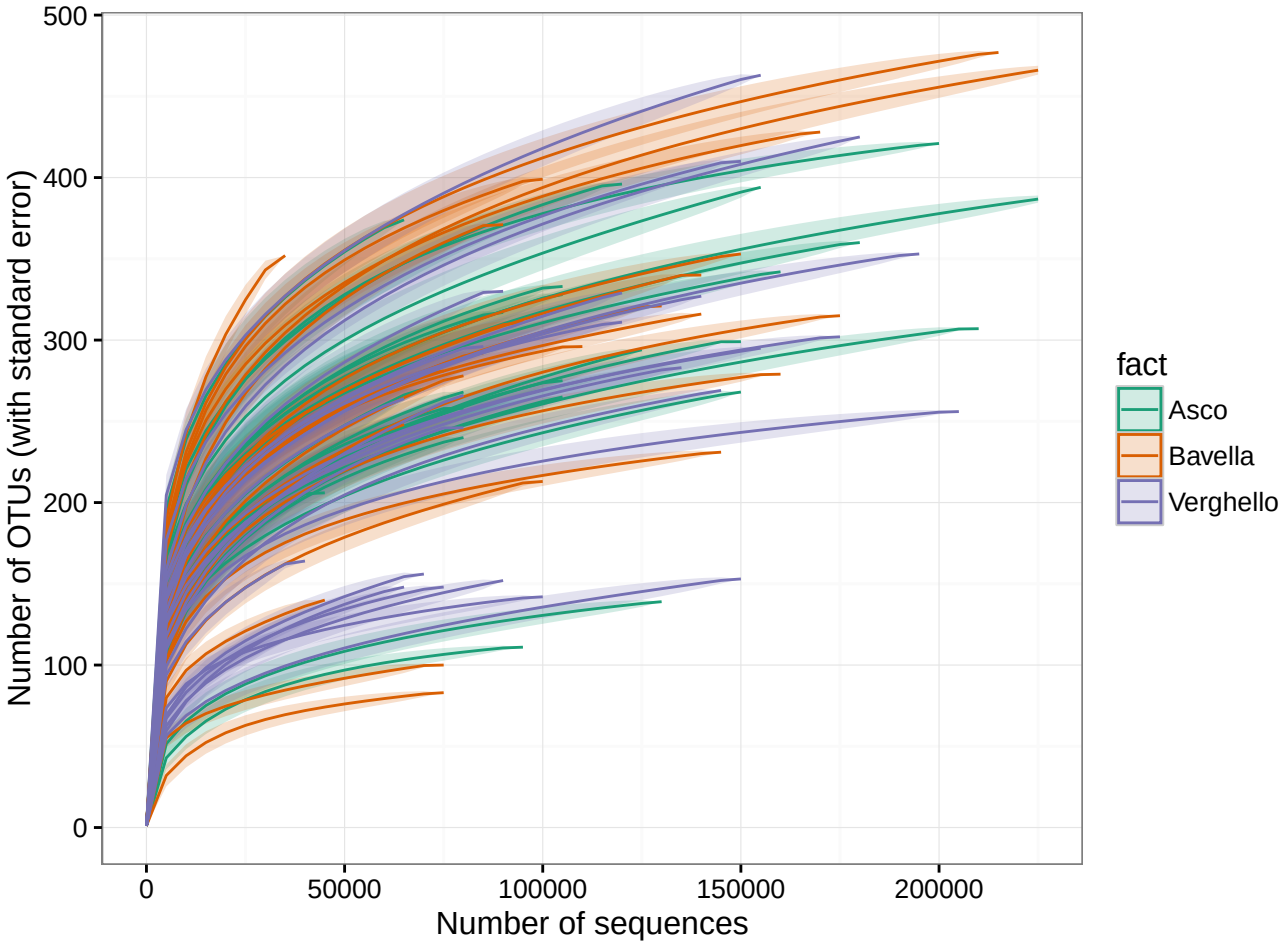

Figure 5.2: Rarefaction curves for each samples using sequences number on x-axes. Notes that if singletons were removed, these curves are biaised.

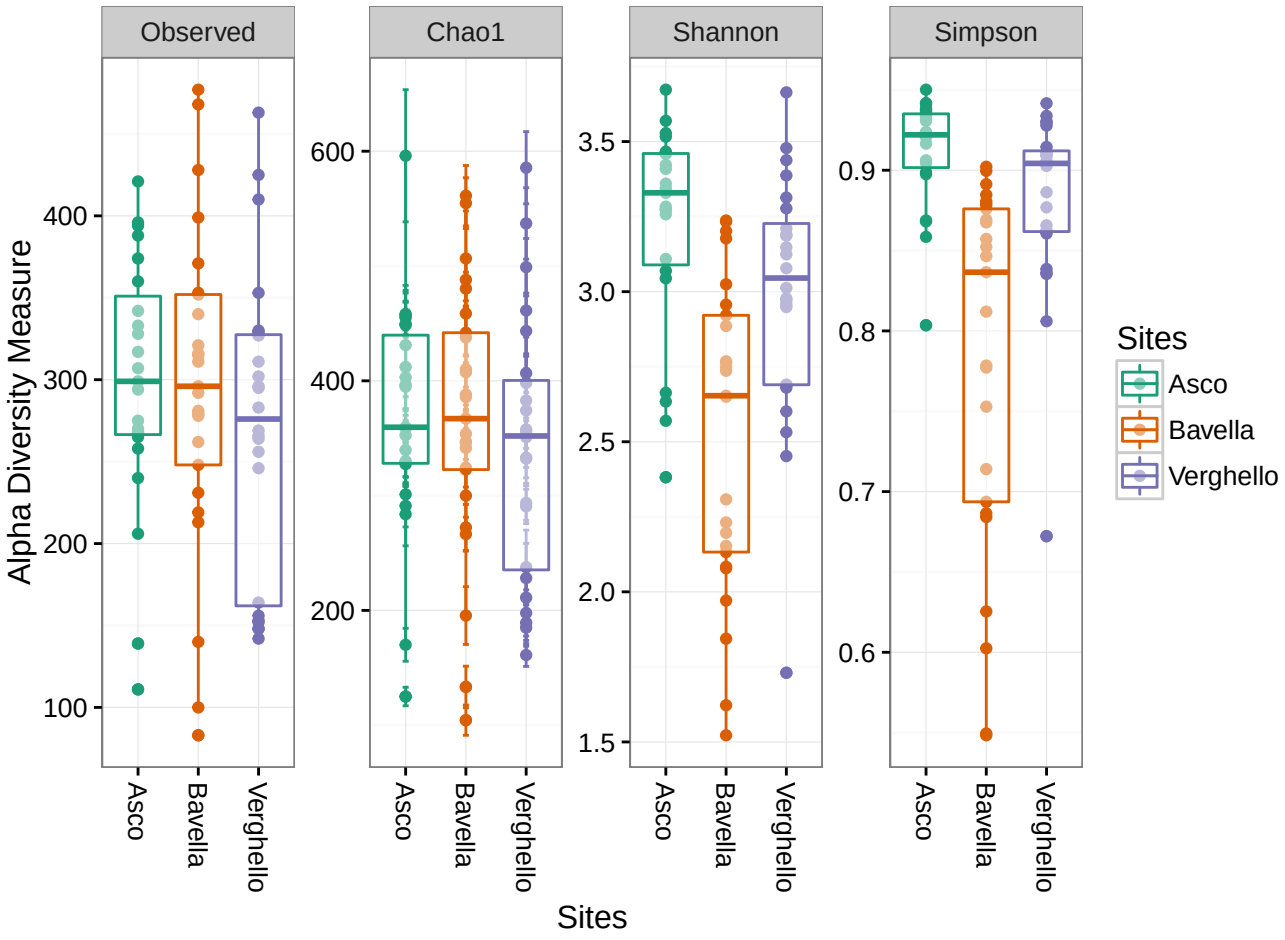

Figure 5.3: Diversity of each sites

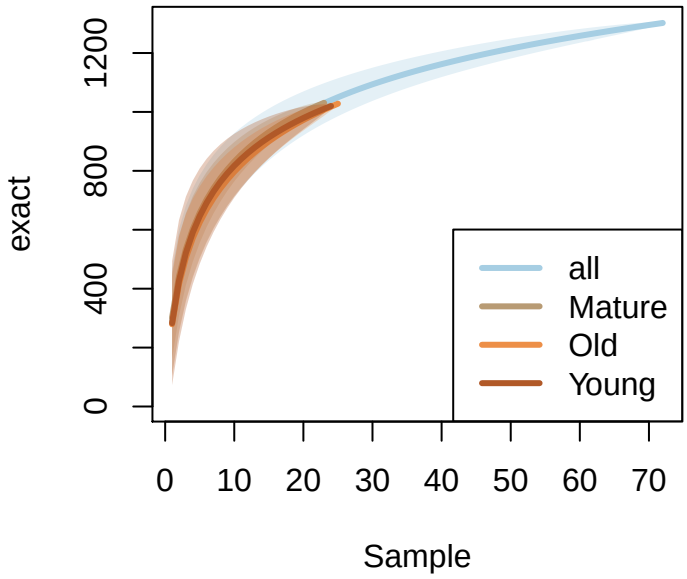

Figure 5.4: Rarefaction curves for each tree age modalities. Notes that if singletons were removed, these curves are biaised.

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

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	62.8502239	49.5872070	1.2674685	0.2095802
sqrt(readNumbers)	0.7288918	0.1267135	5.7522816	0.0000003
data.f3@sam_data$SitesBavella	9.9500272	22.0573697	0.4510976	0.6534440
data.f3@sam_data$SitesVerghello	-24.5733747	22.0808586	-1.1128813	0.2699220
data.f3@sam_data$AgeOld	-13.7853535	21.9625265	-0.6276761	0.5324494
data.f3@sam_data$AgeYoung	-37.0937028	22.4841290	-1.6497727	0.1038895
data.f3@sam_data$ElevationMiddle	20.1789158	22.2394735	0.9073468	0.3676272
data.f3@sam_data$ElevationTop	-5.5313782	21.9385577	-0.2521304	0.8017482

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

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

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

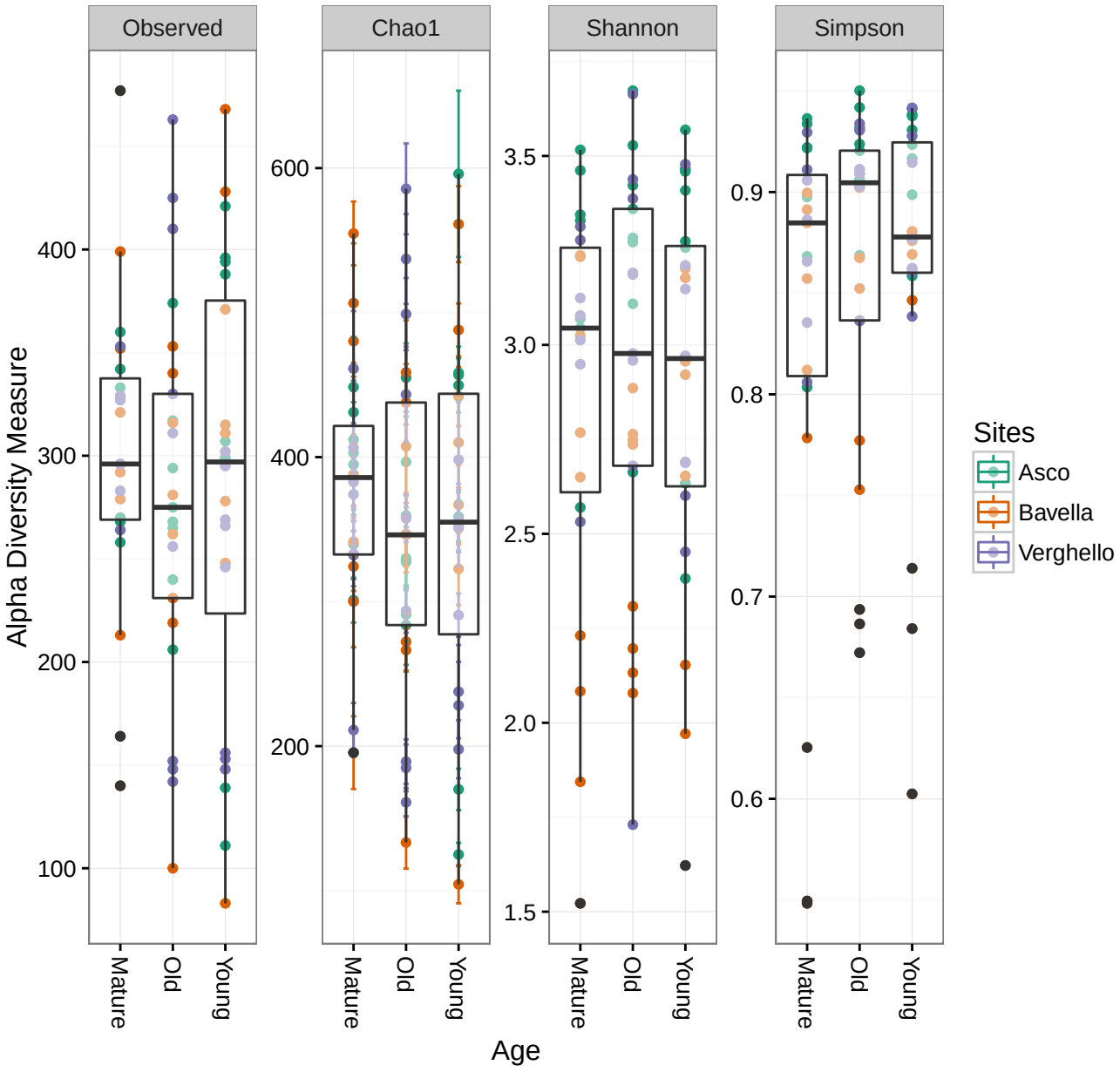

Figure 5.5: Diversity in function of tree age. Color represent sites.

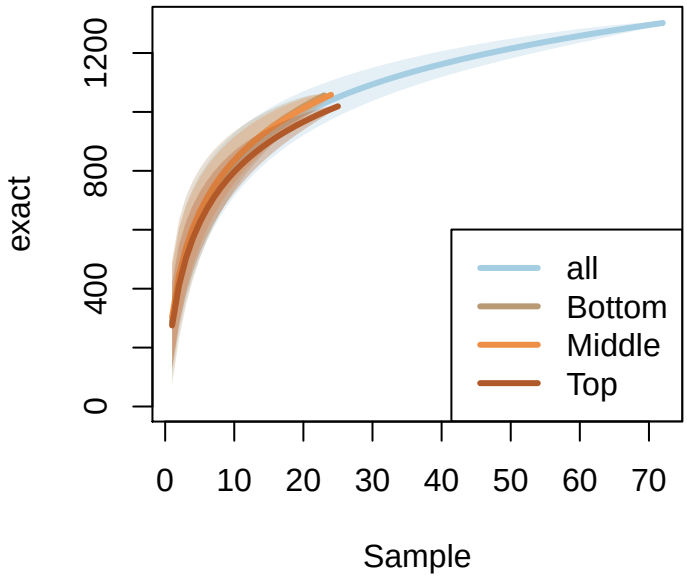

Figure 5.6: Rarefaction curves for each elevation. Notes that if singletons were removed, these curves are biaised.

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

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

6.1 Venn diagramm

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

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

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

6.2 Ordination

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

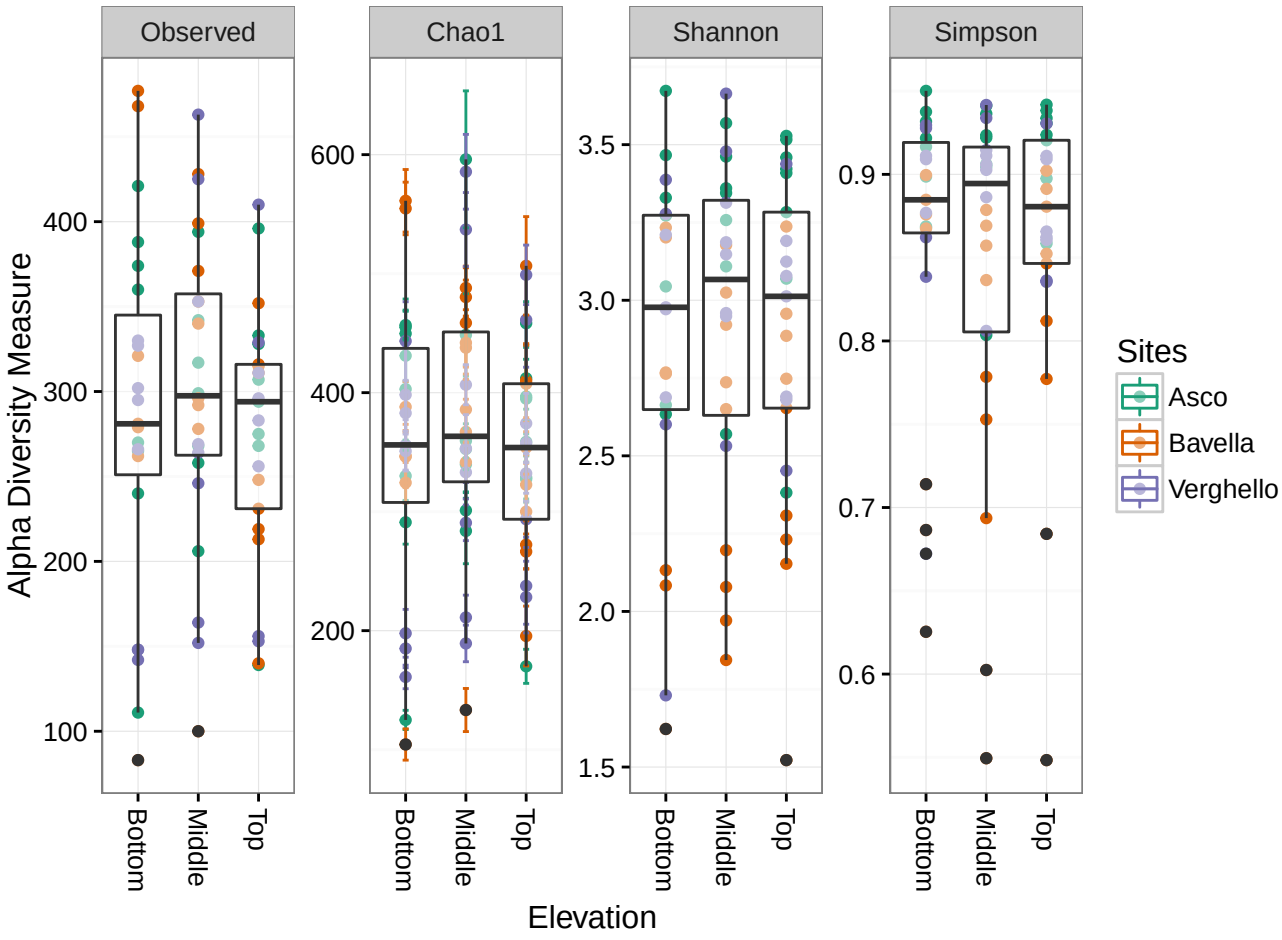

Figure 5.7: Diversity in function of elevation. Color represent sites.

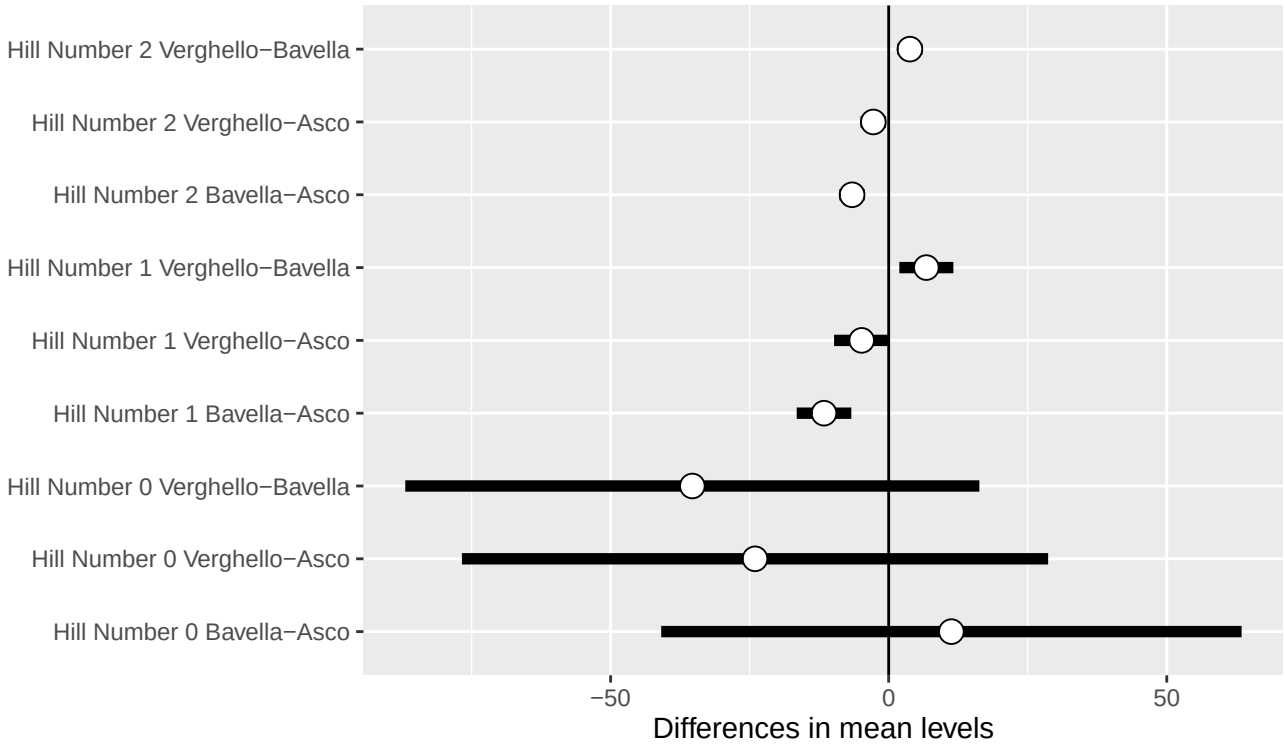

Figure 5.8: Results of the Tuckey HSD testing for differences in mean Hill numbers among pairs of modalities

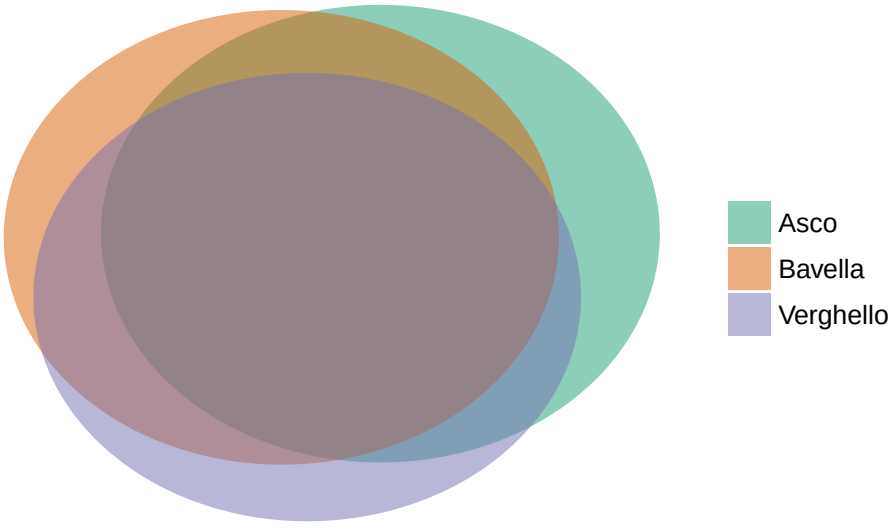

Figure 6.1: Venn diagramm ef the distribution of OTUs among Sites

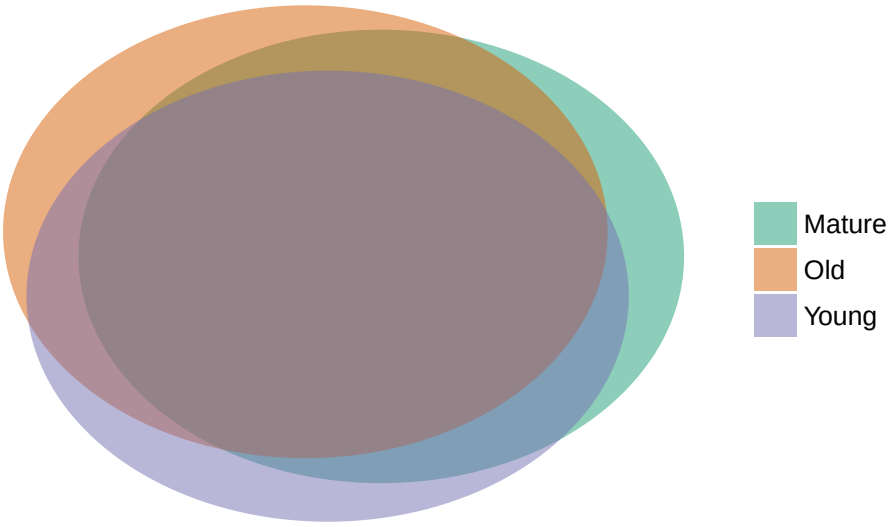

Figure 6.2: Venn diagramm ef the distribution of OTUs among host age

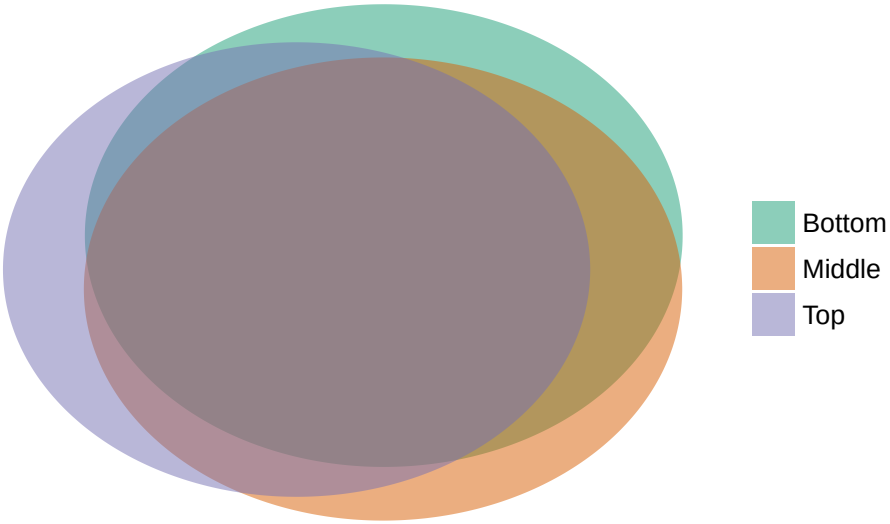

Figure 6.3: Venn diagramm ef the distribution of OTUs among elevation of samples

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	15.9643217	4.7426774	3.3660990	0.0012932
sqrt(readNumbers)	0.0273900	0.0121193	2.2600329	0.0272300
data.f3@sam_data$SitesBavella	-11.8649769	2.1096366	-5.6241804	0.0000004
data.f3@sam_data$SitesVerghello	-4.9746545	2.1118832	-2.3555539	0.0215712
data.f3@sam_data$AgeOld	1.0906065	2.1005655	0.5191966	0.6054145
data.f3@sam_data$AgeYoung	-0.9054816	2.1504532	-0.4210655	0.6751180
data.f3@sam_data$ElevationMiddle	2.1257168	2.1270536	0.9993715	0.3213784
data.f3@sam_data$ElevationTop	1.0378518	2.0982731	0.4946219	0.6225607

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

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.3970926	2.2046679	4.2623619	0.0000678
sqrt(readNumbers)	0.0071962	0.0056337	1.2773406	0.2060965
data.f3@sam_data$SitesBavella	-6.6403487	0.9806799	-6.7711687	0.0000000
data.f3@sam_data$SitesVerghello	-2.8126265	0.9817242	-2.8649865	0.0056353
data.f3@sam_data$AgeOld	1.1277239	0.9764631	1.1549068	0.2524225
data.f3@sam_data$AgeYoung	0.7541150	0.9996537	0.7543762	0.4533902
data.f3@sam_data$ElevationMiddle	0.2513083	0.9887763	0.2541610	0.8001860
data.f3@sam_data$ElevationTop	0.1848083	0.9753974	0.1894698	0.8503243

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

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

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

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

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

## Square root transformation
## Wisconsin double standardization
## Run 0 stress 0.2393401
## Run 1 stress 0.2410864
## Run 2 stress 0.2419549
## Run 3 stress 0.2407944
## Run 4 stress 0.2474657
## Run 5 stress 0.2389204
## ... New best solution
```

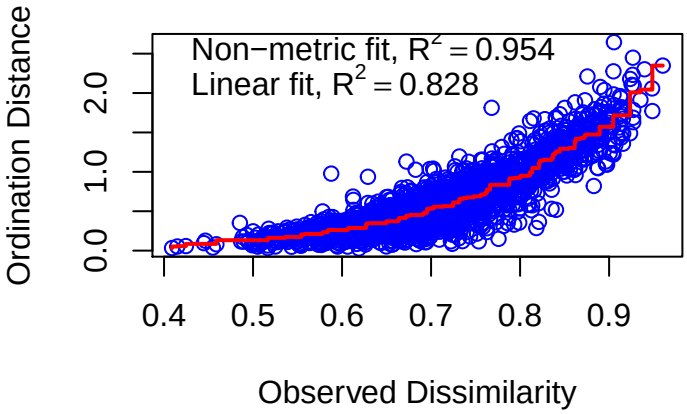

Figure 6.4: Stress plot of the NMDS

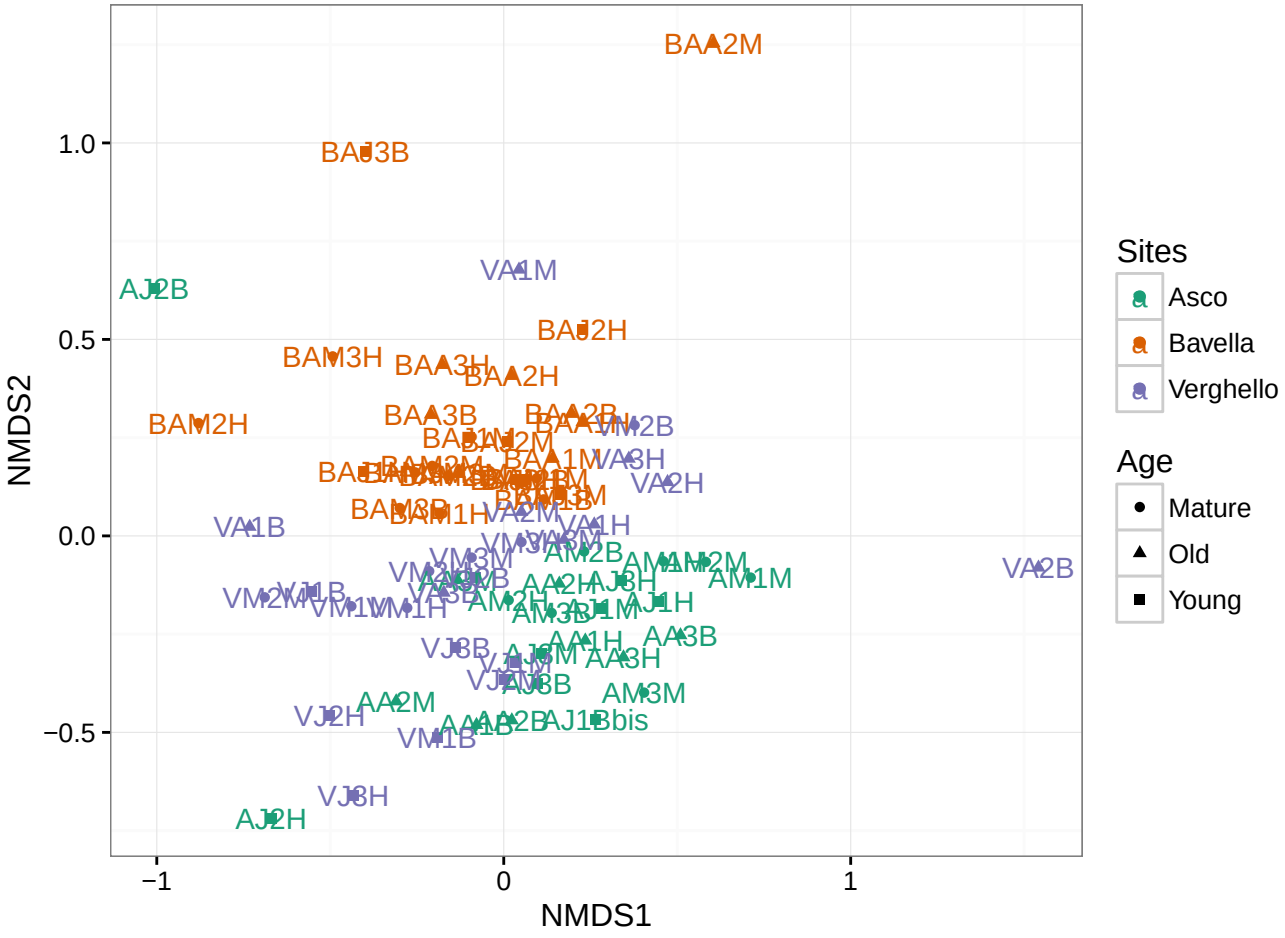

Figure 6.5: NMDS of OTU table. Colors represent sites and shape the age of tree.

```
## ... Procrustes: rmse 0.02217452  max resid 0.08570688
## Run 6 stress 0.2410612
## Run 7 stress 0.2476936
## Run 8 stress 0.239051
## ... Procrustes: rmse 0.01335445  max resid 0.06920975
## Run 9 stress 0.2453805
## Run 10 stress 0.2415455
## Run 11 stress 0.2395415
## Run 12 stress 0.2423029
## Run 13 stress 0.2424918
## Run 14 stress 0.2393948
## ... Procrustes: rmse 0.02140017  max resid 0.08540844
## Run 15 stress 0.2405265
## Run 16 stress 0.2465506
## Run 17 stress 0.2395424
## Run 18 stress 0.2419437
## Run 19 stress 0.2683219
## Run 20 stress 0.2431669
## *** No convergence -- monoMDS stopping criteria:
##      2: no. of iterations >= maxit
##     18: stress ratio > sratmax

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

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

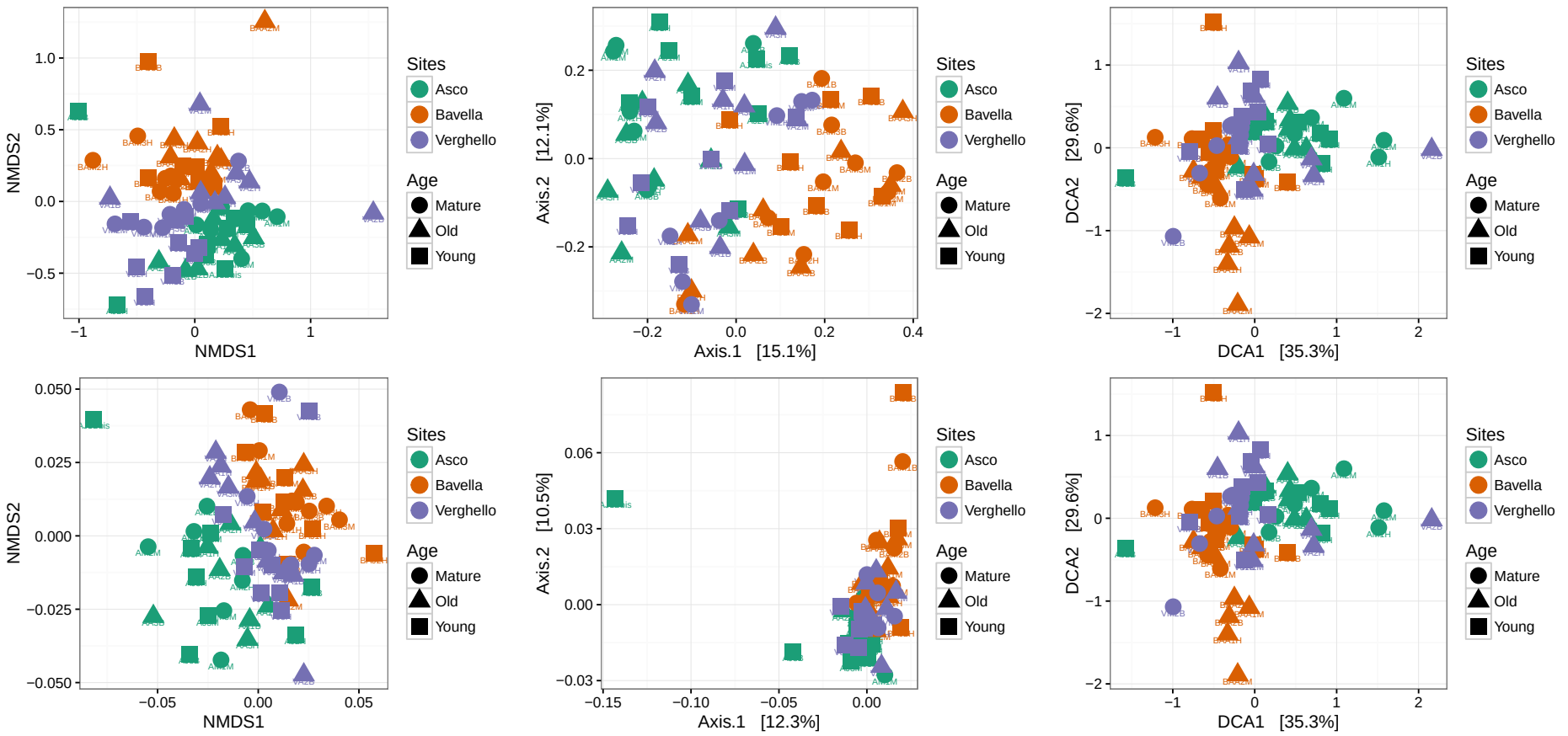

Figure 6.6: Comparison of different distances (bray (up) and gower (bottom)) and ordination methods (NMDS (left), PCoA (center) and DCA (right)).

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

6.3 Permanova on sites, host ages and elevation

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

If we only keep the 264 OTUs present in more than 30 sample, the Permanova results is the following:

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

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

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Sites	2	2.03	1.02	5.41	0.13	0.0001
Age	2	0.63	0.32	1.68	0.04	0.0107
Elevation	2	0.48	0.24	1.29	0.03	0.1232
Sites:Age	4	1.42	0.35	1.88	0.09	0.0004
Sites:Elevation	4	0.68	0.17	0.90	0.04	0.6966
Age:Elevation	4	0.81	0.20	1.07	0.05	0.3269
Sites:Age:Elevation	8	1.43	0.18	0.95	0.09	0.6349
Residuals	45	8.45	0.19		0.53	
Total	71	15.93			1.00	

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

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

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Sites	2	1.98	0.99	5.59	0.13	0.0001
Age	2	0.61	0.30	1.71	0.04	0.0132
Elevation	2	0.47	0.23	1.32	0.03	0.1113
Sites:Age	4	1.37	0.34	1.93	0.09	0.0003
Sites:Elevation	4	0.64	0.16	0.90	0.04	0.6889
Age:Elevation	4	0.77	0.19	1.08	0.05	0.3093
Sites:Age:Elevation	8	1.33	0.17	0.94	0.09	0.6651
Residuals	45	7.98	0.18		0.53	
Total	71	15.14			1.00	

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

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

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

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Sites	2	0.88	0.44	3.66	0.09	0.0001
Age	2	0.44	0.22	1.83	0.05	0.0012
Elevation	2	0.25	0.13	1.04	0.03	0.3749
Sites:Age	4	0.69	0.17	1.43	0.07	0.0057
Sites:Elevation	4	0.39	0.10	0.82	0.04	0.9217
Age:Elevation	4	0.52	0.13	1.08	0.05	0.2703
Sites:Age:Elevation	8	0.88	0.11	0.92	0.09	0.7842
Residuals	45	5.42	0.12		0.57	
Total	71	9.47			1.00	

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

6.4 Permanova on sites, host ages and individual trees

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

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

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Sites	2	2.03	1.02	6.22	0.13	0.0001
Age	2	0.63	0.32	1.93	0.04	0.0023
Sites:Age	4	1.42	0.35	2.17	0.09	0.0001
Sites:Age:IndividualTree	18	4.49	0.25	1.53	0.28	0.0001
Residuals	45	7.35	0.16		0.46	
Total	71	15.93			1.00	

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

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

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Sites	2	1.98	0.99	6.48	0.13	0.0001
Age	2	0.61	0.30	1.98	0.04	0.0024
Sites:Age	4	1.37	0.34	2.24	0.09	0.0001
Sites:Age:IndividualTree	18	4.30	0.24	1.56	0.28	0.0001
Residuals	45	6.88	0.15		0.45	
Total	71	15.14			1.00	

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

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

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Sites	2	0.88	0.44	4.13	0.09	0.0001
Age	2	0.44	0.22	2.07	0.05	0.0004
Sites:Age	4	0.70	0.18	1.65	0.07	0.0004
Sites:Age:IndividualTree	18	2.64	0.15	1.38	0.28	0.0001
Residuals	45	4.80	0.11		0.51	
Total	71	9.47			1.00	

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

6.5 Differences in abundances and OTUs number by Order.

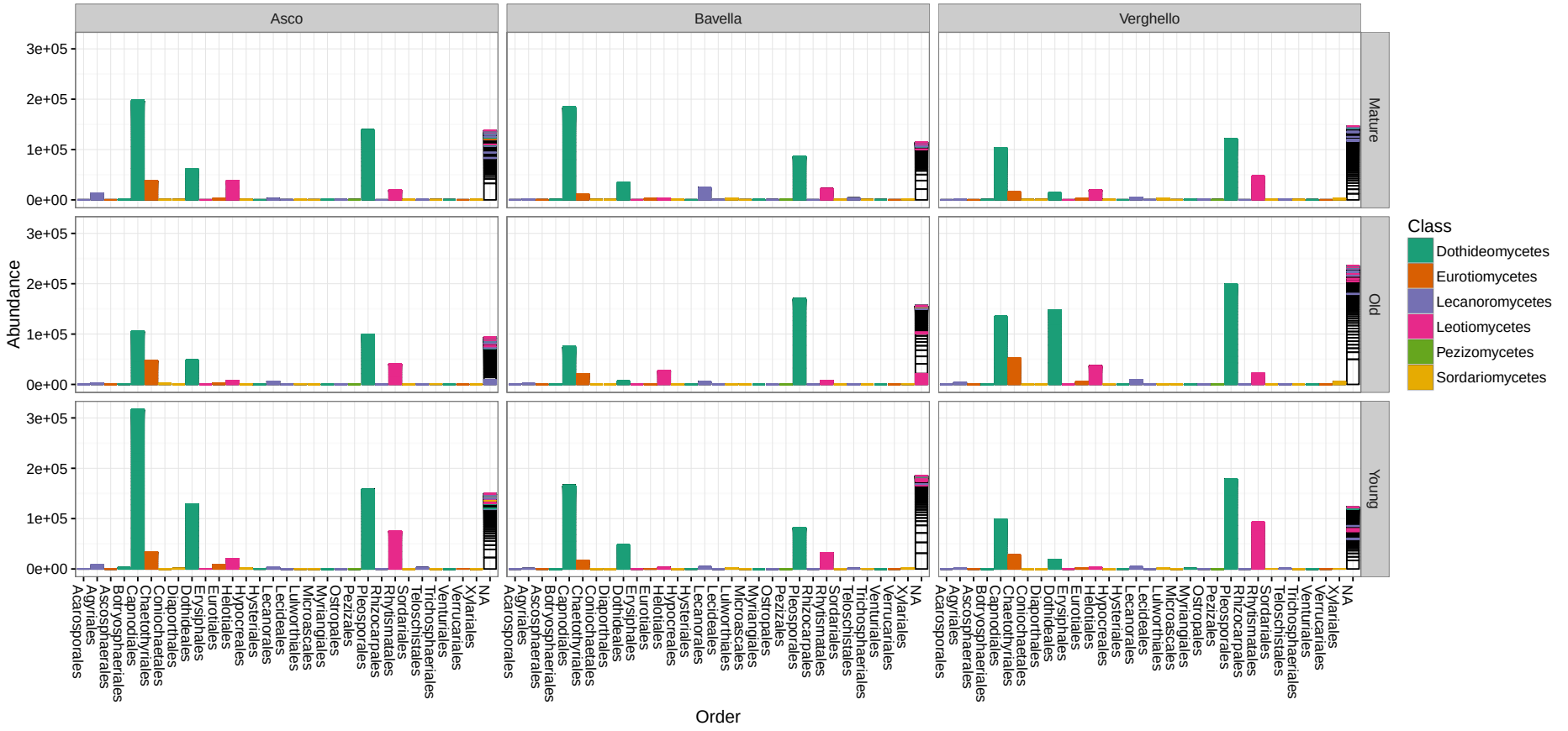

Figure 6.7: Taxonomic distribution of sequences in the different site * age combinaison.

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

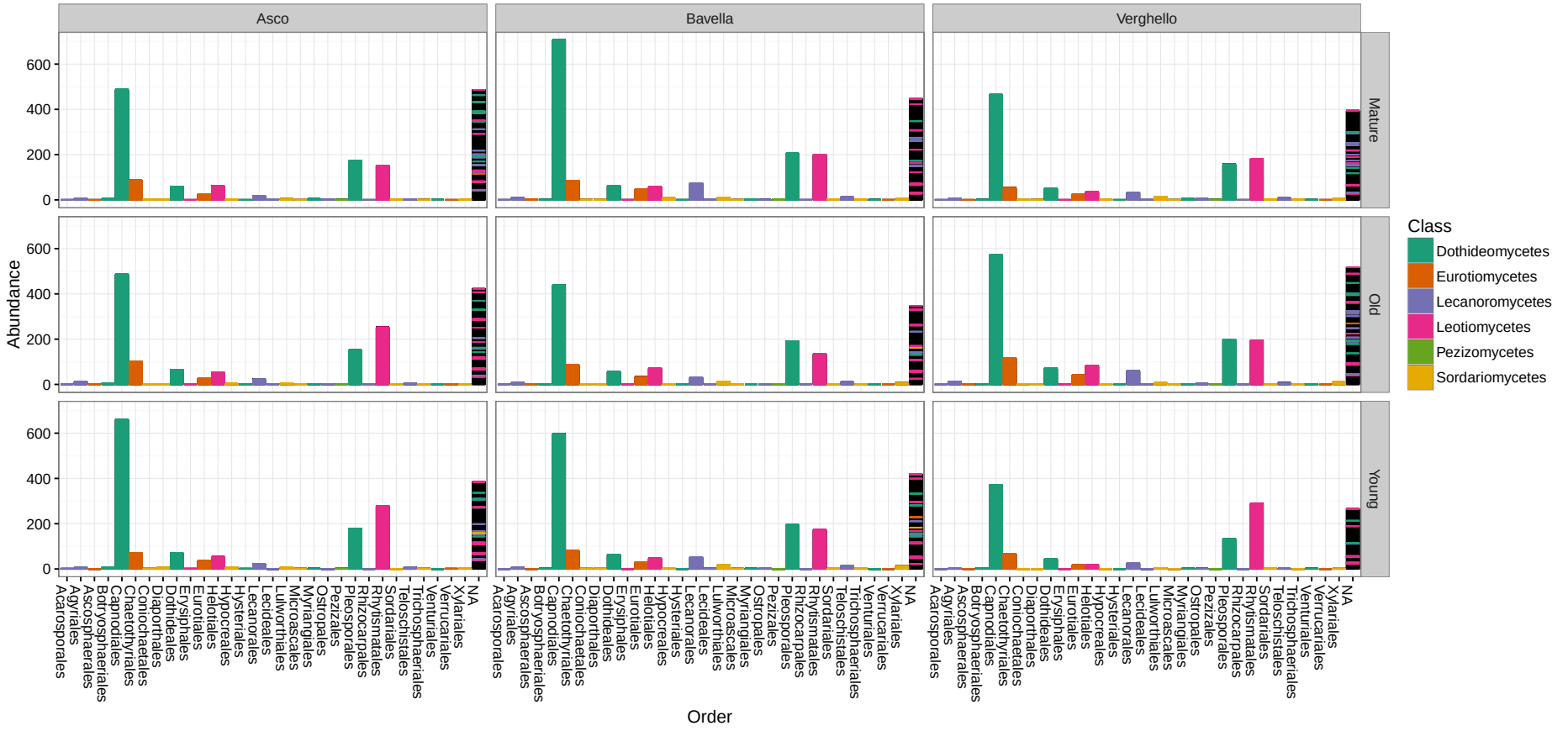

Figure 6.8: Taxonomic distribution of OTUs in the different site * age combinaison.

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

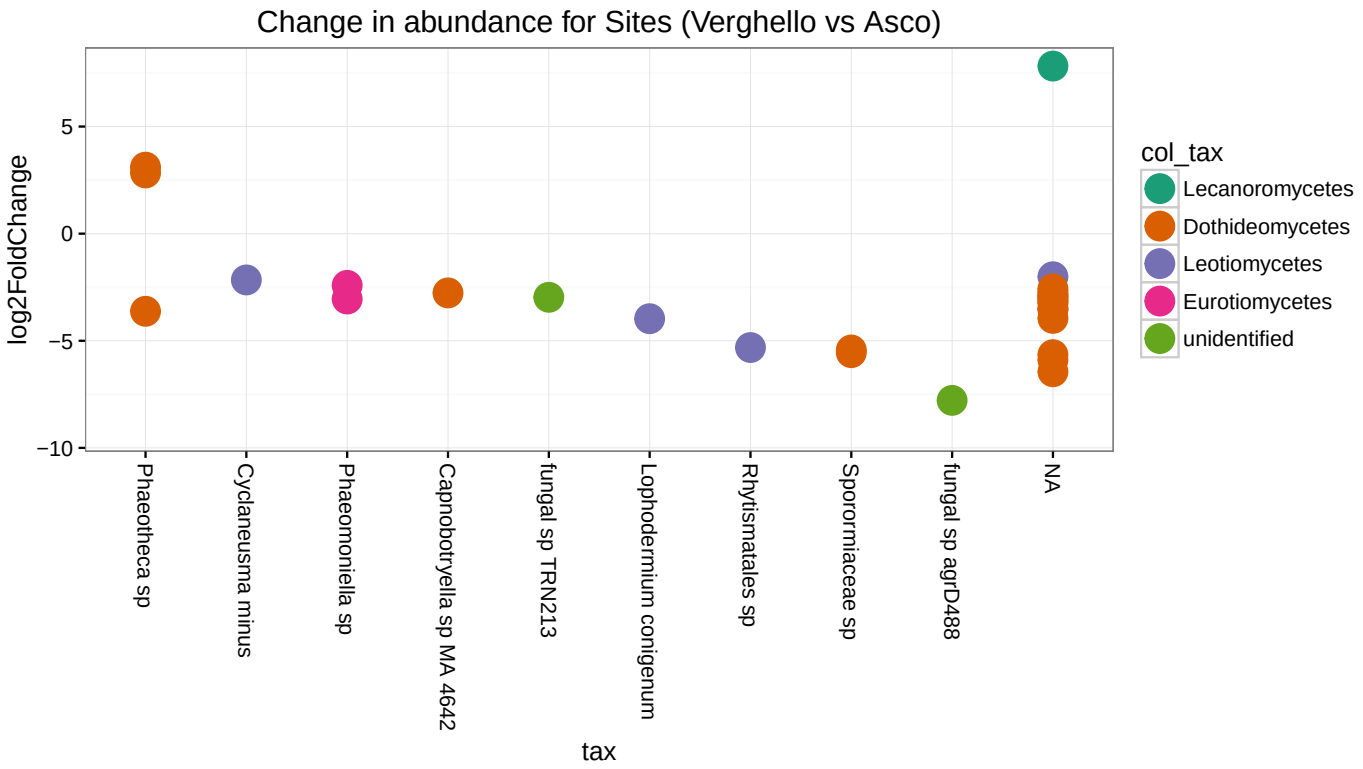

Figure 6.9: OTUs significantly different in terms of abundances between Verghello (positive values) and Asco (negative values)

6.6 Differences in abundances for each OTUs

6.6.1 Pairwise comparison of the OTUs composition by sites

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

## [1] '1.12.3'

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

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

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

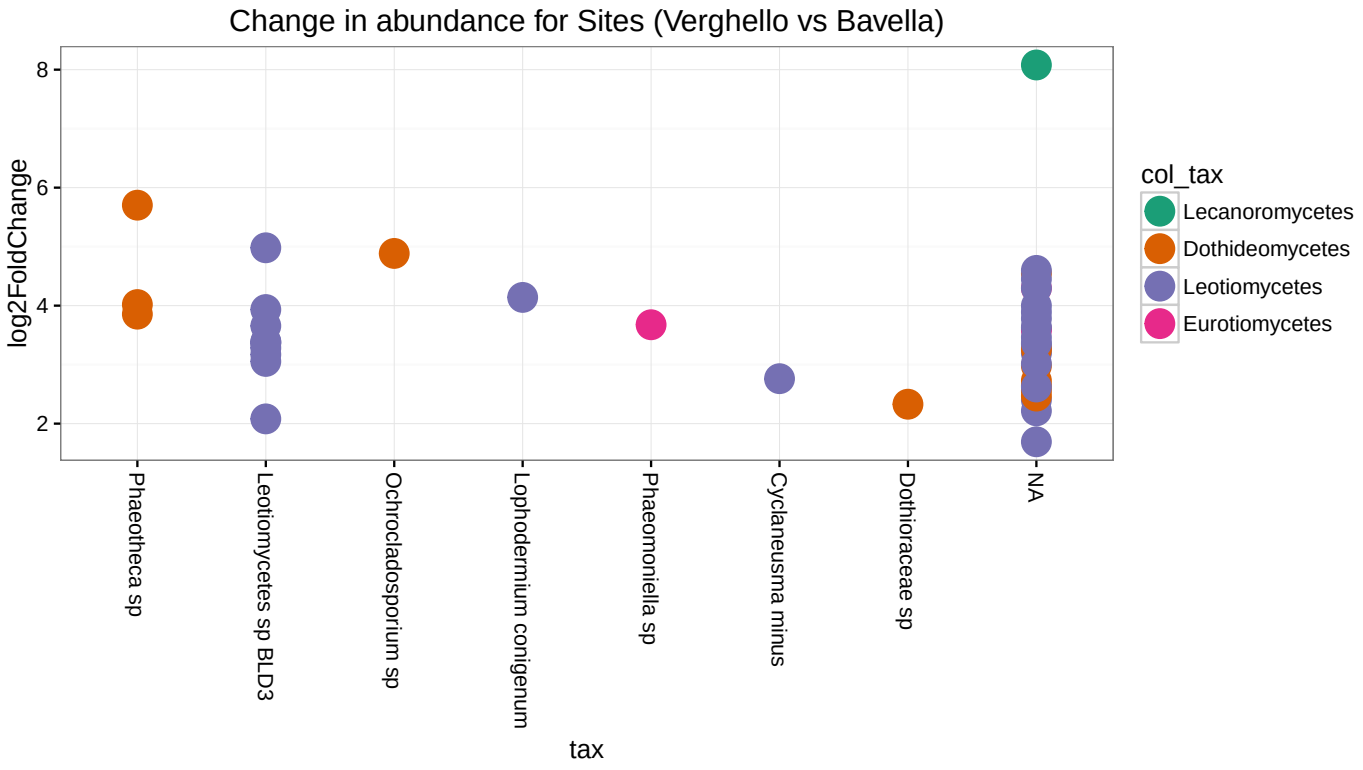

Figure 6.10: OTUs significantly different in terms of abundances between Verghello (positive values) and Bavella (negative values)

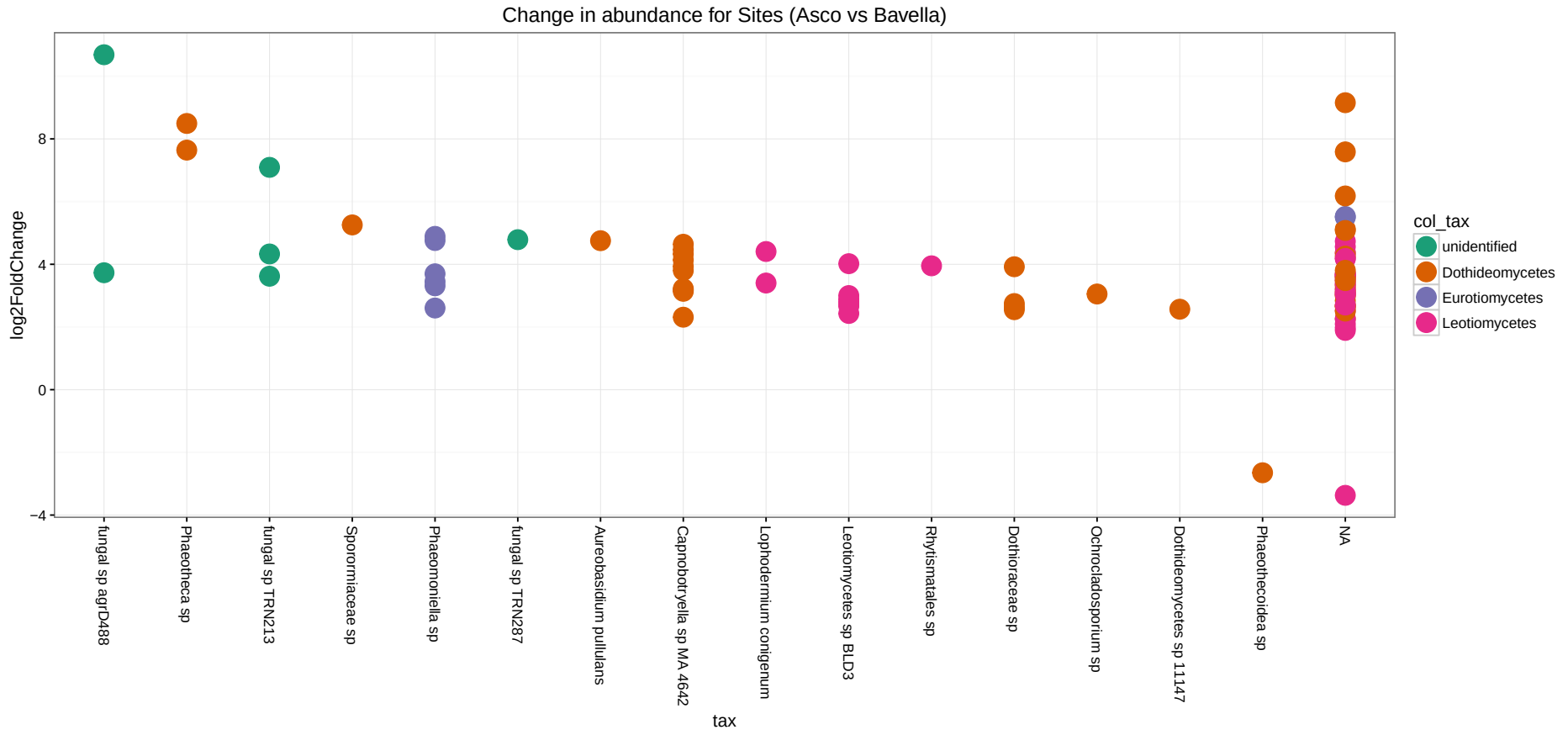

Figure 6.11: OTUs significantly different in terms of abundances between Asco (positive values) and Bavella (negative values)

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

6.6.2 Pairwise comparison of Order composition by sites

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

	Comparison	Species	Class	log2FoldChange (negative = more on second levels)
1	Verghello vs Asco	Phaeomoniella sp	Eurotiomycetes	-3.05107901970215
2	Verghello vs Asco			3.54353869821801
3	Verghello vs Asco		Dothideomycetes	-3.95537713503649
4	Verghello vs Asco		Dothideomycetes	-2.92539665377805
5	Verghello vs Asco	fungal sp TRN213	unidentified	-2.97056468516605
6	Verghello vs Asco			3.32111694969675
7	Verghello vs Asco			-2.79725259034182
8	Verghello vs Asco	Phaeotheca sp	Dothideomycetes	-3.62427001409066
9	Verghello vs Asco			3.20479745524644
10	Verghello vs Asco		Dothideomycetes	-2.88511609729366
11	Verghello vs Asco		Dothideomycetes	-3.24624139821964
12	Verghello vs Asco	Capnobotryella sp MA 4642	Dothideomycetes	-2.76678968604494
13	Verghello vs Asco		Dothideomycetes	-3.1547426069951
14	Verghello vs Asco		Dothideomycetes	-5.89945805866822
15	Verghello vs Asco	Phaeomoniella sp	Eurotiomycetes	-2.41357890454039
16	Verghello vs Asco		Dothideomycetes	-3.03866691469743
17	Verghello vs Asco		Dothideomycetes	-2.73004633308911
18	Verghello vs Asco		Leotiomycetes	-2.01038323987597
19	Verghello vs Asco			2.81448714320522
20	Verghello vs Asco	Phaeotheca sp	Dothideomycetes	3.10428604317115
21	Verghello vs Asco		Dothideomycetes	-5.64743358632746
22	Verghello vs Asco	Cyclaneusma minus	Leotiomycetes	-2.16058557955925
23	Verghello vs Asco		Dothideomycetes	-3.50536259017263
24	Verghello vs Asco			2.66697932664646
25	Verghello vs Asco	Rhytismatales sp	Leotiomycetes	-5.31547426192285
26	Verghello vs Asco	fungal sp agrD488	unidentified	-7.7827414909125
27	Verghello vs Asco	Phaeotheca sp	Dothideomycetes	2.83379010474725
28	Verghello vs Asco			-5.4256098972322
29	Verghello vs Asco		Dothideomycetes	-3.53198145198711
30	Verghello vs Asco			-9.30902629841758
31	Verghello vs Asco	Sporormiaceae sp	Dothideomycetes	-5.55507211796045
32	Verghello vs Asco		Dothideomycetes	-2.57190659977439
33	Verghello vs Asco	Sporormiaceae sp	Dothideomycetes	-5.43050858593381
34	Verghello vs Asco	Lophodermium conigenum	Leotiomycetes	-3.96886297808345
35	Verghello vs Asco		Dothideomycetes	-6.44591426874139
36	Verghello vs Asco		Lecanoromycetes	7.82386689603266
37	Verghello vs Bavella	Phaeomoniella sp	Eurotiomycetes	3.67602138247602
38	Verghello vs Bavella	Lophodermium conigenum	Leotiomycetes	4.13941409359517
39	Verghello vs Bavella		Dothideomycetes	2.55781619384562
40	Verghello vs Bavella	Leotiomycetes sp BLD3	Leotiomycetes	3.35931316171303
41	Verghello vs Bavella			2.58658175868882
42	Verghello vs Bavella	Phaeotheca sp	Dothideomycetes	5.70291208818186
43	Verghello vs Bavella			5.61242510072171
44	Verghello vs Bavella	Dothioraceae sp	Dothideomycetes	2.3289968766658
45	Verghello vs Bavella	Leotiomycetes sp BLD3	Leotiomycetes	2.0804761089416
46	Verghello vs Bavella		Dothideomycetes	2.64456951842101
47	Verghello vs Bavella		Eurotiomycetes	4.29614461305071
48	Verghello vs Bavella	Ochrocladosporium sp	Dothideomycetes	4.88488580265
49	Verghello vs Bavella			6.00787883464121
50	Verghello vs Bavella		Dothideomycetes	4.54638433421036
51	Verghello vs Bavella		Leotiomycetes	2.21687948076651
52	Verghello vs Bavella	Leotiomycetes sp BLD3	Leotiomycetes	3.38301561770858
53	Verghello vs Bavella	Cyclaneusma minus	Leotiomycetes	2.76186385140447
54	Verghello vs Bavella	Leotiomycetes sp BLD3	Leotiomycetes	3.28336704401976
55	Verghello vs Bavella			7.65248964162047
56	Verghello vs Bavella		Leotiomycetes	3.6165362705414
57	Verghello vs Bavella		Leotiomycetes	3.99976239982887
58	Verghello vs Bavella			3.87939724019365
59	Verghello vs Bavella	Phaeotheca sp	Dothideomycetes	4.01806767586435
60	Verghello vs Bavella		Leotiomycetes	3.231468769551
61	Verghello vs Bavella		Leotiomycetes	3.34216157981595
62	Verghello vs Bavella	Leotiomycetes sp BLD3	Leotiomycetes	3.1711576946885
63	Verghello vs Bavella		Leotiomycetes	4.59746832114555
64	Verghello vs Bavella			2.63868092157711
65	Verghello vs Bavella		Leotiomycetes	1.69210308388229
66	Verghello vs Bavella		Leotiomycetes	3.19206757576461
67	Verghello vs Bavella	Leotiomycetes sp BLD3	Leotiomycetes	3.93368546962727
68	Verghello vs Bavella		Leotiomycetes	3.28389051733237
69	Verghello vs Bavella		Leotiomycetes	2.40040923057187
70	Verghello vs Bavella		Leotiomycetes	4.44948476518476
71	Verghello vs Bavella		Dothideomycetes	2.46663451031107
72	Verghello vs Bavella		Dothideomycetes	2.9780384970892
73	Verghello vs Bavella		Dothideomycetes	3.25472555312655
74	Verghello vs Bavella		Leotiomycetes	3.78051751669329
75	Verghello vs Bavella		Eurotiomycetes	3.58566218166602
76	Verghello vs Bavella		Dothideomycetes	2.72951285380043
77	Verghello vs Bavella	Leotiomycetes sp BLD3	Leotiomycetes	3.0543701357961
78	Verghello vs Bavella			2.63163191165821
79	Verghello vs Bavella	Leotiomycetes sp BLD3	Leotiomycetes	3.6563095066927
80	Verghello vs Bavella	Phaeotheca sp	Dothideomycetes	3.85550530560902
81	Verghello vs Bavella		Leotiomycetes	2.62092114924148
82	Verghello vs Bavella		Leotiomycetes	3.89113707070487
83	Verghello vs Bavella		Leotiomycetes	3.36226873879913
84	Verghello vs Bavella			3.82808172973933
85	Verghello vs Bavella		Leotiomycetes	3.01668814871989
86	Verghello vs Bavella		Leotiomycetes	3.46404836937354
87	Verghello vs Bavella			3.52364156996503
88	Verghello vs Bavella		Leotiomycetes	4.29720099936154
89	Verghello vs Bavella			4.60090458278243
90	Verghello vs Bavella			4.76881881507057
91	Verghello vs Bavella	Leotiomycetes sp BLD3	Leotiomycetes	4.98058502602772
92	Verghello vs Bavella			3.01797034070211
93	Verghello vs Bavella		Leotiomycetes	3.63445324388821
94	Verghello vs Bavella	Leotiomycetes sp BLD3	Leotiomycetes	3.38983557555046
95	Verghello vs Bavella		Lecanoromycetes	8.07928909512927
96	Asco vs Bavella	Phaeomoniella sp	Eurotiomycetes	4.75786697111987
97	Asco vs Bavella	Lophodermium conigenum	Leotiomycetes	4.40757402240425
98	Asco vs Bavella	fungal sp TRN287	unidentified	4.78565879681266
99	Asco vs Bavella			5.12588173202422
100	Asco vs Bavella	Capnobotryella sp MA 4642	Dothideomycetes	4.46475578809201
101	Asco vs Bavella	Leotiomycetes sp BLD3	Leotiomycetes	2.66720390814126
102	Asco vs Bavella			3.73496900949493
103	Asco vs Bavella	Phaeotheca sp	Dothideomycetes	8.49269372659997
104	Asco vs Bavella			7.48908997510742
105	Asco vs Bavella	Dothioraceae sp	Dothideomycetes	2.55191727299794
106	Asco vs Bavella	Phaeomoniella sp	Eurotiomycetes	3.45215244403755
107	Asco vs Bavella	Leotiomycetes sp BLD3	Leotiomycetes	3.09237777117663

	Comparison	Order	Class	log2FoldChange (negative = more on second levels)
1	Verghello vs Asco	Xylariales	Sordariomycetes	3.861692475544
2	Verghello vs Bavella	Incertae sedis	Leotiomycetes	-1.3571057701563
3	Verghello vs Bavella	unidentified	unidentified	1.5718779461989
4	Asco vs Bavella	Botryosphaeriales	Dothideomycetes	5.19080395584996
5	Asco vs Bavella	Eurotiales	Eurotiomycetes	1.76365664471148
6	Asco vs Bavella	Incertae sedis	Leotiomycetes	-1.68189384345562
7	Asco vs Bavella	unidentified	unidentified	1.45154956448541
8	Asco vs Bavella	Xylariales	Sordariomycetes	-3.65283955986027

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

List of Figures

2.1 Number of sequences by sample . . . 8
2.2 Number of OTU present in a given number of samples . . . 9
2.3 Number of sequences by OTU (log10 transformed) . . . 10
3.1 Number of OTUs by samples in fonction the number of sequences by samples (log10 axe) 11
3.2 Distribution of reference sequences length . . . 12
3.3 Number of sequences by OTUs (log10 axe) in fonction the number of samples where OTUs were found . . . 13
3.4 Distribution of the number of sequences in the taxonomy . . . 13
3.5 Number of sequences of the 30 more abundant OTUs (number of sequences) . . . 14
3.6 Number of sequences of the 30 most abundant OTUs (number of sequences) . . . 16
3.7 Number of samples of the 30 more frequent OTUs (number of samples) . . . 17
4.1 Distribution of OTUs into functional Guild . . . 19
4.2 Distribution of sequences (log10 transformed) into functional Guild . . . 20
5.1 Rarefaction curves for each sites . . . 21
5.2 Rarefaction curves for each samples using sequences number on x-axes . . . 22
5.3 Diversity of each sites . . . 23
5.4 Rarefaction curves for each tree age modalities . . . 24
5.5 Diversity in function of tree age . . . 25
5.6 Rarefaction curves for each elevation . . . 26
5.7 Diversity in function of elevation . . . 27
5.8 Results of the Tuckey HSD testing for differences in mean Hill numbers among pairs of modalities . . . 28
6.1 Venn diagramm ef the distribution of OTUs among Sites . . . 28
6.2 Venn diagramm ef the distribution of OTUs among host age . . . 29
6.3 Venn diagramm ef the distribution of OTUs among elevation of samples . . . 29
6.4 Stress plot of the NMDS . . . 31
6.5 NMDS of OTU table . . . 31
6.6 Comparison of different distances (bray (up) and gower (bottom)) and ordination methods (NMDS (left), PCoA (center) and DCA (right)) . . . 33
6.7 Taxonomic distribution of sequences in the different site * age combinaison . . . 37
6.8 Taxonomic distribution of OTUs in the different site * age combinaison . . . 38
6.9 OTUs significantly different in terms of abundances between Verghello (positive values) and Asco (negative values) . . . 39
6.10 OTUs significantly different in terms of abundances between Verghello (positive values) and Bavella (negative values) . . . 40
6.11 OTUs significantly different in terms of abundances between Asco (positive values) and Bavella (negative values) . . . 41

List of Tables

1 Number of OTUs, samples and sequences after filtering . . . 10
2 Taxonomie of the 30 more frequent OTUs (number of sequences) . . . 15
3 Taxonomie of the 30 more frequent OTUs (number of samples) . . . 18
4 Summary of the linear model of species richness (Hill number 1 (q = 0) . . . 24
5 Summary of the linear model of the exponential of Shannons entropy index (Hill number 2 (q = 1) . . . 30

6 Summary of the linear model of inverse of Simpsons concentration index (Hill number 3 (q = 2) 30
7 Result of the permanova on abundances (number of sequence). 34
8 Result of the permanova on abundances (number of sequence) using only OTUs present in more than 30 samples 34
9 Result of the permanova on OTUs (each OTU is representing by one sequence)). 35
10 Result of the permanova on abundances (number of sequence). 35
11 Result of the permanova on abundances (number of sequence) using only OTUs present in more than 30 samples 36
12 Result of the permanova on OTUs (each OTU is representing by one sequence)). 36
13 OTUs showing differential abundances in the different sites. 43
14 Order showing differential abundances in the different sites. 44