Appendix S9: results after UPARSE clustering discarding unique sequences keeping sequences classified by ITSx as Tracheophyta

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

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 discarding unique sequences and keeping sequences classified by ITSx as Tracheophyta (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 18:08:04. 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_min2_without_ITSx"

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

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

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

2.2 Load and convert loading

2.2.1 Otu table

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

2.2.2 Taxonomy

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

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

# Format taxonomy for phyloseq
taxRDP <- taxRDP_brut[match(taxa_names(dataBiom),
                            paste(taxRDP_brut[, 1], taxRDP_brut[, 2], "", sep = ";")),
                      c(5, 7, 9, 11, 13, 15, 17)]
taxRDP <- tax_table(as.matrix(taxRDP))
taxa_names(taxRDP) <- taxa_names(dataBiom)
colnames(taxRDP) <- c("Domain", "Phylum", "Class", "Order", "Family",
                      "Genus", "Species")
```

2.2.3 Add FUNguild information to taxonomy Table

```
taxRDP2 <- as.data.frame(taxRDP)
funguild <- read.delim(paste("data/", data_folder, "/FUNGUILD.guilds.txt", sep=""))

match_interm <- match(paste(funguild$OTU_ID,";", sep=""), gsub(";size=", "_",
                                                               rownames(taxRDP2)))

taxRDP2$Trophic_Mode <- NA
taxRDP2$Trophic_Mode[match_interm] <- as.character(funguild$Trophic.Mode)

## Error in taxRDP2$Trophic_Mode[match_interm] <- as.character(funguild$Trophic.Mode):  NAs
interdits dans les affectations indicées

taxRDP2$Guild <- NA
taxRDP2$Guild[match_interm] <- as.character(funguild$Guild)

## Error in taxRDP2$Guild[match_interm] <- as.character(funguild$Guild):  NAs interdits dans
les affectations indicées

taxRDP2$Confidence_Ranking <- NA
taxRDP2$Confidence_Ranking[match_interm] <- as.character(funguild$Confidence.Ranking)
```

```
## Error in taxRDP2$Confidence_Ranking[match_interm] <- as.character(funguild$Confidence.Ranking):
NAs interdits dans les affectations indicées
```

```
taxRDP2$Growth_Morphology <- NA
taxRDP2$Growth_Morphology[match_interm] <- as.character(funguild$Growth.Morphology)
```

```
## Error in taxRDP2$Growth_Morphology[match_interm] <- as.character(funguild$Growth.Morphology):
NAs interdits dans les affectations indicées
```

```
taxRDP2$Trait<-NA
taxRDP2$Trait[match_interm] <- as.character(funguild$Trait)
```

```
## Error in taxRDP2$Trait[match_interm] <- as.character(funguild$Trait):  NAs interdits dans
les affectations indicées
```

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

2.2.4 Representative sequences

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

## Processing map file...

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

2.2.5 Samples information

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

## Processing Reference Sequences...
```

2.2.6 Create the phyloseq object

```
data_all <- merge_phyloseq(dataBiom, repset, taxRDP2)

sample_data(data_all) <- map_endo

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

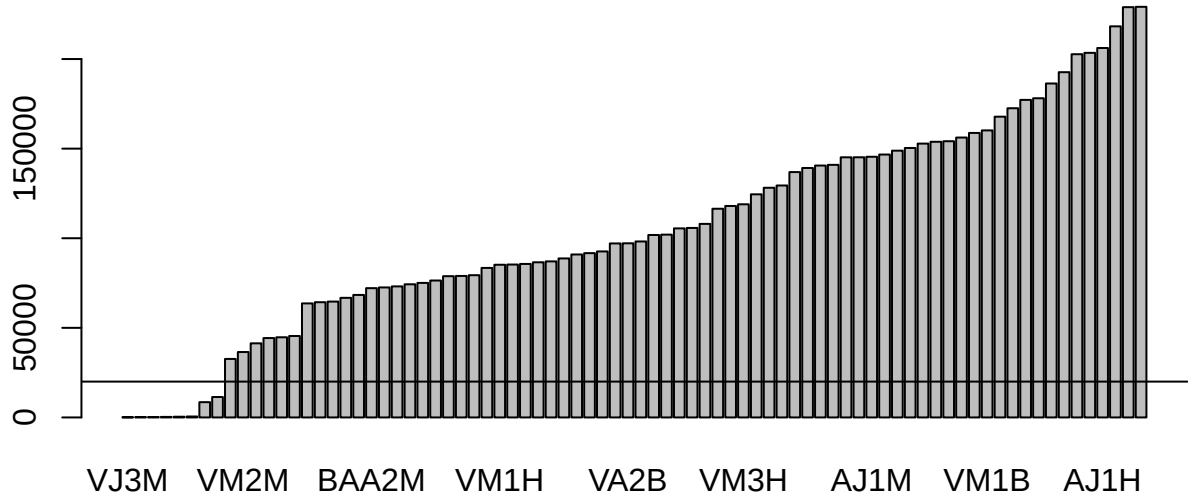

Figure 2.1: Number of sequences by sample

2.2.7 Caracteristics of the phyloseq data

```
data_all

## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 792 taxa and 80 samples ]
## sample_data() Sample Data:       [ 80 samples by 6 sample variables ]
## tax_table()   Taxonomy Table:    [ 792 taxa by 12 taxonomic ranks ]
## refseq()      DNAStringSet:      [ 792 reference sequences ]
```

The data are made of 8.441951×10^6 sequences representing 792 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)
```

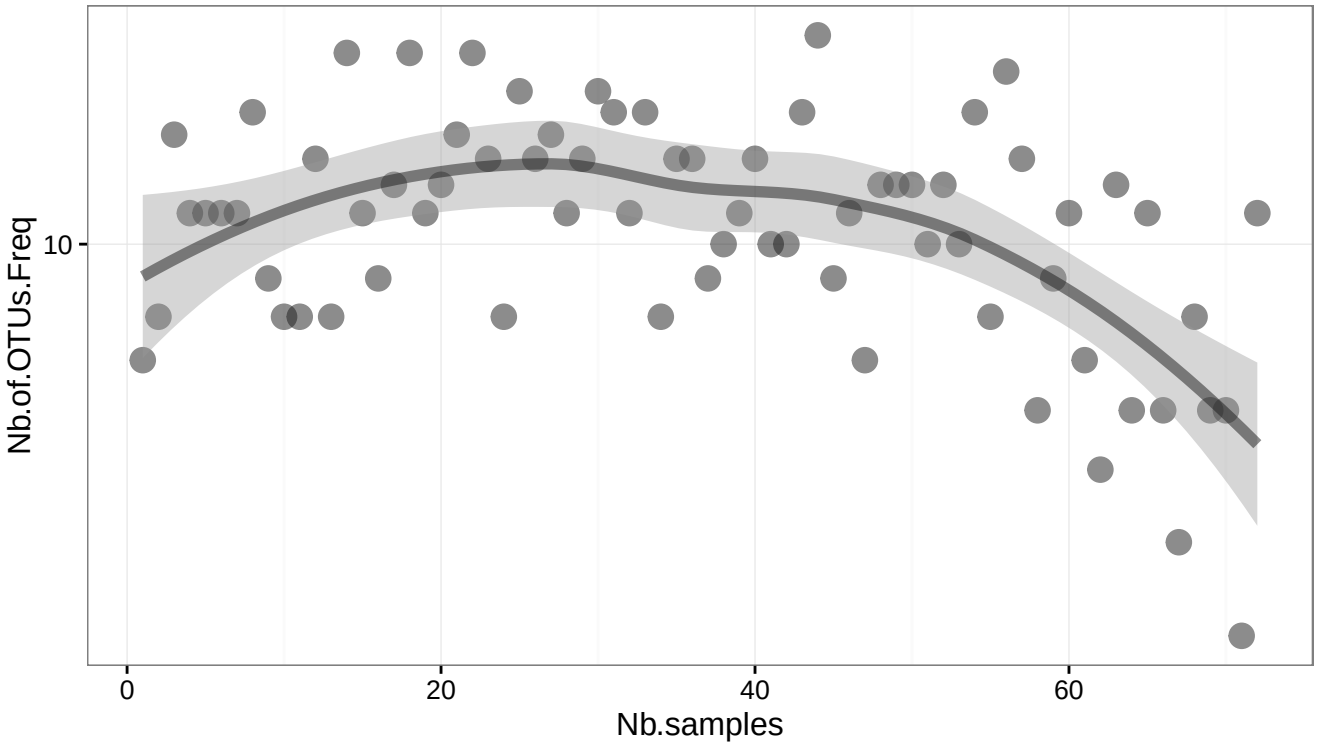

Figure 2.2: Number of OTU present in a given number of samples. Vertical bar illustrate the filtering parameter.

2.4 Filter OTUs by number of samples

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

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

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

summary(df_nbOtu_sample$Nb.samples)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00   18.75   36.50   36.50   54.25   72.00
```

```
N_otu_sam_min

## [1] 1
```

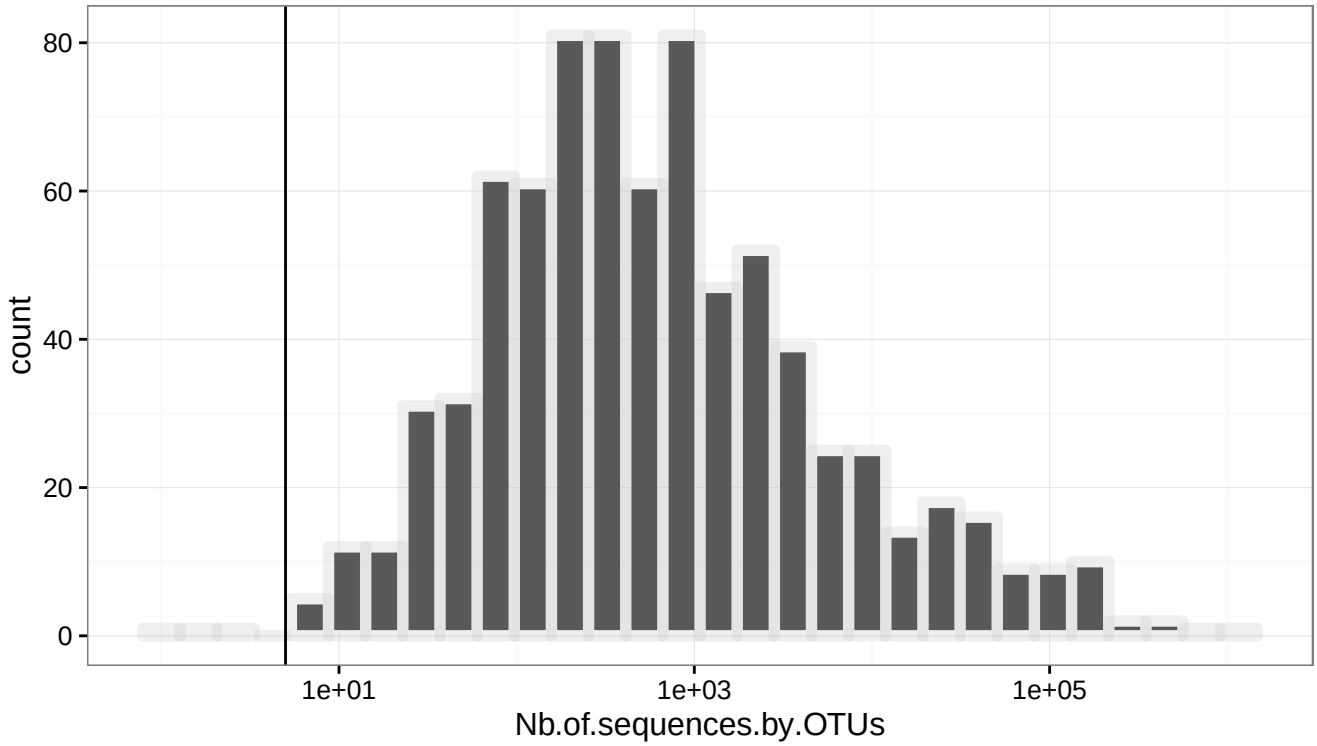

Figure 2.3: Number of sequences by OTU (log10 transformed). Horizontal bar illustrate the filtering parameter.

```
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 792 on the 792 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])

##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##       1.0     135.8     491.0   10630.0    2117.0 1161000.0
```

```
N_seq_otu_min

## [1] 5
```

If we discard OTUs with less than 1 sequences, we keep 789 on the 792 OTUs (99.62%).

```
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 8.420192×10^6 sequences representing 789 OTUs allocate to 72 samples.

	Nb.of.OTUs	Nb.of.samples	Nb.of.sequences
No filter	792	80	8441951.00
Nb of sequences by sample >= 20000	792	72	8420198.00
Nb of sample by OTUs >= 1	792	72	8420198.00
Nb of sequences by OTUs >= 5	789	72	8420192.00

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

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

3.2 Number of sequences and samples for each OTUs

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

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

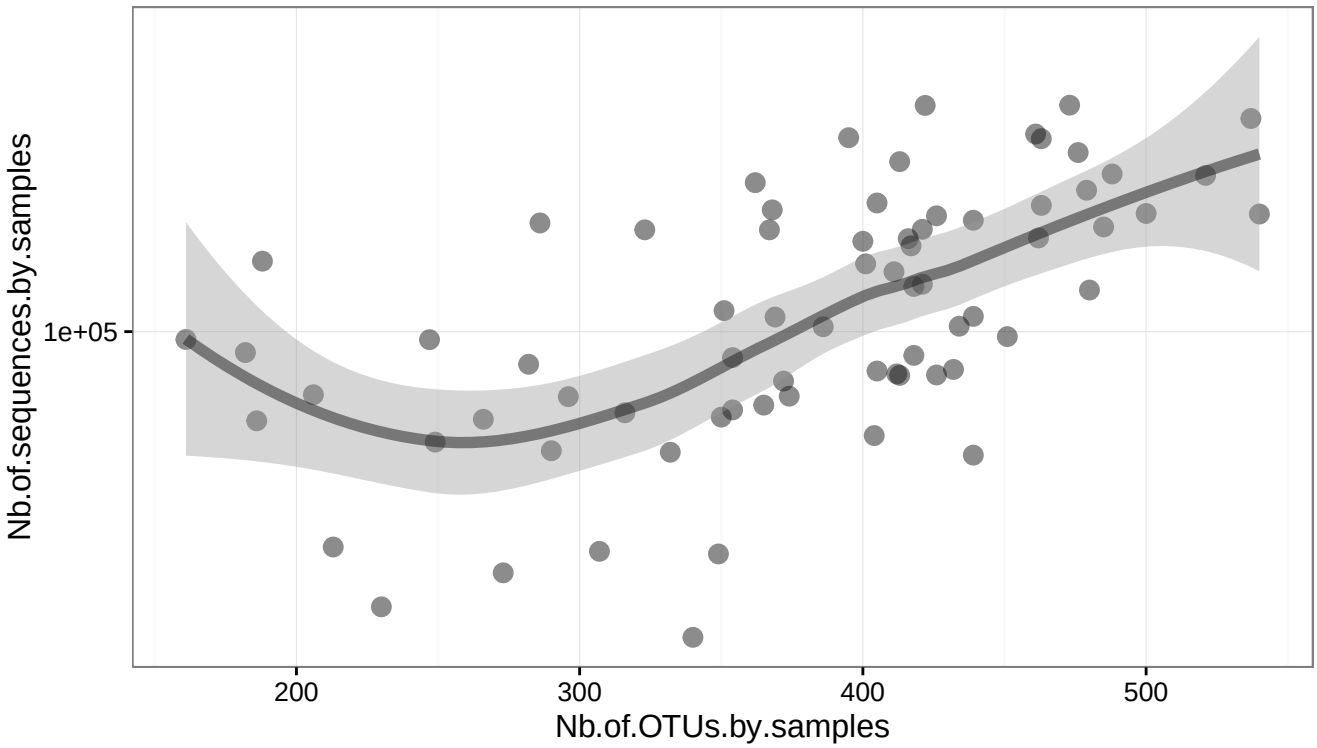

Figure 3.1: Number of OTUs by samples in fonction the number of sequences by samples (log10 axe). The tendency is represented by the line obtain from loess (Local Polynomial Regression Fitting).

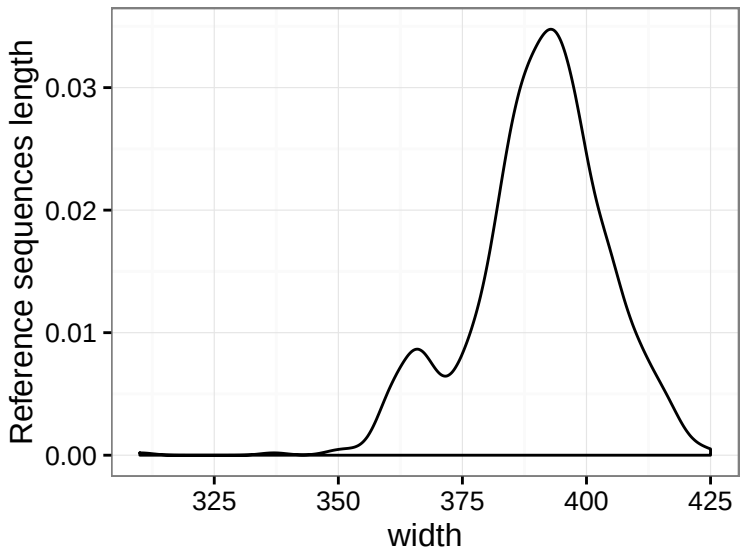

Figure 3.2: Distribution of reference sequences length.

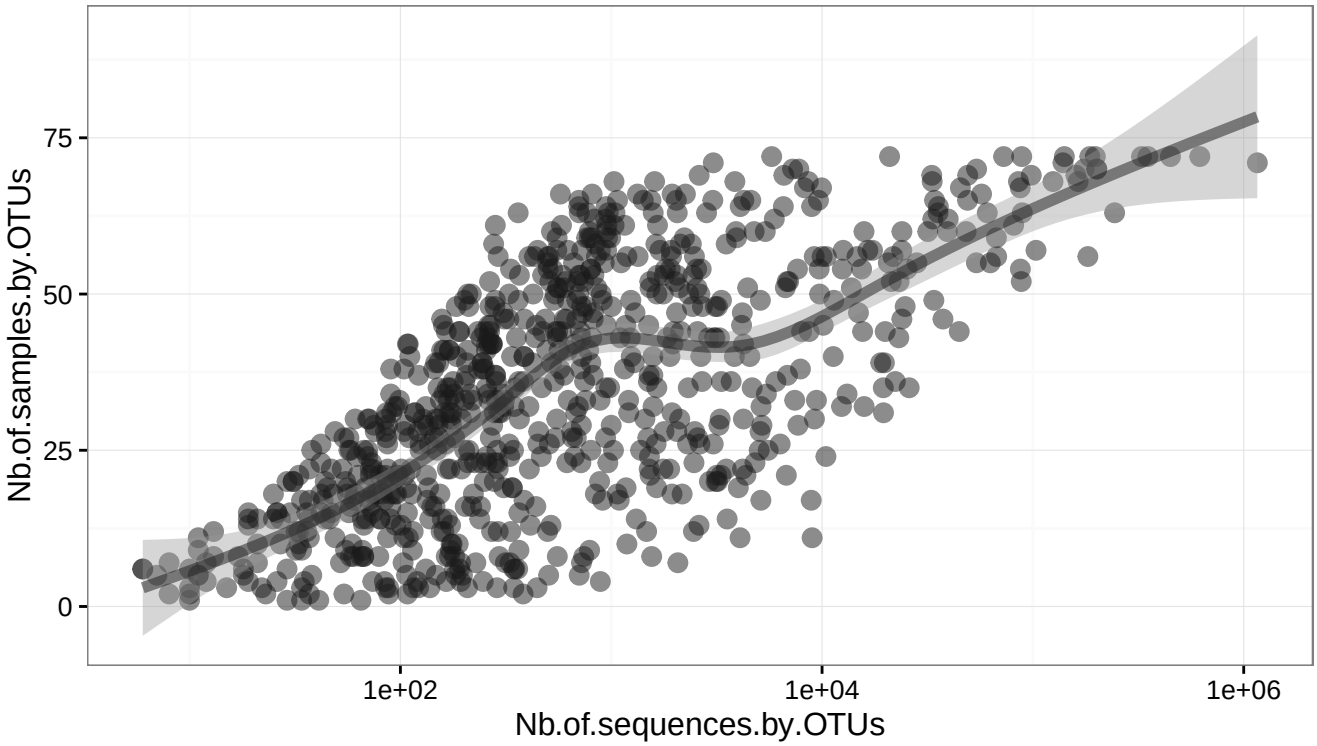

Figure 3.3: Number of sequences by OTUs (log10 axe) in fonction the number of samples where OTUs were found. The tendency is represented by the line obtain from gam (Generalized additive models with integrated smoothness estimation).

3.3 Distribution of sequences in the taxonomy

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

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

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

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

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

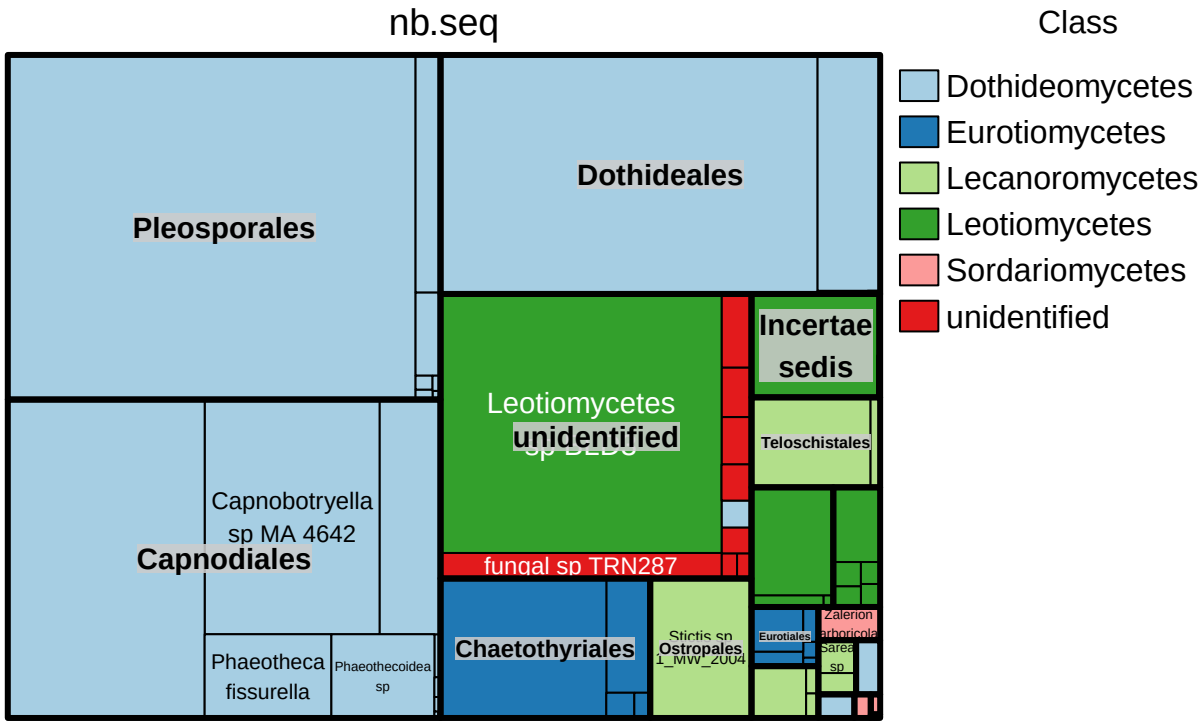

Figure 3.4: Distribution of the number of sequences in the taxonomy. Color represent Class, bold lines delimit Order and thick line delimit species.

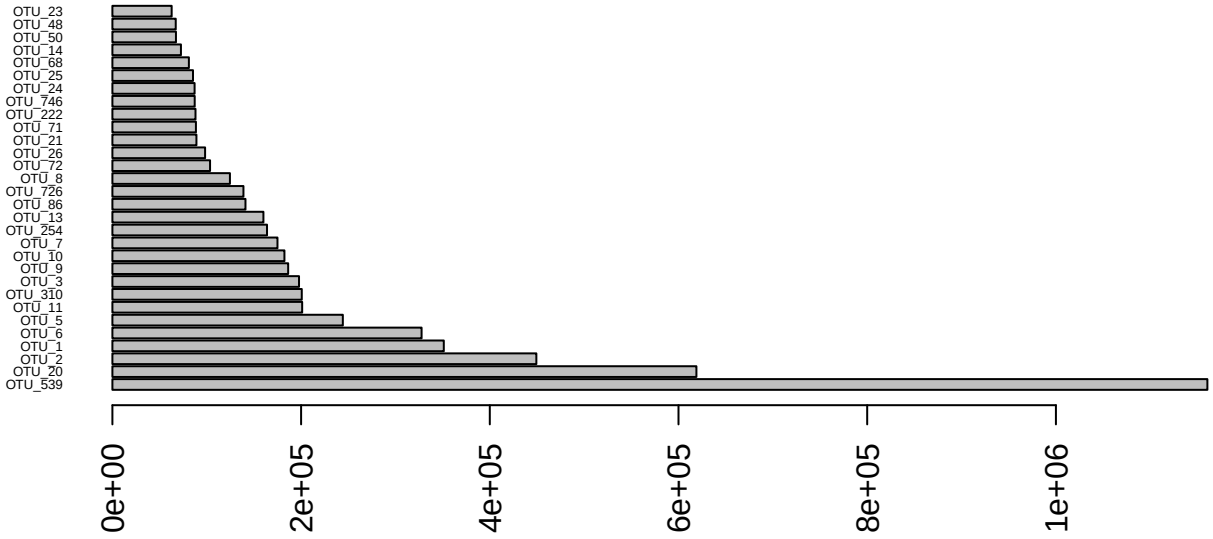

Figure 3.5: Number of sequences of the 30 more abundant OTUs (number of sequences).

```
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							1160548
Ascomycota	Leotiomycetes							618987
Ascomycota								449340
Ascomycota								351230
Ascomycota	Dothideomycetes	Pleosporales						327679
Ascomycota	Dothideomycetes	Capnodiales	Incertae sedis	Capnobotryella	Capnobotryella sp MA 4642			244224
Ascomycota	Leotiomycetes	unidentified	unidentified	unidentified	Leotiomycetes sp BLD3			201133
								200710
Ascomycota	Leotiomycetes							197781
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp			186279
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp			182316
Ascomycota	Dothideomycetes	Pleosporales						174951
Ascomycota	Leotiomycetes	unidentified	unidentified	unidentified	Leotiomycetes sp BLD3			163874
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp			160051
Ascomycota								141027
Ascomycota								138892
Ascomycota								124568
Ascomycota	Dothideomycetes	Pleosporales						103508
Ascomycota	Dothideomycetes	Capnodiales	Davidiellaceae	Cladosporium				98258
Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Phaeomoniella				89068
Ascomycota	Lecanoromycetes	Ostropales	Stictidaceae	Stictis	Stictis sp 1_MW_2004			88531
Ascomycota	Dothideomycetes							88060
Ascomycota								87212
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae					87108
Ascomycota	Dothideomycetes	Capnodiales						85401
Ascomycota	Leotiomycetes	unidentified	unidentified	unidentified	Leotiomycetes sp BLD3			81049
Ascomycota								72643
Ascomycota	Lecanoromycetes	Teloschistales	Teloschistaceae					67357
Ascomycota	Dothideomycetes	Capnodiales	Incertae sedis	Phaeotheca	Phaeotheca fissurella			67126

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

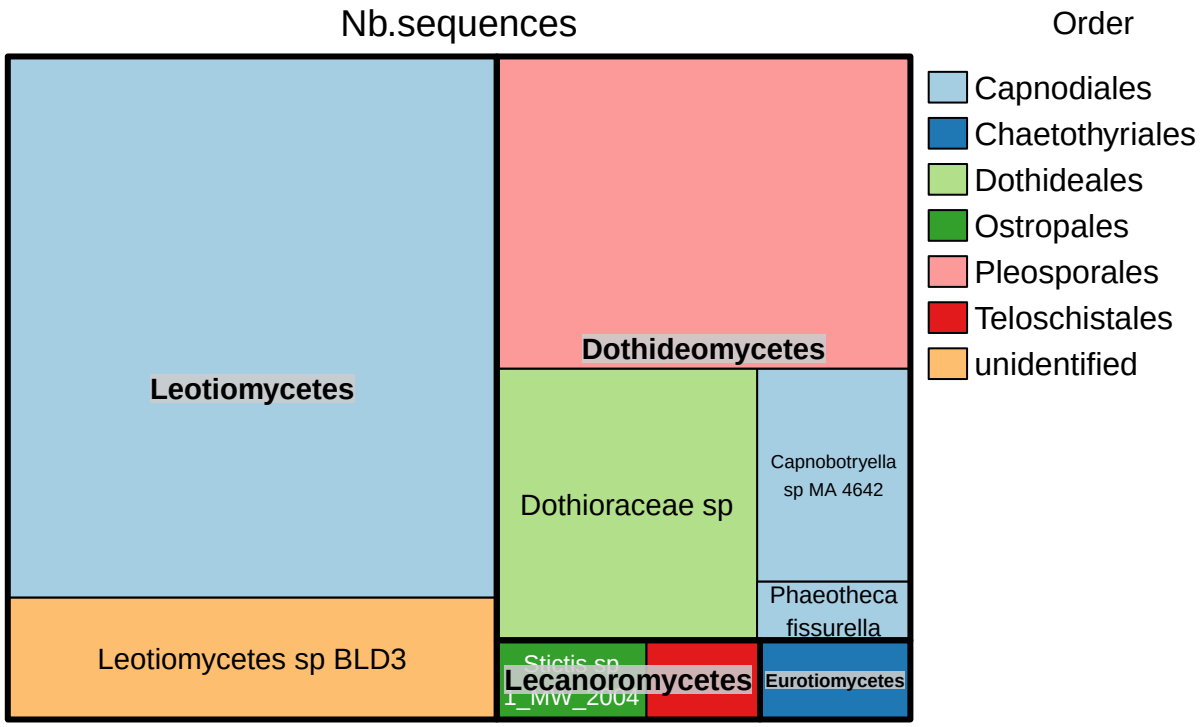

Figure 3.6: Number of sequences of the 30 most abundant OTUs (number of sequences). Colors indicate Order, bold lines delimit Class and thick lines delimit species.

```
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)
```

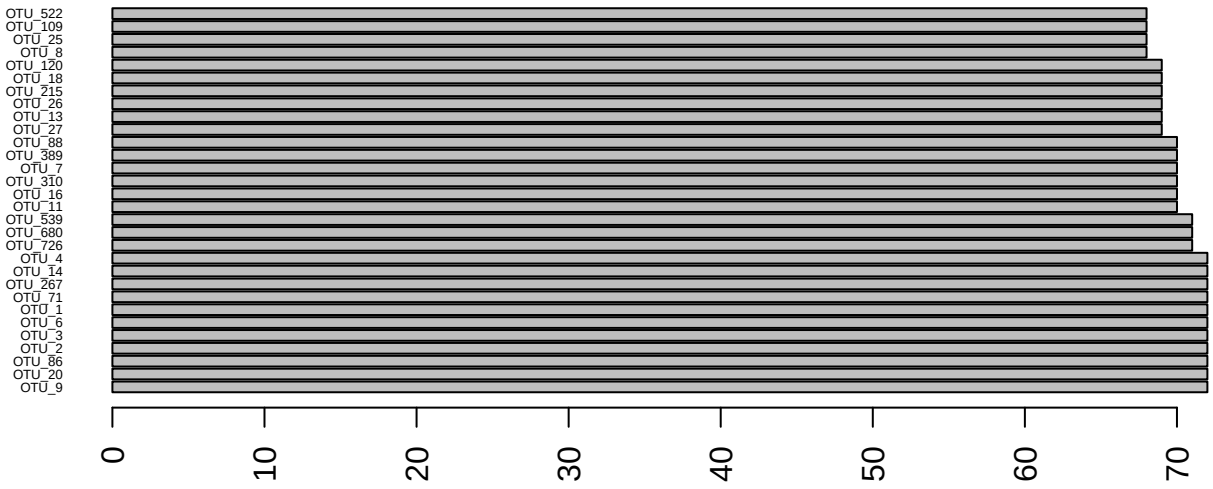

Figure 3.7: Number of samples of the 30 more frequent OTUs (number of samples).

```
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	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp			72
Ascomycota	Leotiomycetes							72
Ascomycota								72
Ascomycota								72
Ascomycota	Leotiomycetes							72
Ascomycota	Dothideomycetes	Pleosporales						72
Ascomycota								72
Ascomycota	Lecanoromycetes	Ostropales	Stictidaceae	Stictis	Stictis sp 1_MW_2004			72
								72
Ascomycota								72
Ascomycota	Dothideomycetes	Pleosporales						72
Ascomycota								71
Ascomycota								71
Ascomycota	Leotiomycetes							71
Ascomycota	Leotiomycetes	unidentified	unidentified	unidentified	Leotiomycetes sp BLD3			70
								70
								70
Ascomycota	Dothideomycetes	Pleosporales						70
Ascomycota	Dothideomycetes	Capnodiales						70
Ascomycota	Leotiomycetes							70
Ascomycota								69
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp			69
Ascomycota	Dothideomycetes	Capnodiales	Davidiellaceae	Cladosporium				69
unidentified	unidentified	unidentified	unidentified	unidentified	fungal sp TRN213			69
Ascomycota	Leotiomycetes							69
Ascomycota	Eurotiomycetes	Eurotiales	Trichocomaceae	Aspergillus				69

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

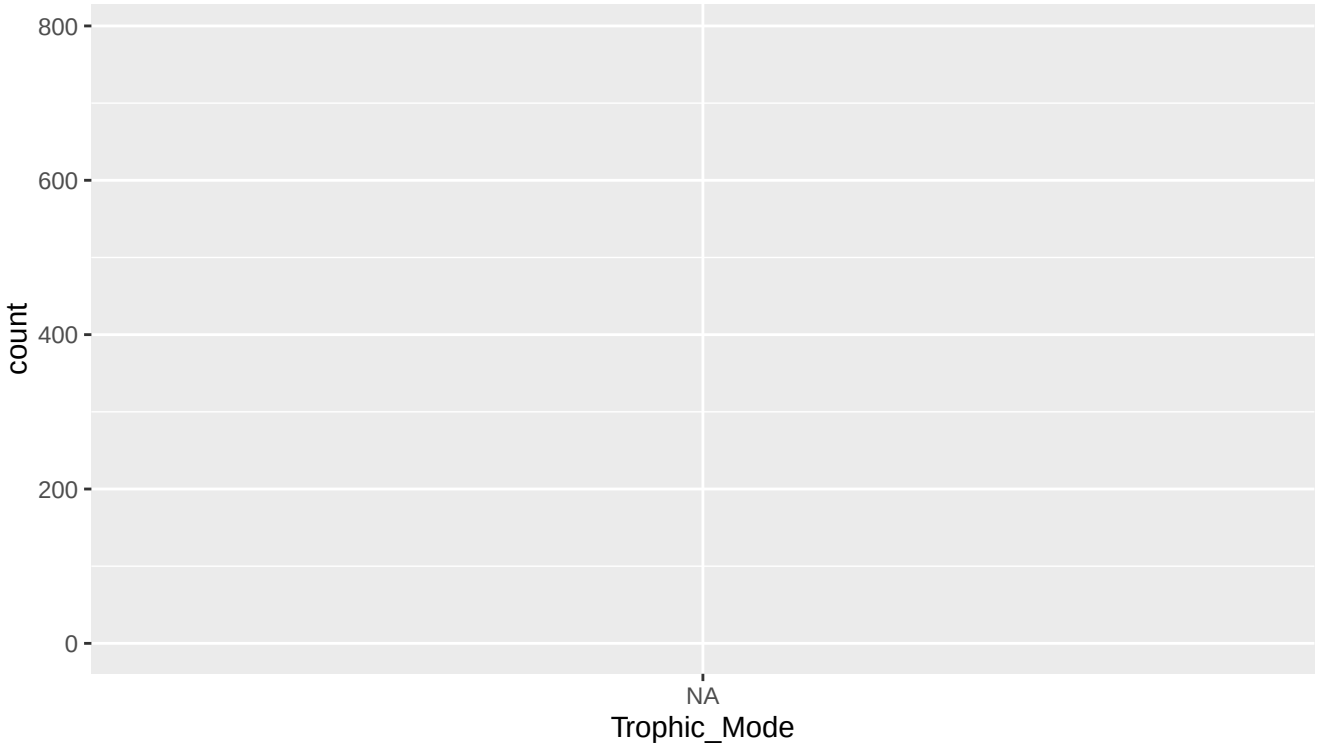

Figure 4.1: Distribution of OTUs into functional Guild.

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] NA

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()
```

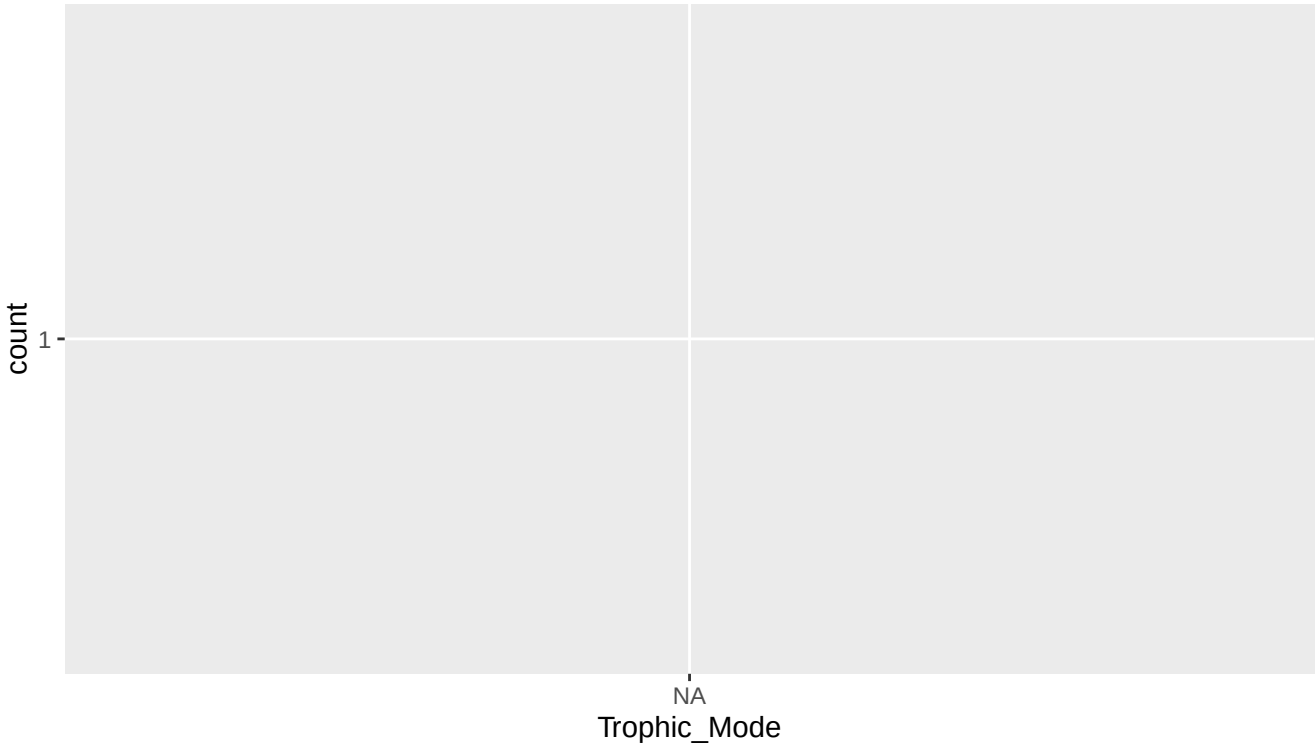

Figure 4.2: Distribution of sequences (log10 transformed) into functional Guild.

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)
```

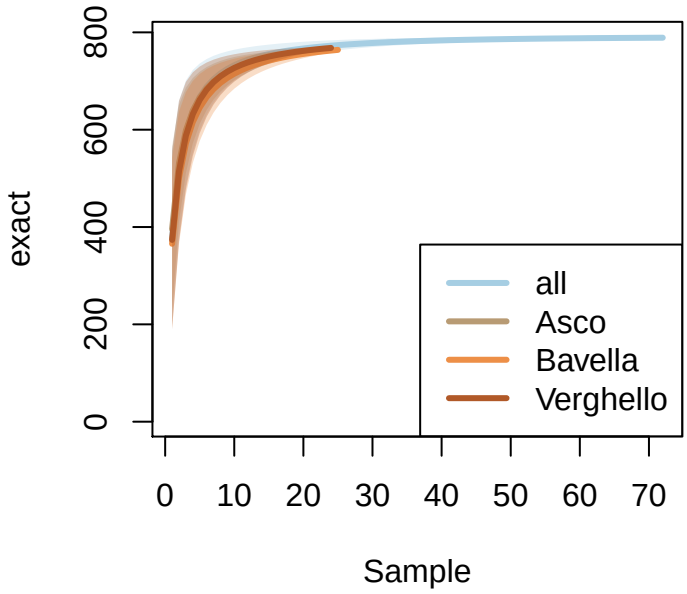

Figure 5.1: Rarefaction curves for each sites. Notes that if singletons were removed, these curves are biaised.

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)
```

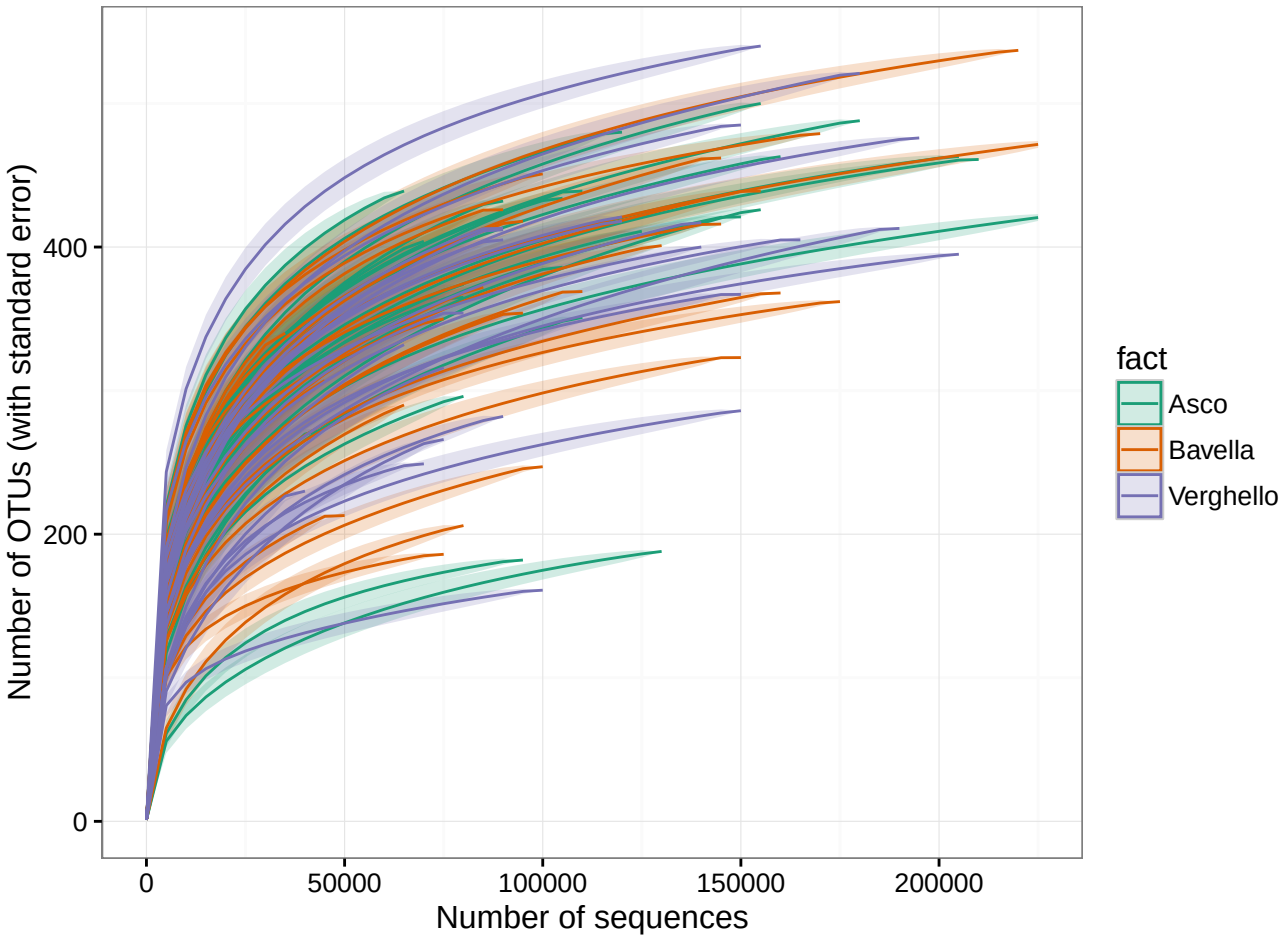

Figure 5.2: Rarefaction curves for each samples using sequences number on x-axes. Notes that if singletons were removed, these curves are biaised.

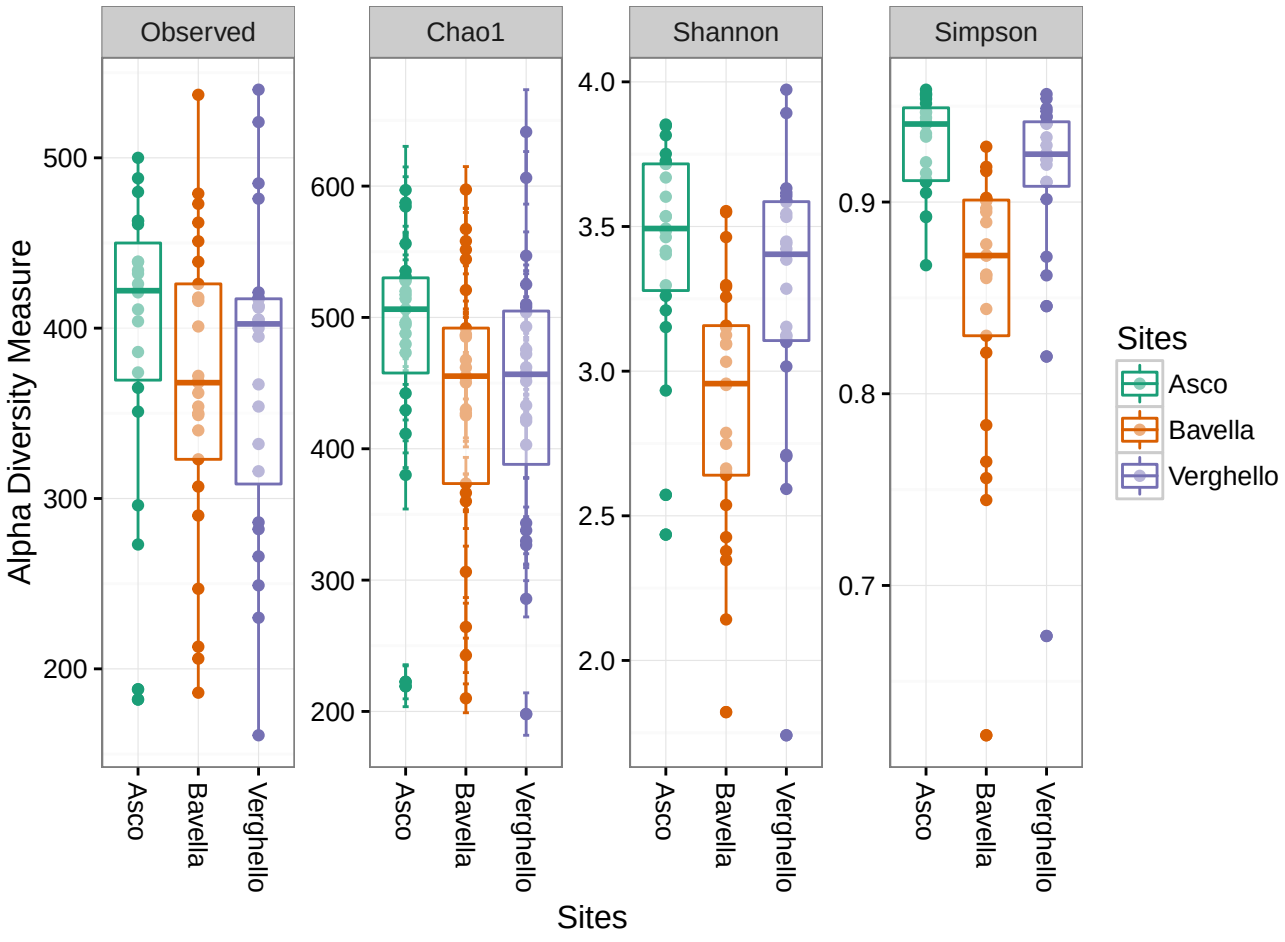

Figure 5.3: Diversity of each sites

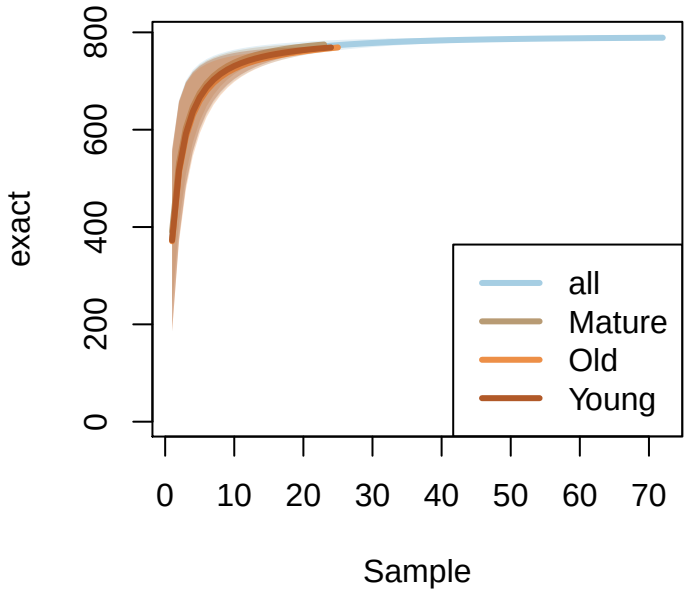

Figure 5.4: Rarefaction curves for each tree age modalities. Notes that if singletons were removed, these curves are biaised.

```
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)	137.1134781	47.9275701	2.8608477	0.0057008
sqrt(readNumbers)	0.7658493	0.1218954	6.2828392	0.0000000
data.f3@sam_data$SitesBavella	-18.6323550	21.0730098	-0.8841810	0.3799081
data.f3@sam_data$SitesVerghello	-20.4021123	21.1161881	-0.9661835	0.3375893
data.f3@sam_data$AgeOld	-11.2020198	21.0175802	-0.5329833	0.5958911
data.f3@sam_data$AgeYoung	-38.3451270	21.5282145	-1.7811569	0.0796320
data.f3@sam_data$ElevationMiddle	34.5733475	21.2819047	1.6245420	0.1091755
data.f3@sam_data$ElevationTop	9.0477061	21.0026184	0.4307894	0.6680689

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))
```

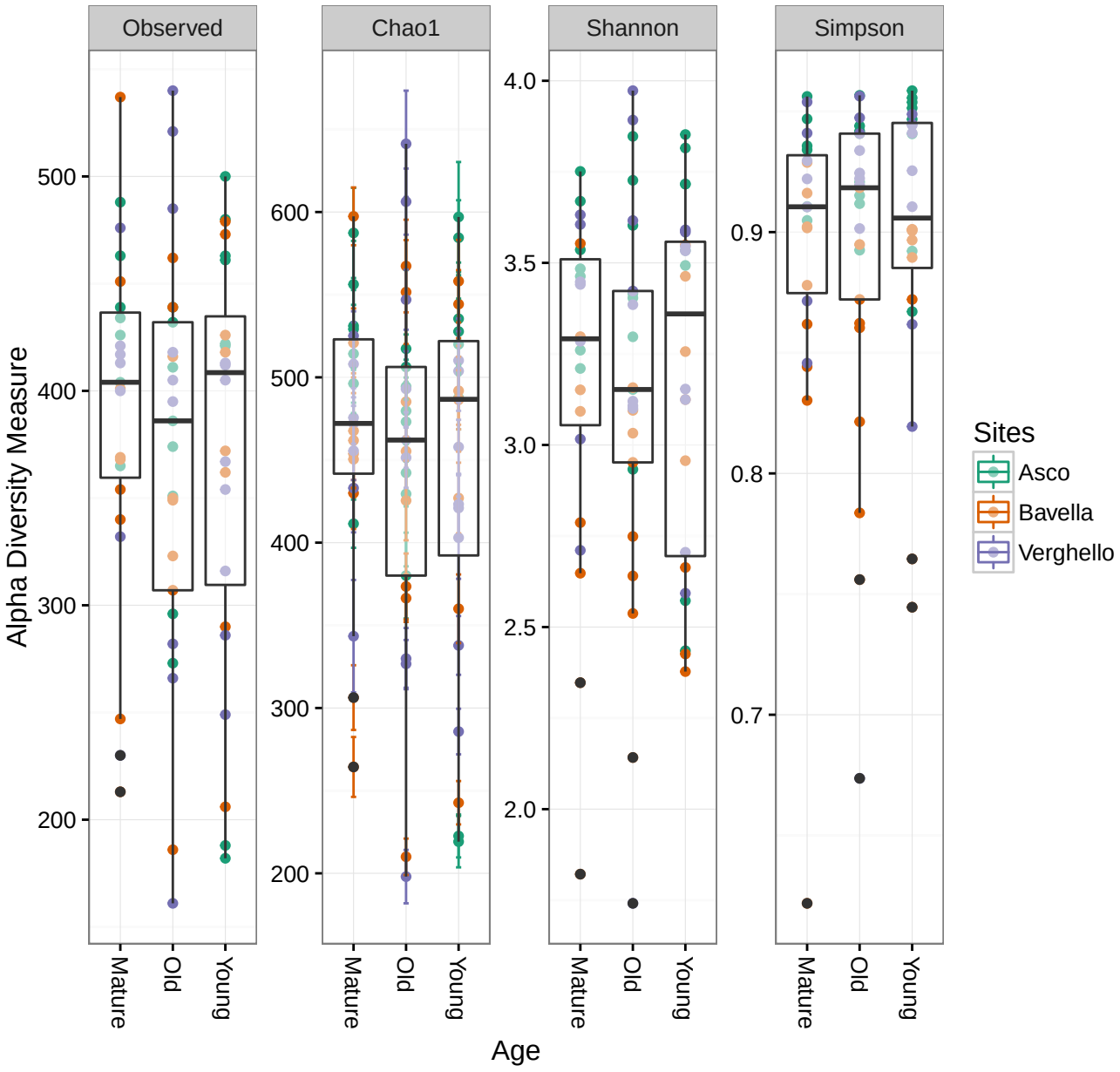

Figure 5.5: Diversity in function of tree age. Color represent sites.

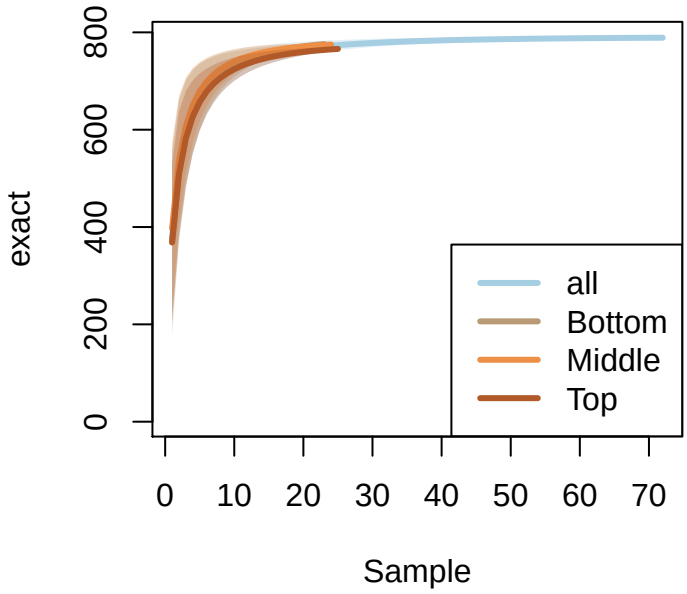

Figure 5.6: Rarefaction curves for each elevation. Notes that if singletons were removed, these curves are biaised.

```
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).

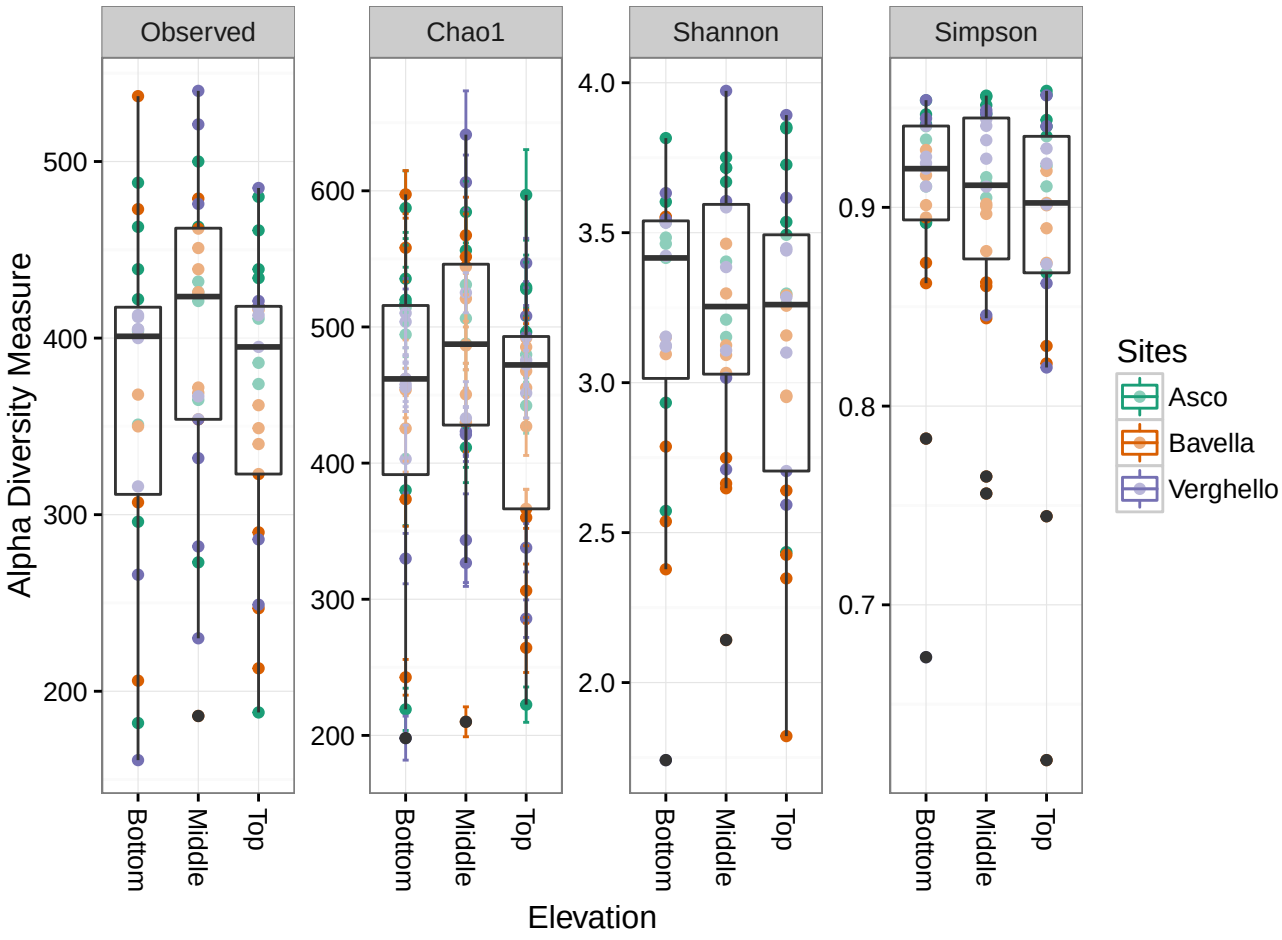

Figure 5.7: Diversity in function of elevation. Color represent sites.

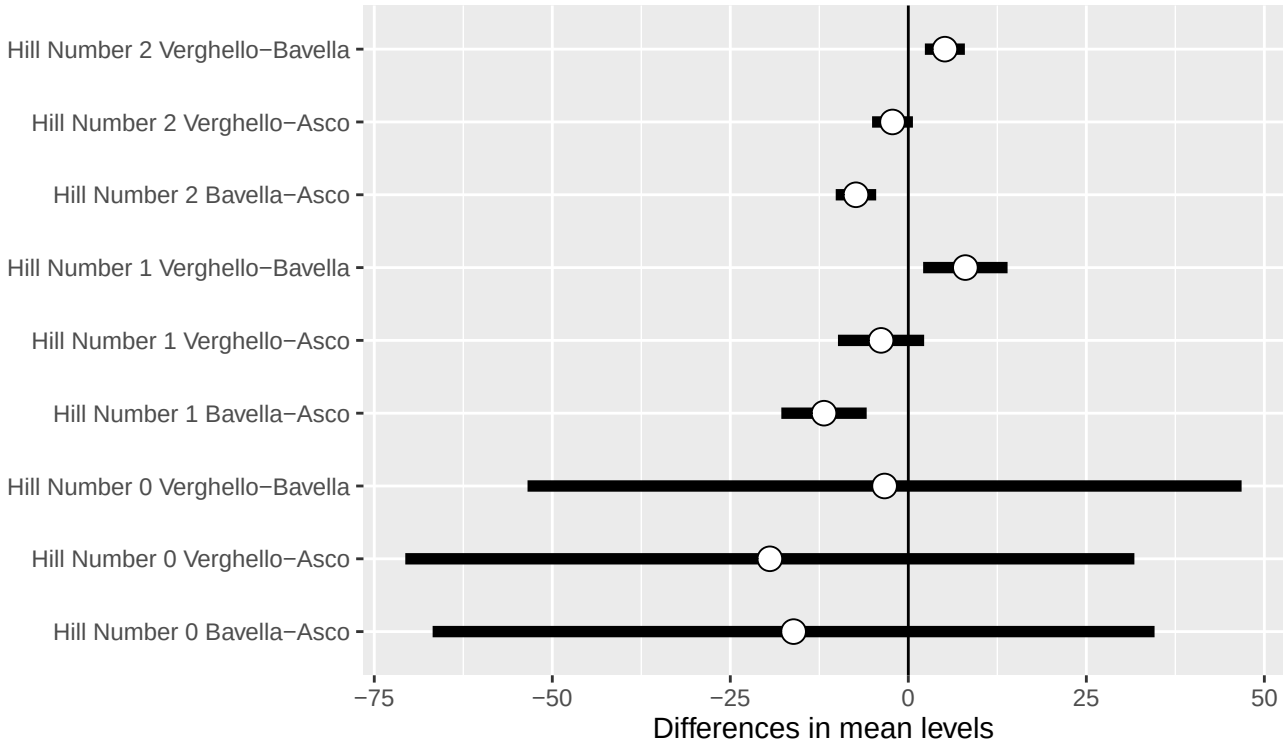

Figure 5.8: Results of the Tuckey HSD testing for differences in mean Hill numbers among pairs of modalities

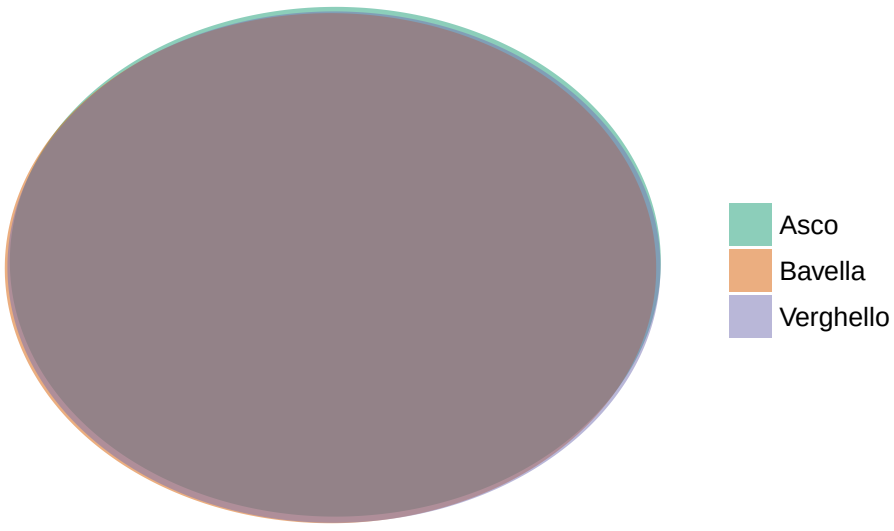

Figure 6.1: Venn diagramm ef the distribution of OTUs among Sites

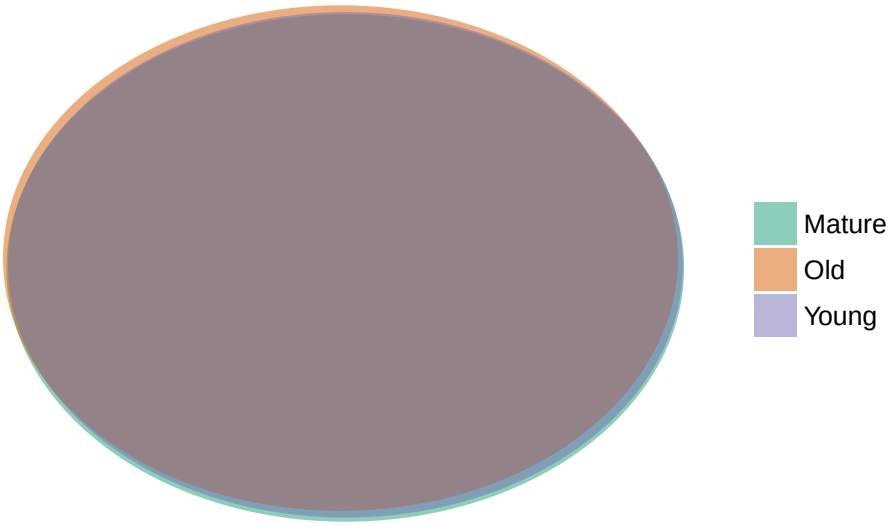

Figure 6.2: Venn diagramm ef the distribution of OTUs among host age

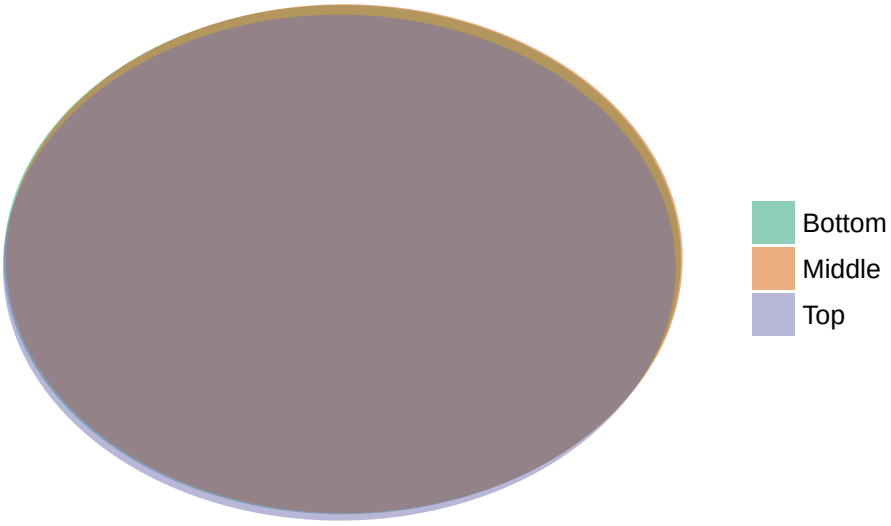

Figure 6.3: Venn diagramm ef the distribution of OTUs among elevation of samples

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	10.1037010	5.8620683	1.7235727	0.0896132
sqrt(readNumbers)	0.0645029	0.0149091	4.3263981	0.0000542
data.f3@sam_data$SitesBavella	-12.1403744	2.5774606	-4.7102076	0.0000138
data.f3@sam_data$SitesVerghello	-3.9386385	2.5827417	-1.5249835	0.1321900
data.f3@sam_data$AgeOld	0.0377035	2.5706809	0.0146668	0.9883437
data.f3@sam_data$AgeYoung	-1.6707637	2.6331371	-0.6345145	0.5280068
data.f3@sam_data$ElevationMiddle	2.7433230	2.6030107	1.0539039	0.2958896
data.f3@sam_data$ElevationTop	0.8276555	2.5688509	0.3221890	0.7483596

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)	7.1131585	2.7945596	2.5453594	0.0133369
sqrt(readNumbers)	0.0251023	0.0071075	3.5318126	0.0007714
data.f3@sam_data$SitesBavella	-7.4811867	1.2287245	-6.0885792	0.0000001
data.f3@sam_data$SitesVerghello	-2.2675179	1.2312422	-1.8416506	0.0701608
data.f3@sam_data$AgeOld	0.1544734	1.2254925	0.1260501	0.9000874
data.f3@sam_data$AgeYoung	0.2371272	1.2552666	0.1889058	0.8507644
data.f3@sam_data$ElevationMiddle	0.7288202	1.2409048	0.5873297	0.5590488
data.f3@sam_data$ElevationTop	-0.3294303	1.2246202	-0.2690061	0.7887903

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 0.2232139
## Run 1 stress 0.2232141
## ... Procrustes: rmse 0.000129336  max resid 0.0006849163
## ... Similar to previous best
## Run 2 stress 0.2232434
## ... Procrustes: rmse 0.005351474  max resid 0.03370459
## Run 3 stress 0.2232202
```

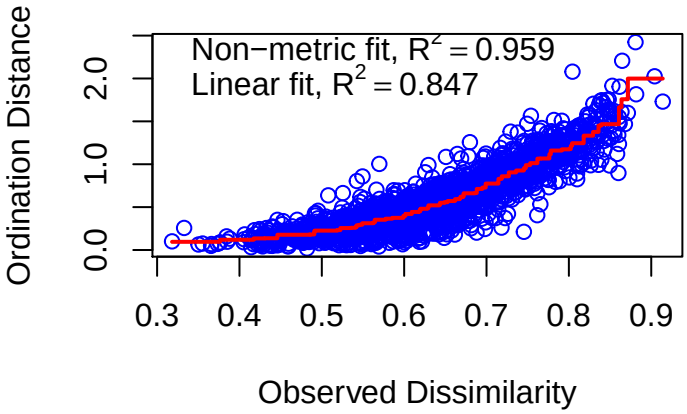

Figure 6.4: Stress plot of the NMDS

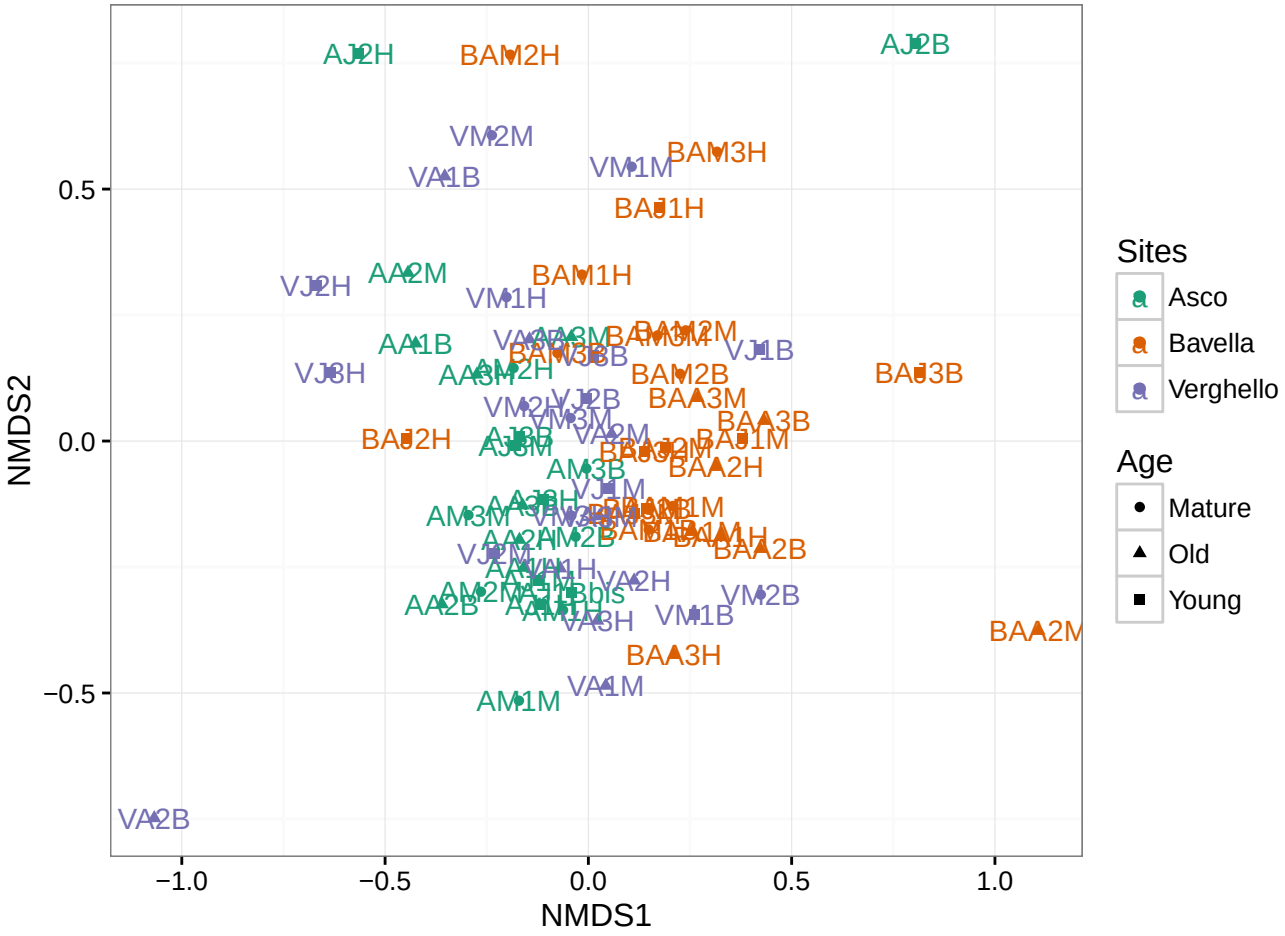

Figure 6.5: NMDS of OTU table. Colors represent sites and shape the age of tree.

```
## ... Procrustes: rmse 0.003189504  max resid 0.02115532
## Run 4 stress 0.2401054
## Run 5 stress 0.2232189
## ... Procrustes: rmse 0.003172698  max resid 0.02128047
## Run 6 stress 0.2232188
## ... Procrustes: rmse 0.00315981  max resid 0.02129892
## Run 7 stress 0.2232022
## ... New best solution
## ... Procrustes: rmse 0.005475788  max resid 0.0333221
## Run 8 stress 0.2232202
## ... Procrustes: rmse 0.005125946  max resid 0.03803663
## Run 9 stress 0.2231817
## ... New best solution
## ... Procrustes: rmse 0.003153037  max resid 0.02133596
## Run 10 stress 0.2401047
## Run 11 stress 0.2232005
## ... Procrustes: rmse 0.003139787  max resid 0.02140712
## Run 12 stress 0.2398462
## Run 13 stress 0.2239184
## Run 14 stress 0.2611217
## Run 15 stress 0.2398473
## Run 16 stress 0.2232638
## ... Procrustes: rmse 0.005691279  max resid 0.03775485
## Run 17 stress 0.2231829
## ... Procrustes: rmse 0.0003537048  max resid 0.00254881
## ... Similar to previous best
## Run 18 stress 0.2232637
## ... Procrustes: rmse 0.005687143  max resid 0.03776935
## Run 19 stress 0.2568044
## Run 20 stress 0.2233548
## ... Procrustes: rmse 0.006289332  max resid 0.05062064
## *** Solution reached

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)
```

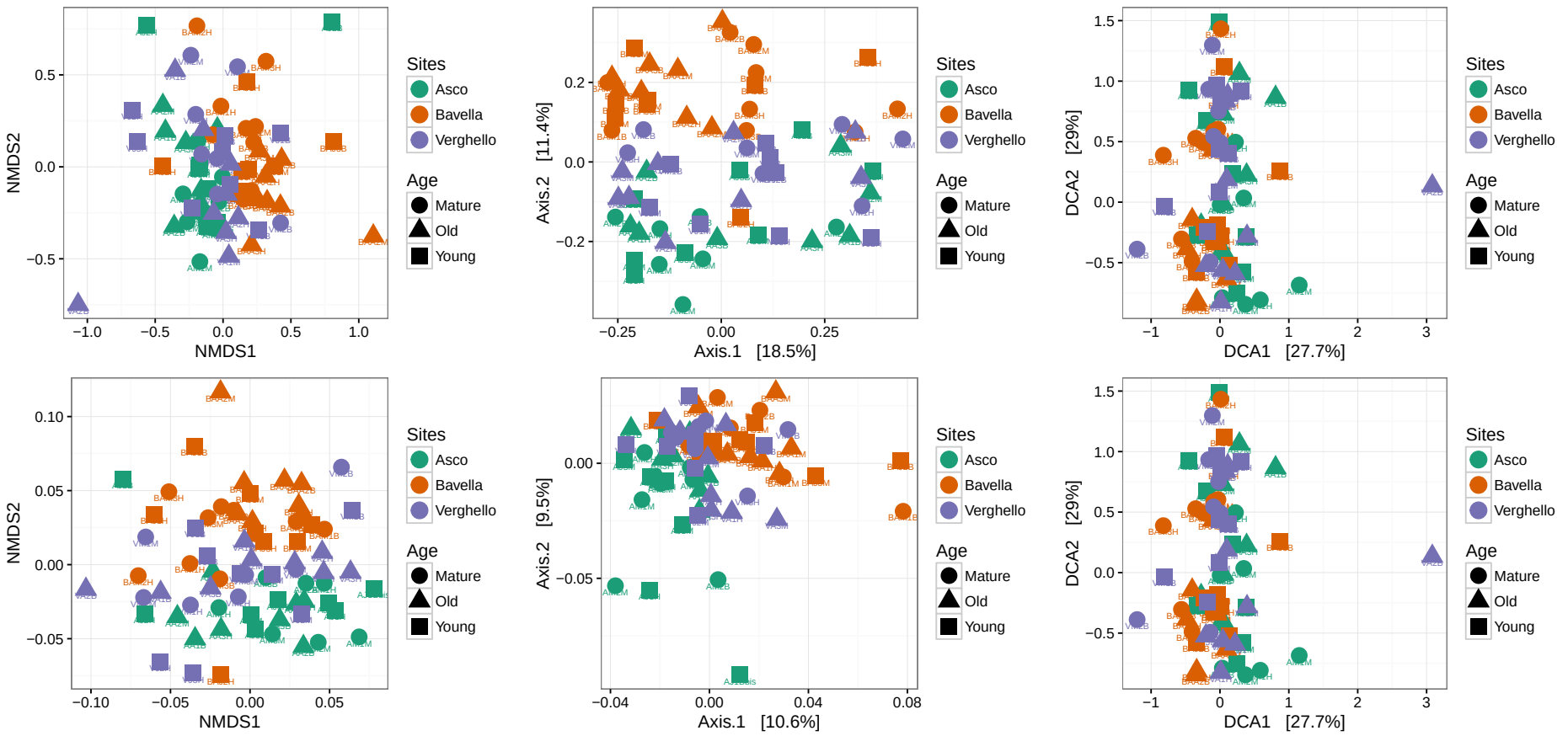

Figure 6.6: Comparison of different distances (bray (up) and gower (bottom)) and ordination methods (NMDS (left), PCoA (center) and DCA (right)).

```
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 447 OTUs present in more than 30 sample, the Permanova results is the following:

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

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

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Sites	2	1.84	0.92	4.40	0.11	0.0001
Age	2	0.57	0.29	1.37	0.03	0.0855
Elevation	2	0.46	0.23	1.11	0.03	0.2887
Sites:Age	4	1.48	0.37	1.77	0.09	0.0010
Sites:Elevation	4	0.78	0.20	0.93	0.04	0.6172
Age:Elevation	4	1.07	0.27	1.28	0.06	0.0749
Sites:Age:Elevation	8	1.77	0.22	1.06	0.10	0.3212
Residuals	45	9.43	0.21		0.54	
Total	71	17.41			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	1.83	0.92	4.46	0.11	0.0001
Age	2	0.56	0.28	1.36	0.03	0.0910
Elevation	2	0.46	0.23	1.11	0.03	0.2958
Sites:Age	4	1.47	0.37	1.79	0.09	0.0010
Sites:Elevation	4	0.76	0.19	0.93	0.04	0.6286
Age:Elevation	4	1.06	0.26	1.29	0.06	0.0703
Sites:Age:Elevation	8	1.74	0.22	1.06	0.10	0.3147
Residuals	45	9.23	0.21		0.54	
Total	71	17.11			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	0.50	0.25	3.52	0.09	0.0001
Age	2	0.20	0.10	1.38	0.03	0.0716
Elevation	2	0.20	0.10	1.39	0.03	0.0700
Sites:Age	4	0.44	0.11	1.55	0.08	0.0069
Sites:Elevation	4	0.27	0.07	0.95	0.05	0.5994
Age:Elevation	4	0.40	0.10	1.40	0.07	0.0259
Sites:Age:Elevation	8	0.60	0.08	1.05	0.10	0.3336
Residuals	45	3.23	0.07		0.55	
Total	71	5.85			1.00	

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

6.4 Permanova on sites, host ages and individual trees

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

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

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Sites	2	1.84	0.92	4.90	0.11	0.0001
Age	2	0.57	0.29	1.52	0.03	0.0375
Sites:Age	4	1.48	0.37	1.97	0.09	0.0003
Sites:Age:IndividualTree	18	5.04	0.28	1.49	0.29	0.0001
Residuals	45	8.48	0.19		0.49	
Total	71	17.41			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	1.83	0.92	4.98	0.11	0.0001
Age	2	0.56	0.28	1.52	0.03	0.0377
Sites:Age	4	1.47	0.37	2.00	0.09	0.0006
Sites:Age:IndividualTree	18	4.97	0.28	1.50	0.29	0.0002
Residuals	45	8.28	0.18		0.48	
Total	71	17.11			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	0.50	0.25	3.73	0.09	0.0001
Age	2	0.20	0.10	1.46	0.03	0.0488
Sites:Age	4	0.45	0.11	1.65	0.08	0.0026
Sites:Age:IndividualTree	18	1.66	0.09	1.36	0.28	0.0016
Residuals	45	3.04	0.07		0.52	
Total	71	5.85			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.

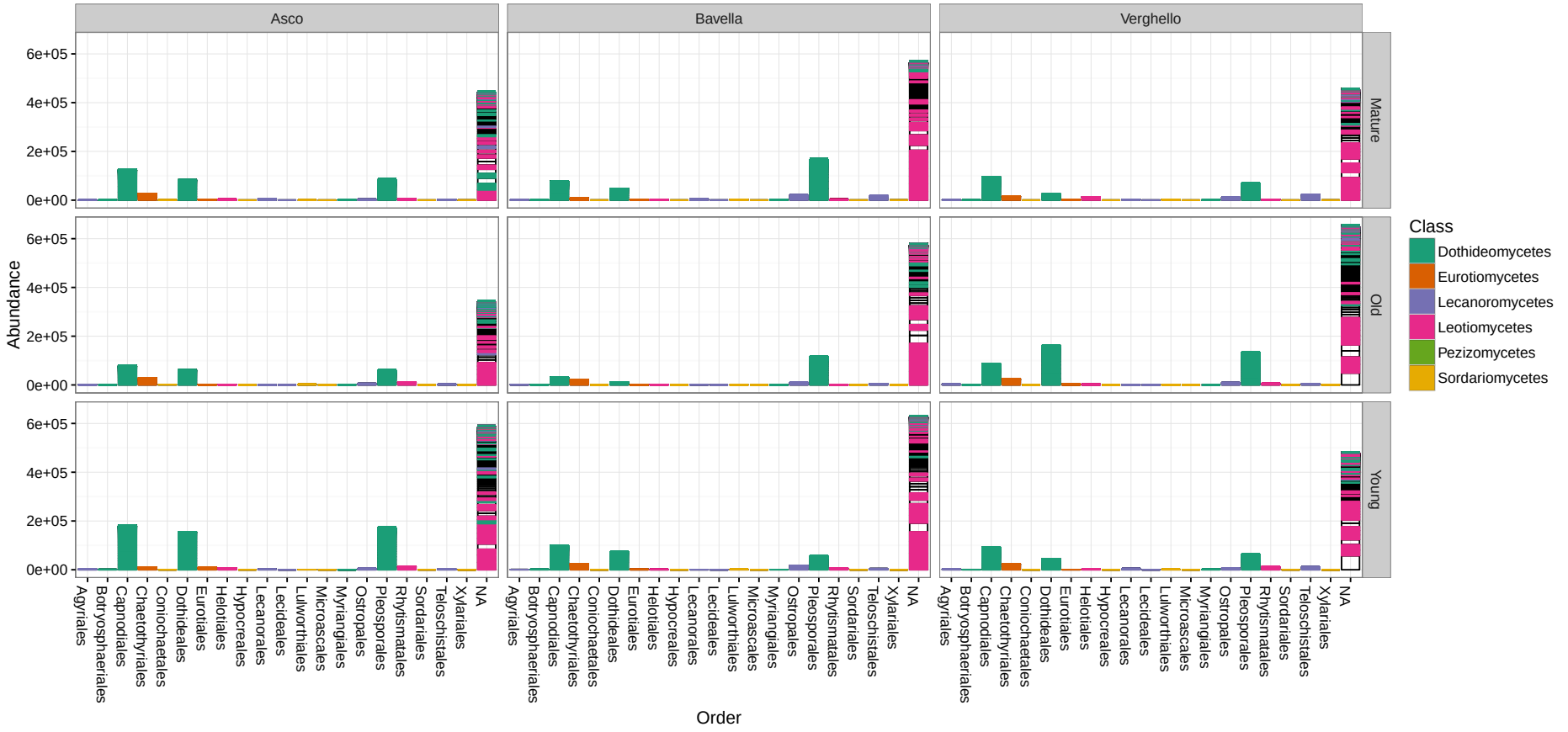

Figure 6.7: Taxonomic distribution of sequences in the different site * age combinaison.

```
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")
```

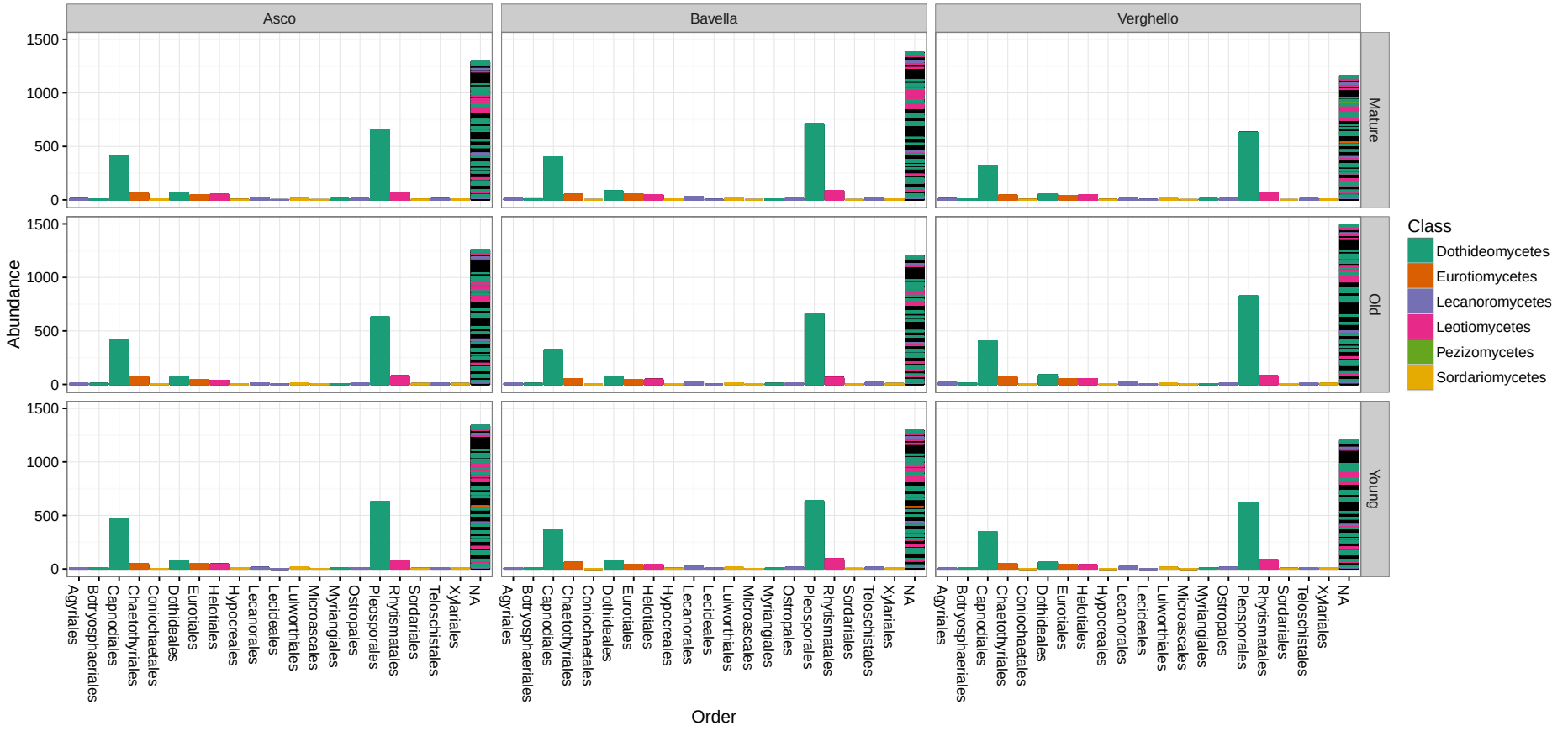

Figure 6.8: Taxonomic distribution of OTUs in the different site * age combinaison.

```
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")
```

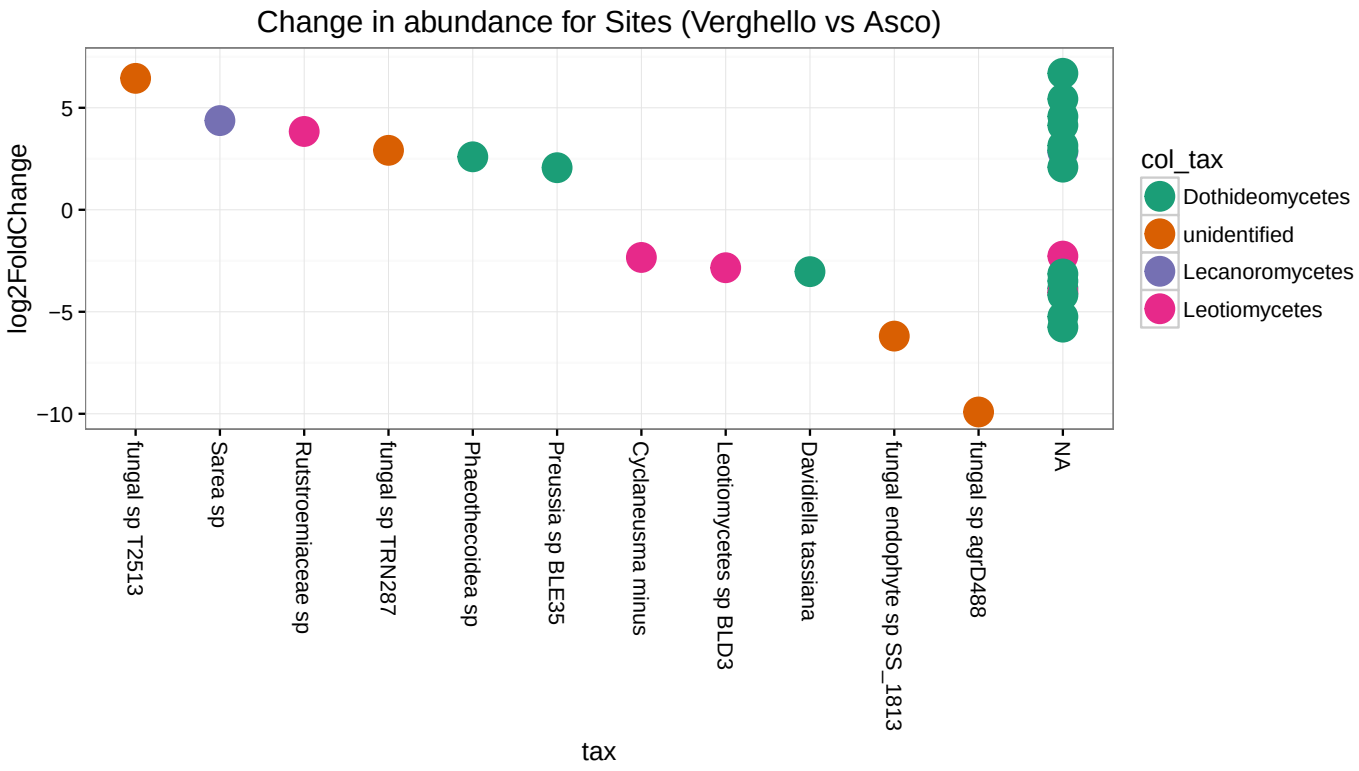

Figure 6.9: OTUs significantly different in terms of abundances between Verghello (positive values) and Asco (negative values)

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
```

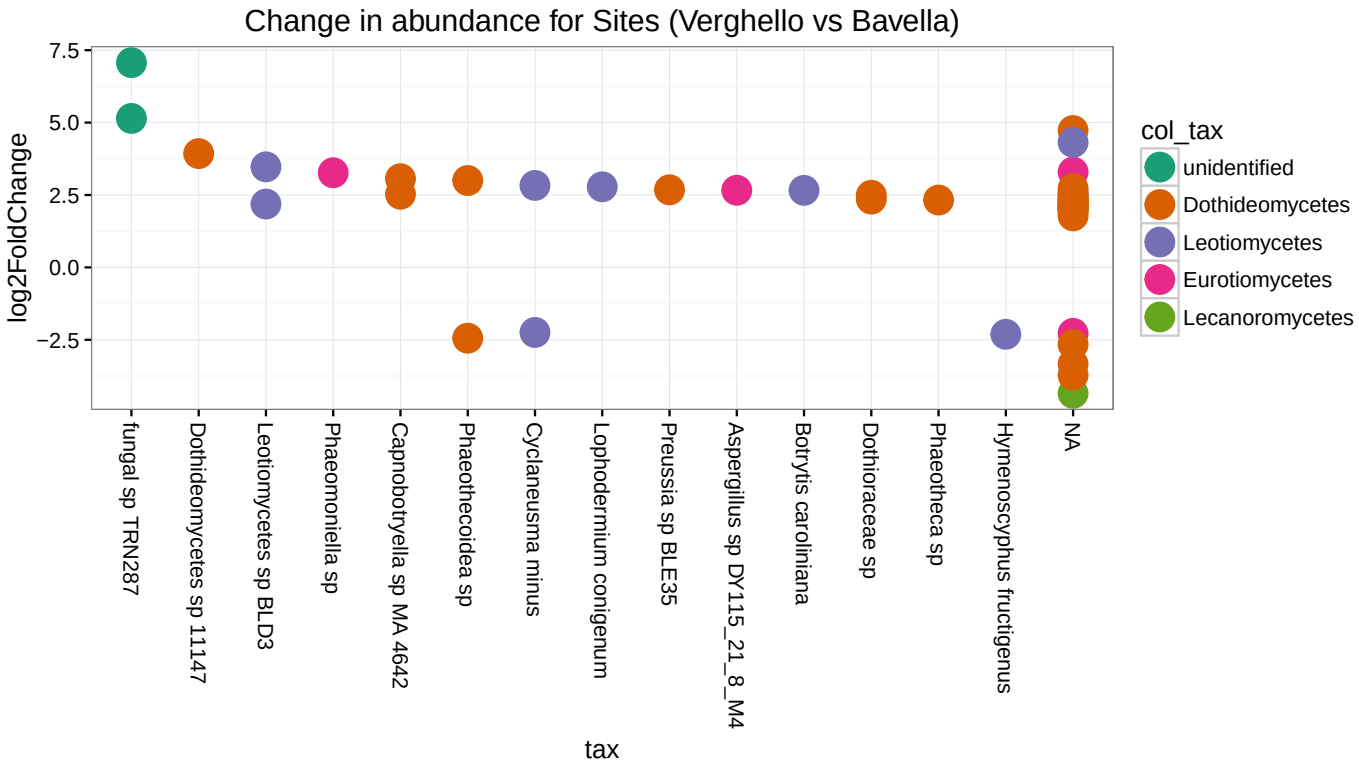

Figure 6.10: OTUs significantly different in terms of abundances between Verghello (positive values) and Bavella (negative values)

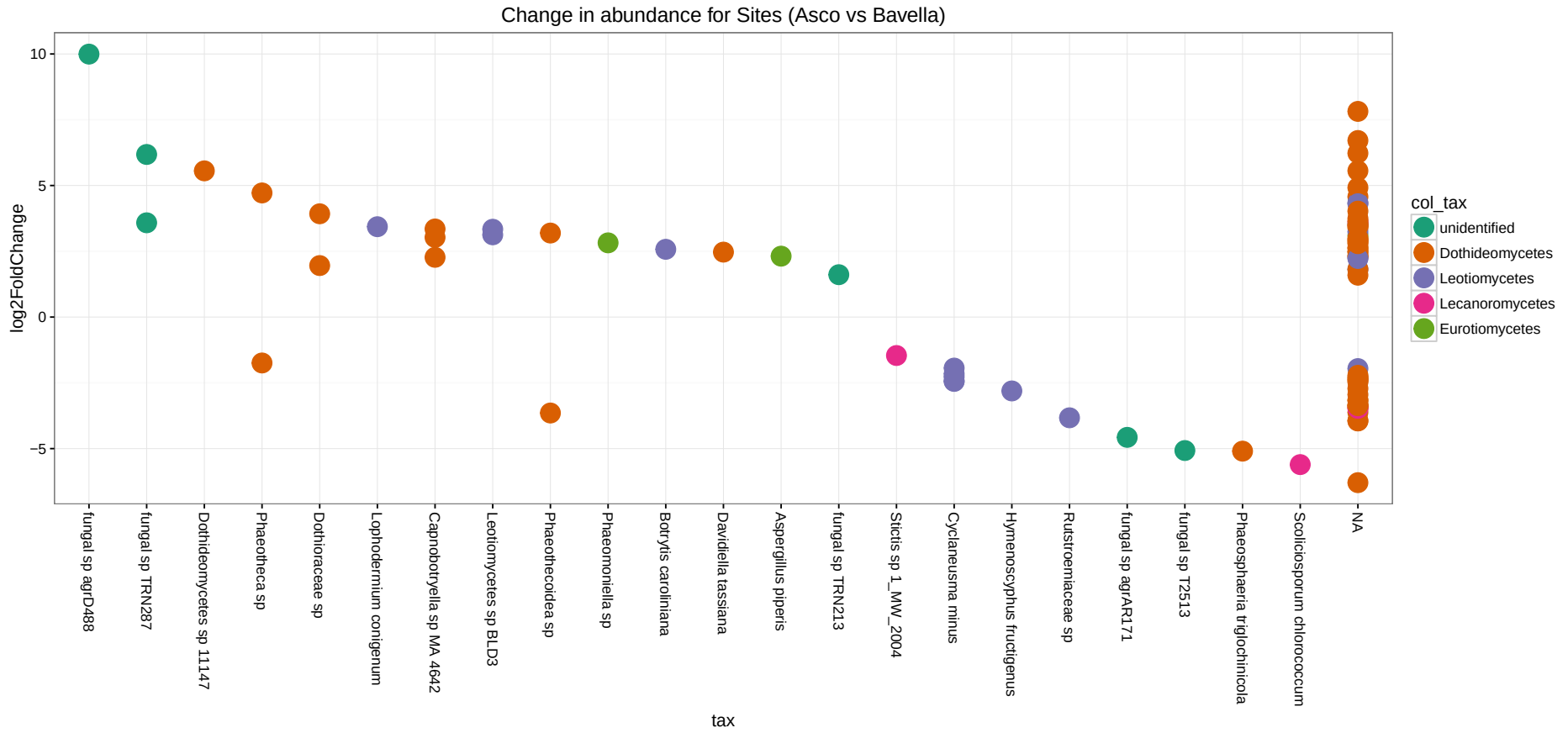

Figure 6.11: OTUs significantly different in terms of abundances between Asco (positive values) and Bavella (negative values)

```
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		Lecanoromycetes	2.83085251825947
2	Verghello vs Asco			2.09846764172597
3	Verghello vs Asco			-3.53903159588761
4	Verghello vs Asco	fungal sp TRN287	unidentified	2.91398352165375
5	Verghello vs Asco			2.64340656887802
6	Verghello vs Asco		Dothideomycetes	2.08917272841107
7	Verghello vs Asco		Dothideomycetes	4.13800547501743
8	Verghello vs Asco	Phaeothecoidea sp	Dothideomycetes	2.59624682559506
9	Verghello vs Asco		Leotiomycetes	-2.27019386227942
10	Verghello vs Asco			3.27406487662501
11	Verghello vs Asco		Dothideomycetes	5.4313886431039
12	Verghello vs Asco	Davidiella tassiana	Dothideomycetes	-3.0325466852721
13	Verghello vs Asco	Preussia sp BLE35	Dothideomycetes	2.06276756639165
14	Verghello vs Asco	Leotiomycetes sp BLD3	Leotiomycetes	-2.84009413085007
15	Verghello vs Asco			-2.79246038710487
16	Verghello vs Asco			-2.7609539773956
17	Verghello vs Asco			-3.10312998236517
18	Verghello vs Asco	Cyclaneusma minus	Leotiomycetes	-2.33694841743386
19	Verghello vs Asco		Dothideomycetes	3.14702241265305
20	Verghello vs Asco		Dothideomycetes	4.57786952257628
21	Verghello vs Asco		Dothideomycetes	-5.75331146501746
22	Verghello vs Asco			3.16009188217377
23	Verghello vs Asco		Dothideomycetes	2.91786260903405
24	Verghello vs Asco	Rutstroemiaceae sp	Leotiomycetes	3.83861833959095
25	Verghello vs Asco		Dothideomycetes	-3.14549517235806
26	Verghello vs Asco		Dothideomycetes	-3.84200772402677
27	Verghello vs Asco	fungal endophyte sp SS_1813	unidentified	-6.19210021229742
28	Verghello vs Asco		Dothideomycetes	-5.23905835484501
29	Verghello vs Asco	fungal sp agrD488	unidentified	-9.9122903179053
30	Verghello vs Asco		Leotiomycetes	-4.04987053986385
31	Verghello vs Asco		Dothideomycetes	-4.18004755763591
32	Verghello vs Asco			7.09494295044463
33	Verghello vs Asco		Dothideomycetes	-3.49203835784756
34	Verghello vs Asco			-8.42063987384668
35	Verghello vs Asco	Sarea sp	Lecanoromycetes	4.37247392420743
36	Verghello vs Asco	fungal sp T2513	unidentified	6.4431429908844
37	Verghello vs Asco		Dothideomycetes	6.68998776604761
38	Verghello vs Bavella	Dothioraceae sp	Dothideomycetes	2.35782415934023
39	Verghello vs Bavella			2.862754171324
40	Verghello vs Bavella	Leotiomycetes sp BLD3	Leotiomycetes	3.4709125325497
41	Verghello vs Bavella			1.98398846890671
42	Verghello vs Bavella		Leotiomycetes	2.36081409779744
43	Verghello vs Bavella			2.43322118108797
44	Verghello vs Bavella			4.16536160727293
45	Verghello vs Bavella	Dothioraceae sp	Dothideomycetes	2.48988745172431
46	Verghello vs Bavella			-2.070083448131
47	Verghello vs Bavella		Dothideomycetes	2.26517997779567
48	Verghello vs Bavella	fungal sp TRN287	unidentified	7.06877026608017
49	Verghello vs Bavella			-2.00772569089781
50	Verghello vs Bavella			3.40772717916203
51	Verghello vs Bavella		Dothideomycetes	2.50005276801135
52	Verghello vs Bavella			-1.34873313957854
53	Verghello vs Bavella	fungal sp TRN287	unidentified	5.14392787668455
54	Verghello vs Bavella		Dothideomycetes	2.04067612898404
55	Verghello vs Bavella		Eurotiomycetes	3.29124578962553
56	Verghello vs Bavella			4.51642010437908
57	Verghello vs Bavella			-1.43246298123001
58	Verghello vs Bavella		Eurotiomycetes	-2.27882985453202
59	Verghello vs Bavella		Dothideomycetes	2.07274948976548
60	Verghello vs Bavella		Dothideomycetes	1.77887611847219
61	Verghello vs Bavella	Phaeomoniella sp	Eurotiomycetes	3.26740812178099
62	Verghello vs Bavella		Leotiomycetes	2.06838322266258
63	Verghello vs Bavella			4.42694821958396
64	Verghello vs Bavella	Aspergillus sp DY115_21_8_M4	Eurotiomycetes	2.66241794838325
65	Verghello vs Bavella		Leotiomycetes	2.01329903114506
66	Verghello vs Bavella	Botrytis caroliniana	Leotiomycetes	2.66052422131409
67	Verghello vs Bavella	Preussia sp BLE35	Dothideomycetes	2.67900181943794
68	Verghello vs Bavella		Leotiomycetes	2.58458235989271
69	Verghello vs Bavella	Phaeotheca sp	Dothideomycetes	2.32566129784107
70	Verghello vs Bavella			-2.02135678557586
71	Verghello vs Bavella		Dothideomycetes	2.15773494931643
72	Verghello vs Bavella			1.86515308621872
73	Verghello vs Bavella		Dothideomycetes	1.99116118442191
74	Verghello vs Bavella			-2.5510050081996
75	Verghello vs Bavella			1.85498903353615
76	Verghello vs Bavella			3.12118452405331
77	Verghello vs Bavella	Leotiomycetes sp BLD3	Leotiomycetes	2.18700430493817
78	Verghello vs Bavella	Phaeothecoidea sp	Dothideomycetes	-2.44490298566578
79	Verghello vs Bavella		Dothideomycetes	-3.34137579992283
80	Verghello vs Bavella			-3.73510347715978
81	Verghello vs Bavella	Lophodermium conigenum	Leotiomycetes	2.78212817325176
82	Verghello vs Bavella	Capnobotryella sp MA 4642	Dothideomycetes	3.06197074179606
83	Verghello vs Bavella			-2.78491066224475
84	Verghello vs Bavella		Dothideomycetes	1.9147867145343
85	Verghello vs Bavella		Dothideomycetes	2.39139120535447
86	Verghello vs Bavella		Dothideomycetes	2.20643534372066
87	Verghello vs Bavella		Dothideomycetes	2.06718122872638
88	Verghello vs Bavella		Dothideomycetes	2.72226342503206
89	Verghello vs Bavella			3.03163677944964
90	Verghello vs Bavella			-1.21033755087119
91	Verghello vs Bavella			-1.99342799280376
92	Verghello vs Bavella			2.81721215259457
93	Verghello vs Bavella		Dothideomycetes	2.14954141833794
94	Verghello vs Bavella			2.57532153514245
95	Verghello vs Bavella	Phaeothecoidea sp	Dothideomycetes	2.9991100185875
96	Verghello vs Bavella			2.46353250084452
97	Verghello vs Bavella	Cyclaneusma minus	Leotiomycetes	2.82997481254788
98	Verghello vs Bavella			3.69141098075678
99	Verghello vs Bavella		Dothideomycetes	2.47992947796516
100	Verghello vs Bavella			4.53923798116494
101	Verghello vs Bavella	Dothideomycetes sp 11147	Dothideomycetes	3.9295104300772
102	Verghello vs Bavella		Dothideomycetes	-2.6603833129523
103	Verghello vs Bavella			-2.03745884793842
104	Verghello vs Bavella		Dothideomycetes	2.32952336673942
105	Verghello vs Bavella	Hymenoscyphus fructigenus	Leotiomycetes	-2.31091811630573
106	Verghello vs Bavella	Capnobotryella sp MA 4642	Dothideomycetes	2.52752967324967
107	Verghello vs Bavella		Lecanoromycetes	-4.34882530967721

	Comparison	Order	Class	log2FoldChange (negative = more on second levels)
1	Verghello vs Asco	Coniochaetales	Sordariomycetes	-5.70138202626319
2	Verghello vs Bavella	Botryosphaeriales	Dothideomycetes	-1.87937519951524
3	Verghello vs Bavella	Helotiales	Leotiomycetes	2.31206551461169
4	Verghello vs Bavella	Incertae sedis	Leotiomycetes	-1.33271140009837
5	Verghello vs Bavella	Ostropales	Lecanoromycetes	-1.22140359771263
6	Verghello vs Bavella	unidentified	Leotiomycetes	1.8058135373692
7	Asco vs Bavella	Coniochaetales	Sordariomycetes	5.50255197627132
8	Asco vs Bavella	Incertae sedis	Leotiomycetes	-1.34776561271851
9	Asco vs Bavella	Ostropales	Lecanoromycetes	-2.20465103548078
10	Asco vs Bavella	Rhytismatales	Leotiomycetes	1.63503979453865
11	Asco vs Bavella	unidentified	Leotiomycetes	1.59300955553037

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

List of Figures

List of Tables