Appendix S9: results after UPARSE clustering discarding 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

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 (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 16:56:58. 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"

#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:         [ 662 taxa and 80 samples ]
## sample_data() Sample Data:       [ 80 samples by 6 sample variables ]
## tax_table()   Taxonomy Table:    [ 662 taxa by 12 taxonomic ranks ]
## refseq()      DNAStringSet:      [ 662 reference sequences ]
```

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

2.3 Filter sample by number of sequences

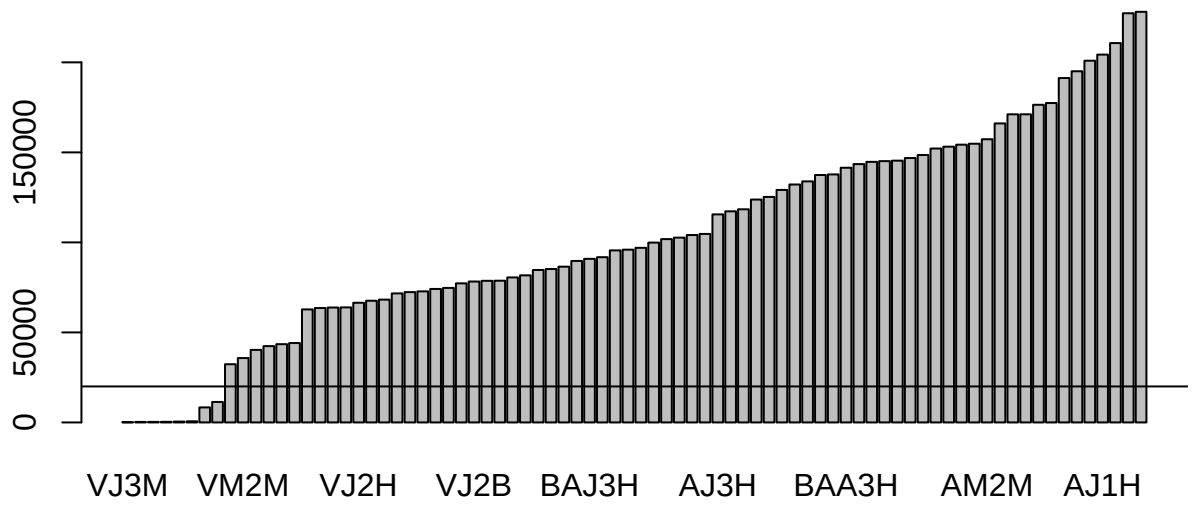

Figure 2.1: Number of sequences by sample

```
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)
```

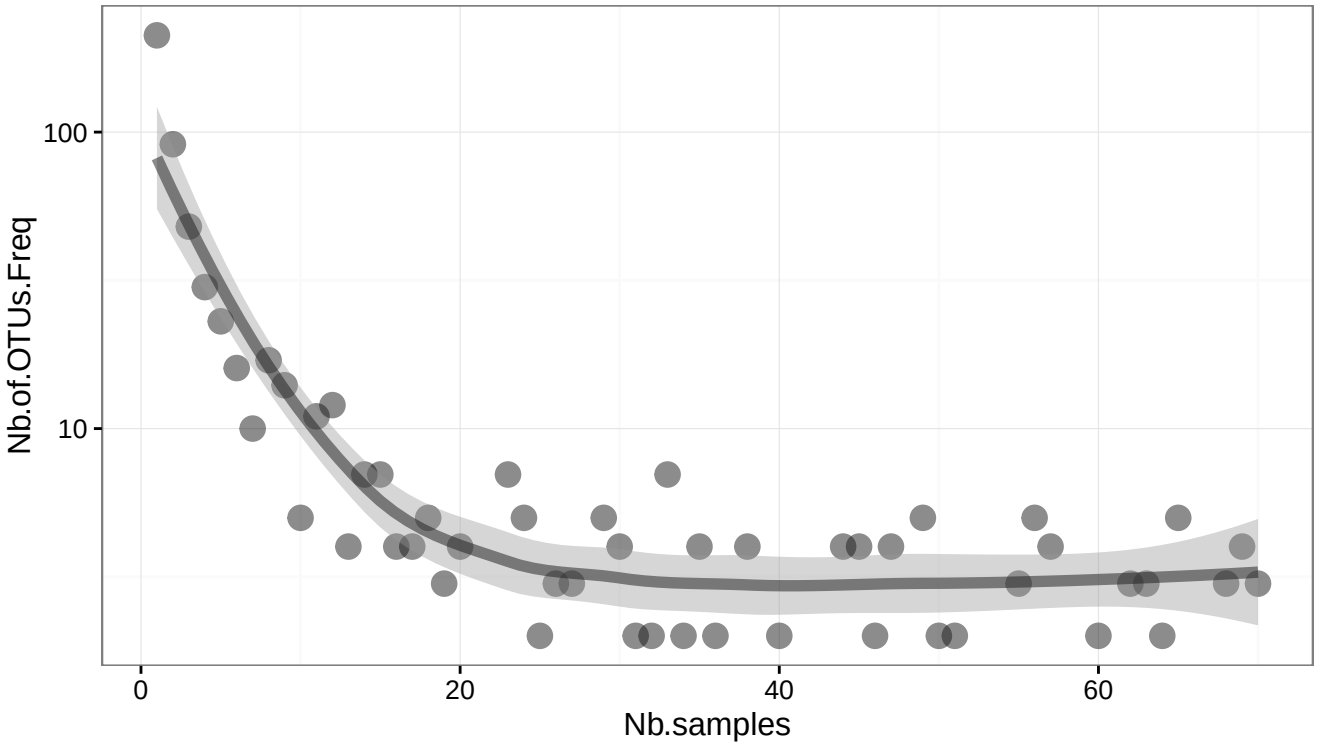

Figure 2.2: Number of OTU present in a given number of samples. Vertical bar illustrate the filtering parameter.

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

```
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 654 on the 654 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])
```

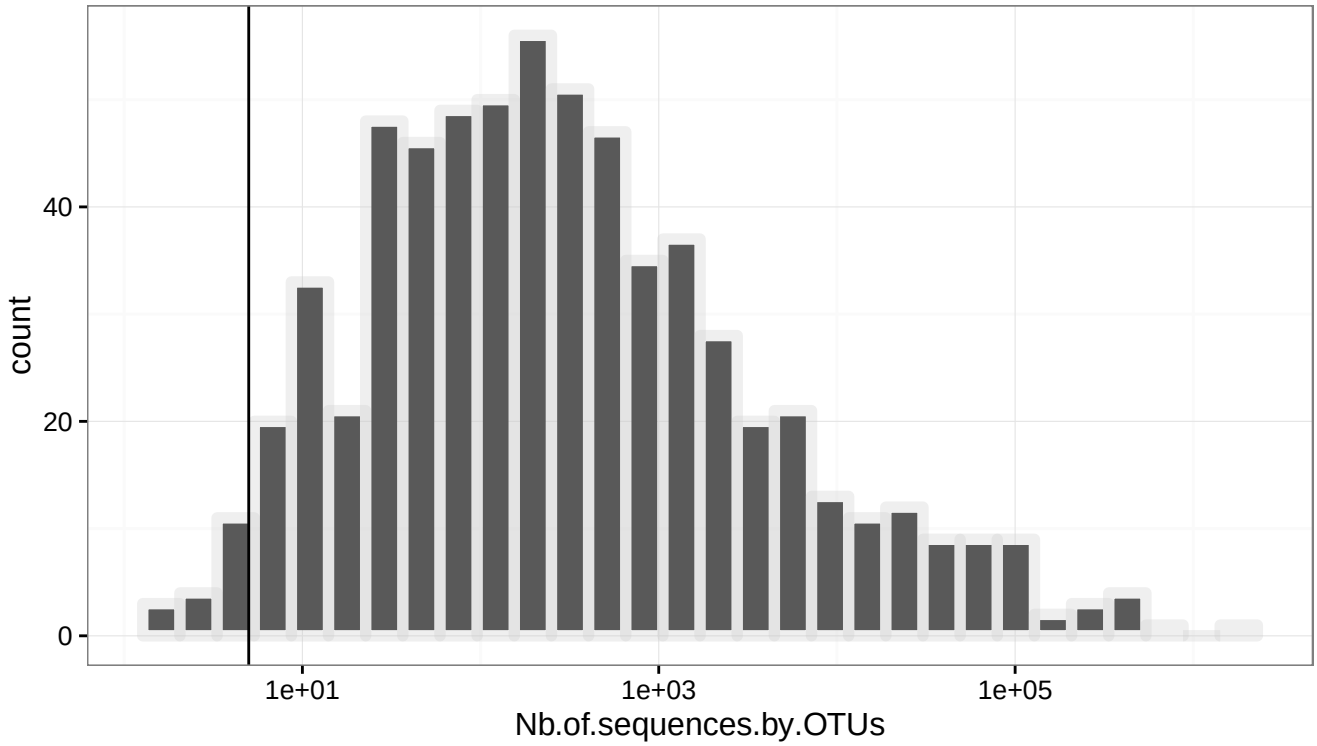

Figure 2.3: Number of sequences by OTU (log10 transformed). Horizontal bar illustrate the filtering parameter.

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##       2.0      46.0     218.5   12600.0    1361.0 2227000.0
```

```
N_seq_otu_min

## [1] 5
```

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

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

2.6 Summary of filtration workflow

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

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

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

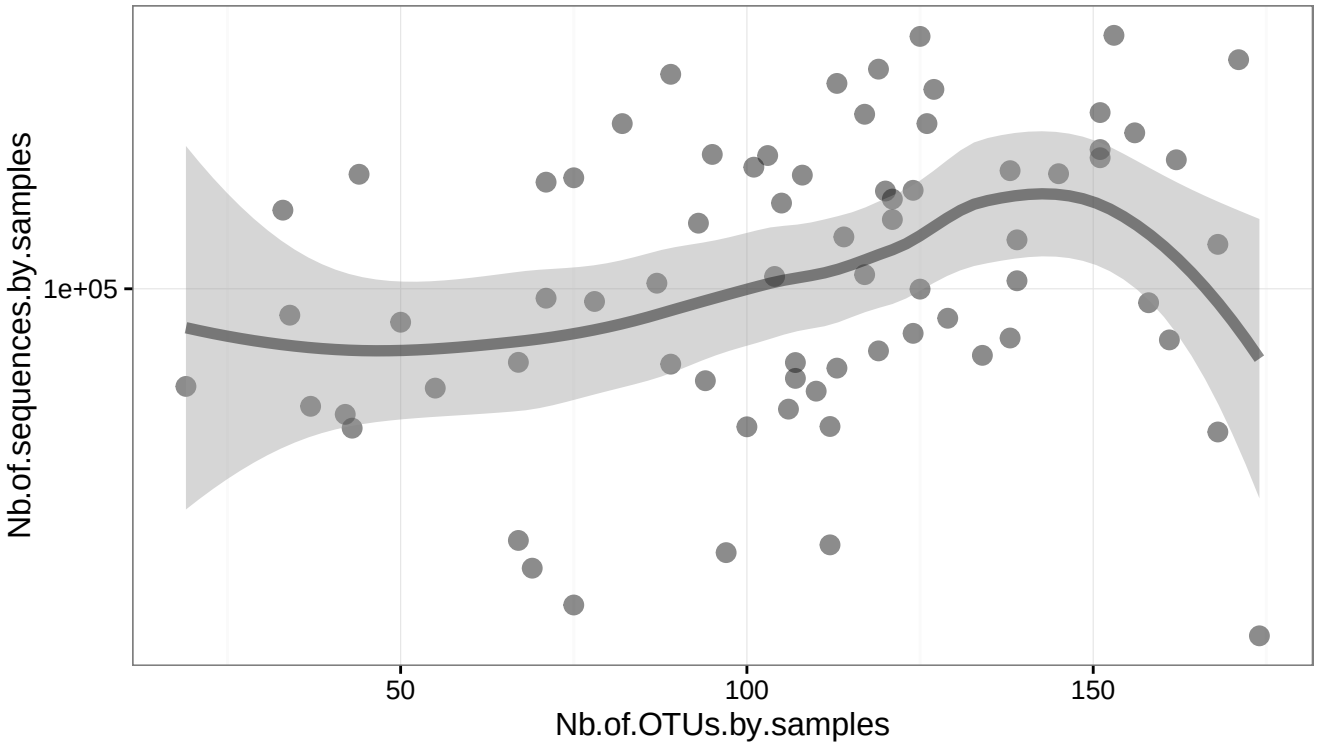

Figure 3.1: Number of OTUs by samples in fonction the number of sequences by samples (log10 axe). The tendency is represented by the line obtain from loess (Local Polynomial Regression Fitting).

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" =
```

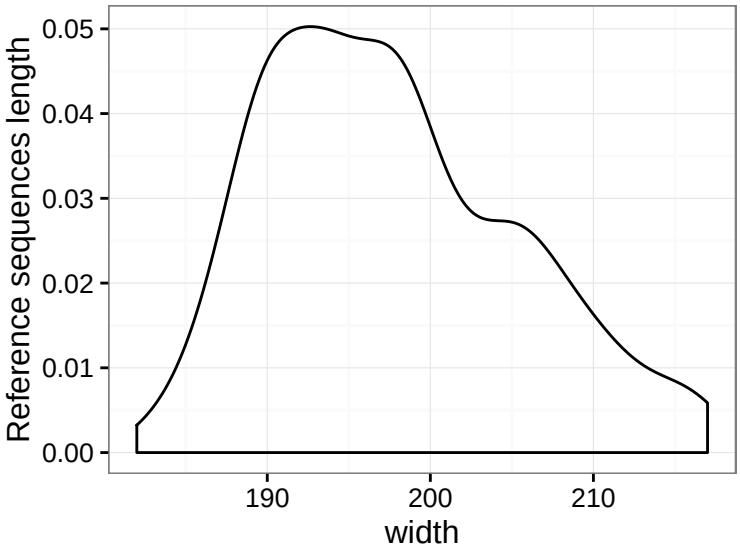

Figure 3.2: Distribution of reference sequences length.

```
                                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)
```

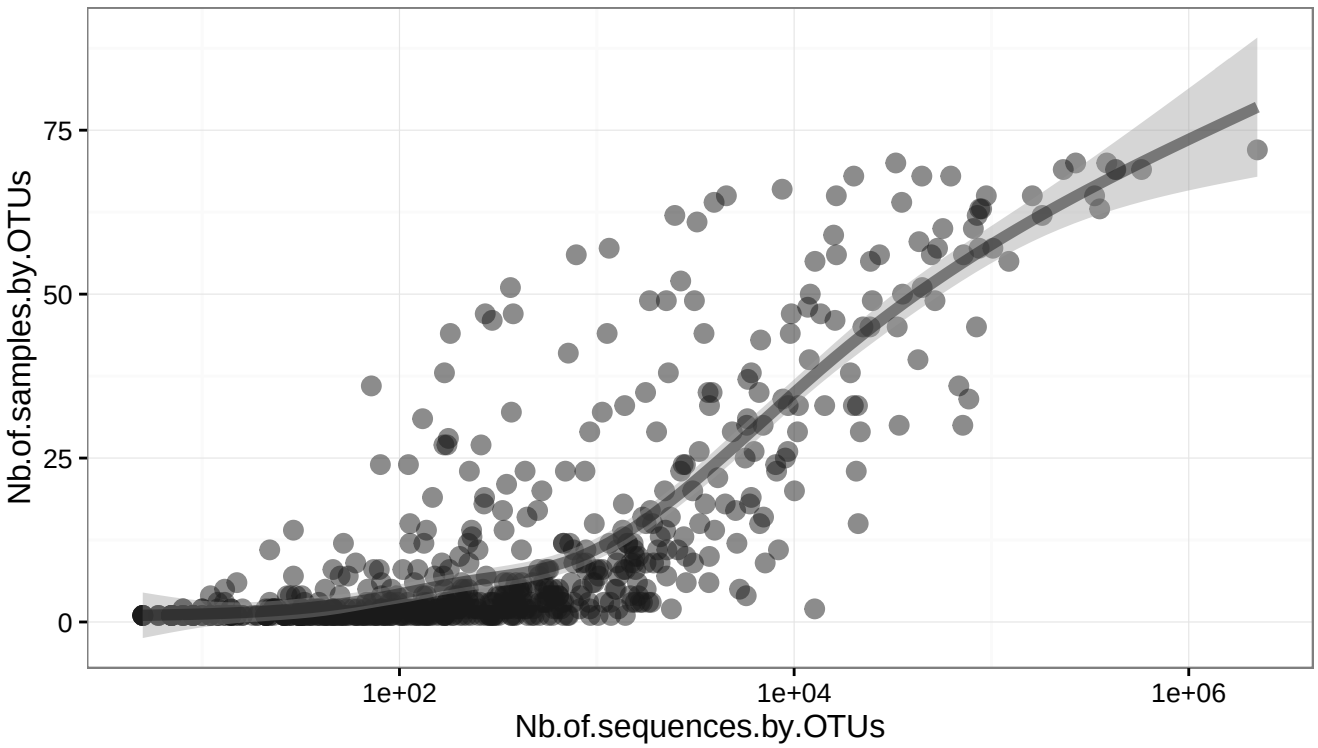

Figure 3.3: Number of sequences by OTUs (log10 axe) in fonction the number of samples where OTUs were found. The tendency is represented by the line obtain from gam (Generalized additive models with integrated smoothness estimation).

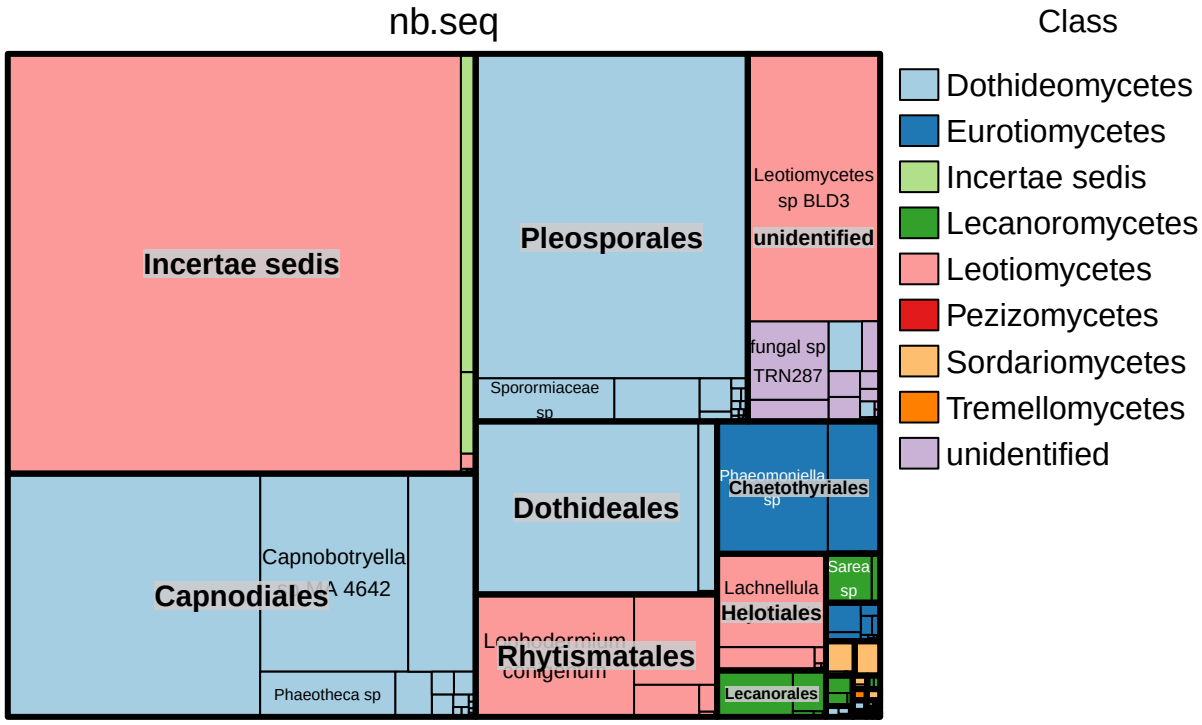

Figure 3.4: Distribution of the number of sequences in the taxonomy. Color represent Class, bold lines delimit Order and thick line delimit species.

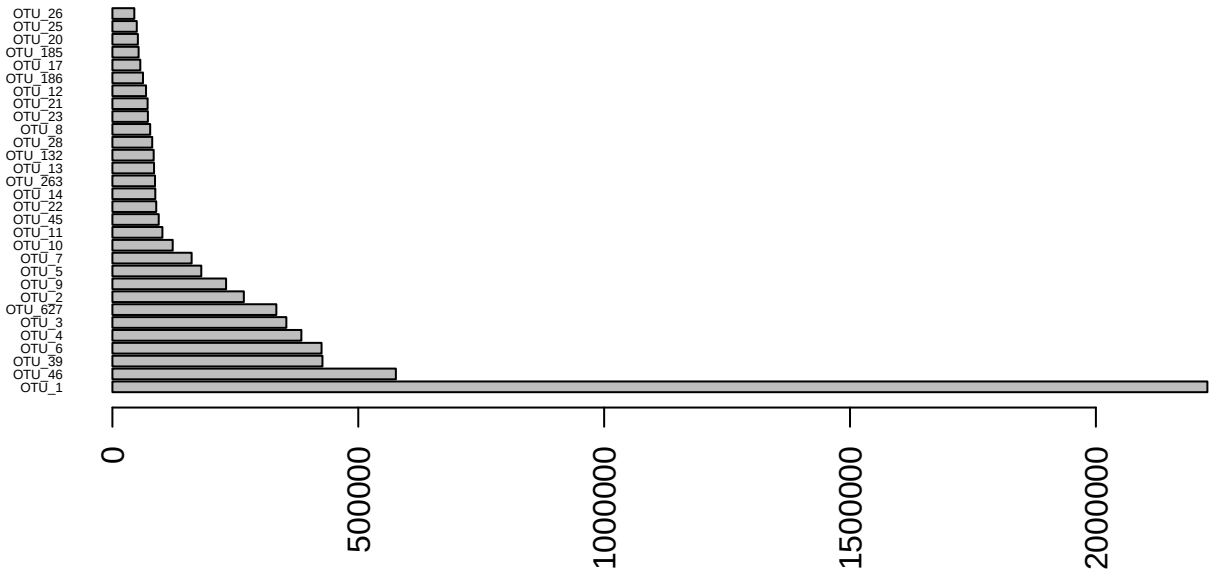

Figure 3.5: Number of sequences of the 30 more abundant OTUs (number of sequences).

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	-	-	2226714
Ascomycota	Dothideomycetes	Pleosporales				-	-	576435
Ascomycota	Leotiomycetes	unidentified	unidentified	unidentified	Leotiomycetes sp BLD3	-	-	427178
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp	-	-	425340
Ascomycota	Dothideomycetes	Capnodiales				-	-	384261
Ascomycota	Dothideomycetes	Capnodiales	Incertae sedis	Capnobotryella	Capnobotryella sp MA 4642	Saprotroph	Undefined Saprotroph	353748
Ascomycota	Dothideomycetes	Pleosporales	Incertae sedis			-	-	333382
						-	-	267278
Ascomycota	Leotiomycetes	Rhytismatales	Rhytismataceae	Lophodermium	Lophodermium conigenum	Pathotroph	Plant Pathogen	231199
						-	-	180724
Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Phaeothecoidea	Phaeothecoidea sp	Saprotroph	Undefined Saprotroph	161122
Ascomycota	Leotiomycetes	Helotiales	Hyaloscyphaceae	Lachnellula	Lachnellula calyciformis	Saprotroph	Undefined Saprotroph	122624
						-	-	101702
Ascomycota	Dothideomycetes	Pleosporales				-	-	94186
Ascomycota	Dothideomycetes	Capnodiales				-	-	89179
Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Phaeomoniella	Phaeomoniella sp	Saprotroph	Undefined Saprotroph	87249
						-	-	86659
Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Phaeomoniella	Phaeomoniella sp	Saprotroph	Undefined Saprotroph	84706
						-	-	83965
Ascomycota	Leotiomycetes	Rhytismatales	Rhytismataceae	Lophodermium	Lophodermium seditiosum	Pathotroph	Plant Pathogen	80980
Ascomycota						-	-	76823
unidentified	unidentified	unidentified	unidentified	unidentified	fungal sp TRN287	-	-	72163
Ascomycota	Dothideomycetes	Pleosporales	Sporormiaceae	unidentified	Sporormiaceae sp	-	-	71602
Ascomycota						-	-	68360
Ascomycota	Dothideomycetes	Capnodiales	Davidiellaceae	Cladosporium		-	-	62215
Ascomycota	Incertae sedis	Incertae sedis	Incertae sedis	Knufia		-	-	56720
Ascomycota	Dothideomycetes	Capnodiales				-	-	53305
Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae			-	-	51765
Ascomycota	Dothideomycetes	Capnodiales	Incertae sedis	Phaeotheca	Phaeotheca sp	-	-	49592

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

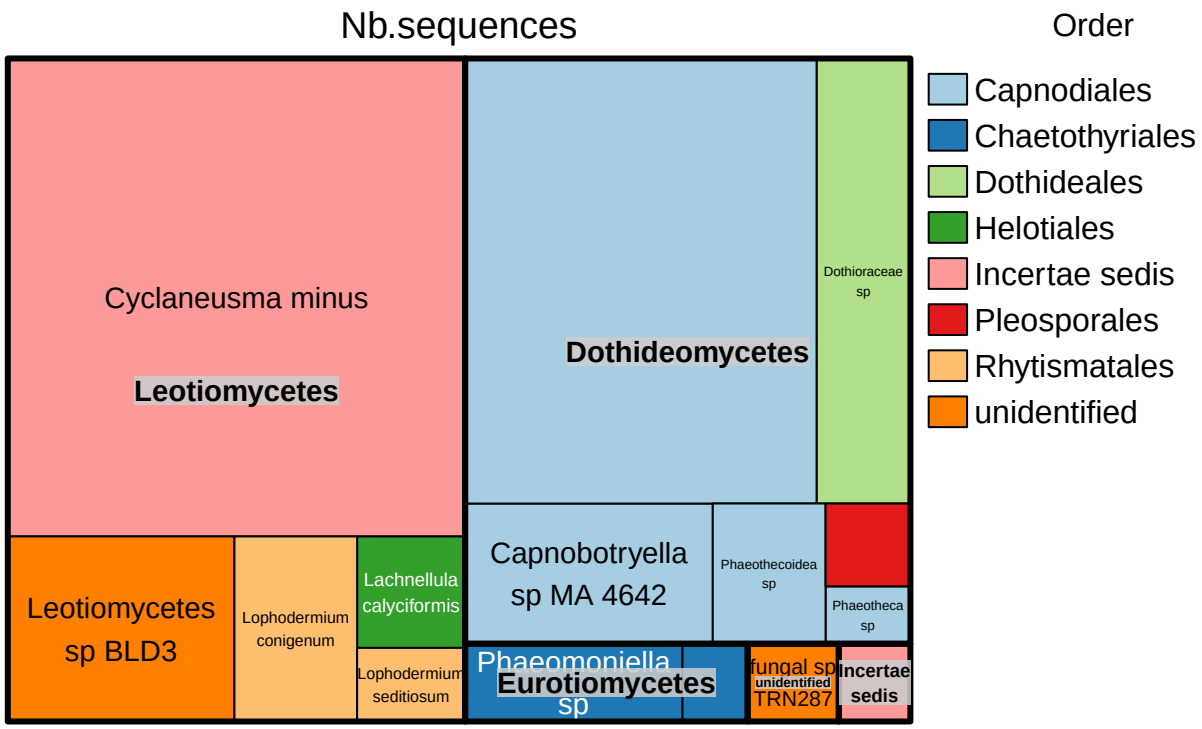

Figure 3.6: Number of sequences of the 30 most abundant OTUs (number of sequences). Colors indicate Order, bold lines delimit Class and thick lines delimit species.

```
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)
```

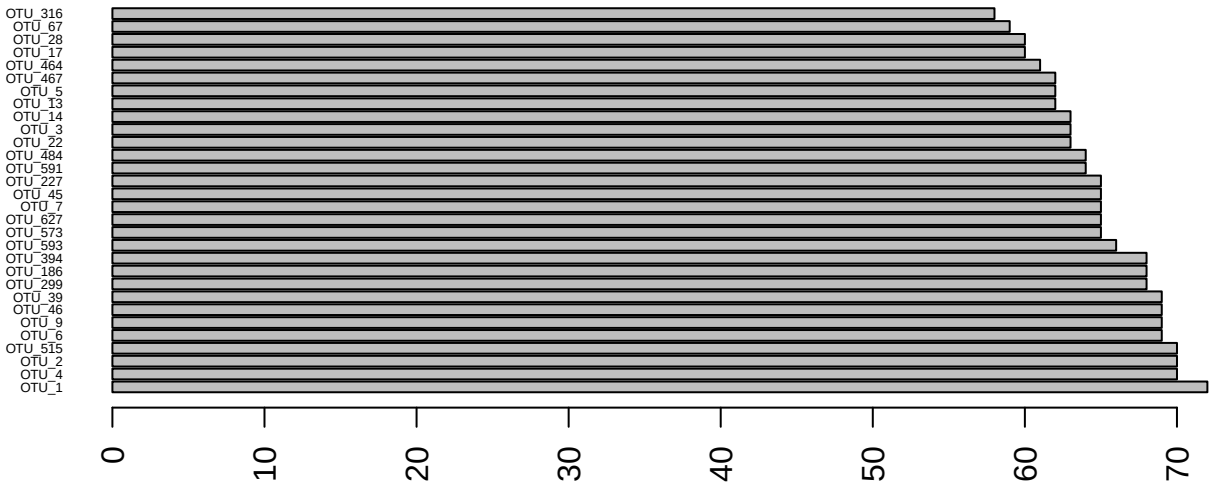

Figure 3.7: Number of samples of the 30 more frequent OTUs (number of samples).

```
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	Dothideomycetes	Capnodiales				-	-	70
						-	-	70
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	70
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp	-	-	69
Ascomycota	Leotiomycetes	Rhytismatales	Rhytismataceae	Lophodermium	Lophodermium conigenum	Pathotroph	Plant Pathogen	69
Ascomycota	Dothideomycetes	Pleosporales				-	-	69
Ascomycota	Leotiomycetes	unidentified	unidentified	unidentified	Leotiomycetes sp BLD3	-	-	69
Ascomycota	Dothideomycetes	Pleosporales	Incertae sedis	Ochrocladosporium	Ochrocladosporium sp	Saprotroph	Undefined Saprotroph	68
Ascomycota	Dothideomycetes	Capnodiales	Davidiellaceae	Cladosporium		-	-	68
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	68
Ascomycota	Dothideomycetes	Capnodiales				-	-	66
Ascomycota	Dothideomycetes	Capnodiales				-	-	65
Ascomycota	Dothideomycetes	Pleosporales	Incertae sedis			-	-	65
Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Phaeothecoidea	Phaeothecoidea sp	Saprotroph	Undefined Saprotroph	65
Ascomycota	Dothideomycetes	Pleosporales				-	-	65
Ascomycota	Leotiomycetes	Rhytismatales	Rhytismataceae	Lophodermium		Pathotroph	Plant Pathogen	65
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp	-	-	64
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp	-	-	64
Ascomycota	Dothideomycetes	Capnodiales				-	-	63
Ascomycota	Dothideomycetes	Capnodiales	Incertae sedis	Capnobotryella	Capnobotryella sp MA 4642	Saprotroph	Undefined Saprotroph	63
Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Phaeomoniella	Phaeomoniella sp	Saprotroph	Undefined Saprotroph	63
Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Phaeomoniella	Phaeomoniella sp	Saprotroph	Undefined Saprotroph	62
						-	-	62
Ascomycota	Leotiomycetes	Rhytismatales	Rhytismataceae	Lophodermium		Pathotroph	Plant Pathogen	62
Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Phaeomoniella	Phaeomoniella sp	Saprotroph	Undefined Saprotroph	61
Ascomycota	Incertae sedis	Incertae sedis	Incertae sedis	Knufia		-	-	60
Ascomycota	Leotiomycetes	Rhytismatales	Rhytismataceae	Lophodermium	Lophodermium seditiosum	Pathotroph	Plant Pathogen	60
Ascomycota	Dothideomycetes	unidentified	unidentified	unidentified	Dothideomycetes sp 11147	-	-	59

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

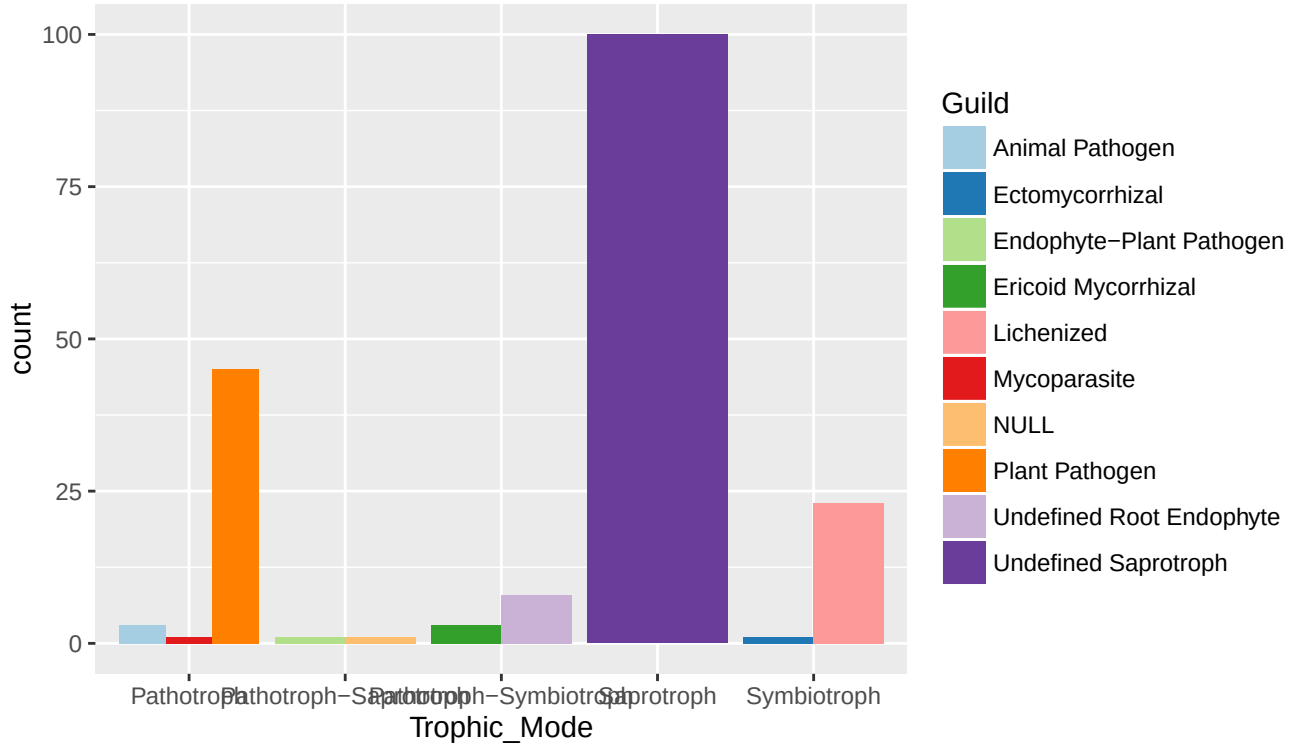

Figure 4.1: Distribution of OTUs into functional Guild.

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

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

## [1] 82.10002

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

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

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

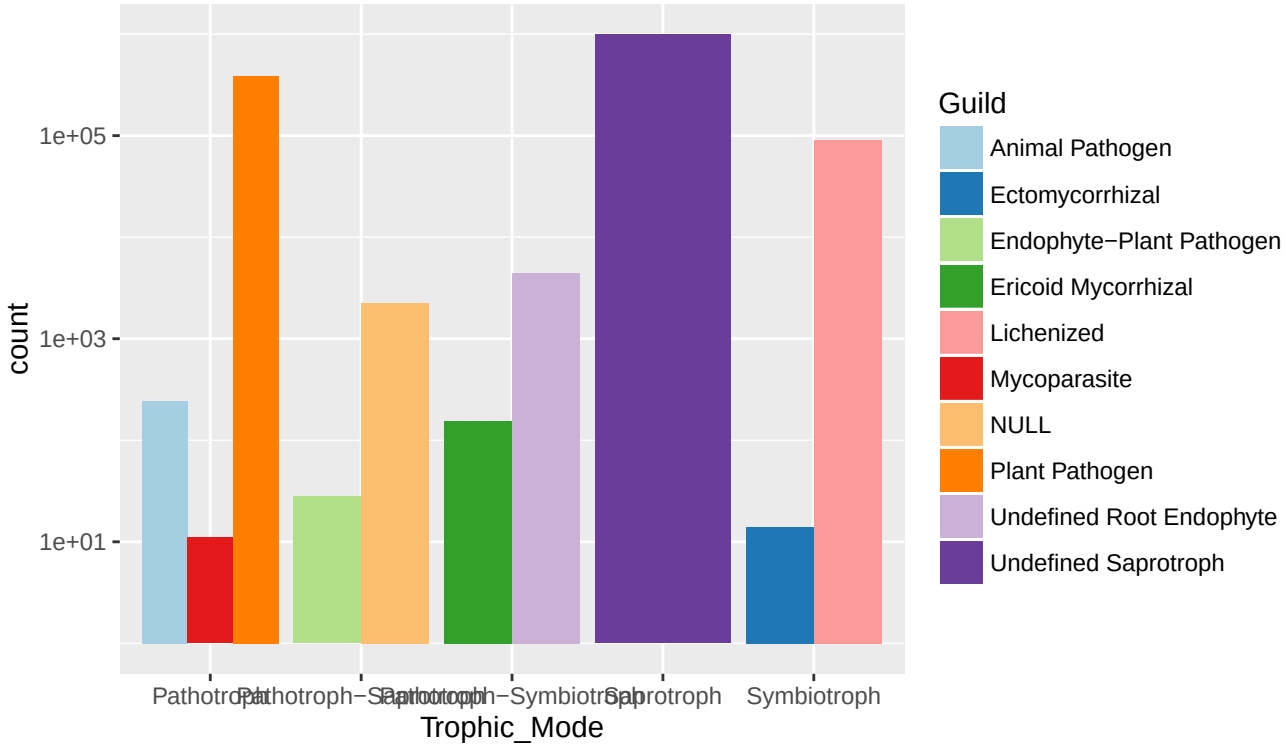

Figure 4.2: Distribution of sequences (log10 transformed) into functional Guild.

5 Distribution of fungal endophytic alpha-biodiversity

5.1 Local diversity = Diversity by sites

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

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

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

5.2 Diversity by age of tree

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

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

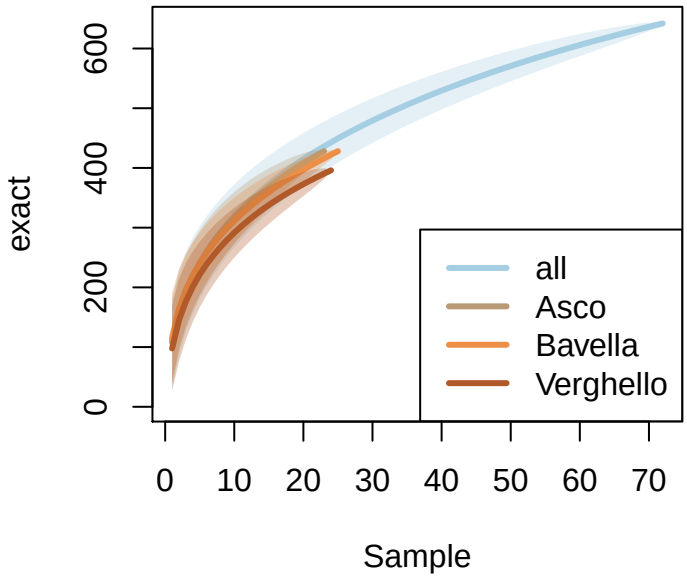

Figure 5.1: Rarefaction curves for each sites. Notes that if singletons were removed, these curves are biaised.

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)
```

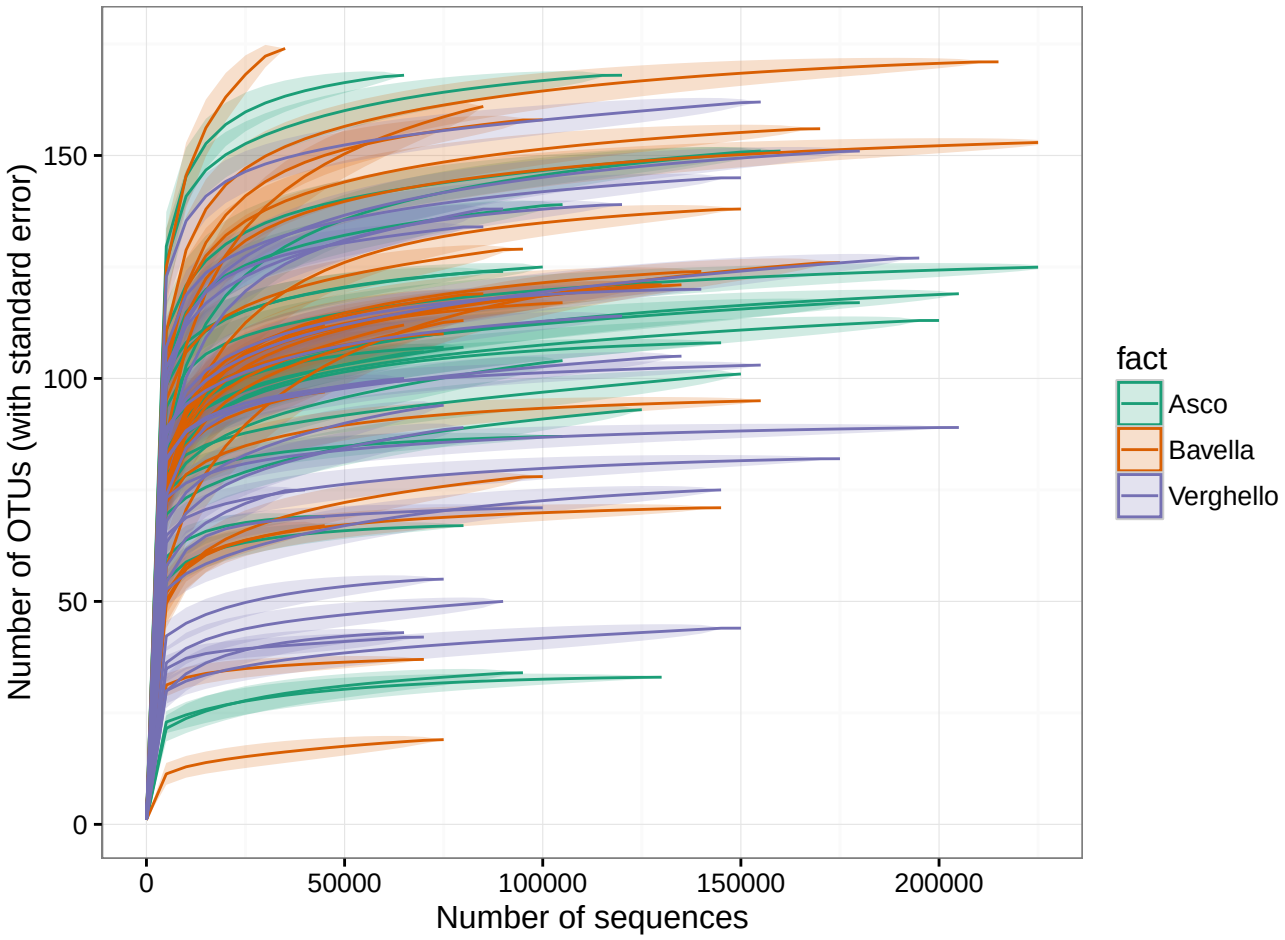

Figure 5.2: Rarefaction curves for each samples using sequences number on x-axes. Notes that if singletons were removed, these curves are biaised.

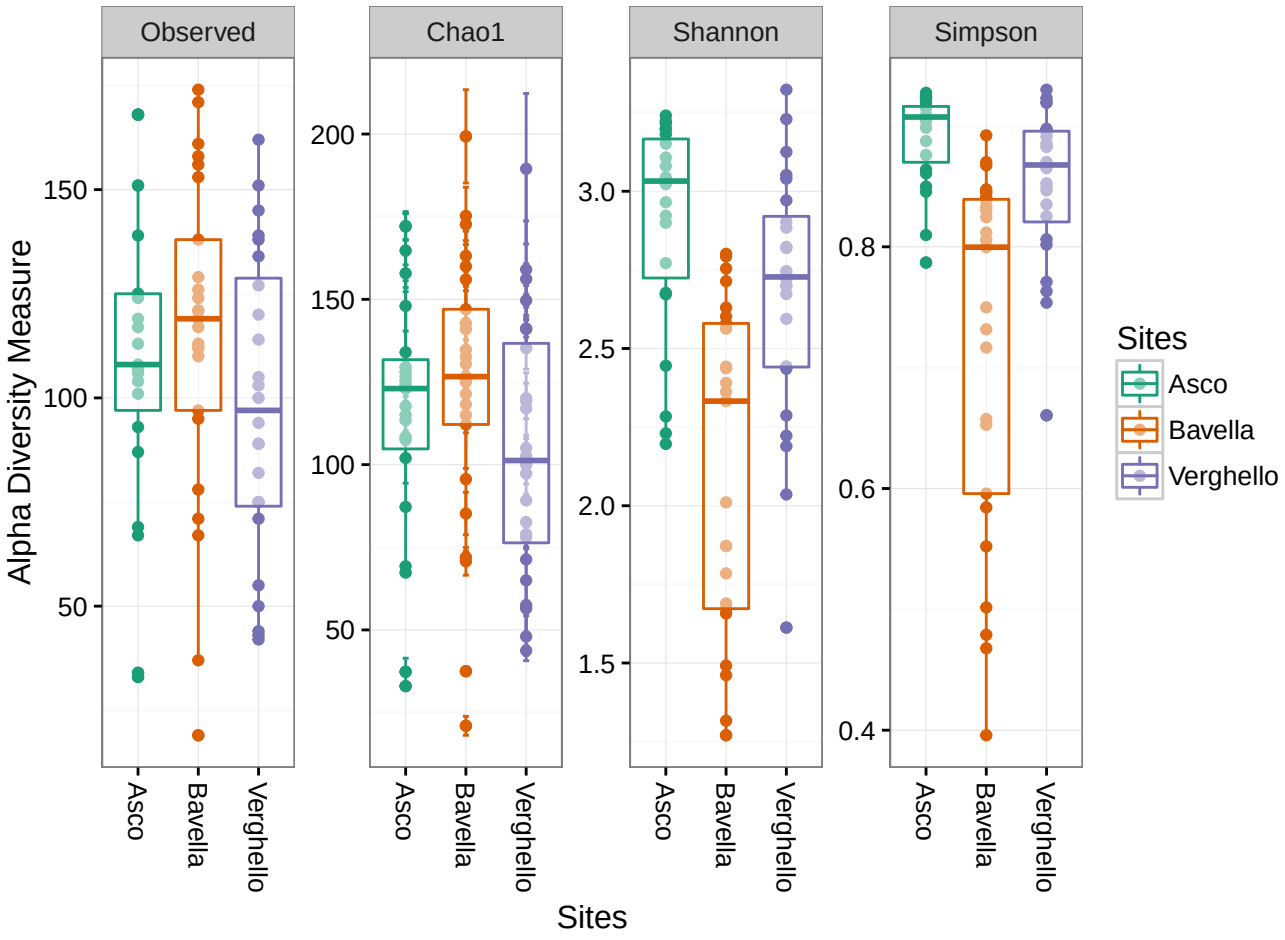

Figure 5.3: Diversity of each sites

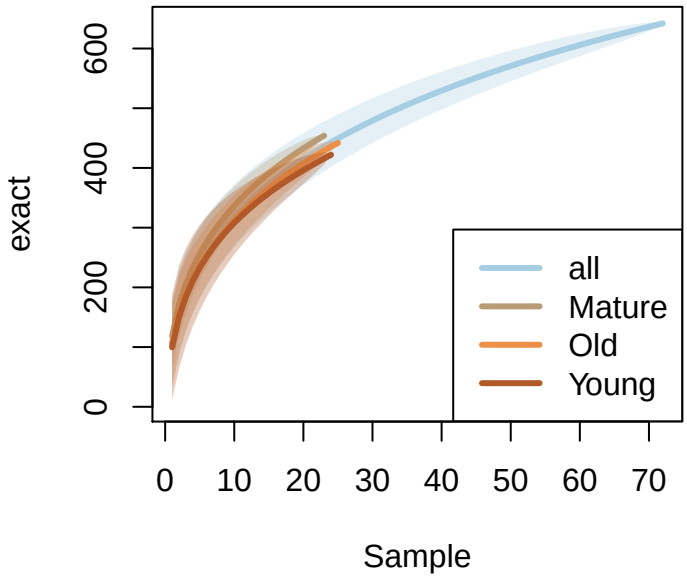

Figure 5.4: Rarefaction curves for each tree age modalities. Notes that if singletons were removed, these curves are biaised.

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

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

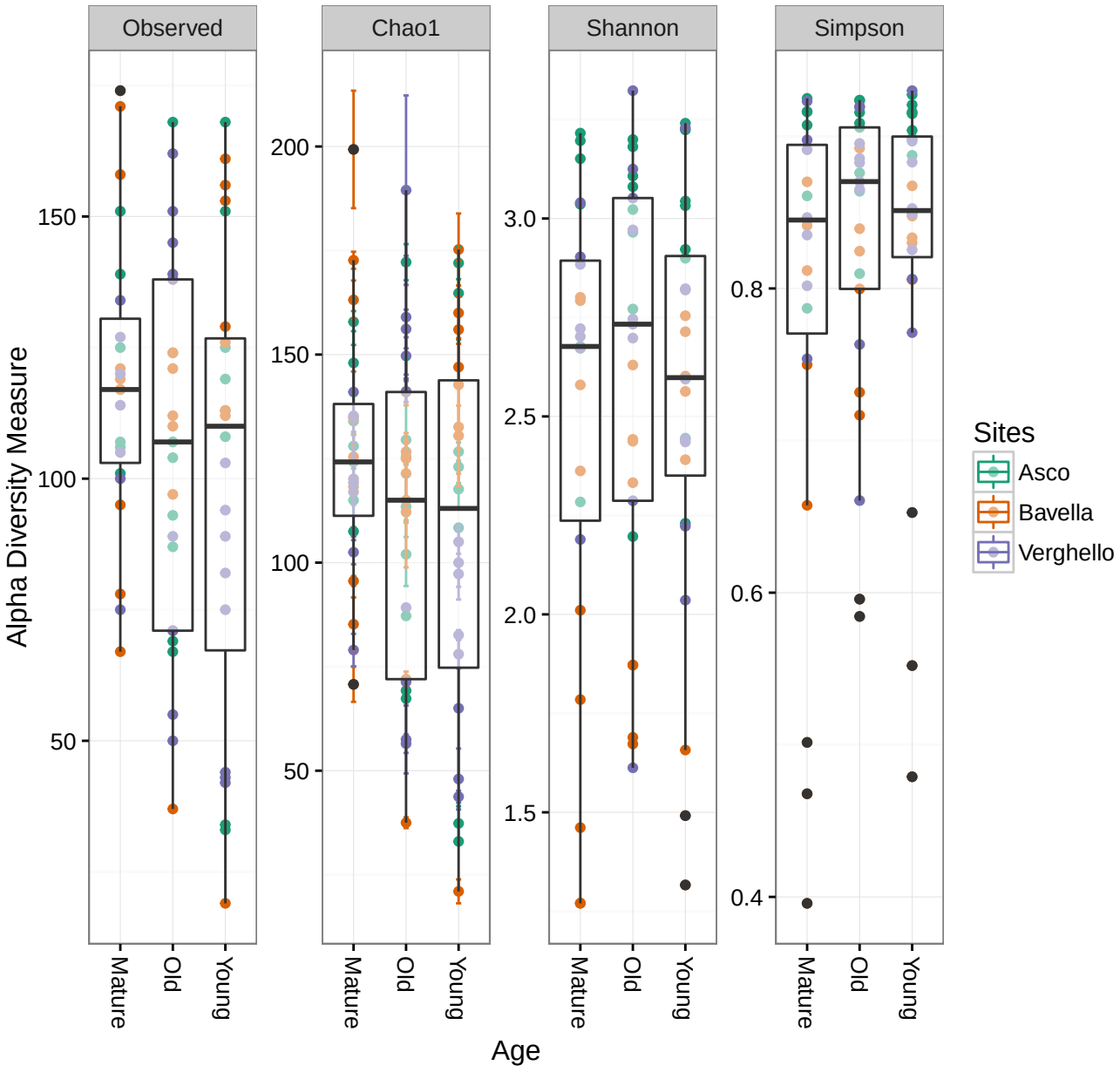

Figure 5.5: Diversity in function of tree age. Color represent sites.

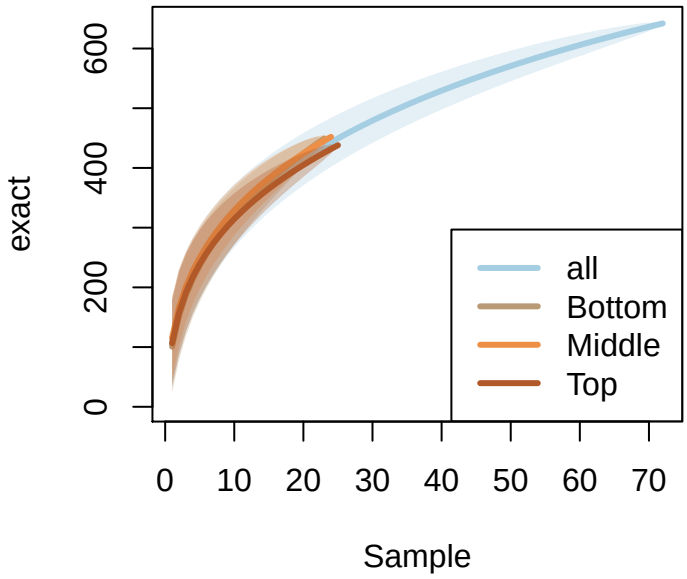

Figure 5.6: Rarefaction curves for each elevation. Notes that if singletons were removed, these curves are biaised.

```
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).

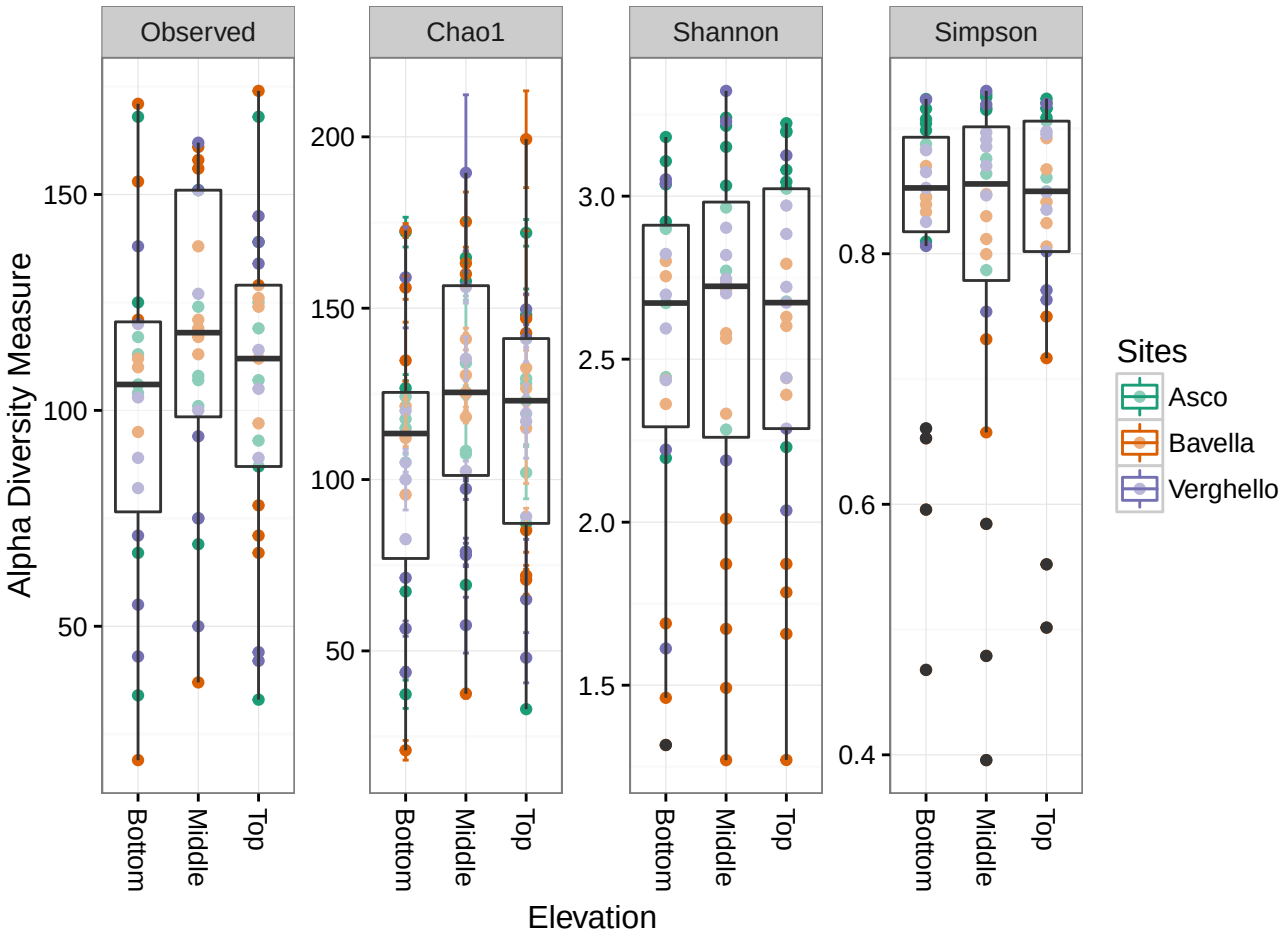

Figure 5.7: Diversity in function of elevation. Color represent sites.

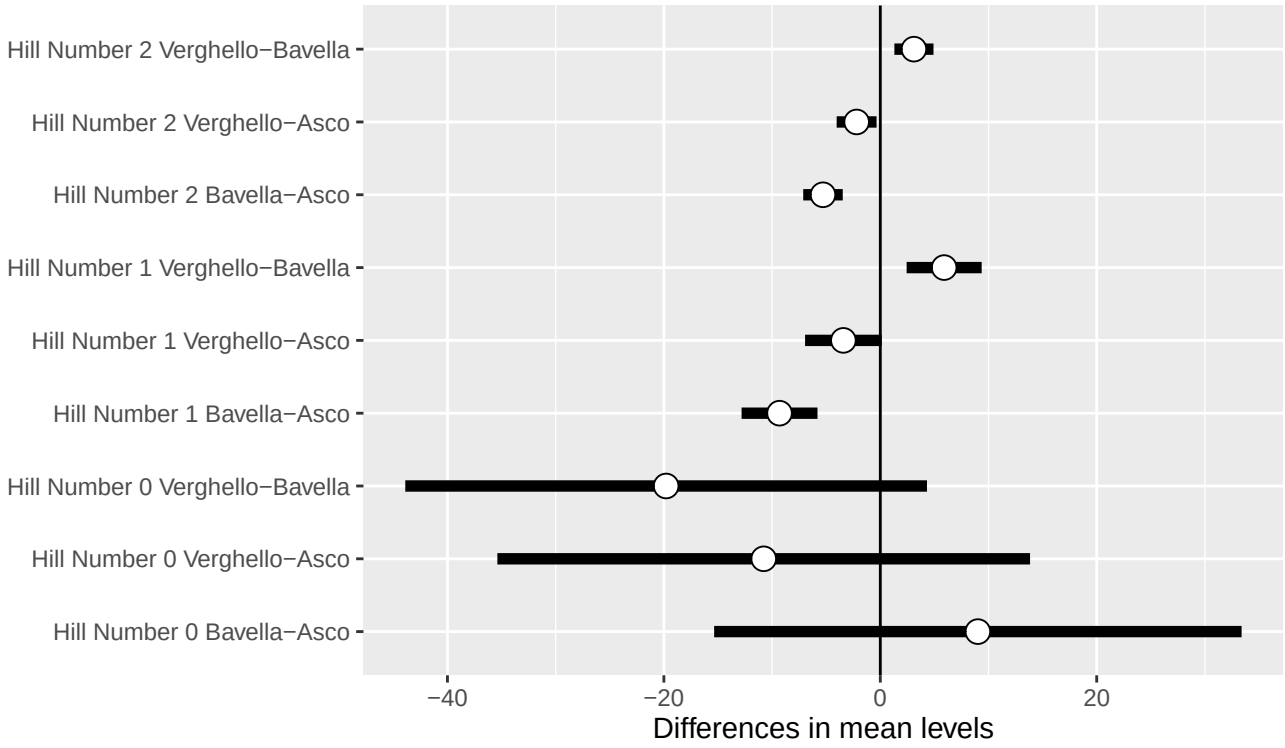

Figure 5.8: Results of the Tuckey HSD testing for differences in mean Hill numbers among pairs of modalities

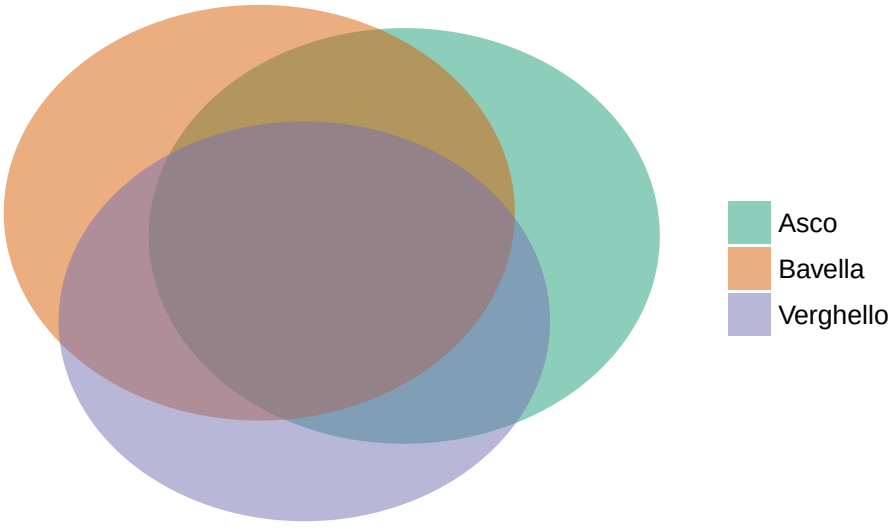

Figure 6.1: Venn diagramm ef the distribution of OTUs among Sites

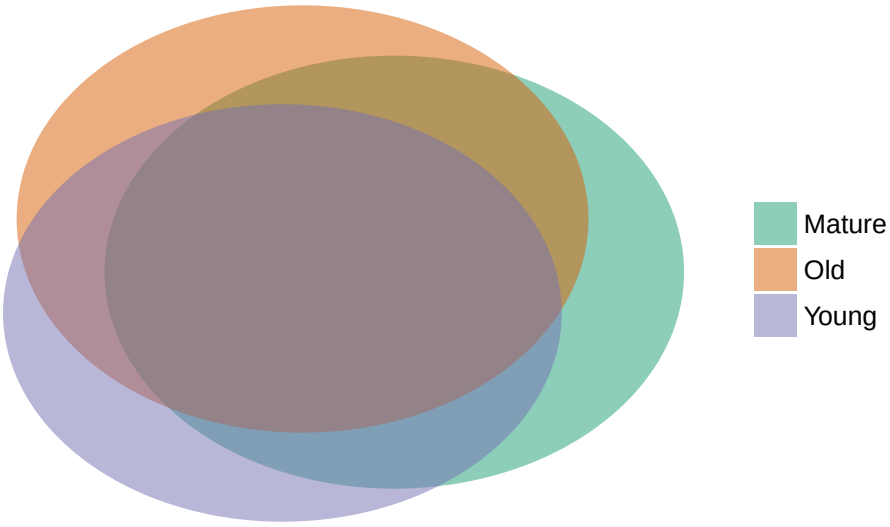

Figure 6.2: Venn diagramm ef the distribution of OTUs among host age

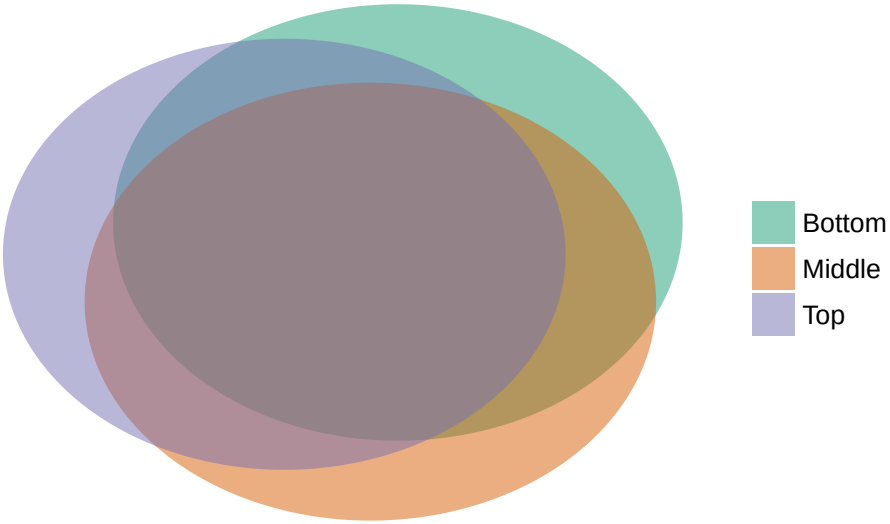

Figure 6.3: Venn diagramm ef the distribution of OTUs among elevation of samples

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

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

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

Table 6: Summary of the linear model of inverse of 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 0.2351585
## Run 1 stress 0.235117
## ... New best solution
## ... Procrustes: rmse 0.06429465  max resid 0.3126958
## Run 2 stress 0.2465396
## Run 3 stress 0.2484362
## Run 4 stress 0.2445164
```

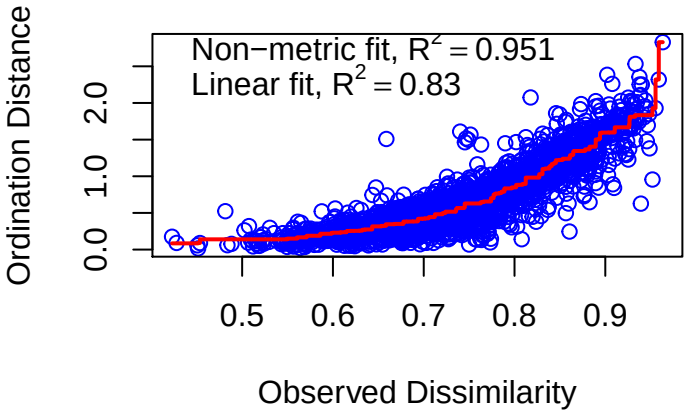

Figure 6.4: Stress plot of the NMDS

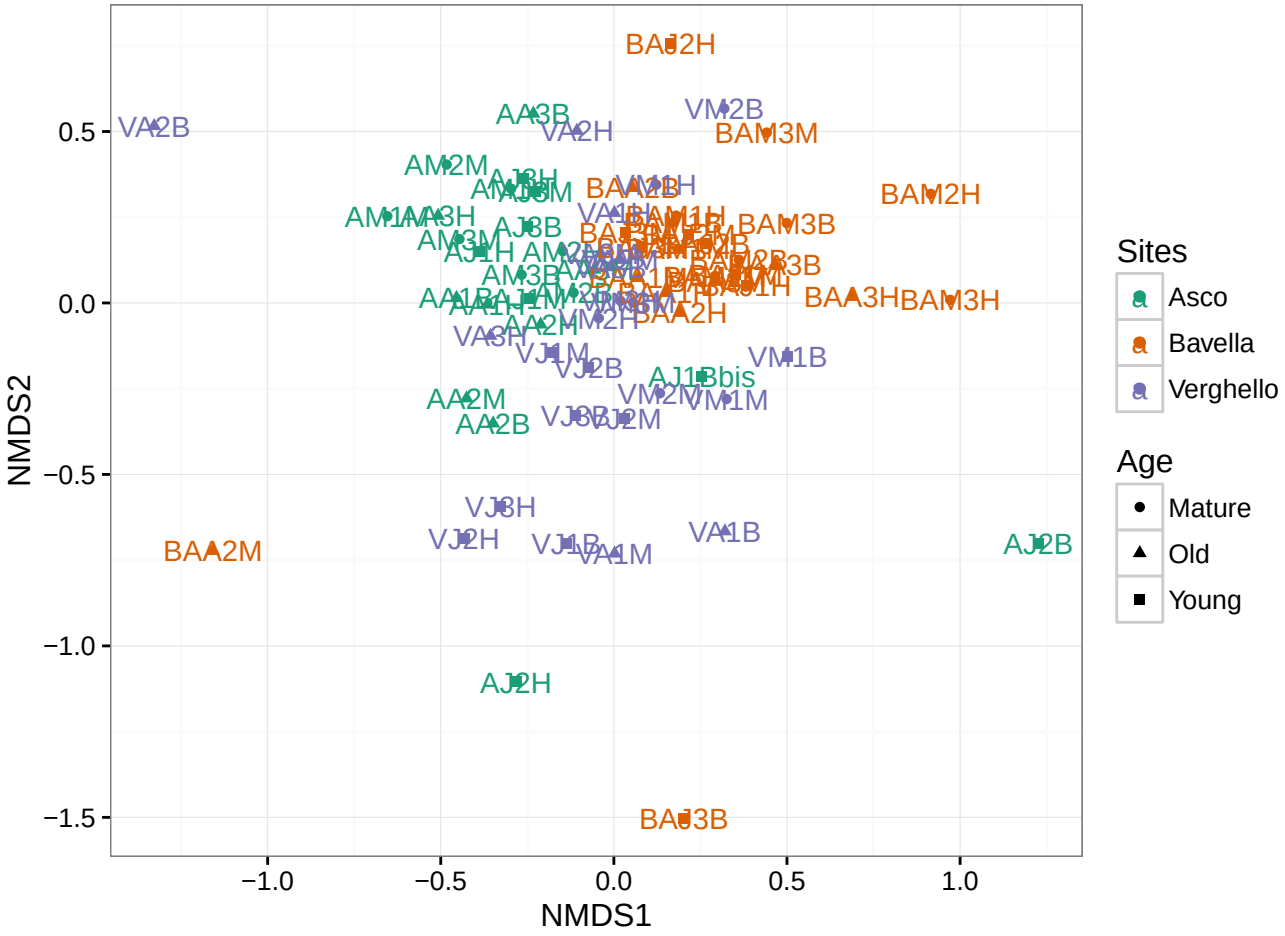

Figure 6.5: NMDS of OTU table. Colors represent sites and shape the age of tree.

```
## Run 5 stress 0.2391762
## Run 6 stress 0.2416561
## Run 7 stress 0.2552093
## Run 8 stress 0.2435231
## Run 9 stress 0.2498757
## Run 10 stress 0.2395833
## Run 11 stress 0.2351381
## ... Procrustes: rmse 0.09200628  max resid 0.3770775
## Run 12 stress 0.2416408
## Run 13 stress 0.2515543
## Run 14 stress 0.2465195
## Run 15 stress 0.2371596
## Run 16 stress 0.246402
## Run 17 stress 0.2450197
## Run 18 stress 0.242819
## Run 19 stress 0.2416565
## Run 20 stress 0.2405462
## *** No convergence -- monoMDS stopping criteria:
##      3: no. of iterations >= maxit
##     17: stress ratio > sratmax

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

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

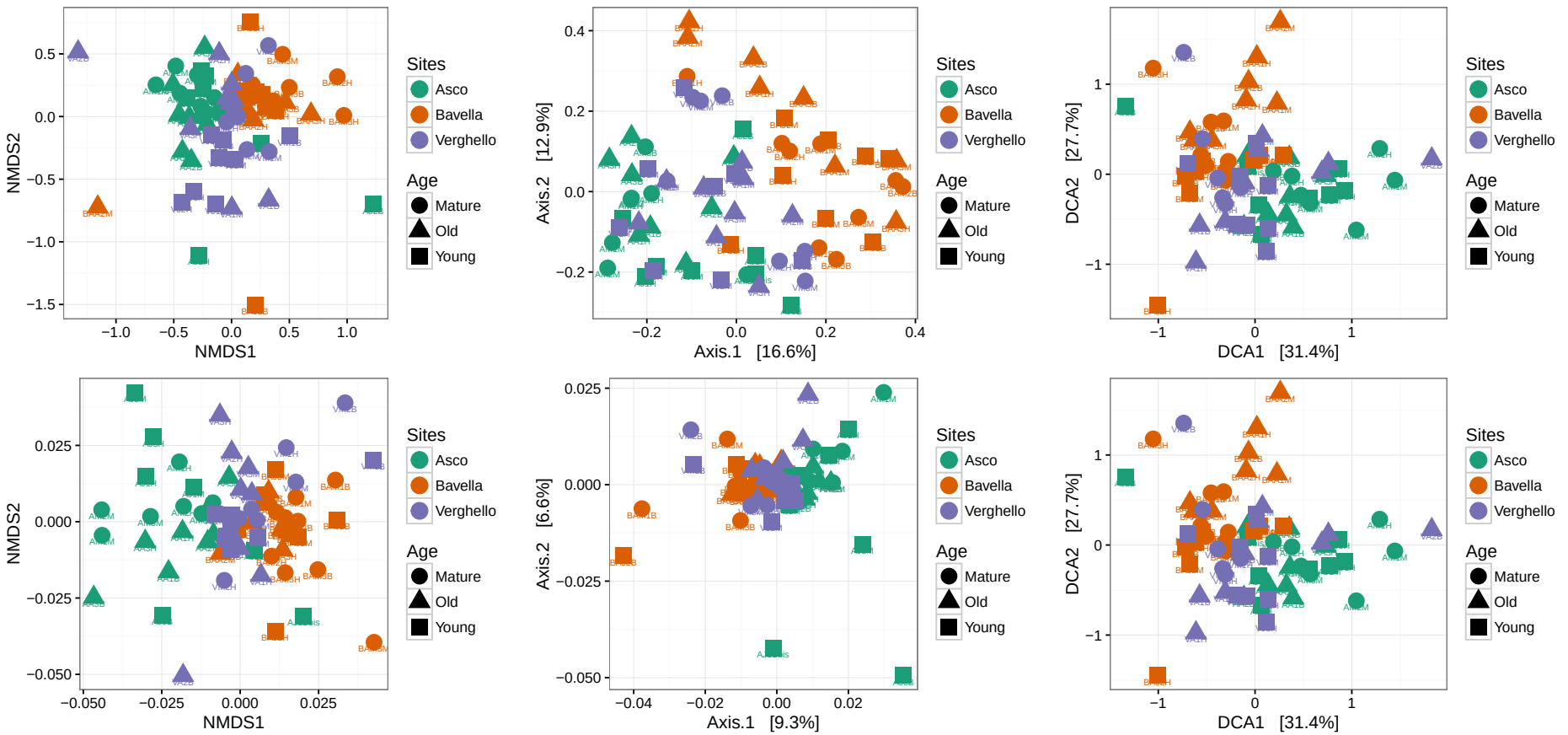

Figure 6.6: Comparison of different distances (bray (up) and gower (bottom)) and ordination methods (NMDS (left), PCoA (center) and DCA (right)).

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

6.3 Permanova on sites, host ages and elevation

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

If we only keep the 99 OTUs present in more than 30 sample, the Permanova results is the following:

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

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

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Sites	2	1.99	1.00	5.78	0.14	0.0001
Age	2	0.58	0.29	1.69	0.04	0.0179
Elevation	2	0.49	0.25	1.43	0.03	0.0676
Sites:Age	4	1.34	0.33	1.94	0.09	0.0001
Sites:Elevation	4	0.66	0.16	0.95	0.04	0.5779
Age:Elevation	4	0.64	0.16	0.93	0.04	0.6175
Sites:Age:Elevation	8	1.20	0.15	0.87	0.08	0.8230
Residuals	45	7.76	0.17		0.53	
Total	71	14.65			1.00	

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

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

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Sites	2	1.94	0.97	5.96	0.14	0.0001
Age	2	0.56	0.28	1.71	0.04	0.0163
Elevation	2	0.48	0.24	1.47	0.03	0.0653
Sites:Age	4	1.29	0.32	1.98	0.09	0.0003
Sites:Elevation	4	0.62	0.15	0.95	0.04	0.5729
Age:Elevation	4	0.60	0.15	0.92	0.04	0.6423
Sites:Age:Elevation	8	1.10	0.14	0.84	0.08	0.8615
Residuals	45	7.33	0.16		0.53	
Total	71	13.90			1.00	

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

```
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	0.81	0.40	3.47	0.09	0.0001
Age	2	0.37	0.18	1.57	0.04	0.0157
Elevation	2	0.24	0.12	1.01	0.03	0.4228
Sites:Age	4	0.67	0.17	1.43	0.07	0.0094
Sites:Elevation	4	0.43	0.11	0.93	0.05	0.6657
Age:Elevation	4	0.51	0.13	1.09	0.06	0.2501
Sites:Age:Elevation	8	0.84	0.10	0.90	0.09	0.7995
Residuals	45	5.24	0.12		0.58	
Total	71	9.10			1.00	

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

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	1.99	1.00	6.88	0.14	0.0001
Age	2	0.58	0.29	2.01	0.04	0.0031
Sites:Age	4	1.33	0.33	2.30	0.09	0.0001
Sites:Age:IndividualTree	18	4.23	0.23	1.62	0.29	0.0001
Residuals	45	6.52	0.14		0.44	
Total	71	14.65			1.00	

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

```
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	1.94	0.97	7.18	0.14	0.0001
Age	2	0.56	0.28	2.07	0.04	0.0029
Sites:Age	4	1.29	0.32	2.39	0.09	0.0001
Sites:Age:IndividualTree	18	4.04	0.22	1.66	0.29	0.0001
Residuals	45	6.08	0.14		0.44	
Total	71	13.90			1.00	

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

```
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	0.81	0.40	3.83	0.09	0.0001
Age	2	0.37	0.18	1.74	0.04	0.0052
Sites:Age	4	0.68	0.17	1.60	0.07	0.0017
Sites:Age:IndividualTree	18	2.49	0.14	1.31	0.27	0.0016
Residuals	45	4.75	0.11		0.52	
Total	71	9.10			1.00	

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

6.5 Differences in abundances and OTUs number by Order.

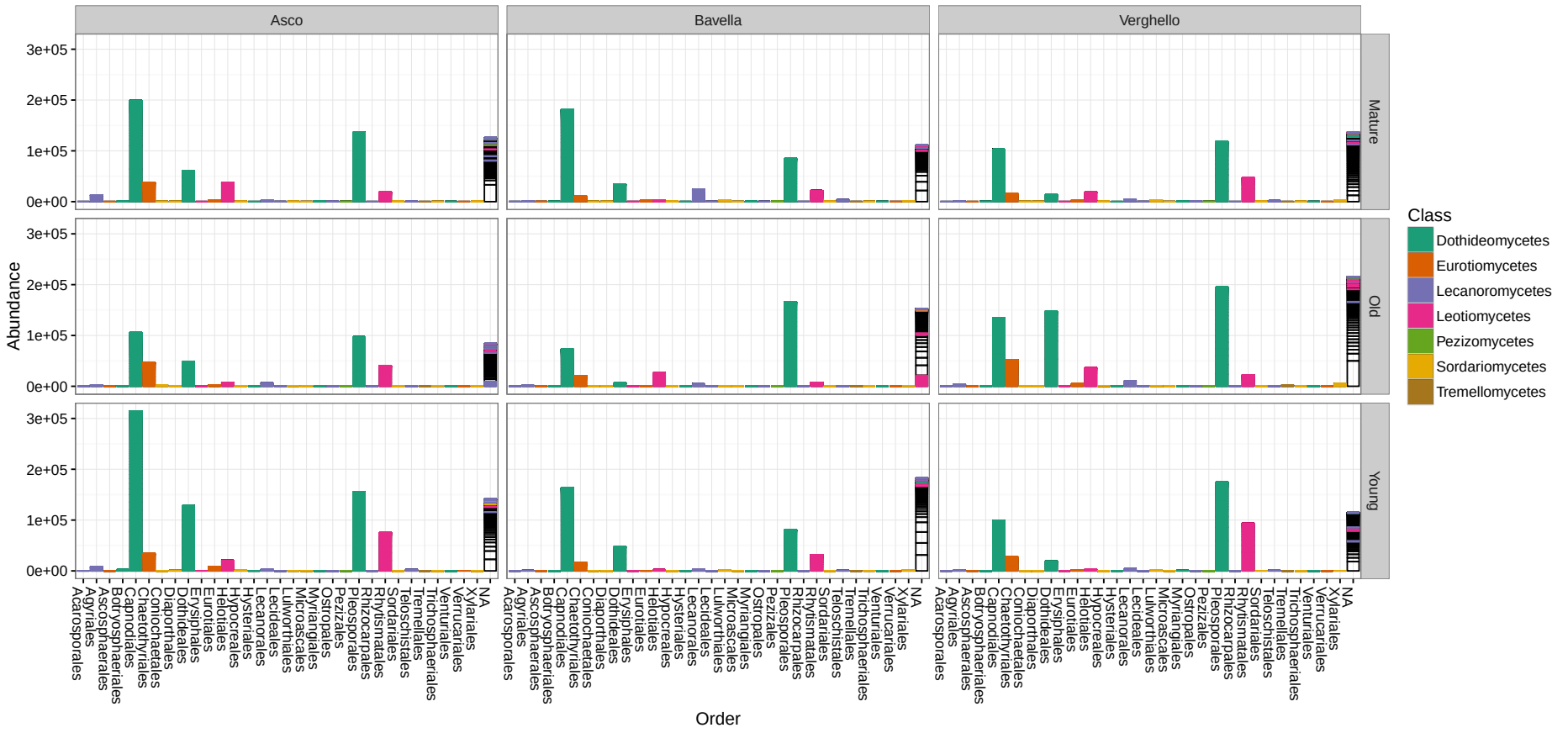

Figure 6.7: Taxonomic distribution of sequences in the different site * age combinaison.

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

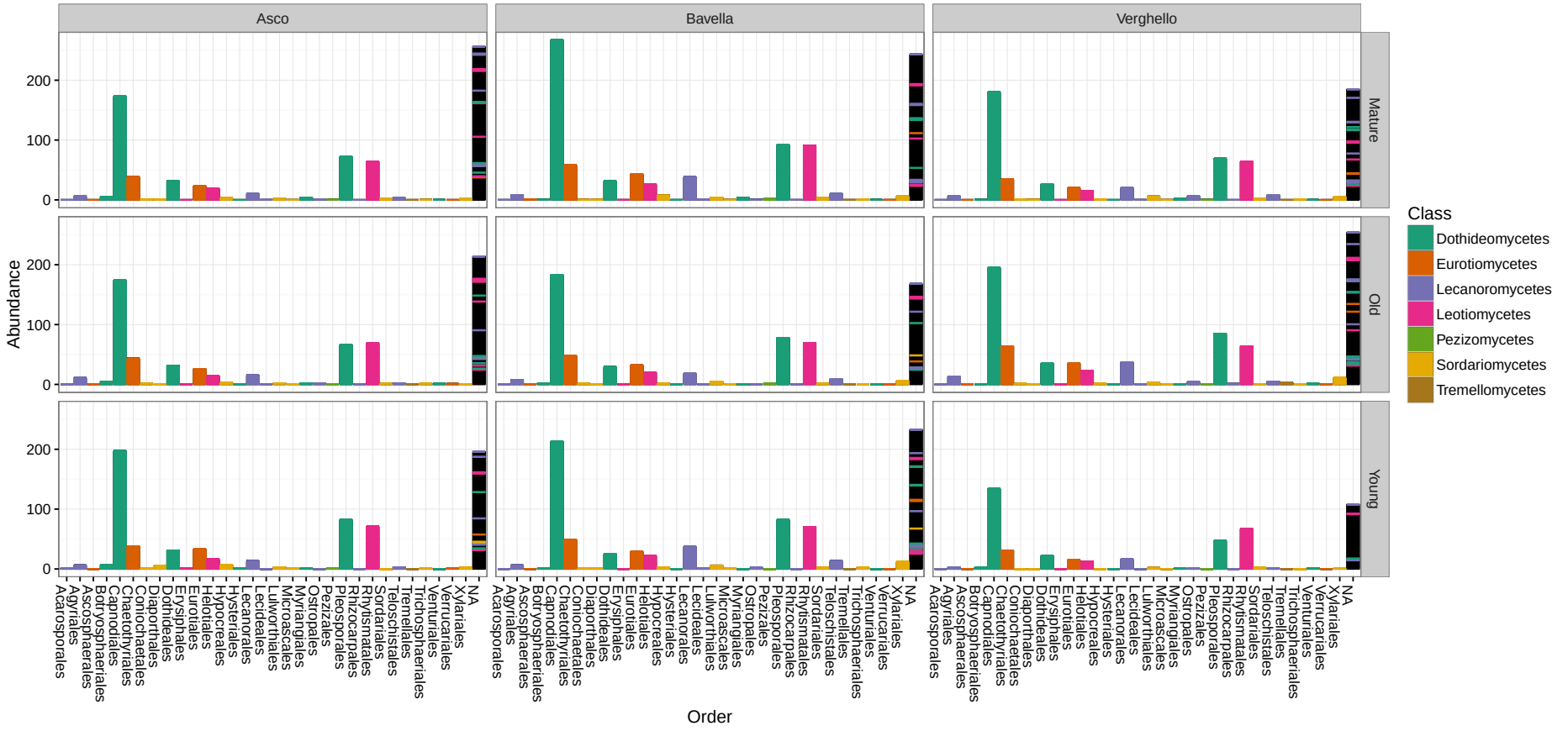

Figure 6.8: Taxonomic distribution of OTUs in the different site * age combinaison.

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

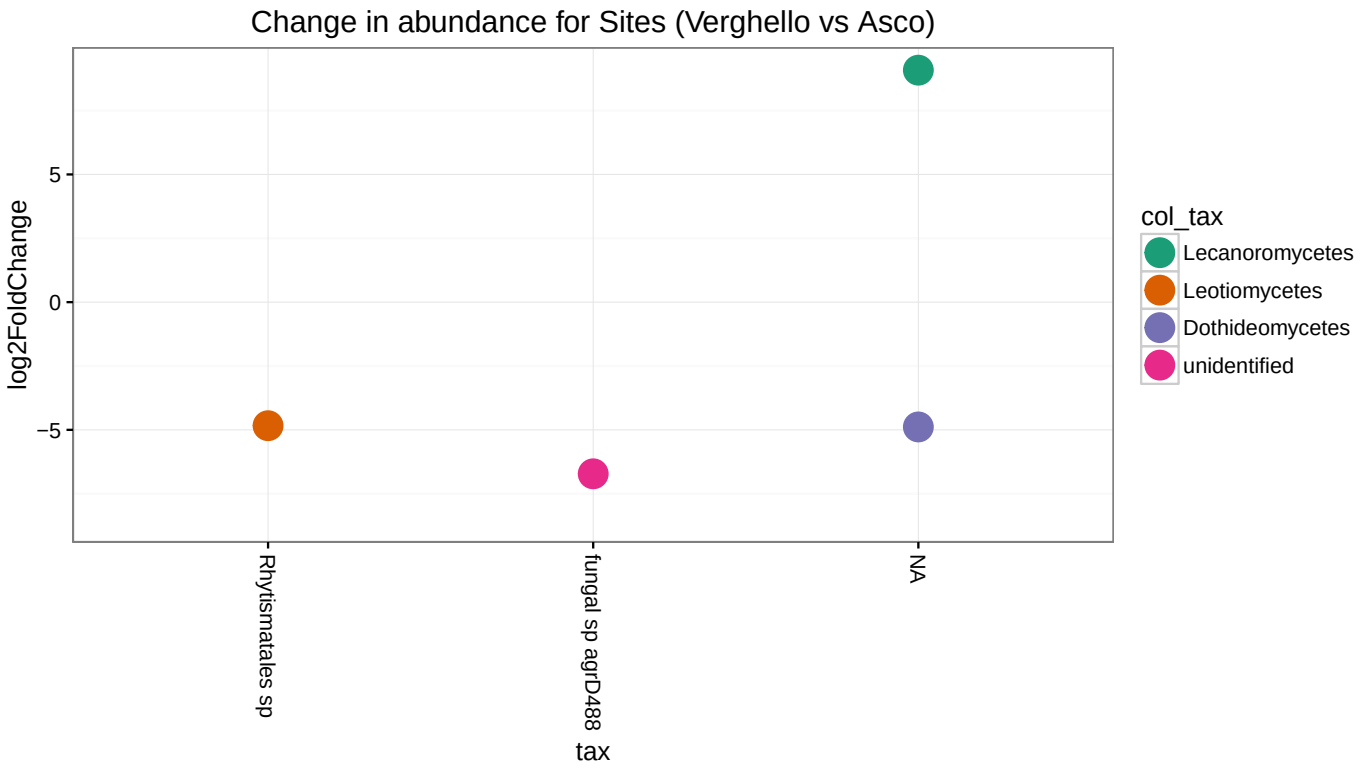

Figure 6.9: OTUs significantly different in terms of abundances between Verghello (positive values) and Asco (negative values)

6.6 Differences in abundances for each OTUs

6.6.1 Pairwise comparison of the OTUs composition by sites

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

## [1] '1.12.3'

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

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

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

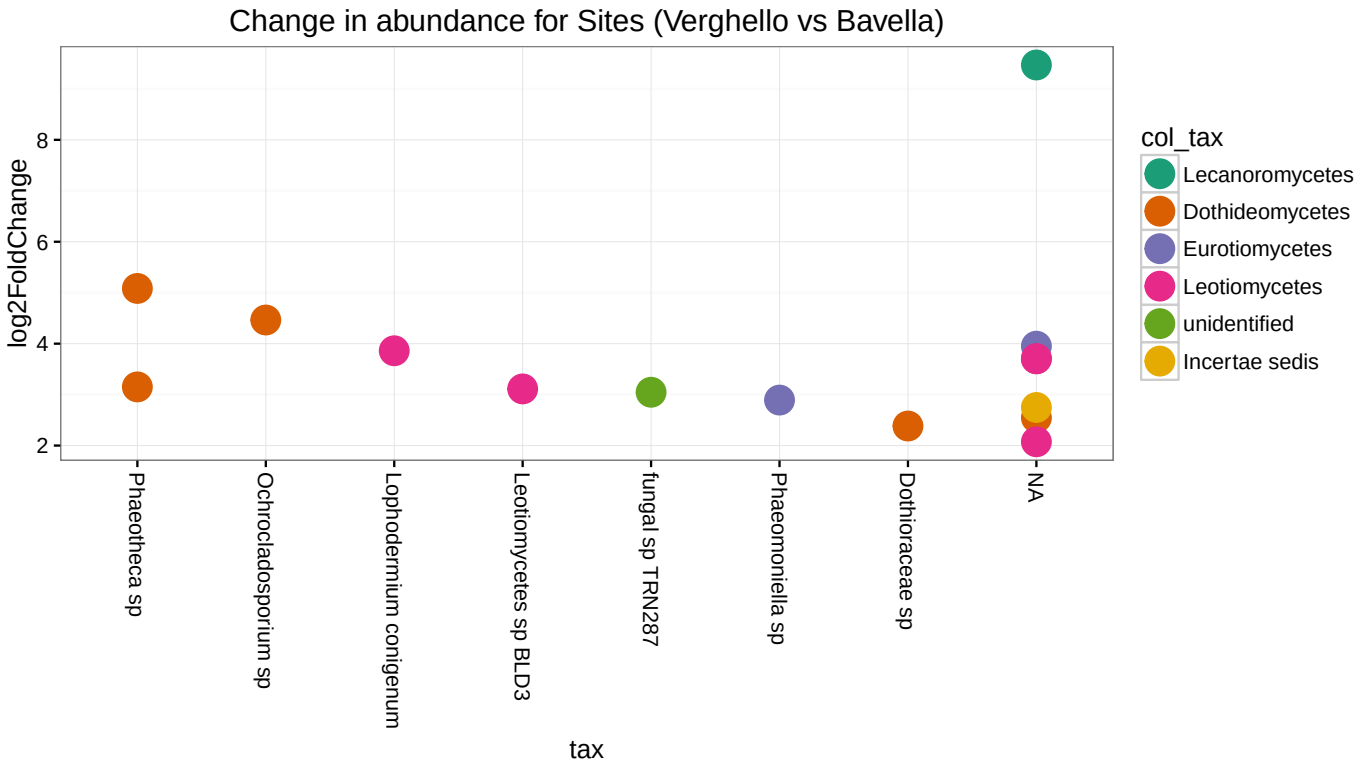

Figure 6.10: OTUs significantly different in terms of abundances between Verghello (positive values) and Bavella (negative values)

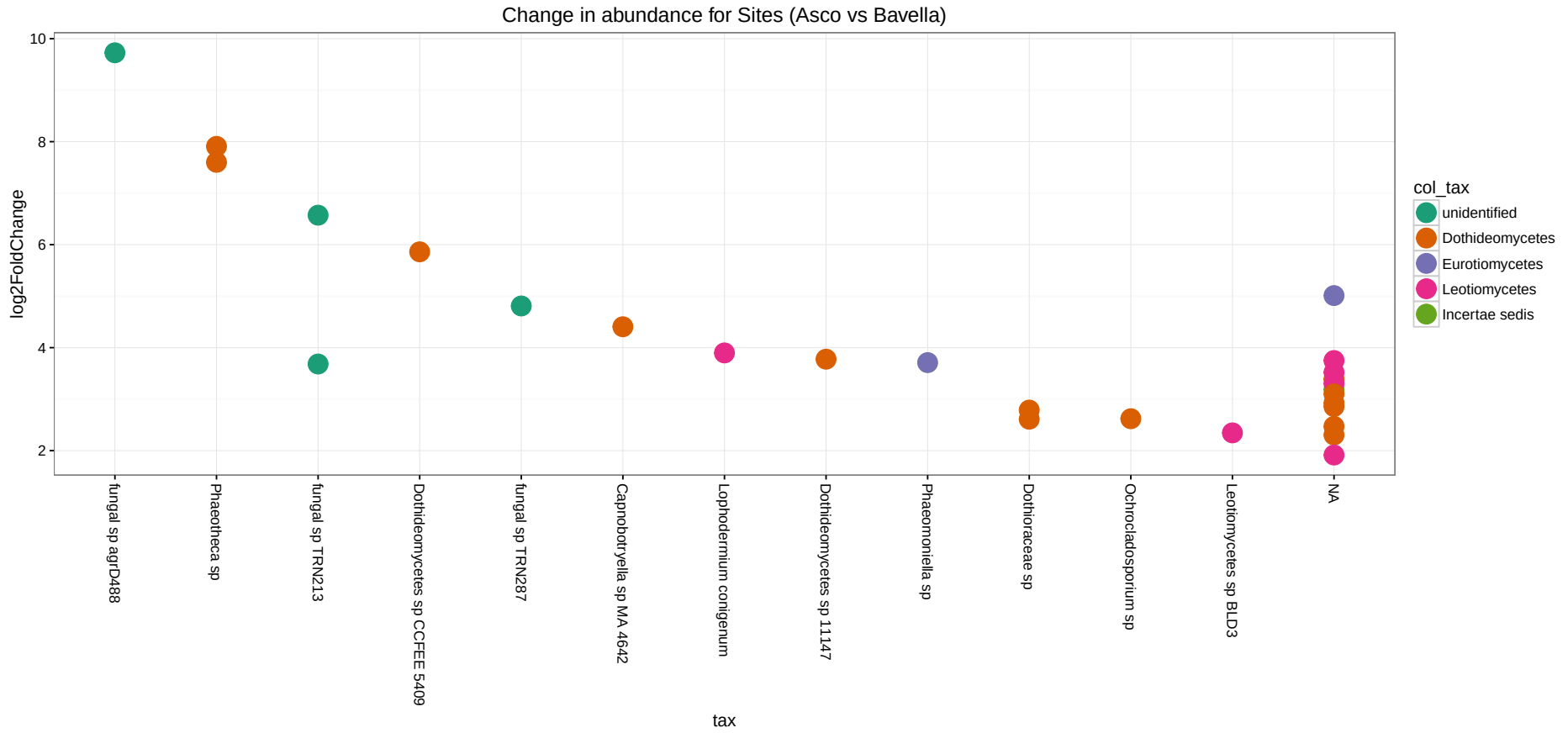

Figure 6.11: OTUs significantly different in terms of abundances between Asco (positive values) and Bavella (negative values)

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

	Comparison	Species	Class	log2FoldChange (negative = more on second levels)
1	Verghello vs Asco			3.34650803573401
2	Verghello vs Asco		Dothideomycetes	-4.88587625995677
3	Verghello vs Asco	Rhytismatales sp	Leotiomycetes	-4.83697217900129
4	Verghello vs Asco	fungal sp agrD488	unidentified	-6.7207967074565
5	Verghello vs Asco			-8.51262065321077
6	Verghello vs Asco		Lecanoromycetes	9.08215005173178
7	Verghello vs Bavella	Dothioraceae sp	Dothideomycetes	2.38357645609838
8	Verghello vs Bavella	Lophodermium conigenum	Leotiomycetes	3.86066845689791
9	Verghello vs Bavella	fungal sp TRN287	unidentified	3.04628000988447
10	Verghello vs Bavella	Phaeomoniella sp	Eurotiomycetes	2.89255911678496
11	Verghello vs Bavella		Dothideomycetes	2.54582241562538
12	Verghello vs Bavella	Leotiomycetes sp BLD3	Leotiomycetes	3.11189246689827
13	Verghello vs Bavella		Incertae sedis	2.74739597769753
14	Verghello vs Bavella	Phaeotheca sp	Dothideomycetes	5.08366508512397
15	Verghello vs Bavella			5.3100020568724
16	Verghello vs Bavella		Eurotiomycetes	3.95069650143599
17	Verghello vs Bavella	Ochrocladosporium sp	Dothideomycetes	4.46408114518602
18	Verghello vs Bavella			4.73753125122742
19	Verghello vs Bavella		Leotiomycetes	2.07416503096289
20	Verghello vs Bavella			7.14204137666655
21	Verghello vs Bavella		Leotiomycetes	3.7067306553489
22	Verghello vs Bavella	Phaeotheca sp	Dothideomycetes	3.15123941889459
23	Verghello vs Bavella		Leotiomycetes	3.70125742114322
24	Verghello vs Bavella			5.52064821578994
25	Verghello vs Bavella		Lecanoromycetes	9.46919308109156
26	Asco vs Bavella	Dothioraceae sp	Dothideomycetes	2.78803933288057
27	Asco vs Bavella		Dothideomycetes	2.92485814340048
28	Asco vs Bavella	Lophodermium conigenum	Leotiomycetes	3.89774298632977
29	Asco vs Bavella	fungal sp TRN287	unidentified	4.80936217436198
30	Asco vs Bavella	Phaeomoniella sp	Eurotiomycetes	3.70962973322757
31	Asco vs Bavella			4.24345169599319
32	Asco vs Bavella	Capnobotryella sp MA 4642	Dothideomycetes	4.40589112634833
33	Asco vs Bavella		Dothideomycetes	2.3039021041349
34	Asco vs Bavella	Leotiomycetes sp BLD3	Leotiomycetes	2.34544553732297
35	Asco vs Bavella		Incertae sedis	3.18386596238751
36	Asco vs Bavella	Phaeotheca sp	Dothideomycetes	7.90813619206671
37	Asco vs Bavella			7.0594168330479
38	Asco vs Bavella		Eurotiomycetes	5.0114514780679
39	Asco vs Bavella	Ochrocladosporium sp	Dothideomycetes	2.61993100125331
40	Asco vs Bavella		Dothideomycetes	2.47113752116495
41	Asco vs Bavella	fungal sp TRN213	unidentified	6.57235691097091
42	Asco vs Bavella			5.44215709752254
43	Asco vs Bavella	fungal sp TRN213	unidentified	3.68211350941658
44	Asco vs Bavella			4.72029827472868
45	Asco vs Bavella		Dothideomycetes	3.38729088371205
46	Asco vs Bavella		Leotiomycetes	1.91509249525582
47	Asco vs Bavella			3.62110923424642
48	Asco vs Bavella	Dothideomycetes sp 11147	Dothideomycetes	3.77612324690022
49	Asco vs Bavella	Dothioraceae sp	Dothideomycetes	2.61074597039296
50	Asco vs Bavella		Leotiomycetes	3.30304241594154
51	Asco vs Bavella	Phaeotheca sp	Dothideomycetes	7.59895969069648
52	Asco vs Bavella		Dothideomycetes	2.85611509262194
53	Asco vs Bavella		Leotiomycetes	3.51959838470288
54	Asco vs Bavella		Dothideomycetes	3.10126651484418
55	Asco vs Bavella		Leotiomycetes	3.74979729080547
56	Asco vs Bavella	fungal sp agrD488	unidentified	9.72582329366518
57	Asco vs Bavella	Dothideomycetes sp CCFEE 5409	Dothideomycetes	5.86086572706484
58	Asco vs Bavella			6.4955349588301
59	Asco vs Bavella			9.14784442723117
60	Asco vs Bavella			8.06590021863711
61	Asco vs Bavella			5.73845001725675
62	Asco vs Bavella			6.13375955939848
63	Asco vs Bavella			6.75863805627879

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

6.6.2 Pairwise comparison of Order composition by sites

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

	Comparison	Order	Class	log2FoldChange (negative = more on second levels)
1	Verghello vs Asco	None	None	None
2	Verghello vs Bavella	Incertae sedis	Leotiomycetes	-1.24027862995376
3	Verghello vs Bavella	unidentified	unidentified	1.5216528045132
4	Asco vs Bavella	Botryosphaeriales	Dothideomycetes	4.74782209869672
5	Asco vs Bavella	Eurotiales	Eurotiomycetes	1.72527688171133
6	Asco vs Bavella	Incertae sedis	Leotiomycetes	-1.63205659617912
7	Asco vs Bavella	unidentified	unidentified	1.46372039108894
8	Asco vs Bavella	Xylariales	Sordariomycetes	-3.39828444715162

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

List of Figures

List of Tables