

Appendix S9: results after SWARM clustering

Adrien Taudiere*

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

July 25, 2016

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

*adrien.taudiere@cefe.cnrs.fr

Contents

1	Introduction	3
1.1	R requirements	3
1.2	System and session informations	3
1.3	Some usefull functions	5
2	Data	5
2.1	Choice of filter parameters	5
2.2	Load and convert loading	5
2.2.1	Otu table	5
2.2.2	Taxonomy	6
2.2.3	Add FUNguild information to taxonomy Table	6
2.2.4	Representative sequences	7
2.2.5	Samples information	7
2.2.6	Create the phyloseq object	7
2.2.7	Characteristics of the phyloseq data	7
2.3	Filter sample by number of sequences	7
2.4	Filter OTUs by number of samples	8
2.5	Filter OTUs by number of sequences	9
2.6	Summary of filtration workflow	10
3	Simple description of the dataset	10
3.1	Number of sequences and OTUs by samples	10
3.2	Number of sequences and samples for each OTUs	11
3.3	Distribution of sequences in the taxonomy	12
3.4	Focus on the 30 more abundant OTUs (number of sequences)	12
3.5	Focus on the 30 more frequent OTUs (number of samples)	16
4	Number of sequences and OTUs in function of putative ecology (using FUNGuild software; Nguyen et al, 2015)	19
5	Distribution of fungal endophytic alpha-biodiversity	20
5.1	Local diversity = Diversity by sites	20
5.2	Diversity by age of tree	20
5.3	Diversity by elevation of the sample	21
5.4	Which factor affect diversity?	21
6	Effect of site, age and elevation on fungal endophytic beta-diversity	26
6.1	Venn diagramm	26
6.2	Ordination	26
6.3	Permanova on sites, host ages and elevation	34
6.4	Permanova on sites, host ages and individual trees	35
6.5	Differences in abundances and OTUs number by Order.	36
6.6	Differences in abundances for each OTUs	39
6.6.1	Pairwise comparison of the OTUs composition by sites	39
6.6.2	Pairwise comparison of Order composition by sites	42

1 Introduction

This supplementary material presents the ecological analysis of endophytic fungal communities in *Pinus nigra* subsp. *laricio*, an endemic species of Corsica. The dataset analyse here was computed using SWARM clustering (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 2016-07-25 09:36:36. 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 <- "Swarm"

#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 <- gsub("_", "\t", taxRDP_brut)
taxRDP_brut <- read.table(textConnection(taxRDP_brut), sep = "\t", fill = TRUE)
```

```
# Format taxonomy for phyloseq
taxRDP <- taxRDP_brut[match(taxa_names(dataBiom),
                           paste(taxRDP_brut[, 1], taxRDP_brut[, 2], sep = "_")),
                     c(4, 6, 8, 10, 12, 14, 16)]
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=""))
funguild <- funguild[!is.na(match(funguild$OTU_ID, rownames(taxRDP2))),]

match_interm <- match(funguild$OTU_ID, rownames(taxRDP2))

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

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

2.2.4 Representative sequences

```
map_endo <-  
  import_qiime(map = "data/map_qiimedata.txt")  
  
## Processing map file...  
  
map_endo <- map_endo[order(rownames(map_endo)),]
```

2.2.5 Samples information

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

2.2.6 Create the phyloseq object

```
data_all <- merge_phyloseq(dataBiom, repset, taxRDP2)  
  
sample_data(data_all) <- map_endo  
  
data_all@tax_table[data_all@tax_table == ""] <- NA
```

2.2.7 Characteristics of the phyloseq data

```
data_all  
  
## phyloseq-class experiment-level object  
## otu_table() OTU Table: [ 15479 taxa and 80 samples ]  
## sample_data() Sample Data: [ 80 samples by 6 sample variables ]  
## tax_table() Taxonomy Table: [ 15479 taxa by 12 taxonomic ranks ]  
## refseq() DNASTringSet: [ 15479 reference sequences ]
```

The data are made of 8.419809×10^6 sequences representing 15479 OTUs allocate to 80 samples.

2.3 Filter sample by number of sequences

```
N_sam_min  
  
## [1] 20000
```

If we discard samples with less than 2×10^4 sequences, we keep 72 on the 80 samples (90%).

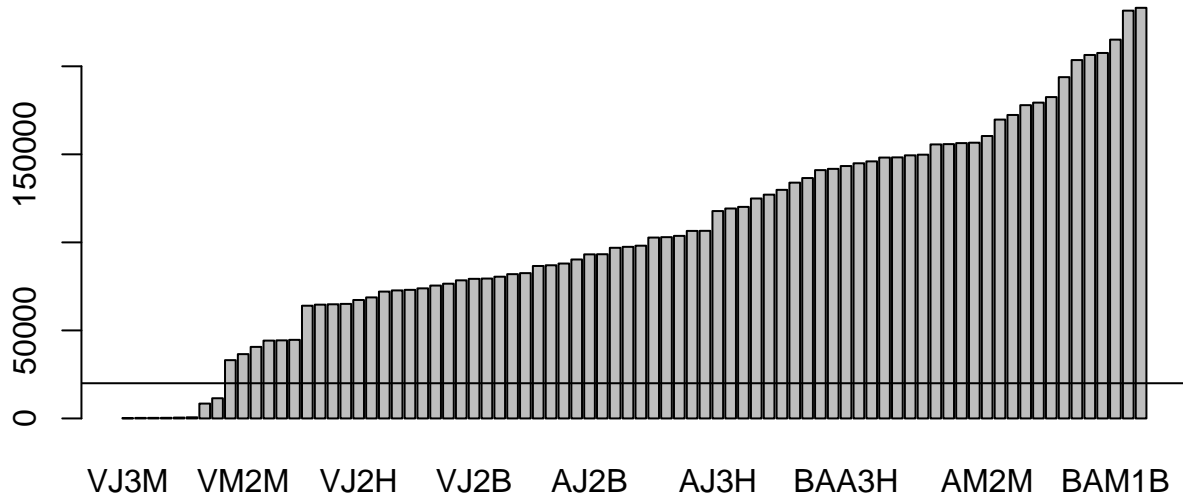


Figure 2.1: Number of sequences by sample

```
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)
```

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

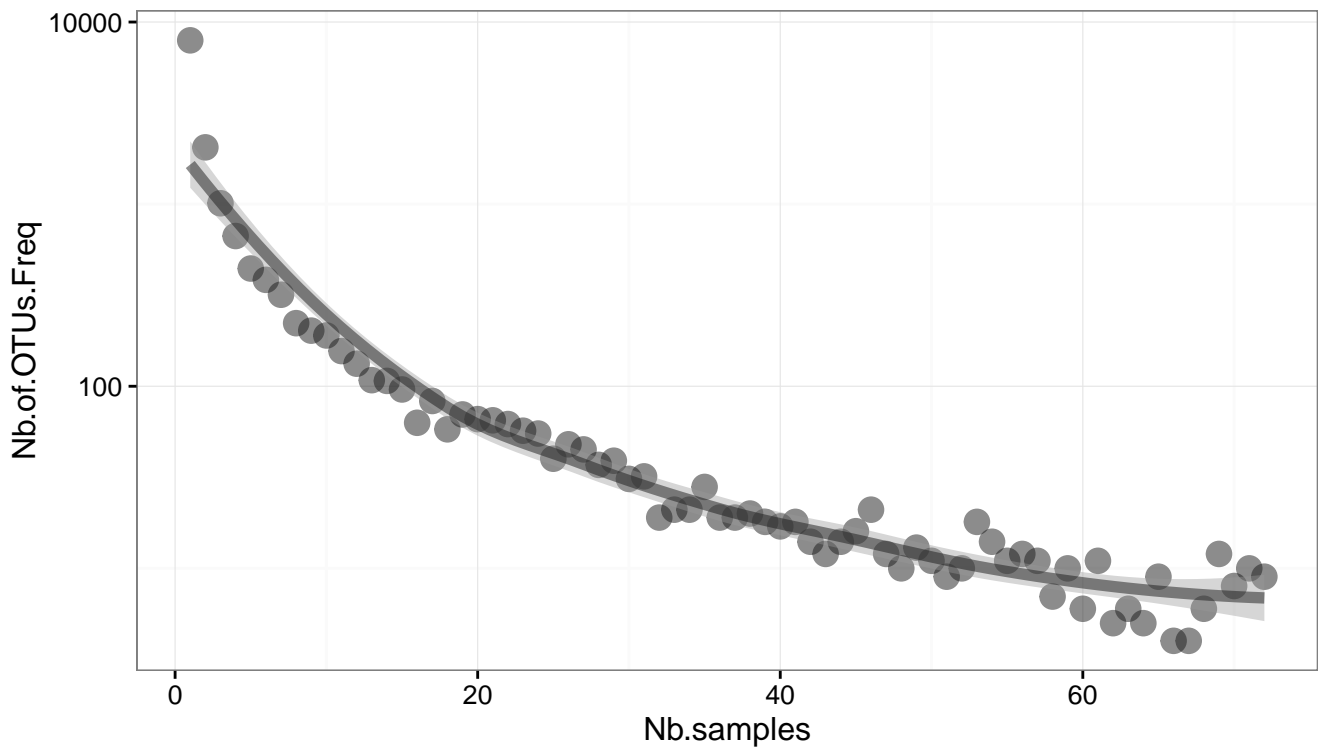


Figure 2.2: Number of OTU present in a given number of samples. Vertical bar illustrate the filtering parameter.

```
N_otu_sam_min
```

```
## [1] 1
```

```
data.f2 <- prune_taxa(rowSums(as.binaryOtuTable(data.f1@otu_table) >=
  N_otu_sam_min, data.f1)
```

If we discard OTUs present in less than 1 sample, we keep 15391 on the 15391 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      1.0      2.0    545.6    14.0 1237000.0
```

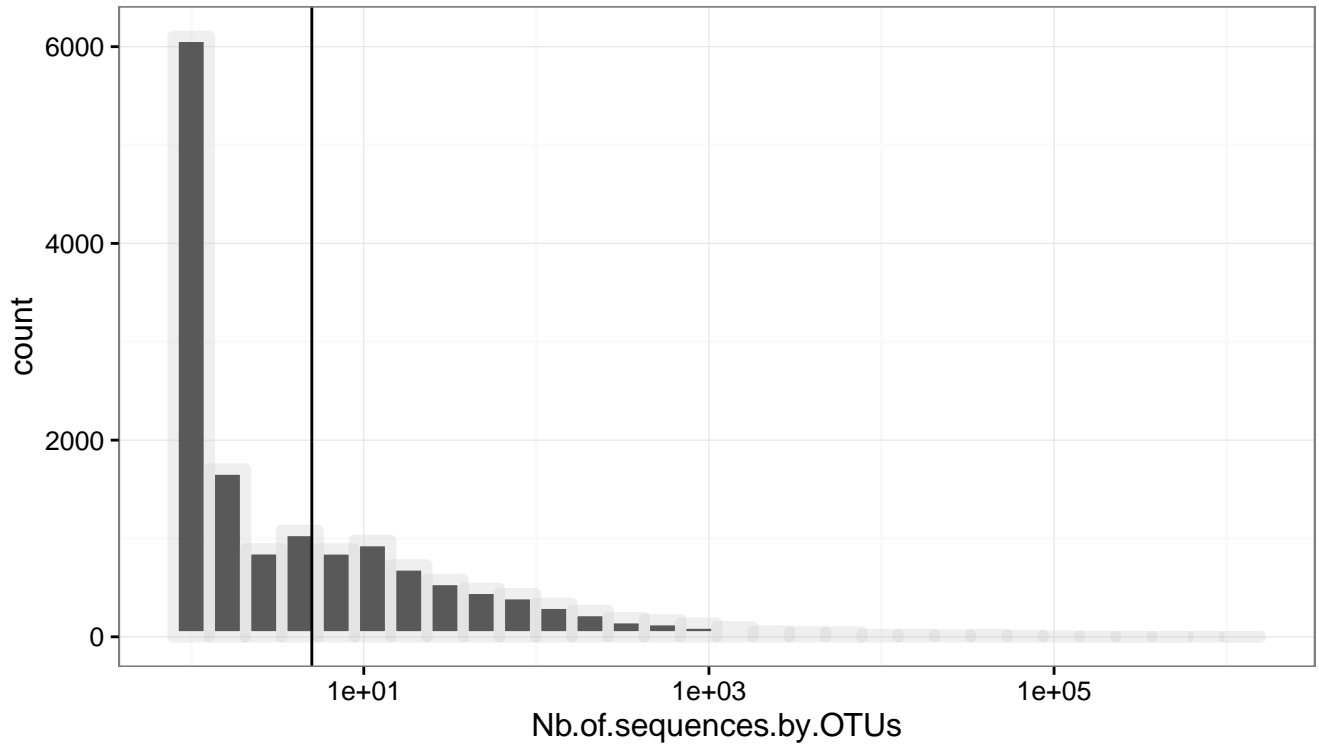


Figure 2.3: Number of sequences by OTU (log10 transformed). Horizontal bar illustrate the filtering parameter.

```
N_seq_otu_min
## [1] 5
```

If we discard OTUs with less than 1 sequences, we keep 6064 on the 15479 OTUs (39.18%).

```
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 8.382948×10^6 sequences representing 6064 OTUs allocate to 72 samples.

	Nb.of.OTUs	Nb.of.samples	Nb.of.sequences
No filter	15479	80	8419809.00
Nb of sequences by sample ≥ 20000	15391	72	8397636.00
Nb of sample by OTUs ≥ 1	15391	72	8397636.00
Nb of sequences by OTUs ≥ 5	6064	72	8382948.00

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

3 Simple description of the dataset

3.1 Number of sequences and OTUs by samples

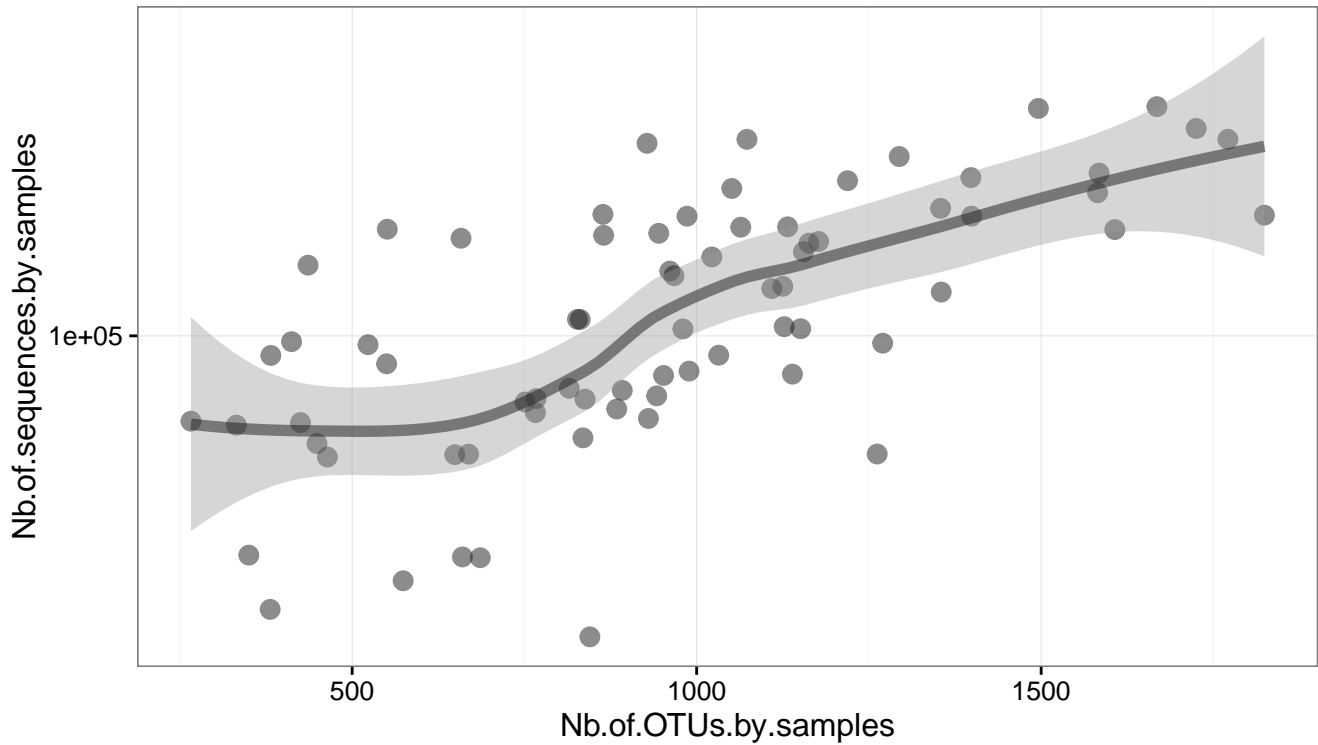


Figure 3.1: Number of OTUs by samples in fonction the number of sequences by samples (log10 axe). The tendency is represented by the line obtain from loess (Local Polynomial Regression Fitting).

```
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))
```

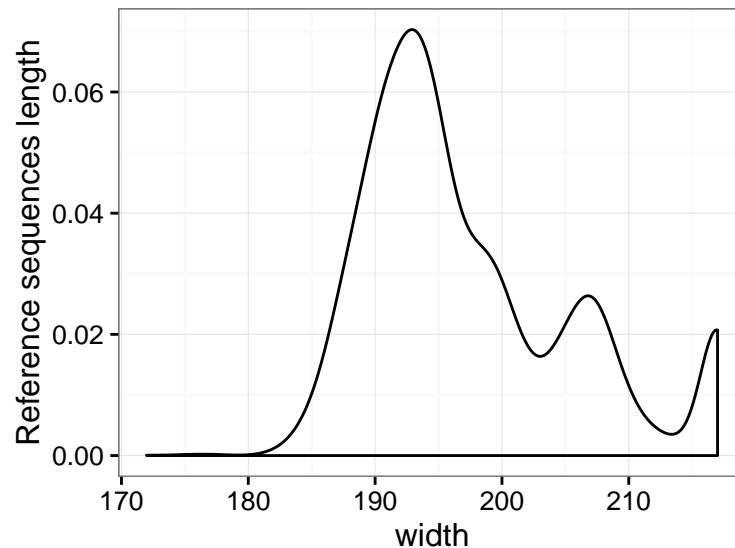


Figure 3.2: Distribution of reference sequences length.

```
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)
```

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

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

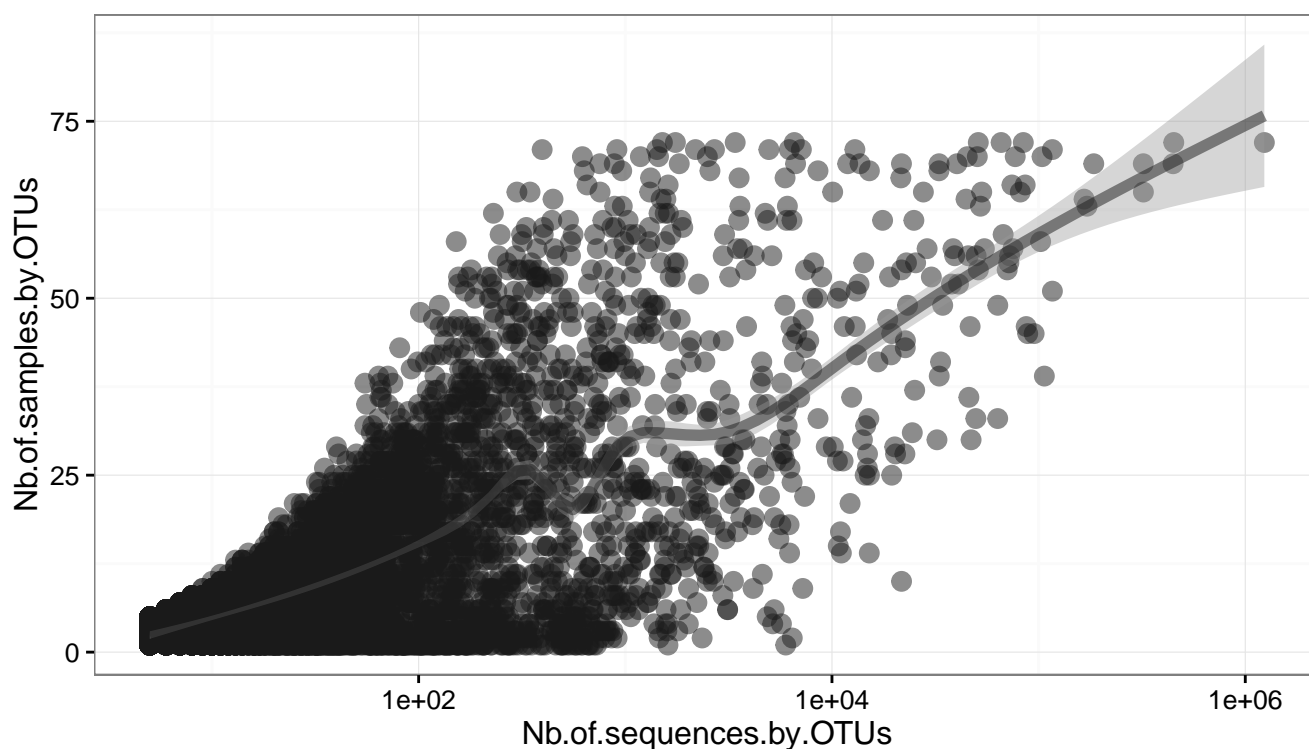


Figure 3.3: Number of sequences by OTUs (log10 axe) in fonction the number of samples where OTUs were found. The tendency is represented by the line obtain from gam (Generalized additive models with integrated smoothness estimation).

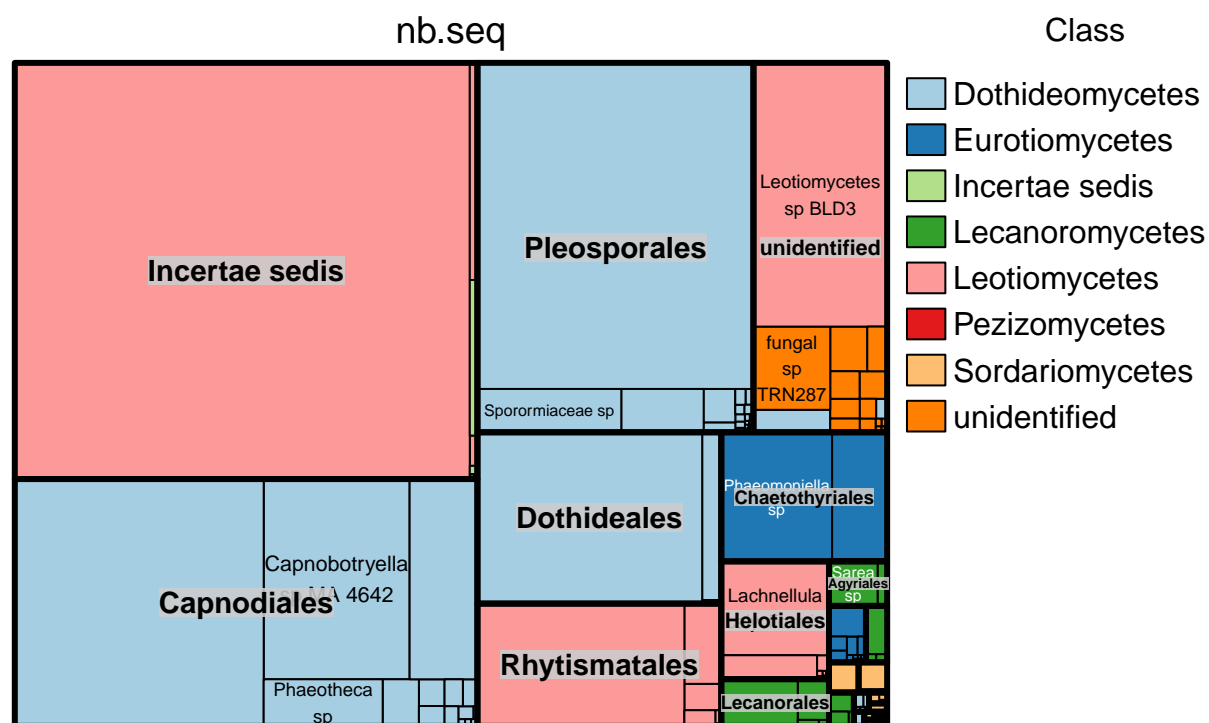


Figure 3.4: Distribution of the number of sequences in the taxonomy. Color represent Class, bold lines delimit Order and thick line delimit species.

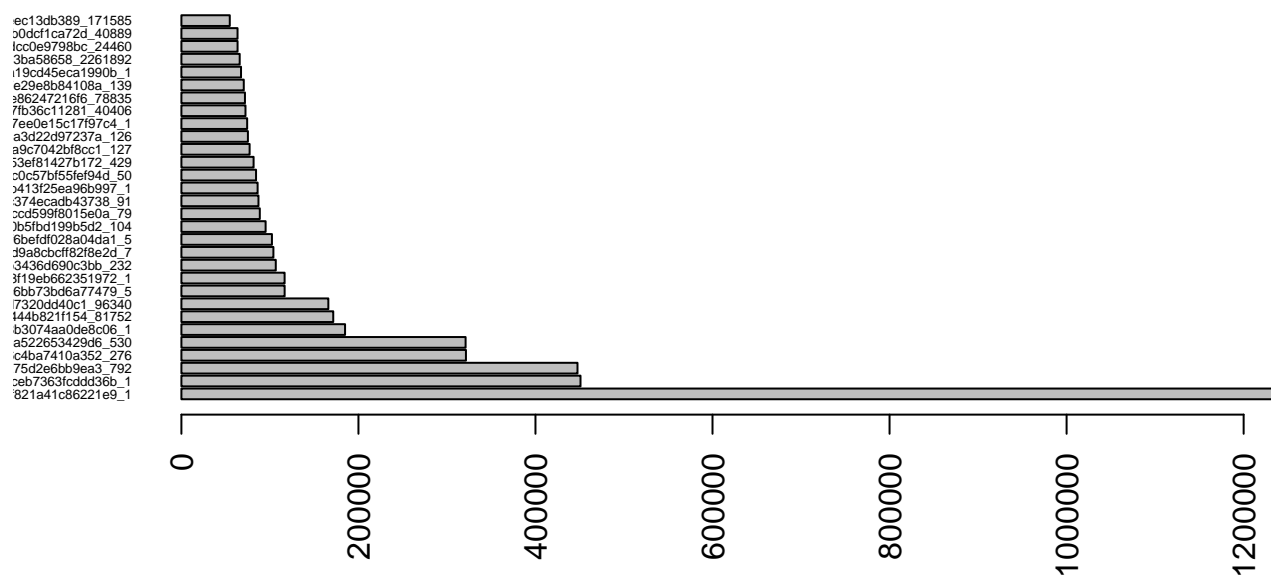


Figure 3.5: Number of sequences of the 30 more abundant OTUs (number of sequences).

```
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	-	-	1236989
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	450861
Ascomycota	Dothideomycetes	Pleosporales				-	-	447439
Ascomycota	Dothideomycetes	Pleosporales				-	-	321439
Ascomycota	Leotiomycetes	unidentified	unidentified	unidentified	Leotiomycetes sp BLD3	-	-	321041
Ascomycota	Dothideomycetes	Capnodiales				-	-	184871
Ascomycota	Dothideomycetes	Capnodiales	Incertae sedis	Capnobotryella	Capnobotryella sp MA 4642	Saprotroph	Undefined Saprotroph	171529
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp	-	-	165977
Ascomycota	Dothideomycetes	Pleosporales				-	-	116606
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	116577
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp	-	-	106621
						-	-	103925
						-	-	102289
Ascomycota	Dothideomycetes	Pleosporales				-	-	95146
Ascomycota	Dothideomycetes	Capnodiales	Incertae sedis	Capnobotryella	Capnobotryella sp MA 4642	Saprotroph	Undefined Saprotroph	88575
						-	-	87076
Ascomycota	Leotiomycetes	Rhytismatales	Rhytismataceae	Lophodermium		Pathotroph	Plant Pathogen	86099
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	84315
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	Dothioraceae sp	-	-	81555
Ascomycota	Dothideomycetes	Capnodiales				-	-	77060
Ascomycota	Leotiomycetes	unidentified	unidentified	unidentified	Leotiomycetes sp BLD3	-	-	75144
Ascomycota	Leotiomycetes	Rhytismatales	Rhytismataceae	Lophodermium		Pathotroph	Plant Pathogen	74295
						-	-	72400
						-	-	71797
Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Phaeothecoidea	Phaeothecoidea sp	Saprotroph	Undefined Saprotroph	70457
Ascomycota	Dothideomycetes	Capnodiales	Incertae sedis	Capnobotryella	Capnobotryella sp MA 4642	Saprotroph	Undefined Saprotroph	67353
Ascomycota	Leotiomycetes	Incertae sedis	Incertae sedis	Cyclaneusma	Cyclaneusma minus	-	-	65868
Ascomycota	Leotiomycetes	Helotiales	Hyaloscyphaceae	Lachnellula	Lachnellula calyciformis	Saprotroph	Undefined Saprotroph	63453
Ascomycota						-	-	63406

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

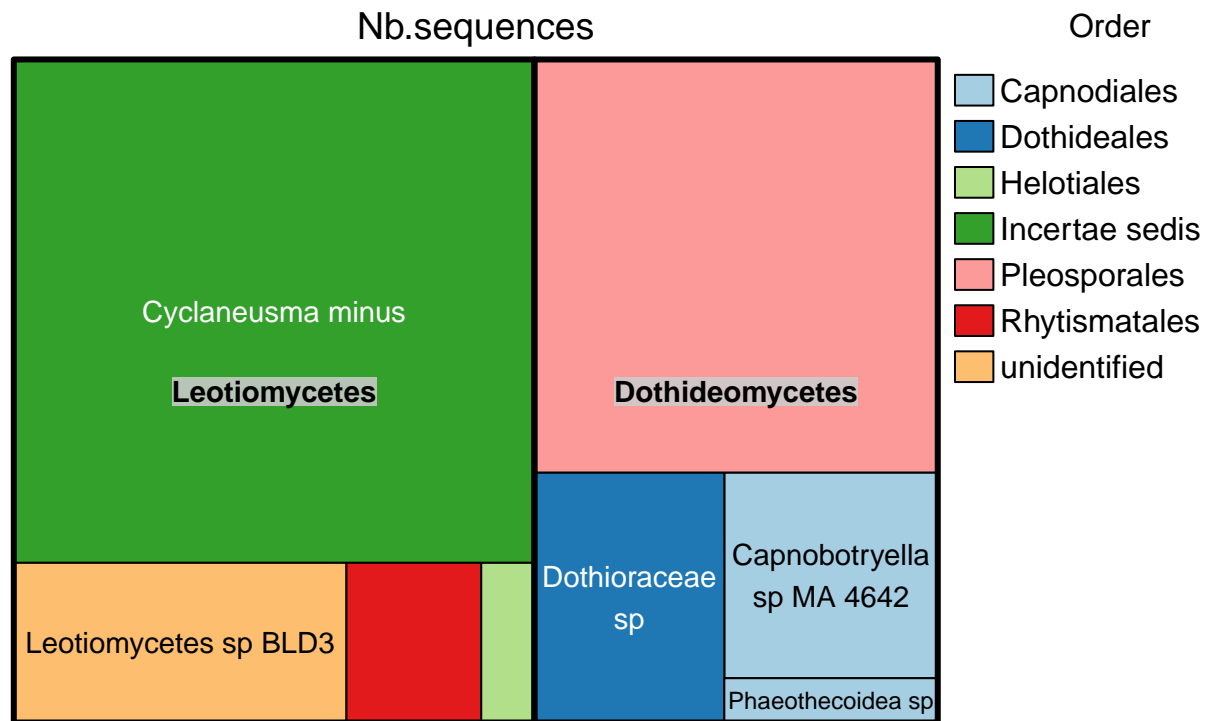


Figure 3.6: Number of sequences of the 30 most abundant OTUs (number of sequences). Colors indicate Order, bold lines delimit Class and thick lines delimit species.

```
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)
```

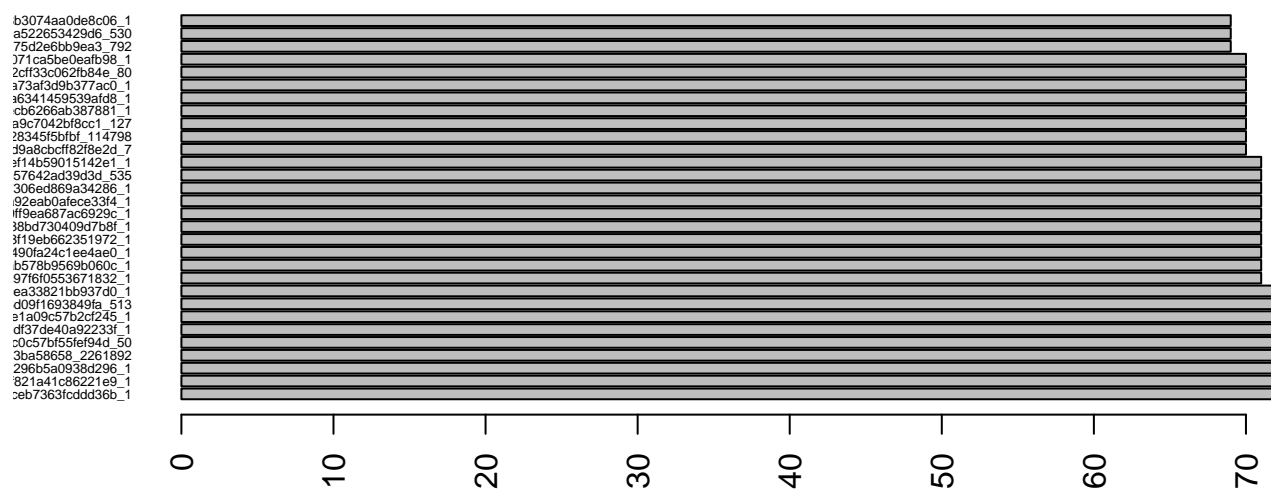



Figure 3.7: Number of samples of the 30 more frequent OTUs (number of samples).

```
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)
```

[illegible]

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

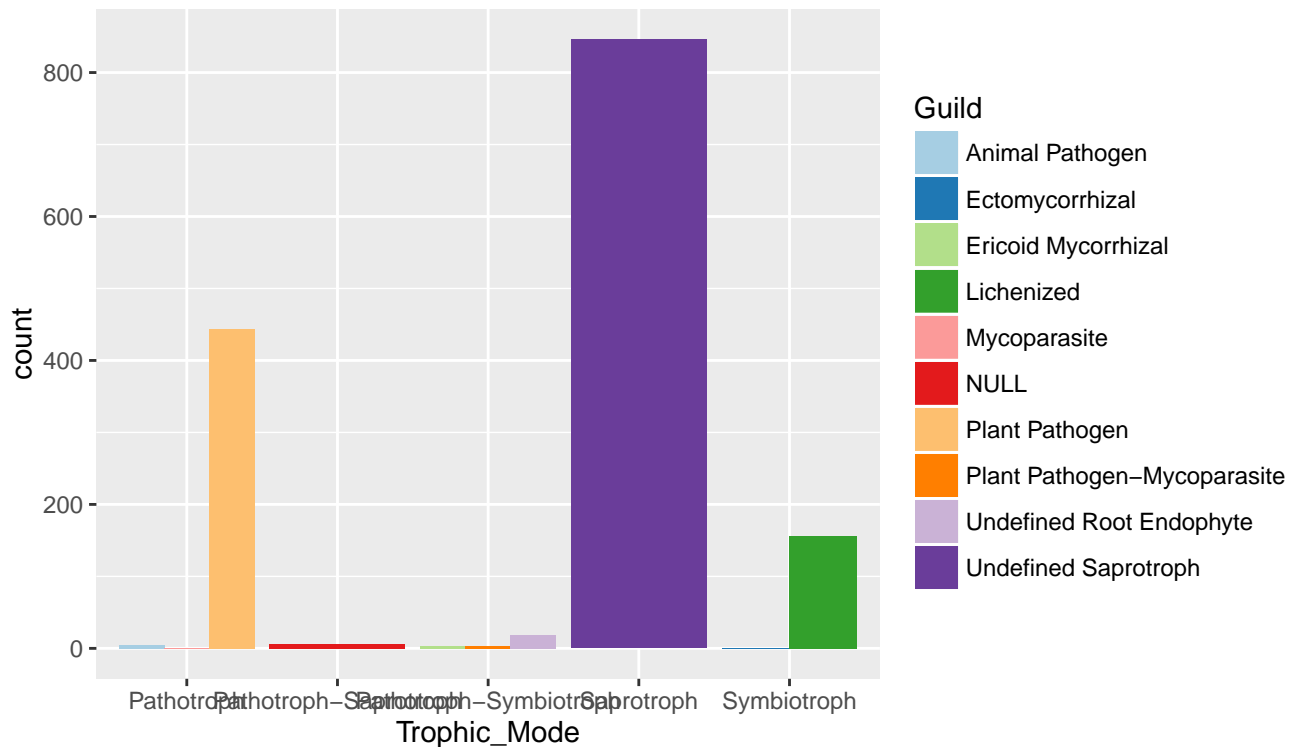


Figure 4.1: Distribution of OTUs into functional Guild.

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] 82.06287

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()
```

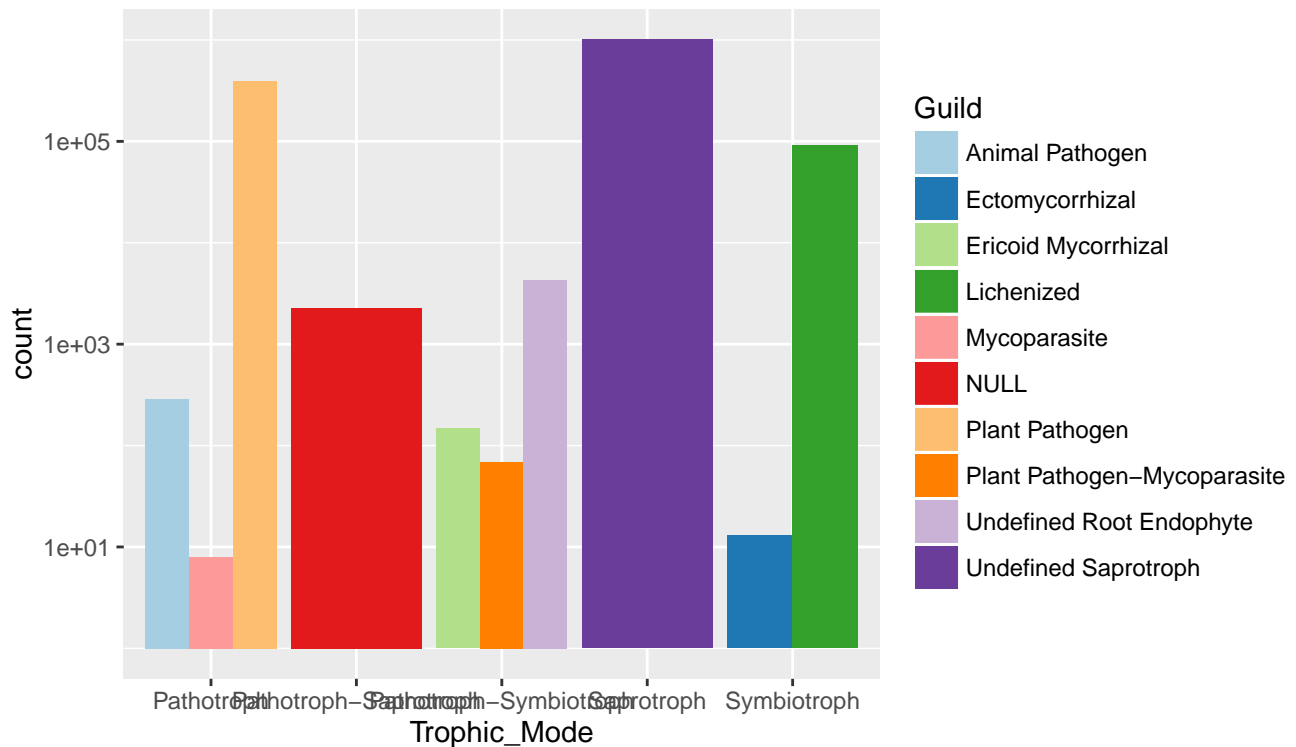


Figure 4.2: Distribution of sequences (log10 transformed) into functional Guild.

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)
```

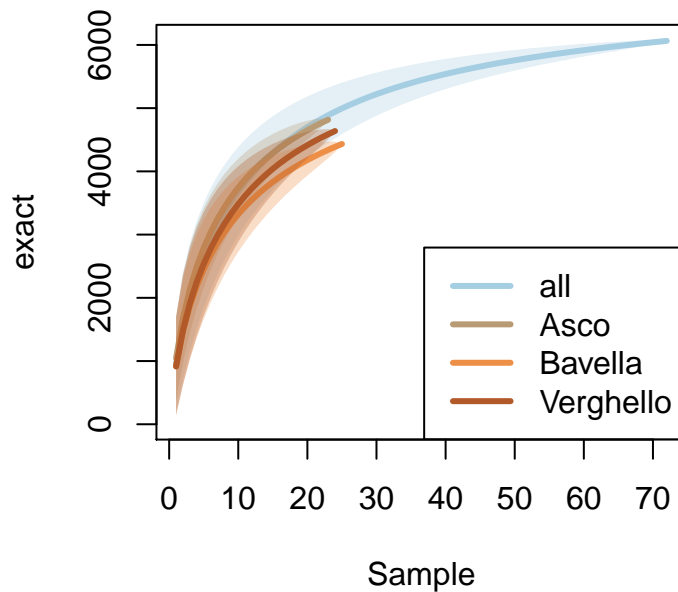


Figure 5.1: Rarefaction curves for each sites. Notes that if singletons were removed, these curves are biased.

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)
```

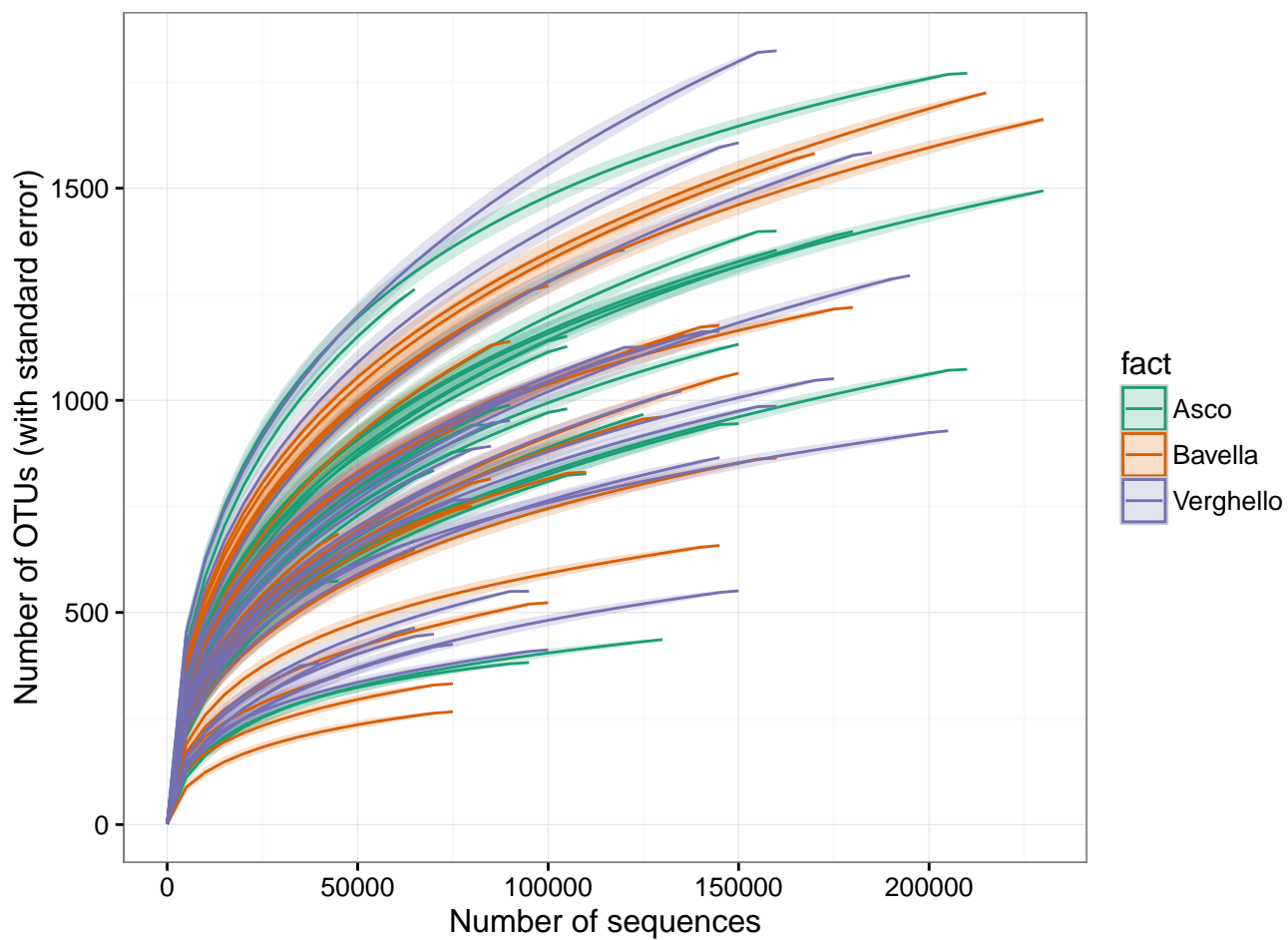


Figure 5.2: Rarefaction curves for each samples using sequences number on x-axes. Notes that if singletons were removed, these curves are biased.

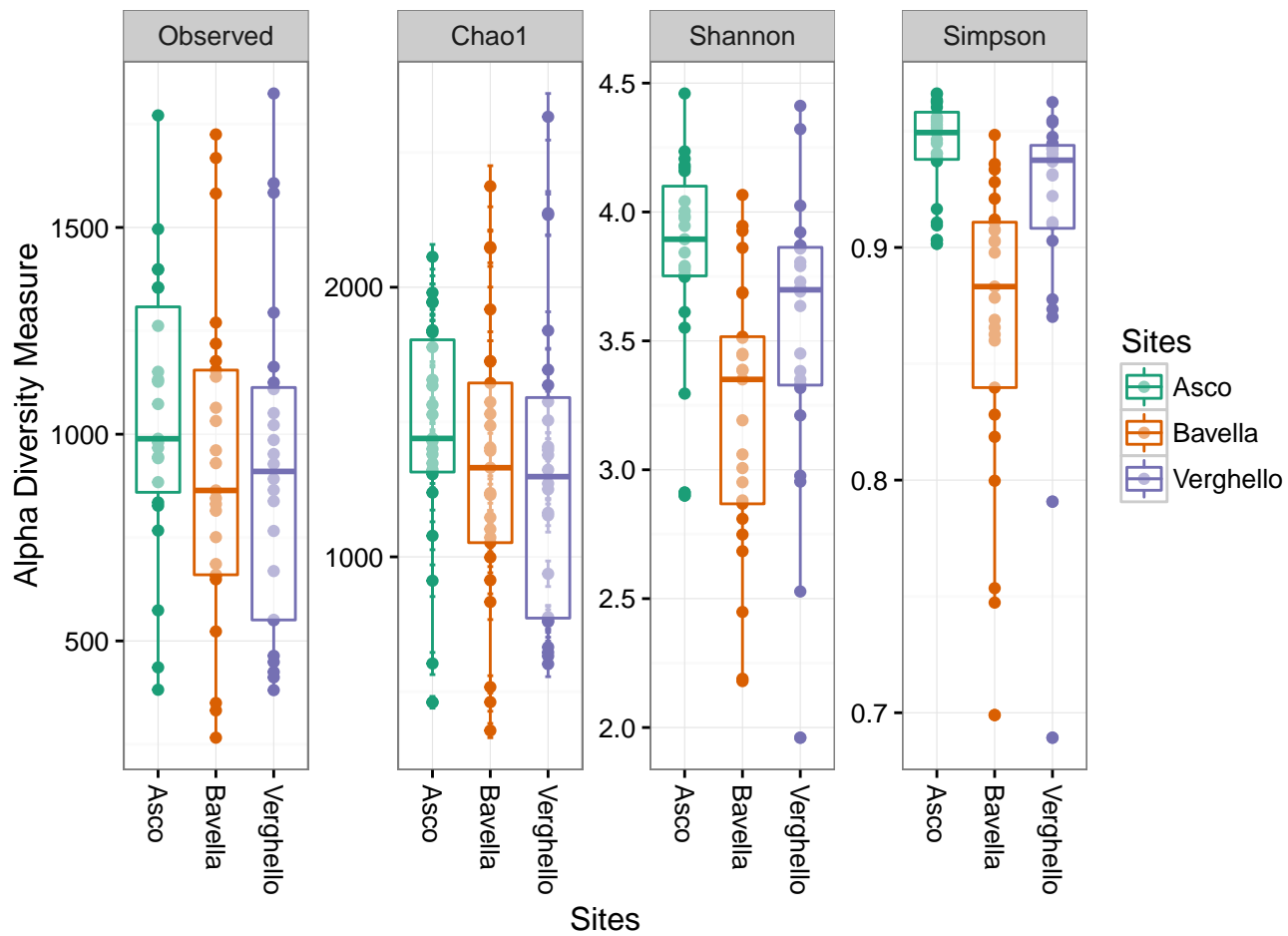


Figure 5.3: Diversity of each sites

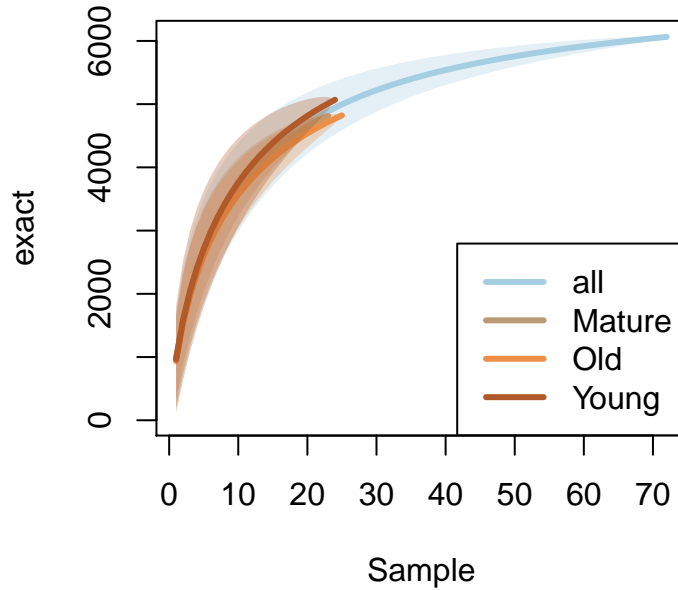


Figure 5.4: Rarefaction curves for each tree age modalities. Notes that if singletons were removed, these curves are biased.

```
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)	-169.7099523	180.4367338	-0.9405510	0.3504716
sqrt(readNumbers)	3.6352080	0.4587897	7.9234730	0.0000000
data.f3@sam_data\$SitesBavella	-49.6428188	80.4392050	-0.6171471	0.5393272
data.f3@sam_data\$SitesVerghello	-114.7273185	80.5393450	-1.4244879	0.1591634
data.f3@sam_data\$AgeOld	-14.6366862	80.0911628	-0.1827503	0.8555716
data.f3@sam_data\$AgeYoung	-120.4101491	82.0227984	-1.4680083	0.1469997
data.f3@sam_data\$ElevationMiddle	70.9260851	81.1127439	0.8744136	0.3851626
data.f3@sam_data\$ElevationTop	-14.1869138	80.0164374	-0.1773000	0.8598327

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))
```

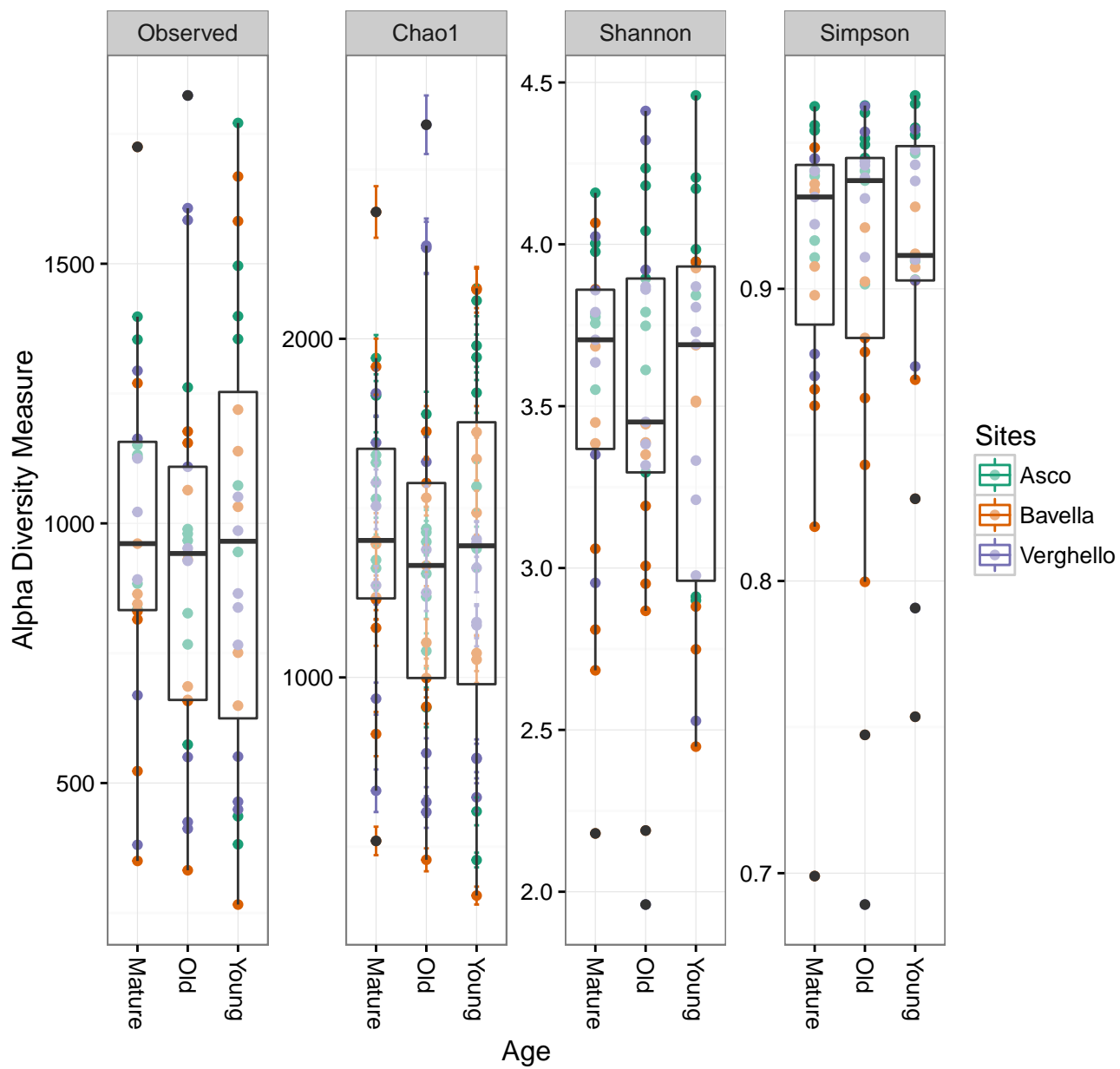



Figure 5.5: Diversity in function of tree age. Color represent sites.

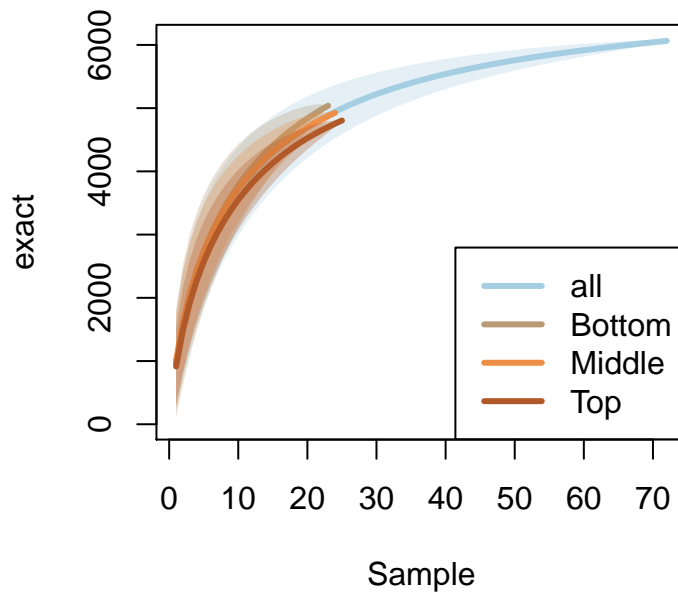


Figure 5.6: Rarefaction curves for each elevation. Notes that if singletons were removed, these curves are biased.

```
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 MultiDimensionstional Scaling).

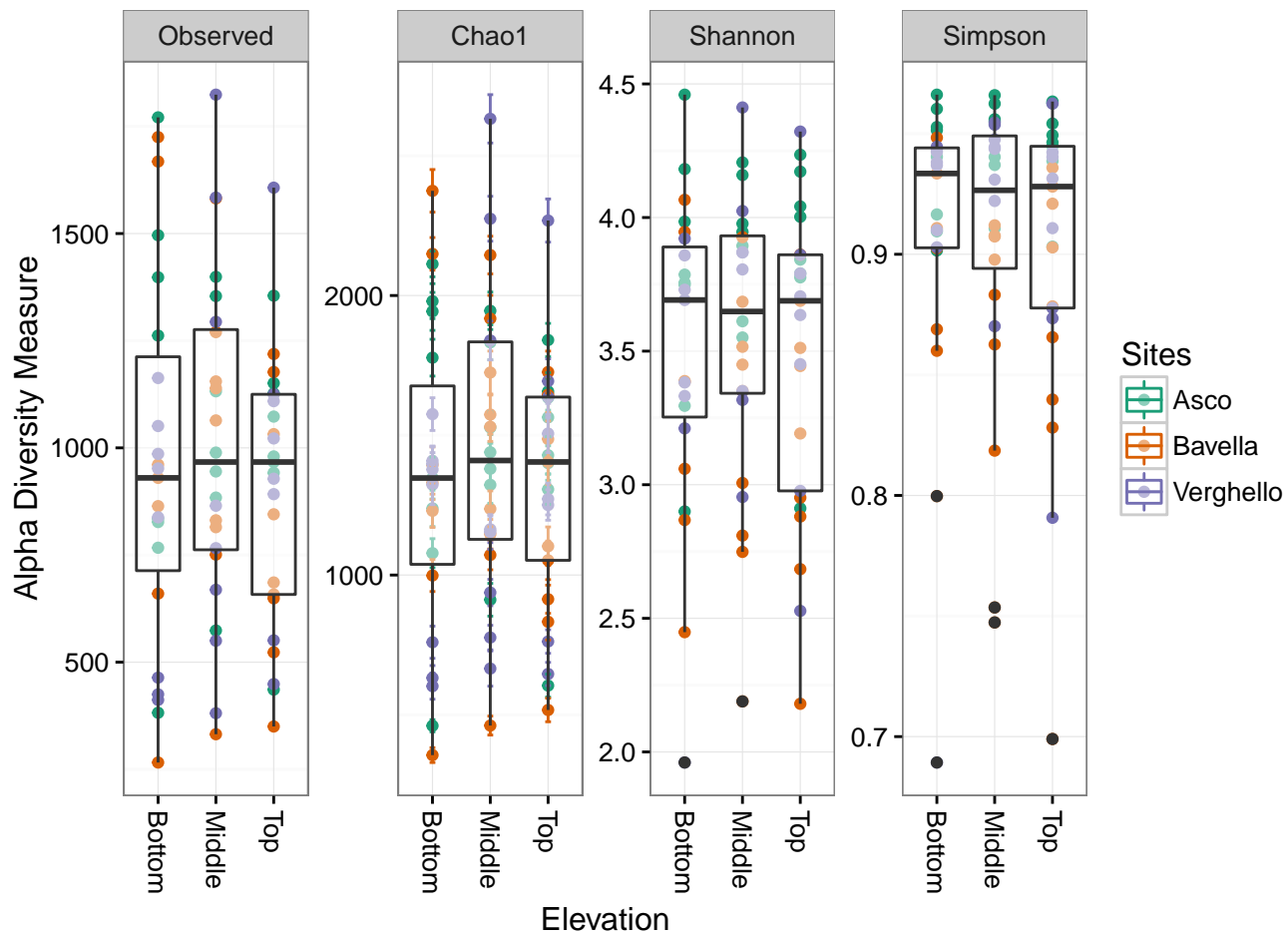


Figure 5.7: Diversity in function of elevation. Color represent sites.

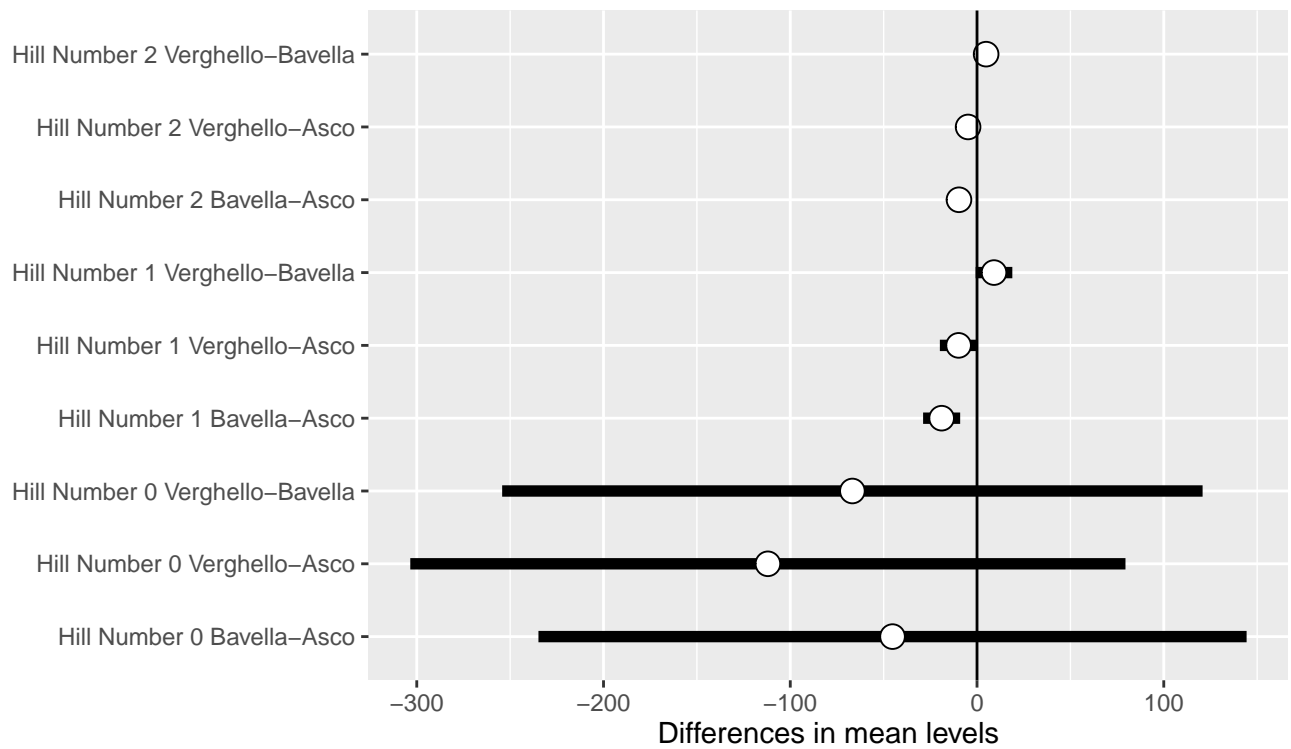


Figure 5.8: Results of the Tuckey HSD testing for differences in mean Hill numbers among pairs of modalities

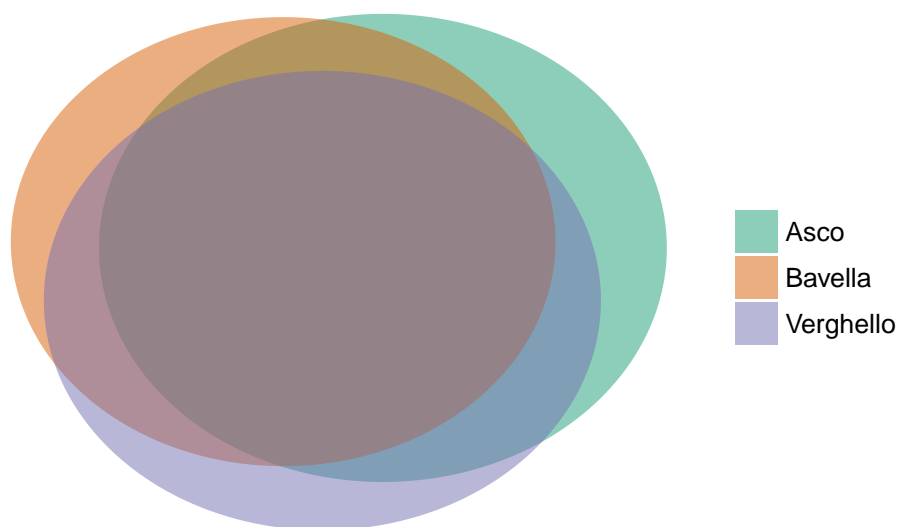


Figure 6.1: Venn diagramm of the distribution of OTUs among Sites

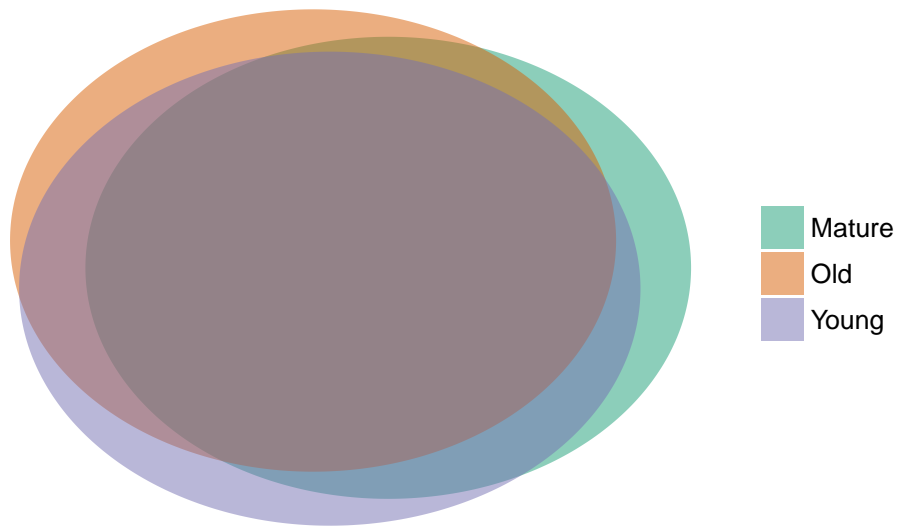


Figure 6.2: Venn diagramm ef the distribution of OTUs among host age

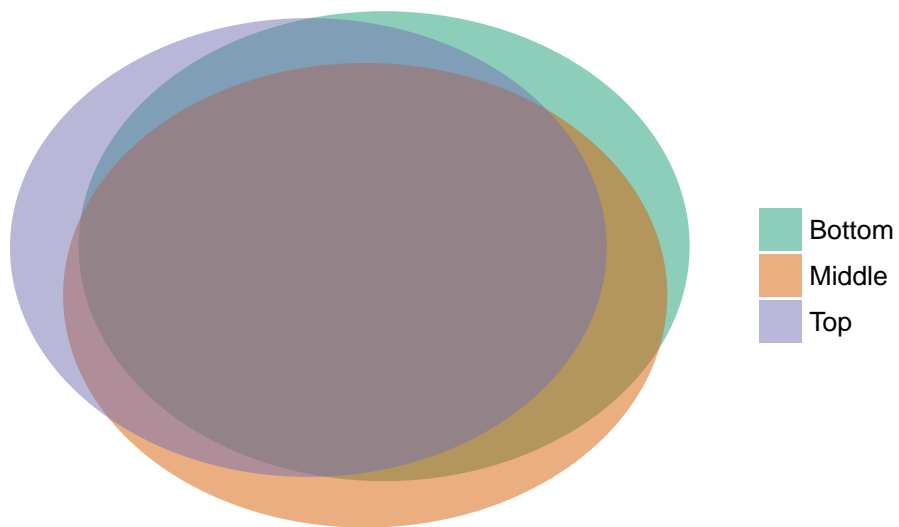


Figure 6.3: Venn diagramm ef the distribution of OTUs among elevation of samples

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	12.9703294	9.5718525	1.3550490	0.1801656
sqrt(readNumbers)	0.1035362	0.0243380	4.2540991	0.0000698
data.f3@sam_data\$SitesBavella	-19.4074863	4.2671589	-4.5481049	0.0000247
data.f3@sam_data\$SitesVerghello	-10.0679214	4.2724711	-2.3564633	0.0215228
data.f3@sam_data\$AgeOld	0.7542946	4.2486958	0.1775356	0.8596484
data.f3@sam_data\$AgeYoung	-3.3431380	4.3511657	-0.7683316	0.4451161
data.f3@sam_data\$ElevationMiddle	4.0435791	4.3028889	0.9397359	0.3508864
data.f3@sam_data\$ElevationTop	1.7596500	4.2447318	0.4145492	0.6798582

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)	10.2095670	3.3737936	3.0261386	0.0035652
sqrt(readNumbers)	0.0258823	0.0085784	3.0171465	0.0036589
data.f3@sam_data\$SitesBavella	-9.8360012	1.5040467	-6.5396913	0.0000000
data.f3@sam_data\$SitesVerghello	-4.8631943	1.5059191	-3.2293861	0.0019593
data.f3@sam_data\$AgeOld	0.4809093	1.4975390	0.3211331	0.7491559
data.f3@sam_data\$AgeYoung	-0.1876484	1.5336566	-0.1223536	0.9030024
data.f3@sam_data\$ElevationMiddle	1.0405173	1.5166405	0.6860672	0.4951487
data.f3@sam_data\$ElevationTop	0.2540789	1.4961418	0.1698227	0.8656852

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

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

```
stressplot(my.ord.nmfs)
```

```
p <- plot_ordination(data.f3, my.ord.nmfs, 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.nmfs_gower <- ordinate(data.f3, distance = "gower", method = "NMDS")

## Square root transformation
## Wisconsin double standardization
## Run 0 stress 0.1900518
## Run 1 stress 0.1895783
## ... New best solution
## ... Procrustes: rmse 0.01072136 max resid 0.05645879
## Run 2 stress 0.1900617
## ... Procrustes: rmse 0.02308659 max resid 0.1179364
## Run 3 stress 0.2146498
```



```

## Run 4 stress 0.1900447
## ... Procrustes: rmse 0.0232604  max resid 0.1194122
## Run 5 stress 0.408333
## Run 6 stress 0.1900308
## ... Procrustes: rmse 0.02318525  max resid 0.1189706
## Run 7 stress 0.194593
## Run 8 stress 0.2212027
## Run 9 stress 0.1907872
## Run 10 stress 0.2167984
## Run 11 stress 0.1908176
## Run 12 stress 0.2208554
## Run 13 stress 0.1909247
## Run 14 stress 0.190119
## Run 15 stress 0.1900087
## ... Procrustes: rmse 0.01190498  max resid 0.05732117
## Run 16 stress 0.1906368
## Run 17 stress 0.2004351
## Run 18 stress 0.1900346
## ... Procrustes: rmse 0.02320311  max resid 0.119131
## Run 19 stress 0.1907752
## Run 20 stress 0.1901383
## *** No convergence -- monoMDS stopping criteria:
##      20: 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)

```

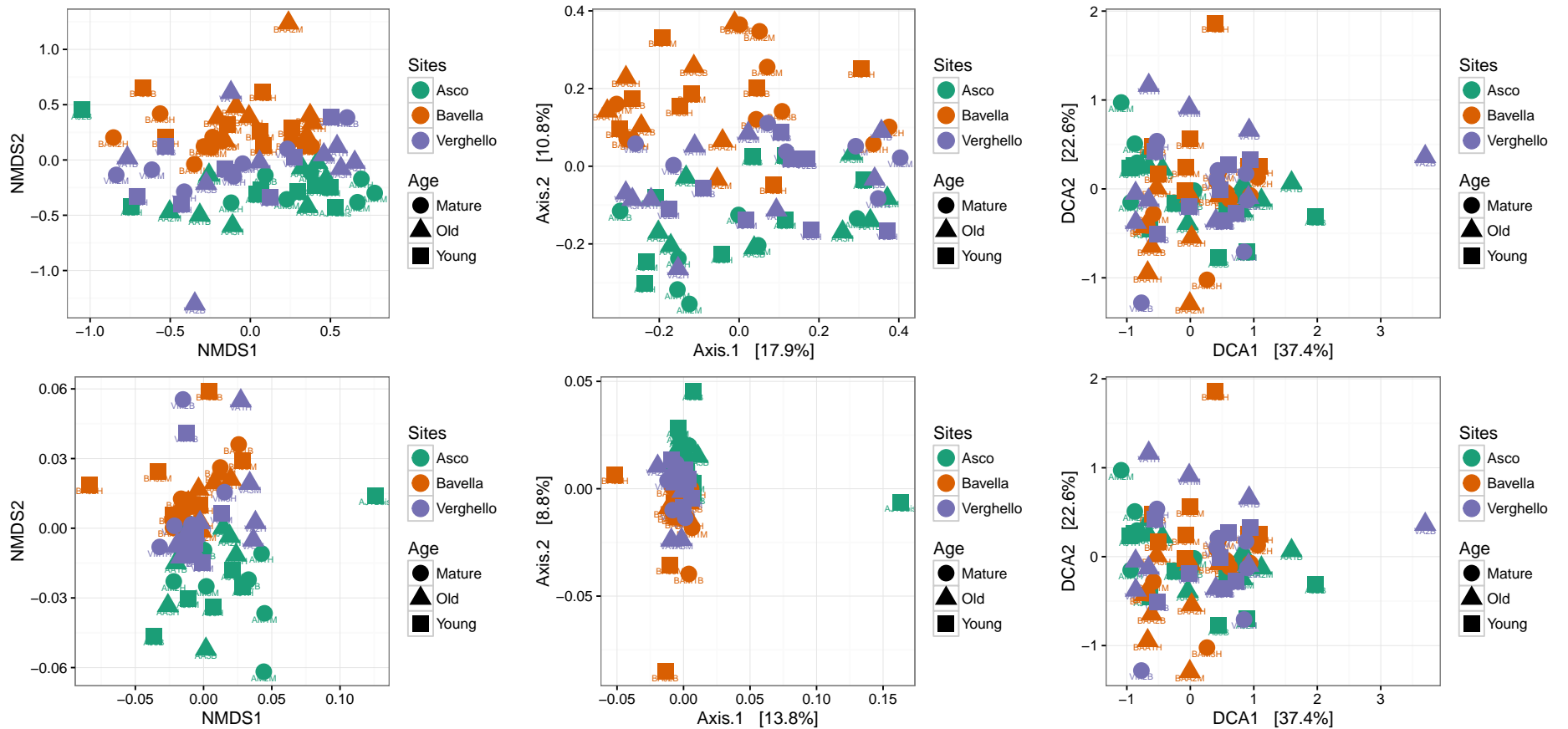



Figure 6.6: Comparison of different distances (bray (up) and gower (bottom)) and ordination methods (NMDS (left), PCoA (center) and DCA (right)).

```
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 583 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	1.91	0.96	4.19	0.10	0.0001
Age	2	0.67	0.34	1.47	0.04	0.0464
Elevation	2	0.54	0.27	1.18	0.03	0.2010
Sites:Age	4	1.55	0.39	1.69	0.08	0.0013
Sites:Elevation	4	0.91	0.23	0.99	0.05	0.4752
Age:Elevation	4	1.10	0.27	1.20	0.06	0.1224
Sites:Age:Elevation	8	1.85	0.23	1.01	0.10	0.4330
Residuals	45	10.27	0.23		0.55	
Total	71	18.79			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	1.86	0.93	4.48	0.11	0.0001
Age	2	0.63	0.32	1.52	0.04	0.0474
Elevation	2	0.51	0.25	1.22	0.03	0.1851
Sites:Age	4	1.48	0.37	1.78	0.09	0.0016
Sites:Elevation	4	0.85	0.21	1.02	0.05	0.4205
Age:Elevation	4	1.04	0.26	1.24	0.06	0.1074
Sites:Age:Elevation	8	1.69	0.21	1.02	0.10	0.4231
Residuals	45	9.35	0.21		0.54	
Total	71	17.42			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	1.28	0.64	3.40	0.09	0.0001
Age	2	0.52	0.26	1.38	0.04	0.0407
Elevation	2	0.40	0.20	1.06	0.03	0.3191
Sites:Age	4	1.09	0.27	1.44	0.07	0.0065
Sites:Elevation	4	0.64	0.16	0.85	0.04	0.8845
Age:Elevation	4	0.88	0.22	1.18	0.06	0.1118
Sites:Age:Elevation	8	1.51	0.19	1.01	0.10	0.4519
Residuals	45	8.45	0.19		0.57	
Total	71	14.76			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	1.91	0.96	4.67	0.10	0.0001
Age	2	0.67	0.34	1.64	0.04	0.0190
Sites:Age	4	1.54	0.39	1.88	0.08	0.0005
Sites:Age:IndividualTree	18	5.45	0.30	1.48	0.29	0.0001
Residuals	45	9.22	0.20		0.49	
Total	71	18.79			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	1.86	0.93	5.00	0.11	0.0001
Age	2	0.63	0.32	1.70	0.04	0.0147
Sites:Age	4	1.48	0.37	1.98	0.08	0.0003
Sites:Age:IndividualTree	18	5.07	0.28	1.51	0.29	0.0001
Residuals	45	8.38	0.19		0.48	
Total	71	17.42			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	1.28	0.64	3.73	0.09	0.0001
Age	2	0.52	0.26	1.52	0.04	0.0162
Sites:Age	4	1.10	0.28	1.61	0.07	0.0005
Sites:Age:IndividualTree	18	4.16	0.23	1.35	0.28	0.0001
Residuals	45	7.70	0.17		0.52	
Total	71	14.76			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.

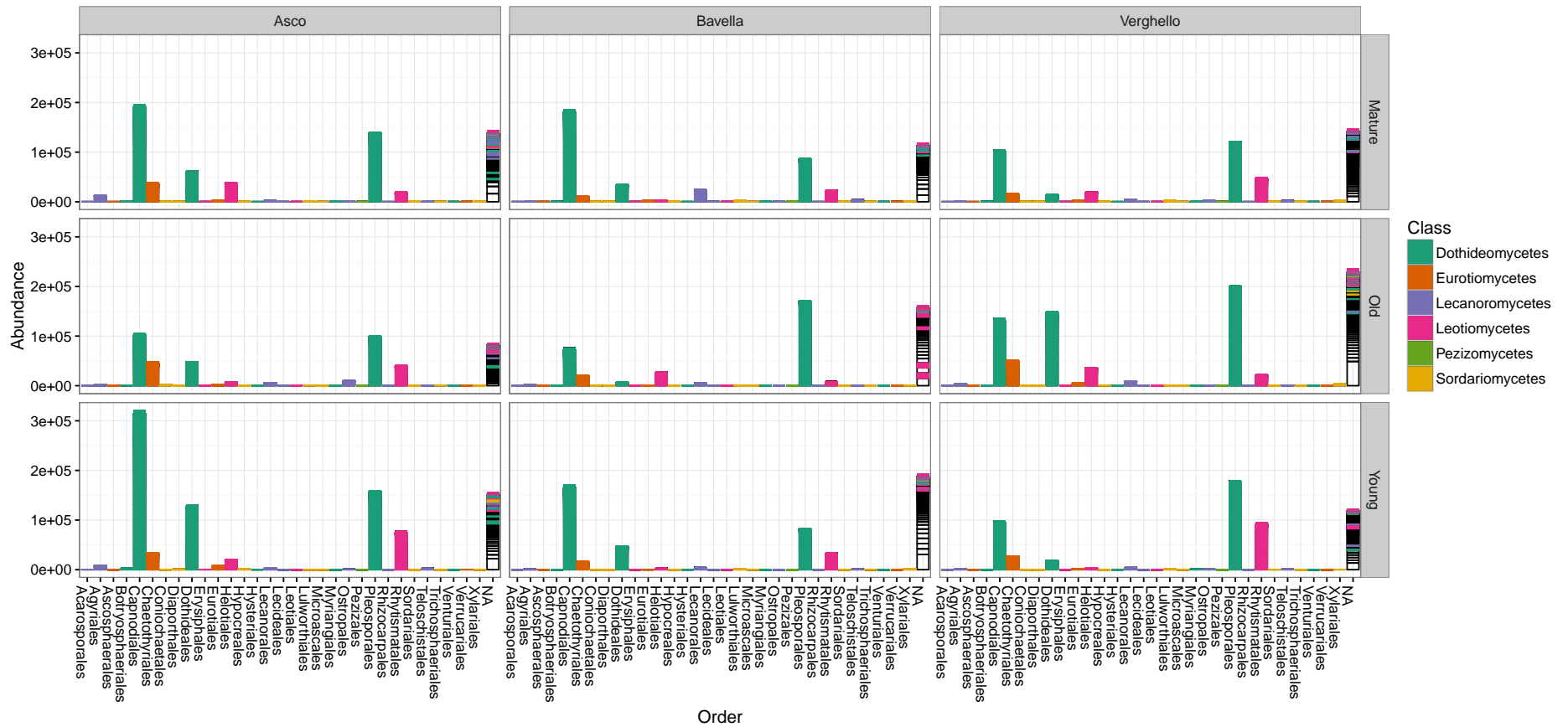


Figure 6.7: Taxonomic distribution of sequences in the different site * age combination.

```
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.binary0tuTable(data.f3_taxo_known), "Order", fill = "Class",
              facet_grid = Age ~ Sites)
p + geom_bar(aes(color = Class, fill = Class), stat = "identity", position = "stack")
```

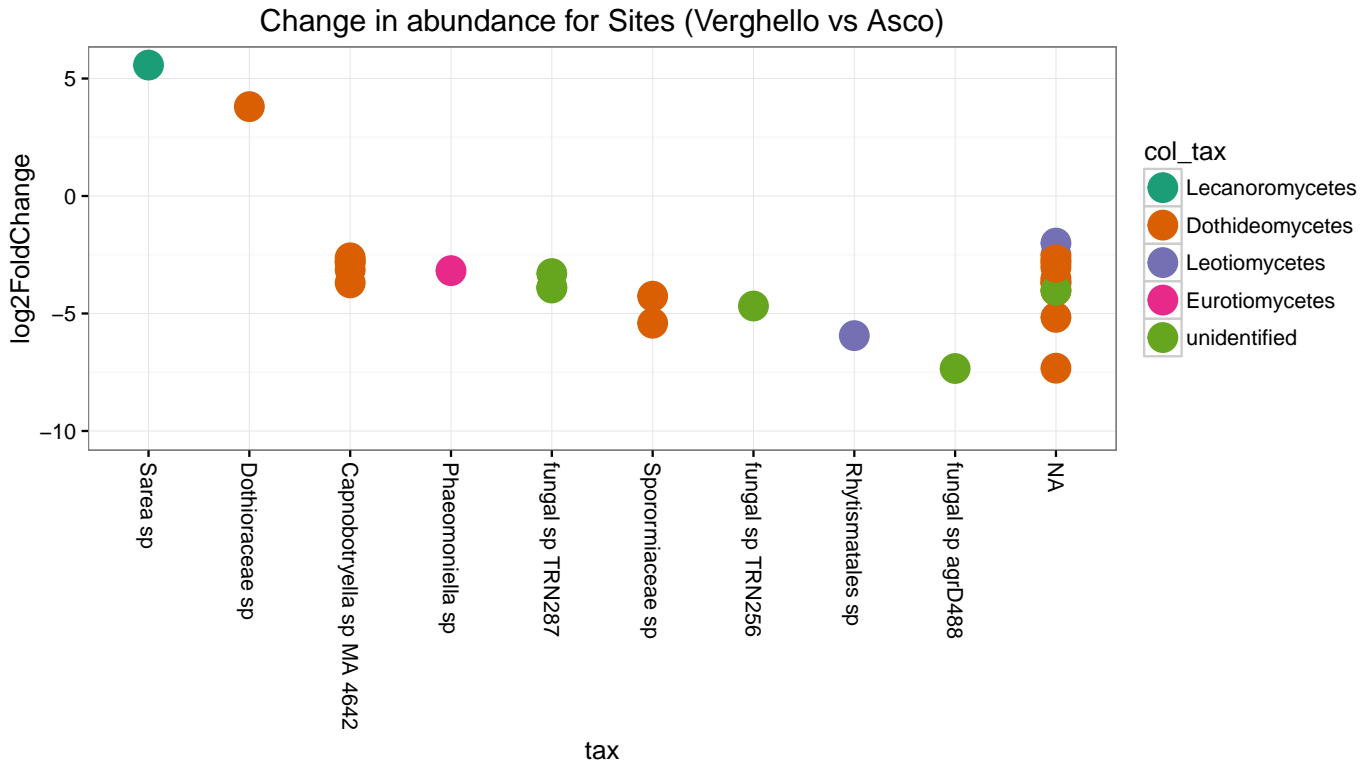


Figure 6.9: OTUs significantly different in terms of abundances between Verghello (positive values) and Asco (negative values)

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
```

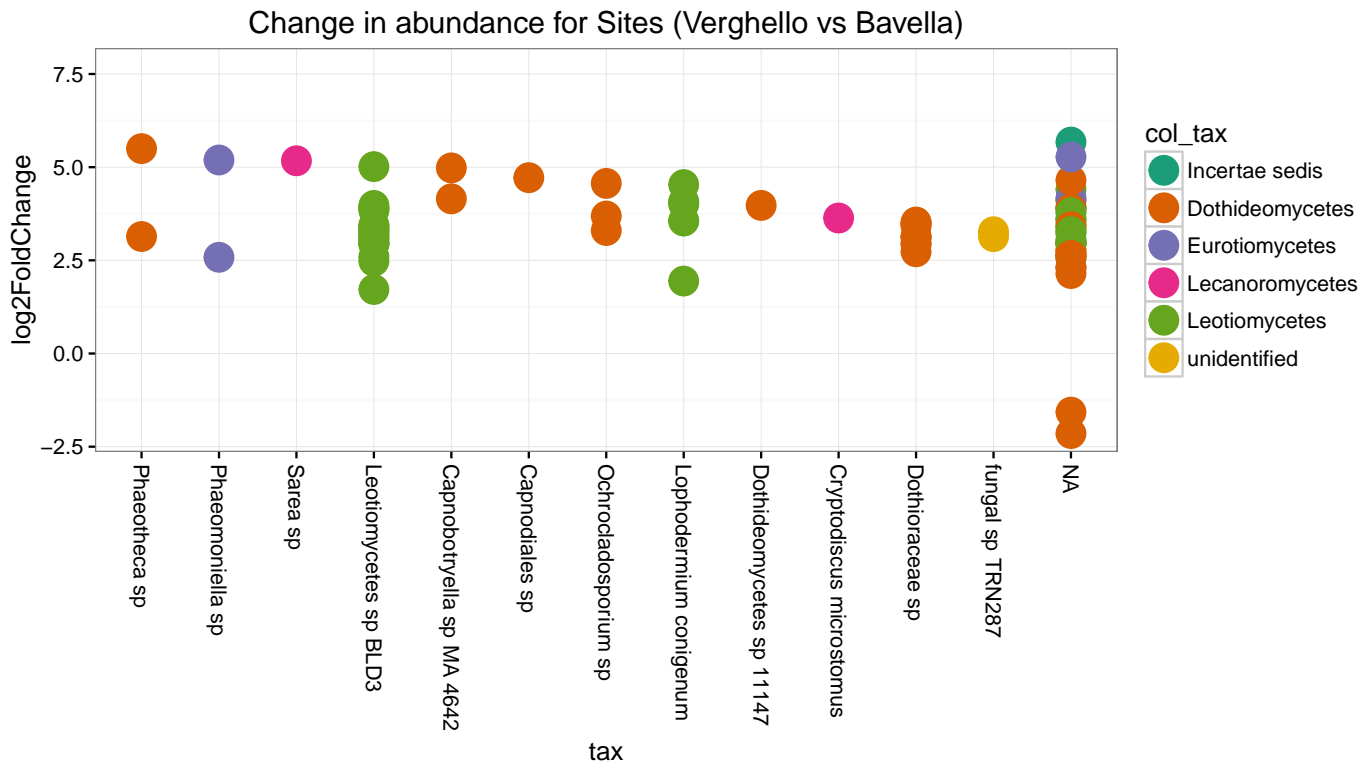


Figure 6.10: OTUs significantly different in terms of abundances between Verghello (positive values) and Bavella (negative values)



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	Capnobotryella sp MA 4642	Dothideomycetes	-2.82646845973802
2	Verghello vs Asco			-4.98457809309708
3	Verghello vs Asco		Dothideomycetes	-2.51650664598345
4	Verghello vs Asco		Dothideomycetes	-4.03376881393835
5	Verghello vs Asco	Capnobotryella sp MA 4642	Dothideomycetes	-2.77438198853472
6	Verghello vs Asco	Phaeomoniella sp	Eurotiomycetes	-3.17622586809418
7	Verghello vs Asco		Dothideomycetes	-3.72669291915564
8	Verghello vs Asco			-2.55672516578965
9	Verghello vs Asco	fungus sp TRN287	unidentified	-3.30082443234941
10	Verghello vs Asco		Dothideomycetes	-3.66001833843314
11	Verghello vs Asco	Capnobotryella sp MA 4642	Dothideomycetes	-2.63077258786098
12	Verghello vs Asco	fungus sp TRN256	unidentified	-4.68188552477027
13	Verghello vs Asco		Dothideomycetes	-5.16302154955906
14	Verghello vs Asco		Dothideomycetes	-7.32777284814396
15	Verghello vs Asco		Dothideomycetes	-2.84709177074018
16	Verghello vs Asco		Leotiomyces	-2.00141879761371
17	Verghello vs Asco	Dothioraceae sp	Dothideomycetes	3.80406500630475
18	Verghello vs Asco	Capnobotryella sp MA 4642	Dothideomycetes	-3.14561353870513
19	Verghello vs Asco		Dothideomycetes	-3.51129939445095
20	Verghello vs Asco	Capnobotryella sp MA 4642	Dothideomycetes	-3.69547275148547
21	Verghello vs Asco	Rhytismatales sp	Leotiomyces	-5.94204822053169
22	Verghello vs Asco	fungus sp TRN287	unidentified	-3.88581221819408
23	Verghello vs Asco	fungus sp agrD488	unidentified	-7.34108022555751
24	Verghello vs Asco	fungus sp TRN287	unidentified	-3.91602588975261
25	Verghello vs Asco			-4.82864555968206
26	Verghello vs Asco			-3.22679203448171
27	Verghello vs Asco			-3.30700257518351
28	Verghello vs Asco		unidentified	-4.01297310479197
29	Verghello vs Asco			-10.0465890664006
30	Verghello vs Asco		Dothideomycetes	-2.71040241828603
31	Verghello vs Asco		Dothideomycetes	-3.03569941110135
32	Verghello vs Asco	Sarea sp	Lecanoromycetes	5.57328226472331
33	Verghello vs Asco	Sporormiaceae sp	Dothideomycetes	-5.41289489049314
34	Verghello vs Asco			-5.73754583795947
35	Verghello vs Asco	Sporormiaceae sp	Dothideomycetes	-4.26353592719463
36	Verghello vs Bavella	Phaeomoniella sp	Eurotiomycetes	5.19257960794333
37	Verghello vs Bavella		Leotiomyces	3.93866418481676
38	Verghello vs Bavella	fungus sp TRN287	unidentified	3.26865081952064
39	Verghello vs Bavella	Capnobotryella sp MA 4642	Dothideomycetes	4.15097921643878
40	Verghello vs Bavella		Dothideomycetes	2.64183958272059
41	Verghello vs Bavella	Cryptodiscus microstomus	Lecanoromycetes	3.63596858179856
42	Verghello vs Bavella		Dothideomycetes	4.10533802187722
43	Verghello vs Bavella	Leotiomyces sp BLD3	Leotiomyces	3.37554264001515
44	Verghello vs Bavella		Incertain sedis	5.67378319467808
45	Verghello vs Bavella	Capnobotryella sp MA 4642	Dothideomycetes	4.98048933156314
46	Verghello vs Bavella	Dothioraceae sp	Dothideomycetes	3.48008812367165
47	Verghello vs Bavella		Leotiomyces	3.87947902464684
48	Verghello vs Bavella		Leotiomyces	4.40205298238785
49	Verghello vs Bavella	Phaeotheca sp	Dothideomycetes	5.49769984959994
50	Verghello vs Bavella			5.56312062256807
51	Verghello vs Bavella		Dothideomycetes	2.551768316934
52	Verghello vs Bavella		Leotiomyces	3.05619301499366
53	Verghello vs Bavella	Capnodiales sp	Dothideomycetes	4.71765201771909
54	Verghello vs Bavella	Leotiomyces sp BLD3	Leotiomyces	1.71818280676113
55	Verghello vs Bavella	fungus sp TRN287	unidentified	3.1360510420606
56	Verghello vs Bavella		Eurotiomycetes	4.16713914508627
57	Verghello vs Bavella	Ochrocladosporium sp	Dothideomycetes	4.56465629351246
58	Verghello vs Bavella	Lophodermium conigenum	Leotiomyces	4.53295672687564
59	Verghello vs Bavella			3.2765757079723
60	Verghello vs Bavella			3.87479020766316
61	Verghello vs Bavella		Dothideomycetes	-1.57292468491607
62	Verghello vs Bavella	Ochrocladosporium sp	Dothideomycetes	3.68889950719099
63	Verghello vs Bavella			3.68566252341367
64	Verghello vs Bavella		Leotiomyces	2.57463197435258
65	Verghello vs Bavella		Dothideomycetes	2.95598887741301
66	Verghello vs Bavella	Lophodermium conigenum	Leotiomyces	1.94416574628666
67	Verghello vs Bavella		Dothideomycetes	2.94456717026919
68	Verghello vs Bavella		Dothideomycetes	3.8856220203889
69	Verghello vs Bavella	Leotiomyces sp BLD3	Leotiomyces	3.42234118047653
70	Verghello vs Bavella			7.70154105845669
71	Verghello vs Bavella	Leotiomyces sp BLD3	Leotiomyces	3.28128666279013
72	Verghello vs Bavella	Dothioraceae sp	Dothideomycetes	3.14509781311276
73	Verghello vs Bavella		Leotiomyces	2.93746971551255
74	Verghello vs Bavella			4.11445735665553
75	Verghello vs Bavella		Dothideomycetes	2.77240989563688
76	Verghello vs Bavella	Leotiomyces sp BLD3	Leotiomyces	3.09420886913011
77	Verghello vs Bavella	Leotiomyces sp BLD3	Leotiomyces	2.92977490277754
78	Verghello vs Bavella			3.15860228260441
79	Verghello vs Bavella	Phaeomoniella sp	Eurotiomycetes	2.58200083354608
80	Verghello vs Bavella		Dothideomycetes	2.30839238767509
81	Verghello vs Bavella		Leotiomyces	3.36206601161872
82	Verghello vs Bavella	Leotiomyces sp BLD3	Leotiomyces	2.47535759023292
83	Verghello vs Bavella	Dothioraceae sp	Dothideomycetes	3.12572669535196
84	Verghello vs Bavella			4.04626486803813
85	Verghello vs Bavella	Leotiomyces sp BLD3	Leotiomyces	3.97289554230687
86	Verghello vs Bavella			3.08852454092464
87	Verghello vs Bavella	Leotiomyces sp BLD3	Leotiomyces	2.95328218194723
88	Verghello vs Bavella	Leotiomyces sp BLD3	Leotiomyces	3.90071070730146
89	Verghello vs Bavella		Leotiomyces	3.60707405870496
90	Verghello vs Bavella		Leotiomyces	3.47955179226687
91	Verghello vs Bavella	Leotiomyces sp BLD3	Leotiomyces	5.01459729397156
92	Verghello vs Bavella	Leotiomyces sp BLD3	Leotiomyces	3.20489417303019
93	Verghello vs Bavella		Dothideomycetes	2.75005373078387
94	Verghello vs Bavella			3.08764319088441
95	Verghello vs Bavella	Lophodermium conigenum	Leotiomyces	4.07821418521343
96	Verghello vs Bavella	Leotiomyces sp BLD3	Leotiomyces	2.96902863592455
97	Verghello vs Bavella		Dothideomycetes	2.14597644808271
98	Verghello vs Bavella	Lophodermium conigenum	Leotiomyces	4.00793865152818
99	Verghello vs Bavella			3.03971179808403
100	Verghello vs Bavella			3.37570089351978
101	Verghello vs Bavella			4.49422397277345
102	Verghello vs Bavella		Leotiomyces	2.9572482723965
103	Verghello vs Bavella	Leotiomyces sp BLD3	Leotiomyces	3.16351187980077
104	Verghello vs Bavella		Leotiomyces	3.78213477447987
105	Verghello vs Bavella	Ochrocladosporium sp	Dothideomycetes	3.30058293279581
106	Verghello vs Bavella	Dothioraceae sp	Dothideomycetes	2.72972523716246
107	Verghello vs Bavella		Leotiomyces	3.50968807407409

	Comparison	Order	Class	log2FoldChange (negative = more on second levels)
1	Verghello vs Asco	Xylariales	Sordariomycetes	4.1558818804718
2	Verghello vs Bavella	Capnodiales	Dothideomycetes	-0.831462437740621
3	Verghello vs Bavella	Incertae sedis	Leotiomycetes	-1.24211406337419
4	Verghello vs Bavella	Ostropales	Lecanoromycetes	3.47293291293366
5	Verghello vs Bavella	unidentified	unidentified	1.54674641158209
6	Asco vs Bavella	Botryosphaerales	Dothideomycetes	5.78231220589885
7	Asco vs Bavella	Eurotiales	Eurotiomycetes	1.80504515685601
8	Asco vs Bavella	Incertae sedis	Leotiomycetes	-1.66112382659302
9	Asco vs Bavella	unidentified	unidentified	1.46739532752367
10	Asco vs Bavella	Xylariales	Sordariomycetes	-4.09136027089075

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

List of Figures

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

List of Tables

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

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