Metagenomics Analysis on Fungal Communities in Stone Fruit Ecosystems in Canada

Minuka Hewapathirana

2024-07-19

Contents

1	Loading Packages	1
2	Read Depth Analysis	2
3	Singleton Removal	3
4	Observation of Ubiqitous species	5
5	Alpha Diversity Analysis	8
6	Heatmap of Relative Abundance of Fungal Detections	9
7	Specific Phylogenetic Subsets	11

1 Loading Packages

A comprehensive list of packages required to perform metagenomics analysis for ION torrent data

```
library("knitr")
library("qiime2R") # devtools::install_qithub("jbisanz/qiime2R")
library("phyloseq")
library("readxl")
library("tibble")
library("vegan")
library("DESeq2") #BiocManager::install("DESeq2")
library("speedyseq") # remotes::install github("mikemc/speedyseq")
library("ape")
library("ggstar")
library("forcats")
library("patchwork")
library("ggpubr")
library("plotROC")
library("viridis")
library("cowplot")
library("ggplot2")
library("microbiome") # BiocManager::install("microbiome")
library("microbiomeutilities")
library("ggtree") # BiocManager::install("ggtree")
```

```
library("ggtreeExtra") #install.packages("ggExtra")
library('MicrobiotaProcess') # BiocManager::install("MicrobiotaProcess")
#library("tidytree")
library("file2meco")
library("microeco")

#install.packages('tinytex')
#tinytex::install_tinytex() # install TinyTeX
```

2 Read Depth Analysis

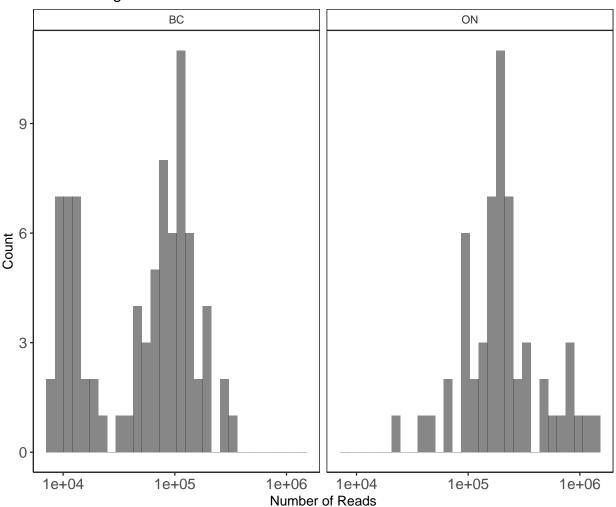
Image is loaded from Ch1_Setup.md , thus we have to remove a number of PS objects and retain only applicable PS Objects for this analysis

Observe read depth to make informed decisions about normalization or standardization of data

```
#load data from previous Ch1 Setup
load("D:\\Grad_School\\R_Projects\\BeeProject_Metagenomics\\Images\\Setup.RData")
#Remove unnessecary PSObjects
rm(PS_16S_BC_ON,PS_16S_BC_ON_BB,PS_16S_BC_ON_SF,PS_16S_Global)
#visualize the objects
PS_ITS1R_BC_ON_SF
## phyloseq-class experiment-level object
## otu_table()
                 OTU Table:
                                   [ 11005 taxa and 138 samples ]:
## sample_data() Sample Data:
                 Sample Data: [ 138 samples by 15 sample variables Taxonomy Table: [ 11005 taxa by 7 taxonomic ranks ]:
                                     [ 138 samples by 15 sample variables ]:
## tax_table()
                  Phylogenetic Tree: [ 11005 tips and 10923 internal nodes ]:
## phy_tree()
## taxa are rows
#Visualize the metadatafile
#head(meta(PS ITS1R BC ON SF))
\#meta(PS_ITS1R_BC_ON_SF)
head((meta(PS_ITS1R_BC_ON_SF))[,c(4,5,6,7)])
```

	Description	RunNumber	${\bf Sample Name}$	Number
A01	ITS2_A_B01	BCC_R1	BCCV1-AP-1B	1
A02	$ITS2_A_B02$	BCC_R1	BCCV1-AP-2B	2
A03	$ITS2_A_B03$	BCC_R1	BCCV1-AP-3B	3
A04	$ITS2_A_B04$	BCC_R1	BCCV2-AP-1B	4
A05	$ITS2_A_B05$	BCC_R1	BCCV2-AP-2B	5
A06	$ITS2_A_B06$	BCC_R1	BCCV2-AP-3B	6

Read Histogram Distribution



3 Singleton Removal

A singleton is defined as a read that occurs uniquely in the dataset. Normally, we remove singletons as they are usually artifacts generated from sequencing error.

Here we identify 4563 singletons and we remove them to create a **Pruned** PS Object

```
#Summarize the PS Object to observe singletons
summarize_phyloseq(PS_ITS1R_BC_ON_SF)
```

```
## [[1]]
## [1] "1] Min. number of reads = 7904"
##
## [[2]]
## [1] "2] Max. number of reads = 1422317"
##
```

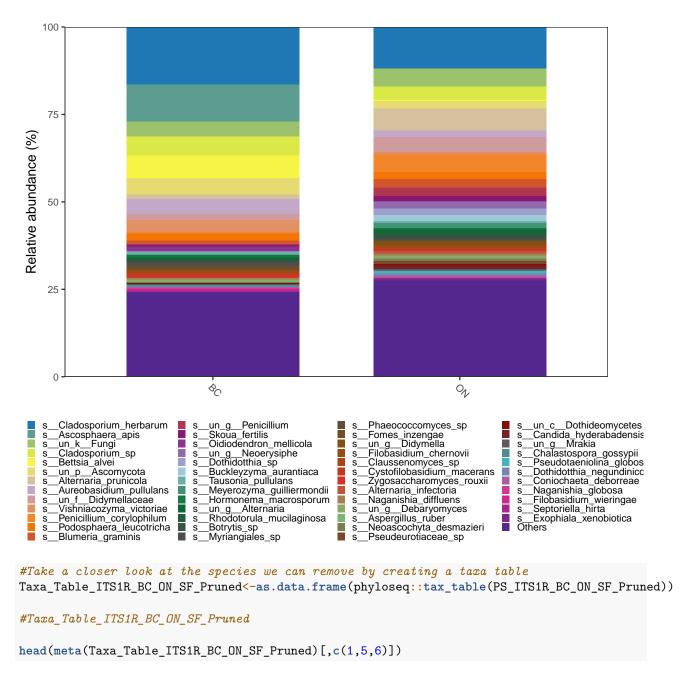
```
## [[3]]
## [1] "3] Total number of reads = 22291019"
## [[4]]
## [1] "4] Average number of reads = 161529.123188406"
## [[5]]
## [1] "5] Median number of reads = 108408.5"
##
## [[6]]
## [1] "7] Sparsity = 0.993258663716756"
## [[7]]
## [1] "6] Any OTU sum to 1 or less? YES"
## [[8]]
## [1] "8] Number of singletons = 4563"
## [[9]]
## [1] "9] Percent of OTUs that are singletons \n
                                                       (i.e. exactly one read detected across all sam
##
## [[10]]
## [1] "10] Number of sample variables are: 15"
## [[11]]
## [1] "barcode.seqeunce"
                                 "linker.primer.seqeunce" "InputFileName"
## [4] "Description"
                                                          "SampleName"
                                 "RunNumber"
## [7] "Number"
                                 "Province"
                                                          "Location"
## [10] "Plant"
                                 "Sample.type"
                                                          "Site"
## [13] "DateCollected"
                                 "Colony"
                                                          "DNANg"
#Number of singletons = 4563
#We need to remove the so we remove any taxa sums that are less than 2 reads, so we use the prune_taxa
PS_ITS1R_BC_ON_SF_Pruned = prune_taxa(taxa_sums(PS_ITS1R_BC_ON_SF) > 02, PS_ITS1R_BC_ON_SF)
PS_ITS1R_BC_ON_SF_Pruned
## phyloseq-class experiment-level object
                            [ 6377 taxa and 138 samples ]:
## otu table()
                OTU Table:
## sample_data() Sample Data:
                                   [ 138 samples by 15 sample variables ]:
                Taxonomy Table:
                                    [ 6377 taxa by 7 taxonomic ranks ]:
## tax_table()
## phy_tree()
                 Phylogenetic Tree: [ 6377 tips and 6338 internal nodes ]:
## taxa are rows
summarize_phyloseq(PS_ITS1R_BC_ON_SF_Pruned)
## [[1]]
## [1] "1] Min. number of reads = 7904"
## [[2]]
## [1] "2] Max. number of reads = 1422313"
## [[3]]
## [1] "3] Total number of reads = 22290889"
```

```
##
## [[4]]
## [1] "4] Average number of reads = 161528.18115942"
## [[5]]
## [1] "5] Median number of reads = 108408.5"
##
## [[6]]
## [1] "7] Sparsity = 0.988440114269351"
##
## [[7]]
## [1] "6] Any OTU sum to 1 or less? NO"
## [[8]]
## [1] "8] Number of singletons = 0"
##
## [[9]]
## [1] "9] Percent of OTUs that are singletons \n
                                                           (i.e. exactly one read detected across all sam
##
## [[10]]
## [1] "10] Number of sample variables are: 15"
## [[11]]
   [1] "barcode.sequence"
                                  "linker.primer.sequence" "InputFileName"
##
                                  "RunNumber"
   [4] "Description"
                                                            "SampleName"
   [7] "Number"
                                  "Province"
                                                            "Location"
                                  "Sample.type"
## [10] "Plant"
                                                            "Site"
## [13] "DateCollected"
                                  "Colony"
                                                            "DNANg"
```

4 Observation of Ubiqitous species

Singletons were successfully removed as singleton count = 0

Now I generated a **relative abundance** profile for all fungal hits in the pruned object. We observe a number of species that are considered **ubiquitous** and not necessary for the goals of the project.



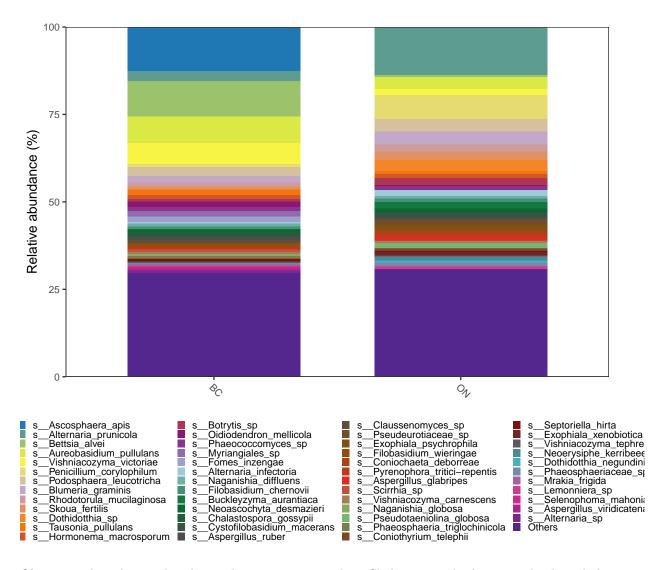
Kingdom	Family	Genus
Fungi	Phaffomycetaceae	Cyberlindnera

Based on the above relative abundance profile, we observe high abundance of $Cladosporium\ herbarum\ and\ through\ taxa_tables$ we identify high levels of Saccharomycetes and N/A hits.

We aim to remove these in the next code chunk We remove them using a subsetting strategy, where we create

subsets of the pruned object and keep removing specific organisms based on phylum.

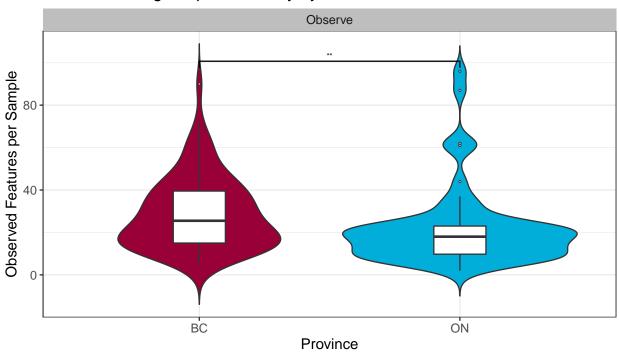
```
# Removal of ubiqitous species and unknown phylums
# We generate a new Filtered object by continiously removing certain species and phylum
PS_ITS1R_BC_ON_SF_Filtered <- subset_taxa(PS_ITS1R_BC_ON_SF_Pruned, Species !="Cladosporium_herbarum")
PS ITS1R BC ON SF Filtered <- subset taxa(PS ITS1R BC ON SF Filtered, Species !="Cladosporium sp")
PS_ITS1R_BC_ON_SF_Filtered <- subset_taxa(PS_ITS1R_BC_ON_SF_Filtered, Class!="Saccharomycetes")
PS_ITS1R_BC_ON_SF_Filtered <- subset_taxa(PS_ITS1R_BC_ON_SF_Filtered, Phylum!="N/A")
PS ITS1R BC ON SF Filtered
## phyloseq-class experiment-level object
## otu table()
                OTU Table:
                                   [ 3916 taxa and 138 samples ]:
## sample data() Sample Data:
                                    [ 138 samples by 15 sample variables ]:
                Taxonomy Table: [ 3916 taxa by 7 taxonomic ranks ]:
## tax_table()
                Phylogenetic Tree: [ 3916 tips and 3896 internal nodes ]:
## phy_tree()
## taxa are rows
#Re-checking the Filtered object to check if filtering via subsetting worked?
classtaxa_ITS1R_BC_ON_SF_Filtered <- get_taxadf(obj=PS_ITS1R_BC_ON_SF_Filtered, taxlevel=7)</pre>
# The 30 most abundant taxonomy will be visualized by default (parameter `topn=30`).
TopTaxa_ITS1R_BC_ON_SF_Filtered<- ggbartax(obj=classtaxa_ITS1R_BC_ON_SF_Filtered,
                   facetNames="Province",
                   plotgroup=TRUE,
                   topn=50) +
  xlab(NULL) +
  ylab("Relative abundance (%)") +
  guides(fill= guide_legend(keywidth = 0.5, keyheight = 0.5, ncol=4))
TopTaxa_ITS1R_BC_ON_SF_Filtered
```



Observing the relative abundance charts we can see that *Cladosporium herbarum* and other phylum not required are removed successfully.

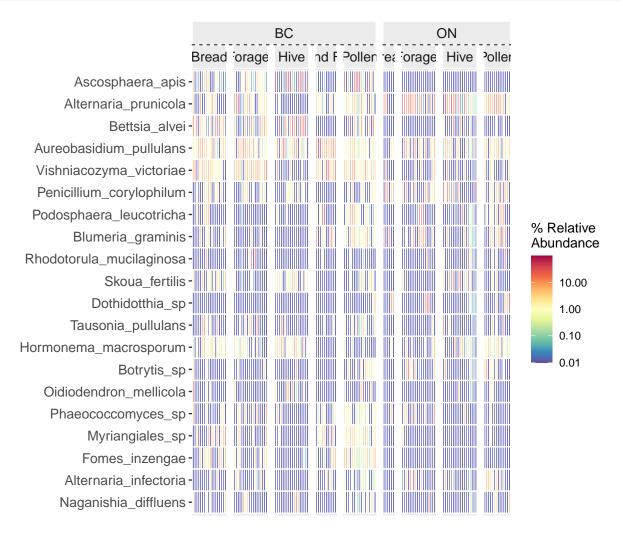
5 Alpha Diversity Analysis

Observed Fungal Alpha Diversity by Provinces



6 Heatmap of Relative Abundance of Fungal Detections

We use functions to convert the PS object to a Meco object and represent this object using *microeco* package Heatmaps offer a visual representation of taxonomic detections based on experiment parameters



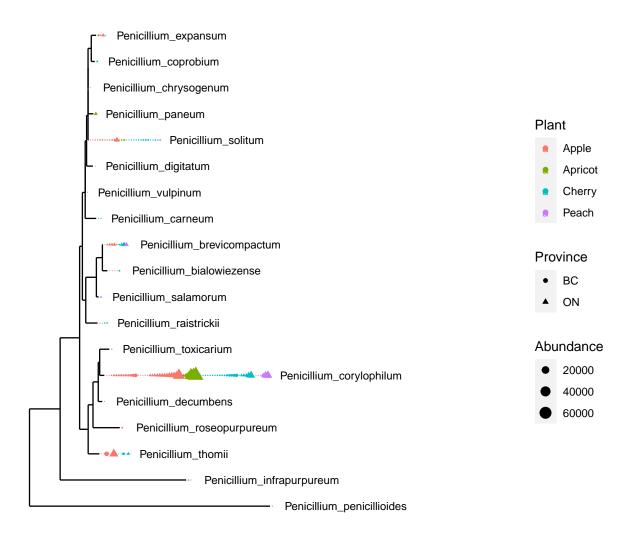
Based on the above heatmap we can identify some pathogenic species to \mathbf{bees} : - $Ascospahera\ apis$ - $Bettsia\ alvei$

Based on the above heatmap we can identify some pathogenic species and genus to plants: - Alternaria

7 Specific Phylogenetic Subsets

Based on knowledge of plant pathogens , we can create sub objects of the filtered PS object to consist of a specific genus and attempt to target certain pathogenic species

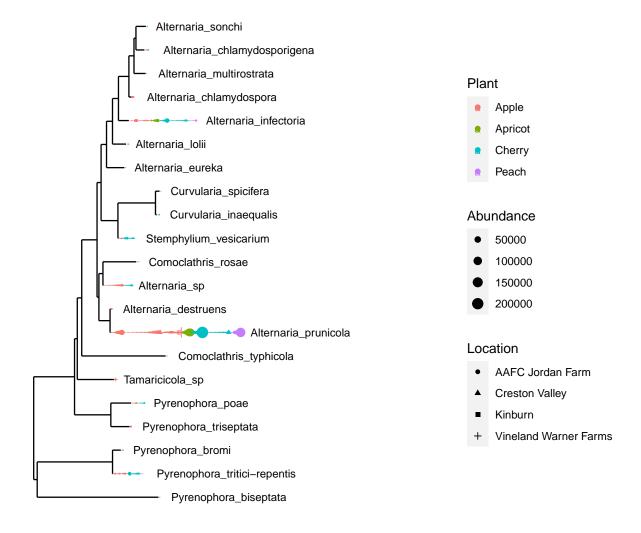
(1) Penicillium Expansum - blue mold of apples



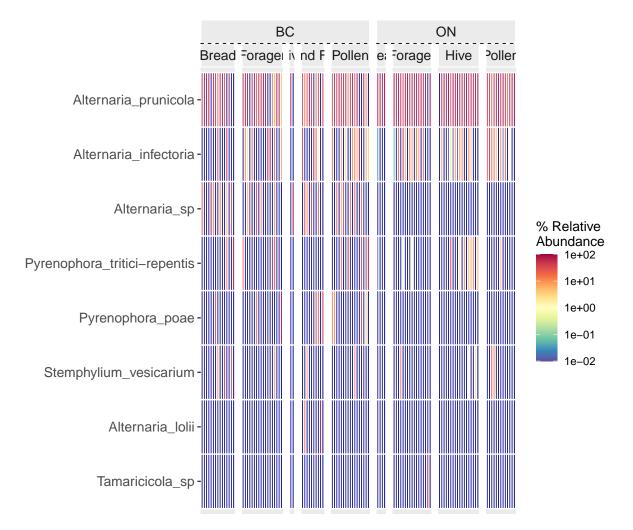
(2) Alternaria prunicola - Cherry leaf spot

For a more complete phylogenetic tree we subset based on the family Pleosporaceae

```
label.tips="Species",
  text.size=3,
  base.spacing=0.01,
  justify="jagged",
  shape="Location",
  size="abundance",
  plot.margin =0.9)+
scale_size_continuous(range = c(0.0001, 4))
```



microtable-class object:

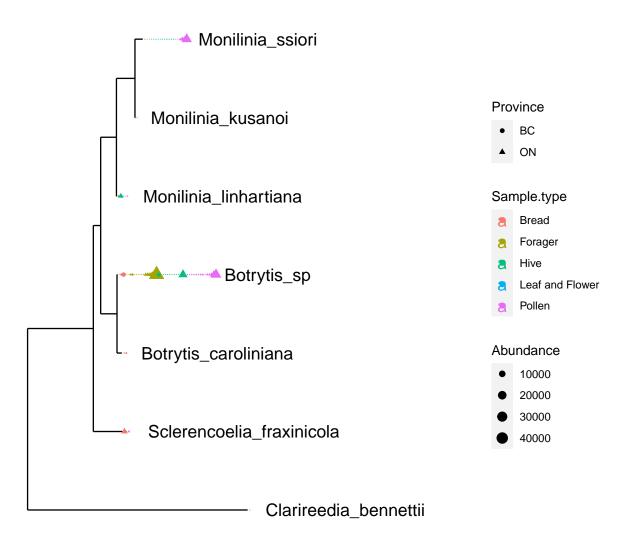


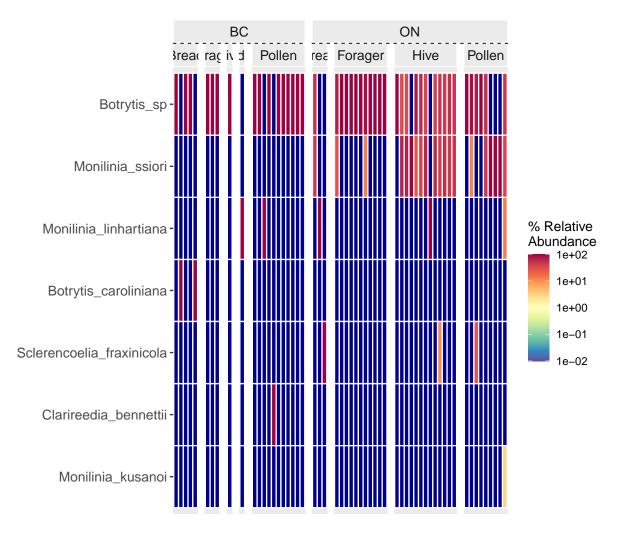
NOTE : The heatmaps may look difficult to visualize here hwoever when run in Rstudio environment, it is possible to export to PNG and have a clearer output

(2) Monilinia_sp - Brown Rot

For a more complete phylogenetic tree we subset based on the family Sclerotiniaceae

```
method = "sampledodge",
    ladderize="left",
    nodelabf=nodeplotblank,
    color="Sample.type",
    label.tips="Species",
    text.size=5,
    base.spacing=0.01,
    justify="jagged",
    shape="Province",
    size="abundance",
    plot.margin =0.9)+
scale_size_continuous(range = c(0.0001, 4))
```



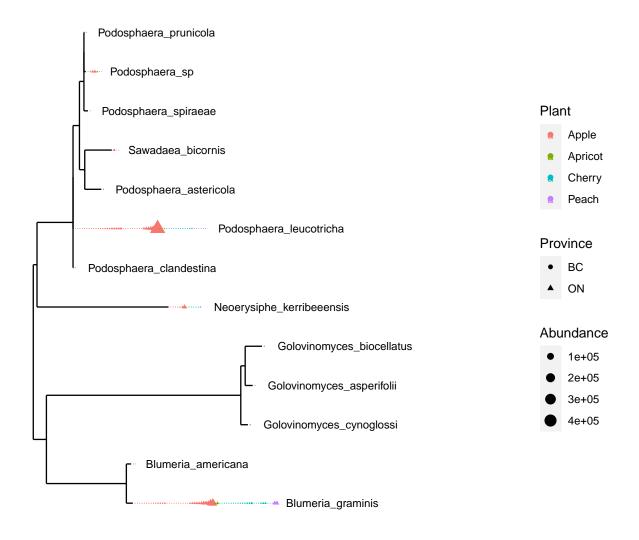


(3) Podosphaera Leucotricha - Powdery Mildew

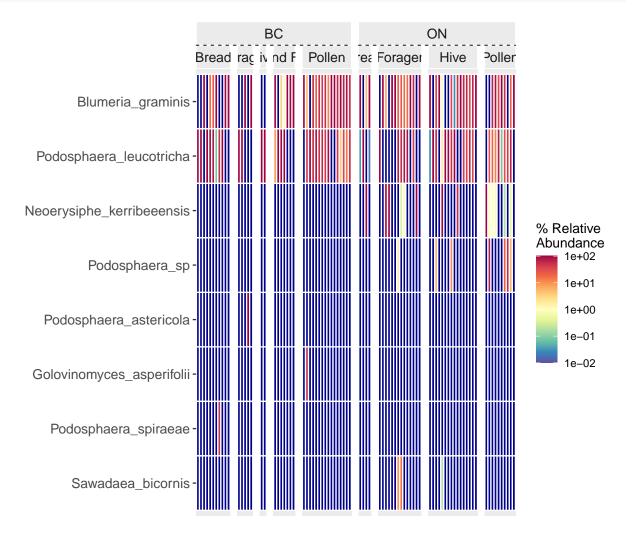
For a more complete phylogenetic tree we subset based on the family Erysiphaceae

```
PS_ITS2_BC_ON_SF_Erysiphaceae <- subset_taxa(PS_ITS1R_BC_ON_SF_Filtered, Family == "Erysiphaceae")
plot_tree(tax_glom(PS_ITS2_BC_ON_SF_Erysiphaceae,
```

```
taxrank="Species"),
method = "sampledodge",
ladderize="left",
nodelabf=nodeplotblank,
color="Plant",
label.tips="Species",
text.size=3,
base.spacing=0.01,
justify="jagged",
shape="Province",
size="abundance",
plot.margin =0.9)+
scale_size_continuous(range = c(0.0001, 4))
```

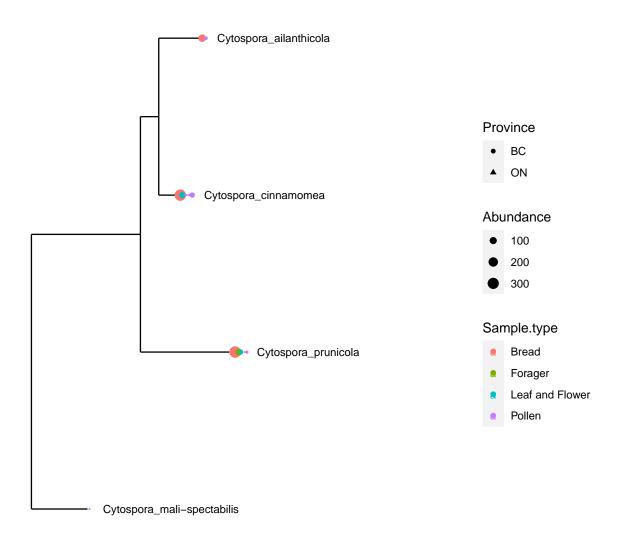


```
##HEATMAP
meco_ITS2_BC_ON_SF_Erysiphaceae<- phyloseq2meco(PS_ITS2_BC_ON_SF_Erysiphaceae)
b1_ITS2_BC_ON_SF_Erysiphaceae <- trans_abund$new(dataset = meco_ITS2_BC_ON_SF_Erysiphaceae,</pre>
```

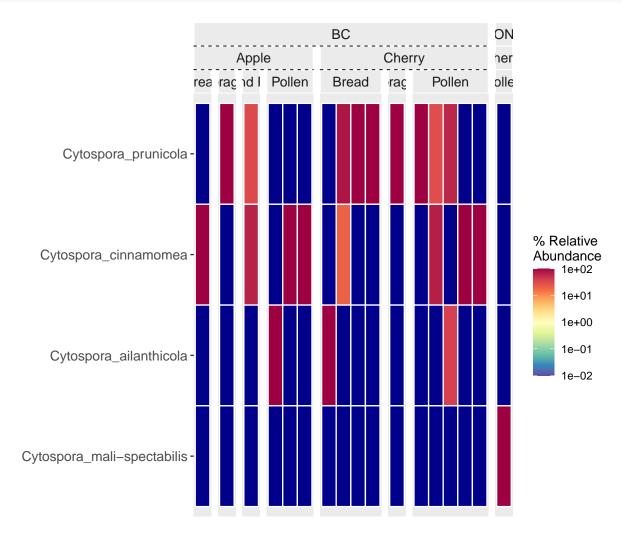


(4) $Cytospora\ Cinnamomea$ - Cherry Canker Disease

For a more complete phylogenetic tree we subset based on the genus Cytospora



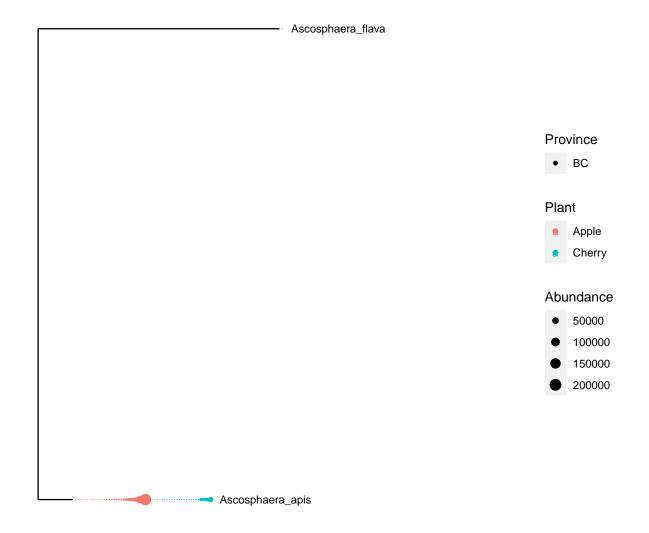
```
##HEATMAP
meco_ITS1R_Cytospora<- phyloseq2meco(PS_ITS1R_Cytospora)</pre>
```

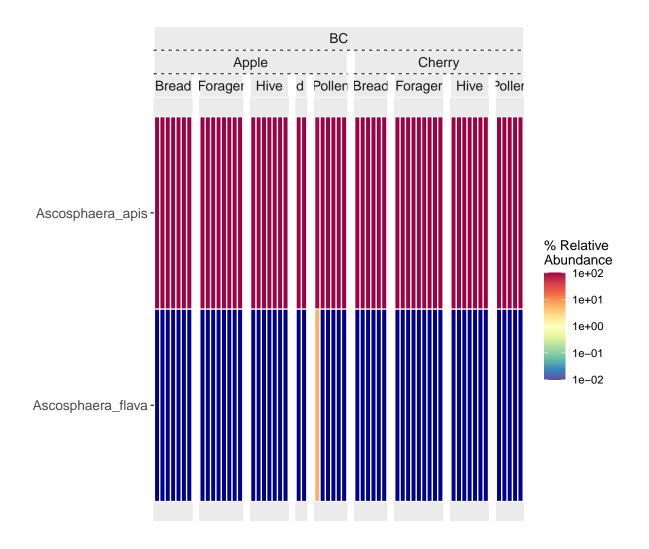


(5) Ascosphaera Apis - Chalkbrood Disease

For a more complete phylogenetic tree we subset based on the family Ascosphaeraceae

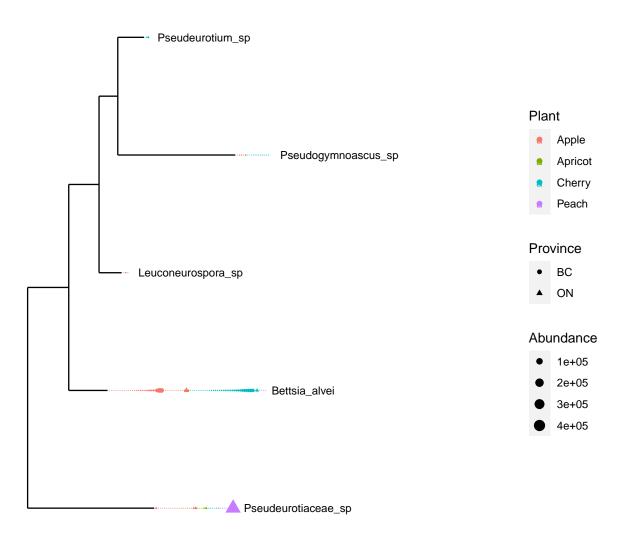
```
##BEE PATHOGENS
PS_ITS2_BC_ON_SF_Ascosphaeraceae <- subset_taxa(PS_ITS1R_BC_ON_SF_Filtered, Family == "Ascosphaeraceae"
plot_tree(tax_glom(PS_ITS2_BC_ON_SF_Ascosphaeraceae,
                   taxrank="Species"),
          method = "sampledodge",
          ladderize="left",
          nodelabf=nodeplotblank,
          color="Plant",
          label.tips="Species",
          text.size=3,
          base.spacing=0.01,
          justify="jagged",
          shape="Province",
          size="abundance",
          plot.margin =0.9)+
  scale_size_continuous(range = c(0.0001, 4))
```





(6) Bettsia alvei - Pollen Mould Disease

For a more complete phylogenetic tree we subset based on the family Pseudeurotiaceae



withmargin = TRUE)

