

phyloseq Testing

Ella Crotty

Contents

Documentation Link	2
Make a phyloseq object	2
Subset by taxa	2
Data transformation functions	2
Plotting functions	2
Variance analysis	2

```
library(igraph)
```

```
## Warning: package 'igraph' was built under R version 4.3.1
##
## Attaching package: 'igraph'
##
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
##
## The following object is masked from 'package:base':
##
##     union
```

```
library(phyloseq)
```

```
# Problem
# Google says it depends on "igraph"
# Actual problem was that it needed BiocManager
```

```
library(vegan)
```

```
## Warning: package 'vegan' was built under R version 4.3.3
## Loading required package: permute
##
## Attaching package: 'permute'
##
## The following object is masked from 'package:igraph':
##
##     permute
## Loading required package: lattice
## Warning: package 'lattice' was built under R version 4.3.1
```

```
## This is vegan 2.6-6.1
##
## Attaching package: 'vegan'
## The following object is masked from 'package:igraph':
##
##      diversity
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.3.1
```

Documentation Link

Make a phyloseq object

```
ps <- phyloseq(otu_table(ASV table, taxa_are_rows=FALSE),
               sample_data(Sample Data),
               tax_table(as.matrix(Taxonomy Table$Taxon)))
```

Subset by taxa

```
ps.fungi = subset_taxa(ps, Kingdom==k__Fungi")
```

Data transformation functions

Phyloseq has a variety of data cleaning and transformation functions.

Data exploration functions:

- `ntaxa()`
- `nsamples()`
- `rank_names()` lists the taxonomic levels represented
- `sample_variables()`
- `otu_table()`
- `tax_table()`

Plotting functions

- `plot_bar(ps, x = "Sample", y = "Abundance", fill = pick a taxon level, ex. "Phylum")`
← This is a ggplot object that can be added to
- `plot_ordination(agglomerated sp, ordered sp, type = "samples", color = "horizon")`
quick plots NMDS for clustering
 - This can get more complicated and you can draw `stat_ellipse()` for clustering

The function `ordinate()` can be used to order by dissimilarity

Variance analysis

```
metadata <- as(sample_data(ps.fungi.nosd.log.sp), "data.frame") # Export data
```

```
adonis(phyloseq::distance(physeq = ps.fungi.nosd.log.sp, method="bray") ~ horizon * forest, strata = metadata,
       data = metadata,
```

```
permutations = 99999)
```

- `betadisper()` tests for dispersion homogeneity using different methods, including `method = "bray"` for Bray-Curtis dissimilarity
- `ordinate()` can be used with Redundancy Analysis (RDA). `ord.rda <- ordinate(ps.fungi.nosd.log.sp, formula = , method = , distance =)`, then `RsquareAdj(ord.rda)` and `anova.cca(ord.rda)`
- `vif.cca(ord.rda)` checks for Variance Inflation Factors, $VIF < 10$ should be fine
 - `plot_ordination()` can also plot RDA
 - `ord.rda` has eigenvectors in it