

WORKSHOP FRIBOURG 2025

TARGETED OR AMPLICON DOWNSTREAM ANALYSIS



Laurent Falquet

Swiss Institute of
Bioinformatics

Possible downstream analyses with microbiome data

Usually one begins with some previsualization and preliminary analysis like absolute, relative, or rank abundance representations and QC like rarefaction curves.

Then one can perform community composition (α -, β - and γ -diversity) or multivariate analysis (ordination) and other Machine Learning techniques.

Other possible analysis are differentially abundant OTUs/ASVs or Linear discriminant Effect Size analysis.

Various tools and platforms available

With R

main packages: **phyloseq, microbiome, microbiomeSeq, dada2, yingtools2, vegan, biomformat**

Standalone

MEGAN6, micca, Mothur, QIIME2, PICRUSt2, Phinch, MetaPhiAn2

Web based

<https://www.microbiomeanalyst.ca/>

<https://www.ebi.ac.uk/metagenomics/>

<https://vamps2.mbl.edu/>

<https://www.mg-rast.org/> (not maintained anymore)



Output from Mothur, QIIME or other methods

Usually you get 2 main tables: the taxonomy and the OTU tables

ws.an.cons.taxonomy

```
OTU      Size     Taxonomy
Otu00001  530053  Bacteria(100);Proteobacteria(100);Gammaproteobacteria(100);Pseudomonadales(100);Pseudomonadaceae(100);Pseudomonas(100);
Otu00002  298855  Bacteria(100);Proteobacteria(100);Betaproteobacteria(100);Methylophilales(100);Methylophilaceae(100);Methylphilus(90);
Otu00003  546152  Bacteria(100);Proteobacteria(100);Alphaproteobacteria(100);Rhizobiales(100);Rhizobiaceae(100);Rhizobium(100);
Otu00004  269985  Bacteria(100);Proteobacteria(100);Gammaproteobacteria(100);Pseudomonadales(100);Pseudomonadaceae(100);Pseudomonas(99);
Otu00005  206811  Bacteria(100);Proteobacteria(100);Betaproteobacteria(100);Burkholderiales(100);Burkholderiaceae(100);Burkholderia(100);
Otu00006  100666  Bacteria(100);Proteobacteria(100);Betaproteobacteria(100);Burkholderiales(100);Comamonadaceae(100);Variovorax(96);
Otu00007  54191   Bacteria(100);Proteobacteria(100);Gammaproteobacteria(100);Xanthomonadales(100);Xanthomonadaceae(100);Stenotrophomonas
...

```

ws.an.shared

```
label Group    numOtu      Otu00001    Otu00002    Otu00003    Otu00004    Otu00005    Otu00006    Otu00007    Otu00008 ...
0.03 bdg-ws-1a 11867      17387      27426      11854      207       53        5808      5964      2006 ...
0.03 bdg-ws-1b ...
...
```

And other tables like metadata of the samples.

To export results in other tools, convert to BIOM:
`make.biom(shared=ws.an.shared, constaxonomy=ws.an.cons.taxonomy)`

The BIOM exists in 2 main versions:
v1 (JSON) and v2 (HDF5).

Some software accept only the v1, others accept only the v2, others accept every version of BIOM.

Software using BIOM format:

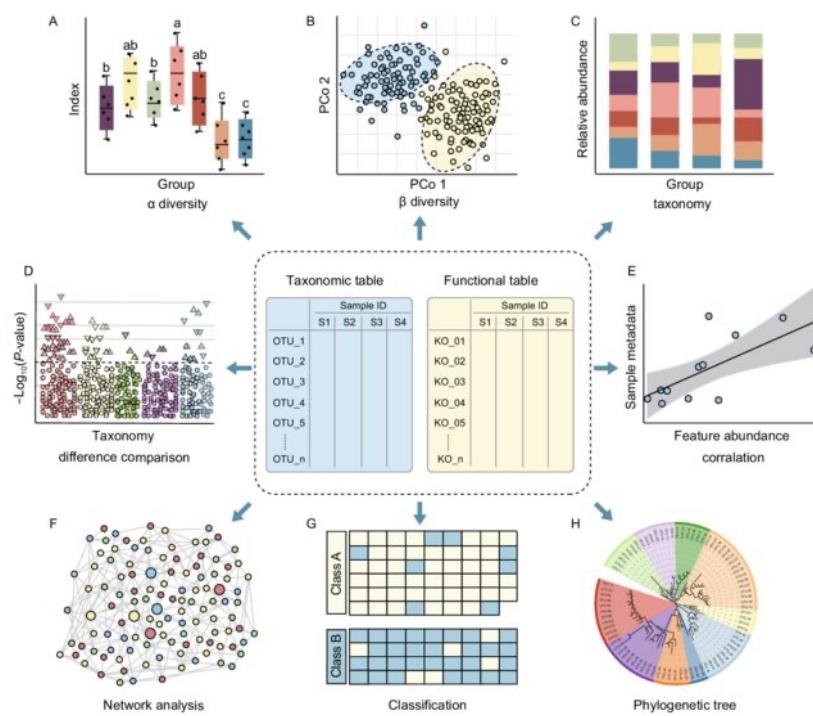
QIIME, MG-RAST, PICRUSt, Mothur, phyloseq, MEGAN, VAMPS, metagenomeSeq, Phinch, RDP Classifier, USEARCH, PhyloToAST, EBI Metagenomics, GCModeller, MetaPhiAn2, MetagenomeAnalyst

Downstream analysis

Once you have the OTUs/ASVs taxonomic tables or BIOM files, you will probably like to obtain some downstream analysis like:

- Rarefaction curves
- Relative and Absolute abundance
- Diversity indices: Alpha, beta, gamma
- Multivariate analysis
- Differentially abundant OTUs/ASVs (DESeq2, EdgeR)
- LefSe (Linear discriminant Effect Size analysis)
- Build trees
- Network analysis
- Machine learning
 - Classification (SVM, RF, NN, etc.)
 - Regression (GLM, ANOVA, ANOSIM etc.)

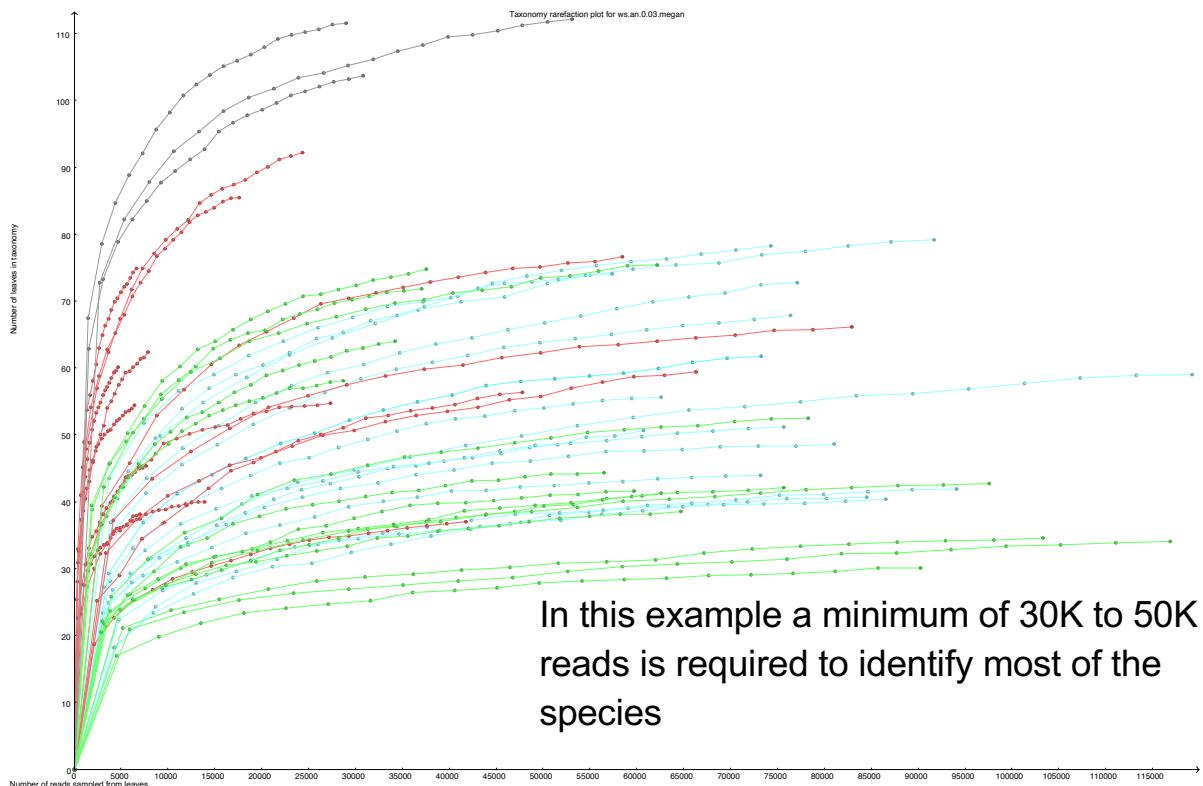
Diverse representations of the data



UNIVERSITÉ DE FRIBOURG / UNIVERSITÄT FREIBURG | Biochemistry/Bioinformatics Unit | Falquet Laurent



Important to check rarefaction curves (e.g., with MEGAN)

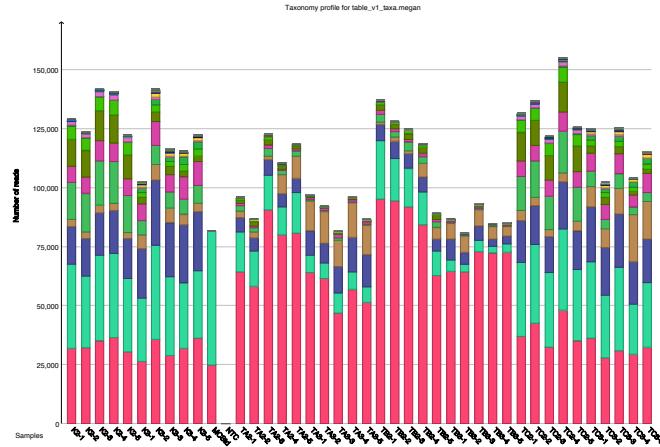


UNIVERSITÉ DE FRIBOURG / UNIVERSITÄT FREIBURG | Biochemistry/Bioinformatics Unit | Falquet Laurent

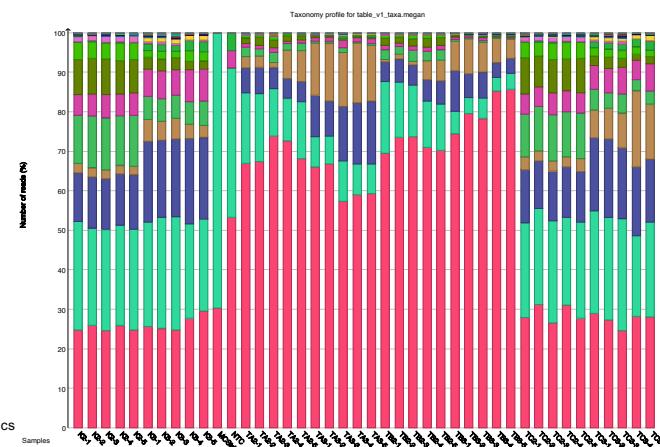


Relative vs Absolute abundance

Absolute abundance (counts)

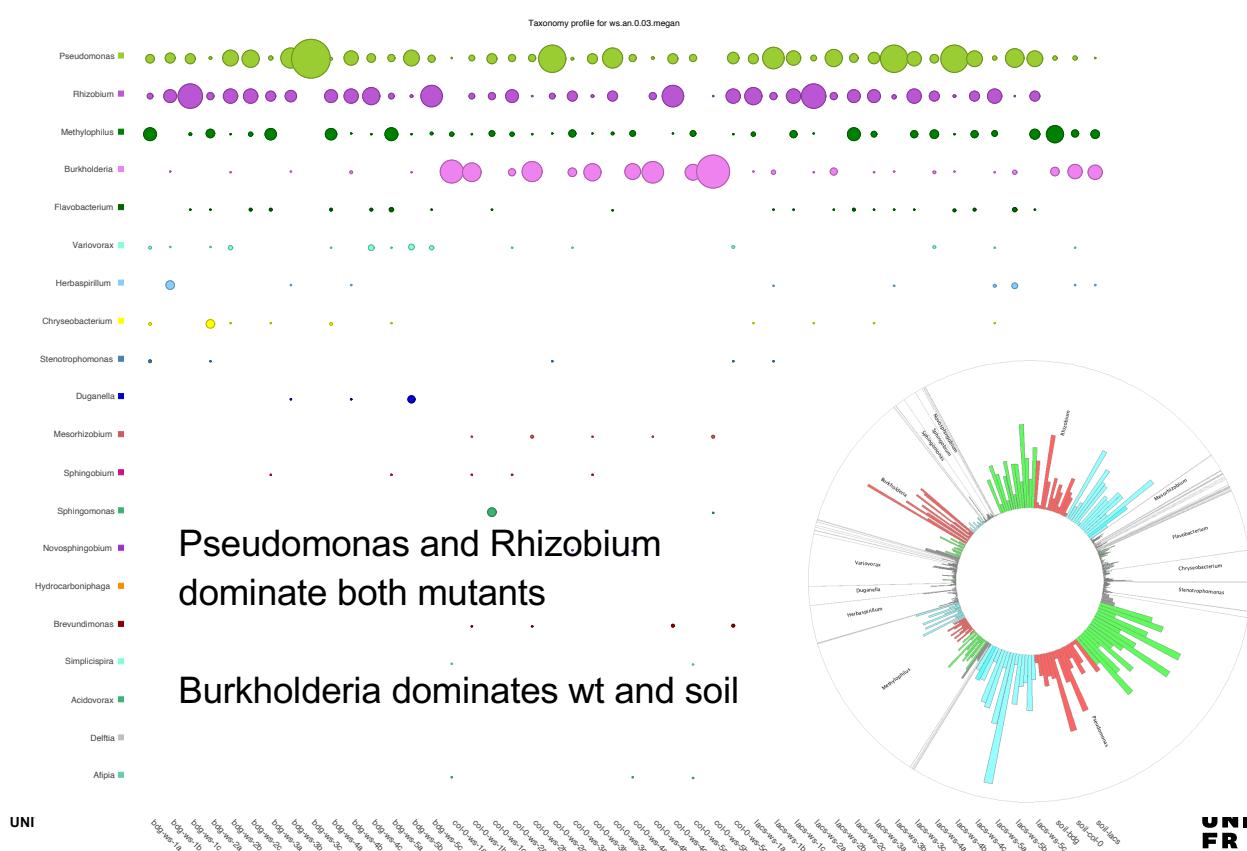


Relative abundance (percent)



UNIVERSITÉ DE FRIBOURG / UNIVERSITÄT FREIBURG | Biochemistry/Bioinformatics

MEGAN example: abundance plot & taxonomy chart



Statistical Analysis - Diversity

Contributed diversity

alpha

diversity inside an area or ecosystem (species richness)

beta

diversity between ecosystems

gamma

overall diversity of all ecosystems in a particular area

Diversity can be measured with different indexes, e.g.:

Shannon entropy

Inverse Simpson

Chao1

Count of species (but depends on sampling depth, to be checked using rarefaction curves)

...

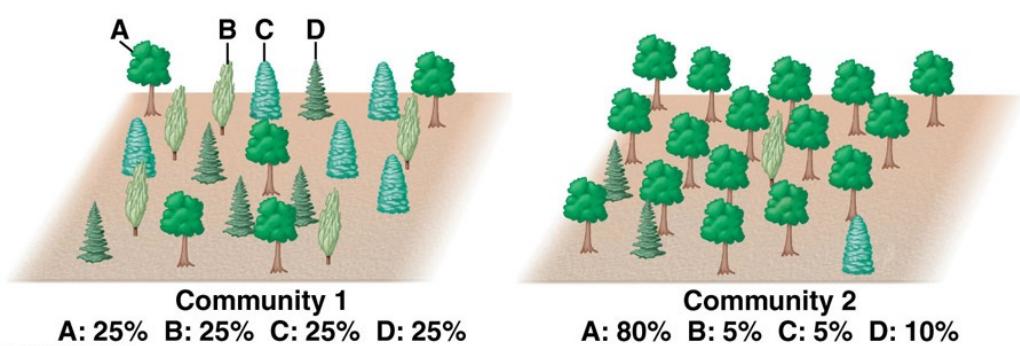
UNIVERSITÉ DE FRIBOURG / UNIVERSITÄT FREIBURG | Biochemistry/Bioinformatics Unit | Falquet Laurent



Richness vs diversity

Richness: Number of observed species (OTUs/ASVs)

Diversity: Number of observed species (OTUs/ASVs) scaled to evenness

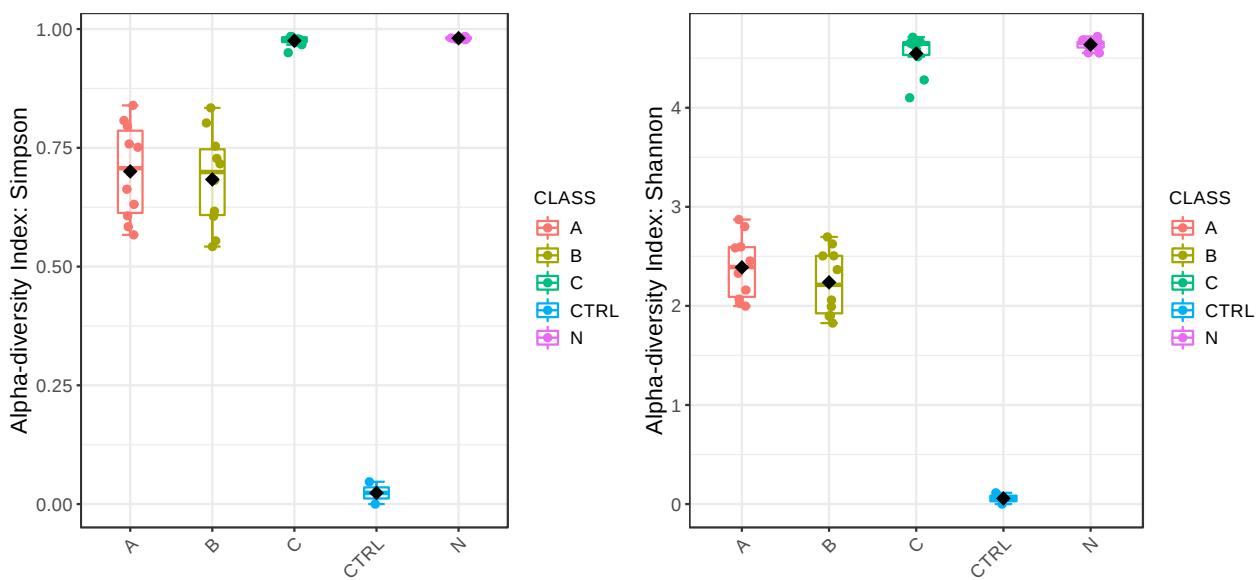


Same richness, different diversity

UNIVERSITÉ DE FRIBOURG / UNIVERSITÄT FREIBURG | Biochemistry/Bioinformatics Unit | Falquet Laurent



Alpha diversity indexes example



Downstream analysis

Once you have the OTUs/ASVs taxonomic tables or BIOM files, you will probably like to obtain some downstream analysis like:

- Rarefaction curves
- Relative and Absolute abundance
- Diversity indices: Alpha, beta, gamma
- Multivariate analysis
- Differentially abundant OTUs/ASVs (DESeq2, EdgeR)
- LefSe (Linear discriminant Effect Size analysis)
- Build trees
- Network analysis
- Machine learning
 - Classification (SVM, RF, NN, etc.)
 - Regression (GLM, ANOVA, ANOSIM etc.)

Ordination methods and distances

Ordination methods are unsupervised exploratory methods based on distance matrices and aim at dimension reduction.

Distance matrices examples:

Unifrac

Weighted unifrac

Bray-Curtis

Jaccard

...

Ordination methods examples:

NMDS (non metric multidimensional scaling)

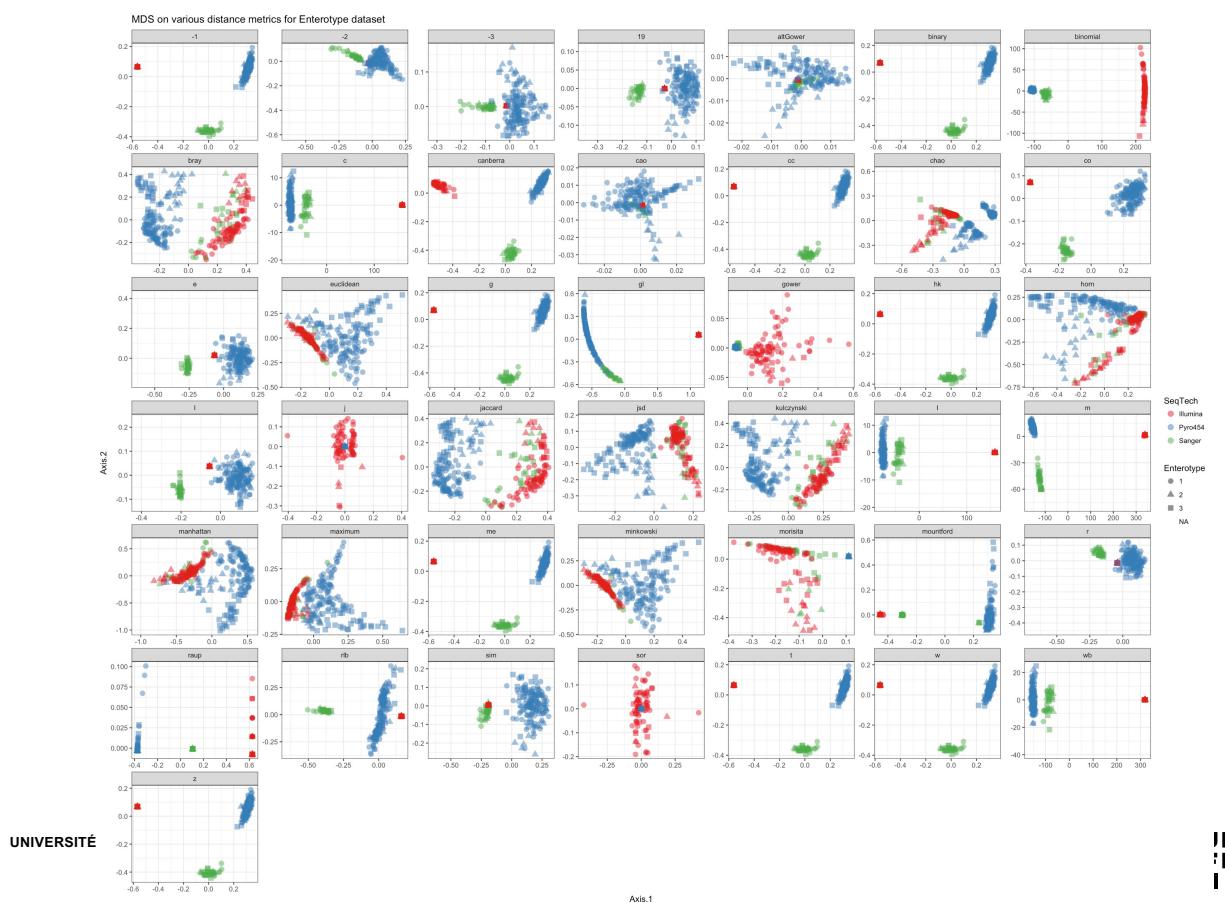
MDS (multidimensional scaling) or PCoA (extension of PCA with any distance metric)

PCA (= PCoA with Euclidian distance metric)

UNIVERSITÉ DE FRIBOURG / UNIVERSITÄT FREIBURG | Biochemistry/Bioinformatics Unit | Falquet Laurent



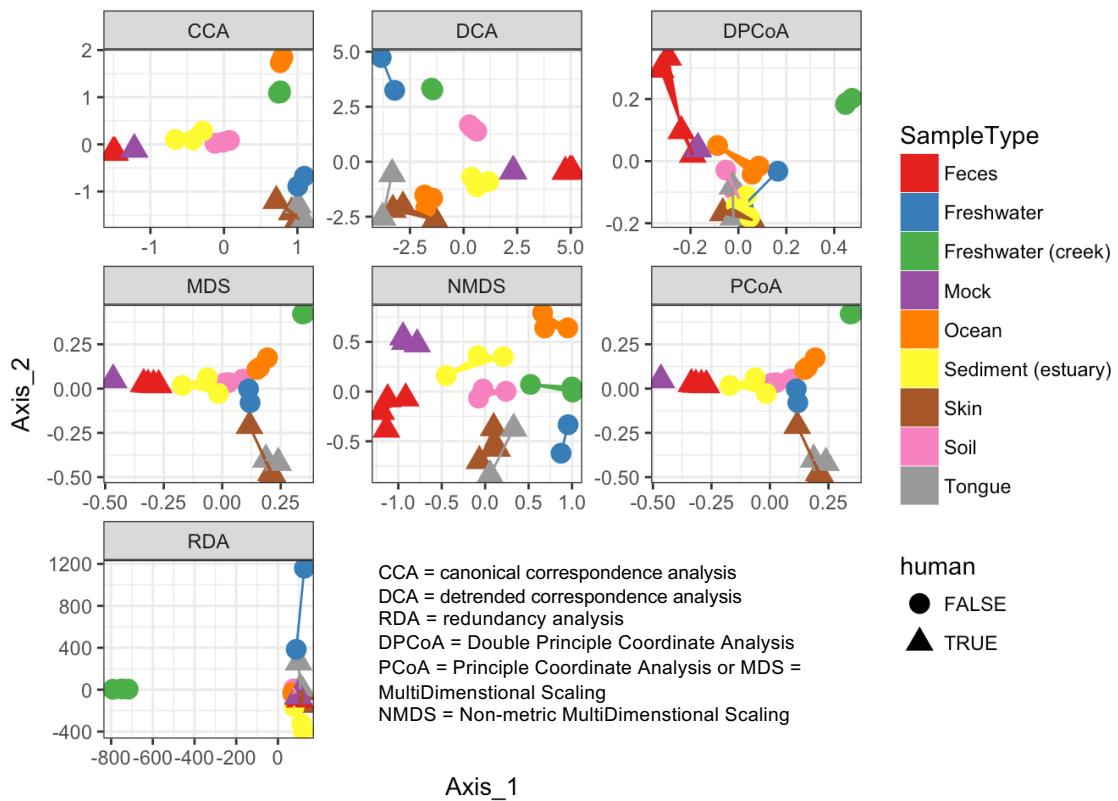
MDS on multiple distance matrices (phyloseq)



UNIVERSITÉ



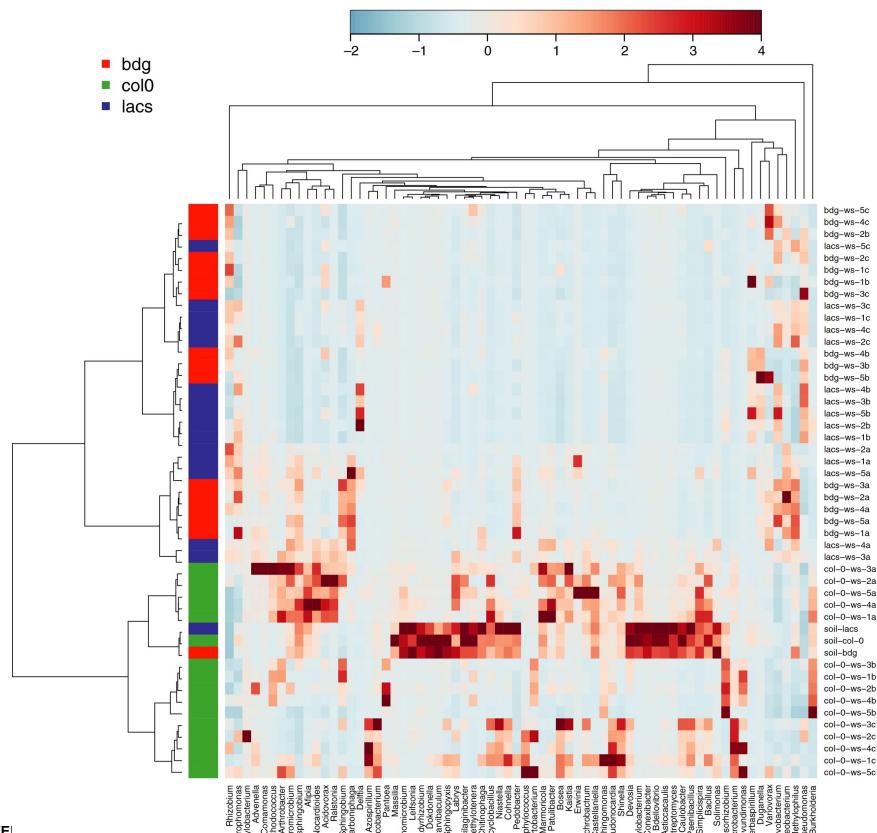
Bray-Curtis distance on multiple ordination methods (phyloseq)



UNIVERSITÉ DE FRIBOURG / UNIVERSITÄT FREIBURG | Biochemistry/Bioinformatics Unit | Falquet Laurent



Heatmap and clustering



UNIVERSITÉ DE FRIBOURG / UNIVE



Downstream analysis

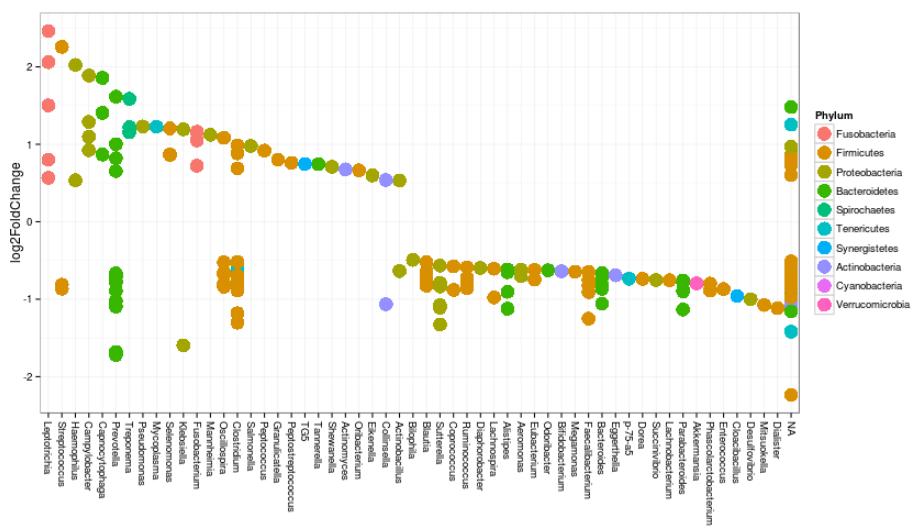
Once you have the OTUs/ASVs taxonomic tables or BIOM files, you will probably like to obtain some downstream analysis like:

- Rarefaction curves
- Relative and Absolute abundance
- Diversity indices: Alpha, beta, gamma
- Multivariate analysis
- Differentially abundant OTUs/ASVs (DESeq2, EdgeR)
- LefSe (Linear discriminant Effect Size analysis)
- Build trees
- Network analysis
- Machine learning
 - Classification (SVM, RF, NN, etc.)
 - Regression (GLM, ANOVA, ANOSIM etc.)

Differential abundance of OTUs /ASVs using phyloseq extensions

You can try to identify the differences in OTUs or ASV between pairs of samples (e.g., cancer vs normal)

Both DESeq2 and EdgeR are packages from the RNAseq analysis domain and can be used with microbiome data.



Differential abundance of OTUs /ASVs using phyloseq extensions

It gives you a table similar to DEG, but for OTUs/ASVs

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	Kingdom
## 304309	1.686	-0.6409	0.1823	-3.516	4.380e-04	0.0018912	Bacteria
## 16076	13.805	-0.8583	0.2094	-4.099	4.144e-05	0.0002678	Bacteria
## 561483	1.866	-0.6382	0.1809	-3.529	4.178e-04	0.0018193	Bacteria
## 177005	8.598	-0.7467	0.1905	-3.920	8.865e-05	0.0004876	Bacteria
## 469778	9.049	-0.6633	0.2209	-3.003	2.673e-03	0.0080976	Bacteria
## 308873	4.165	-0.5360	0.1794	-2.988	2.809e-03	0.0084662	Bacteria
	Phylum	Class	Order				
## 304309	Firmicutes	Clostridia	Clostridiales				
## 16076	Firmicutes	Clostridia	Clostridiales				
## 561483	Actinobacteria	Actinobacteria (class)	Bifidobacteriales				
## 177005	Firmicutes	Clostridia	Clostridiales				
## 469778	Bacteroidetes	Bacteroidia	Bacteroidales				
## 308873	Firmicutes	Clostridia	Clostridiales				
	Family	Genus	Species				
## 304309	Lachnospiraceae	Blautia	Blautia producta				
## 16076	Ruminococcaceae	Ruminococcus	Ruminococcus bromii				
## 561483	Bifidobacteriaceae	Bifidobacterium	Bifidobacterium longum				
## 177005	Lachnospiraceae	Blautia	<NA>				
## 469778	Bacteroidaceae	Bacteroides	Bacteroides coprophilus				
## 308873	Ruminococcaceae	Clostridium	Clostridium orbiscindens				

FR
■

Using Linear discriminant Effect Size analysis (LEfSe) with yingtools2 or Lefser packages

LEfSe (Linear discriminant analysis Effect Size) determines the features (organisms, clades, operational taxonomic units, genes, or functions) most likely to explain differences between classes by coupling standard tests for statistical significance with additional tests encoding biological consistency and effect relevance.

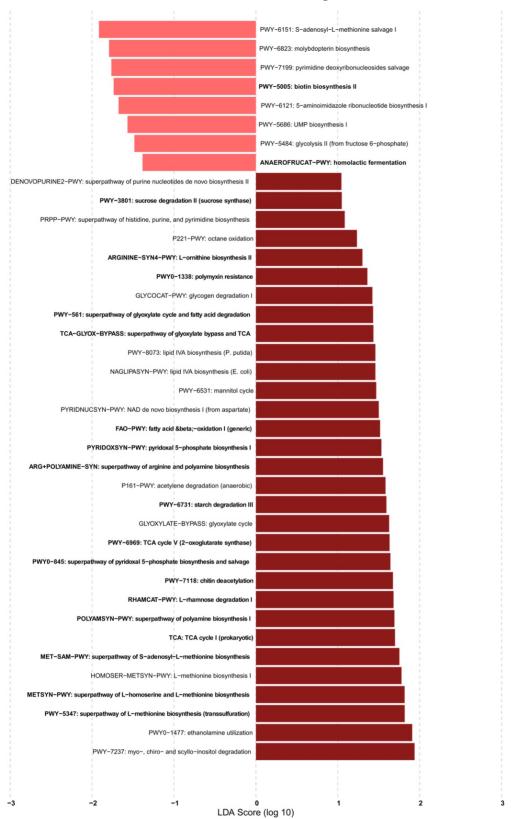


UNI
■

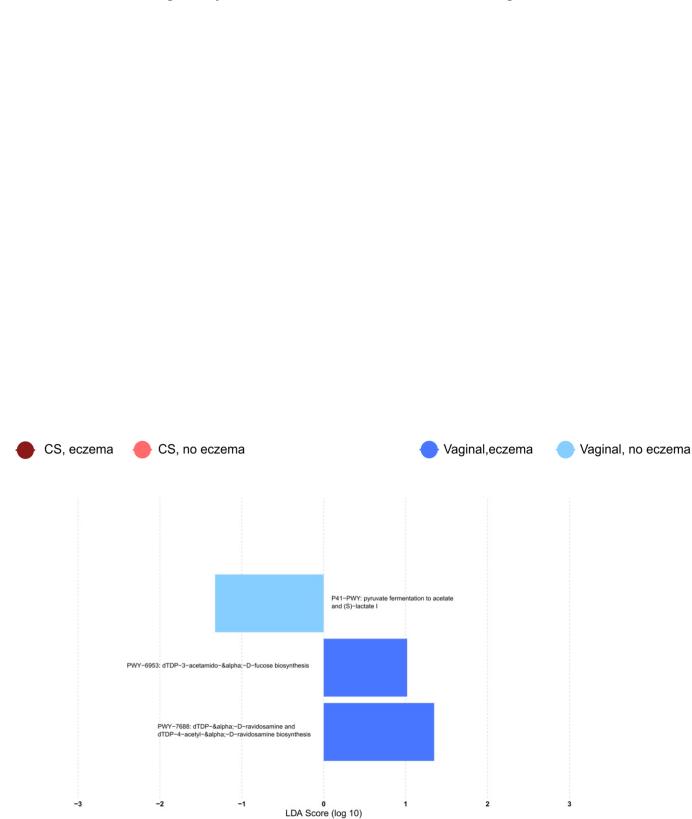
LEfSe comparing pathways

Leo et al, Front. Microbiomes 2:1147082. doi: 10.3389/fmabi.2023.1147082

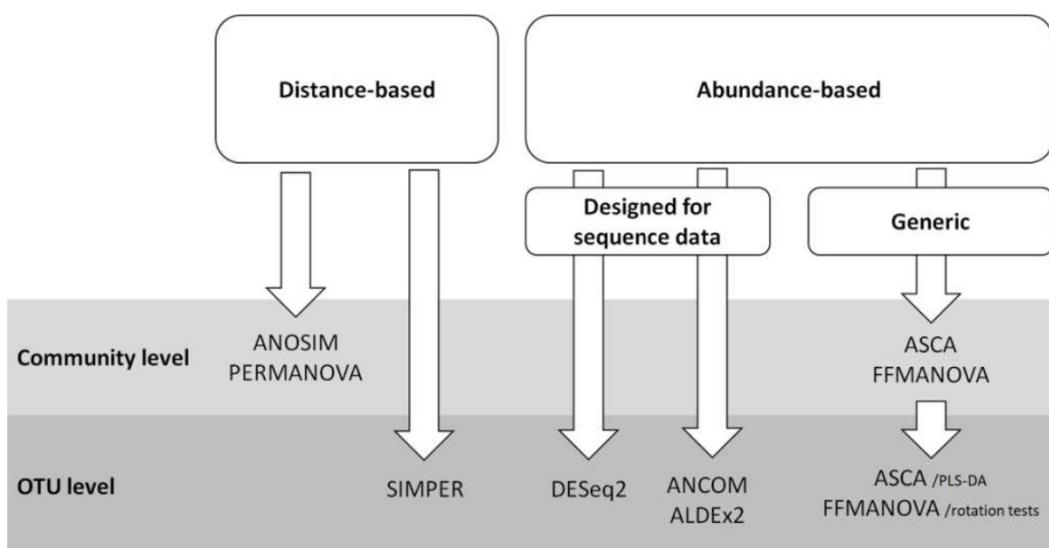
CS-born infants with UK-tool-diagnosed eczema



Vaginally-born infants with UK-tool-diagnosed eczema

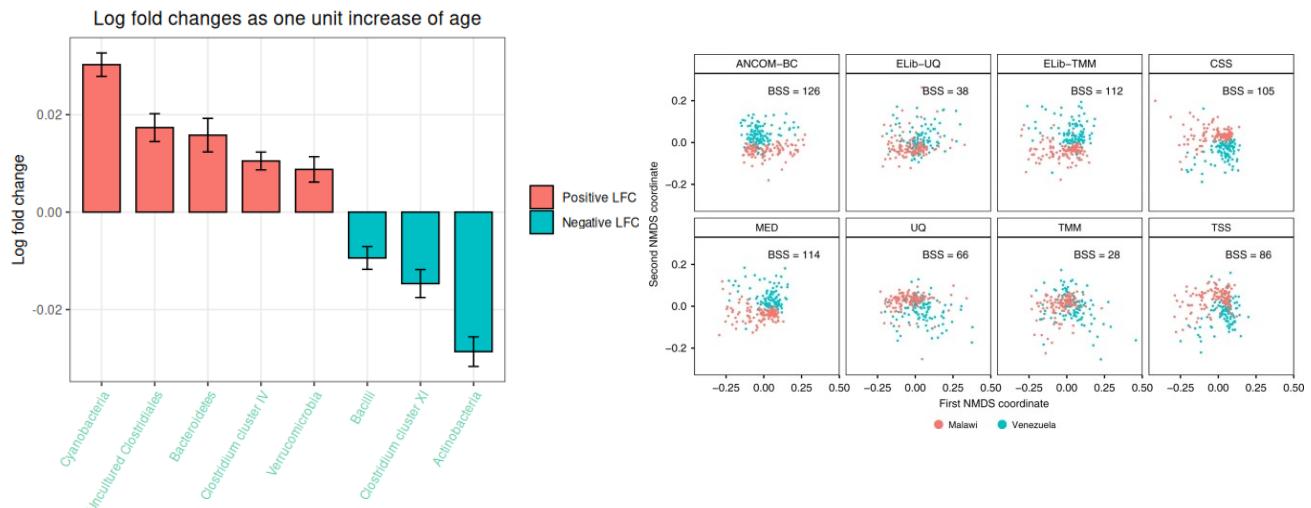


Other multivariate analysis



ANCOM-BC: Microbiome differential abundance and correlation analyses with bias correction

Microbiome data are typically subject to two sources of biases: unequal sampling fractions (sample-specific biases) and differential sequencing efficiencies (taxon-specific biases). ANCOM-BC corrects for those biases.



Lin, H., Peddada, S.D. Analysis of compositions of microbiomes with bias correction. *Nat Commun* 11, 3514 (2020). <https://doi.org/10.1038/s41467-020-17041-7>

UNIVERSITÉ DE FRIBOURG / UNIVERSITÄT FREIBURG | Biochemistry/Bioinformatics Unit | Falquet Laurent

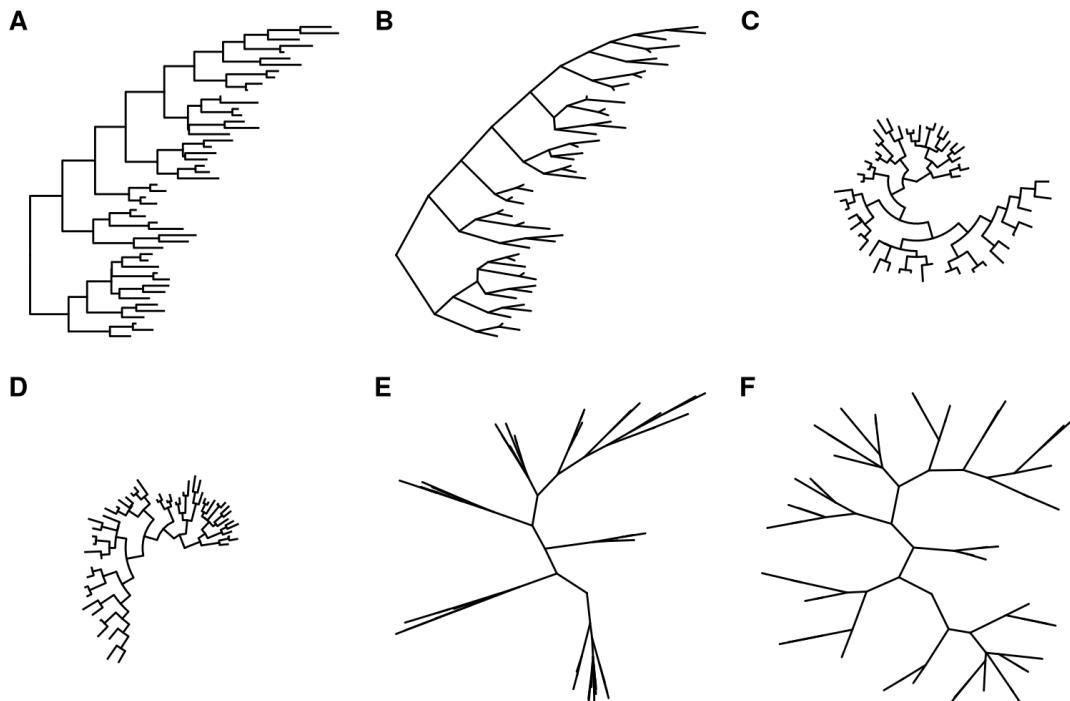


Downstream analysis

Once you have the OTUs/ASVs taxonomic tables or BIOM files, you will probably like to obtain some downstream analysis like:

- Rarefaction curves
- Relative and Absolute abundance
- Diversity indices: Alpha, beta, gamma
- Multivariate analysis
- Differentially abundant OTUs/ASVs (DESeq2, EdgeR)
- LefSe (Linear discriminant Effect Size analysis)
- Build trees
- Network analysis
- Machine learning
 - Classification (SVM, RF, NN, etc.)
 - Regression (GLM, ANOVA, ANOSIM etc.)

Tree representation of the data



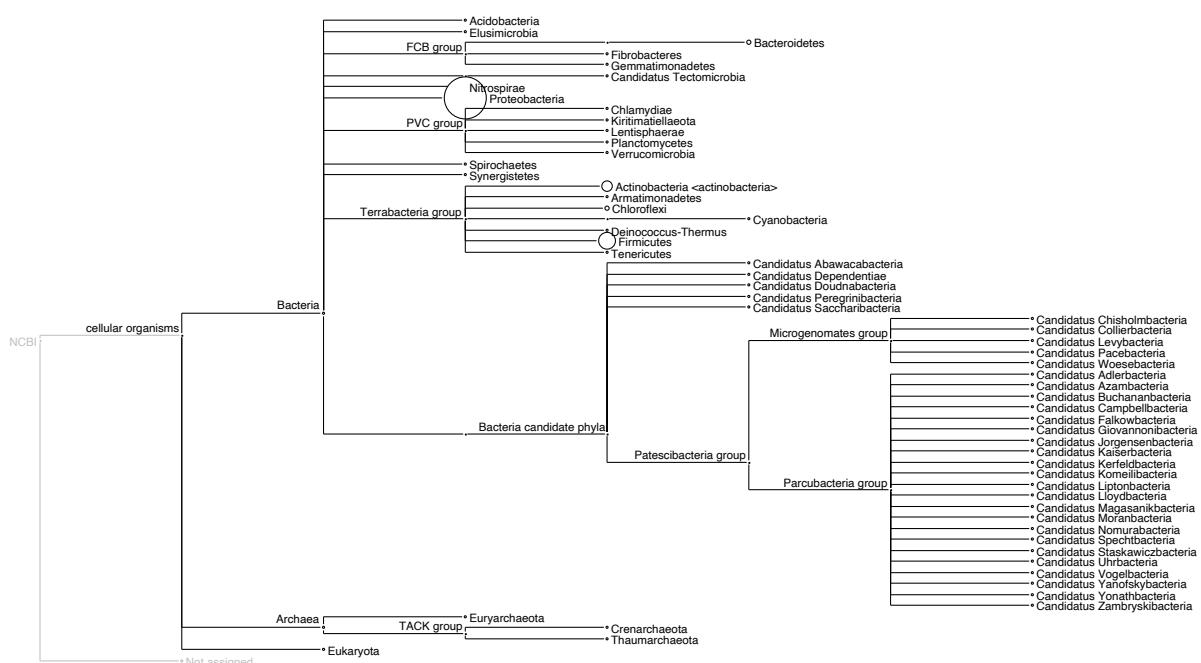
rectangular layout (A), slanted layout (B), circular layout (C) and fan layout (D).
Unrooted: equal-angle method (E) and daylight method (F).

UNIVERSITÉ DE FRIBOURG / UNIVERSITÄT FREIBURG | Biochemistry/Bioinformatics Unit | Falquet Laurent

<https://yulab-smu.github.io/treedata-book/>



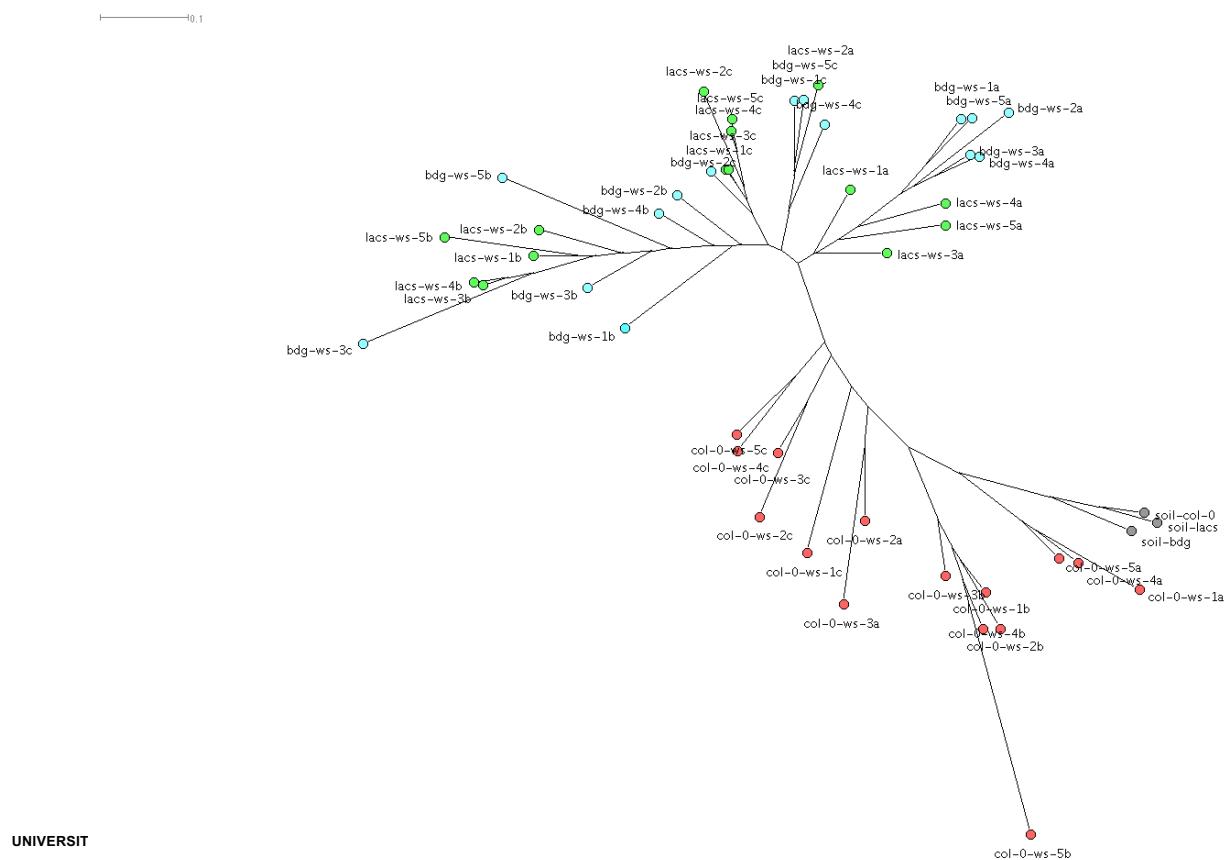
Trees: phylogenetic



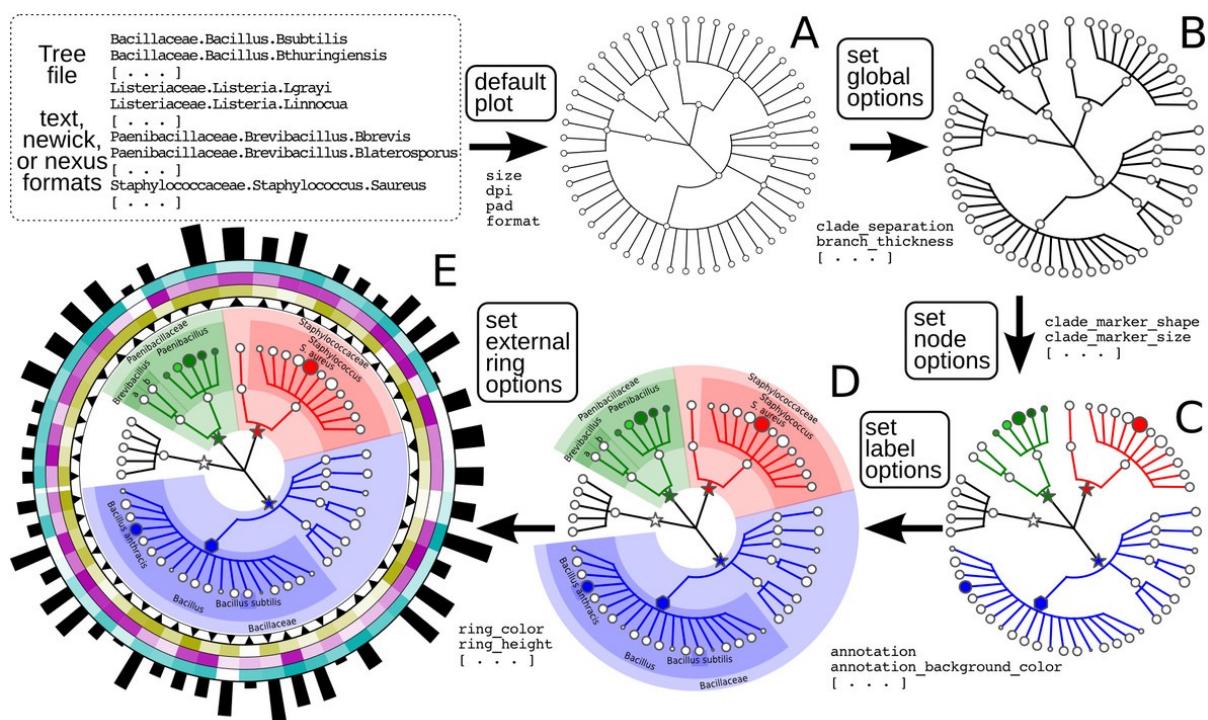
UNIVERSITÉ DE FRIBOURG / UNIVERSITÄT FREIBURG | Biochemistry/Bioinformatics Unit | Falquet Laurent



Trees: Neighbor Joining



Complex representation with GraPhlAn

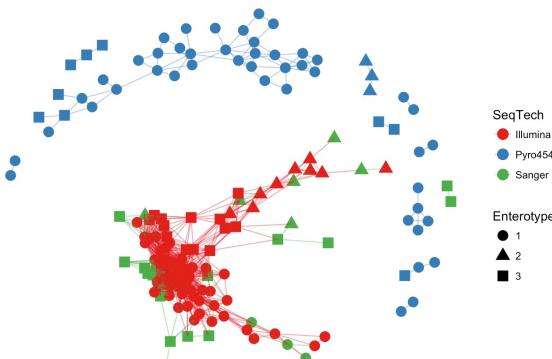


PeerJ. 2015 Jun 18;3:e1029. doi: 10.7717/peerj.1029. eCollection 2015.

UNIVERSITÉ DE FRIBOURG / UNIVERSITÄT FREIBURG | Biochemistry/Bioinformatics Unit | Falquet Laurent

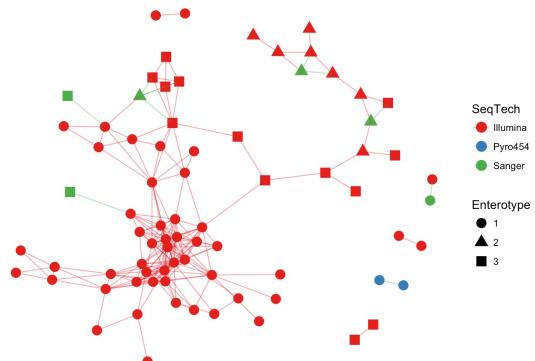
Network analysis with phyloseq

Create an igraph-based network based on the default distance method, "Jaccard", and a maximum distance between connected nodes of 0.3



Maximum-distance
default 0.3

Same with Maximum-distance
lowered to 0.2



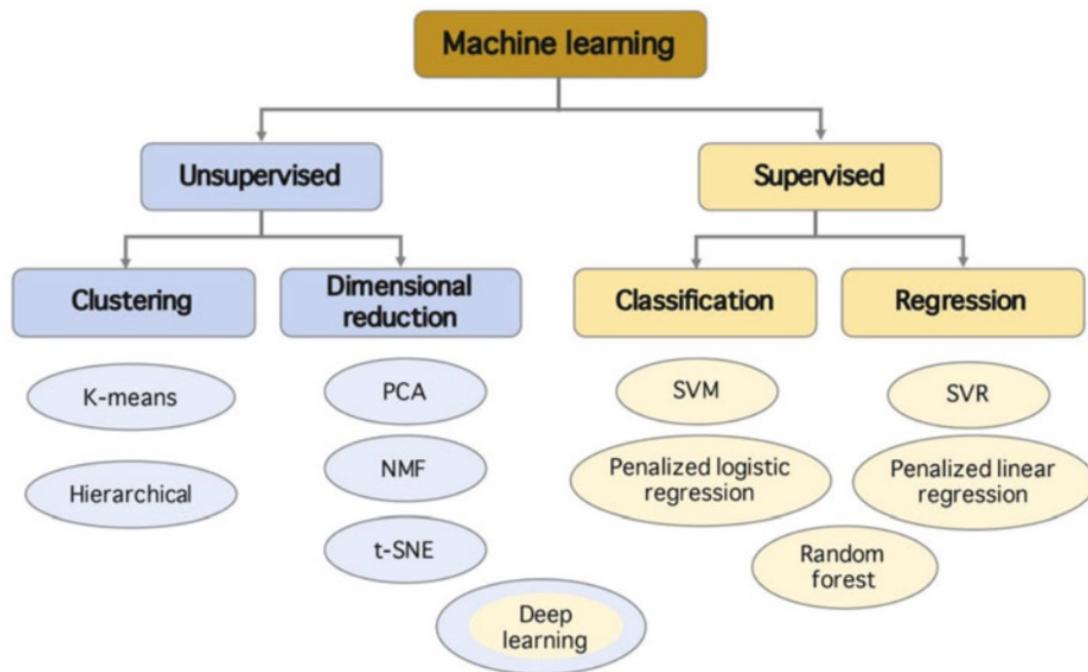
PhyloGeo mapping

If you have metadata containing geographic coordinates, Phylogeo can now be used to map a phyloseq object on a map.



<https://zachcp.github.io/phylogeo/>

Many of these are part of Machine Learning techniques



UNIVERSITÉ DE FRIBOURG / UNIVERSITÄT FREIBURG | Biochemistry/Bioinformatics Unit | Falquet Laurent



Research goal	Assumed relationship	Input data	Technique
<ul style="list-style-type: none"> Explore main gradients of variation Reveal patterns of object similarity 	<ul style="list-style-type: none"> Linear Unimodal Any^{DM} 	<ul style="list-style-type: none"> Raw Raw Distance matrix 	<ul style="list-style-type: none"> PCA CA/DCA PCoA NMDS
<ul style="list-style-type: none"> Define groups of similar variables or objects 	<ul style="list-style-type: none"> Any^{DM} 	<ul style="list-style-type: none"> Distance matrix 	<ul style="list-style-type: none"> CLA
<ul style="list-style-type: none"> Reveal relationships between sets of variables 	<ul style="list-style-type: none"> Linear Any^{ORD} Any 	<ul style="list-style-type: none"> Raw Ordination output Any 	<ul style="list-style-type: none"> CCoA CIA PA
<ul style="list-style-type: none"> Identify gradients of variation in a set of measured variables explained by another set of variables 	<ul style="list-style-type: none"> Linear Unimodal Any^{LF} Any^{DM} 	<ul style="list-style-type: none"> Raw Raw Raw Distance matrix 	<ul style="list-style-type: none"> RDA PRC CCA GLM db-RDA
<ul style="list-style-type: none"> Discriminate object classes based on values of measured variables 	<ul style="list-style-type: none"> Linear Any^{KF} Any 	<ul style="list-style-type: none"> Raw Raw Raw 	<ul style="list-style-type: none"> OPLS-DA DFA SVM RF

Paliy & Shankar

Mol Ecol. 2016 March ; 25(5):

1032–1057.

doi:10.1111/mec.13536.

Figure 8. Diagram of potential choices of multivariate techniques based on the research goal, assumed relationship among variables, and input data structure

DM – assumed relationship depends on the chosen distance metric; ORD – assumed relationship depends on the ordination technique used; LF – assumed relationship depends on the chosen link function; KF – SVM model can be linear or non-linear if a non-linear kernel function is used.

Summary of the day

BIOM is a universal format (warning BIOM1 <> BIOM2)

Data Exploration is important to understand the results

Rarefaction curves are useful to evaluate coverage

Alpha and beta diversity are useful to detect biases

Multiple representations of the data are possible (plots, networks, heatmap, trees, etc.)

Differential analysis is possible with RNAseq tools or LEfSE

Choose the right tool to your question

UNIVERSITÉ DE FRIBOURG / UNIVERSITÄT FREIBURG | Biochemistry/Bioinformatics Unit | Falquet Laurent



Thank you for your attention. Questions?



UNIVERSITÉ DE FRIBOURG / UNIVERSITÄT FREIBURG | Biochemistry/Bioinformatics Unit | Falquet Laurent

