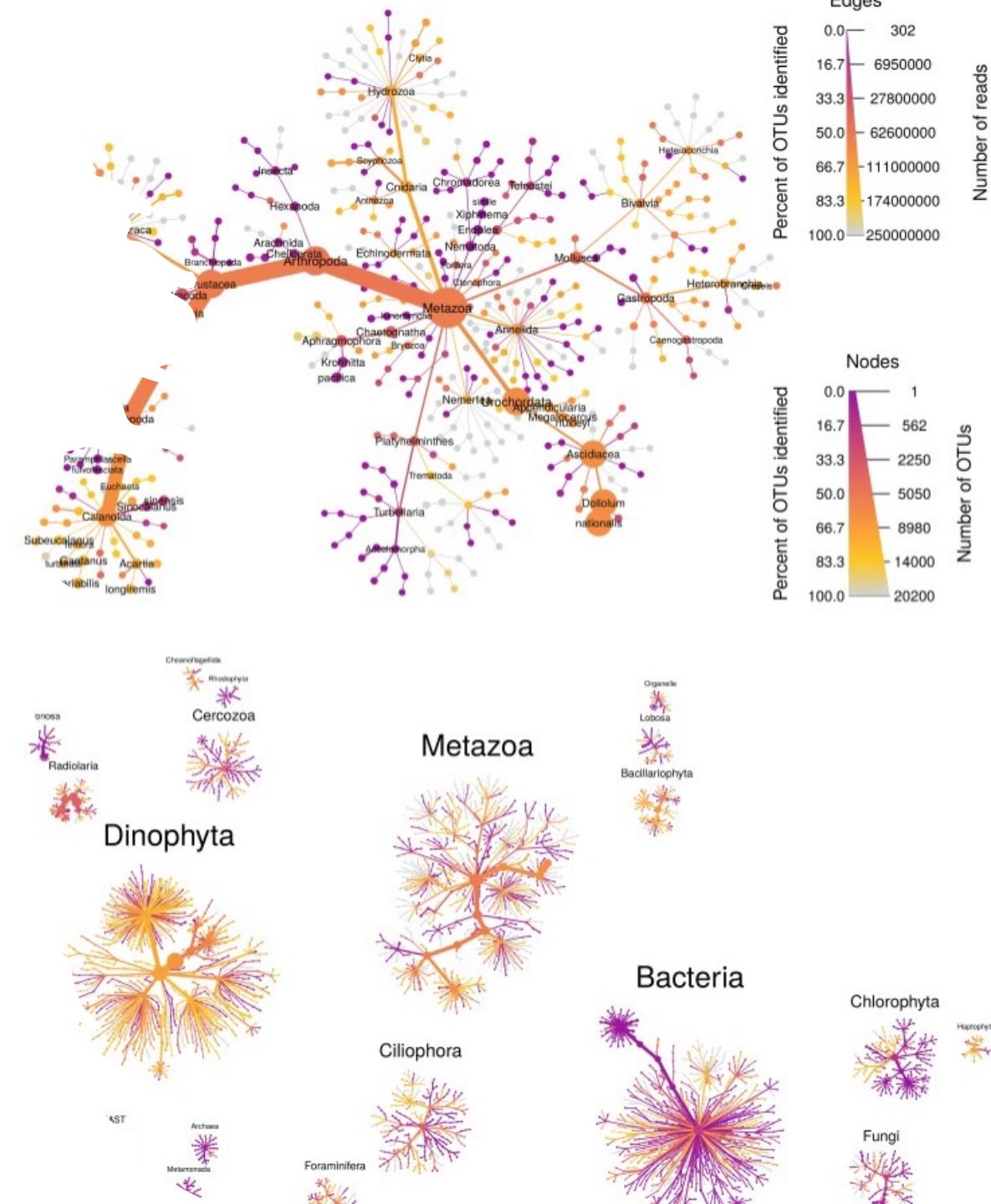


# Intro to metacoder

11.04.25, Ella Thoen



RESEARCH ARTICLE

# Metacoder: An R package for visualization and manipulation of community taxonomic diversity data

Zachary S. L. Foster<sup>1</sup>, Thomas J. Sharpton<sup>2,3,4</sup>, Niklaus J. Grünwald<sup>5\*</sup>

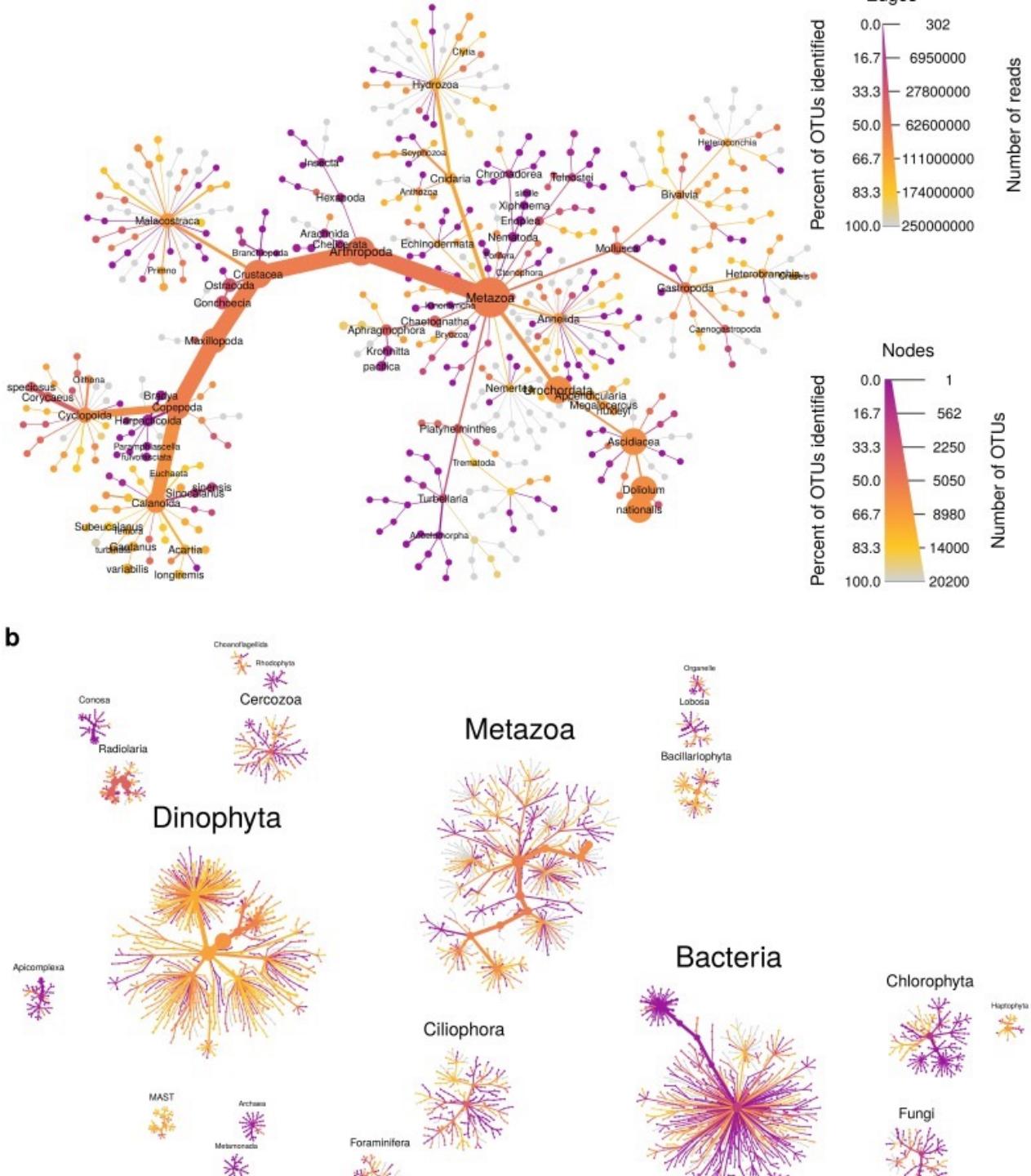
**1** Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon, United States of America, **2** Department of Microbiology, Oregon State University, Corvallis, Oregon, United States of America, **3** Department of Statistics, Oregon State University, Corvallis, Oregon, United States of America, **4** Center for Genome Research and Biocomputing, Oregon State University, Corvallis, Oregon, United States of America, **5** Horticultural Crops Research Laboratory, USDA-ARS, Corvallis, Oregon, United States of America



\* [nik.grunwald@ars.usda.gov](mailto:nik.grunwald@ars.usda.gov)

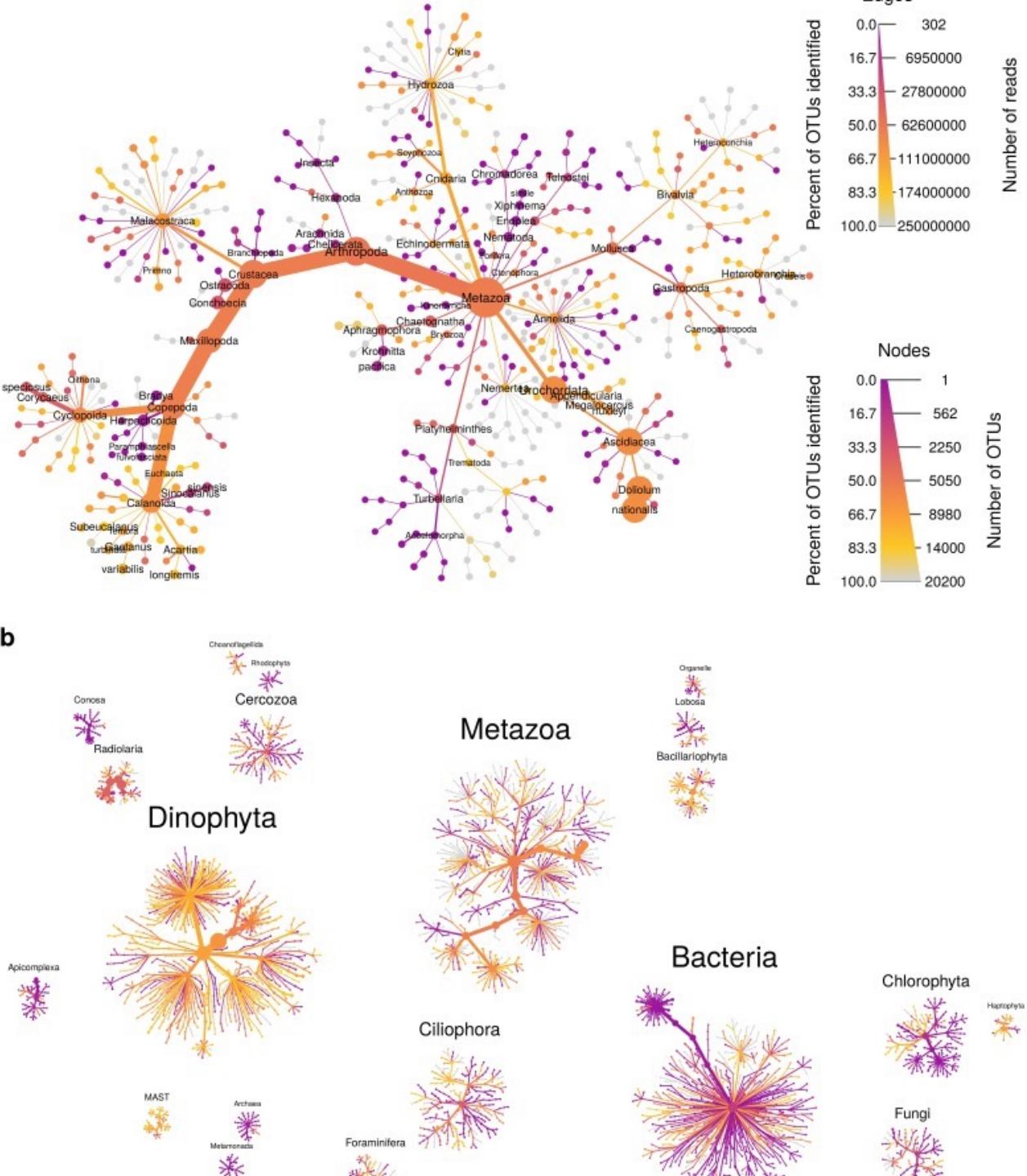
# R package for plotting taxonomic data

- Parsing, plotting and manipulating large taxonomic datasets
  - Includes
    - Summing reads per taxon
    - rarefaction curves
    - Converting to proportions
    - Combining data for group variables
    - Converting to and from *phyloseq* and *metacoder* formats
    - Plotting taxonomic trees!



# R package for plotting taxonomic data

- Uses Tidyverse data structure (e.g. tibbles)
  - Communicates well with *phyloseq*
  - Uses tree terminology
    - Nodes and edges



# Very good documentation!

Check out:

[https://grunwaldlab.github.io/metacoder\\_documentation/index.html](https://grunwaldlab.github.io/metacoder_documentation/index.html)



## Metacoder documentation

Zachary S. L. Foster and Niklaus J. Grünwald

2021-07-23

### Introduction

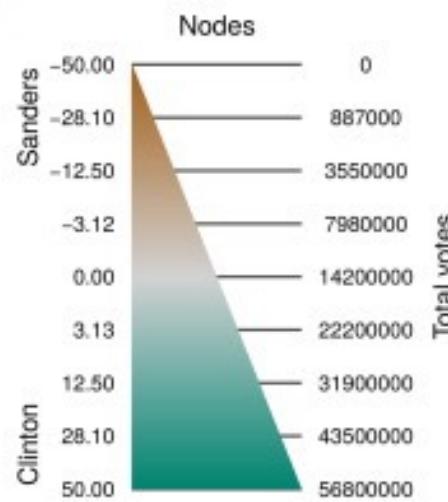
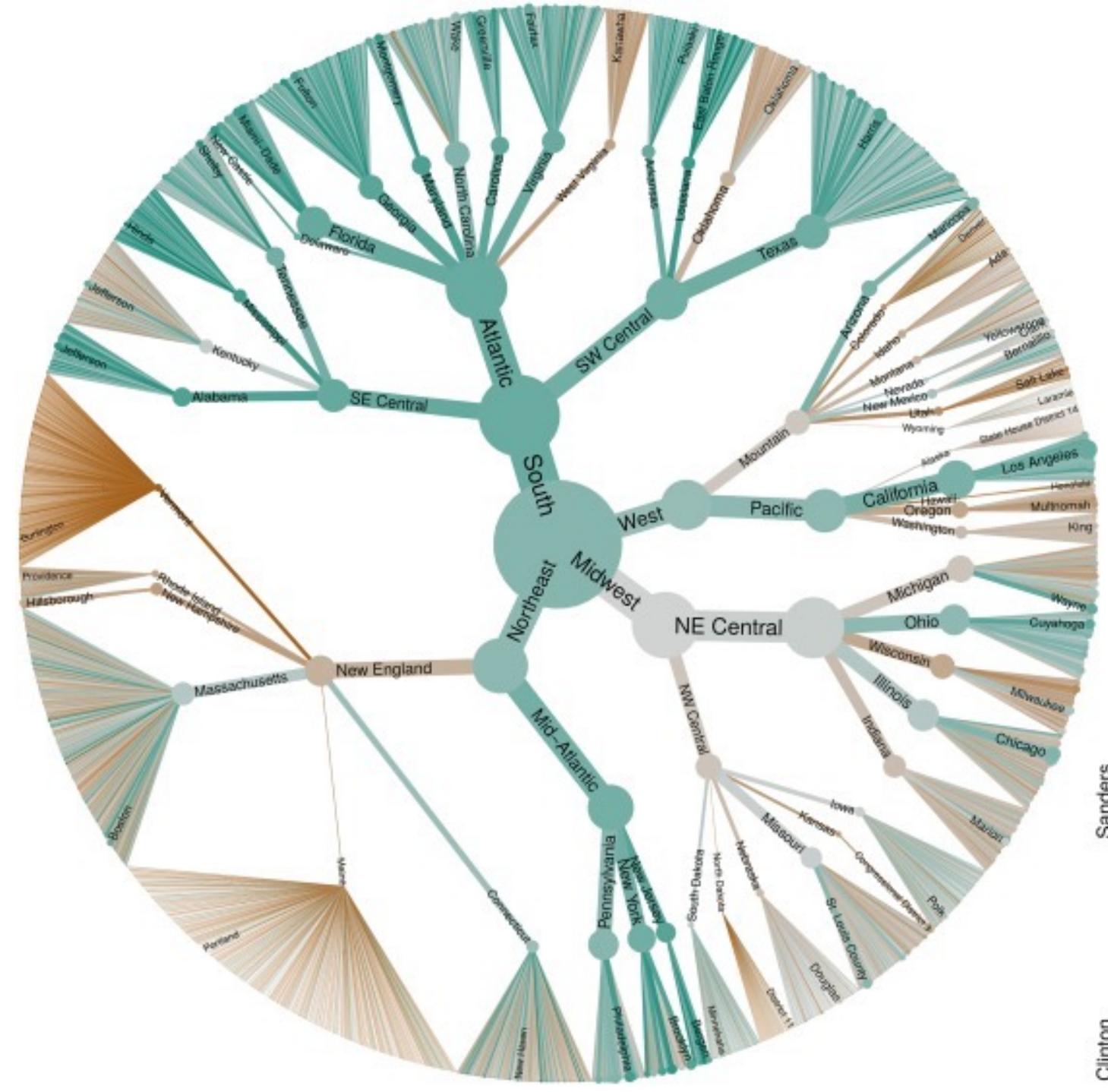
Metacoder is an R package for `parsing`, plotting, and manipulating large **taxonomic** data sets, like those generated from metagenomic sequencing, like metabarcoding (i.e. amplification metagenomics, 16S metagenomics, etc). It provides a tree-based visualization of “phylogenetic trees” used to depict statistics for every taxon in a taxonomy using color and size. It also provides various functions to do common microbiome bioinformatics on data in the `taxmap` format defined by the `metacoder` package, such as:

- Summing read counts/abundance per taxon
- Converting counts to proportions and `rarefaction` of counts using `vegan`
- Comparing the abundance (or other characteristics) of groups of samples (e.g., experimental treatments) per taxon
- Combining data for groups of samples
- Simulated PCR, via EMBOSS primersearch, for testing primer specificity and coverage of taxonomic groups
- Converting common microbiome formats for data and reference databases into the objects defined by the `metacoder` package
- Converting to and from the `phyloseq` format and the `metacoder` format

### Relationship with other packages

Many of these operations can be done using other packages like `phyloseq`, which also provides tools for diversity analysis. The strength of `metacoder` is that its functions use the flexible data types defined by `metacoder`, which has powerful parsing and subsetting functions that take into account the hierarchical relationship between taxa and user-defined data. In general, `metacoder` is more of an abstract package, while `phyloseq` has more specialized functions for community diversity data, but they both can do similar things. I encourage you to try both and see which fits your needs and style best. You can also combine the two in a single analysis by converting between the two data formats using the `parse_phyloseq` and `as_phyloseq` functions.

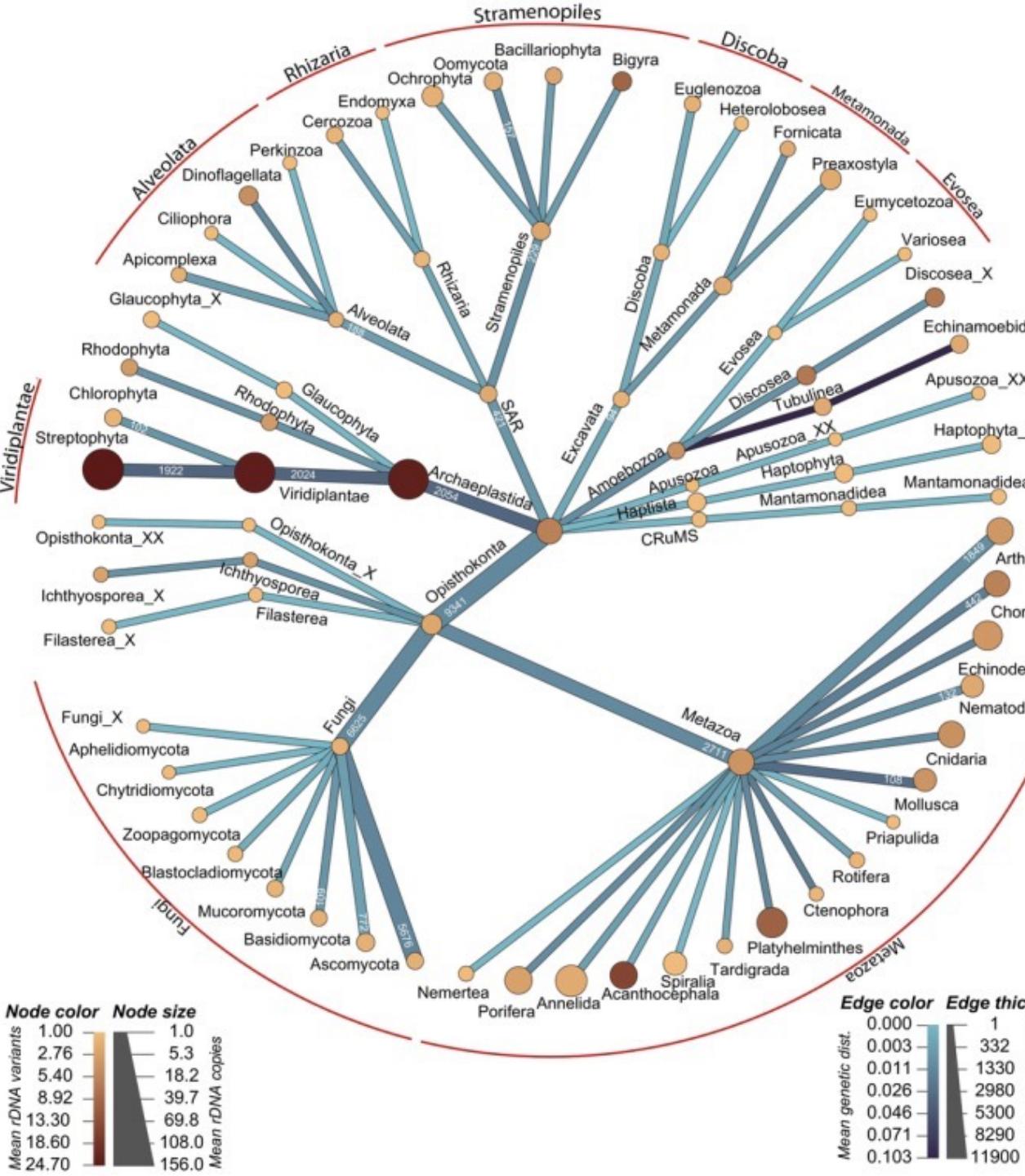
Metacoder can be used with any data that can be ordered by hierarchy



# Possible to show many layers of information

Figure from **The Ribosomal Operon Database: A Full-Length rDNA Operon Database Derived From Genome Assemblies (Krabberød et al 2025)**, where

- node color indicates mean number of ribosomal operon variants
- node size indicates number of copies
- Edge colour indicates genetic distance
- Edge thickness indicates number of genomes





# R example: Roots and soil from Vestland climate grid

---



# Dataset

- 3 sample types: soil, *P. erecta* roots and *B. vivipara* roots in 12 sites
- CTAB+Chloroform extractions followed by E.Z.N.A. soil kit
- ITS1 region from all samples using Illumina miSeq
- DADA2 pipeline followed by VSEARCH clustering at 97% and LULU

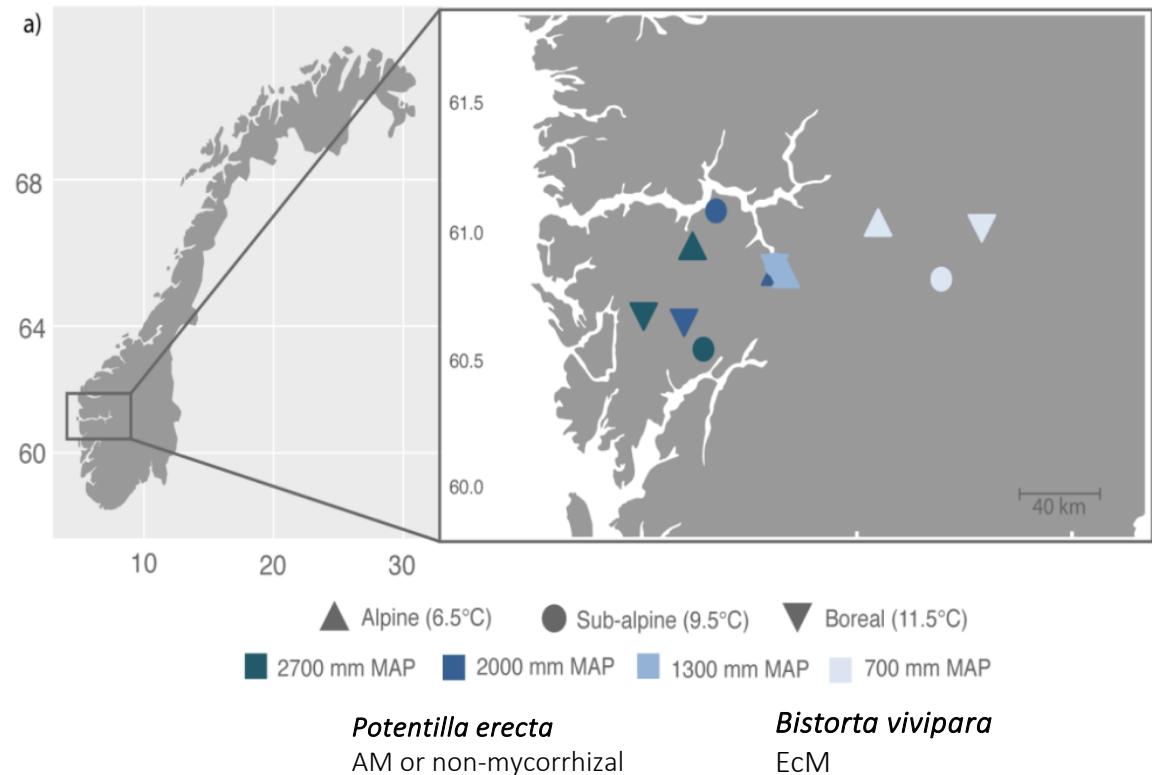


Figure credit: Erlend Y. Fines

# Dataset

- Reduced example dataset consists of
  - 94 samples,
  - 1362 OTUs
  - Rarified to 500 reads

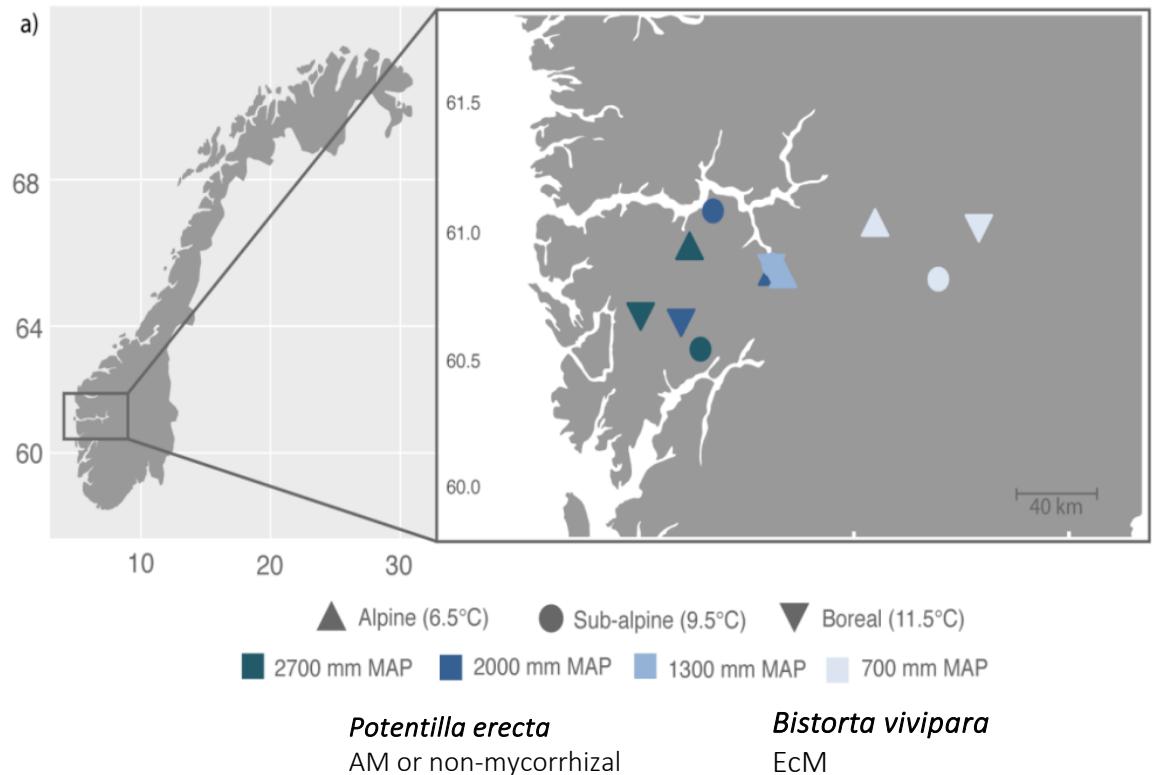


Figure credit: Erlend Y. Fines

## OUT table characteristics

- OTUs as rows
- Samples as columns
- Needs to contain a single column with taxonomy in the format:  
k\_Fungi;p\_Basidiomycota;c\_Agaricomycetes;o\_Agaricales;f\_Clavariaceae;g\_Clavulinopsis;s\_Clavulinopsis\_corniculat

	A	CC	CD	CE	CF	CG	CH	CI	CJ	CK	CL	CM	CN	CO	CP	CQ	CR	
1	OTUid	DS31	DS32	DS34	DS37	DS38	DS39	DS47	DS48	DS50	DS52	DS54	DS55	DS58	DS60	DS61	size	taxonomy
2	OTU0778	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
3	OTU1150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
4	OTU1576	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
5	OTU0517	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
6	OTU0729	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
7	OTU0958	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
8	OTU0730	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	3 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
9	OTU0249	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
10	OTU0788	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	3	4 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
11	OTU2041	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
12	OTU2014	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	5 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
13	OTU0282	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	6 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
14	OTU1577	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
15	OTU0283	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
16	OTU0726	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
17	OTU0957	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
18	OTU0576	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	14 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
19	OTU0766	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
20	OTU0769	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	16 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
21	OTU0731	0	0	0	0	1	4	4	0	0	0	0	0	0	0	0	0	17 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
22	OTU1072	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom
23	OTU0061	0	0	0	0	0	0	0	0	0	0	0	0	0	8	1	0	22 k_Fungi;p_Ascomycota;c_Archaeorhizomycetes;o_Archaeorhizom

## Metadata table characteristics

- Sample IDs as columns (needs to match sample IDs in the OTU table)
- Metadata as columns – can be numerical or characters

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
1	SampleID	Sample	Type	Location	Temperatu	Precipitati	pH	N	C	C:N	Free_ergost	Total_ergos	Long	Lat	Wind	Perc	Cloud	Temp_1	Mo
2	DB04	Bistorta	Plantroot	Ves	mid	wet	0.718	0.367	0.223	0.414	0.192	0.524	0	0.328	0.37	0.691	0.092	0.328	
3	DB05	Bistorta	Plantroot	Ves	mid	wet	0.718	0.367	0.223	0.414	0.192	0.524	0	0.328	0.37	0.691	0.092	0.328	
4	DB06	Bistorta	Plantroot	Ves	mid	wet	0.718	0.367	0.223	0.414	0.192	0.524	0	0.328	0.37	0.691	0.092	0.328	
5	DB07	Bistorta	Plantroot	Ves	mid	wet	0.718	0.367	0.223	0.414	0.192	0.524	0	0.328	0.37	0.691	0.092	0.328	
6	DB08	Bistorta	Plantroot	Arh	low	semi-wet	0.385	0.396	0.602	0.742	0.518	0.395	0.158	0.239	0.372	0.851	0	0.995	
7	DB09	Bistorta	Plantroot	Arh	low	semi-wet	0.385	0.396	0.602	0.742	0.518	0.395	0.158	0.239	0.372	0.851	0	0.995	
8	DB10	Bistorta	Plantroot	Arh	low	semi-wet	0.385	0.396	0.602	0.742	0.518	0.395	0.158	0.239	0.372	0.851	0	0.995	
9	DB14	Bistorta	Plantroot	Ulv	alpine	dry	0.663	0.626	0.485	0.402	0.693	0.554	0.84	0.823	0.481	0.32	0.64	0.007	
10	DB15	Bistorta	Plantroot	Ulv	alpine	dry	0.663	0.626	0.485	0.402	0.693	0.554	0.84	0.823	0.481	0.32	0.64	0.007	
11	DB16	Bistorta	Plantroot	Ulv	alpine	dry	0.663	0.626	0.485	0.402	0.693	0.554	0.84	0.823	0.481	0.32	0.64	0.007	
12	DB20	Bistorta	Plantroot	Lav	alpine	semi-dry	0.744	0.422	0.739	0.774	0.343	0.497	0.412	0.61	0.578	0.476	0.679	0.049	
13	DB21	Bistorta	Plantroot	Lav	alpine	semi-dry	0.744	0.422	0.739	0.774	0.343	0.497	0.412	0.61	0.578	0.476	0.679	0.049	
14	DB22	Bistorta	Plantroot	Lav	alpine	semi-dry	0.744	0.422	0.739	0.774	0.343	0.497	0.412	0.61	0.578	0.476	0.679	0.049	
15	DB26	Bistorta	Plantroot	Alr	mid	dry	0.513	0.46	0.629	0.69	0.638	0.487	0.409	0.936	0.13	0.493	1	0.321	
16	DB27	Bistorta	Plantroot	Alr	mid	dry	0.513	0.46	0.629	0.69	0.638	0.487	0.409	0.936	0.13	0.493	1	0.321	
17	DB28	Bistorta	Plantroot	Alr	mid	dry	0.513	0.46	0.629	0.69	0.638	0.487	0.409	0.936	0.13	0.493	1	0.321	

# Let's move to R!

