

Introduction to metacoder

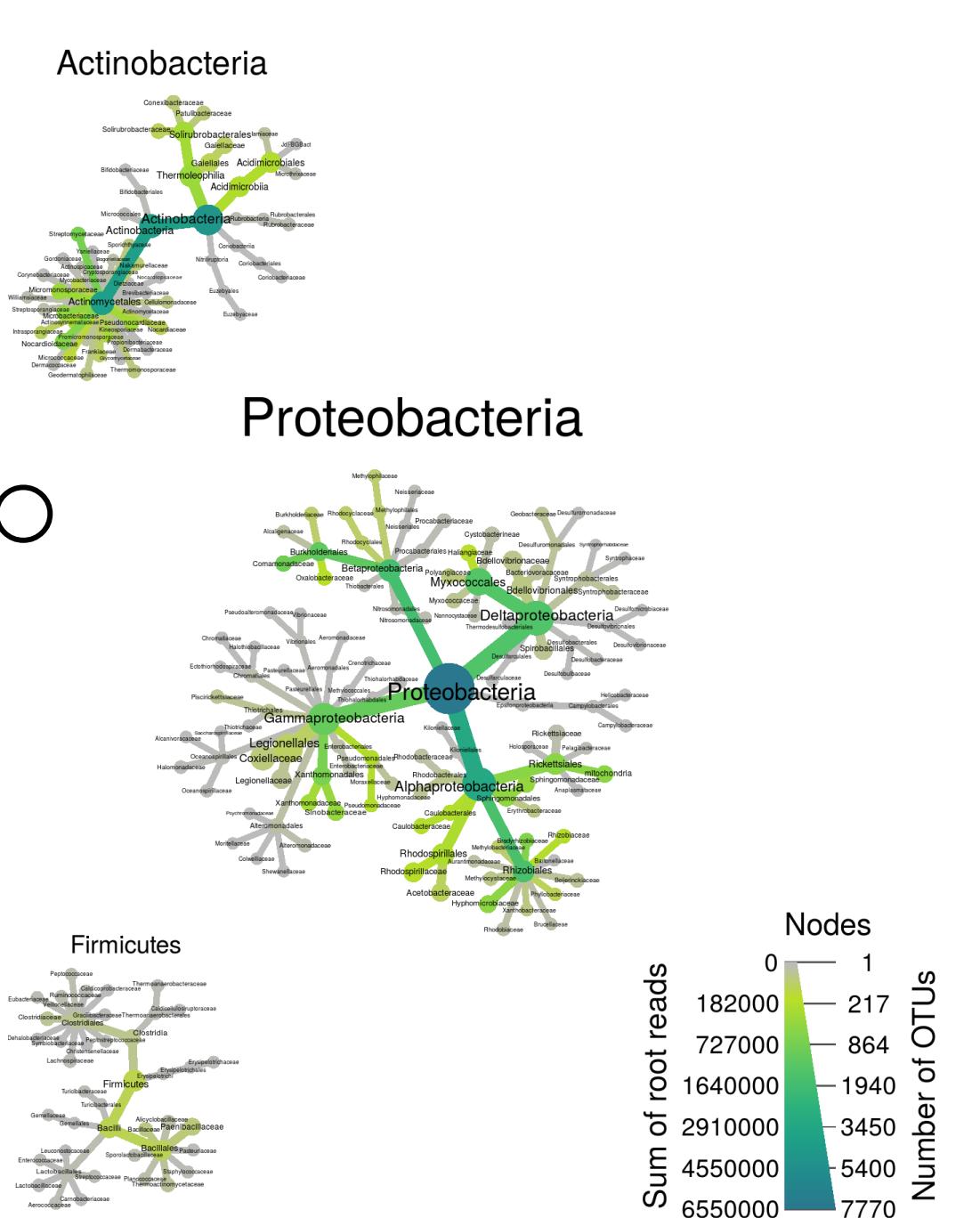
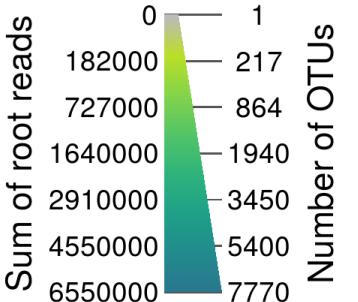
Ella Thoen
21.04.23

Actinobacteria

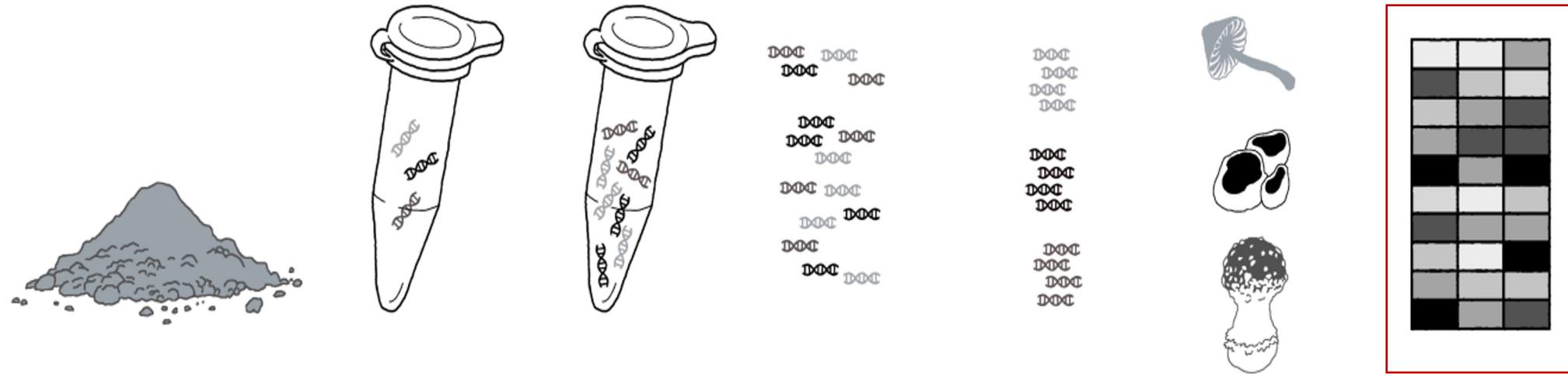
Proteobacteria

Firmicutes

Nodes



How to visualize the taxonomic data from your metabarcoding study?

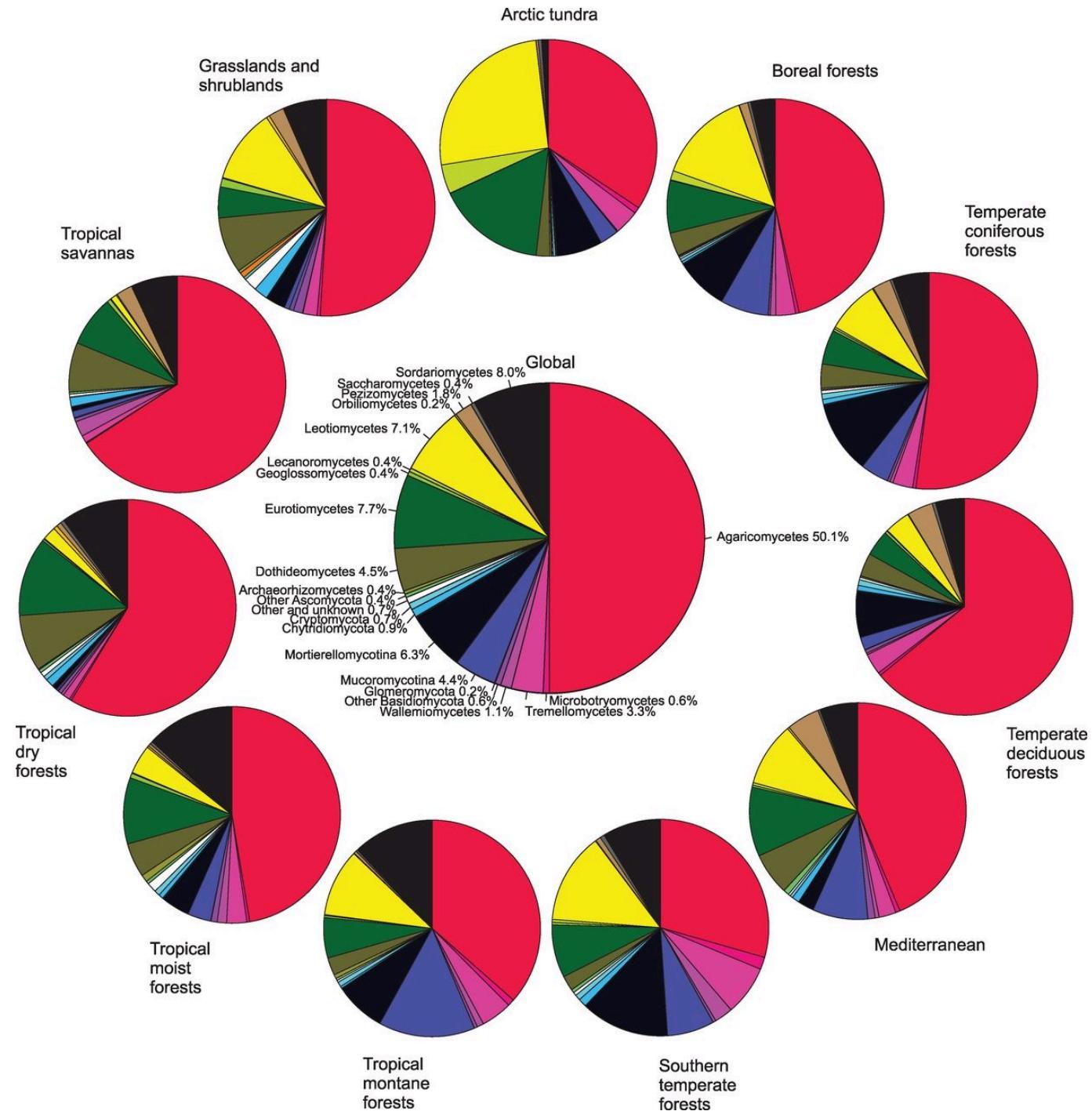


OTU table

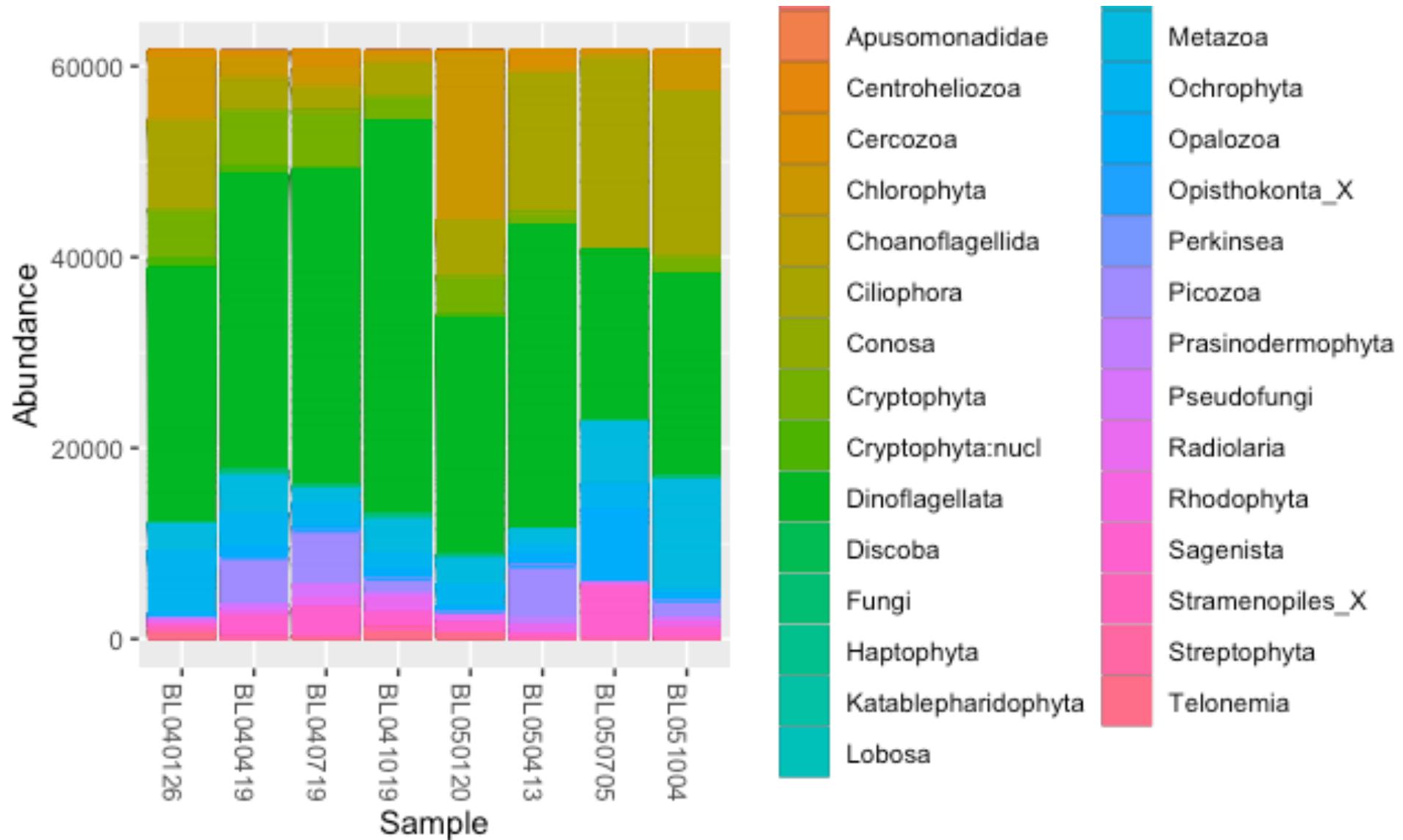
OTUid	DB04	DB05	DB06	DB07	DB08	DB09	DB10	DB14	DB15	DB16	DB20	DB21	DB22	I
OTU2586	6	19	73	20	10	28	12	24	87	16	10	4	9	
OTU1871	21	6	31	32	1	22	35	16	8	2	41	72	0	
OTU1503	1	1	2	1	10	21	6	7	1	2	5	3	0	
OTU2202	18	31	5	7	6	6	6	2	6	1	2	10	0	
OTU1996	94	1	5	3	7	13	2	143	104	76	99	3	0	
OTU0649	0	1	0	0	0	0	0	6	3	0	0	0	0	
OTU1017	0	0	1	0	0	0	0	35	9	89	8	2	0	
OTU0607	140	11	1	138	3	16	45	0	0	1	1	9	2	
OTU2328	0	0	2	0	0	9	2	0	0	0	1	0	0	
OTU1799	1	2	5	1	1	15	0	0	0	0	0	1	0	
OTU2260	11	0	0	0	0	0	1	0	0	0	0	0	0	
OTU1311	0	0	4	0	17	16	14	3	2	2	0	2	0	
OTU0617	0	0	0	1	12	0	6	0	2	0	0	1	0	
OTU2320	3	0	2	3	4	8	79	0	0	0	0	1	0	
OTU1243	0	0	0	3	2	0	0	1	3	5	0	0	0	
OTU1145	0	3	1	0	1	3	1	4	0	0	0	0	0	
OTU2059	0	1	1	0	0	2	0	4	0	3	0	2	0	
OTU1396	0	0	1	1	5	2	0	0	0	0	0	0	0	
OTU0851	1	0	4	2	2	12	1	0	0	0	0	1	0	
OTU2359	0	1	1	0	0	1	0	2	0	0	0	0	0	
OTU0735	0	0	0	0	0	0	0	0	1	0	0	0	0	
OTU1628	0	1	3	3	3	9	0	0	0	0	0	0	0	
OTU2548	2	0	0	1	0	2	1	0	0	0	1	0	0	
OTU1990	0	3	0	0	0	0	0	5	0	0	0	0	196	
OTU2100	4	0	0	0	15	5	1	0	0	0	0	0	0	
OTU2336	0	0	0	0	1	0	1	0	0	0	0	0	0	
OTU0799	0	0	0	0	0	0	0	0	0	0	0	0	0	
OTU2424	0	0	1	0	0	1	46	0	0	0	0	0	0	
OTU1131	0	0	0	0	0	8	1	0	0	0	0	0	0	
OTU1258	6	8	0	0	0	4	0	0	0	0	0	1	0	

SampleID	Sample	Type	Location	Temperat	Precipitat	pH	N	C	C:N	Free_ergc	Total_erg
DB04	Bistorta	Plantroot	Ves	mid	wet	0.718	0.367	0.223	0.414	0.192	0.524
DB05	Bistorta	Plantroot	Ves	mid	wet	0.718	0.367	0.223	0.414	0.192	0.524
DB06	Bistorta	Plantroot	Ves	mid	wet	0.718	0.367	0.223	0.414	0.192	0.524
DB07	Bistorta	Plantroot	Ves	mid	wet	0.718	0.367	0.223	0.414	0.192	0.524
DB08	Bistorta	Plantroot	Arh	low	semi-wet	0.385	0.396	0.602	0.742	0.518	0.395
DB09	Bistorta	Plantroot	Arh	low	semi-wet	0.385	0.396	0.602	0.742	0.518	0.395
DB10	Bistorta	Plantroot	Arh	low	semi-wet	0.385	0.396	0.602	0.742	0.518	0.395
DB14	Bistorta	Plantroot	Ulv	alpine	dry	0.663	0.626	0.485	0.402	0.693	0.554
DB15	Bistorta	Plantroot	Ulv	alpine	dry	0.663	0.626	0.485	0.402	0.693	0.554
DB16	Bistorta	Plantroot	Ulv	alpine	dry	0.663	0.626	0.485	0.402	0.693	0.554
DB20	Bistorta	Plantroot	Lav	alpine	semi-dry	0.744	0.422	0.739	0.774	0.343	0.497
DB21	Bistorta	Plantroot	Lav	alpine	semi-dry	0.744	0.422	0.739	0.774	0.343	0.497
DB22	Bistorta	Plantroot	Lav	alpine	semi-dry	0.744	0.422	0.739	0.774	0.343	0.497
DB26	Bistorta	Plantroot	Alr	mid	dry	0.513	0.46	0.629	0.69	0.638	0.487
DB27	Bistorta	Plantroot	Alr	mid	dry	0.513	0.46	0.629	0.69	0.638	0.487
DB28	Bistorta	Plantroot	Alr	mid	dry	0.513	0.46	0.629	0.69	0.638	0.487

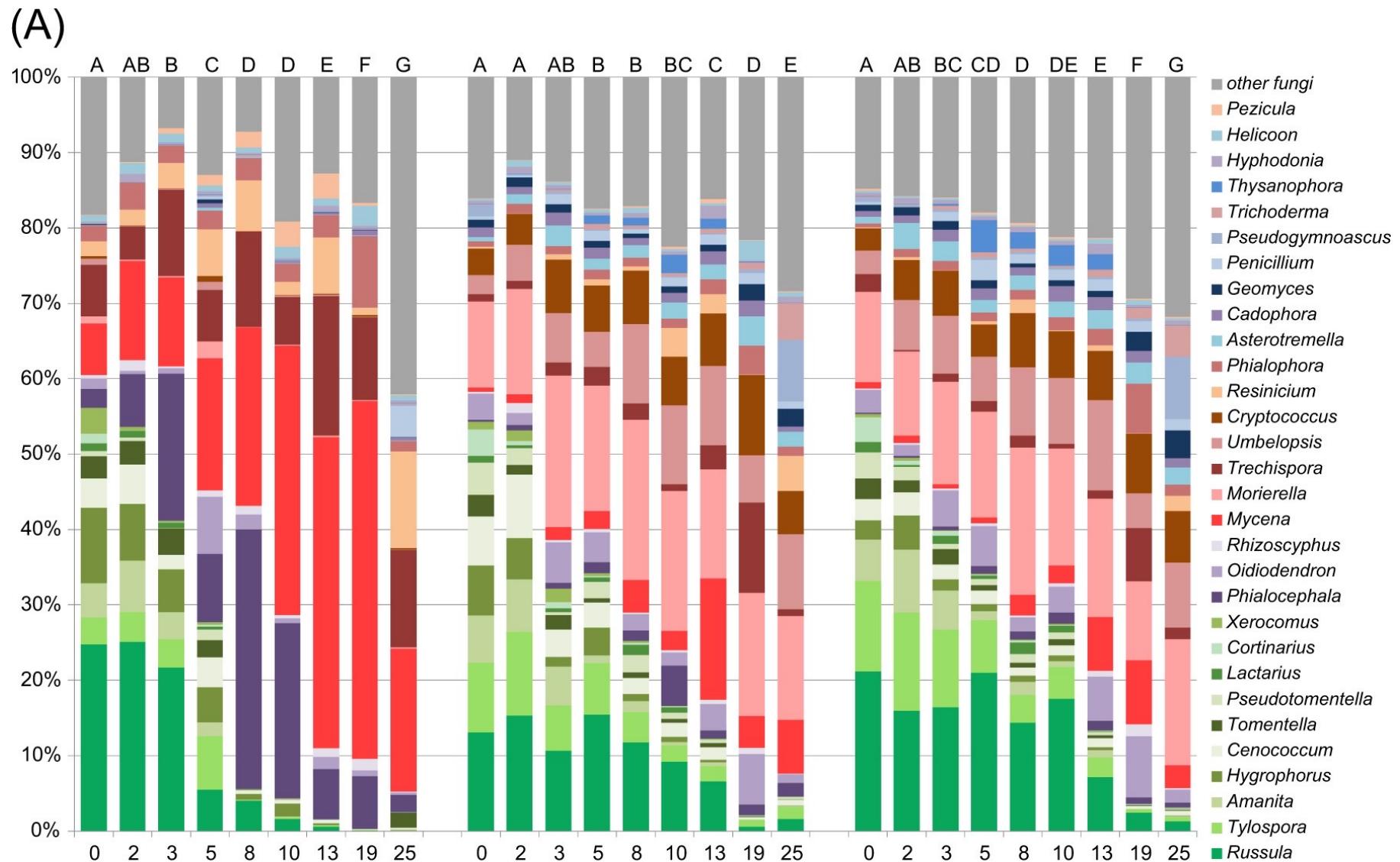
Pie chart?



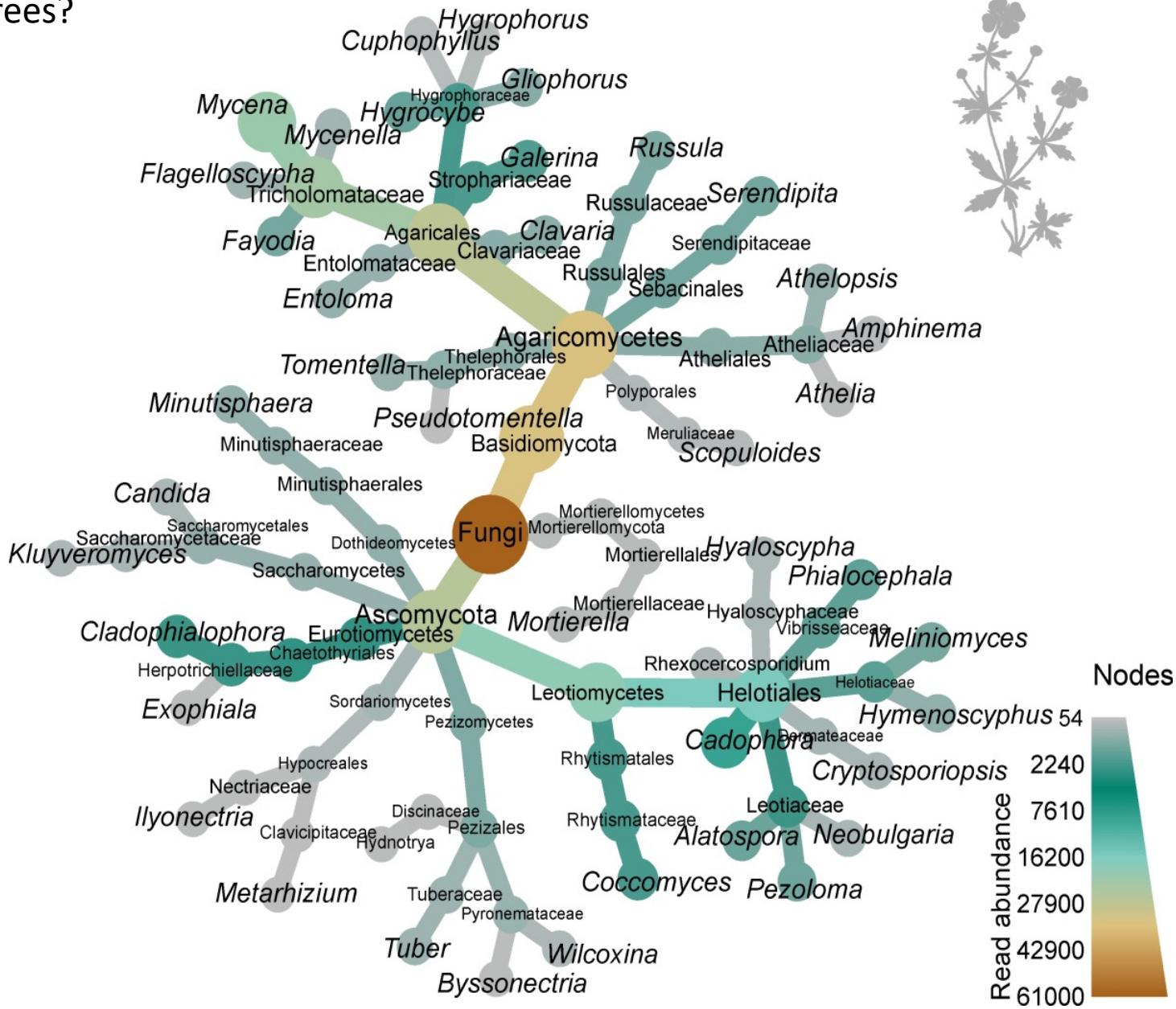
Stacked bar chart?

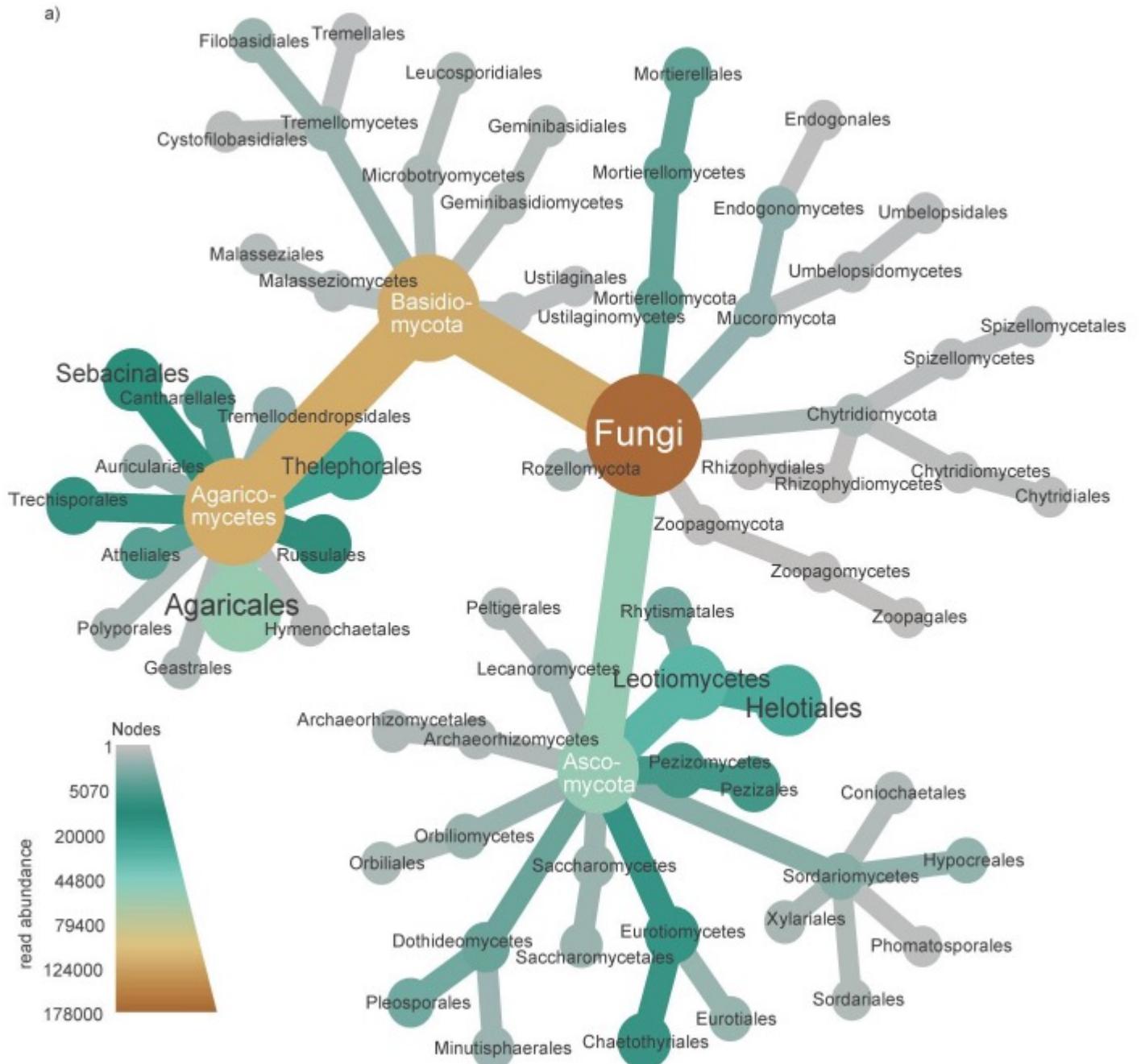


Stacked bar chart?



Heat trees?





Heat trees with metacoder

Metacoder uses the phylogenetic backbone of your taxonomic data to construct heat trees

=NOT phylogenetic trees!

Depict statistics for every taxon in a taxonomy using color and size

- Nodes (circles) scaled and colored according to read abundance
- Edges (branches) scaled and colored by read abundance

Barcharts vs Heat trees with metacoder

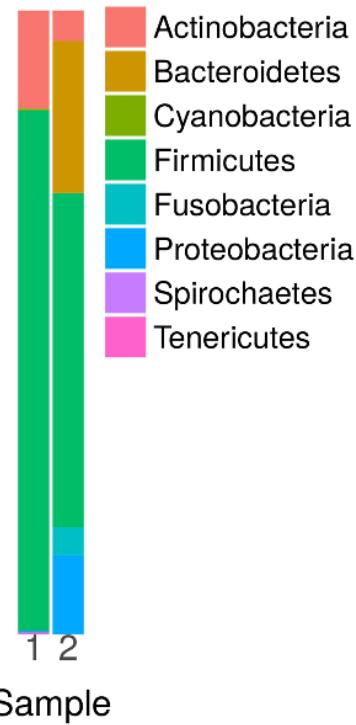


Fig 2. from Foster *et al* 2017

Heat trees allow for a better understanding of community structure than stacked bar charts

Heat trees with metacoder

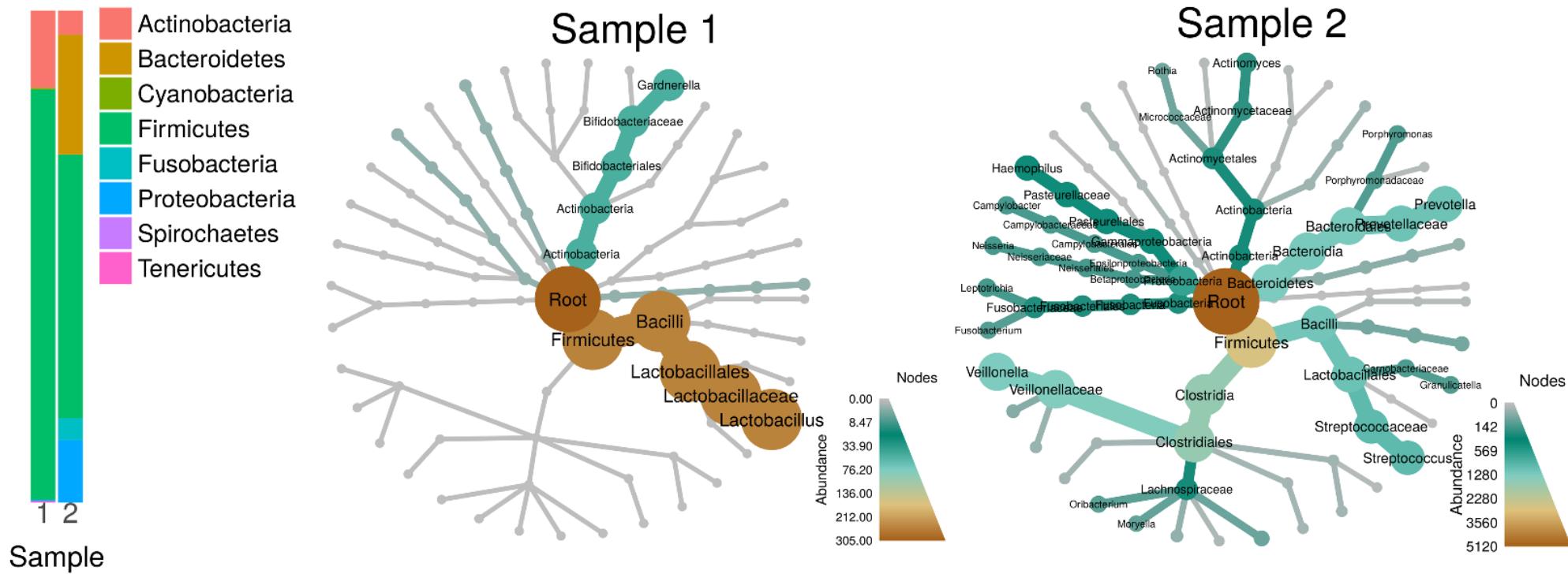


Fig 2. from Foster *et al* 2017

Heat trees allow for a better understanding of community structure than stacked bar charts

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

Zachary S. L. Foster, Thomas J. Sharpton, Niklaus J. Grünwald 

An R package for parsing, **plotting** and manipulating large taxonomic datasets

What can you do with metacoder?

- Summing read counts/relative abundance per taxon
- Converting counts to proportions and rarefaction of counts using `vegan`
- Comparing the abundance of groups of samples per taxon
- Combining data for groups of samples
- Simulated PCR (primer specificity and coverage of taxonomic groups)
- Converting common microbiome formats for data and reference databases into the objects defined by the `taxa` package.
- Converting to and from the `phyloseq` format and the `taxa` format

But most important: data visualization with heat trees!



Metacoder documentation

Zachary S. L. Foster and Niklaus J. Grünwald

2021-07-23

Introduction

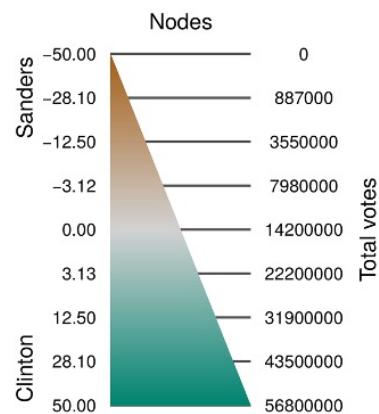
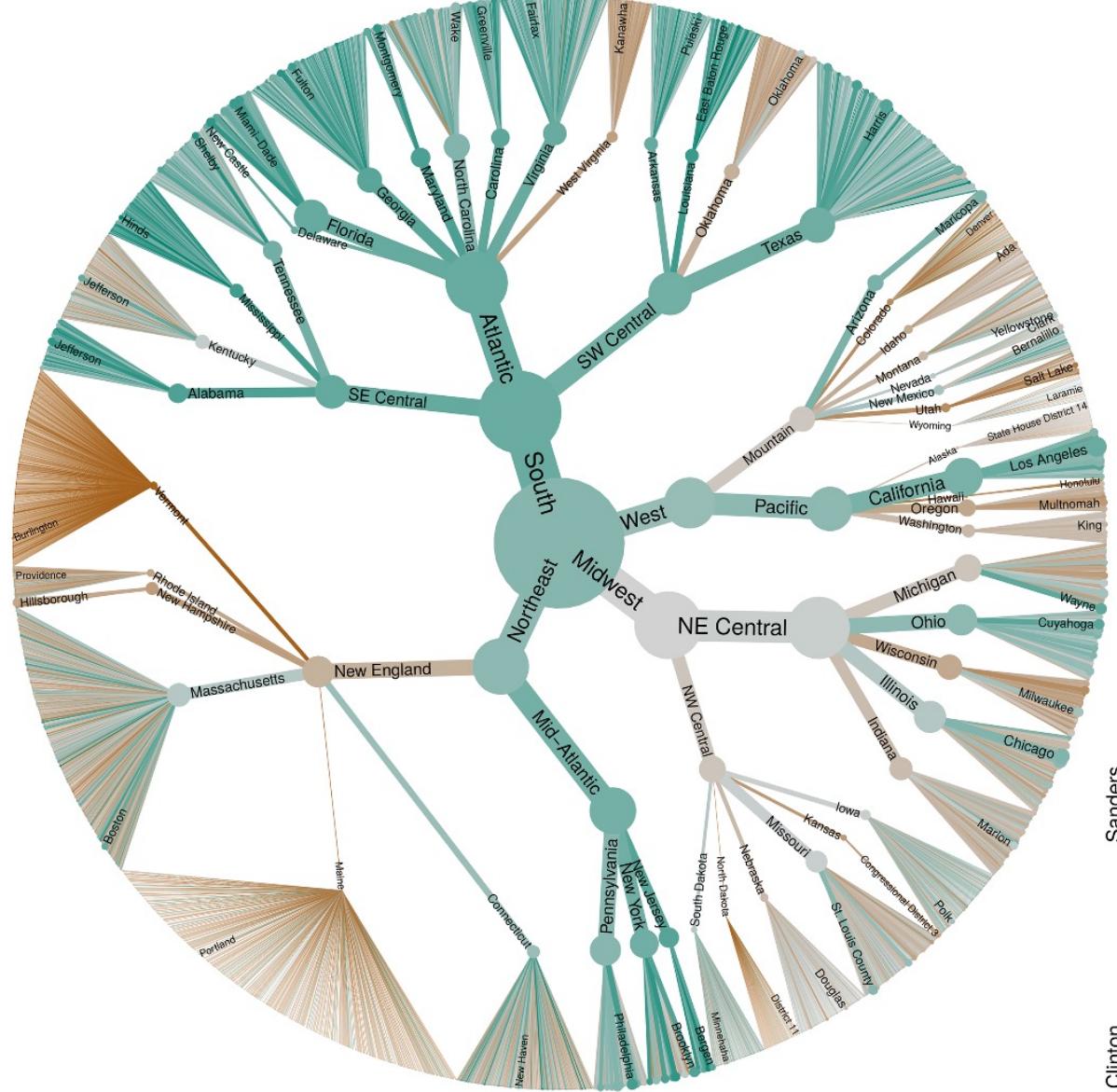
Metacoder is an R package for **parsing**, plotting, and manipulating large **taxonomic** data sets, like those generated from modern high-throughput sequencing, like metabarcoding (i.e. amplification metagenomics, 16S metagenomics, etc). It provides a tree-based visualization called "heat trees" used to depict statistics for every taxon in a taxonomy using color and size. It also provides various functions to do common tasks in 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 main strength of `metacoder` is that its functions use the flexible data types defined by `metacoder`, which has powerful parsing and subsetting abilities that take into account the hierarchical relationship between taxa and user-defined data. In general, `metacoder` is more of an abstracted tool kit, whereas `phyloseq` has more specialized functions for community diversity data, but they both can do similar things. I encourage you to try both to see which fits your needs and style best. You can also combine the two in a single analysis by converting between the two data types when needed.

https://grunwaldlab.github.io/metacoder_documentation/index.html



metacoder can be used with any type of data that can be organized hierarchically

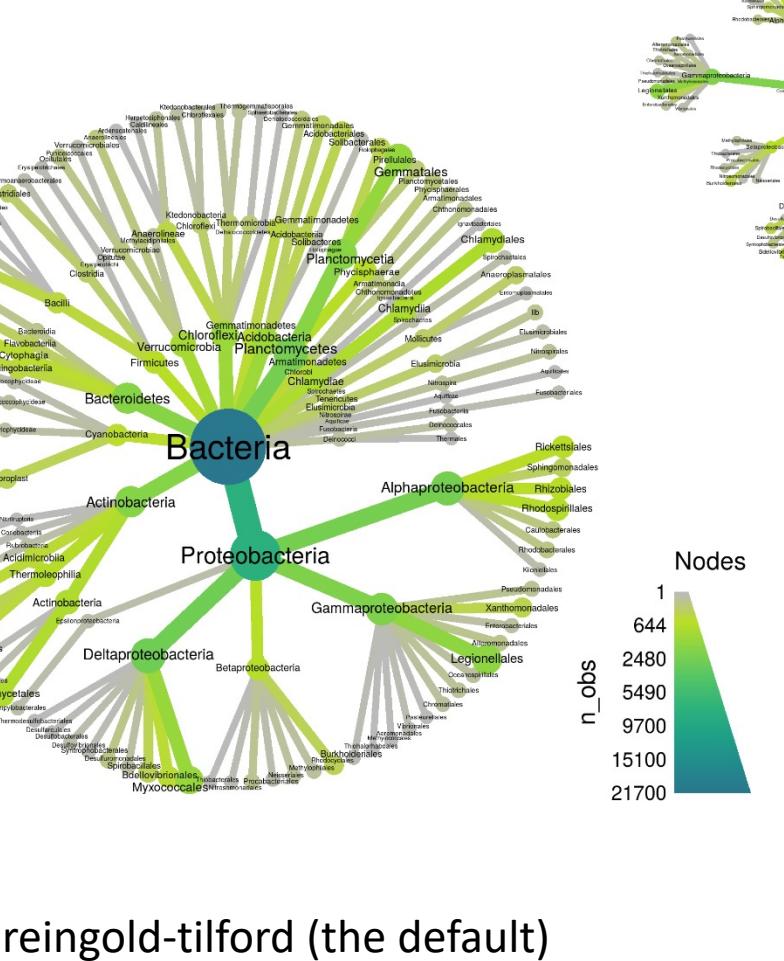
Fig. 6. Foster et al 2017

Relationship with other packages

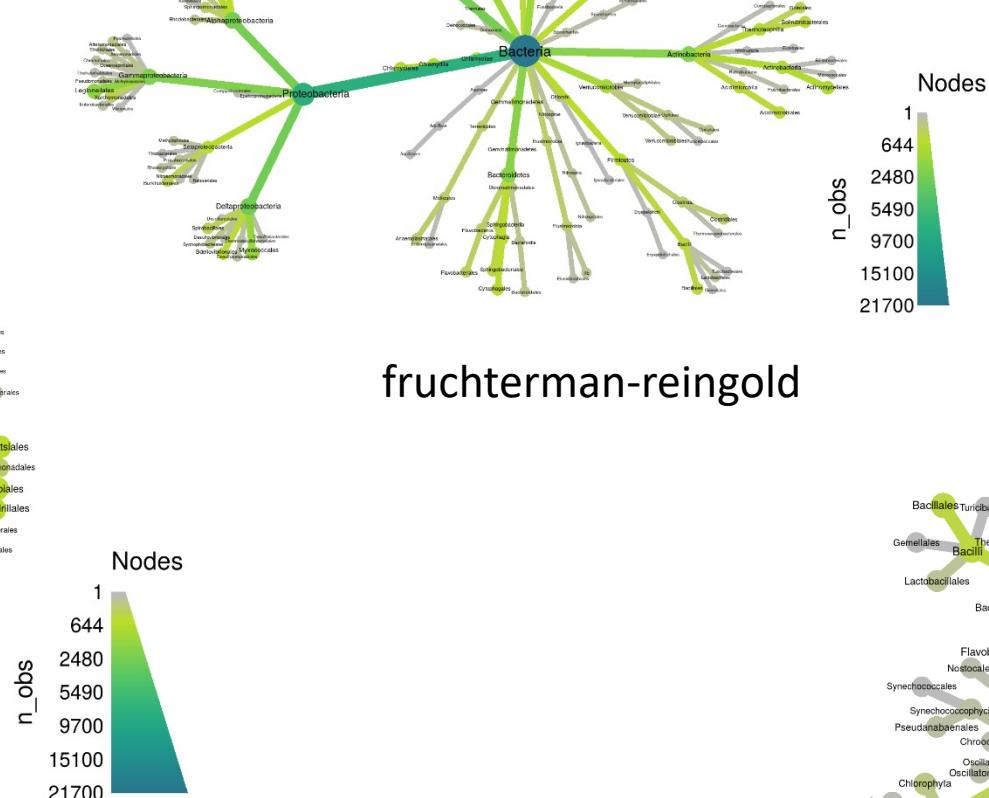
- metacoder can do similar things to e.g. phyloseq
- Uses the flexible data types defined by taxa
- Plotting is using ggplot2 – graphics
- Your data is organized in tibbles
(Tidyverse)
→ easy to use in other analyses

```
> print(obj)
<Taxmap>
1310 taxa: aab. Fungi, aac. Ascomycota ... byj. Rozellomycota_sp, byk. Fungi_sp
1310 edges: NA->aab, aab->aac, aab->aae ... bdh->byi, bdi->byj, bdj->byk
2 data sets:
  tax_data:
    # A tibble: 1,361 x 119
      taxon_id OTUID DB04 DB05 DB06 DB07 DB08 DB09 DB10 DB11 DB12 DB13 DB14 DB15 DB16 DB20
      <chr>   <chr> <dbl> <dbl>
    1 bdk     OTU0778 0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
    2 bdk     OTU1150 0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
    3 bdk     OTU1576 0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
    # ... with 1,358 more rows, and 106 more variables: DB21 <dbl>, DB22 <dbl>,
    # DB26 <dbl>, DB27 <dbl>, DB28 <dbl>, DB32 <dbl>, DB33 <dbl>, DB34 <dbl>,
    # DB38 <dbl>, DB39 <dbl>, ...
  class_data:
    # A tibble: 9,527 x 5
      taxon_id input_index tax_rank name          regex_match
      <chr>       <int>   <chr>   <chr>           <chr>
      1 aab         1       k       Fungi          k_Fungi
      2 aac         1       p       Ascomycota    p_Ascomycota
      3 aao         1       c       Archaeorhizomycetes c_Archaeorhizomycetes
    # ... with 9,524 more rows
  0 functions:
> |
```

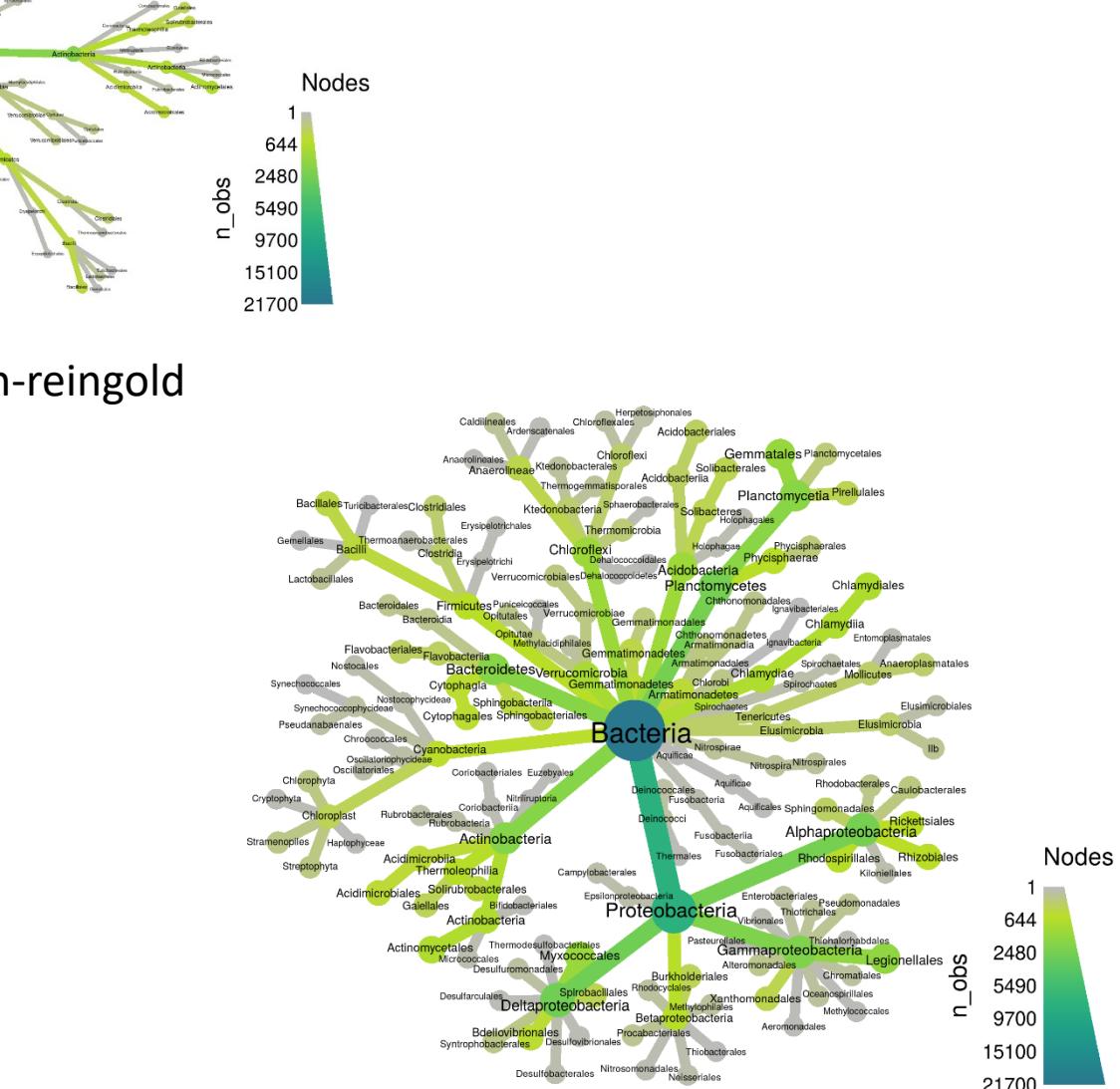
Standard heat tree heat_tree()



reingold-tilford (the default)



fruchterman-reingold



Combination of "reingold-tilford" and "davidson-harel"

Visualization of «standard» metabarcoding datasets

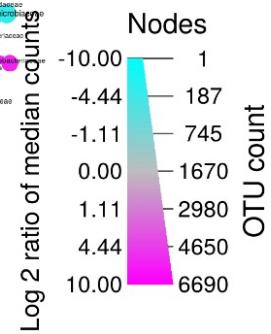
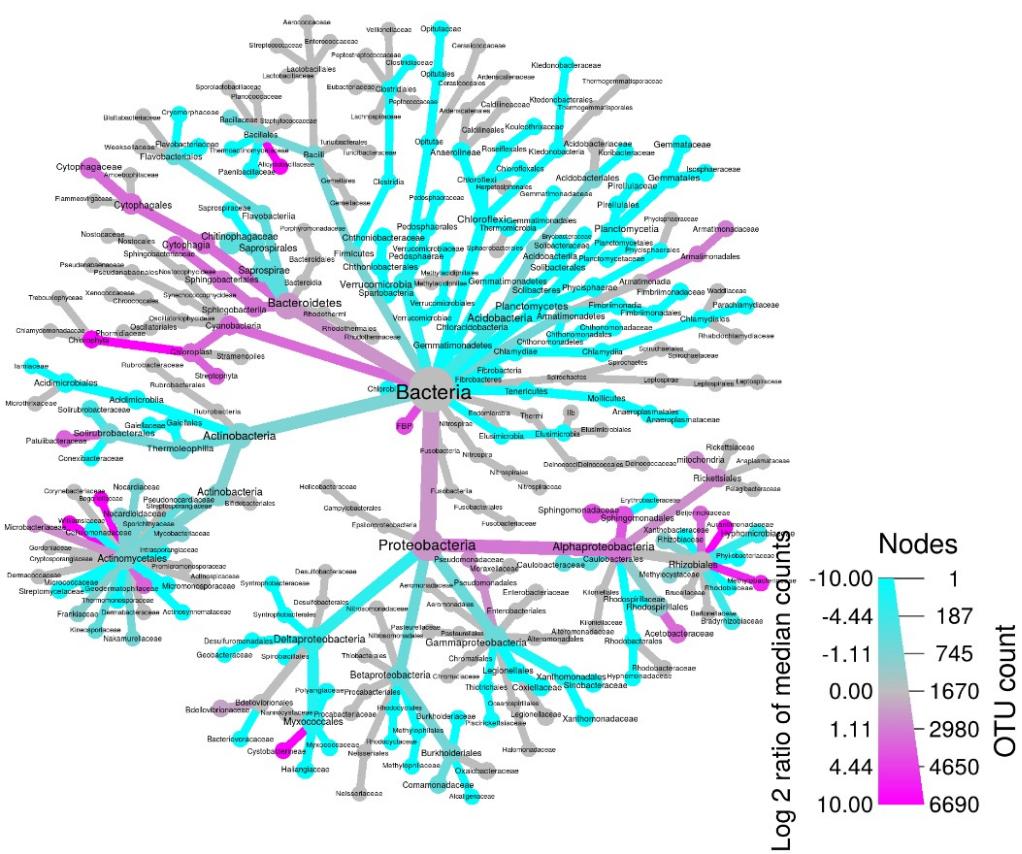
- OTU table
- Taxonomy for each OTU
- (Sample metadata)

OTUid	DS31	DS32	DS34	DS37	DS38	DS39	DS47	DS48	DS50	DS52	DS54	DS55	DS58	DS60	DS61	size	taxonomy
OTU0778	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU1150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU1576	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0517	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0729	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0958	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0730	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	3 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0249	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0788	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	4 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU2041	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU2014	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	5 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0282	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	6 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU1577	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0283	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0726	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0357	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0576	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	14 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0766	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0763	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	16 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0731	0	0	0	1	4	4	0	0	0	0	0	0	0	0	0	0	17 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU1072	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0061	0	0	0	0	0	0	0	0	0	0	0	0	8	1	0	0	23 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0768	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	29 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0636	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU1192	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	54 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0771	2	0	0	0	3	0	0	0	3	0	4	0	0	0	0	0	55 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0440	3	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	56 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0770	0	0	0	0	0	0	27	10	17	0	0	0	0	0	0	0	58 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU2015	0	0	0	0	0	0	0	0	0	0	0	0	54	0	0	0	62 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU0518	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	63 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU1218	0	0	0	0	0	0	0	0	0	0	0	0	39	0	0	0	65 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_
OTU1219	0	0	0	0	0	0	0	0	0	0	0	0	45	0	0	1	65 k_Fungi;p_Ascomycota;o_Archaeohizomycetes;o_Archaeohizomycetales;f_

SampleID	Sample	Type	Location	Temperat	Precipit	pH	N	C	C:N	Free_erg	Total_erg
DB04	Bistorta	Plantroot	Ves	mid	wet	0.718	0.367	0.223	0.414	0.192	0.524
DB05	Bistorta	Plantroot	Ves	mid	wet	0.718	0.367	0.223	0.414	0.192	0.524
DB06	Bistorta	Plantroot	Ves	mid	wet	0.718	0.367	0.223	0.414	0.192	0.524
DB07	Bistorta	Plantroot	Ves	mid	wet	0.718	0.367	0.223	0.414	0.192	0.524
DB08	Bistorta	Plantroot	Arh	low	semi-wet	0.385	0.396	0.602	0.742	0.518	0.395
DB09	Bistorta	Plantroot	Arh	low	semi-wet	0.385	0.396	0.602	0.742	0.518	0.395
DB10	Bistorta	Plantroot	Arh	low	semi-wet	0.385	0.396	0.602	0.742	0.518	0.395
DB14	Bistorta	Plantroot	Ulv	alpine	dry	0.663	0.626	0.485	0.402	0.693	0.554
DB15	Bistorta	Plantroot	Ulv	alpine	dry	0.663	0.626	0.485	0.402	0.693	0.554
DB16	Bistorta	Plantroot	Ulv	alpine	dry	0.663	0.626	0.485	0.402	0.693	0.554
DB20	Bistorta	Plantroot	Lav	alpine	semi-dry	0.744	0.422	0.739	0.774	0.343	0.497
DB21	Bistorta	Plantroot	Lav	alpine	semi-dry	0.744	0.422	0.739	0.774	0.343	0.497
DB22	Bistorta	Plantroot	Lav	alpine	semi-dry	0.744	0.422	0.739	0.774	0.343	0.497
DB26	Bistorta	Plantroot	Air	mid	dry	0.513	0.46	0.629	0.69	0.638	0.487
DB27	Bistorta	Plantroot	Air	mid	dry	0.513	0.46	0.629	0.69	0.638	0.487
DB28	Bistorta	Plantroot	Ram	mid	semi-wet	0.56	0.406	0.686	0.769	0.26	0.572
DB39	Bistorta	Plantroot	Ram	mid	semi-wet	0.56	0.406	0.686	0.769	0.26	0.572
DB40	Bistorta	Plantroot	Ram	mid	semi-wet	0.56	0.406	0.686	0.769	0.26	0.572
DB44	Bistorta	Plantroot	Vik	low	semi-dry	0.194	0.784	0.422	0	0.499	0.52
DB45	Bistorta	Plantroot	Vik	low	semi-dry	0.194	0.784	0.422	0	0.499	0.52
DB46	Bistorta	Plantroot	Fau	low	semi-dry	0.194	0.784	0.422	0	0.499	0.52
DB50	Bistorta	Plantroot	Fau	low	dry	0.56	0.574	0.673	0.597	0.371	0.436
DB51	Bistorta	Plantroot	Fau	low	dry	0.56	0.574	0.673	0.597	0.371	0.436
DB52	Bistorta	Plantroot	Fau	low	dry	0.56	0.574	0.673	0.597	0.371	0.436

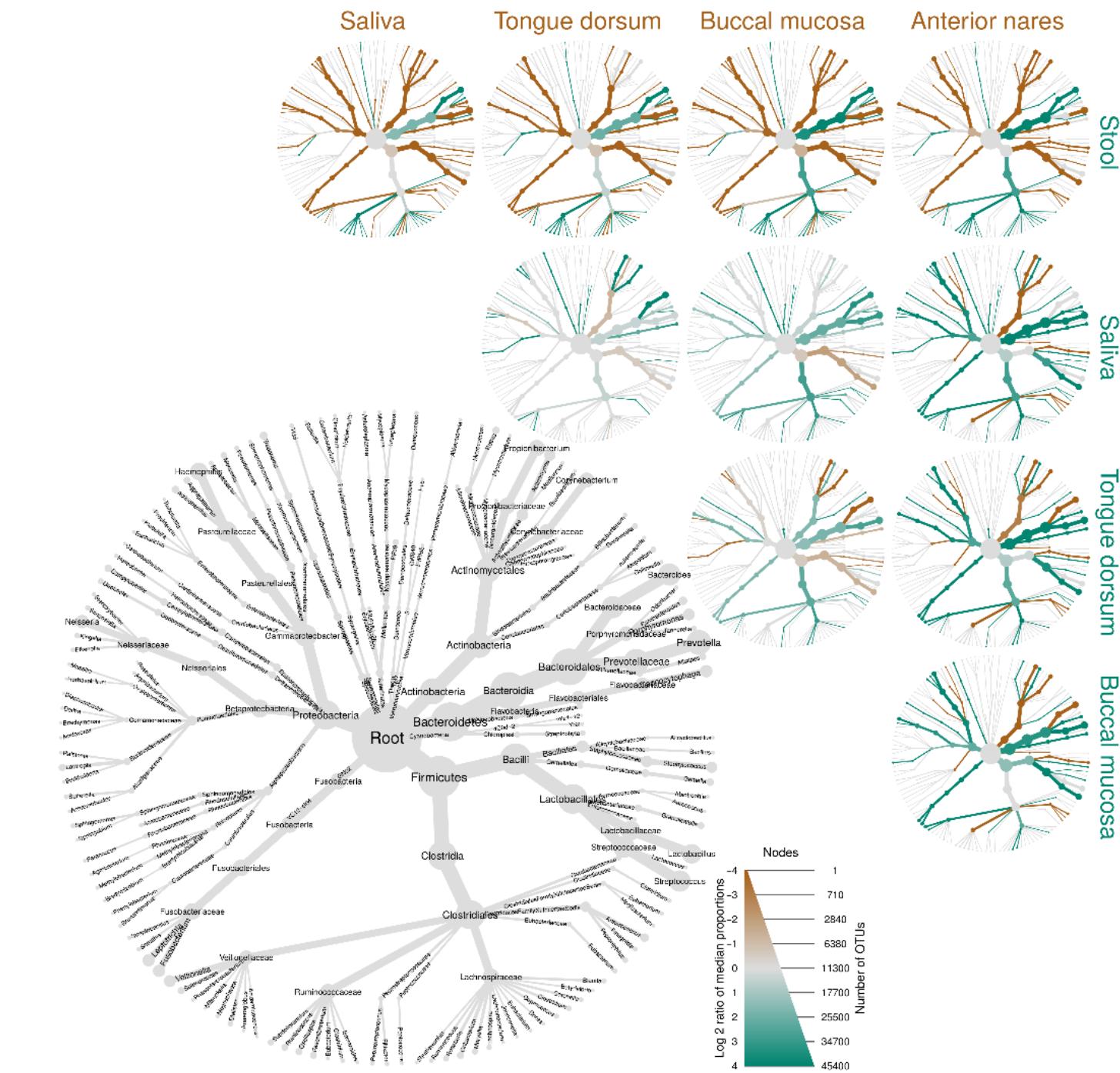
Comparing two treatments

leaf vs root samples



compare_groups()

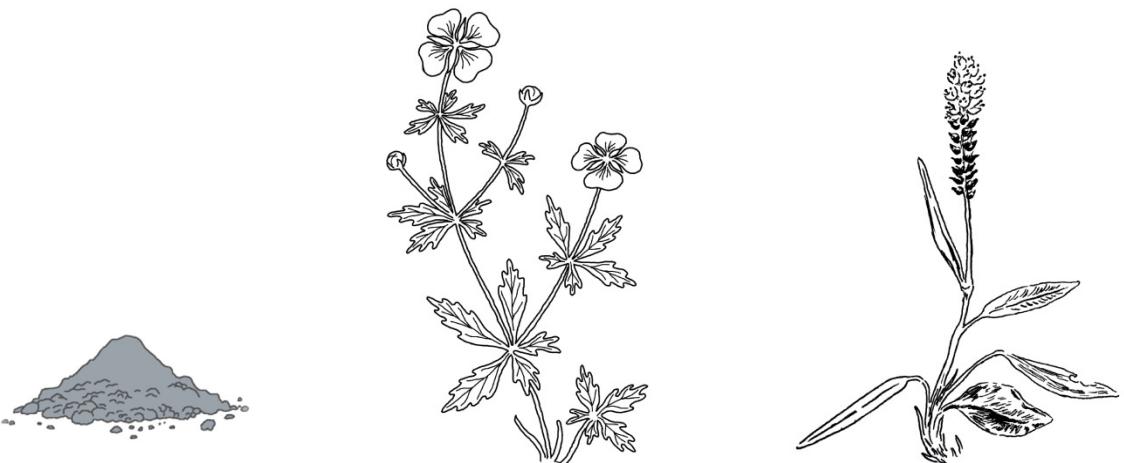
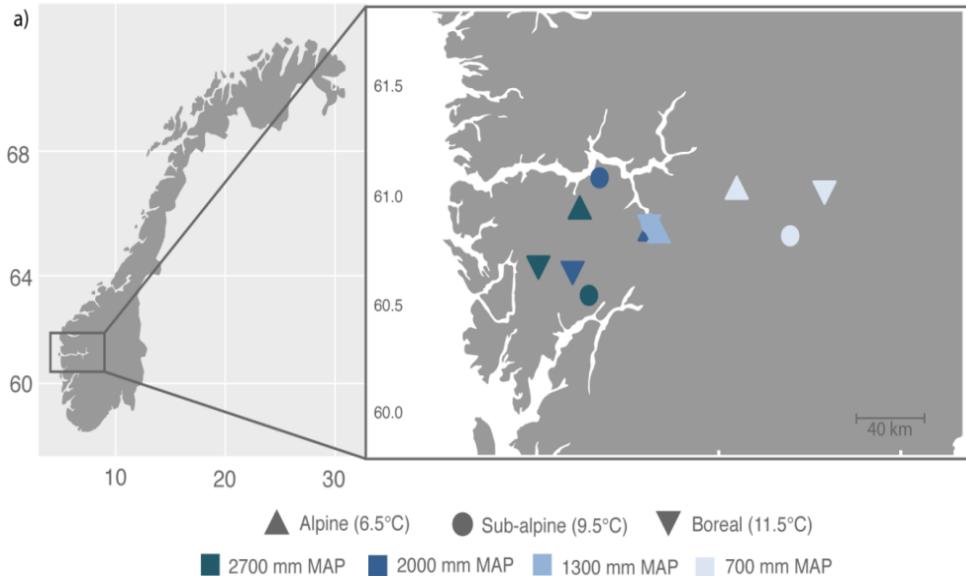
Comparing three or more treatments



= heat tree matrix

Case study (for R)

- Three sample types collected over a climate gradient in seminatural grasslands in West Norway; soil samples and two different plant roots
- The dataset has been reduced and modified – we will only look at how to make heat trees and compare differences in sample type



Soil
«All» fungi

Potentilla erecta
AM or non-
mycorrhizal fungi

Bistorta vivipara
EcM fungi

Data format for *this* metacoder script

- OTU data with samples as columns and OTUs as rows
- Taxonomy (or other hierarchical information) with a “denominator” for each level, here included as a single column; taxonomy:
k_Fungi;p_Basidiomycota;c_Agaricomycetes;o_Agaricales;f_Inocybaceae;g_Inocybe;s_Inocybe_castanea
- Sample data/metadata organized with samples as rows and variables as columns

The dataset

Three files:

1. The script called: `IntroMetacoder.R`
2. The OTU table including taxonomic assignment
`(rarotutable_small.txt)`
3. Sample data containing information for each sample (`sampledata.txt`)

These are available at the github pages if you wish to try this yourself!