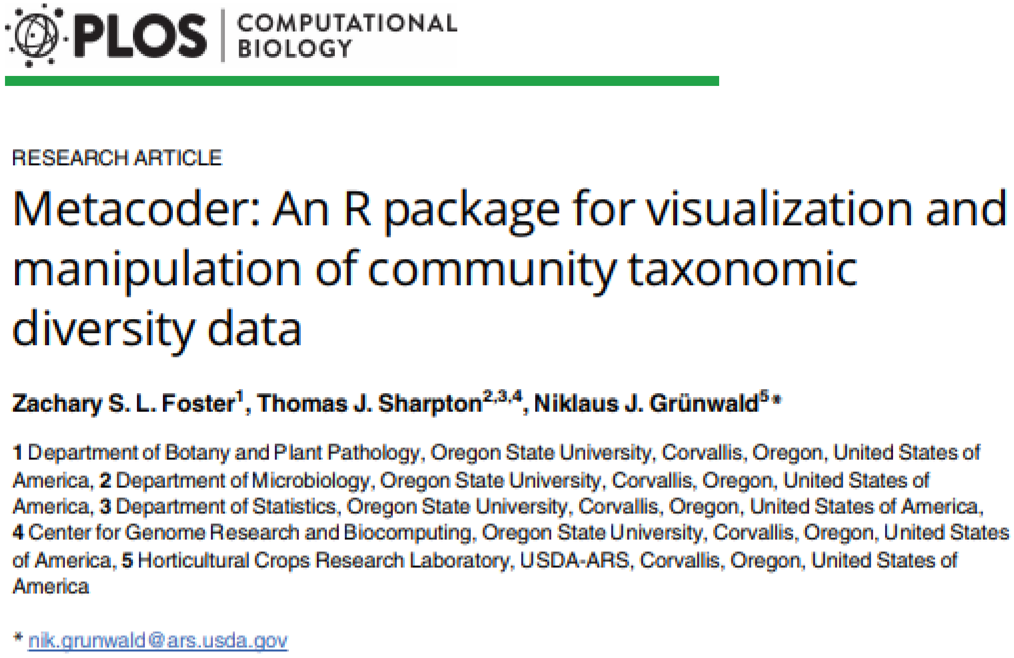
METACODER

Laurent Cauquil - Biopuce 12 avril 2019



PLOS | COMPUTATIONAL BIOLOGY

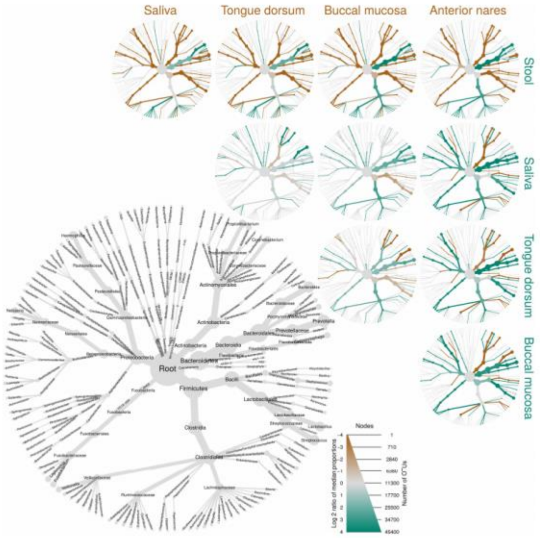
RESEARCH ARTICLE

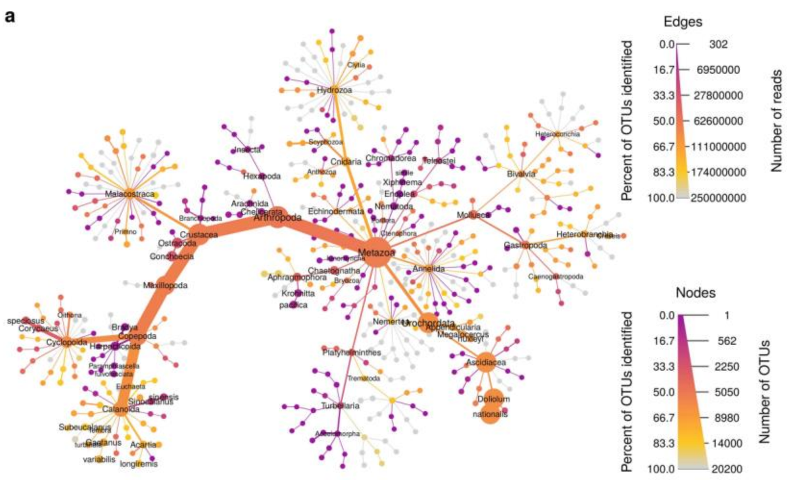
Metacoder: An R package for visualization and manipulation of community taxonomic diversity data

Zachary S. L. Foster[1], **Thomas J. Sharpton**[2,3,4], **Niklaus J. Grünwald**[5]*

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

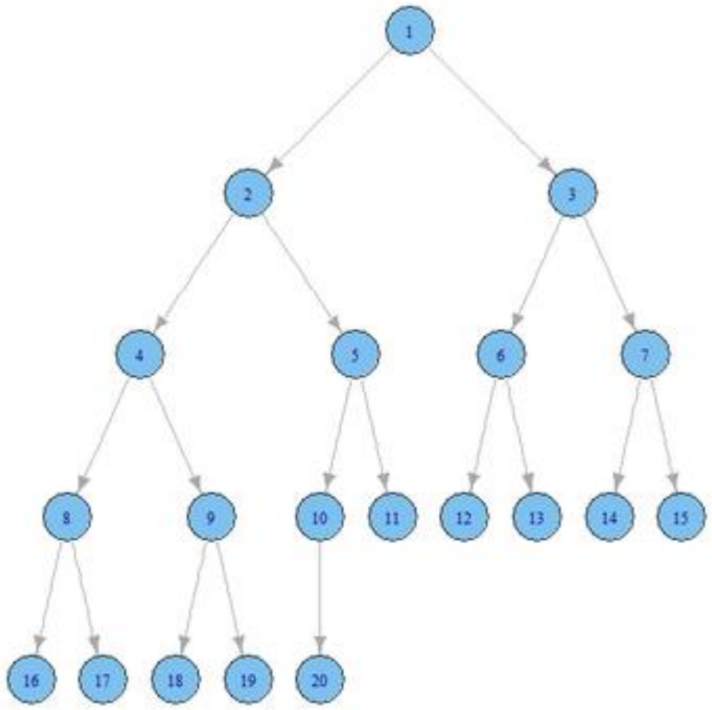




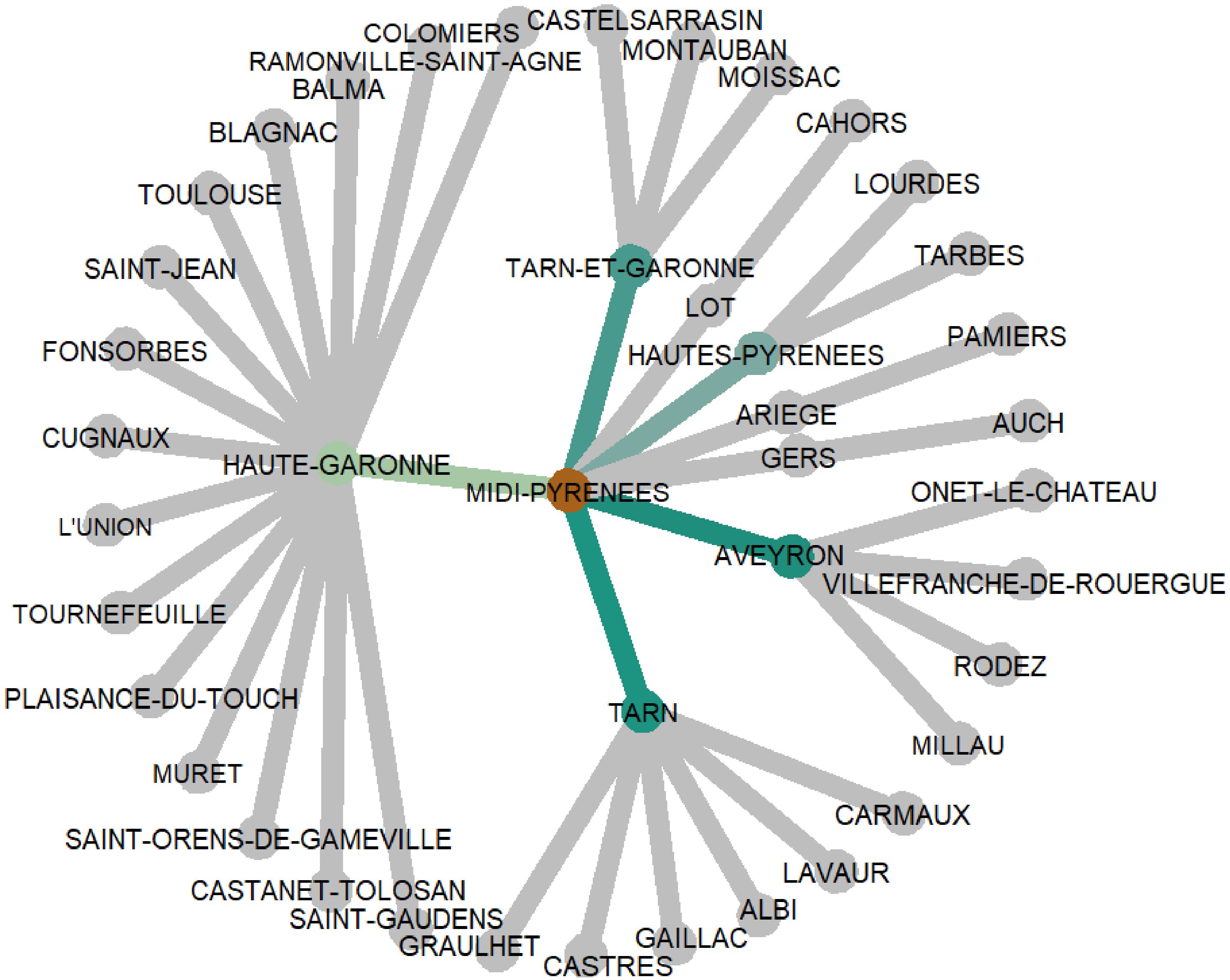
Metacoder can be applied to any dataset that can be organized hierarchically such as:

- community taxonomic diversity data
- gene expression
- geographic data
- …

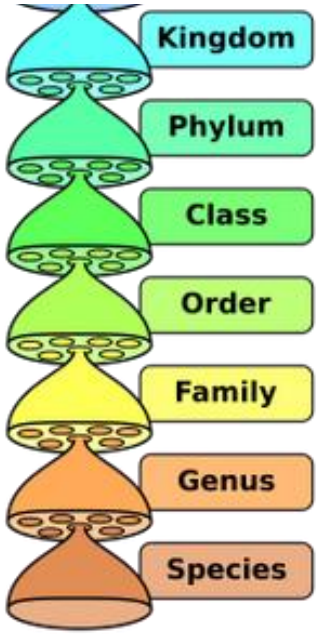
HIERARCHICAL DATA

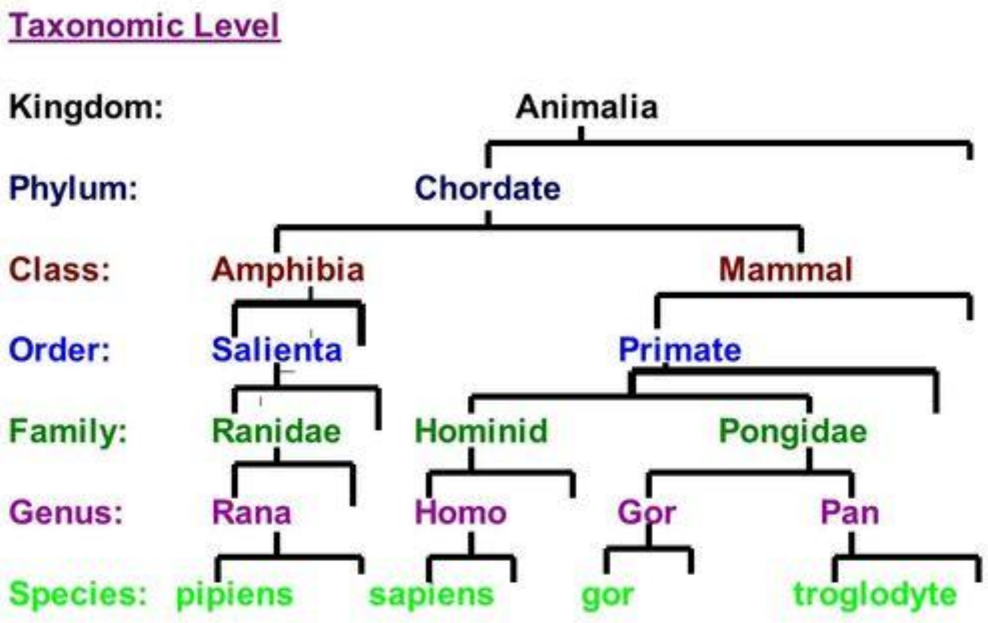


Geographical data with metacoder

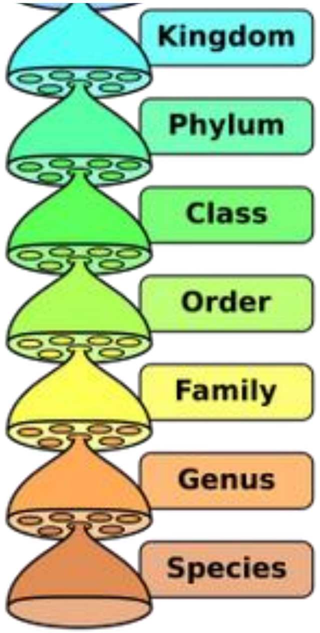


DATASETS WITH HIERARCHICAL COMPONENT: TAXONOMIC DATA





DATASETS WITH HIERARCHICAL COMPONENT: TAXONOMIC DATA



BACTERIA

PROTEOBACTERIA

GAMMAPROTEOBACTERIA

ENTEROBACTERIALES

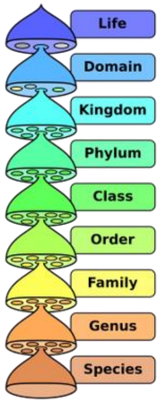
ENTEROBACTERIACEAE

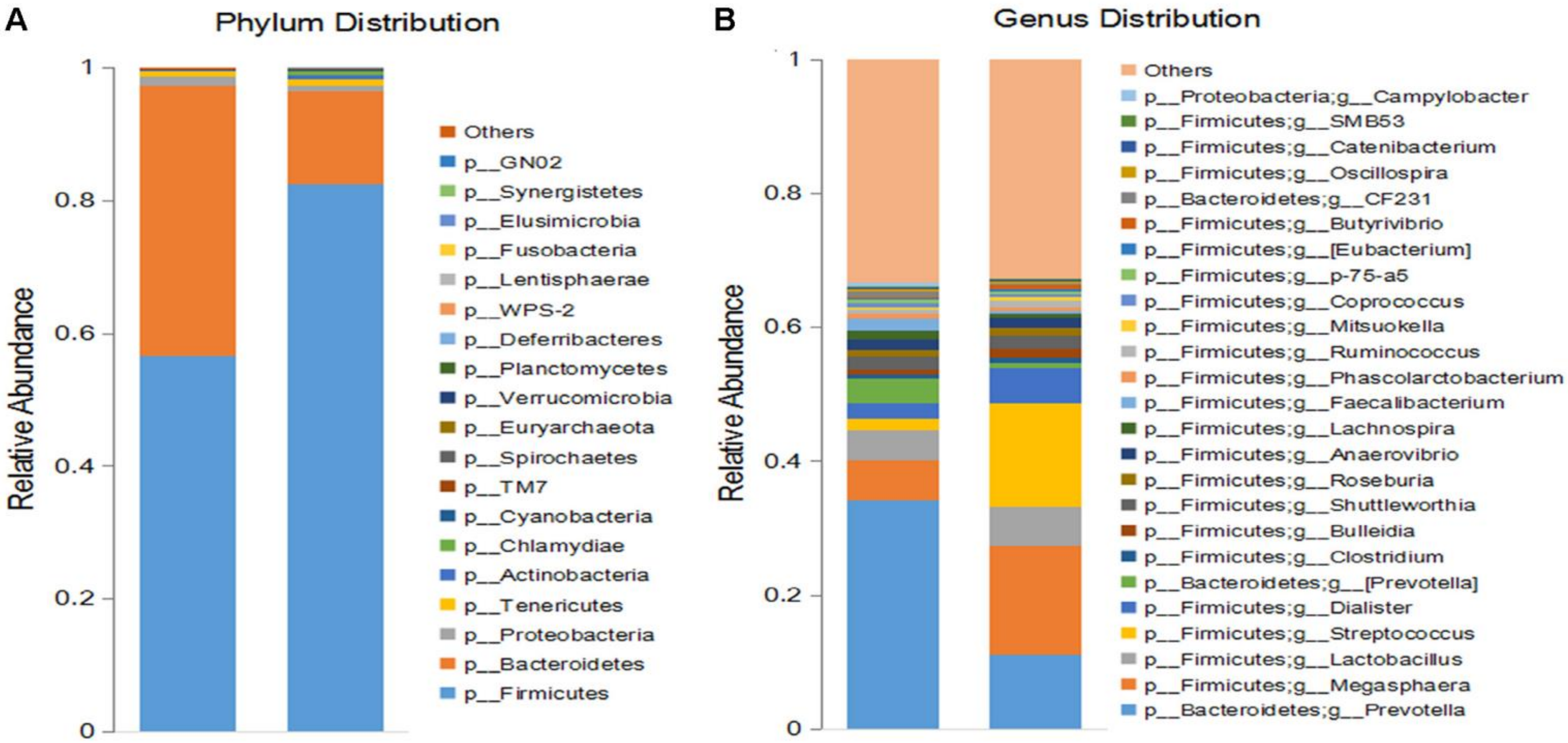
ESCHERICHIA-SHIGELLA

ESCHERICHIA COLI

EXPLORING GUT MICROBIOTA

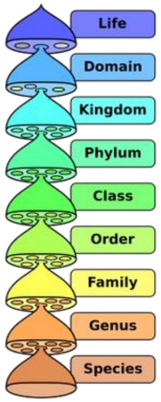
BACTERIA COMPOSITION: BARPLOT

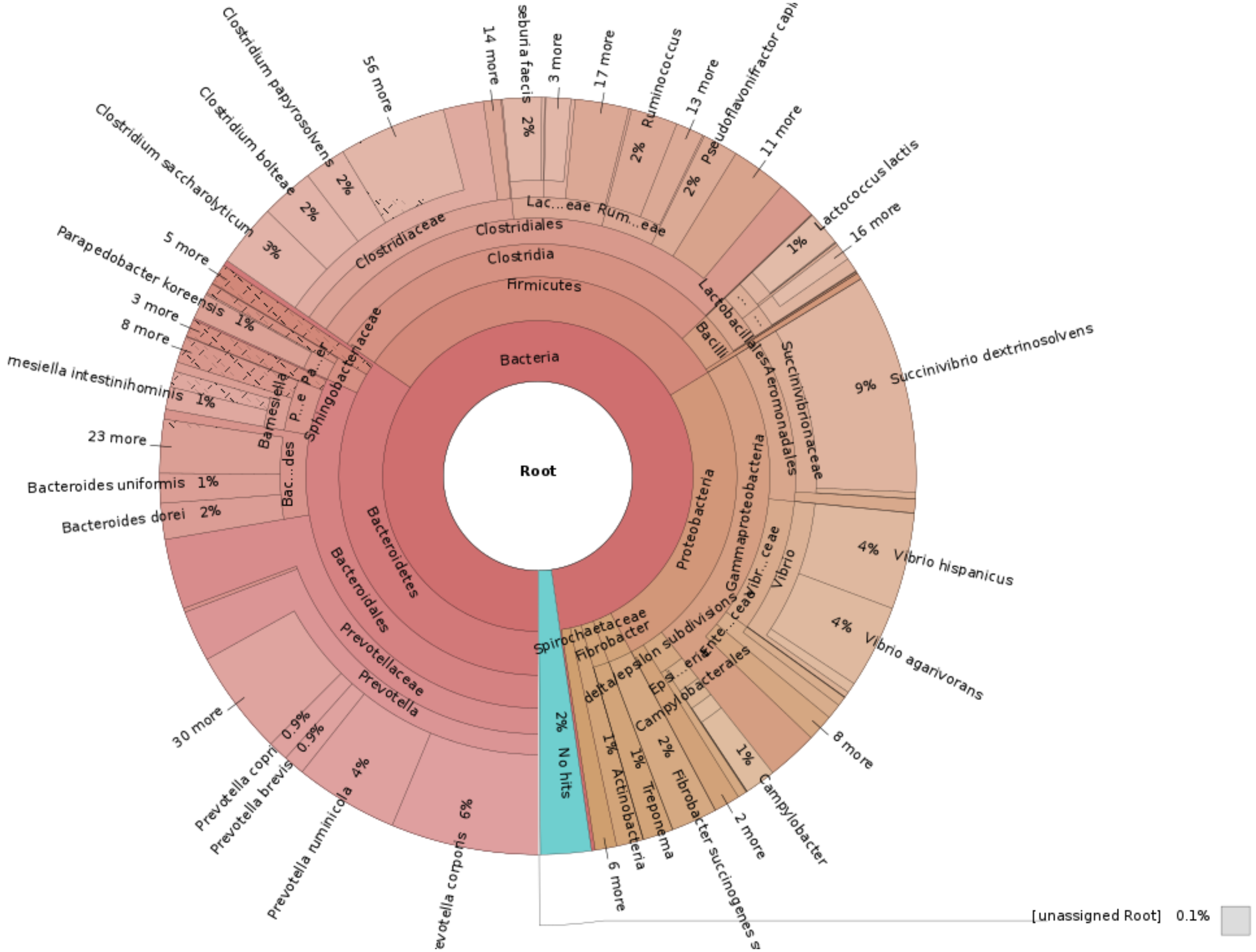




EXPLORING GUT MICROBIOTA

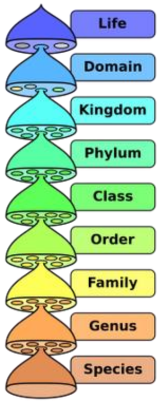
BACTERIA COMPOSITION: SUNBURST, KRONA

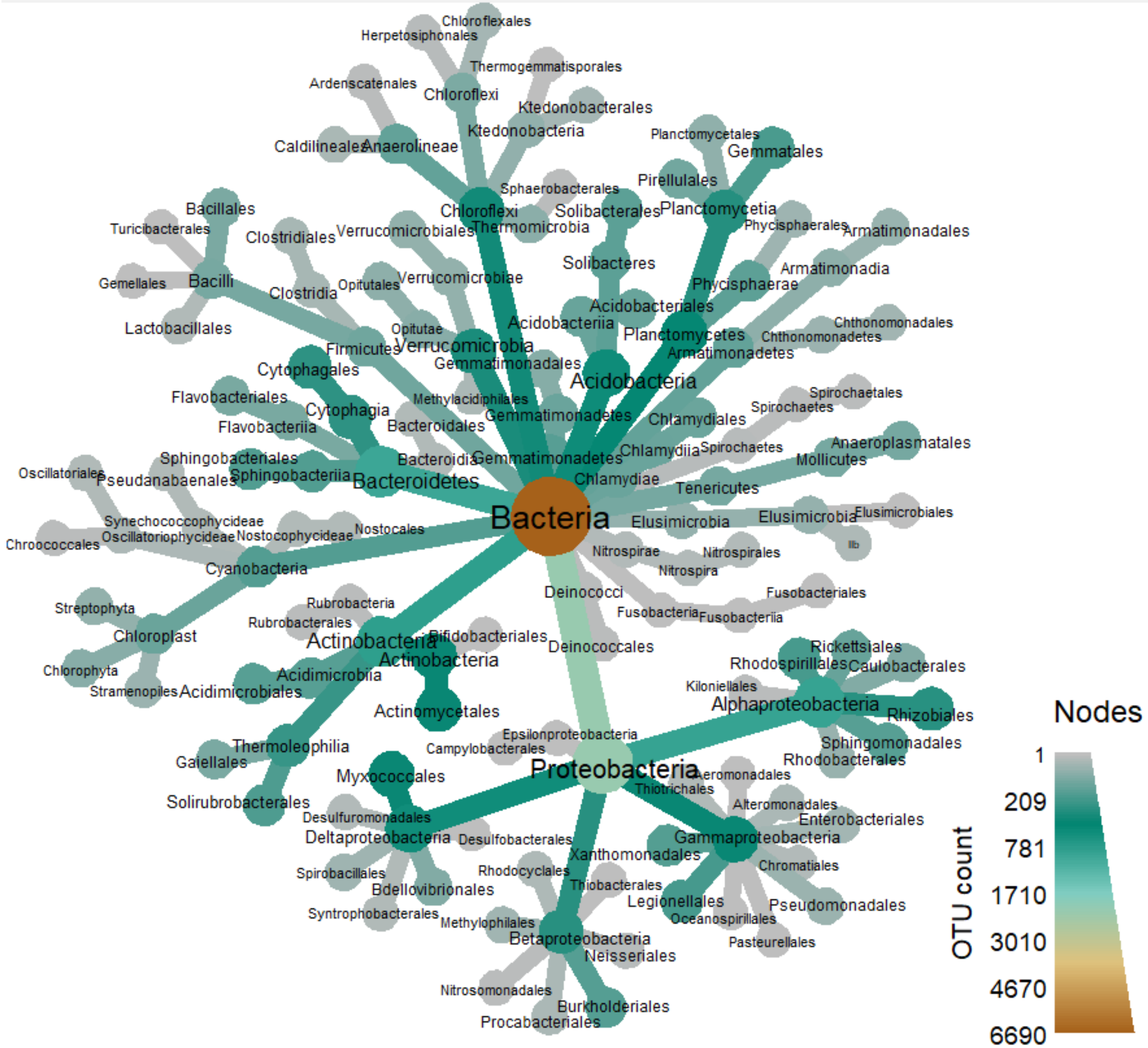




EXPLORING GUT MICROBIOTA

BACTERIA COMPOSITION WITH METACODER





PACKAGE METACODER
(EXTENSION OF TAXA PACKAGE)

THE TAXA PACKAGE IS INTENDED TO:

- PROVIDE A SET OF CLASSES TO STORE TAXONOMIC DATA AND ANY USER-SPECIFIC DATA ASSOCIATED WITH IT
- PROVIDE FUNCTIONS TO CONVERT COMMONLY USED FORMATS TO THESE CLASSES
- PROVIDE GENERALLY USEFUL FUNCTIONALITY, SUCH AS FILTERING AND MAPPING FUNCTIONS

→ **`taxmap class object`**

PACKAGE METACODER
(EXTENSION OF TAXA PACKAGE)

FROM A *taxmap* OBJECT (TAXA PACKAGE)

- R6 CLASS OBJECT TO HOLD TAXONOMIC AND ASSOCIATED DATA
- PARSING SPECIFIC FILE FORMATS USED IN METAGENOMICS RESEARCH (MOTHER, QIIME, PHYLOSEQ, GREENGENES, RDP, SILVA)
- SUBSETTING COMPLEX HIERARCHICAL DATA SETS USING DPLYR DATA-MANIPULATION PHILOSOPHY
- PLOTTING FUNCTION ENABLES QUANTITATIVE REPRESENTATION OF UP TO 4 ARBITRARY STATISTICS SIMULTANEOUSLY IN A TREE FORMAT BY MAPPING STATISTICS TO THE COLOR AND SIZE OF TREE NODES AND EDGES

PACKAGE METACODER
(EXTENSION OF TAXA PACKAGE)

HOW IT WORKS:

- `parse_tax_data`(): CREATE TAX_MAP OBJECT
- `heat_tree`(): TO VISUALIZE TREE

PACKAGE METACODER
(EXTENSION OF TAXA PACKAGE)

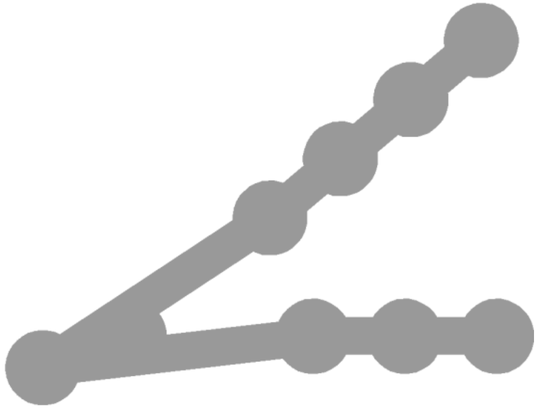
HOW IT WORKS: A SIMPLE CASE

```
x <- c("Mammalia;Theria;Metatheria;Diplodontia;Macropodiformes",
       "Mammalia;Theria;Eutheria;Primates;Haloplorrhini;Simiiformes")

obj <- parse_tax_data(x, class_sep = ";")

heat_tree(obj)
```

PACKAGE METACODER
(EXTENSION OF TAXA PACKAGE)

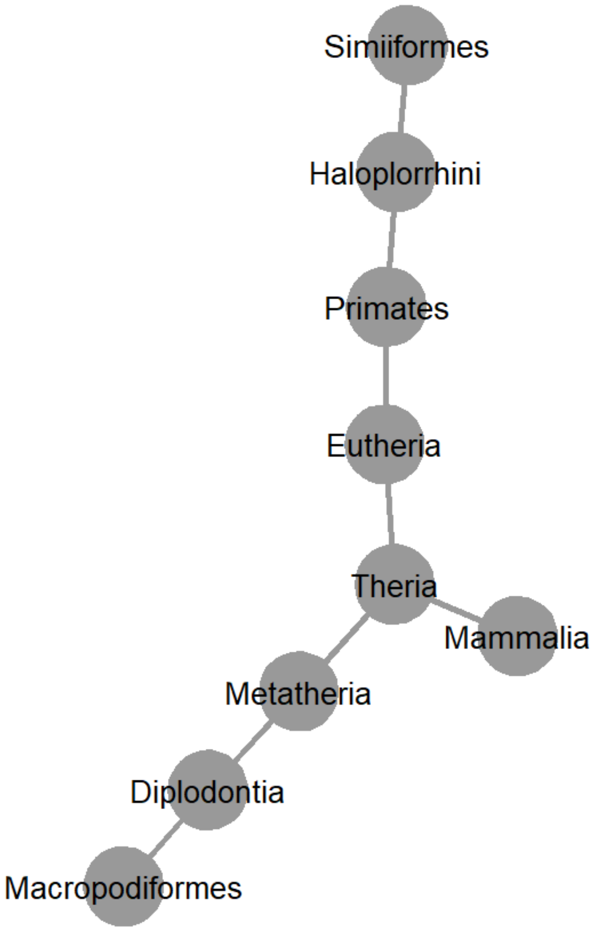
HOW IT WORKS: A SIMPLE CASE

```
x <- c("Mammalia;Theria;Metatheria;Diplodontia;Macropodiformes",
       "Mammalia;Theria;Eutheria;Primates;Haloplorrhini;Simiiformes")

obj <- parse_tax_data(x, class_sep = ";")

heat_tree(obj)
```



PACKAGE METACODER
(EXTENSION OF TAXA PACKAGE)

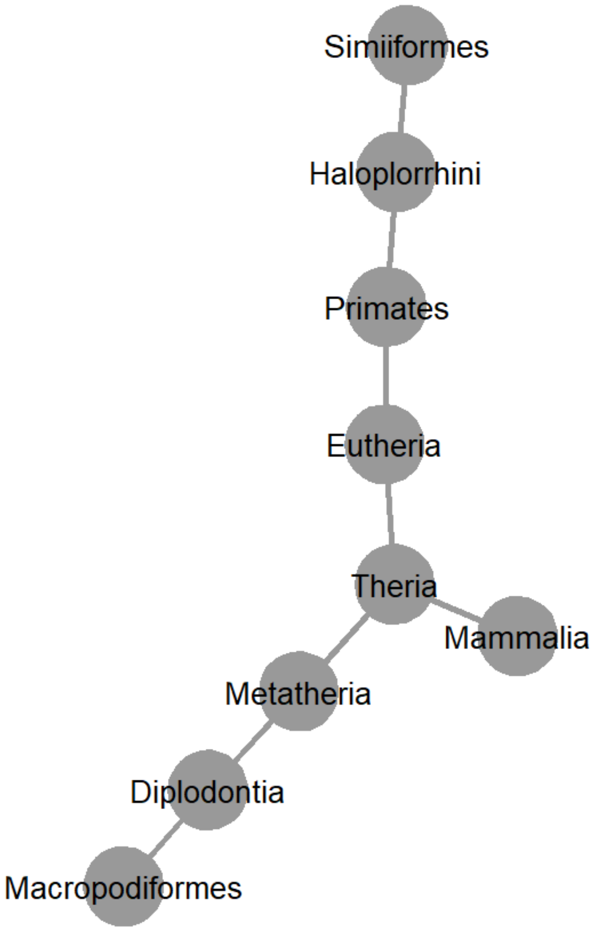
HOW IT WORKS: A SIMPLE CASE



PACKAGE METACODER
(EXTENSION OF TAXA PACKAGE)

HOW IT WORKS: A SIMPLE CASE

```
heat_tree(obj,
          node_size_range = c(0.06, 0.06),
          node_label = taxon_names,
          edge_size_range = c(0.005, 0.005),
          initial_layout = "reingold-tilford",
          layout = "davidson-harel")
```



LAYOUT = IGRAPH PARAMETERS

Package metacoder

(extension of taxa package)

How it works: a simple case

```
> obj
<Taxmap>
  9 taxa: b. Mammalia, c. Theria ... i. Haloplorrhini, j. Simiiformes
  9 edges: NA->b, b->c, c->d, c->e, d->f, e->g, f->h, g->i, i->j
  1 data sets:
    tax_data: a named vector of 'character' with 2 items
       h. Mammalia;Theria;[truncated] ... j. Mammalia;Theria;[truncated]
  0 functions:
```

- *Taxmap* object
- 9 different taxa
- 9 edges

Package metacoder

(extension of taxa package)

How it works: a simple case

```
> obj
<Taxmap>
  9 taxa: b. Mammalia, c. Theria ... i. Haloplorrhini, j. Simiiformes
  9 edges: NA->b, b->c, c->d, c->e, d->f, e->g, f->h, g->i, i->j
  1 data sets:
    tax_data: a named vector of 'character' with 2 items
       h. Mammalia;Theria;[truncated] ... j. Mammalia;Theria;[truncated]
  0 functions:
```

- *Taxmap* object
- 9 different taxa
- 9 edges

```
> length(obj)
[1] 65
```

Package metacoder

(extension of taxa package)

How it works: a simple case

```
> obj
<Taxmap>
  9 taxa: b. Mammalia, c. Theria ... i. Haloplorrhini, j. Simiiformes
  9 edges: NA->b, b->c, c->d, c->e, d->f, e->g, f->h, g->i, i->j
  1 data sets:
    tax_data: a named vector of 'character' with 2 items
       h. Mammalia;Theria;[truncated] ... j. Mammalia;Theria;[truncated]
  0 functions:
```

- *taxmap* object
- 9 different taxa
- 9 edges

```
> str(obj)
Classes 'Taxmap', 'Taxonomy', 'R6' <Taxmap>
  Inherits from: <Taxonomy>
  Public:
    all_names: function (tables = TRUE, funcs = TRUE, others = TRUE,
builtin_funcs = TRUE,
    arrange_obs: function (data, ..., target = NULL)
    arrange_taxa: function (...)
    branches: function (subset = NULL, value = "taxon_indexes")
    classifications: function (value = "taxon_names", sep = ";")
    clone: function (deep = FALSE)
    data: list
    data_used: function (...)
    …
    …
```

Package metacoder
(extension of taxa package)

How it works: a simple case

```
> obj
<Taxmap>
  9 taxa: b. Mammalia, c. Theria ... i. Haloplorrhini, j. Simiiformes
  9 edges: NA->b, b->c, c->d, c->e, d->f, e->g, f->h, g->i, i->j
  1 data sets:
    tax_data: a named vector of 'character' with 2 items
       h. Mammalia;Theria;[truncated] ... j. Mammalia;Theria;[truncated]
  0 functions:
```

- *TAXMAP* OBJECT
- 9 DIFFERENT TAXA
- 9 EDGES

```
> obj$data$tax_data
                                                            h 
     "Mammalia;Theria;Metatheria;Diplodontia;Macropodiformes" 
                                                            j 
"Mammalia;Theria;Eutheria;Primates;Haloplorrhini;Simiiformes" 
```

Package metacoder

(extension of taxa package)

How it works: a simple case

```
> obj
<Taxmap>
  9 taxa: b. Mammalia, c. Theria ... i. Haloplorrhini, j. Simiiformes
  9 edges: NA->b, b->c, c->d, c->e, d->f, e->g, f->h, g->i, i->j
  1 data sets:
    tax_data: a named vector of 'character' with 2 items
       h. Mammalia;Theria;[truncated] ... j. Mammalia;Theria;[truncated]
  0 functions:
```

- *TAXMAP* OBJECT
- 9 DIFFERENT TAXA
- 9 EDGES

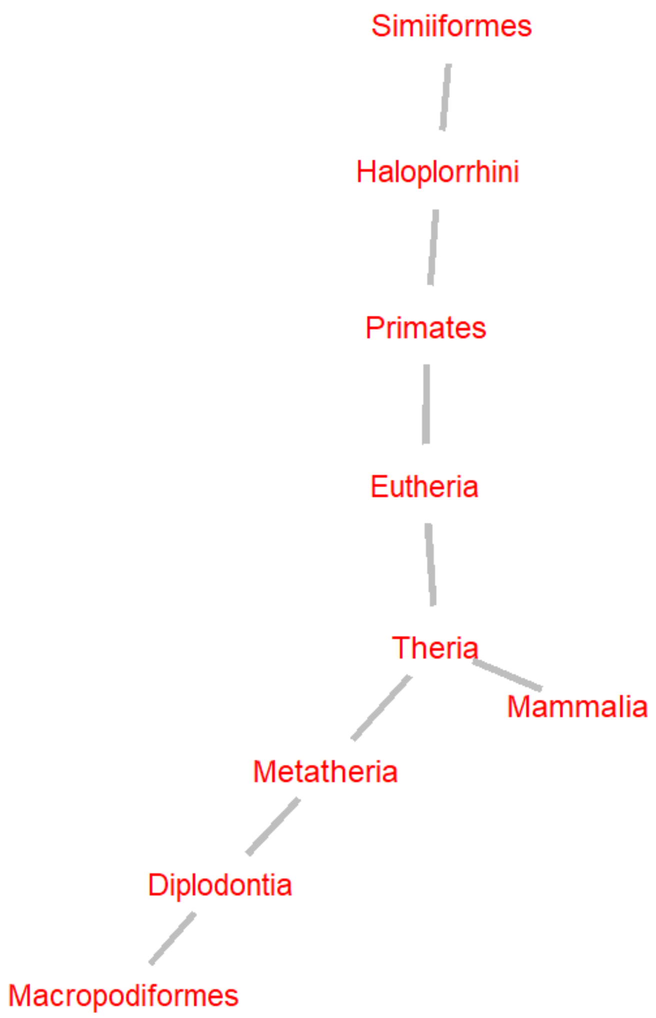
```
> obj$edge_list
  from to
1 <NA>   b
2    b   c
3    c   d
4    c   e
5    d   f
6    e   g
7    f   h
8    g   i
9    i   j
```

PACKAGE METACODER
(EXTENSION OF TAXA PACKAGE)

HOW IT WORKS: A SIMPLE CASE + <u>TUNING</u>

```
heat_tree(obj,
          node_size_range = c(0.06, 0.06),
          node_label = taxon_names,
          node_label_color = "red",
          edge_size_range = c(0.005, 0.005),
          node_color = "white",
          edge_color = "gray",
          initial_layout = "reingold-tilford",
          layout = "davidson-harel")
```

(~ 70 PARAMETERS)



PACKAGE METACODER
(EXTENSION OF TAXA PACKAGE)

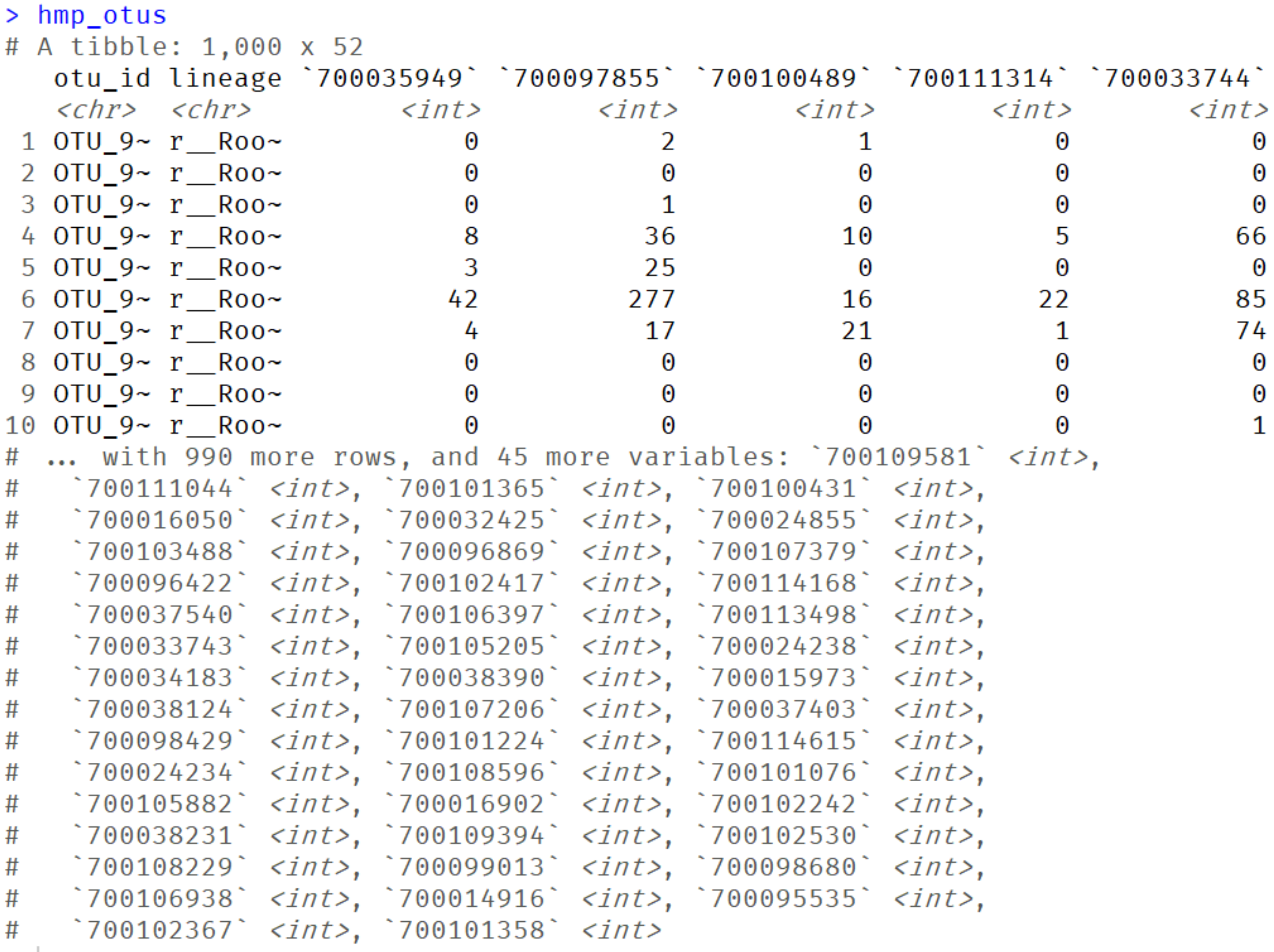
HOW IT WORKS: A REAL CASE

THE HUMAN MICROBIOME PROJECT (SUBSET):

- ➢ 50 SAMPLES FROM HUMAN
- ➢ 1000 OTU (CLUSTERS) IDENTIFIED
- ➢ SAMPLES INFORMATION:
  - SEX: MALE, FEMALE
  - BODY SITE: SALIVA, SKIN, STOOL, THROAT, NOSE

PACKAGE METACODER
(EXTENSION OF TAXA PACKAGE)

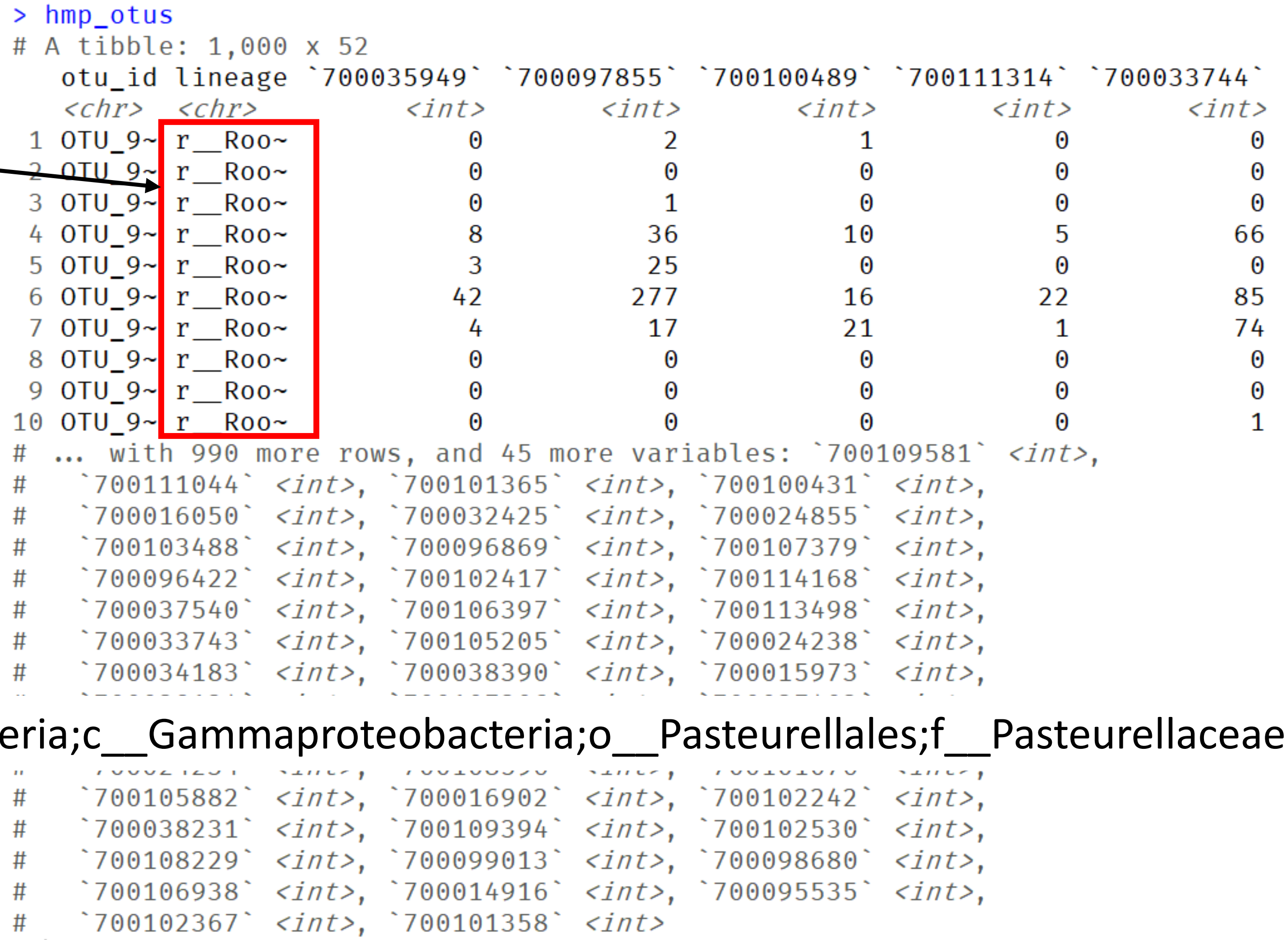
CREATING *TAXMAP* OBJECT: ABUNDANCE MATRIX



```
> hmp_otus
# A tibble: 1,000 x 52
   otu_id lineage `700035949` `700097855` `700100489` `700111314` `700033744`
   <chr>  <chr>         <int>       <int>       <int>       <int>       <int>
 1 OTU_9~ r__Roo~           0           2           1           0           0
 2 OTU_9~ r__Roo~           0           0           0           0           0
 3 OTU_9~ r__Roo~           0           1           0           0           0
 4 OTU_9~ r__Roo~           8          36          10           5          66
 5 OTU_9~ r__Roo~           3          25           0           0           0
 6 OTU_9~ r__Roo~          42         277          16          22          85
 7 OTU_9~ r__Roo~           4          17          21           1          74
 8 OTU_9~ r__Roo~           0           0           0           0           0
 9 OTU_9~ r__Roo~           0           0           0           0           0
10 OTU_9~ r__Roo~           0           0           0           0           1
# ... with 990 more rows, and 45 more variables: `700109581` <int>,
#   `700111044` <int>, `700101365` <int>, `700100431` <int>,
#   `700016050` <int>, `700032425` <int>, `700024855` <int>,
#   `700103488` <int>, `700096869` <int>, `700107379` <int>,
#   `700096422` <int>, `700102417` <int>, `700114168` <int>,
#   `700037540` <int>, `700106397` <int>, `700113498` <int>,
#   `700033743` <int>, `700105205` <int>, `700024238` <int>,
#   `700034183` <int>, `700038390` <int>, `700015973` <int>,
#   `700038124` <int>, `700107206` <int>, `700037403` <int>,
#   `700098429` <int>, `700101224` <int>, `700114615` <int>,
#   `700024234` <int>, `700108596` <int>, `700101076` <int>,
#   `700105882` <int>, `700016902` <int>, `700102242` <int>,
#   `700038231` <int>, `700109394` <int>, `700102530` <int>,
#   `700108229` <int>, `700099013` <int>, `700098680` <int>,
#   `700106938` <int>, `700014916` <int>, `700095535` <int>,
#   `700102367` <int>, `700101358` <int>
```

PACKAGE METACODER
(EXTENSION OF TAXA PACKAGE)

CREATING *TAXMAP* OBJECT: ABUNDANCE MATRIX



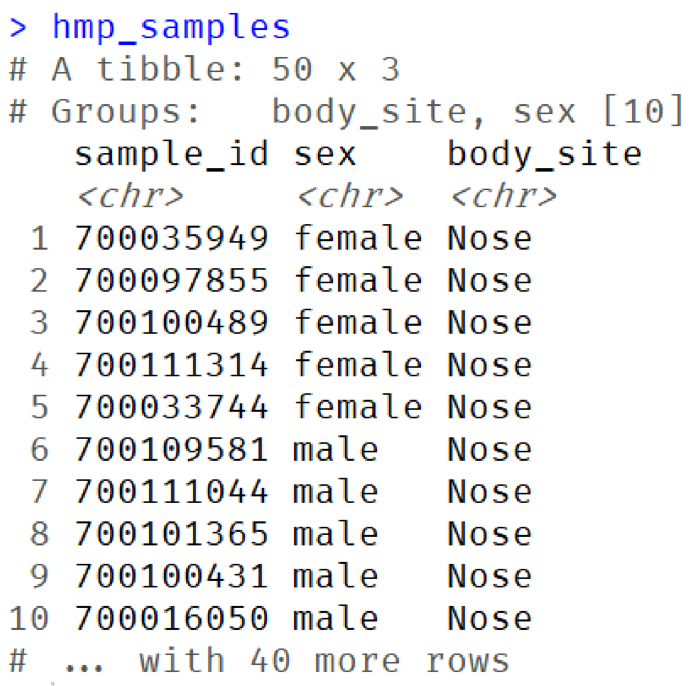
Taxonomy

```
> hmp_otus
# A tibble: 1,000 x 52
   otu_id lineage `700035949` `700097855` `700100489` `700111314` `700033744`
   <chr>  <chr>         <int>       <int>       <int>       <int>       <int>
 1 OTU_9~ r__Roo~           0           2           1           0           0
 2 OTU_9~ r__Roo~           0           0           0           0           0
 3 OTU_9~ r__Roo~           0           1           0           0           0
 4 OTU_9~ r__Roo~           8          36          10           5          66
 5 OTU_9~ r__Roo~           3          25           0           0           0
 6 OTU_9~ r__Roo~          42         277          16          22          85
 7 OTU_9~ r__Roo~           4          17          21           1          74
 8 OTU_9~ r__Roo~           0           0           0           0           0
 9 OTU_9~ r__Roo~           0           0           0           0           0
10 OTU_9~ r__Roo~           0           0           0           0           1
# ... with 990 more rows, and 45 more variables: `700109581` <int>,
#   `700111044` <int>, `700101365` <int>, `700100431` <int>,
#   `700016050` <int>, `700032425` <int>, `700024855` <int>,
#   `700103488` <int>, `700096869` <int>, `700107379` <int>,
#   `700096422` <int>, `700102417` <int>, `700114168` <int>,
#   `700037540` <int>, `700106397` <int>, `700113498` <int>,
#   `700033743` <int>, `700105205` <int>, `700024238` <int>,
#   `700034183` <int>, `700038390` <int>, `700015973` <int>,
#   `700105882` <int>, `700016902` <int>, `700102242` <int>,
#   `700038231` <int>, `700109394` <int>, `700102530` <int>,
#   `700108229` <int>, `700099013` <int>, `700098680` <int>,
#   `700106938` <int>, `700014916` <int>, `700095535` <int>,
#   `700102367` <int>, `700101358` <int>
```

r__Root;p__Proteobacteria;c__Gammaproteobacteria;o__Pasteurellales;f__Pasteurellaceae;g__Haemophilus

Package metacoder
(extension of taxa package)

Creating *taxmap* object: sample data



```
> hmp_samples
# A tibble: 50 x 3
# Groups:   body_site, sex [10]
   sample_id sex    body_site
   <chr>     <chr>  <chr>    
 1 700035949 female Nose     
 2 700097855 female Nose     
 3 700100489 female Nose     
 4 700111314 female Nose     
 5 700033744 female Nose     
 6 700109581 male   Nose     
 7 700111044 male   Nose     
 8 700101365 male   Nose     
 9 700100431 male   Nose     
10 700016050 male   Nose     
# ... with 40 more rows
```

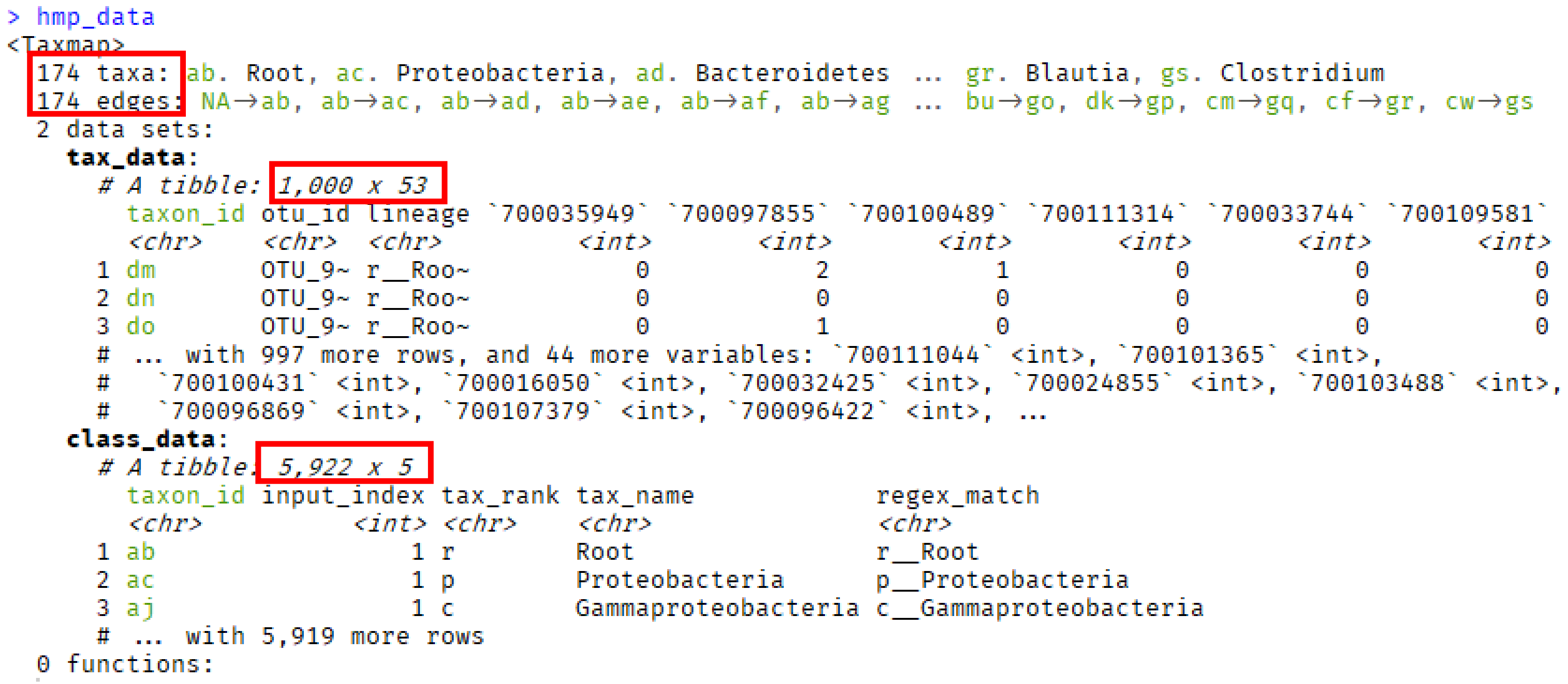
PACKAGE METACODER
(EXTENSION OF TAXA PACKAGE)

CREATING *TAXMAP* OBJECT:

```
hmp_data <- parse_tax_data(hmp_otus,
                      # the column that contains taxonomic information
                         class_cols = "lineage",
                      # The character used to separate taxa in the classification
                         class_sep = ";",
                      # Regex identifying where the data for each taxon is
                         class_regex = "^(.+)__(.+)$",
                      # A key describing each regex capture group
                         class_key = c(tax_rank = "info",
                                    tax_name = "taxon_name"))
```

PACKAGE METACODER
(EXTENSION OF TAXA PACKAGE)

CREATING *TAXMAP* OBJECT:



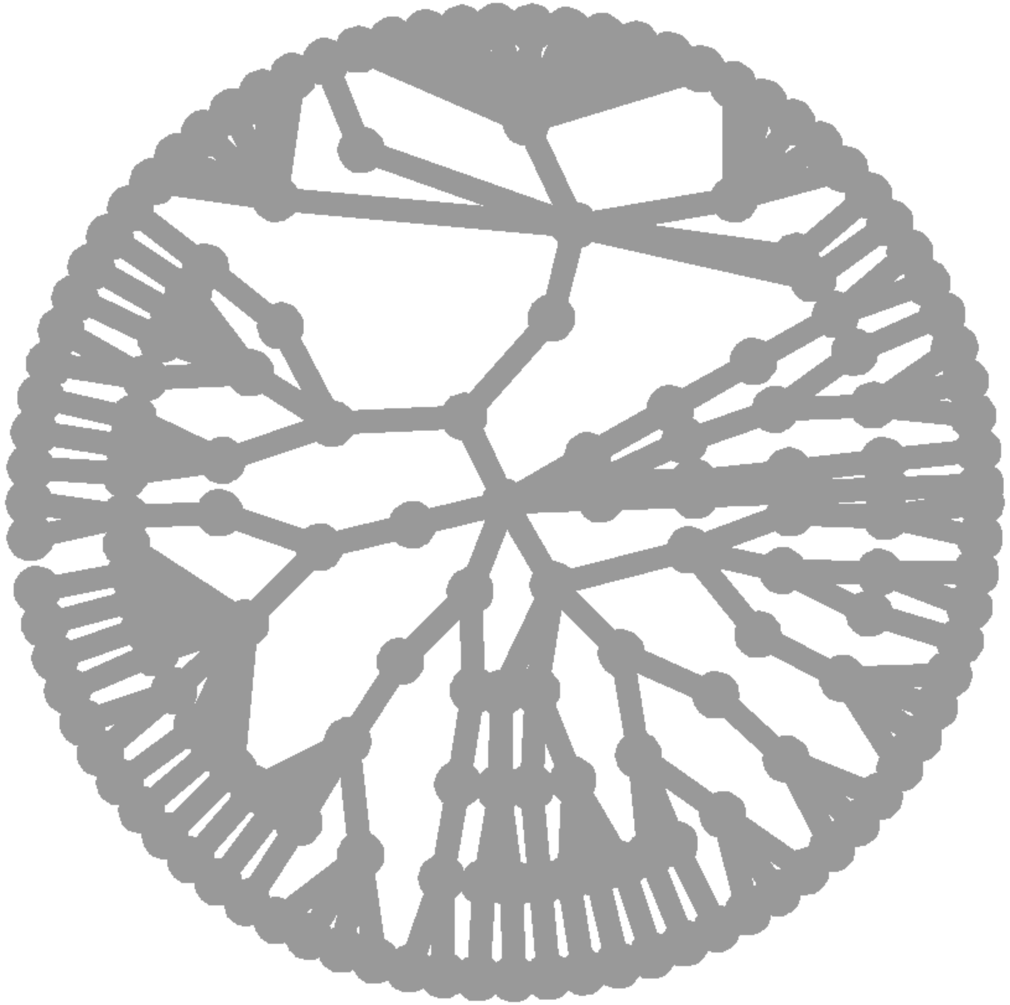
```
> hmp_data
<Taxmap>
  174 taxa: ab. Root, ac. Proteobacteria, ad. Bacteroidetes ... gr. Blautia, gs. Clostridium
  174 edges: NA->ab, ab->ac, ab->ad, ab->ae, ab->af, ab->ag ... bu->go, dk->gp, cm->gq, cf->gr, cw->gs
  2 data sets:
    tax_data:
      # A tibble: 1,000 x 53
        taxon_id otu_id lineage `700035949` `700097855` `700100489` `700111314` `700033744` `700109581`
        <chr>    <chr>  <chr>         <int>       <int>       <int>       <int>       <int>       <int>
      1 dm       OTU_9~ r__Roo~           0           2           1           0           0           0
      2 dn       OTU_9~ r__Roo~           0           0           0           0           0           0
      3 do       OTU_9~ r__Roo~           0           1           0           0           0           0
      # ... with 997 more rows, and 44 more variables: `700111044` <int>, `700101365` <int>,
      #   `700100431` <int>, `700016050` <int>, `700032425` <int>, `700024855` <int>, `700103488` <int>,
      #   `700096869` <int>, `700107379` <int>, `700096422` <int>, ...
    class_data:
      # A tibble: 5,922 x 5
        taxon_id input_index tax_rank tax_name            regex_match
        <chr>          <int> <chr>    <chr>               <chr>
      1 ab                 1 r        Root                r__Root
      2 ac                 1 p        Proteobacteria      p__Proteobacteria
      3 aj                 1 c        Gammaproteobacteria c__Gammaproteobacteria
      # ... with 5,919 more rows
  0 functions:
```

Package metacoder

(extension of taxa package)

Visualizing *taxmap* object:

```
heat_tree(hmp_data)
```

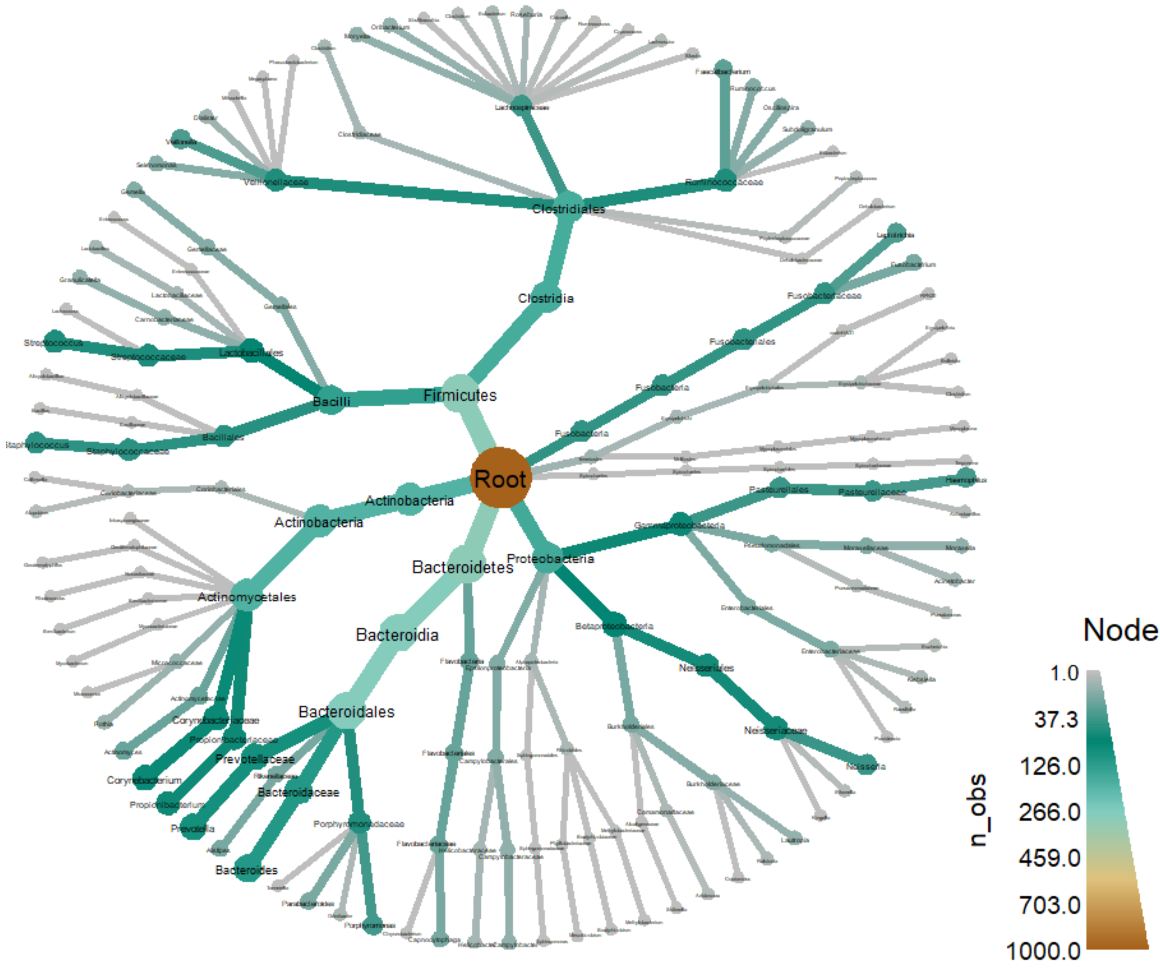


PACKAGE METACODER

(EXTENSION OF TAXA PACKAGE)

VISUALIZING *TAXMAP* OBJECT:

```
heat_tree(hmp_data,
          node_label = taxon_names,
          node_size = n_obs,
          node_color = n_obs)
```



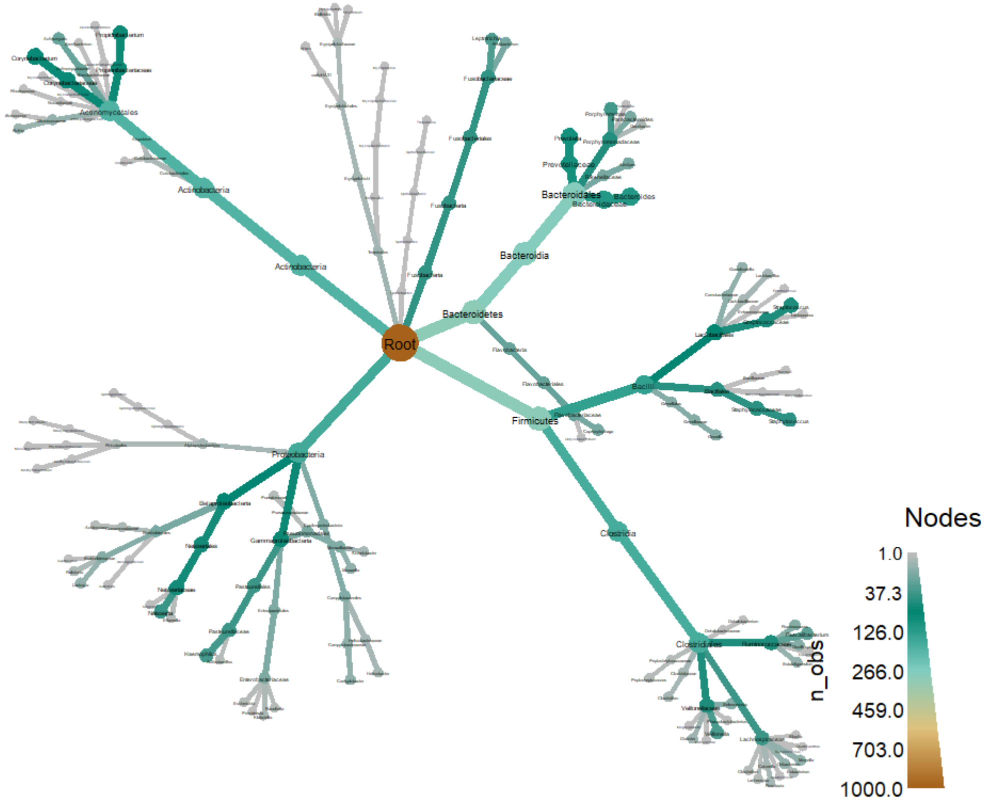
Package metacoder

(extension of taxa package)

Visualizing *taxmap* object:

```
heat_tree(hmp_data,
            node_label = taxon_names,
            node_size = n_obs,
            node_color = n_obs,
            layout = "fr",
            output_file = "plot_example.pdf")
```

SAVE THE PLOT USING GGSAVE



PACKAGE METACODER
(EXTENSION OF TAXA PACKAGE)

MANIPULATING *TAXMAP* OBJECT: DPLYR-LIKE FUNCTIONS

FILTERING:
- FILTER_TAXA
- FILTER_OBS

SUBSETTING:
- SELECT_OBS

ADDING COLUMNS:
- MUTATE_OBS

SAMPLING:
- SAMPLE_N_TAXA
- SAMPLE_N_OBS
- SAMPLE_FRAC_TAXA
- SAMPLE_FRAC_OBS

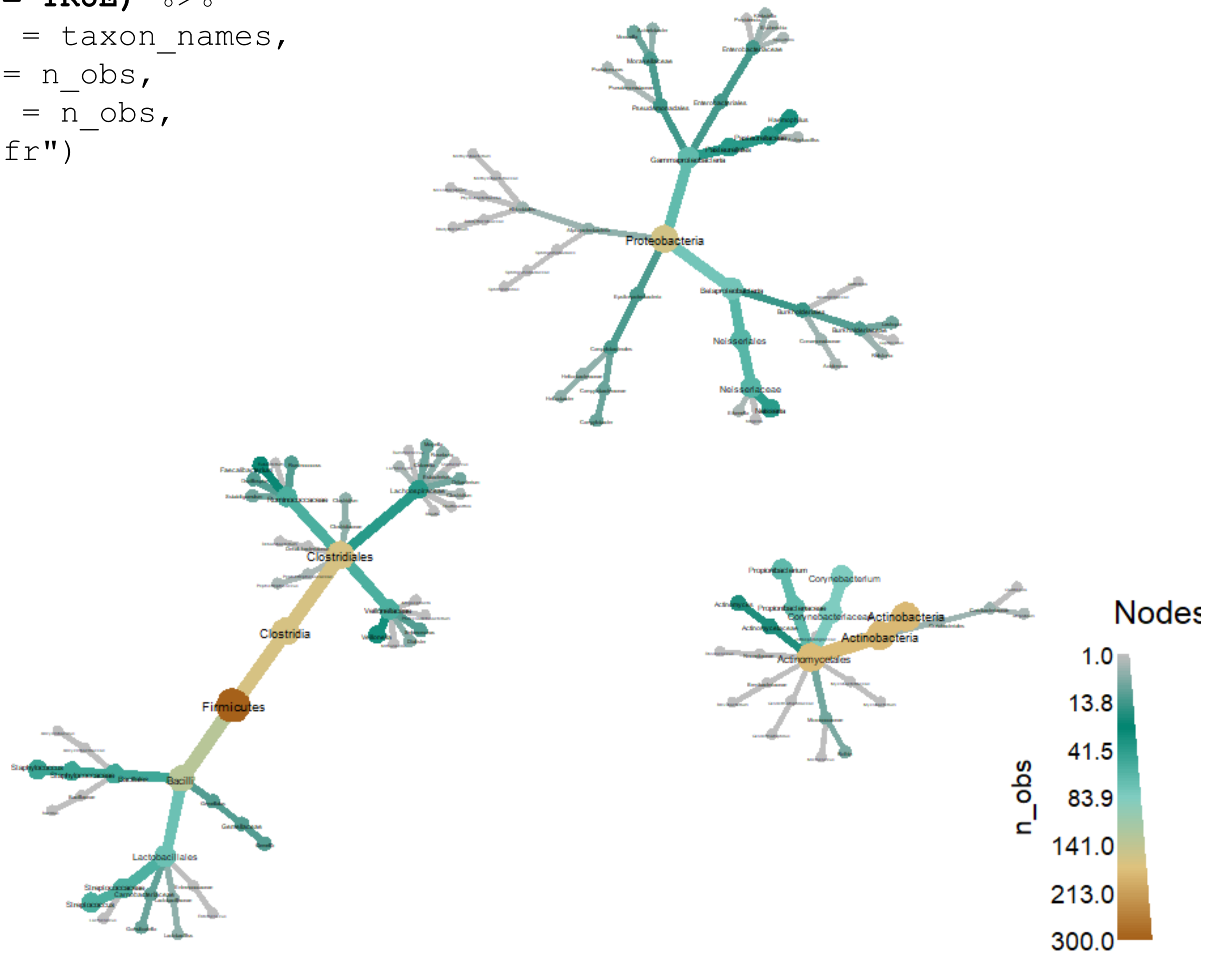
SORTING:
- ARRANGE_TAXA
- ARRANGE_OBS

PACKAGE METACODER

(EXTENSION OF TAXA PACKAGE)

VISUALIZING *TAXMAP* OBJECT:

```
hmp_data %>%
  filter_taxa(taxon_names %in% c("Proteobacteria", "Actinobacteria", "Firmicutes"),
              subtaxa = TRUE) %>%
  heat_tree(node_label = taxon_names,
            node_size = n_obs,
            node_color = n_obs,
            layout = "fr")
```



PACKAGE METACODER

(EXTENSION OF TAXA PACKAGE)

TRANSFORMING *TAXMAP* OBJECT:

RELATIVE ABUNDANCE:

```
obj$data$tax_data <- calc_obs_props(obj, "tax_data")
```

SUMMING PER-TAXON COUNTS:

```
obj$data$tax_abund <- calc_taxon_abund(obj, "tax_data",
                                       cols = hmp_samples$sample_id)
```

Package metacoder
(extension of taxa package)

```
> obj
<Taxmap>
  174 taxa: ab. Root, ac. Proteobacteria ... gr. Blautia, gs. Clostridium
  174 edges: NA->ab, ab->ac, ab->ad, ab->ae, ab->af ... dk->gp, cm->gq, cf->gr, cw->gs
  3 data sets:
    tax_data:
      # A tibble: 1,000 x 53
        taxon_id otu_id lineage `700035949` `700097855` `700100489` `700111314` `700033744`
        <chr>    <chr>  <chr>         <int>       <int>       <int>       <int>       <int>
      1 dm       OTU_9~ r__Roo~           0           2           1           0           0
      2 dn       OTU_9~ r__Roo~           0           0           0           0           0
      3 do       OTU_9~ r__Roo~           0           1           0           0           0
      # ... with 997 more rows, and 45 more variables: `700109581` <int>,
      #   `700111044` <int>, `700101365` <int>, `700100431` <int>, `700016050` <int>,
      #   `700032425` <int>, `700024855` <int>, `700103488` <int>, `700096869` <int>,
      #   `700107379` <int>, ...
    class_data:
      # A tibble: 5,922 x 5
        taxon_id input_index tax_rank tax_name            regex_match
        <chr>          <int> <chr>    <chr>               <chr>
      1 ab                 1 r        Root                r__Root
      2 ac                 1 p        Proteobacteria      p__Proteobacteria
      3 aj                 1 c        Gammaproteobacteria c__Gammaproteobacteria
      # ... with 5,919 more rows
    tax_abund:
      # A tibble: 174 x 51
        taxon_id `700035949` `700097855` `700100489` `700111314` `700033744` `700109581`
        <chr>          <dbl>       <dbl>       <dbl>       <dbl>       <dbl>       <dbl>
      1 ab               955        4330         887         598        6303        3379
      2 ac               193         144          17         132        1388          30
      3 ad                 5          43           6          21          20          15
      # ... with 171 more rows, and 44 more variables: `700111044` <dbl>,
      #   `700101365` <dbl>, `700100431` <dbl>, `700016050` <dbl>, `700032425` <dbl>,
      #   `700024855` <dbl>, `700103488` <dbl>, `700096869` <dbl>, `700107379` <dbl>,
      #   `700096422` <dbl>, ...
  0 functions:
```

PACKAGE METACODER
(EXTENSION OF TAXA PACKAGE)

STATISTICS ON *TAXMAP* OBJECT: COMPARE_GROUPS() FUNCTION

IT APPLIES A FUNCTION TO COMPARE DATA, USUALLY ABUNDANCE, FROM PAIRS OF TREATMENTS/GROUPS

BY DEFAULT: WILCOXON RANK SUM TEST ON THE DIFFERENCES IN MEDIAN ABUNDANCE FOR THE SAMPLES

PACKAGE METACODER

(EXTENSION OF TAXA PACKAGE)

STATISTICS ON *TAXMAP* OBJECT: COMPARE_GROUPS() FUNCTION

DIFFERENCES IN ABUNDANCE BETWEEN MICROBIOME COMMUNITIES IN DIFFERENT PARTS OF THE HUMAN BODY

```
hmp_data$data$diff_table <- compare_groups(hmp_data,
                                           data = "tax_prop",
                                           cols = hmp_samples$sample_id,
                                           groups = hmp_samples$body_site)
```

(by default wilcox.test)

CREATE NEW DATA IN HMP_DATA → `diff_table`

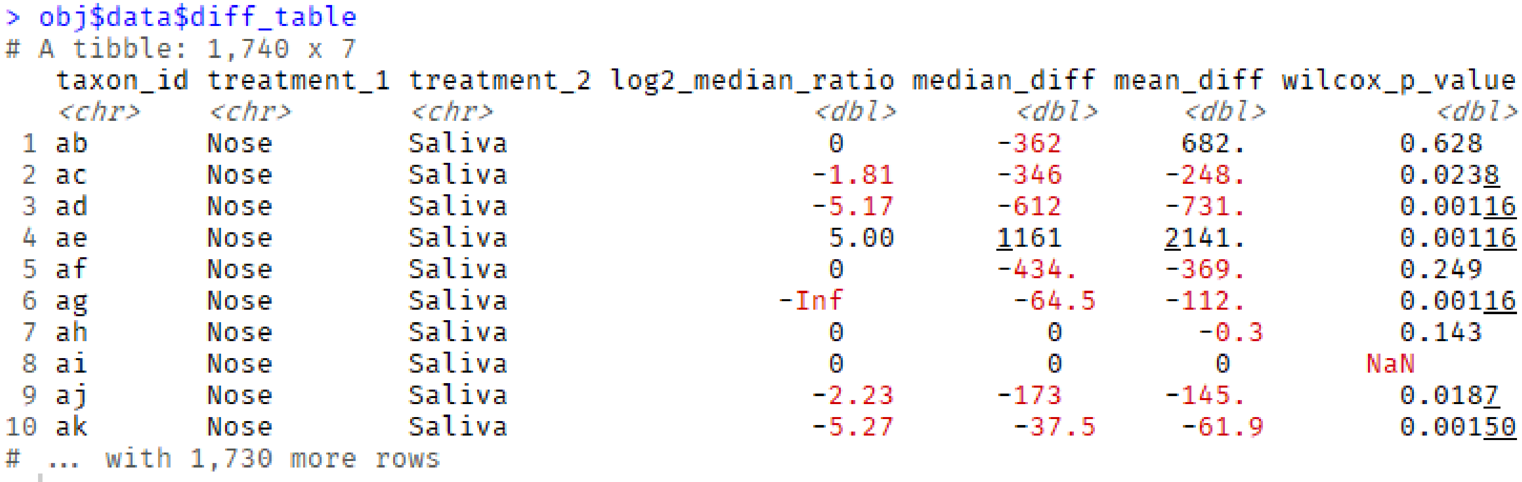
Package metacoder

(extension of taxa package)

Statistics on *taxmap* object: compare_groups() function

1740 tests correction "fdr"

```
hmp_data <- mutate_obs(hmp_data, "diff_table",
                       wilcox_p_value = p.adjust(wilcox_p_value, method = "fdr"),
                       log2_median_ratio = ifelse(wilcox_p_value < 0.05 | is.na(wilcox_p_value),
                       log2_median_ratio, 0))
```



```
> obj$data$diff_table
# A tibble: 1,740 x 7
   taxon_id treatment_1 treatment_2 log2_median_ratio median_diff mean_diff wilcox_p_value
   <chr>    <chr>       <chr>                   <dbl>       <dbl>     <dbl>          <dbl>
 1 ab       Nose        Saliva                   0         -362       682.         0.628
 2 ac       Nose        Saliva                  -1.81      -346      -248.         0.0238
 3 ad       Nose        Saliva                  -5.17      -612      -731.         0.00116
 4 ae       Nose        Saliva                   5.00      1161      2141.         0.00116
 5 af       Nose        Saliva                   0         -434.     -369.         0.249
 6 ag       Nose        Saliva                -Inf          -64.5    -112.         0.00116
 7 ah       Nose        Saliva                   0            0        -0.3        0.143
 8 ai       Nose        Saliva                   0            0         0        NaN
 9 aj       Nose        Saliva                  -2.23      -173      -145.         0.0187
10 ak       Nose        Saliva                  -5.27       -37.5     -61.9        0.00150
# ... with 1,730 more rows
```

Package metacoder

(extension of taxa package)

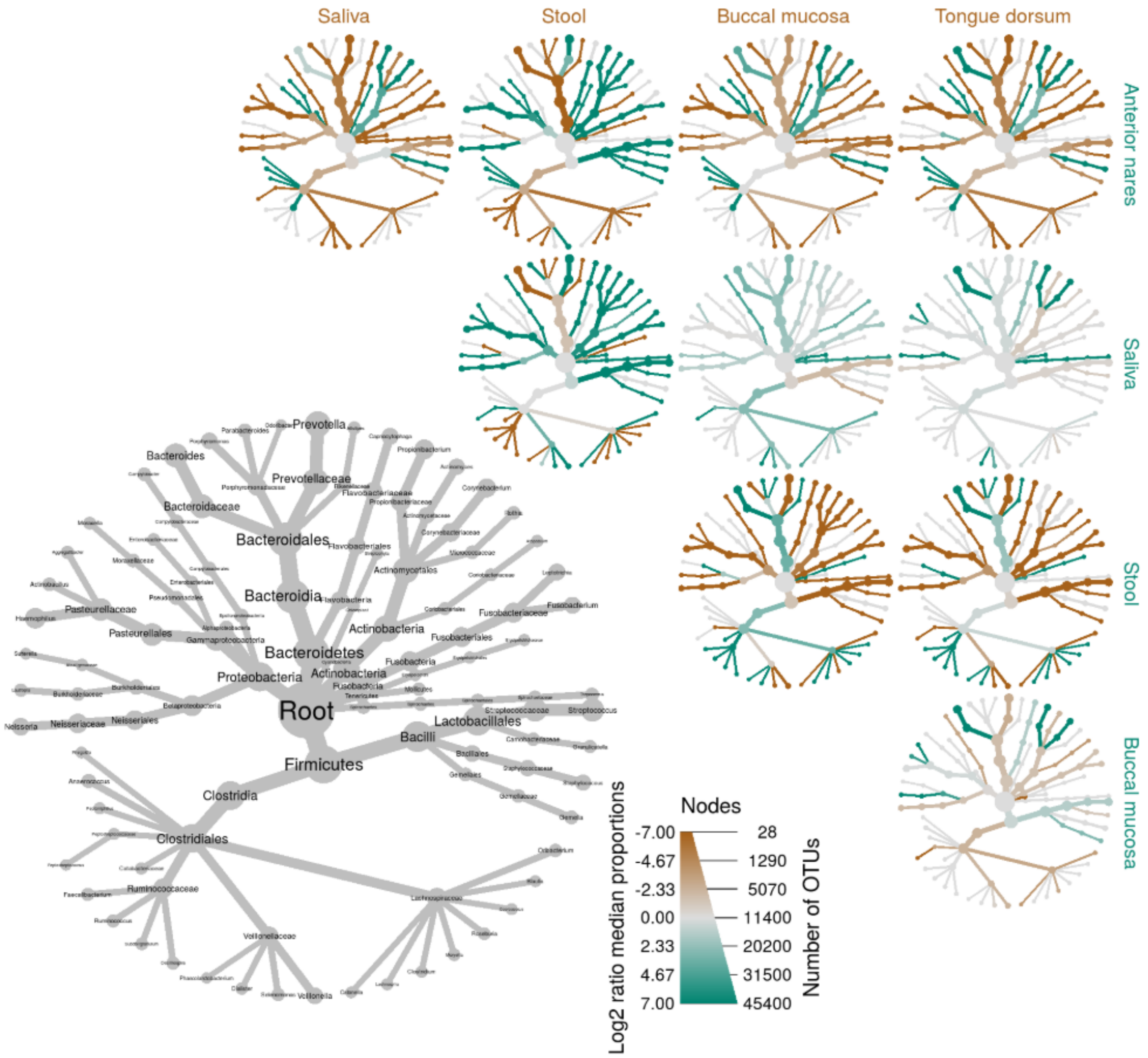
Visualizing comparison: heat_tree_matrix() function

```
heat_tree_matrix(hmp_data,
                 data = "diff_table",
                 node_size = n_obs,
                 node_label = taxon_names,
                 node_color = log2_median_ratio
                 node_color_range = diverging_palette(),
                 node_color_trans = "linear",
                 node_color_interval = c(-3, 3),
                 edge_color_interval = c(-3, 3),
                 node_size_axis_label = "Number of OTUs",
                 node_color_axis_label = "Log2 ratio median proportions",
                 layout = "davidson-harel",
                 initial_layout = "reingold-tilford")
```

Package metacoder

(extension of taxa package)

Visualizing comparison: heat_tree_matrix() function



Package metacoder

(extension of taxa package)

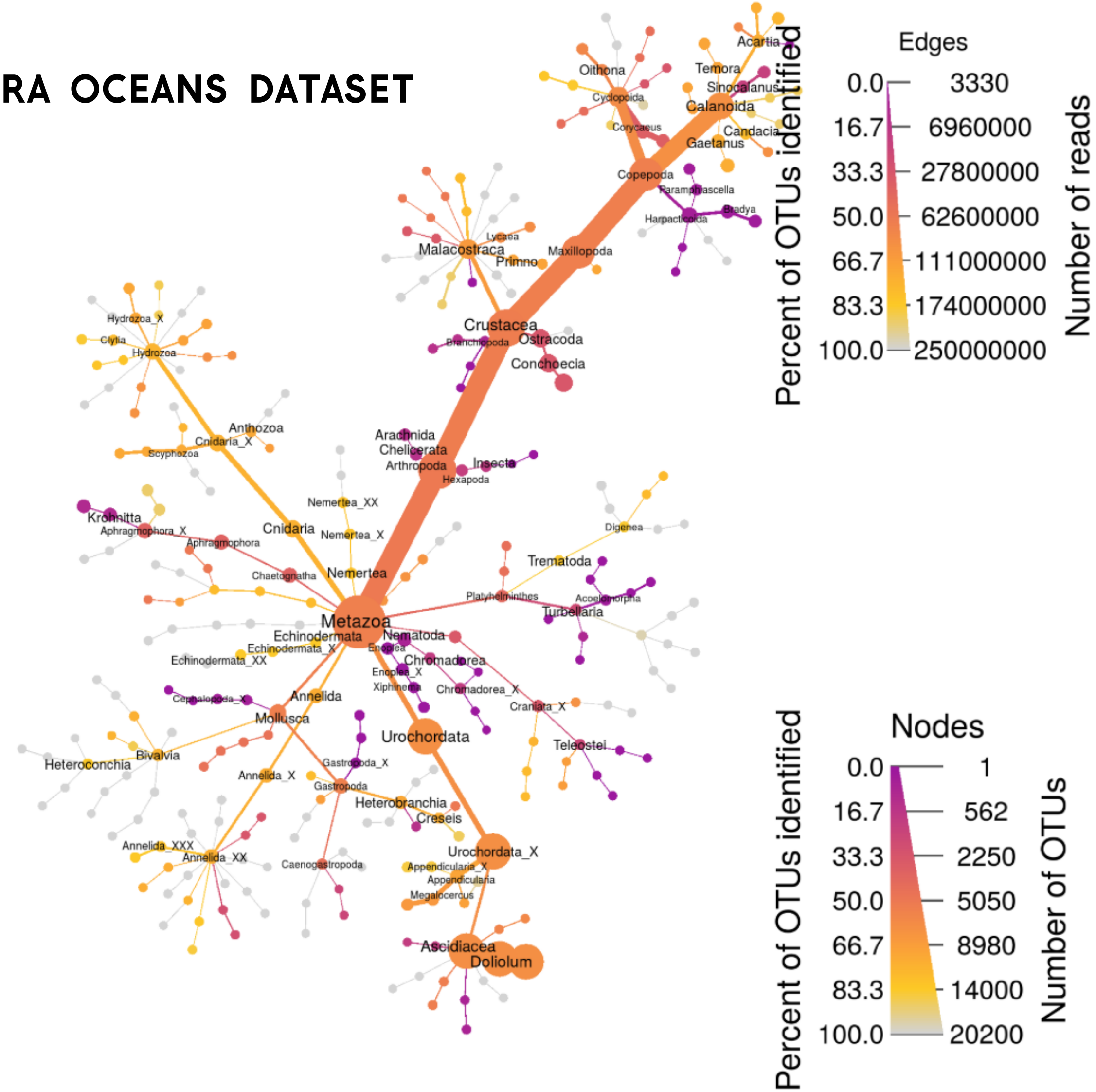
Metacoder and Tara oceans dataset

(20 200 OTU)

10Gb RAM

few min …

3 statistics



PACKAGE METACODER

(EXTENSION OF TAXA PACKAGE)

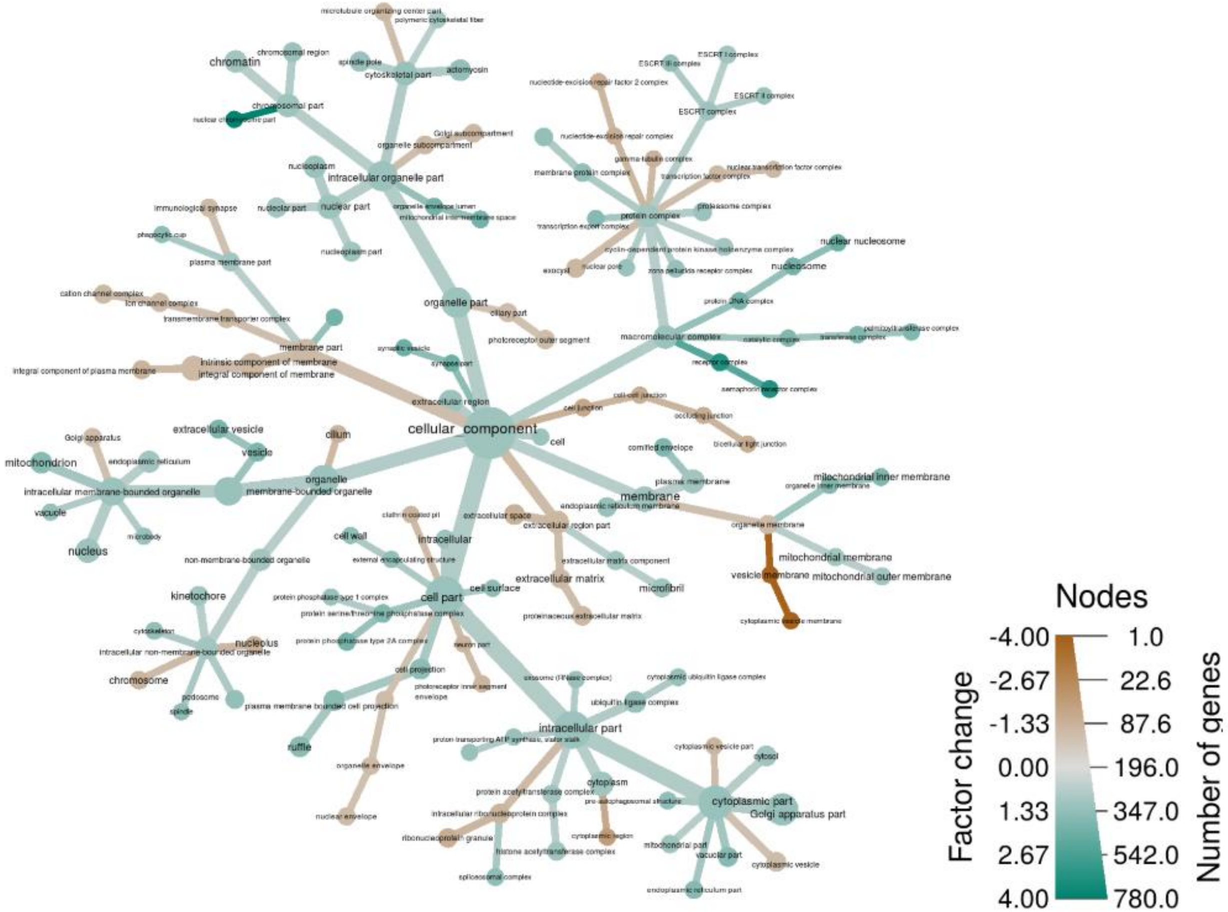
METACODER GENE EXPRESSION

DISPLAYING THE RESULTS OF GENE EXPRESSION STUDIES BY ASSOCIATING DIFFERENTIAL EXPRESSION WITH GENE ONTOLOGY (GO) ANNOTATIONS

PACKAGE METACODER
(EXTENSION OF TAXA PACKAGE)

METACODER GENE EXPRESSION

CELLULAR COMPONENT

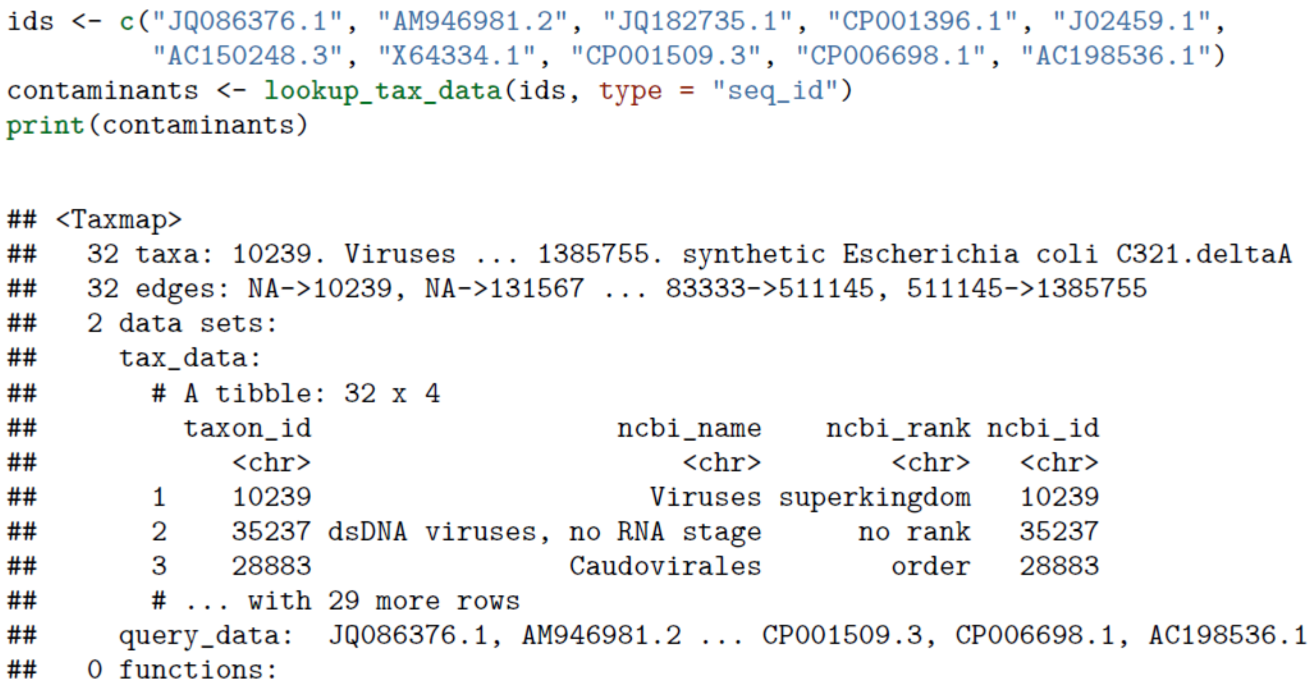


Package metacoder

(extension of taxa package)

Creating object from public database (NCBI, …):

ID → LOOKUP_TAX_DATA → TAXMAP



```
ids <- c("JQ086376.1", "AM946981.2", "JQ182735.1", "CP001396.1", "J02459.1",
         "AC150248.3", "X64334.1", "CP001509.3", "CP006698.1", "AC198536.1")
contaminants <- lookup_tax_data(ids, type = "seq_id")
print(contaminants)
```

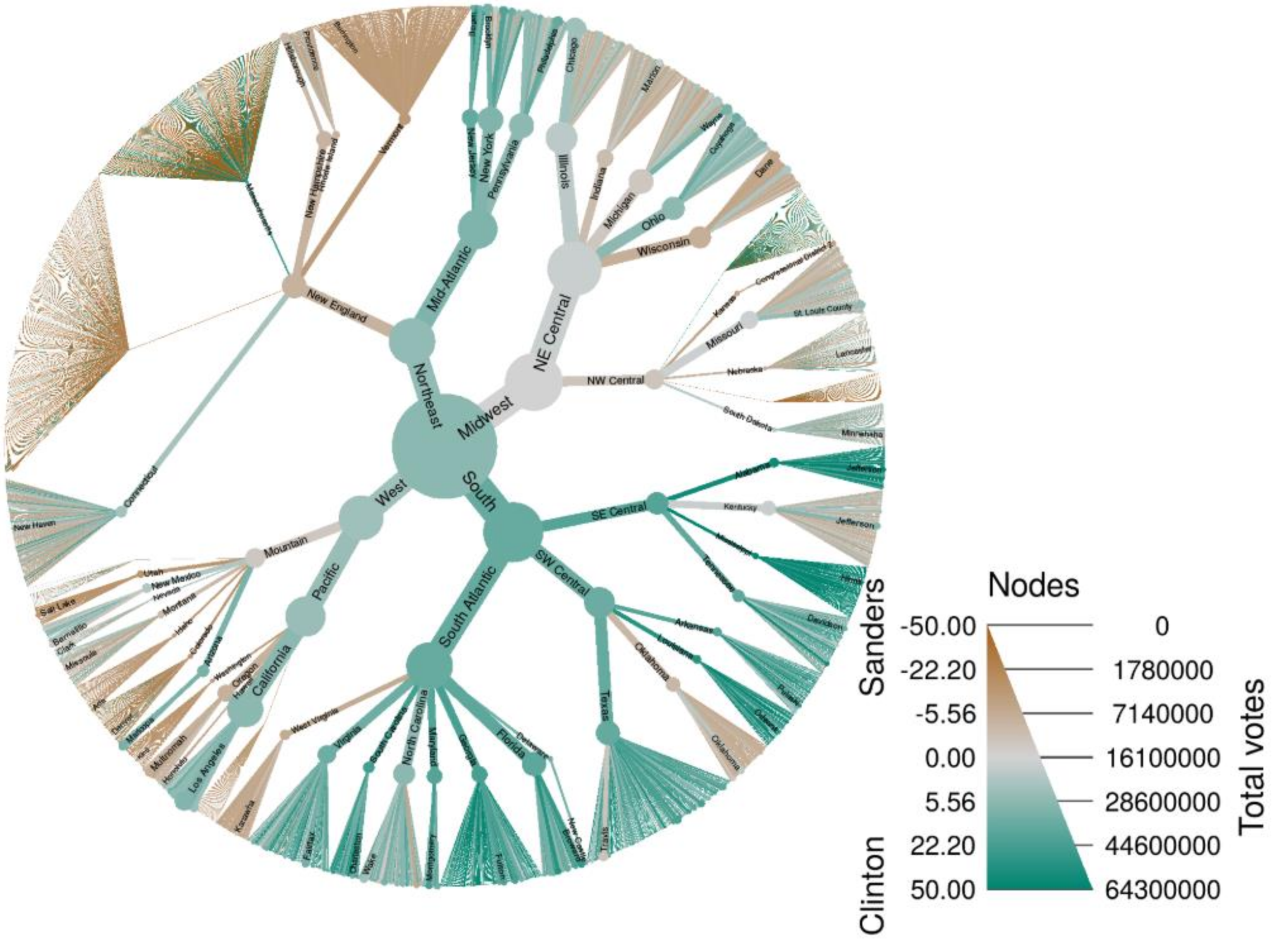
```
## <Taxmap>
##   32 taxa: 10239. Viruses ... 1385755. synthetic Escherichia coli C321.deltaA
##   32 edges: NA->10239, NA->131567 ... 83333->511145, 511145->1385755
##   2 data sets:
##     tax_data:
##       # A tibble: 32 x 4
##         taxon_id                   ncbi_name    ncbi_rank ncbi_id
##            <chr>                       <chr>        <chr>   <chr>
##       1    10239                     Viruses superkingdom   10239
##       2    35237 dsDNA viruses, no RNA stage      no rank   35237
##       3    28883                Caudovirales        order   28883
##       # ... with 29 more rows
##     query_data:  JQ086376.1, AM946981.2 ... CP001509.3, CP006698.1, AC198536.1
##   0 functions:
```

PACKAGE METACODER
(EXTENSION OF TAXA PACKAGE)

METACODER AND GEOGRAPHICAL DATA

RESULTS OF THE 2016 DEMOCRATIC PRIMARY ELECTION IN US

HIERARCHY:
- REGION
  - DIVISION
    - STATE
      - COUNTY



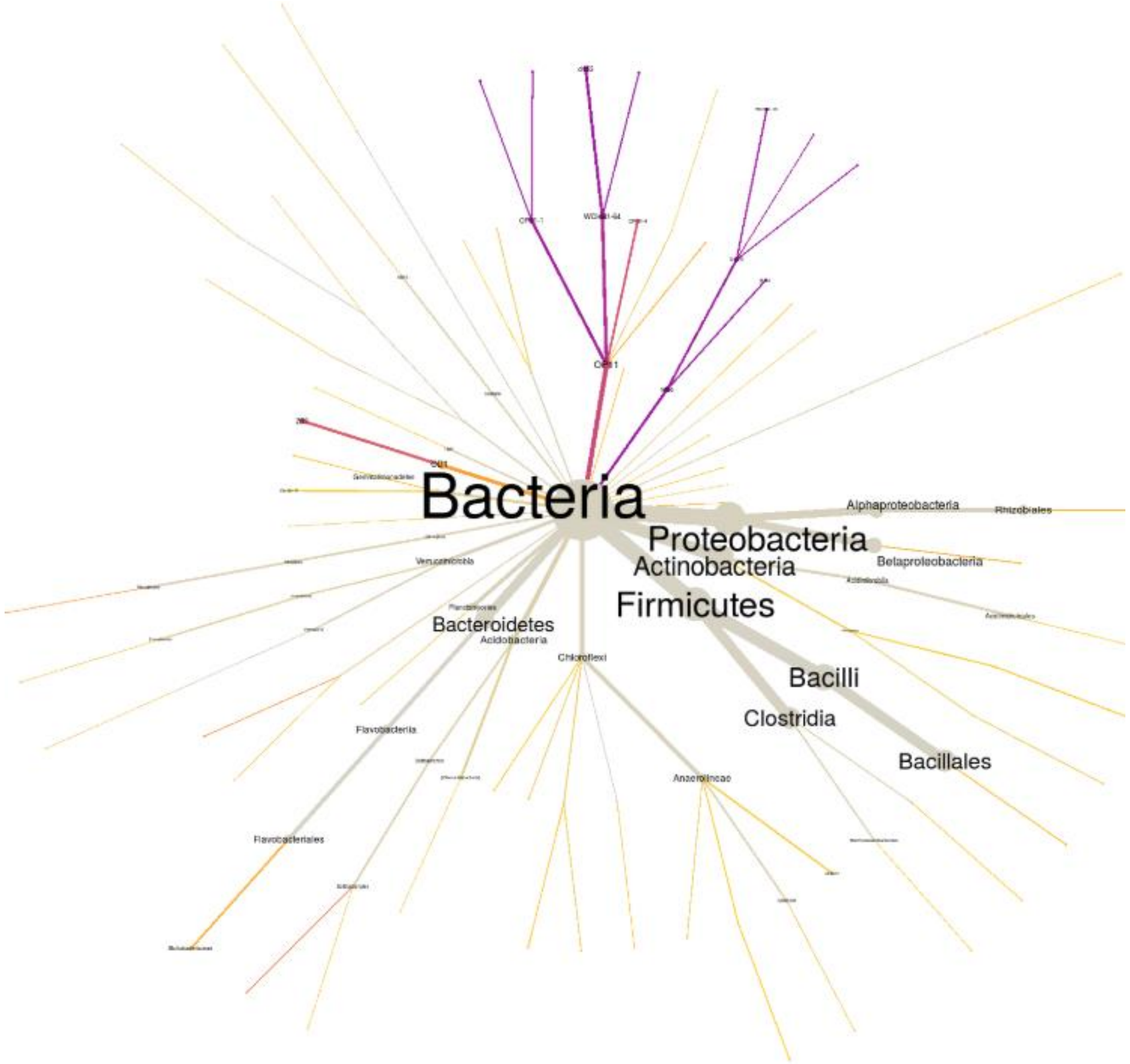
Package metacoder
(extension of taxa package)

To conclude:

Pros:
- Can handle any hierarchical dataset
- Provides a large panel of functions for manipulating data (based on dplyr)
- Creates customizable graph (based on ggplot), ~ 70 parameters
- Allows to add any type of data linked with hierarchical dataset
- Compare_groups() function

Cons:
- Requires time at the beginning (*taxa* package environment)
- *taxmap* object complexity
- Many parameters (~ 70 for heat_tree)
- Can be slow for large datasets

QUESTIONS ?


Bacteria
Proteobacteria
Actinobacteria
Firmicutes
Alphaproteobacteria
Rhizobiales
Betaproteobacteria
Bacteroidetes
Acidobacteria
Chloroflexi
Verrucomicrobia
Bacilli
Clostridia
Bacillales
Flavobacteria
Flavobacteriales
Anaerolineae

Article:
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005404

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

Taxa package article:
https://f1000research.com/articles/7-272/v2