MEDICAL MICROBIOLOGY AND INFECTIOUS DISEASES BIOINFORMATICS WORKSHOP

Presents

Introduction to tree visualization and annotation using ggtree

INSTRUCTED BY
Taylor Davedow, PhD Student
Department of Medical Microbiology and Infectious Diseases
University of Manitoba

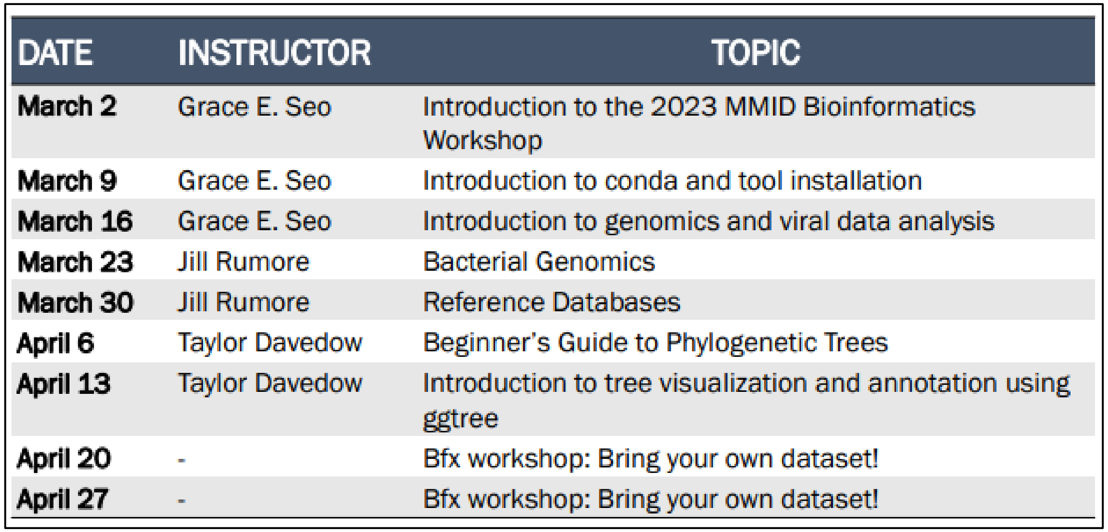
Educational series supported by the Department of MMID to promote knowledge translation and dissemination among students and researchers

INFORMATION FOR PARTICIPANTS

All workshops are being recorded and posted to the MMID Bioinformatics Workshop – YouTube

For live Q&A, go to slido.com and use participant code #3323315

2023 MMID Bioinformatics Workshop Schedule



DATE	INSTRUCTOR	TOPIC
March 2	Grace E. Seo	Introduction to the 2023 MMID Bioinformatics Workshop
March 9	Grace E. Seo	Introduction to conda and tool installation
March 16	Grace E. Seo	Introduction to genomics and viral data analysis
March 23	Jill Rumore	Bacterial Genomics
March 30	Jill Rumore	Reference Databases
April 6	Taylor Davedow	Beginner's Guide to Phylogenetic Trees
April 13	Taylor Davedow	Introduction to tree visualization and annotation using ggtree
April 20	-	Bfx workshop: Bring your own dataset!
April 27	-	Bfx workshop: Bring your own dataset!

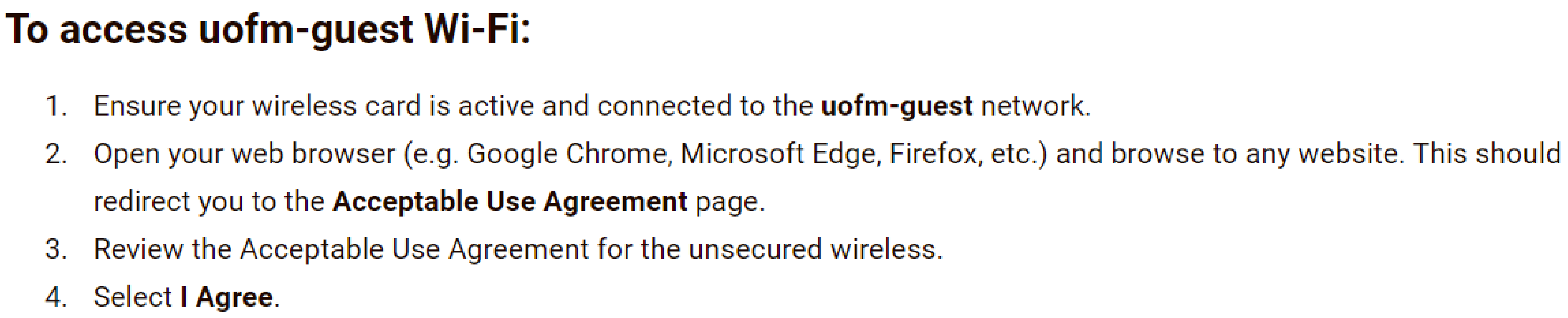
April 20 and April 27 in-person sessions are open to the public (up to 100 people)!

Work with your colleagues/friends to analyze data together.

SET UP WI-FI (IN-PERSON PARTICIPANTS)

1. ***Connect to UofM-secure (if you are a student or staff)***
   - Use your @myumanitoba.ca or @umanitoba.ca login and password

2. ***Connect to UofM-guest***



To access uofm-guest Wi-Fi:

1. Ensure your wireless card is active and connected to the **uofm-guest** network.
2. Open your web browser (e.g. Google Chrome, Microsoft Edge, Firefox, etc.) and browse to any website. This should redirect you to the **Acceptable Use Agreement** page.
3. Review the Acceptable Use Agreement for the unsecured wireless.
4. Select **I Agree**.

LEARNING OBJECTIVES

1. Create a tree using ggtree
2. Use geometric functions and aesthetic mappings to annotate tree
3. Learn how to customize legends and add themes

DISCLAIMER

To provide a basic working instruction, all tools will be run with default settings. HOWEVER, careful consideration of analysis parameters in the context of the research question should be taken into account when analyzing your own datasets, as default parameters do not always provide the most optimal result.

FILES FOR WORKSHOP

1. msa.fasta.tree (newick format)
   - Created by downloading a subset of genome accessions listed in manuscript, skesa assembly, multiple sequence alignment and running iqtree

2. metadata.xlsx
   - Information compiled from manuscript (Tables 1 & 2)

3. RScripts
   - Follow along with today's workshop using our R code

Please download and save materials to your workshop directory: https://github.com/mmid-bioinformatics-workshop/2023-04-13-Intro-to-tree-visualization-and-annotation-using-ggtree

DEMONSTRATION

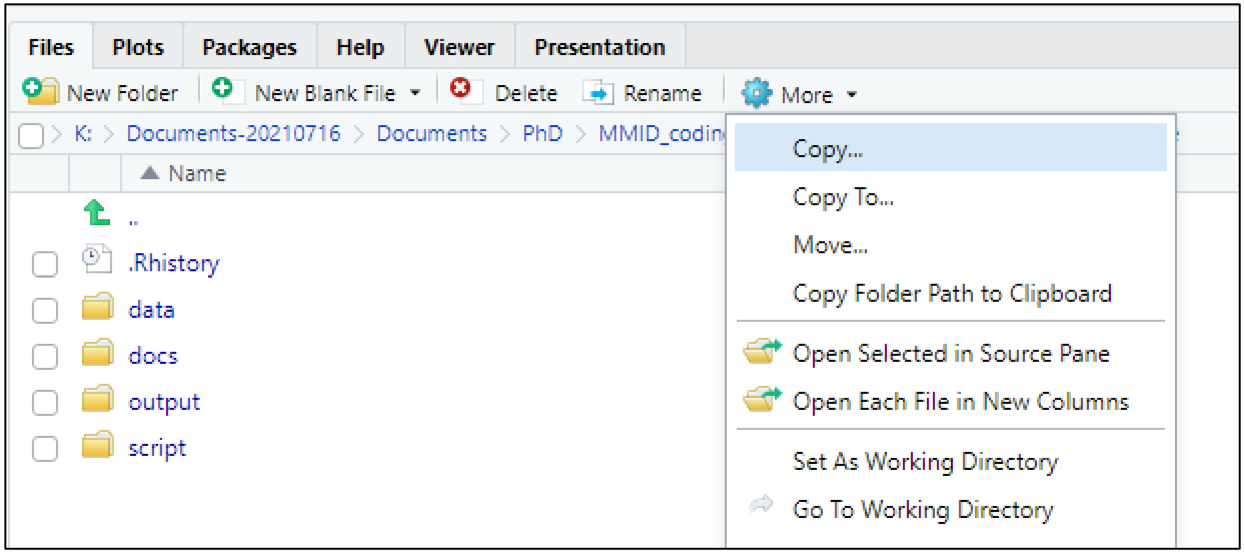
GETTING STARTED…

1. Open up RStudio
2. Set working directory to your workshop folder:

   In the files pane, navigate to your workshop folder, then click: more > set as working directory
3. Create four sub directories in the workshop directory:

   Move the tree and metadata to "data" directory

   Move script files to "script" directory



GETTING STARTED…

4. Open up Script and install packages

```
install.packages(“readxl”)
install.packages("BiocManager")
install.packages("treeio")
install.packages("tidyverse")
install.packages("phytools")
```

5. Load BiocManager package then install ggtree

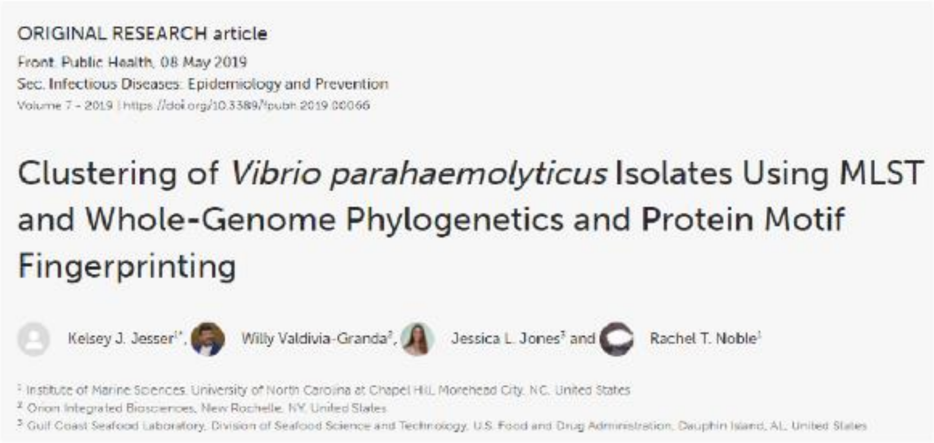
```
library(BiocManager)
BiocManager::install("ggtree")
```

** Note: if you already have ggtree (or other packages) installed you can skip installation**

Detailed instructions for installing and loading packages can be found here: https://www.rdocumentation.org/packages/utils/versions/3.6.2/topics/install.packages

PUBLICALLY AVAILABLE DATASET

https://doi.org/10.3389/fpubh.2019.00066



ORIGINAL RESEARCH article

Front. Public Health, 08 May 2019

Sec. Infectious Diseases: Epidemiology and Prevention

Volume 7 - 2019 | https://doi.org/10.3389/fpubh.2019.00066

Clustering of *Vibrio parahaemolyticus* Isolates Using MLST and Whole-Genome Phylogenetics and Protein Motif Fingerprinting

Kelsey J. Jesser[1]*, Willy Valdivia-Granda[2], Jessica L. Jones[3] and Rachel T. Noble[1]

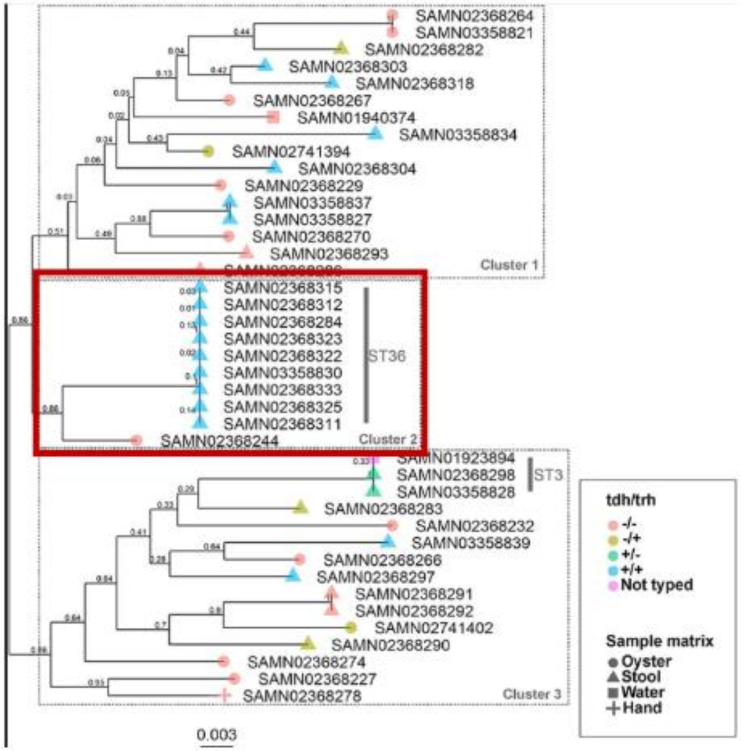
[1] Institute of Marine Sciences, University of North Carolina at Chapel Hill, Morehead City, NC, United States

[2] Orion Integrated Biosciences, New Rochelle, NY, United States

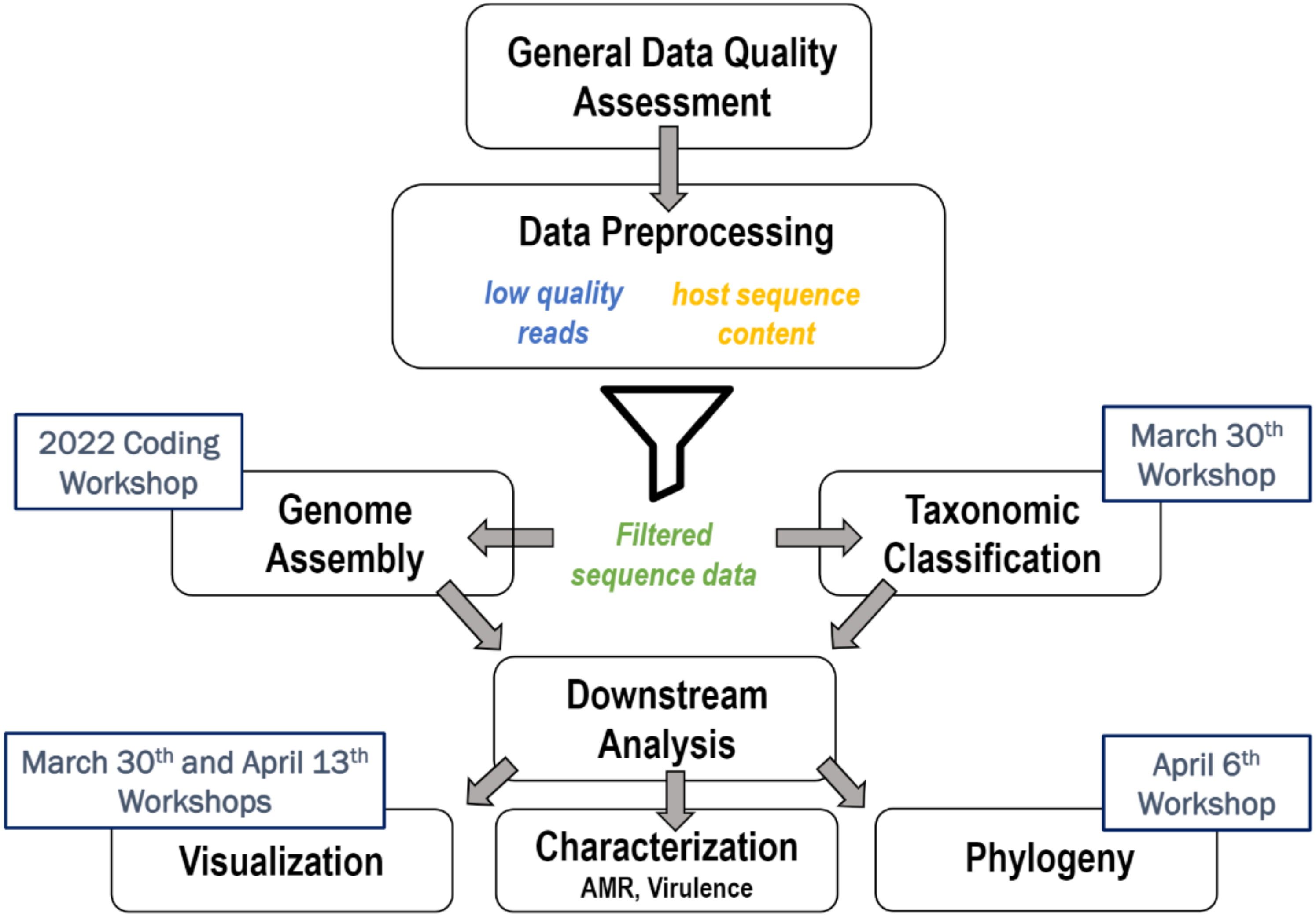
[3] Gulf Coast Seafood Laboratory, Division of Seafood Science and Technology, U.S. Food and Drug Administration, Dauphin Island, AL, United States

Uploaded to the MMID Bioinformatics GitHub Repository https://github.com/mmid-bioinformatics-workshop

Data was sequenced using the Illumina HiSeq 2000



BASIC WORKFLOW



General Data Quality Assessment

Data Preprocessing

low quality reads

host sequence content

Filtered sequence data

2022 Coding Workshop

Genome Assembly

March 30th Workshop

Taxonomic Classification

Downstream Analysis

March 30th and April 13th Workshops

Visualization

Characterization

AMR, Virulence

April 6th Workshop

Phylogeny

ggtree

IQtree

GGTREE

- Package for R programming language
- Under Bioconductor project
- Creator: Guangchuang Yu
- Extension of ggplot2
- Data integration, manipulation and visualization of phylogenetic trees
- Customized annotation of tree

More information about ggtree can be found here:

- https://yulab-smu.top/treedata-book/
- https://github.com/YuLab-SMU/ggtree

FOLLOW ALONG IN RSCRIPT

GETTING STARTED…

Load packages

```
library(readxl)   # for reading in xl files
library(ggtree)   # for building tree
library(treeio)   # for read.newick function
library(phytools) # for midpoint.root (also has read.newick option)
library(tidyverse)# to assist with data tidying
```

** We will also be using ggplot2 which should automatically load in with ggtree**

LOAD IN FILES

Load tree and metadata file from data directory

```
# tree file
tree <- read.newick("data/msa.fasta.tree")

# metadata file
metadata <- read_xlsx("data/metadata.xlsx")
```

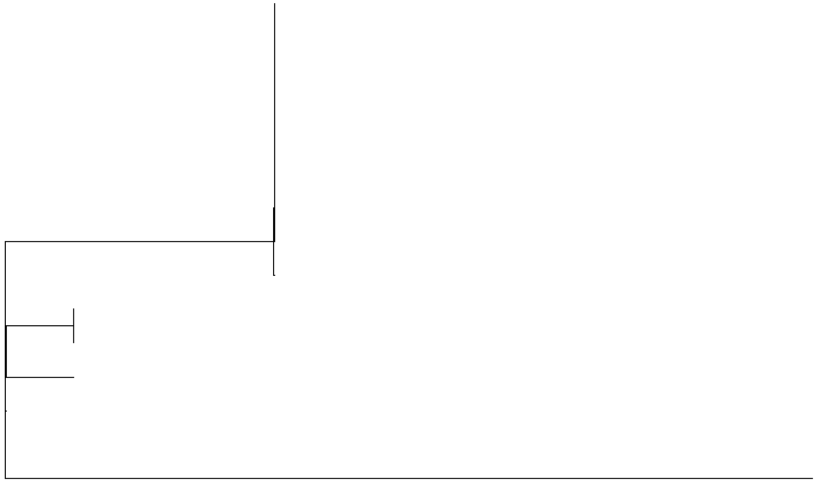
CREATE A BASIC TREE

Create a basic tree using the ggtree() function:

```
ggtree(tree)
```

- Tree is our phylo object
- For a list of other arguments check out the ggtree help page:

```
?ggtree
```

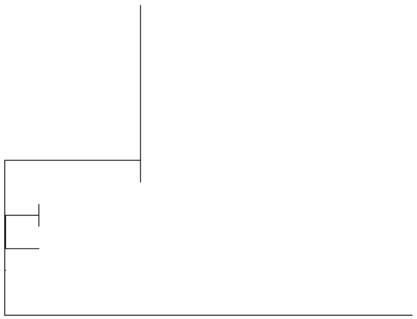


CHANGING TREE LAYOUT

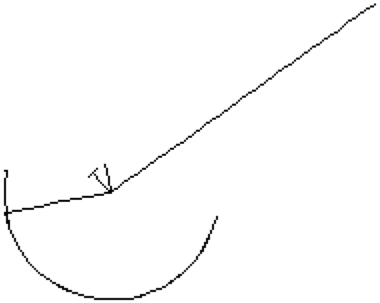
You can change the tree layout by using the “layout” argument

```
ggtree(tree, layout = “rectangular”)
```

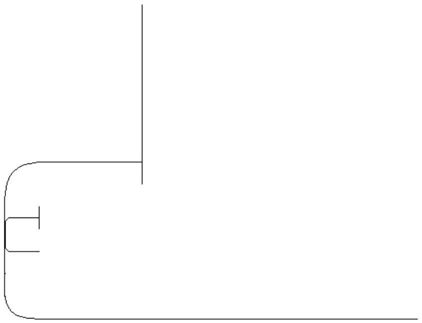
rectangular



circular



roundrect



CHANGING TREE LAYOUT

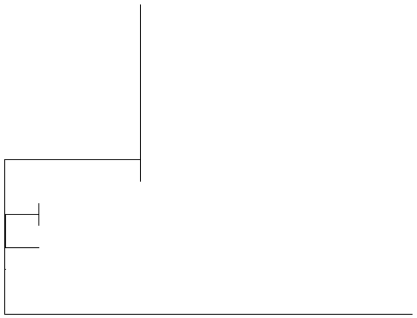
A **cladogram** will show topology without branch length information

```
ggtree(tree, branch.length = "none")
```

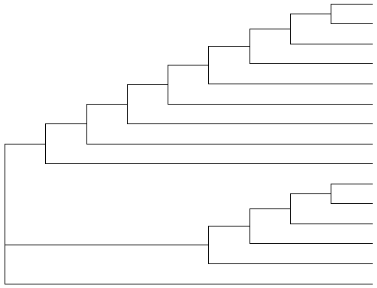
Midpoint root: roots the tree at the midpoint of the longest point between two tips.

```
ggtree(midpoint.root(tree))
```

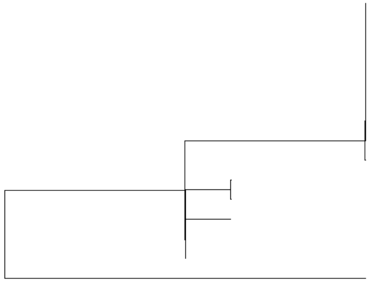
rectangular



cladogram



midpoint.root



You could also try changing position of the root node using **root.position** argument

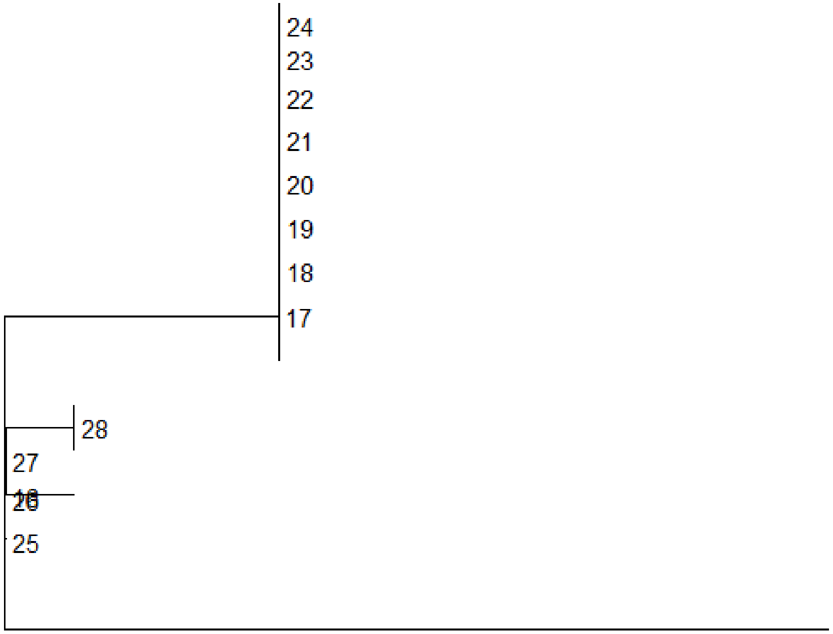
**Caution: rooting *can* drastically change tree topology and you must use an appropriate method based on your data **

IDENTIFYING NODES

We can display the node numbers by using geom_text2 argument:

```
ggtree(tree) +
geom_text2(aes(subset=!isTip,
                        label=node),
                  hjust = -.3)
```

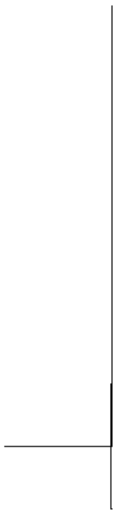
This will help us refer to a specific nodes while using other functions later on



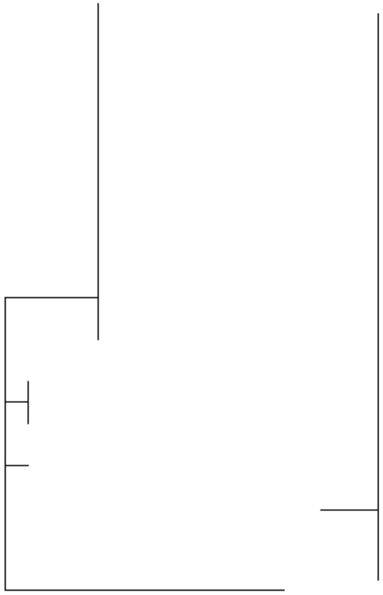
TREE MANIPULATION

- To view a particular clade, we can use viewClade()
- Notice the difference b/w operators
  - + Plus
  - %>% Pipe

```
> ggtree(tree) %>%
  viewClade(node = 17)
```

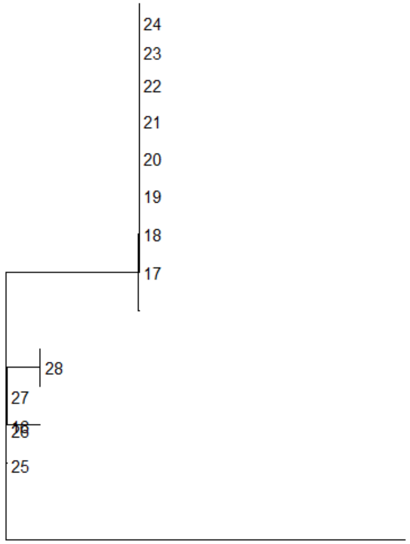


```
> ggtree(tree) +
  viewClade(node = 17)
```

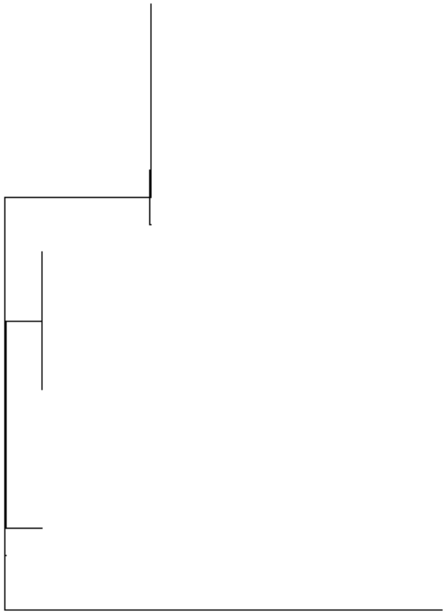


TREE MANIPULATION

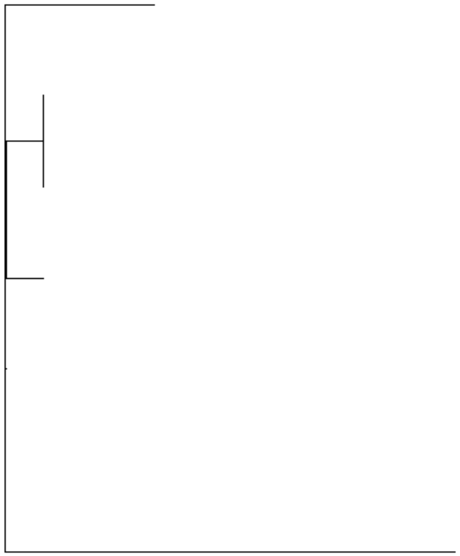
Original tree with
node labels



```
> ggtree(tree) %>%
  scaleClade(node = 27,
             scale = 5)
```



```
> ggtree(tree) %>%
  collapse(node = 17)
```



ADDING TIP LABELS

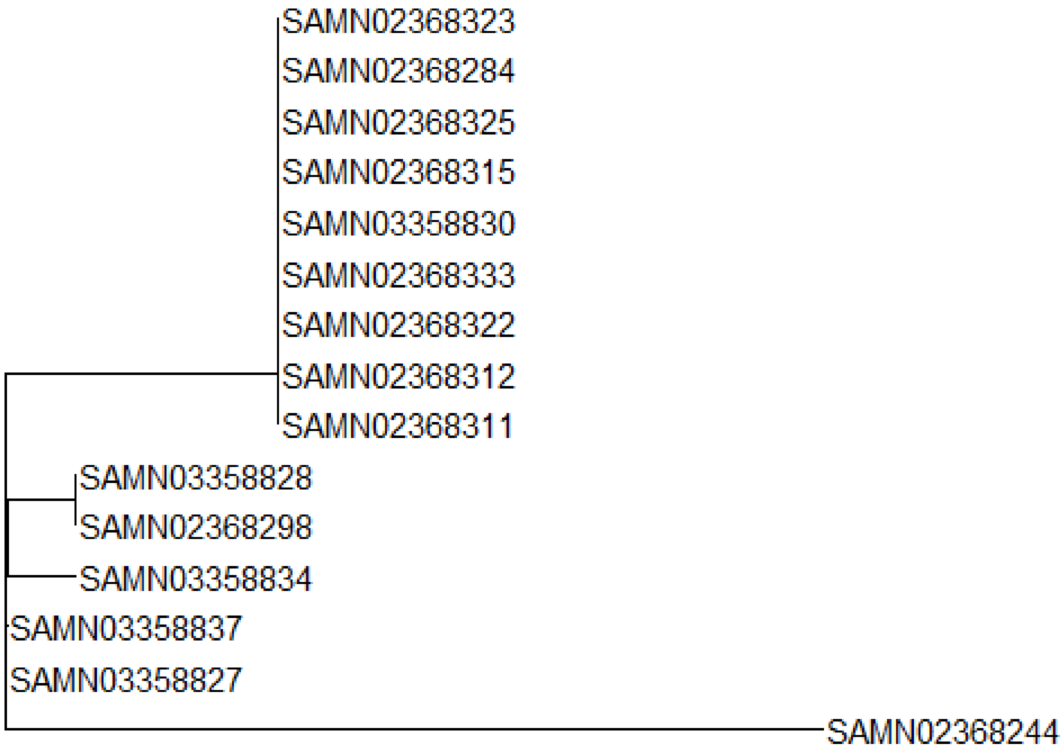
1. You can check the name of the tip labels using head()

```
head(tree$tip.label)
```

2. Add tip labels using geom_tiplab

```
ggtree(tree) +
  geom_tiplab(size = 4) + # displaying tip labels
  coord_cartesian(clip = 'off')+ # allows us to draw outside the
plot
  theme(plot.margin = margin(1,3,1,1, "cm")) # add space around
the plot
```

ADDING TIP LABELS



CUSTOMIZING TIP LABELS

1. Since we will be linking the metadata file to the tree, we can use a vector to check if there are any biosample_id observations that are not in the tree:

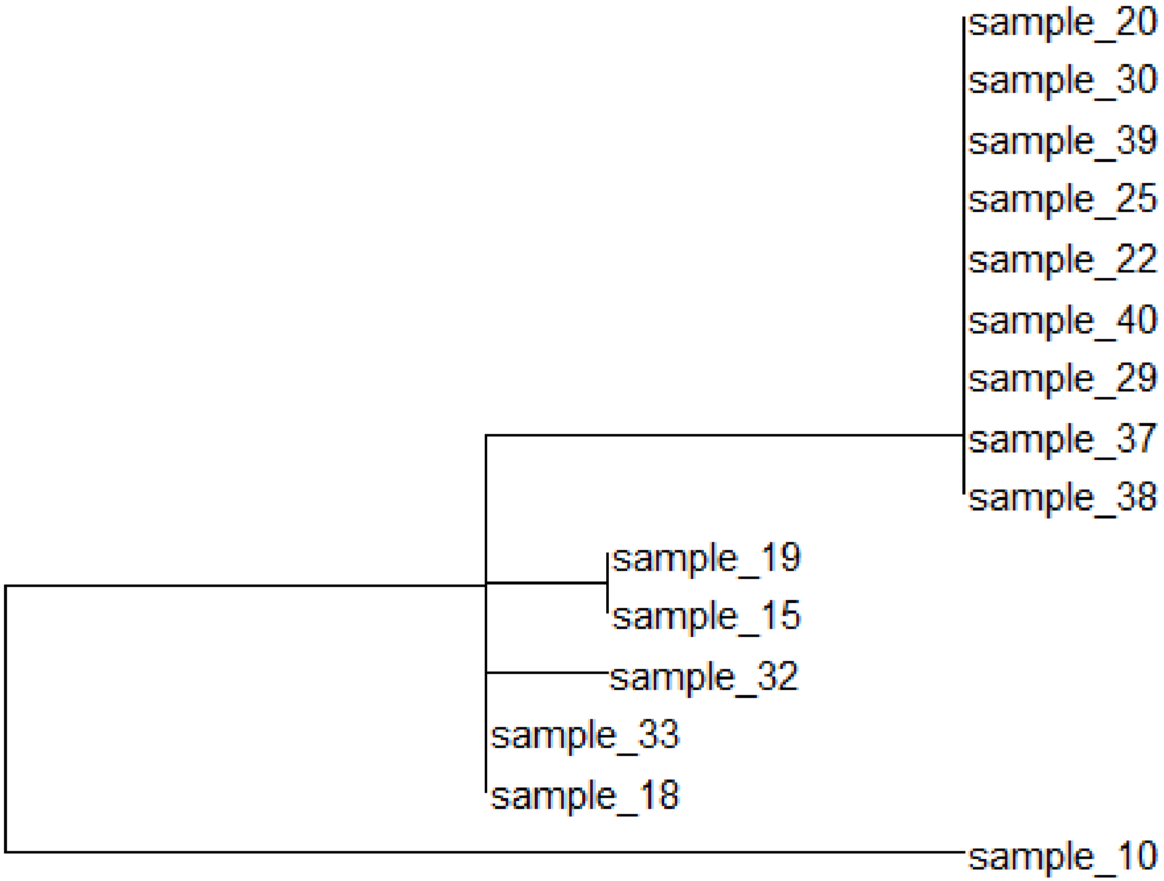
```
metadata$biosample_id[!tree$tip.label %in% metadata$biosample_id]
[1]  character(0)
# all observations match b/w tree and metadata
```

2. Link metadata and change the tip label using geom_tiplab:

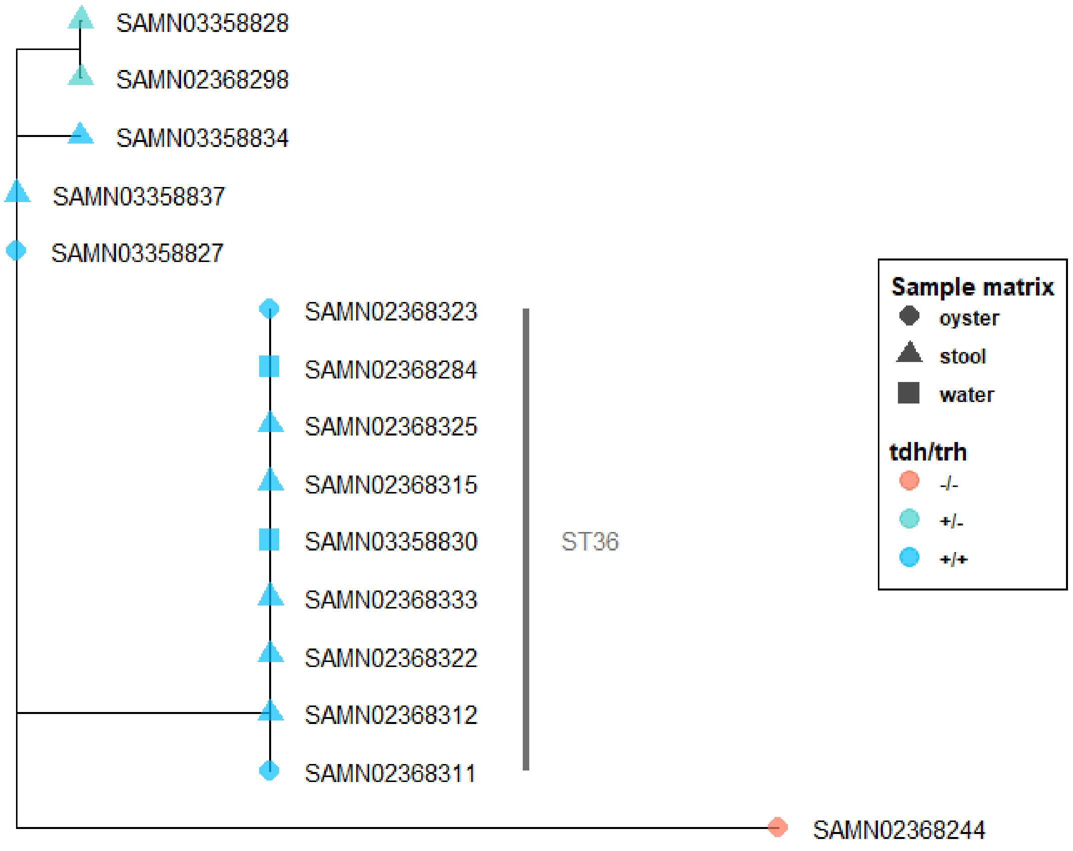
```
ggtree(tree) %<+% # to attach annotation data to tree
  metadata + # our metadata
  geom_tiplab(aes(label = sample_id)) + # change to strain_ID
  coord_cartesian(clip = 'off')+
  theme(plot.margin = margin(1,3,1,1, "cm"))
```

Any data we want linked to the tree must have a column that **EXACTLY matches** the tip.label

CUSTOMIZING TIP LABELS



EXAMPLE 1


SAMN03358828
SAMN02368298
SAMN03358834
SAMN03358837
SAMN03358827
SAMN02368323
SAMN02368284
SAMN02368325
SAMN02368315
SAMN03358830
SAMN02368333
SAMN02368322
SAMN02368312
SAMN02368311
ST36
SAMN02368244
Sample matrix
oyster
stool
water
tdh/trh
-/-
+/-
+/+

DEMONSTRATION

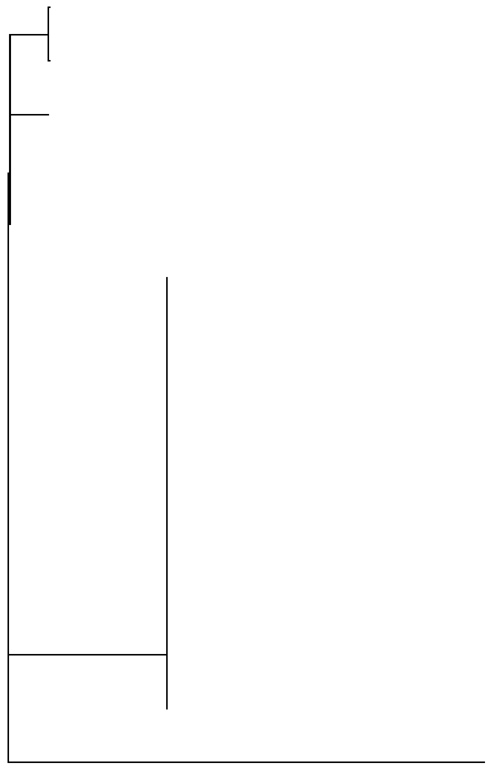
EXAMPLE 1

1. create a simple tree and save it as an object

```
gg_simple <- ggtree(tree) %<+%
  metadata + # link our metadata file
here
  coord_cartesian(clip = 'off')+
  theme(plot.margin = margin(1,4,1,1,
"cm"))
```

2. Reorient the tree using flip(), and save as new object

```
gg_flip <- gg_simple %>%
  flip(25, 17)
```



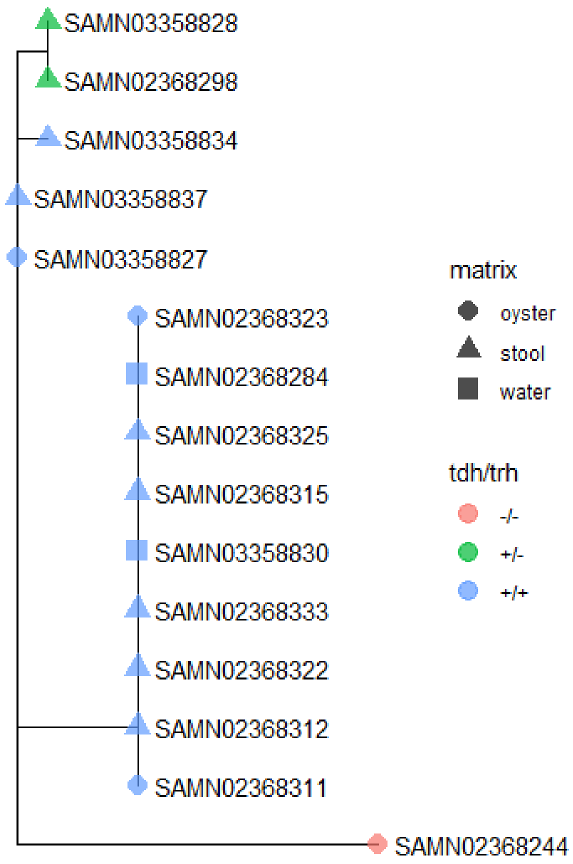
EXAMPLE 1

3. Add tip labels

```
gg_flip +
  geom_tiplab(offset = 0.0001)
```

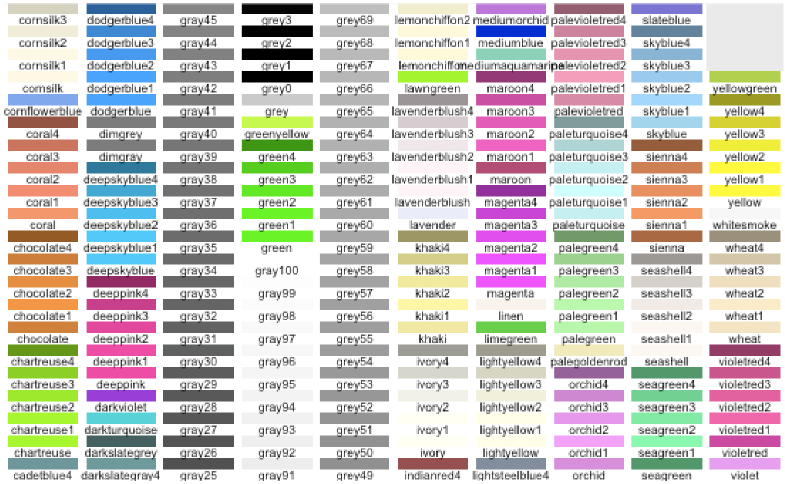
4. Add tip points

```
gg_flip +
  geom_tiplab(offset = 0.0001)+
  geom_tippoint(aes(color = `tdh/trh`,
shape = matrix),
                size = 4,
                alpha = 0.7)
```



SCALE_MANUAL

Create your own discrete scale



cornsilk3
cornsilk2
cornsilk1
cornsilk
cornflowerblue
coral4
coral3
coral2
coral1
coral
chocolate4
chocolate3
chocolate2
chocolate1
chocolate
chartreuse4
chartreuse3
chartreuse2
chartreuse1
chartreuse
cadetblue4

dodgerblue4
dodgerblue3
dodgerblue2
dodgerblue1
dodgerblue
dimgrey
dimgray
deepskyblue4
deepskyblue3
deepskyblue2
deepskyblue1
deepskyblue
deeppink4
deeppink3
deeppink2
deeppink1
deeppink
darkviolet
darkturquoise
darkslategrey
darkslategray4

gray45
gray44
gray43
gray42
gray41
gray40
gray39
gray38
gray37
gray36
gray35
gray34
gray33
gray32
gray31
gray30
gray29
gray28
gray27
gray26
gray25

grey3
grey2
grey1
grey0
grey
greenyellow
green4
green3
green2
green1
green
gray100
gray99
gray98
gray97
gray96
gray95
gray94
gray93
gray92
gray91

grey69
grey68
grey67
grey66
grey65
grey64
grey63
grey62
grey61
grey60
grey59
grey58
grey57
grey56
grey55
grey54
grey53
grey52
grey51
grey50
grey49

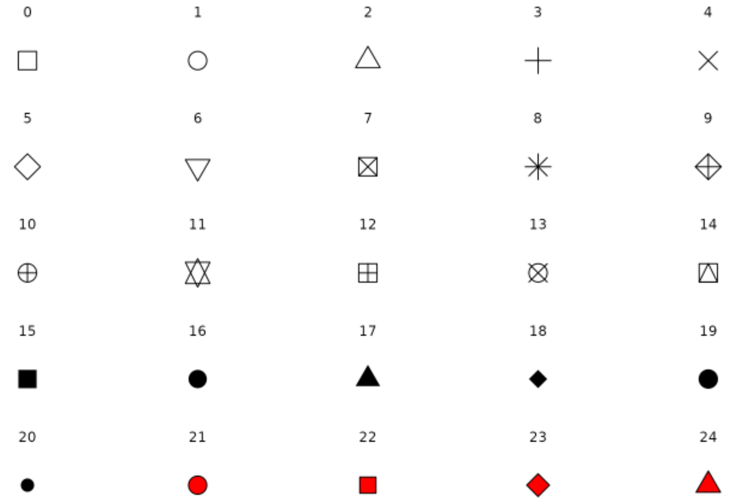
lemonchiffon2
lemonchiffon1
lemonchiffon
lawngreen
lavenderblush4
lavenderblush3
lavenderblush2
lavenderblush1
lavenderblush
lavender
khaki4
khaki3
khaki2
khaki1
khaki
ivory4
ivory3
ivory2
ivory1
ivory
indianred4

mediumorchid
mediumblue
mediumaquamarine
maroon4
maroon3
maroon2
maroon1
maroon
magenta4
magenta3
magenta2
magenta1
magenta
linen
limegreen
lightyellow4
lightyellow3
lightyellow2
lightyellow1
lightyellow
lightsteelblue4

palevioletred4
palevioletred3
palevioletred2
palevioletred1
palevioletred
paleturquoise4
paleturquoise3
paleturquoise2
paleturquoise1
paleturquoise
palegreen4
palegreen3
palegreen2
palegreen1
palegreen
palegoldenrod
orchid4
orchid3
orchid2
orchid1
orchid

slateblue
skyblue4
skyblue3
skyblue2
skyblue1
skyblue
sienna4
sienna3
sienna2
sienna1
sienna
seashell4
seashell3
seashell2
seashell1
seashell
seagreen4
seagreen3
seagreen2
seagreen1
seagreen

yellowgreen
yellow4
yellow3
yellow2
yellow1
yellow
whitesmoke
wheat4
wheat3
wheat2
wheat1
wheat
violetred4
violetred3
violetred2
violetred1
violetred
violet



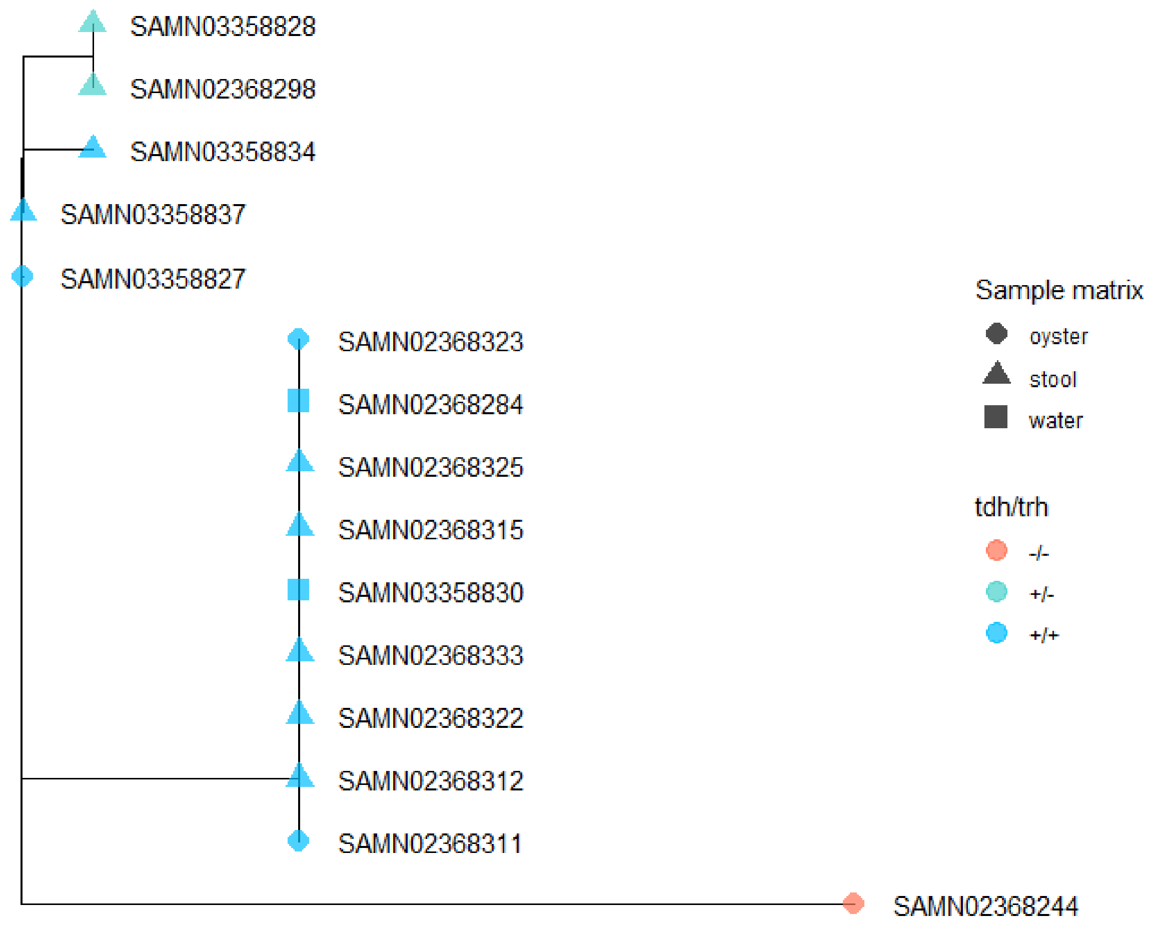
To find codes for ggplot2 colors and shapes, go to:
http://sape.inf.usi.ch/quick-reference/ggplot2/colour
Or search "ggplot2 colors" in google

EXAMPLE 1

5. Use scale_manual to specific shape and color

```
gg_flip +
  geom_tiplab(offset = 0.0001)+
  geom_tippoint(aes(color = `tdh/trh`, shape = matrix),
                size = 4,
                alpha = 0.7)+
  scale_color_manual(values = c("+/+" = "deepskyblue1",
                                "+/-" = "mediumturquoise",
                                "-/-" = "coral1"))+
  scale_shape_manual(values = c("oyster" = 16,
                                "stool" = 17,
                                "water" = 15),
                     name = "Sample matrix")
```

EXAMPLE 1

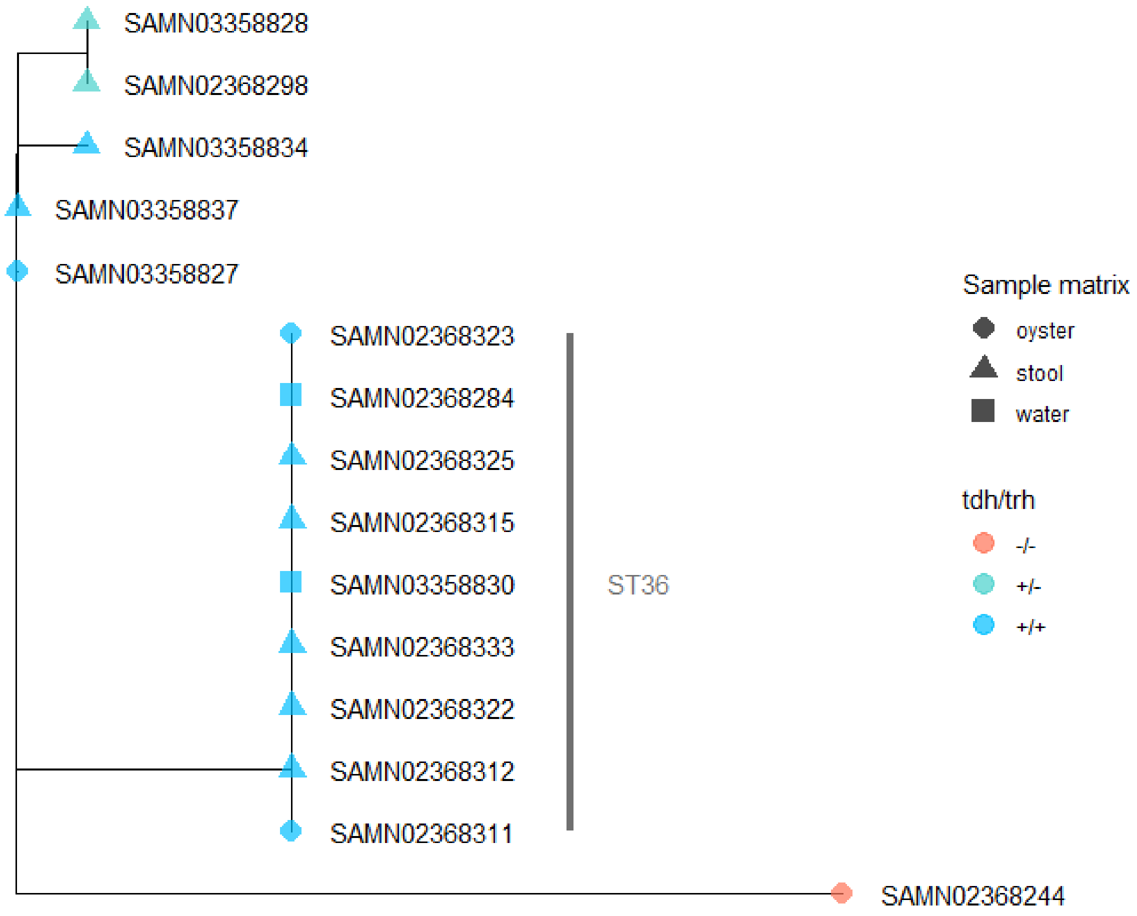

SAMN03358828
SAMN02368298
SAMN03358834
SAMN03358837
SAMN03358827
SAMN02368323
SAMN02368284
SAMN02368325
SAMN02368315
SAMN03358830
SAMN02368333
SAMN02368322
SAMN02368312
SAMN02368311
SAMN02368244
Sample matrix
oyster
stool
water
tdh/trh
-/-
+/-
+/+

EXAMPLE 1

6. Add a clade label

```
gg_flip +
  geom_tiplab(offset = 0.0001)+
  geom_tippoint(aes(color = `tdh/trh`, shape = matrix),
                size = 4,
                alpha = 0.7)+
  scale_color_manual(values = c("+/+" = "deepskyblue1",
                                "+/-" = "mediumturquoise",
                                "-/-" = "coral1"))+
  scale_shape_manual(values = c("oyster" = 16,
                                "stool" = 17,
                                "water" = 15),
                     name = "Sample matrix") +
  geom_cladelab(node = 17, label = "ST36",
                offset = 0.0008,
                barsize = 1.5,
                barcolor = 'grey44',
                textcolor = 'grey44',
                offset.text = 0.0001)
```

EXAMPLE 1

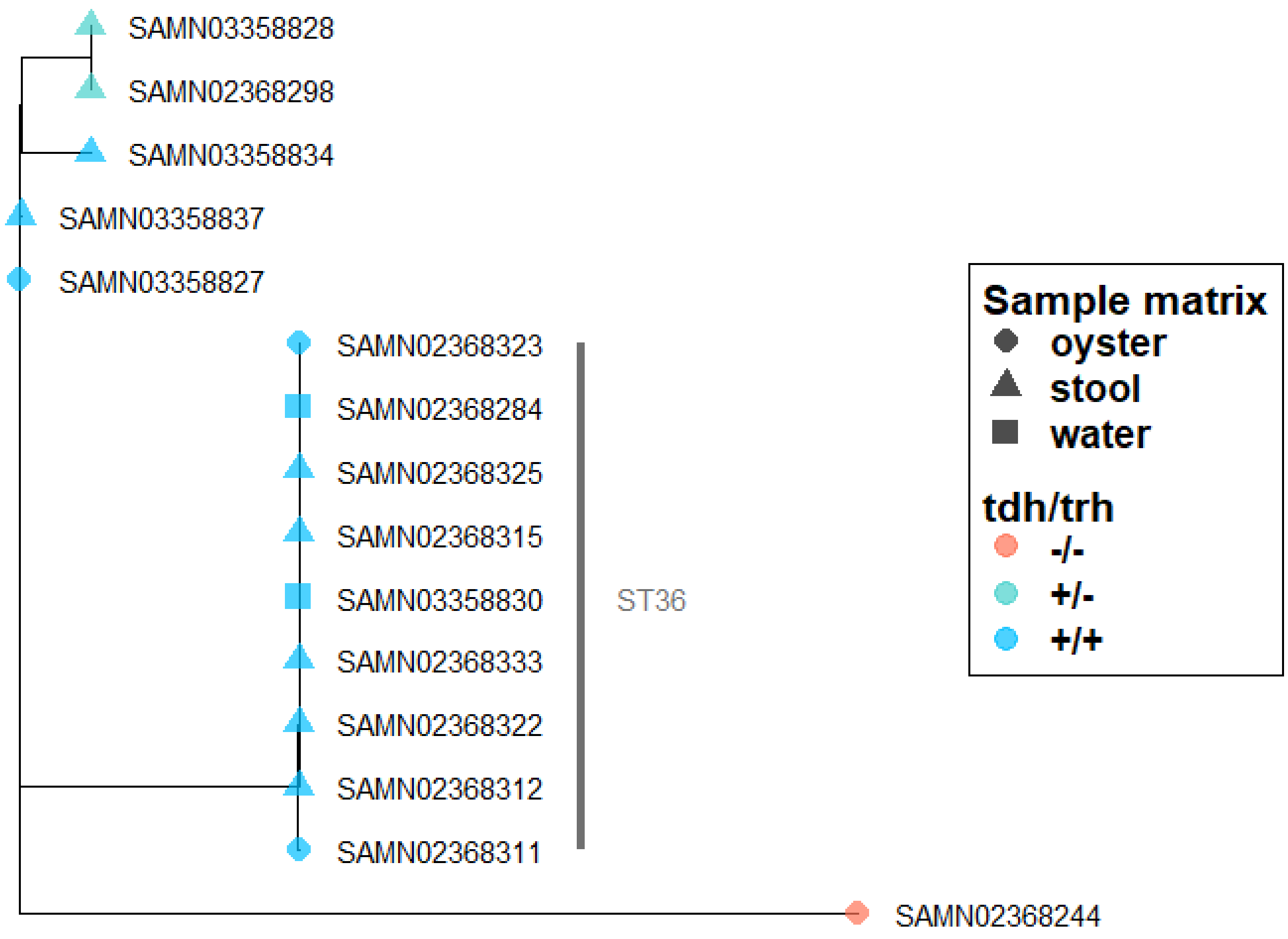

SAMN03358828
SAMN02368298
SAMN03358834
SAMN03358837
SAMN03358827
SAMN02368323
SAMN02368284
SAMN02368325
SAMN02368315
SAMN03358830
SAMN02368333
SAMN02368322
SAMN02368312
SAMN02368311
ST36
SAMN02368244
Sample matrix
oyster
stool
water
tdh/trh
-/-
+/-
+/+

EXAMPLE 1

7. Use theme() to add final touches to the plot & save as new object

```
example1 <- gg_flip +
  geom_tiplab(offset = 0.0001)+
  geom_tippoint(aes(color = `tdh/trh`, shape = matrix),
                size = 4,
                alpha = 0.7)+
  scale_color_manual(values = c("+/+" = "deepskyblue1",
                                "+/-" = "mediumturquoise",
                                "-/-" = "coral1"))+
  scale_shape_manual(values = c("oyster" = 16,
                                "stool" = 17,
                                "water" = 15),
                     name = "Sample matrix") +
  geom_cladelab(node = 17, label = "ST36",
                offset = 0.0008,
                barsize = 1.5,
                barcolor = 'grey44',
                textcolor = 'grey44',
                offset.text = 0.0001)+
  theme(legend.text = element_text(size = 14, face = "bold"),
        legend.title = element_text(size = 16, face = "bold"),
        legend.spacing.y = unit(0, "mm"),
        panel.border = element_blank(),
        aspect.ratio = 1, axis.text = element_text(colour = 1, size = 12),
        legend.background = element_blank(),
        legend.box.background = element_rect(colour = "black"))
```

EXAMPLE 1


SAMN03358828
SAMN02368298
SAMN03358834
SAMN03358837
SAMN03358827
SAMN02368323
SAMN02368284
SAMN02368325
SAMN02368315
SAMN03358830
SAMN02368333
SAMN02368322
SAMN02368312
SAMN02368311
ST36
SAMN02368244
Sample matrix
oyster
stool
water
tdh/trh
-/-
+/-
+/+

EXAMPLE 1

8. Export final tree into output directory

```
ggsave("output/example1.jpeg", example1, dpi = 300)
```

Additional arguments

- **Scale, width, height, units, dpi**

Details:

- **Currently supports pdf, jpeg, tiff, png and more…**

Learn more about ggsave() in r documentation:
https://www.rdocumentation.org/packages/ggplot2/versions/0.9.0/topics/ggsave

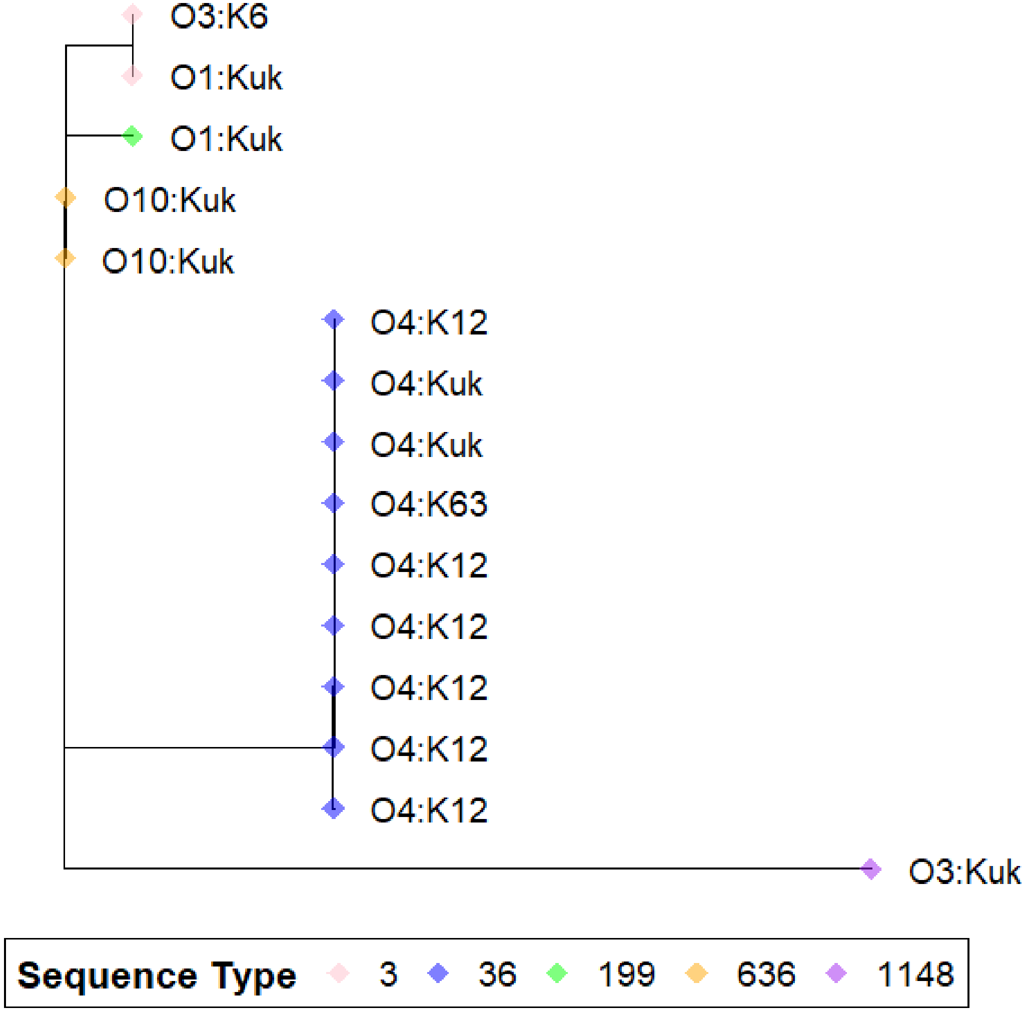
EXERCISE 1

Let’s apply everything we’ve learned so far:

1) Create a tree using the serovar as the tip labels
2) Set your own color of the nodes based on sequence type (st) and change their shape
3) Change the font size of the tip labels and legend
4) (Bonus) Move the legend to the bottom of the tree (hint, use help page)

Time: 10 minutes

EXERCISE 1


O3:K6
O1:Kuk
O1:Kuk
O10:Kuk
O10:Kuk
O4:K12
O4:Kuk
O4:Kuk
O4:K63
O4:K12
O4:K12
O4:K12
O4:K12
O4:K12
O3:Kuk
Sequence Type
3
36
199
636
1148

EXERCISE 1

Let’s apply everything we’ve learned so far:

1) Create a tree using the serovar as the tip labels

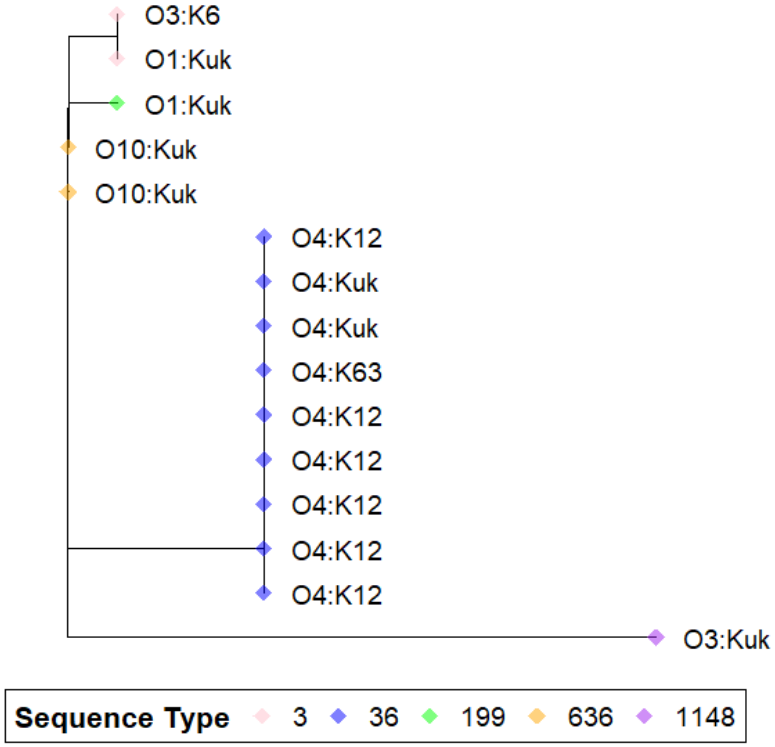
2) Set your own color of the nodes based on sequence type (st) and change their shape

3) Change the font size of the tip labels and legend

4) (Bonus) Move the legend to the bottom of the tree

Time: 10 minutes

HINT: use AS.FACTOR() around a continuous variable to read as a discrete scale



To find codes for ggplot2 colors and shapes, go to:
http://sape.inf.usi.ch/quick-reference/ggplot2/colour
Or search “ggplot2 colors” in google

DEMONSTRATION

EXERCISE 1 ANSWER

```
exercise1 <- gg_flip +
  geom_tiplab(aes(label = serovar),
              offset = 0.0001,
              size = 5)+
  geom_tippoint(aes(color = as.factor(st)),
                shape = 18,
                size = 4,
                alpha = 0.5)+
  scale_color_manual(values =
    c("pink", "blue", "green", "orange", "purple"),
                     name = "Sequence Type")+
  theme(legend.text = element_text(size = 14),
        legend.title = element_text(size = 16, face = "bold"),
        legend.position = "bottom",
        legend.spacing = unit(0, "cm"),
        panel.border = element_blank(),
        aspect.ratio = 1, axis.text = element_text(colour = 1,
         size = 12),
        legend.background = element_blank(),
        legend.box.background = element_rect(colour = "black"))
```

FOLLOW ALONG IN RSCRIPT

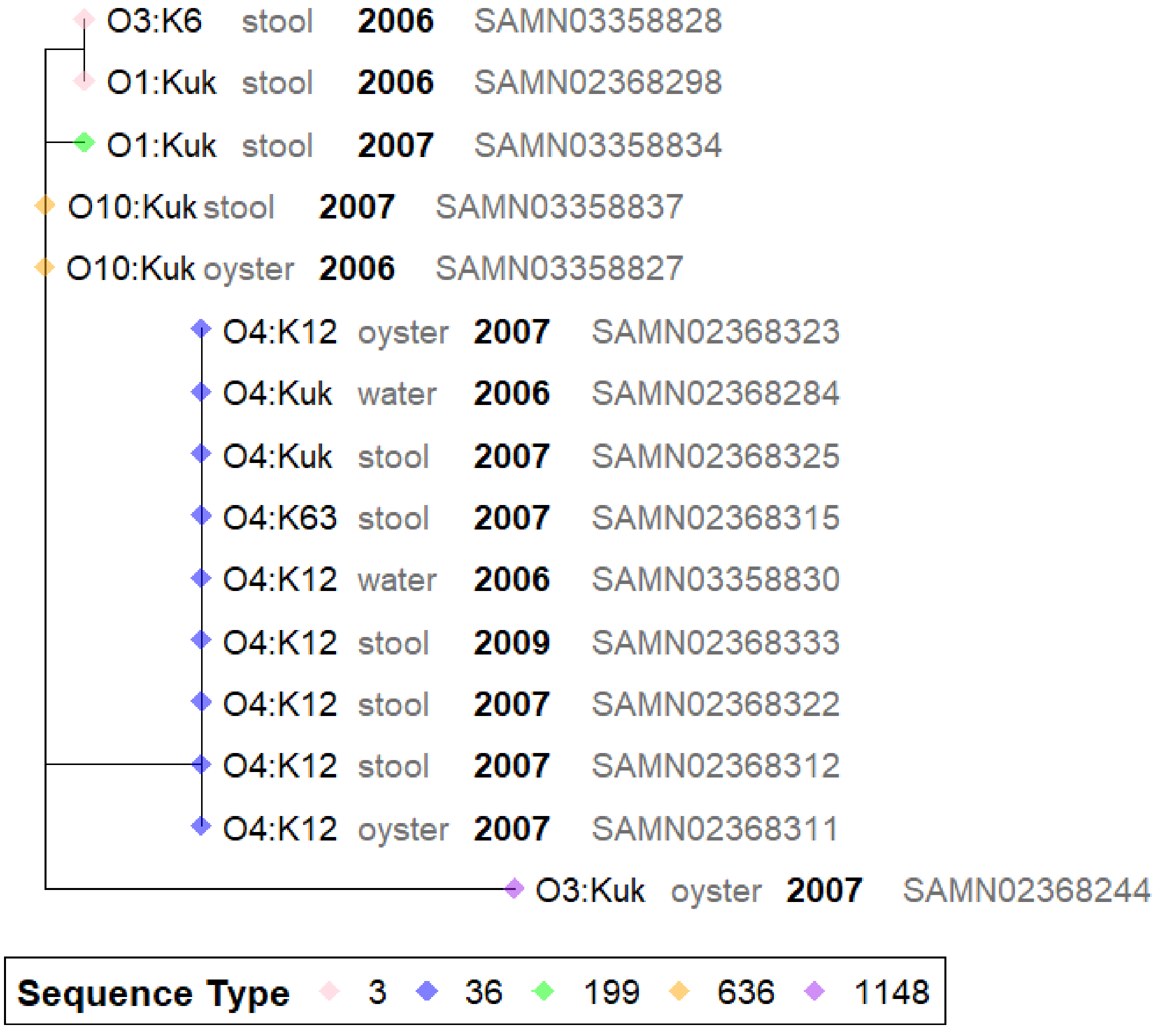
LAYERS FOR TREE ANNOTATION

Add geom_tiplab layer onto the tree from exercise 1

```
exercise1 +
  geom_tiplab(aes(label = matrix),
              offset = 0.0006,
              size = 5,
              color = "grey44")
```

You can continue to add layers, while changing the offset each time

LAYERS FOR TREE ANNOTATION



O3:K6 stool **2006** SAMN03358828

O1:Kuk stool **2006** SAMN02368298

O1:Kuk stool **2007** SAMN03358834

O10:Kuk stool **2007** SAMN03358837

O10:Kuk oyster **2006** SAMN03358827

O4:K12 oyster **2007** SAMN02368323

O4:Kuk water **2006** SAMN02368284

O4:Kuk stool **2007** SAMN02368325

O4:K63 stool **2007** SAMN02368315

O4:K12 water **2006** SAMN03358830

O4:K12 stool **2009** SAMN02368333

O4:K12 stool **2007** SAMN02368322

O4:K12 stool **2007** SAMN02368312

O4:K12 oyster **2007** SAMN02368311

O3:Kuk oyster **2007** SAMN02368244

Sequence Type ◆ 3 ◆ 36 ◆ 199 ◆ 636 ◆ 1148

HIGHLIGHTING CLADES

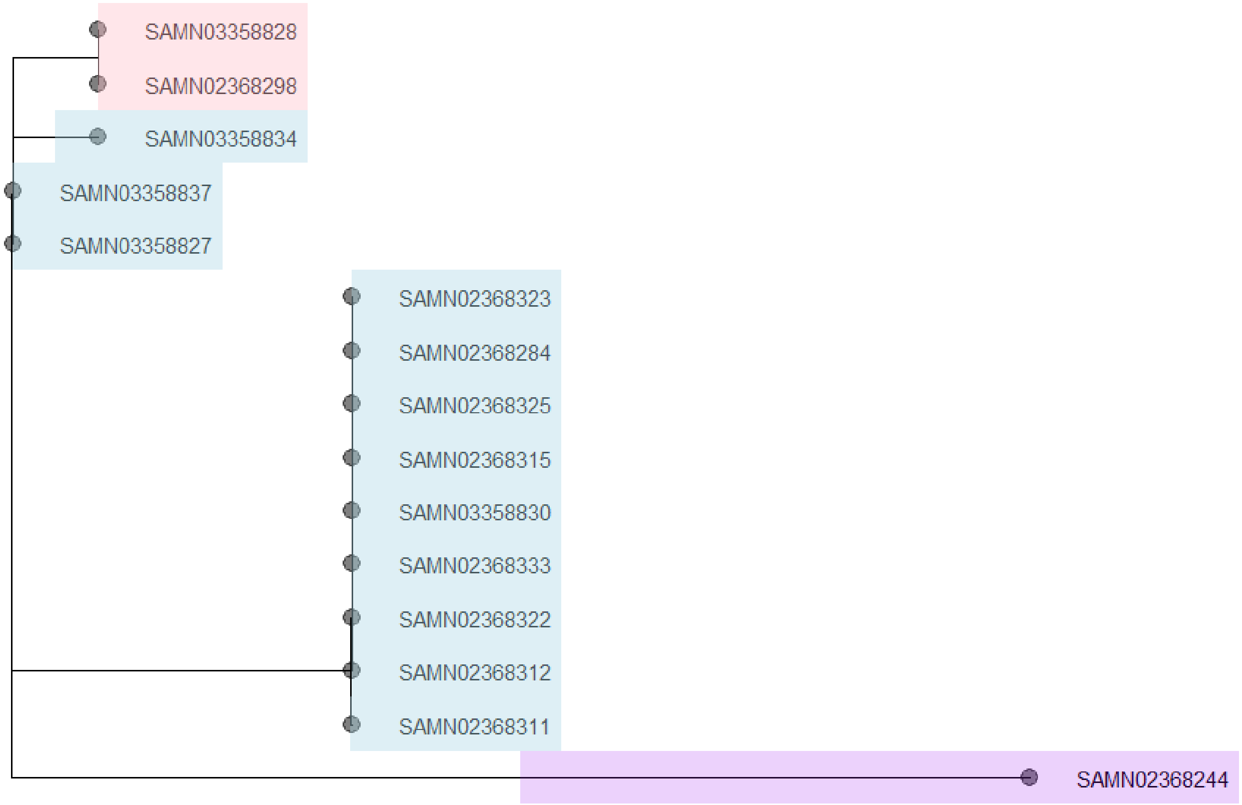
1. By specifying node

```
gg_flip +
  geom_tiplab(offset = 0.0001)+
  geom_tippoint(color = 'black',
                size = 4,
                alpha = 0.5)+
  geom_hilight(node = c(11,12),
               fill = "pink",
               alpha = 0.4,
               extend = 0.0005)
```

2. By subsetting by a condition

```
gg_flip +
  geom_tiplab(offset = 0.0001)+
  geom_tippoint(color = 'black',
                size = 4,
                alpha = 0.5)+
  geom_hilight(mapping=
aes(subset = wg_cluster %in% 1),
               fill = "pink",
               alpha = 0.4,
               extend = 0.0005)
```

HIGHLIGHTING CLADES



MORE EXAMPLES !

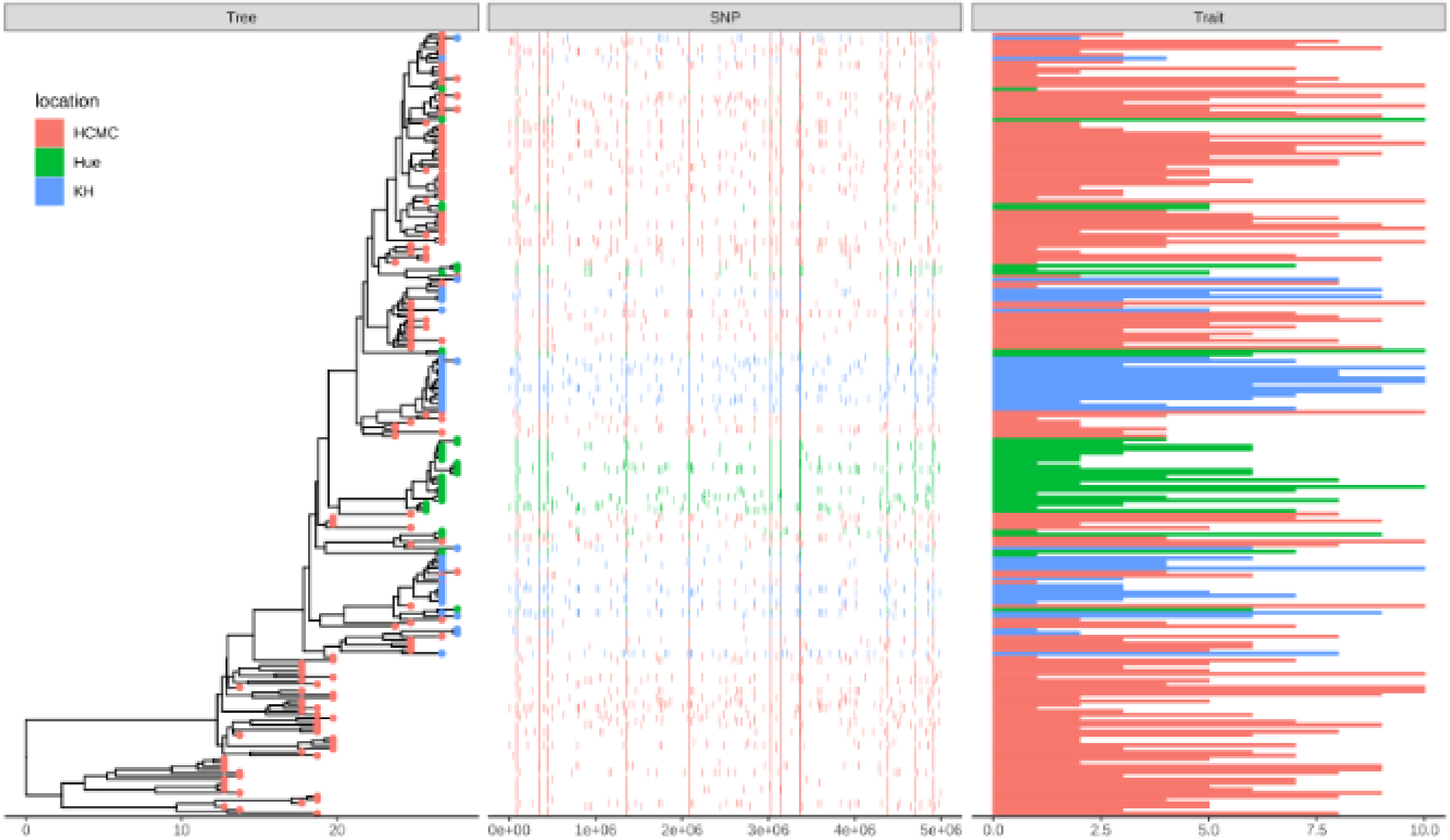


FIGURE 7.2: **Example of plotting SNP and trait data.** The 'location' information was attached to the tree and used to color tip symbols (Tree panel), and other datasets. SNP and Trait data were visualized as dot chart (SNP panel) and bar chart (Trait panel).

https://yulab-smu.top/treedata-book/chapter7.html

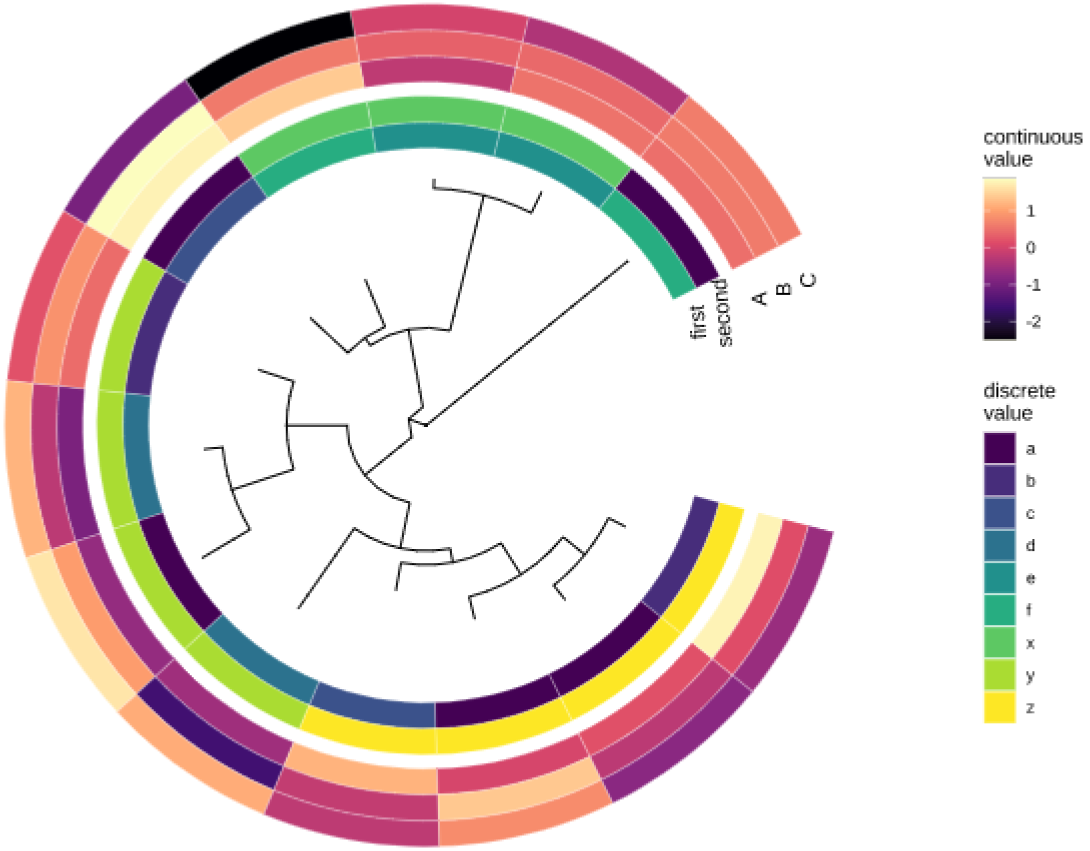


FIGURE 7.4: **Example of plotting matrix with `gheatmap()`**. A H3 influenza tree with a genotype table visualized as a heatmap (A). Tips were aligned and with a tailored *x*-axis for divergence times (tree) and genomic segments (heatmap) (B).

https://yulab-smu.top/treedata-book/chapter7.html

MICROBIAL GENOMICS

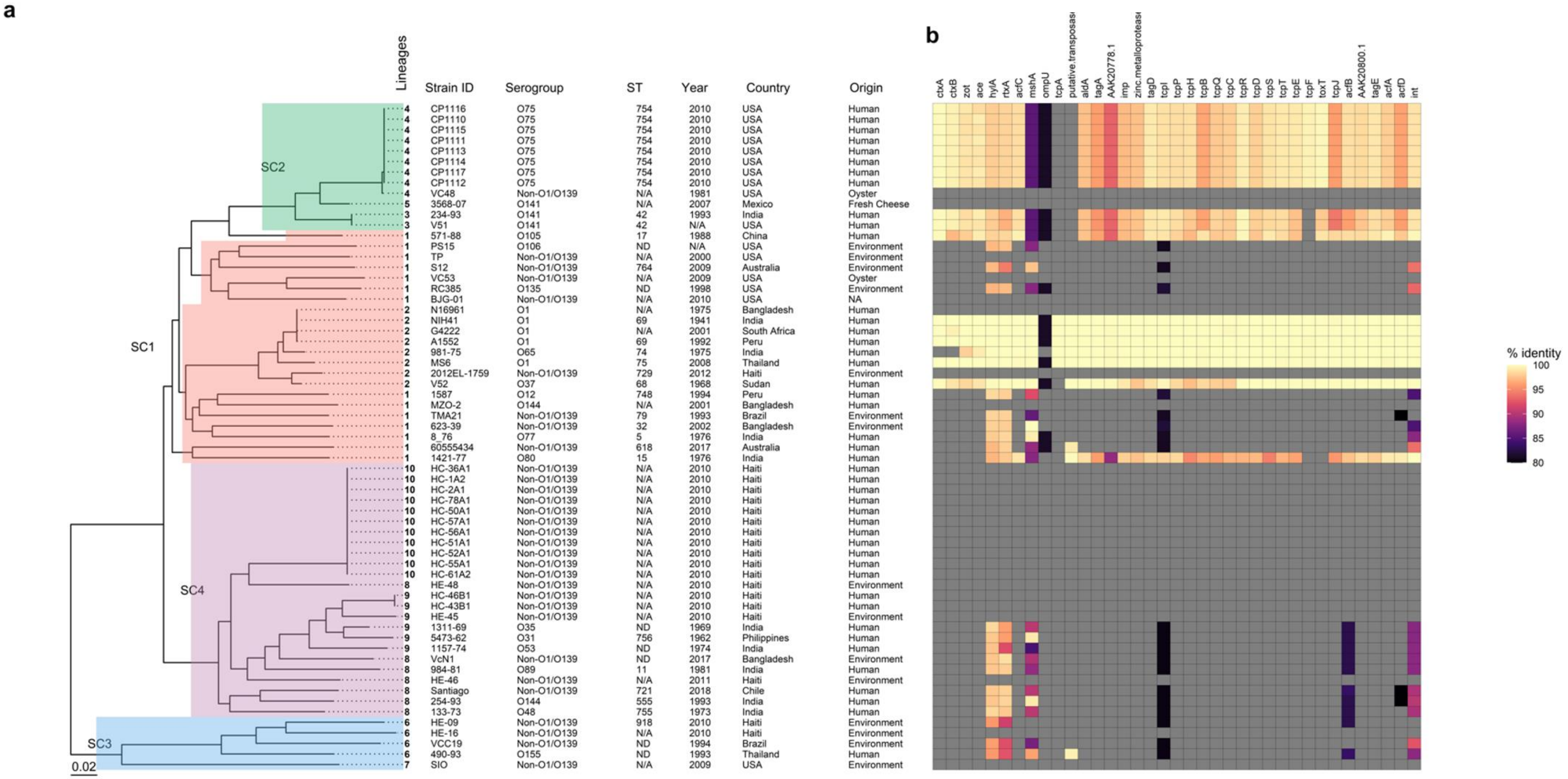
SHORT COMMUNICATION
Arteaga *et al.*, *Microbial Genomics*
DOI 10.1099/mgen.0.000340

MICROBIOLOGY SOCIETY

OPEN DATA OPEN MICROBIOLOGY

Genomic characterization of the non-O1/non-O139 *Vibrio cholerae* strain that caused a gastroenteritis outbreak in Santiago, Chile, 2018

Mónica Arteaga[1]‡, Juliana Velasco[1]‡, Shelly Rodriguez[1], Maricel Vidal[2], Carolina Arellano[3], Francisco Silva[4], Leandro J. Carreño[5,6], Roberto Vidal[3,6,*] and David A. Montero[3,5,*]



Lineages	Strain ID	Serogroup	ST	Year	Country	Origin
4	CP1116	O75	754	2010	USA	Human
4	CP1110	O75	754	2010	USA	Human
4	CP1115	O75	754	2010	USA	Human
4	CP1111	O75	754	2010	USA	Human
4	CP1113	O75	754	2010	USA	Human
4	CP1114	O75	754	2010	USA	Human
4	CP1117	O75	754	2010	USA	Human
4	CP1112	O75	754	2010	USA	Human
4	VC48	Non-O1/O139	N/A	1981	USA	Oyster
5	3568-07	O141	N/A	2007	Mexico	Fresh Cheese
3	234-93	O141	42	1993	India	Human
3	V51	O141	42	N/A	USA	Human
1	571-88	O105	17	1988	China	Human
1	PS15	O106	ND	N/A	USA	Environment
1	TP	Non-O1/O139	N/A	2000	USA	Environment
1	S12	Non-O1/O139	764	2009	Australia	Environment
1	VC53	Non-O1/O139	N/A	2009	USA	Oyster
1	RC385	O135	ND	1998	USA	Environment
1	BJG-01	Non-O1/O139	N/A	2010	USA	NA
2	N16961	O1	N/A	1975	Bangladesh	Human
2	NIH41	O1	69	1941	India	Human
2	G4222	O1	N/A	2001	South Africa	Human
2	A1552	O1	69	1992	Peru	Human
2	981-75	O65	74	1975	India	Human
2	MS6	O1	75	2008	Thailand	Human
2	2012EL-1759	Non-O1/O139	729	2012	Haiti	Environment
2	V52	O37	68	1968	Sudan	Human
1	1587	O12	748	1994	Peru	Human
1	MZO-2	O144	N/A	2001	Bangladesh	Human
1	TMA21	Non-O1/O139	79	1993	Brazil	Environment
1	623-39	Non-O1/O139	32	2002	Bangladesh	Environment
1	8_76	O77	5	1976	India	Human
1	60555434	Non-O1/O139	618	2017	Australia	Human
1	1421-77	O80	15	1976	India	Human
10	HC-36A1	Non-O1/O139	N/A	2010	Haiti	Human
10	HC-1A2	Non-O1/O139	N/A	2010	Haiti	Human
10	HC-2A1	Non-O1/O139	N/A	2010	Haiti	Human
10	HC-78A1	Non-O1/O139	N/A	2010	Haiti	Human
10	HC-50A1	Non-O1/O139	N/A	2010	Haiti	Human
10	HC-57A1	Non-O1/O139	N/A	2010	Haiti	Human
10	HC-56A1	Non-O1/O139	N/A	2010	Haiti	Human
10	HC-51A1	Non-O1/O139	N/A	2010	Haiti	Human
10	HC-52A1	Non-O1/O139	N/A	2010	Haiti	Human
10	HC-55A1	Non-O1/O139	N/A	2010	Haiti	Human
10	HC-61A2	Non-O1/O139	N/A	2010	Haiti	Human
8	HE-48	Non-O1/O139	N/A	2010	Haiti	Environment
9	HC-46B1	Non-O1/O139	N/A	2010	Haiti	Human
9	HC-43B1	Non-O1/O139	N/A	2010	Haiti	Human
9	HE-45	Non-O1/O139	N/A	2010	Haiti	Environment
9	1311-69	O35	ND	1969	India	Human
9	5473-62	O31	756	1962	Philippines	Human
9	1157-74	O53	ND	1974	India	Human
8	VcN1	Non-O1/O139	ND	2017	Bangladesh	Environment
8	984-81	O89	11	1981	India	Human
8	HE-46	Non-O1/O139	N/A	2011	Haiti	Environment
8	Santiago	Non-O1/O139	721	2018	Chile	Human
8	254-93	O144	555	1993	India	Human
8	133-73	O48	755	1973	India	Human
6	HE-09	Non-O1/O139	918	2010	Haiti	Environment
6	HE-16	Non-O1/O139	N/A	2010	Haiti	Environment
6	VCC19	Non-O1/O139	ND	1994	Brazil	Environment
6	490-93	O155	ND	1993	Thailand	Human
7	SIO	Non-O1/O139	N/A	2009	USA	Environment

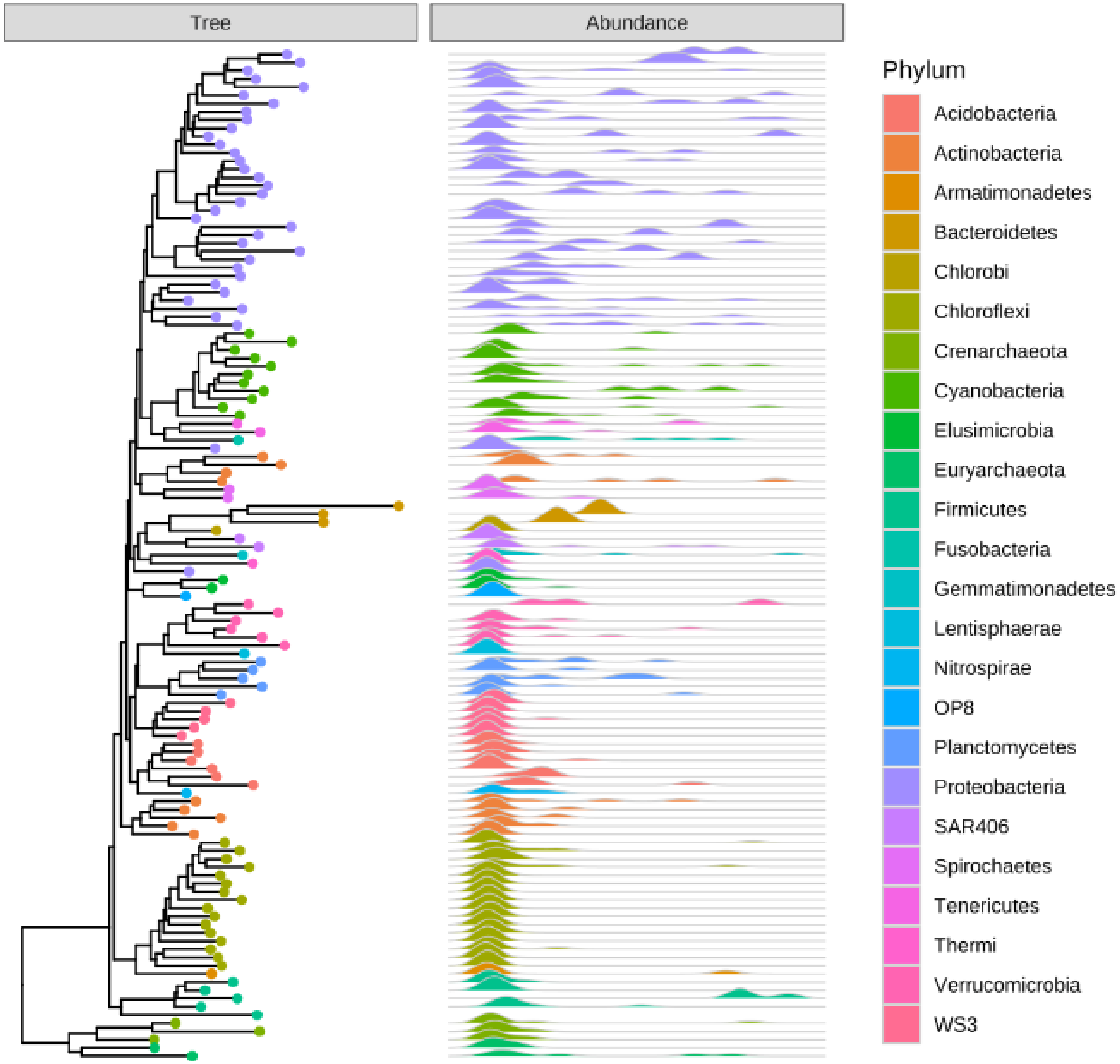


FIGURE 9.4: **Phylogenetic tree with OTU abundance densities.** Tips were colored by Phylum, and the corresponding abundances across different samples were visualized as density ridgelines and sorted according to the tree structure.

https://yulab-smu.top/treedata-book/chapter7.html

HELPFUL RESOURCES

Data integration, manipulation and visualization of phylogenetic trees:

https://yulab-smu.top/treedata-book/index.html

ggtree github:

https://github.com/YuLab-SMU/ggtree

Enhanced annotation practice:

http://www.randigriffin.com/2017/05/11/primate-phylogeny-ggtree.html

ggtreeExtra:

https://yulab-smu.top/treedata-book/chapter10.html
https://github.com/YuLab-SMU/ggtreeExtra

NEXT WEEK…

Bring your own data set!

THANK YOU FOR ATTENDING!

Please make sure to fill out the Exit Survey at https://docs.google.com/forms/d/e/1FAIpQLScSMU-AOMMru9CUz-UP1pJhrz0npverR6PpDcvF6jdrw-QDSA/viewform We value your feedback!

More questions? Please email us at mmid.bioinformatics.workshop@gmail.com or post them to the workshop slack channel