

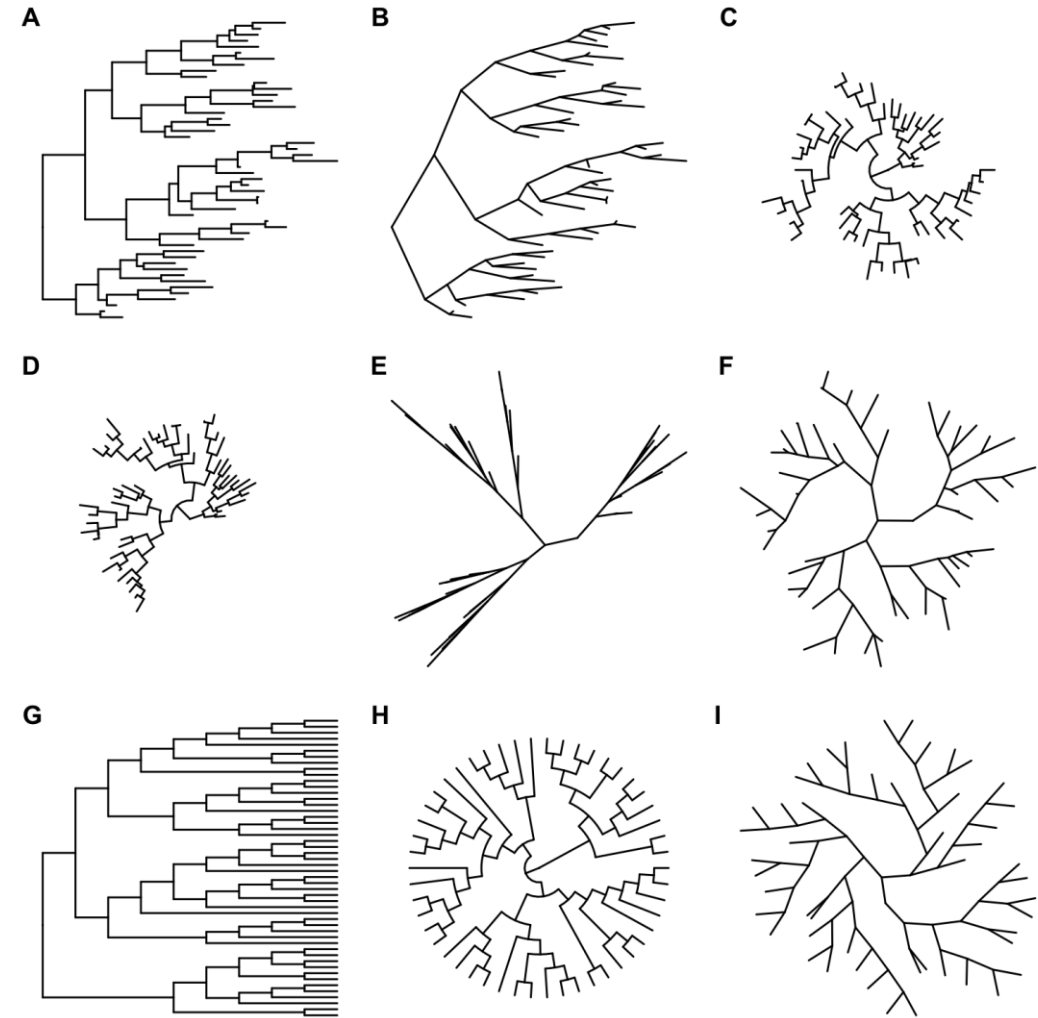
— / Ǻ ǻ ǻ ǻ —

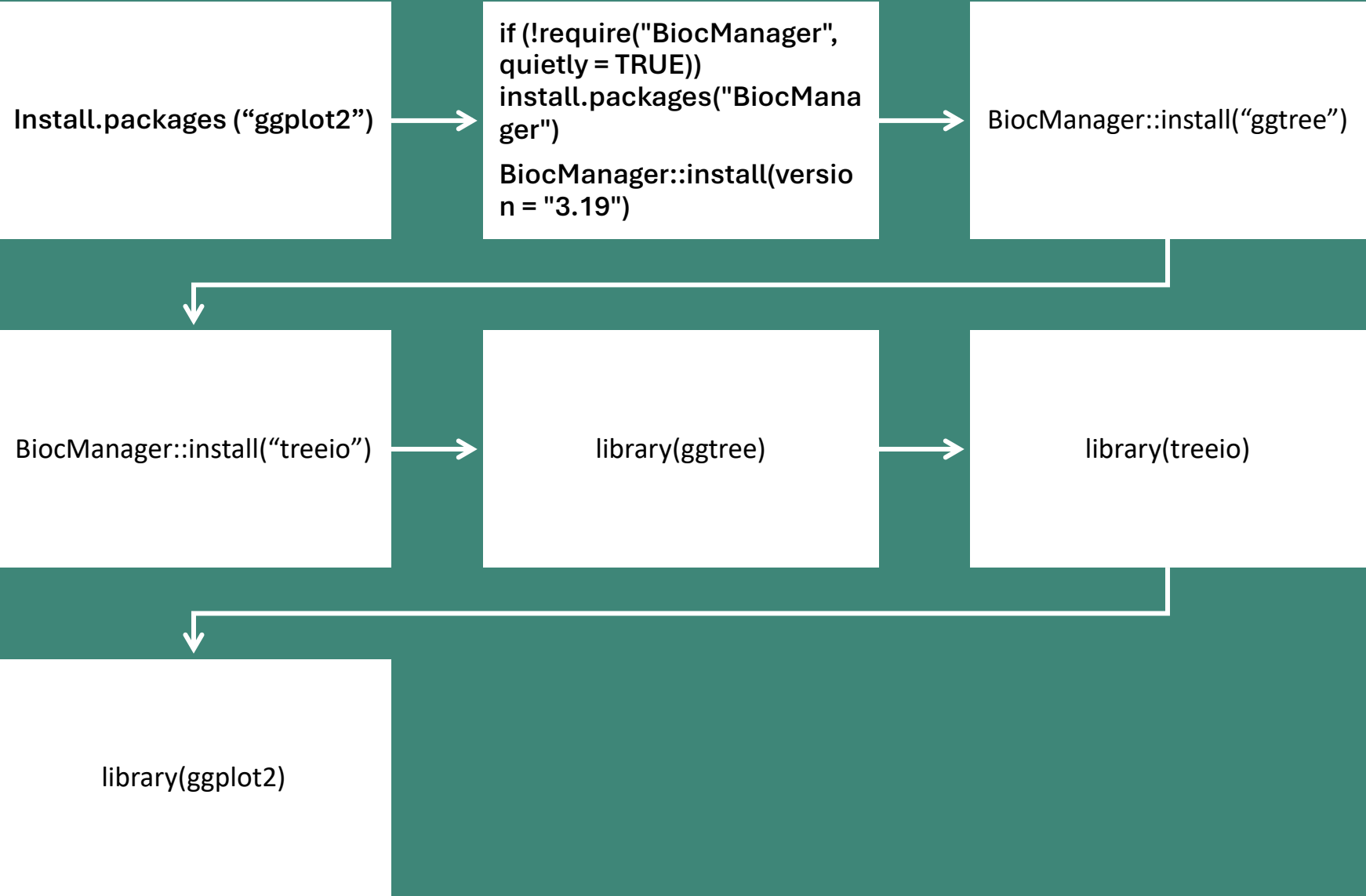
o o o i i

Ǻ ĩ E ǻ ĩ ĩ Ǻ v ù ģ ° ǻ ǻ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ  
L v ° ǻ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ

# Overview

- An R package designed for visualizing phylogenetic trees
- Extension of ggplot2
- Key Applications: Genomics, Evolutionary Biology and Phylogenetics





# Installation and Load

Example: Use *Salmonella enterica* datasets that were retrieved from a public database to make a phylogenetic tree using the Talbot Pipeline.

## Application

# Application Cont.

1) Retrieve the dataset from  
Pathogen Detection in NCBI  
(In this case, we chose  
*Salmonella enterica*)

2) Retrieve the data using  
the SRA accession number  
in HiPerGator

3) Use Flaq AMR pipeline to  
create gff files from the  
dataset

4) Use the gff files to create  
a Newick file with Talbot  
pipeline

5) Retrieve the Newick file  
that is generated and  
import it to R

6) Generate a phylogenetic  
tree using ggtree



# Application Cont.

sampleID	speciesID_mash	nearest_neighbor_mash	mash_distance	speciesID_kraken	kraken_percent	mist_scheme	mist_st	num_clean_reads	avg_readlength	avg_read_qual	est_coverage	num_contig	longest_contig	N50	L50	total_length	gc_content	annotated_cds
SRR240765	Salmonella_enterica	GCF_000623775.1	0.0011856	Salmonella_enterica	97.05	senterica	11	1027736	225.85	36.03	49.14	32	1508692	4E+05	3	4722943	52	4433
SRR253350	Salmonella_enterica	GCF_000623775.1	0.00060662	Salmonella_enterica	97.74	senterica	11	937312	221.7	36.44	44.04	26	1508692	4E+05	3	4717664	52	4420
SRR258436	Salmonella_enterica	GCF_000623775.1	0.000931334	Salmonella_enterica	97.29	senterica	11	1460314	222.61	36.64	69.75	55	634220	2E+05	7	4660154	52	4356
SRR607714	Salmonella_enterica	GCF_000623775.1	0.00213483	Salmonella_enterica	96.82	senterica	11	1131246	210.83	36.38	50.66	36	1215647	4E+05	4	4707886	52	4411
SRR607714	Salmonella_enterica	GCF_000623775.1	0.00343111	Salmonella_enterica	96.83	senterica	11	2255536	211.83	36.39	101.3	27	1508692	5E+05	3	4716521	52	4421
SRR635106	Salmonella_enterica	GCF_000623775.1	0.00213483	Salmonella_enterica	96.27	senterica	11	1506344	200.02	36.15	63.89	32	1024859	4E+05	4	4715769	52	4422
SRR635107	Salmonella_enterica	GCF_000623775.1	0.00157453	Salmonella_enterica	96.24	senterica	11	1124130	183.18	35.64	44.13	29	1024859	4E+05	4	4665490	52	4365



SRR2407650.gff  
SRR2533501.gff  
SRR2584382.gff  
SRR6077140.gff  
SRR6077143.gff  
SRR6351069.gff  
SRR6351075.gff



```
((SRR6077140:0.000000005,(SRR2584382:1.144617642,SRR6351075:0.579009679)0.376:0.000000005)0.996:0.150768689,  
(SRR6351069:0.089264090,SRR6077143:0.010586278)0.916:0.046706152,(SRR2407650:0.112240552,SRR2533501:0.066495  
582)0.853:0.030665257);
```

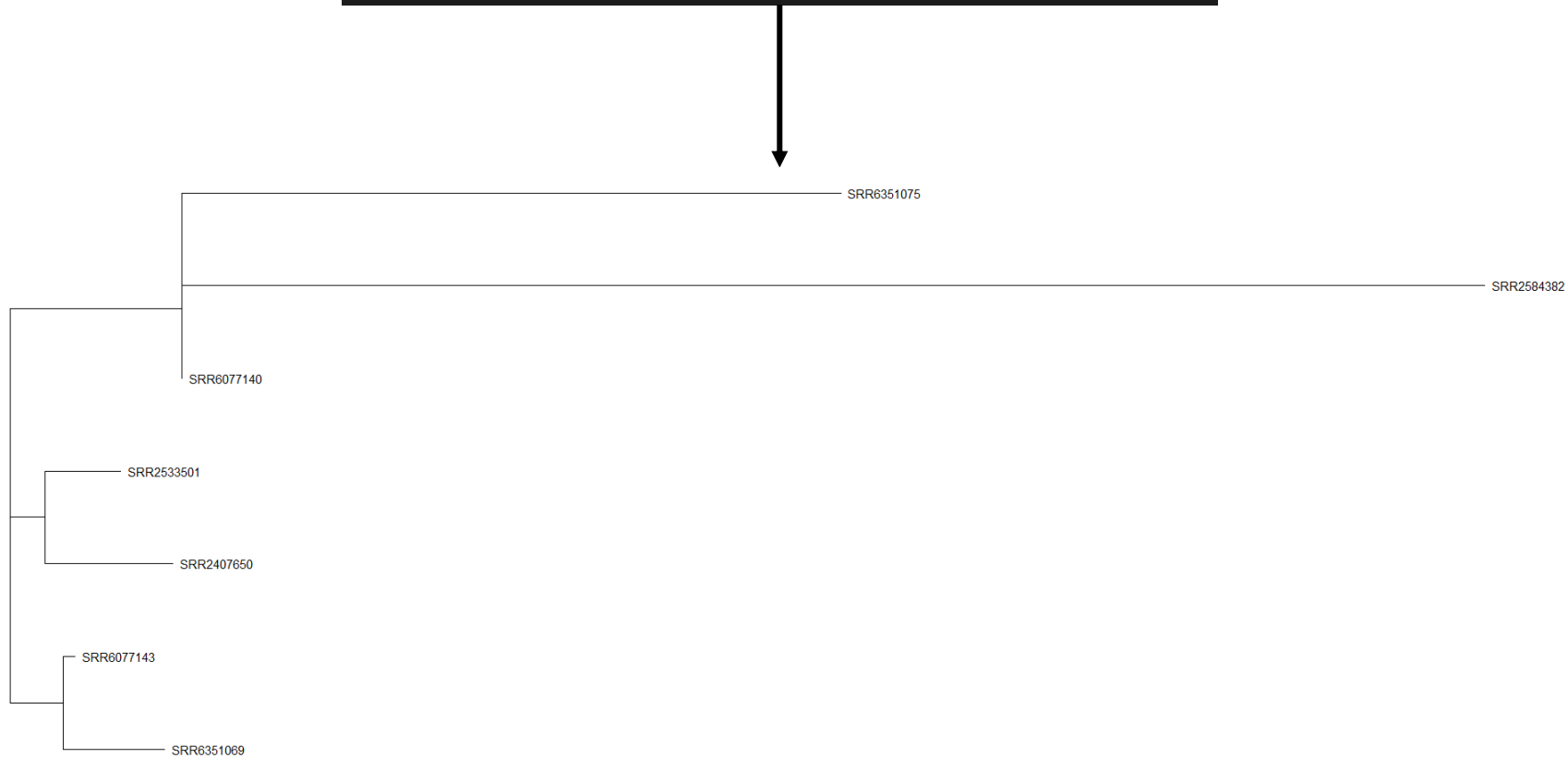
# Application Cont.



```
library(ggtree)
library(treeio)
library(ggplot2)

maintree <- read.tree("accessory_binary_genes.fa.newick")

ggtree(maintree)+geom_tiplab()
```



# Application Cont.



# Application Cont.

```
ggtree(maintree)+ labs(color="Samples")+ geom_tiplab(aes(color=label), size=3) + theme_tree2()
```

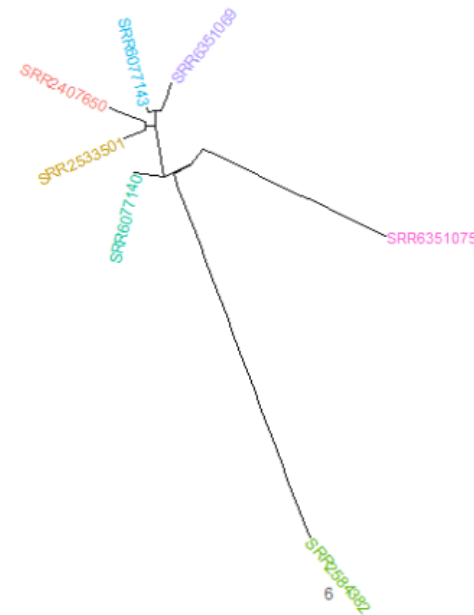


# Application Cont.

```
ggtree(maintree, layout="circular")+ labs(color="Samples")+ geom_tiplab(aes(color=label), size=3) + theme_tree2()
```

2

0



Samples

- SRR2407650
- SRR2533501
- SRR2584382
- SRR6077140
- SRR6077143
- SRR6351069
- SRR6351075

4

6





Understood the basic principles of ggtree



Installation and loading the ggtree packages in R



Successfully applied ggtree with real world dataset using Flaq AMR and Talbot Pipeline

## Conclusion



# Citation

1. U.S. National Library of Medicine. (n.d.). *Home - Pathogen Detection - NCBI*. National Center for Biotechnology Information.

<https://www.ncbi.nlm.nih.gov/pathogens/>

2. Yu, G. (n.d.). *Ggtree: Elegant graphics for phylogenetic tree visualization and annotation*. Guangchuang Yu. <https://guangchuangyu.github.io/ggtree-book/chapter-ggtree.html>



# Advanced Molecular Detection Southeast Region Bioinformatics

## Questions?

[bphl-sebioinformatics@flhealth.gov](mailto:bphl-sebioinformatics@flhealth.gov)

**Molly Mitchell, PhD**

Bioinformatician Supervisor

[Molly.Mitchell@flhealth.gov](mailto:Molly.Mitchell@flhealth.gov)

**Nikhil Reddy, MS**

Bioinformatician

[Nikhil.Yengala@flhealth.gov](mailto:Nikhil.Yengala@flhealth.gov)

**Sam Bernhoft, MPH**

Bioinformatician

[Samantha.bernhoft@flhealth.gov](mailto:Samantha.bernhoft@flhealth.gov)