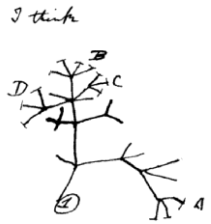
30 Sept 2024

ggtree

Advanced Molecular Detection
Southeast Region Bioinformatics



Updates

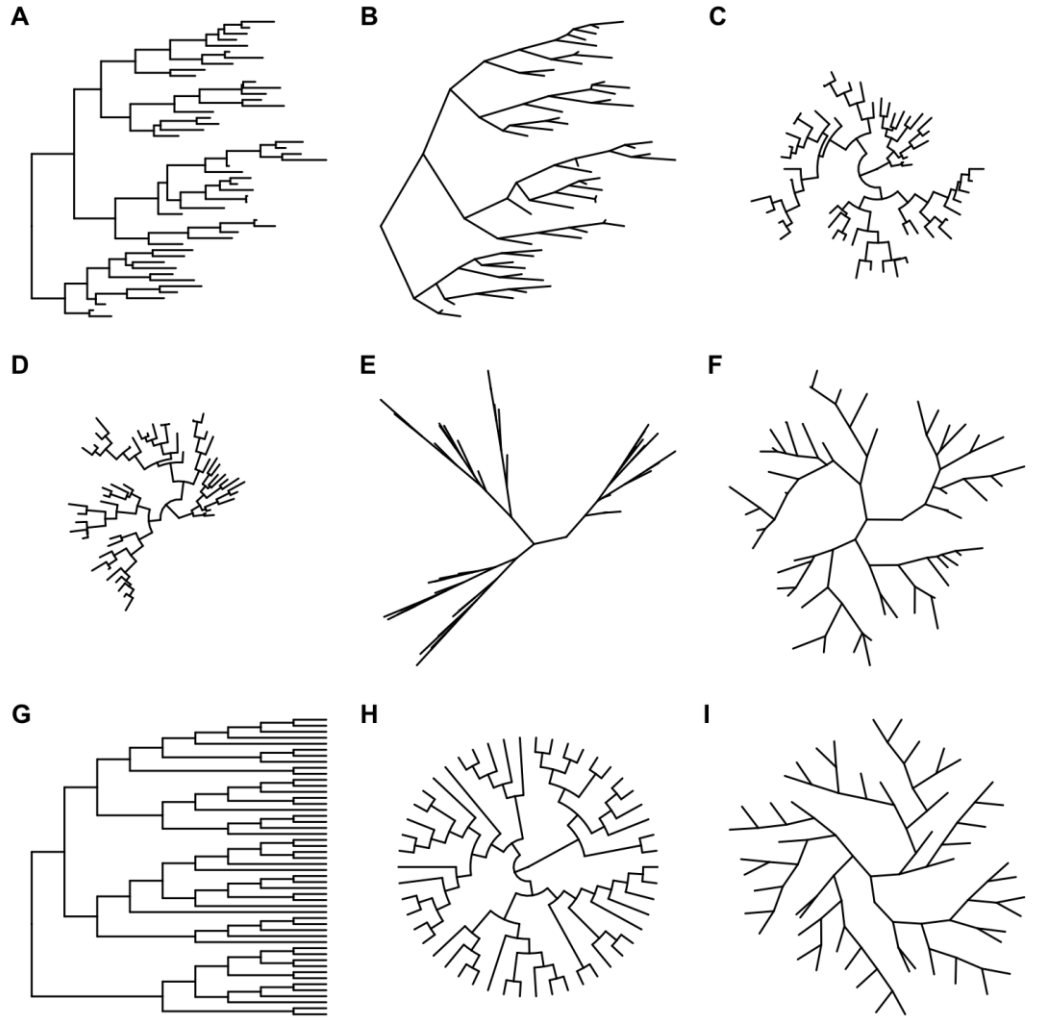
Office Hours

- **October 14, 2024 -** FLAQ-Antimicrobial Resistance
- **October 28, 2024 -** FLAQ-SARS-CoV-2

Advanced Molecular Detection
Southeast Region Bioinformatics

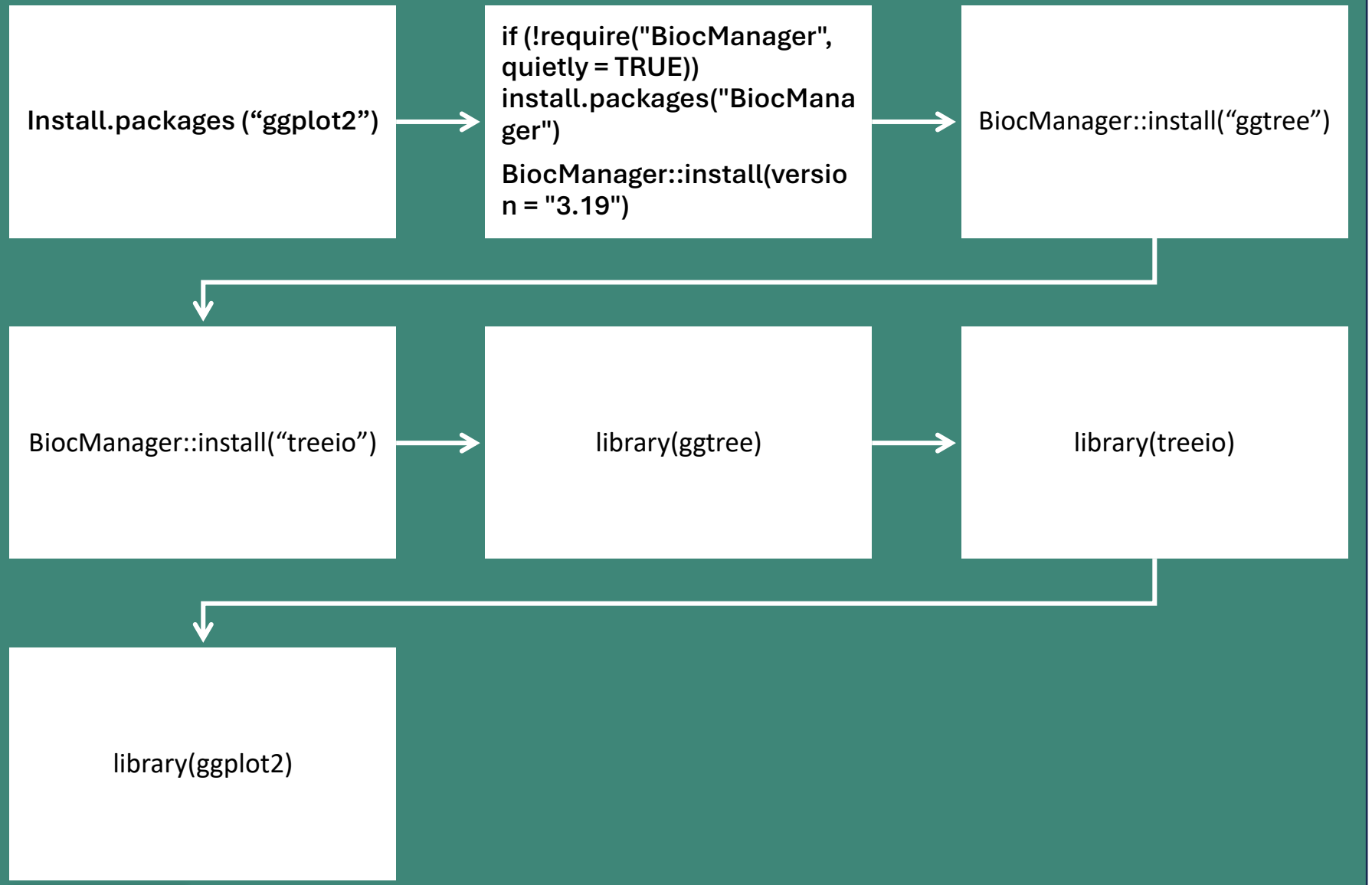
Overview

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



(1)

Advanced Molecular Detection
Southeast Region Bioinformatics



Installation and Load

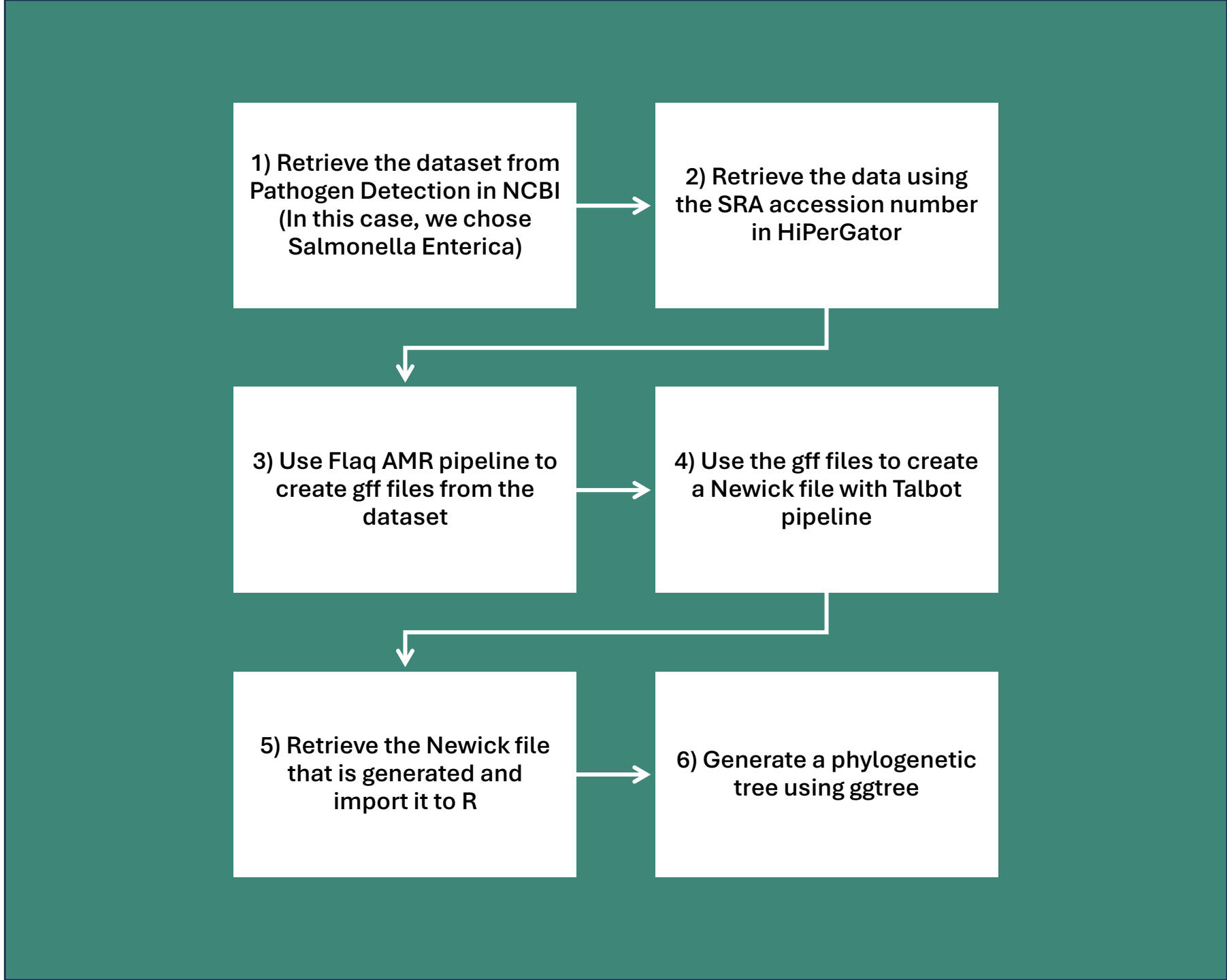
Advanced Molecular Detection
Southeast Region Bioinformatics

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

Application

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

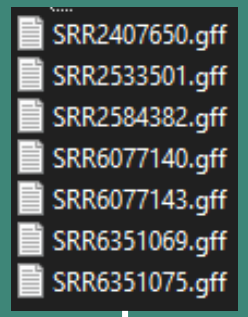


Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

sampleID	speciesID_mash	nearest_neighbor_mash	mash_distance	speciesID_kraken	kraken_percent	mlst_scheme	mlst_st	num_clean_reads	avg_readlength	avg_read_qual	est_coverage	num_contigs	longest_contig	N50	L50	total_length	gc_conten	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
SRR25335(	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
SRR25843(	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

Advanced Molecular Detection
Southeast Region Bioinformatics




```
((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);
```

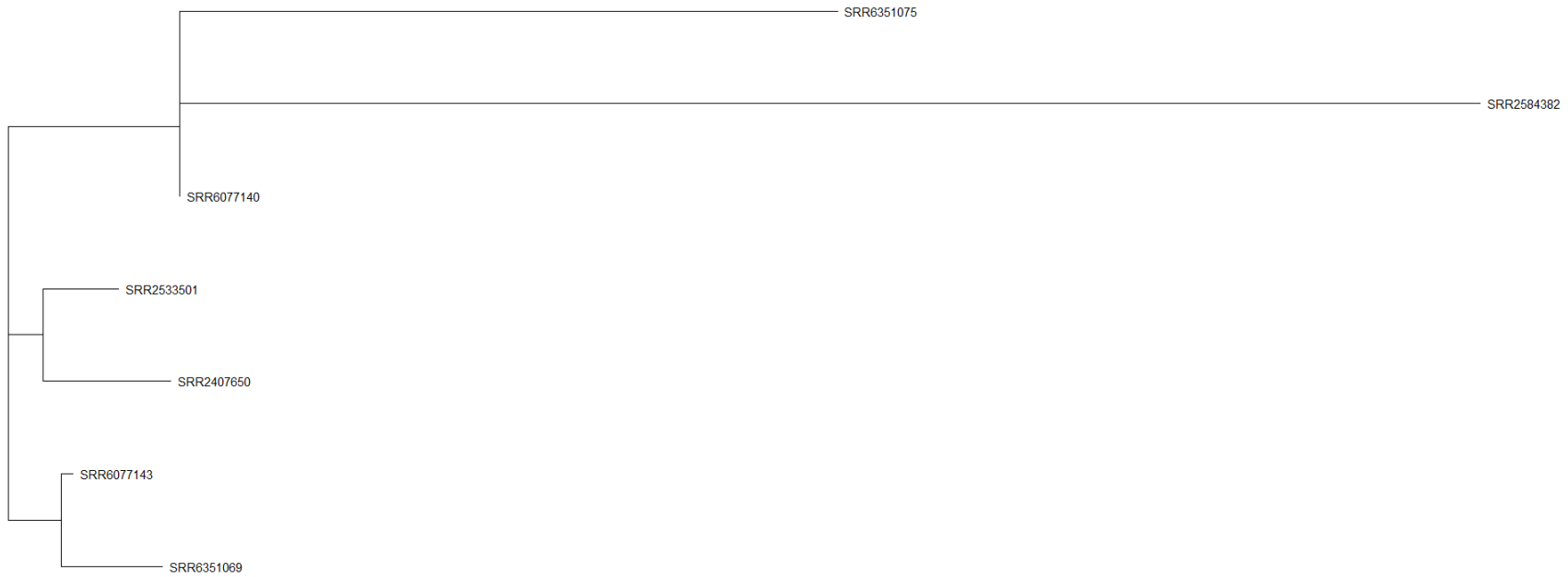
Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

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

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

ggtree(maintree)+geom_tiplab()
```



Application Cont.

Advanced Molecular Detection
Southeast Region Bioinformatics

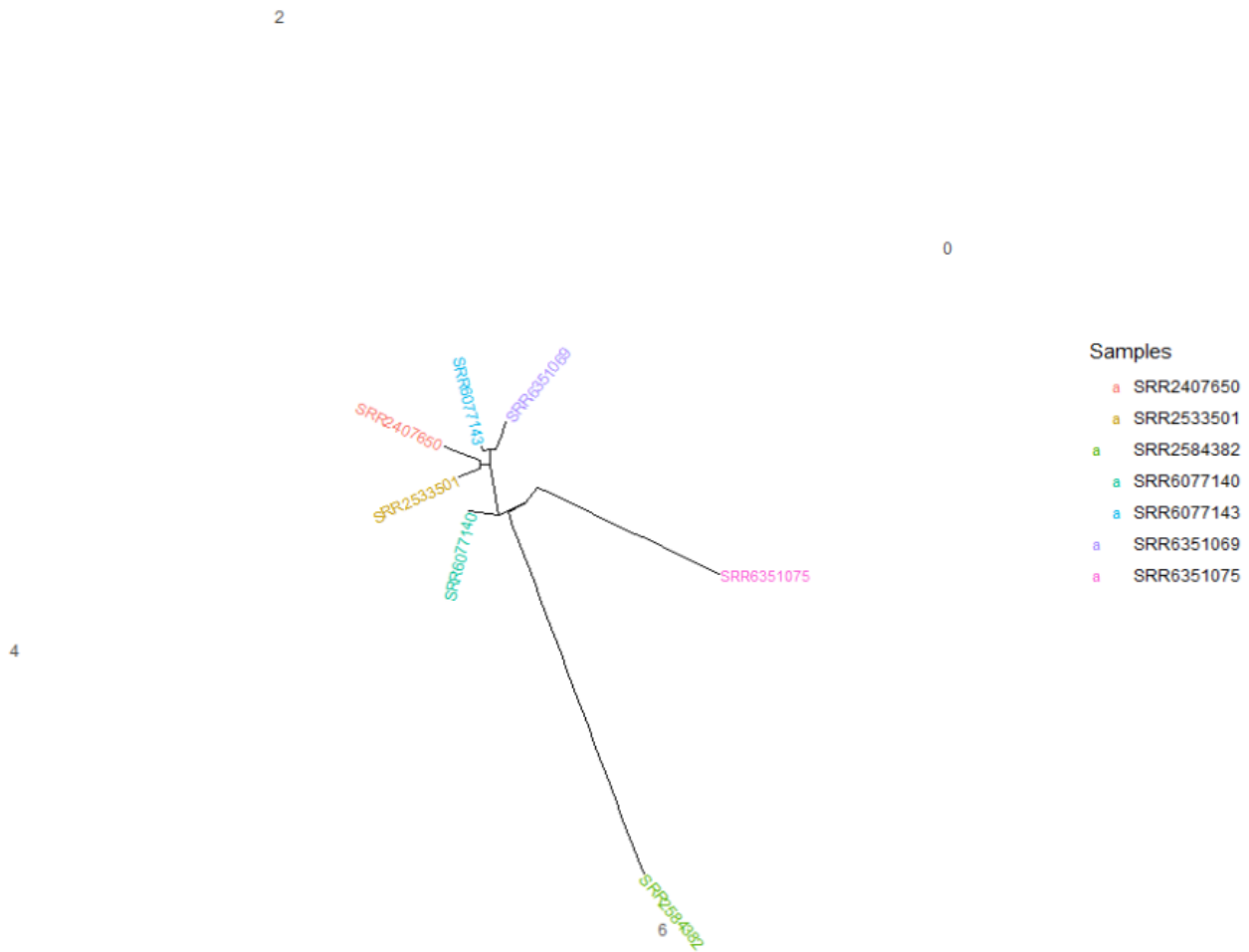
Application Cont.

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

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

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



Advanced Molecular Detection
Southeast Region Bioinformatics

Conclusion

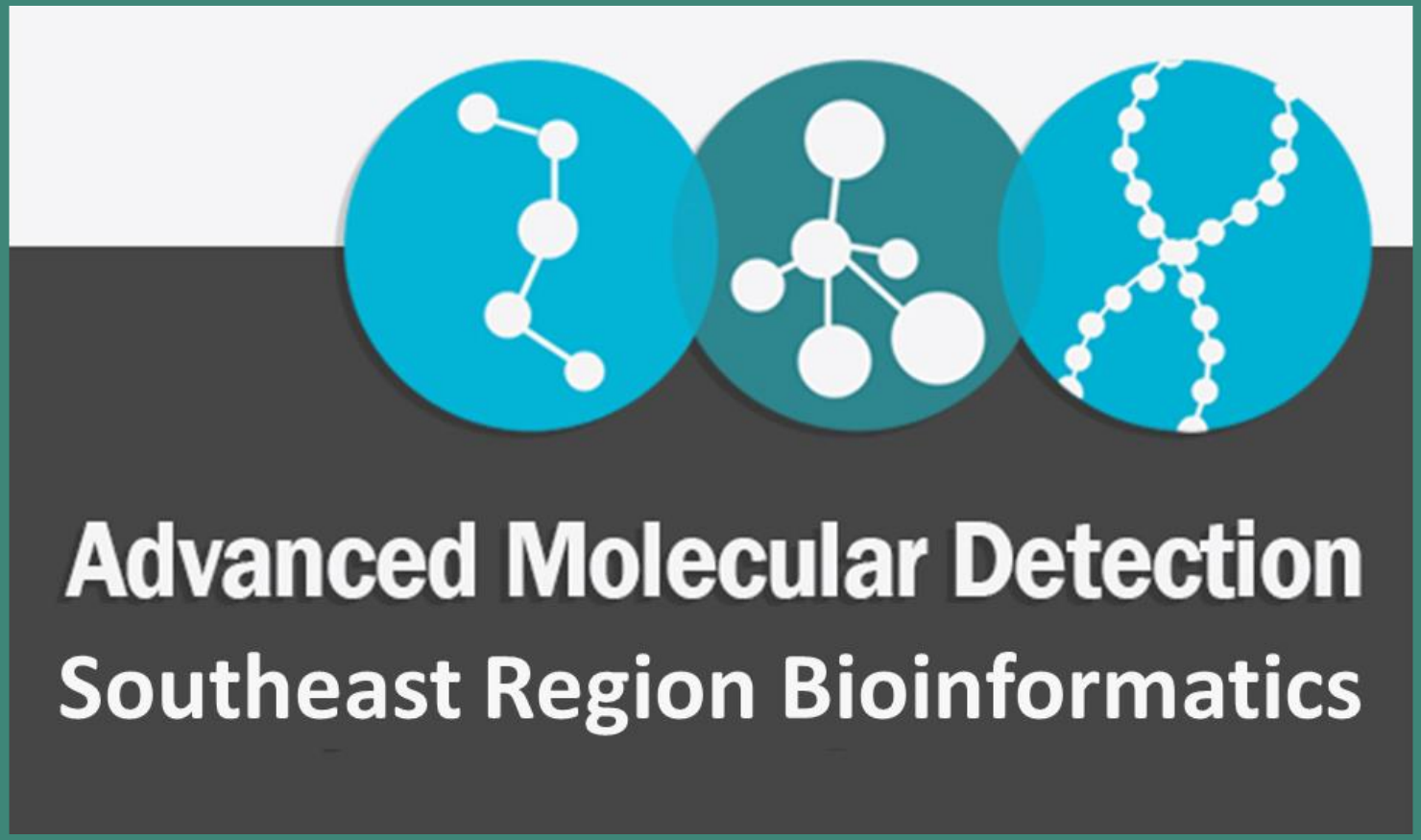
- Understood the basic principles of ggtree
- Installation and loading the ggtree packages in R
- Successfully applied ggtree with real world dataset using Flaq AMR and Talbot Pipeline

Advanced Molecular Detection
Southeast Region Bioinformatics

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



Questions?

bphl-sebioinformatics@flhealth.gov

Molly Mitchell, PhD
Bioinformatician Supervisor
Molly.Mitchell@flhealth.gov

Nikhil Reddy, MS
Bioinformatician
Nikhil.Yengala@flhealth.gov

Sam Marcellus, MPH
Bioinformatician
Samantha.marcellus@flhealth.gov