OUTBREAK REPORTING

Molly Mitchell, PhD, Bioinformatics Supervisor

FUNDING STATEMENT

This resource was made possible through funding provided under the Epidemiology and Laboratory Capacity for Prevention and Control of Emerging Infectious Diseases (ELC) Cooperative Agreement (CK24-0002), Project D: Advanced Molecular Detection to the Florida Department of Health. The conclusions, findings, and opinions expressed by authors do not necessarily reflect the official position of the U.S. Department of Health and Human Services, the Public Health Service, or the Centers for Disease Control and Prevention.

Advanced Molecular Detection
Southeast Region Bioinformatics

INTROS – BRR TEAM

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

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

Sam Bernhoft, MPH
Domain Lead Bioinformatician
Samantha.Bernhoft@flhealth.gov

For any BRR requests, please send an email to bphl-sebioinformatics@flhealth.gov
Based on the requests one of us from the team will respond as soon as possible.

Advanced Molecular Detection
Southeast Region Bioinformatics

OVERVIEW

When creating an outbreak report, consider the following:

- The best reports include a cumulation of genotyping, AMR profiling, phylogenetic analysis, and corresponding metadata
- Make conclusions based on considerations from all analyses results
- Use audience appropriate language when making the summary report
- Make sure to omit PPI or identifiable information
  - **Only include what may be helpful for drawing conclusions (i.e., collection dates and source)**

Advanced Molecular Detection
Southeast Region Bioinformatics

FLAQ-AMR

- Using raw ***.fastq.gz*** *(zipped .fastq)*, run FL's FLAQ-AMR pipeline
  - BPHL-Molecular FLAQ-AMR
  - Follow the directions on this GitHub page for git cloning the repository and executing the script.
  - This pipeline also needs python to run, make sure to load using **module load python**
- The FLAQ-AMR output will produce a directory called 'amrfinder_results'
  - Use this directory to analyze AMR genes found within your sample set
- The output from FLAQ-AMR will also produce a directory called 'annotations' with a **.gff** file for each isolate.
  - Copy and paste these files to a new directory using the code below

Advanced Molecular Detection
Southeast Region Bioinformatics

```
$ cp *.gff /path/to/new/directory/ .
```

AMRFINDERPLUS

- Used to identify AMR genes and resistance-associated point mutations
- Has an organism flag which will provide extra info related to stress, heat, biocide, and virulence

For the following organisms, run amrfinder –organism

- https://github.com/evolarjun/amr/wiki/Curated-organisms

Either use singularity or conda to install and run AMRFinderPlus

```
$ module load conda
$ conda create -yp /blue/bphl-<state>/<user>/conda_envs/amrfinder
$ conda activate /blue/bphl-<state>/<user>/conda_envs/amrfinder
$ conda install -c bioconda ncbi-amrfinderplus
```

Advanced Molecular Detection
Southeast Region Bioinformatics

AMRFINDERPLUS

We also covered this on the Office Hours on June 6, 2023

- This is on our GitHub - StaPH-B Southeast-region page if you need further information

I wrote a loop to run AMRFinderPlus in a directory of *.fasta* files

- This is in /blue/bphl-<state>/share/amrfinderplus/directory
- Edit the –organism, your email, and path to amrfinder conda env then sbatch amrfinder_plus_loop.sh

Advanced Molecular Detection
Southeast Region Bioinformatics

FL_CG_SNP

In this newly created directory, run FL's FL-cgSNP pipeline

- BPHL-Molecular FL-cgSNP
- Follow the directions on this GitHub page for git cloning the repository and executing the script.
- This pipeline requires python to run; load using **module load python** prior to running the pipeline
- **Note: this script needs to be run on the group of isolates you're comparing. This pipeline will produce comparative output that can't be merged.**

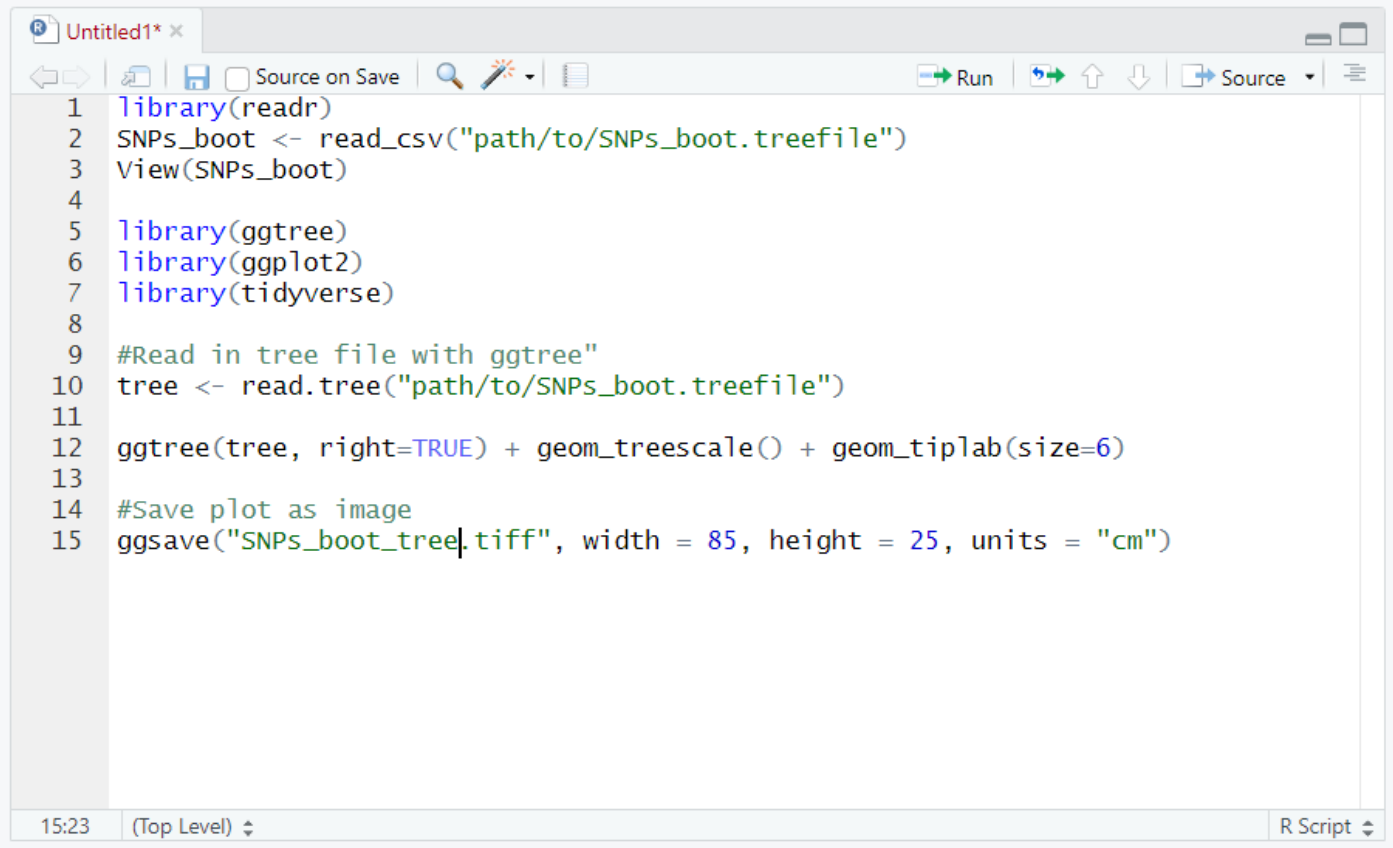
The output of this pipeline includes a **SNPs_boot.treefile**

- This file will produce a ML tree using the cgSNP's from a multiple sequence alignment of the shared core genes across your isolates.
- Upload this file to R/RStudio for visualization

Advanced Molecular Detection
Southeast Region Bioinformatics

R/RSTUDIO FOR TREE GENERATION

Using the following as a guide, use ggtree to visualize your SNPs_boot.treefile



Advanced Molecular Detection
Southeast Region Bioinformatics

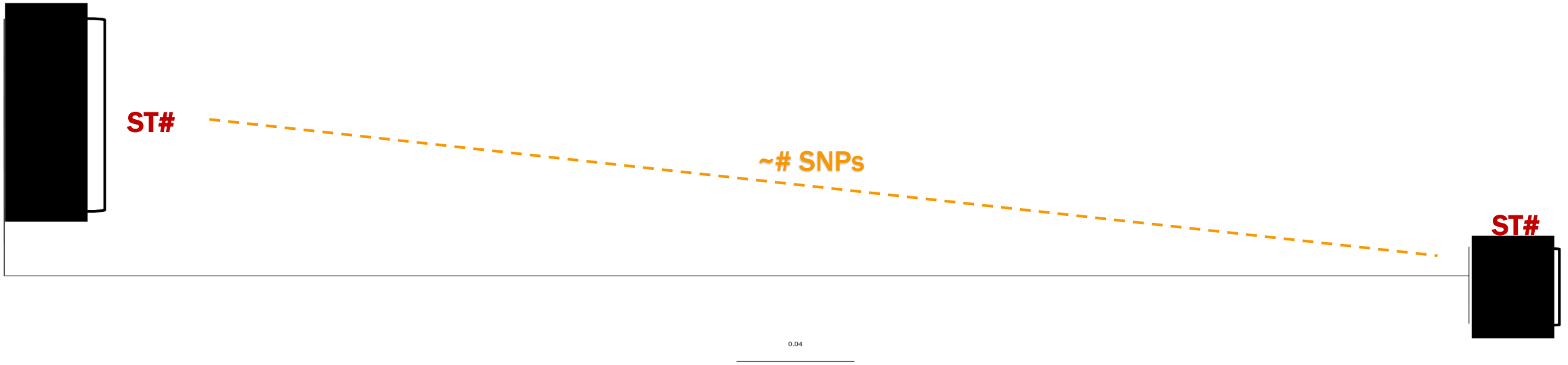
ADVANCED R/RSTUDIO FOR TREE GENERATION

If you're proficient in R, use this script written by Dr. Schmedes for labeling clades, etc.

Data visualization - R scripts (ggtree) · GitHub

Advanced Molecular Detection
Southeast Region Bioinformatics

ML CGSNP PHYLOGENETIC TREE



- An example of the ggtree output
- Adding ST values and numbers of SNPs between clades can help with concluding results

**Strain numbers blacked out for privacy

Advanced Molecular Detection
Southeast Region Bioinformatics

SNP MATRIX

- Use the **pairwise_matrix.tsv** file from the FL-cgSNP output to determine SNP differences between your isolates
- Copy this matrix to an excel file for easy visualization
- Then, use conditional formatting to add a heatmap to the values

snp-dists 0.6.2	Strain 1	Strain 2	Strain 3
Strain 1	0	22	13
Strain 2	22	0	17
Strain 3	13	17	0

Advanced Molecular Detection
Southeast Region Bioinformatics

FASTANI

Start an interactive session

```
$ srun –qos=bphl-umbrella –account=bphl-umbrella –cpus-per-task=10 –mem=50gb –time=10:00:00 –pty bash -i
```

Use conda to install fastANI

```
$ module load conda
$ conda create -yp /your/path/to/conda/envs/fastani
$ conda activate /your/path/to/conda/envs/fastani
$ conda install -c bioconda fastani
```

Advanced Molecular Detection
Southeast Region Bioinformatics

FASTANI

Change to the directory of *.fasta* files and run fastANI

```
$ ls *.fasta > strains.txt
$ fastANI –ql strains.txt –rl strains.txt –output ANIresults.txt –matrix
```

Advanced Molecular Detection
Southeast Region Bioinformatics

FASTANI

Copy the results of the ANIresults.txt.matrix to excel and apply conditional formatting for easy visualization of the results

	Strain 1	Strain 2	Strain 3
Strain 1	100		
Strain 2	99.99758	100	
Strain 3	99.9997	99.99881	100

Advanced Molecular Detection
Southeast Region Bioinformatics

SANIBEL AND TALBOT

- Nextflow versions of both FLAQ-AMR and FL_cg_SNP are Sanibel and Talbot – same output, slightly different format
- Both available on our Github: BPHL-Molecular · GitHub

Advanced Molecular Detection
Southeast Region Bioinformatics

OTHER TOOLS – SPECIES DEPENDENT

- mycoSNP - CDC pipeline for *C. auris*
- Ksnp3 or Ksnp4 – highly divergent/variable (*Legionella*)

- Important to research pipelines before using them!
  - Pick what's best for your species or the question you're asking!

Advanced Molecular Detection
Southeast Region Bioinformatics

REPORT GENERATION - BEGINNER

1. Use PowerPoint presentation or Microsoft publisher
2. Have an introductory slide with the main points
3. Include a methodology slide with pipelines and versions used in the analysis
4. List table with relevant metadata, without PPI or identifiable information
5. Also include AMR results from the FLAQ-AMR report if applicable
6. Include a ML phylogeny, SNP, and ANI matrix within report
   - Use a slide for each
7. Snipping Tool/screenshots help for imaging

Advanced Molecular Detection
Southeast Region Bioinformatics

REPORT GENERATION - INTERMEDIATE

For those of you proficient in R, there is a report generation package called R Markdown

- R Markdown (rstudio.com)
- Fully reproducible
- HTML and PDF options

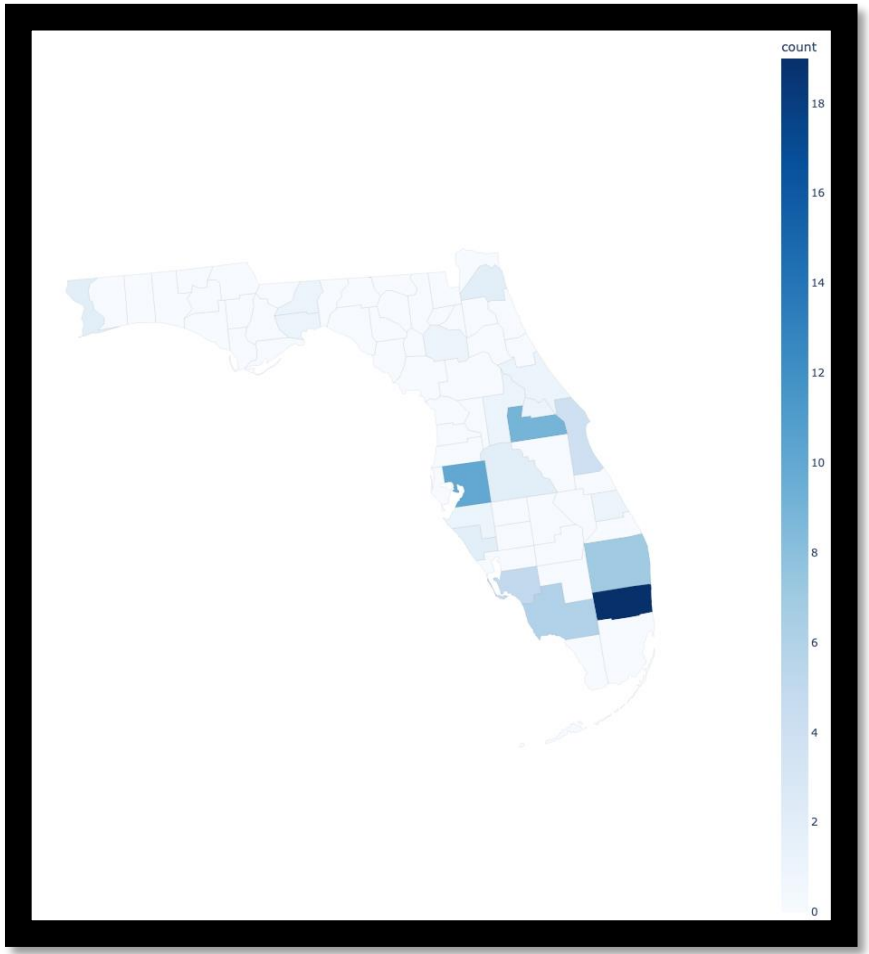
Advanced Molecular Detection
Southeast Region Bioinformatics

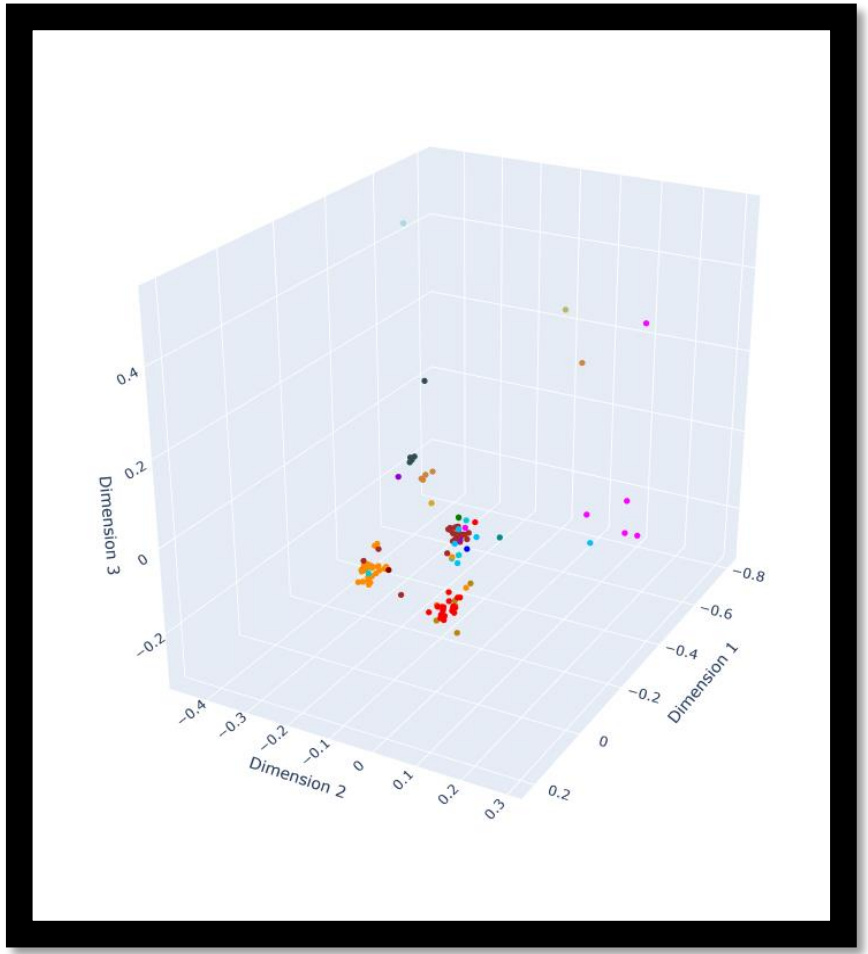
REPORT GENERATION - ADVANCED

- Jupyter Notebook & plotly to create an interactive html report
- Geojson interactive maps where we add # of cases
- Plot phylogenies with
- Add 3D MDS modeling plots – another cool way of displaying distance between samples!

Advanced Molecular Detection
Southeast Region Bioinformatics

SOME EXAMPLES





SOME THINGS TO REMEMBER

Use audience appropriate language

- Use "similar" or "related" instead of "identical" or "from the same source"
- Be ready to describe in detail your analyses and findings to epis or providers

Add a "surveillance only" disclaimer

- Our NGS analyses and bioinformatics pipelines are not clinically validated and are for surveillance only
- Here's ours: "The results in this report were obtained by WGS, using research use only procedures that were not CLIA validated. They may be used for surveillance and research purposes only. They are not to be used for diagnosis, treatment, or management of patient care."

Advanced Molecular Detection
Southeast Region Bioinformatics

NEXT STEPS

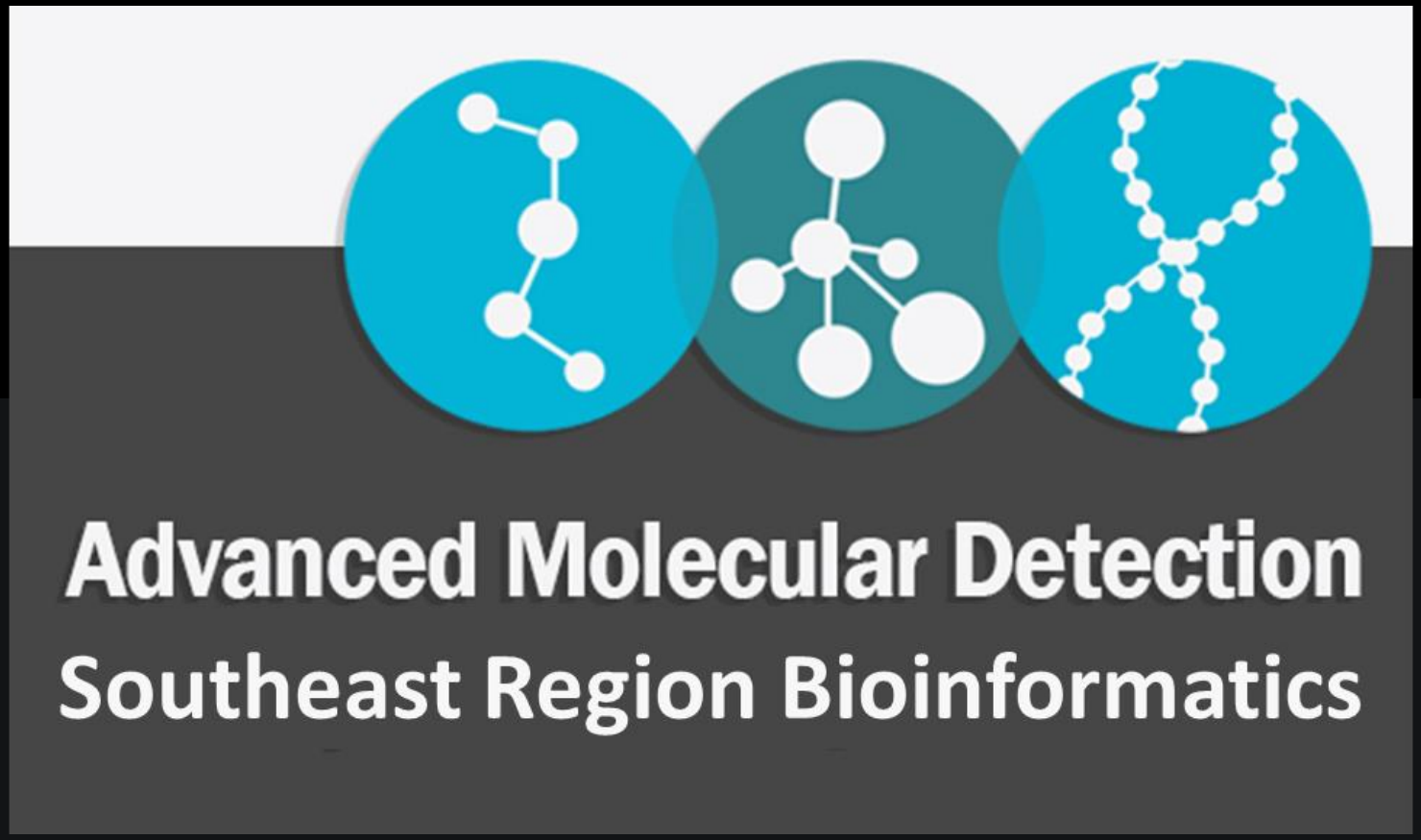
An internal SOP that will be adjusted to external use

Files will be hosted on our GitHub - BPHL-Molecular · GitHub

- Geojson, conda environment yml, data processing and report generation ipynb

We will provide a training via our Office Hours hosted by Nikhil

Advanced Molecular Detection
Southeast Region Bioinformatics



QUESTIONS?

bphl-sebioinformatics@flhealth.gov

Molly Mitchell, PhD

Bioinformatics Supervisor

Molly.Mitchell@flhealth.gov

Nikhil Yengala Reddy, MS

Bioinformatician

Nikhil.Yengala@flhealth.gov