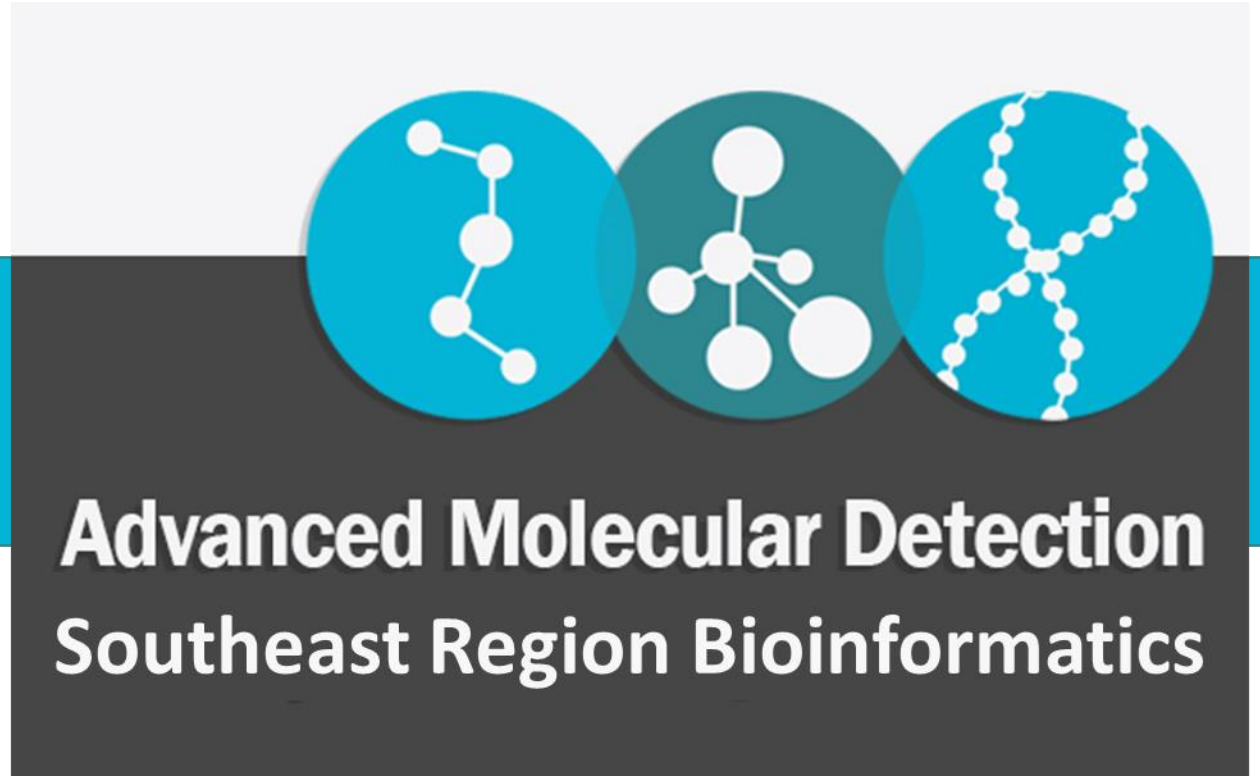


Outbreak/Cluster Report Training

06/12/2023

Outline

- Updates - APHL
- Agenda
- Overview
- FLAQ-AMR
- AMRFinderPlus
- FL-cgSNP
- Tree Generation
- FastANI
- Report Generation
- Questions

Updates - APHL

- Easy Genomics and Nextflow Training Application
  - APHL Easy Genomics and Nextflow Training Application Survey (surveymonkey.com)
  - Potentially one person per jurisdiction
  - In-person training in Arlington, VA - **August 14-16**
  - Applications are due by **June 15**
  - Must complete 8-10 hours of pre-work
  - Requires command line knowledge – proficient
- DataCamp subscription renewal/application
  - DataCamp Membership Application 2023 Survey (surveymonkey.com)

Advanced Molecular Detection
Southeast Region Bioinformatics

Agenda

June 26 – Git & GitHub
July 10 – Bactopia tools

Future Trainings
- ONT & FL's Flisochar pipeline
- StaPH-B Toolkit Programs/Pipelines
- GISAID flagged SARS-CoV-2
- R Training Series
- Dryad pipeline
- Generating R figures
- …and more

Advanced Molecular Detection
Southeast Region Bioinformatics

Overview

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

Advanced Molecular Detection
Southeast Region Bioinformatics

FLAQ-AMR

- Using raw ***.fastq.gz***, run FL's FLAQ-AMR pipeline
  - BPHL-Molecular FLAQ-AMR
  - Follow the directions on this GitHub page for git cloning the repository and steps to 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 to analyze AMR genes found within your sample set
- The output from FLAQ-AMR will produce a directory called 'annotations' with a **.gff** file for each isolate.
  - Copy and paste these files to a new directory

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

Advanced Molecular Detection
Southeast Region Bioinformatics

AMRFinderPlus

- AMRFinderPlus 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**

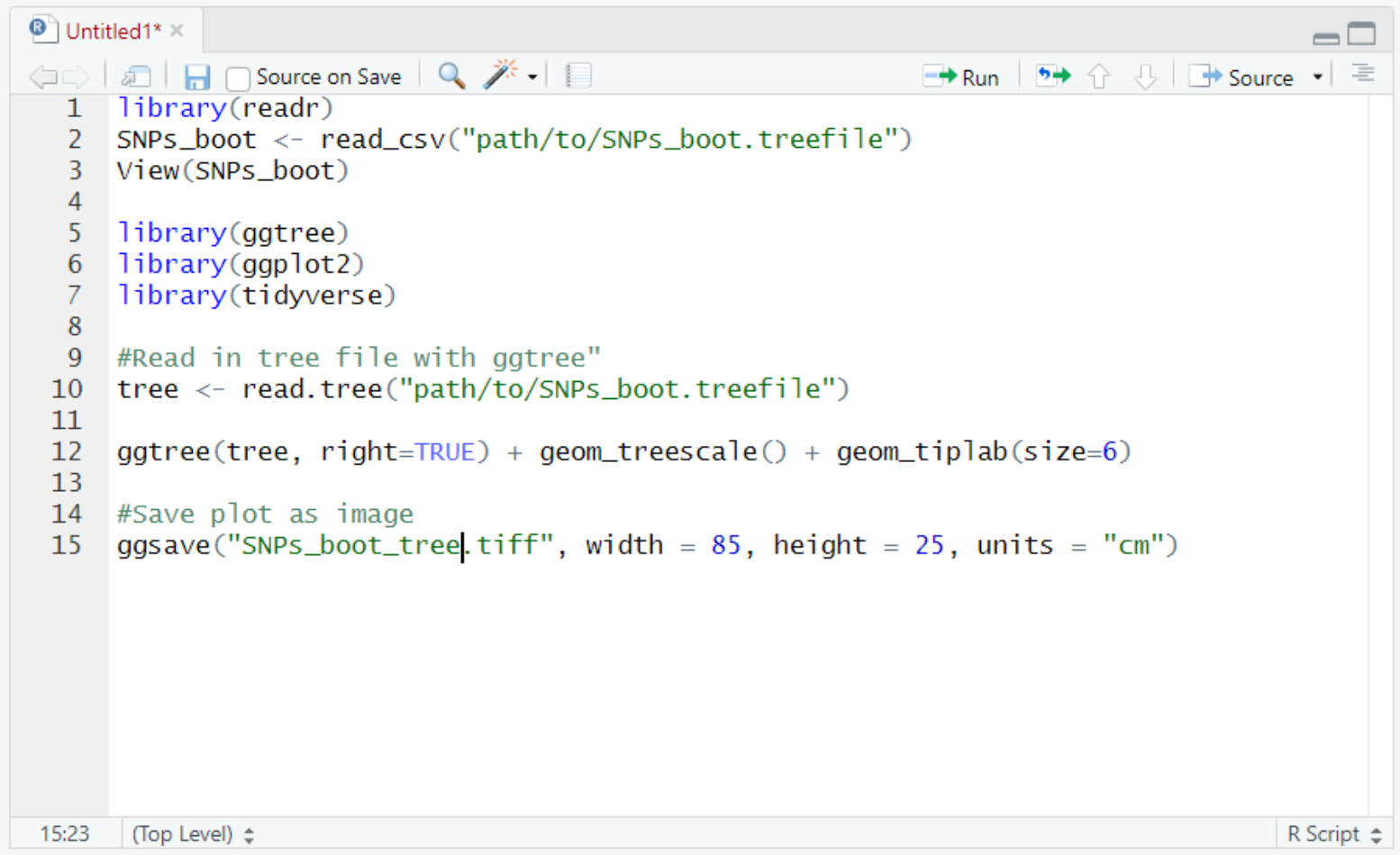
FL-cgSNP

- In this new directory, run FL's FL-cgSNP pipeline
  - BPHL-Molecular FL-cgSNP
  - Follow the directions on this GitHub page for git cloning the repository and steps to executing the script.
  - This pipeline also needs python to run, make sure to load using **module load python**
  - **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

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**



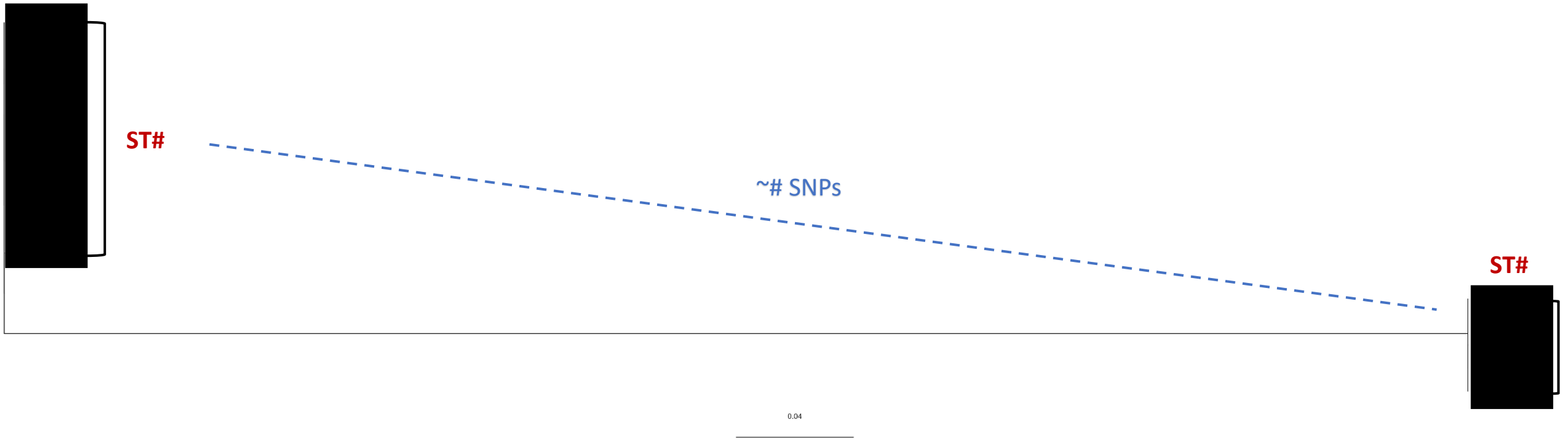
Advanced Molecular Detection
Southeast Region Bioinformatics

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

Report Generation

- Use PowerPoint presentation
- Have an introductory slide with the main points
- Include a methodology slide with pipelines and versions used in the analysis
- List table with relevant metadata, without PPI or identifiable information
  - Also include AMR results from the FLAQ-AMR report if applicable
- Include a ML phylogeny, SNP, and ANI matrix within report
  - Use a slide for each
  - Snip Tool helps for imaging
- Use appropriate language
  - Use "similar" or "related" instead of "identical" or "from the same source"

Advanced Molecular Detection
Southeast Region Bioinformatics

R Markdown

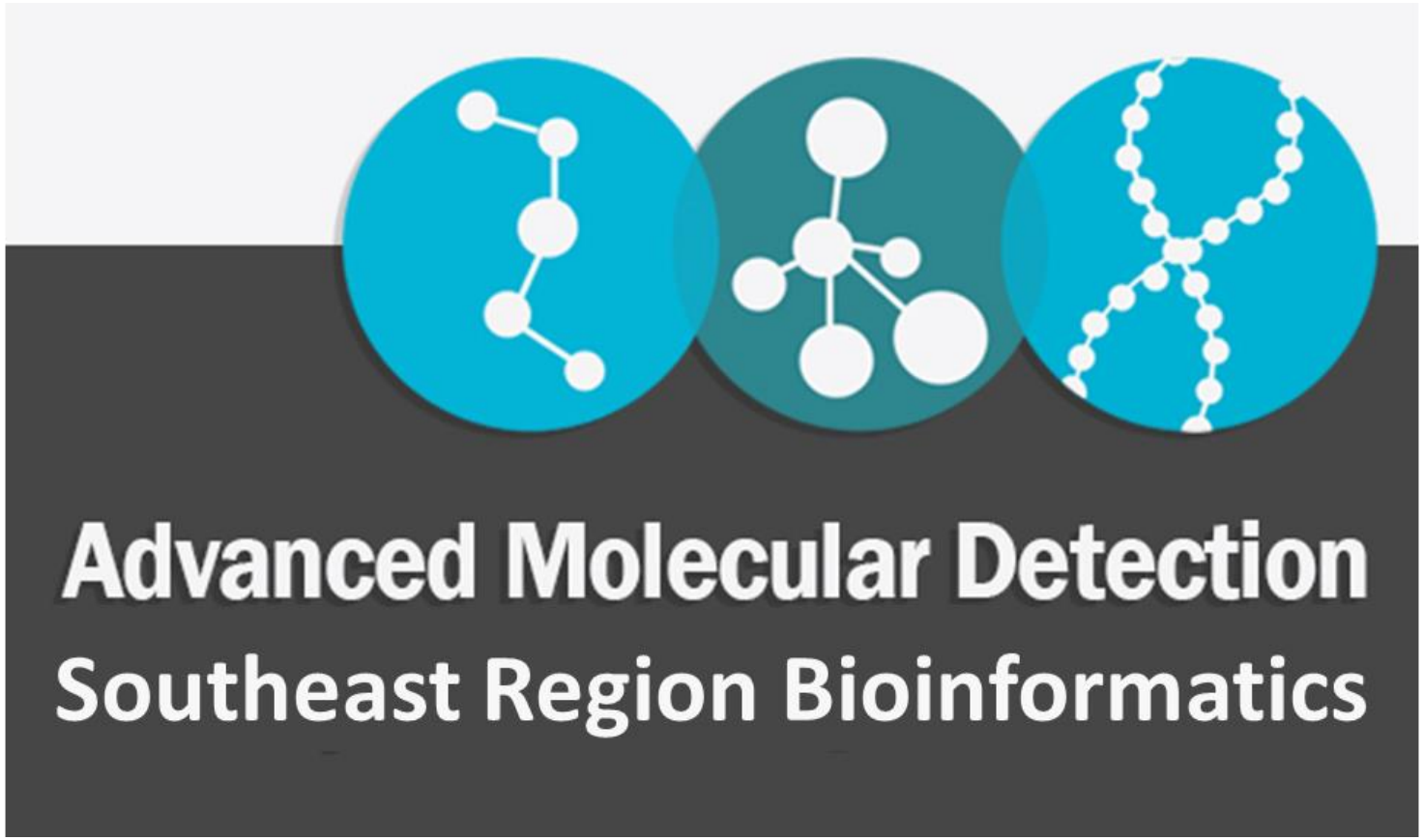
- For those of you proficient in R, there is a report generation package called R Markdown
  - R Markdown (rstudio.com)
  - Fully reproducible
  - HTML and PDF options
- Here in Florida, we're in the process of transitioning from PowerPoint to R Markdown
- Look for an Office Hours on R Markdown

Advanced Molecular Detection
Southeast Region Bioinformatics

Time for Questions & Feedback

- Questions?
  - Do you need help with anything?
  - Requests for separate trainings?
- Feedback
  - What would you like to see?

Advanced Molecular Detection
Southeast Region Bioinformatics



Questions?

bphl-sebioinformatics@flhealth.gov

Lakshmi Thsaliki, MS
Bioinformatician
Lakshmi.Thsaliki@flhealth.gov

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

Sarah Schmedes, PhD
Bioinformatics Supervisor
Sarah.Schmedes@flhealth.gov