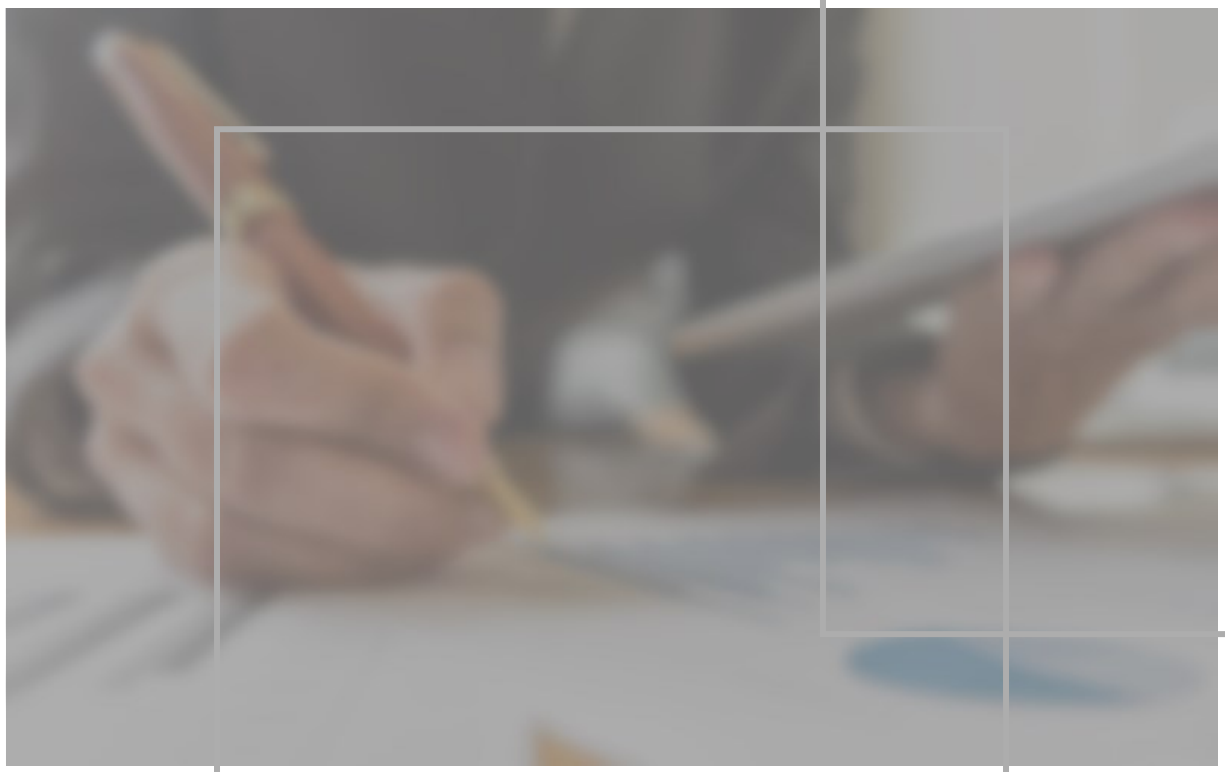
Report Generation

Advanced Molecular Detection

Southeast Region Bioinformatics



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.

Updates

- In September we'll be scheduling our next check-in calls
- Next presentations:
  - 09/15- Legionella Tools
  - 09/29- Neissflow

Overview

Purpose

- This pipeline tool is developed by the Florida BPHL team to streamline and generate genomic epidemiology reports which consist of data preprocessing and report generation of phylogenetic tree, 3D plots, generation of AMR genes, and Geomap plot.

Usage

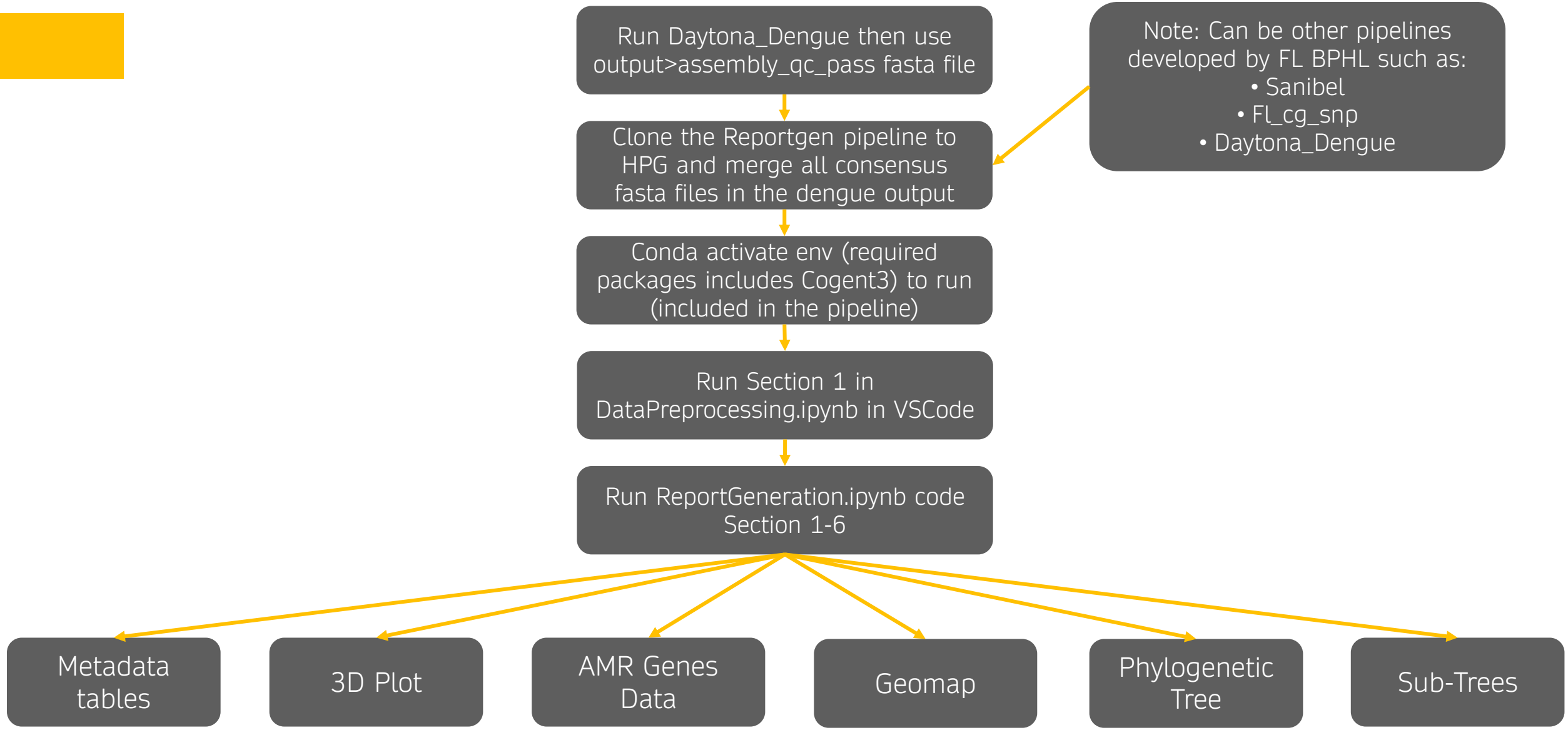
- Can be used by public health labs, researchers, and sequencing centers to generate standardized epidemiology reports for surveillance and research purposes only.

Dependencies

- Conda; Pandas, Plotly, Bio, Matplotlib, Numpy, Sklearn, Warnings, Json
- VScode
- Cogent3

Advanced Molecular Detection
Southeast Region Bioinformatics

Workflow



Advanced Molecular Detection
Southeast Region Bioinformatics

Application

Objective

Use a Dengue_Daytona output dataset and generate a genomic epidemiology report from the Report Generation pipeline.

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

1. BPHL-Molecular Github:

https://github.com/BPHL-Molecular/ReportGen

2. Clone the Code and copy the Data to your Analysis Directory

3. Create an environment in conda using the environmental.yml file that is provided:

conda env create --name ReportGen -f environment.yml

4. Run HPG on VS Code:

Note: Review VS Code Office Hours on connecting HPG to VS Code

5. Install all the Required Packages in the ReportGen Env:

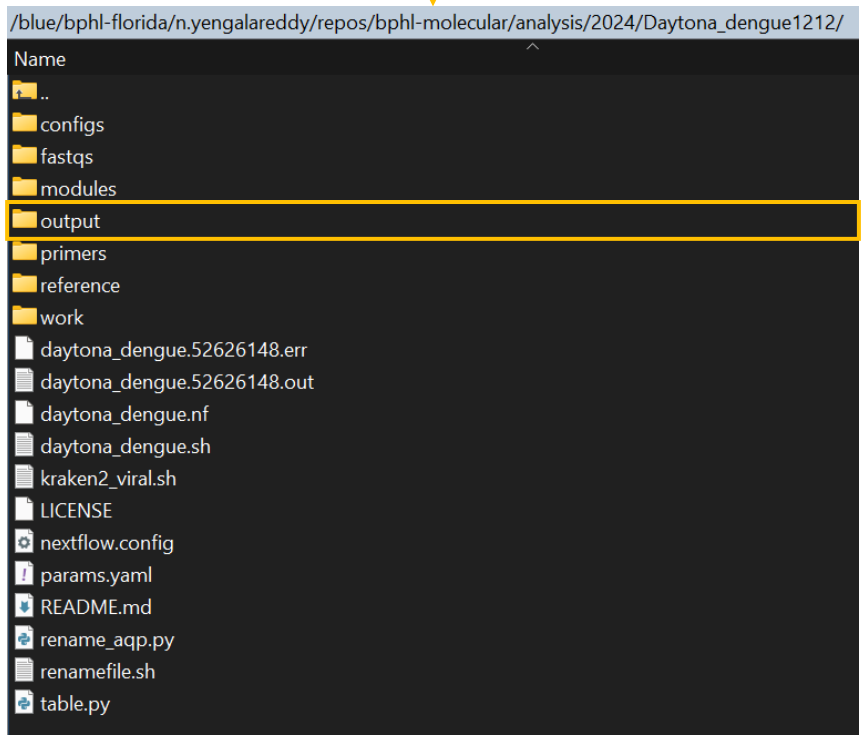
Cogent3, Pandas, Plotly, Bio, Matplotlib, Numpy, Sklearn, Warnings, Json

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

Run Daytona_Dengue

Note: Review Previous Office Hours on How to Run Daytona_Dengue



Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/2024/Daytona_dengue1212/

Name
..
configs
fastqs
modules
output
primers
reference
work
daytona_dengue.52626148.err
daytona_dengue.52626148.out
daytona_dengue.nf
daytona_dengue.sh
kraken2_viral.sh
LICENSE
nextflow.config
params.yaml
README.md
rename_aqp.py
renamefile.sh
table.py

cat *.fa > merged.consensus.fa
Note: Run this command in the assembly_qc_pass directory

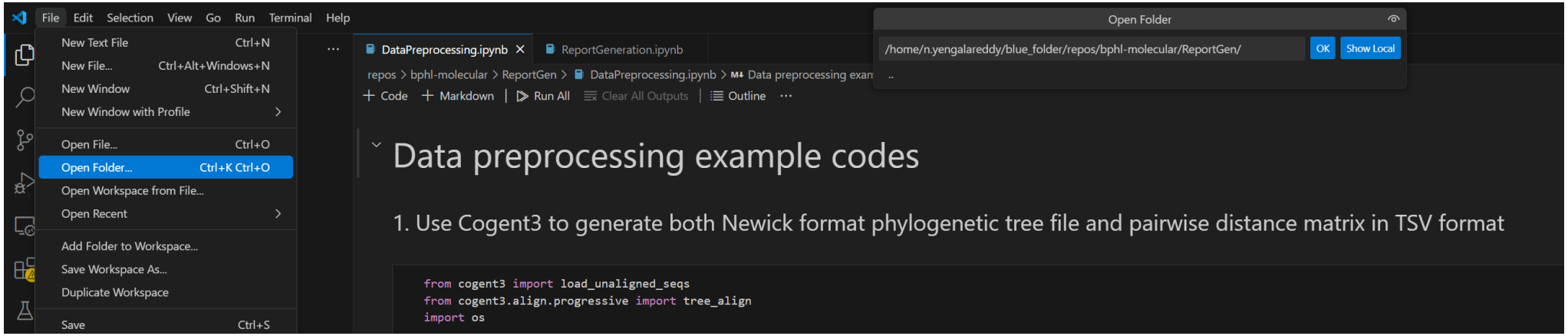
/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/2024/Daytona_dengue1212/output/

Name
..
alignment
assembly
assembly_qc_pass
bbduk
fastqc
fastqc_clean
humanscrubber
kraken_out_broad
multiqc
report
variant
variant_qc_pass

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/2024/Daytona_dengue1212/output/assembly_qc_pass/

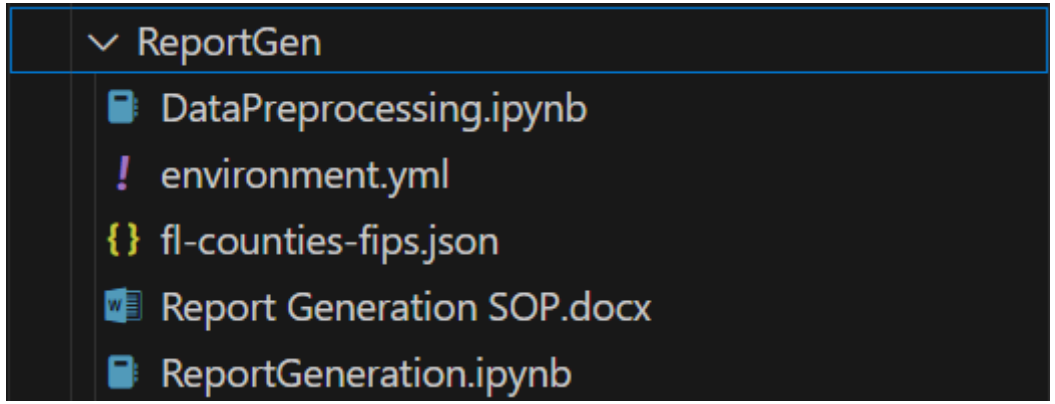
Name
..
SER_del
J137.consensus.fa
J138.consensus.fa
J196.consensus.fa
J198.consensus.fa
J210.consensus.fa
J248.consensus.fa
J320.consensus.fa
J339.consensus.fa
J357.consensus.fa
J369.consensus.fa
merged.consensus.fa
T91.consensus.fa
T93.consensus.fa
T94.consensus.fa
T107.consensus.fa

Application Cont.



Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.



Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

repos > bphl-molecular > ReportGen > DataPreprocessing.ipynb > Data preprocessing example codes

+ Code + Markdown | Run All Restart Clear All Outputs | Jupyter Variables Outline ···

cogent3 (Python 3.11.13)

Data preprocessing example codes

1. Use Cogent3 to generate both Newick format phylogenetic tree file and pairwise distance matrix in TSV format

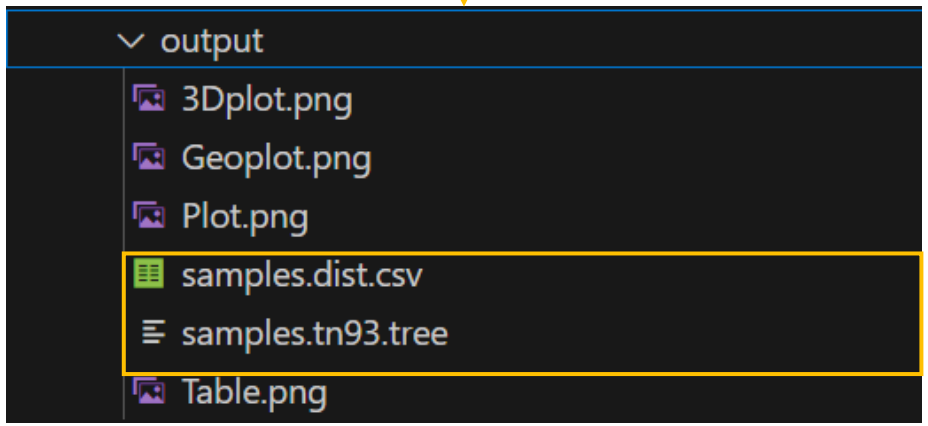
```
from cogent3 import load_unaligned_seqs
from cogent3.align.progressive import tree_align
import os

consensus_fasta_file = "merged.consensus.fa"  # Replace with your file path
tree_newick_file = "samples.tn93.tree"  # Replace with your file path
dists_csv_file = "samples.dist.csv"  # Replace with your file path

seqs = load_unaligned_seqs(consensus_fasta_file, moltype="dna")
aln, tree = tree_align("TN93", seqs, show_progress=False)
tree.write(tree_newick_file)

dists = aln.distance_matrix(calc="tn93", show_progress=False)
dists_df = dists.to_table().to_dataframe()
dists_df.to_csv(dists_csv_file)
```

[ ] Python

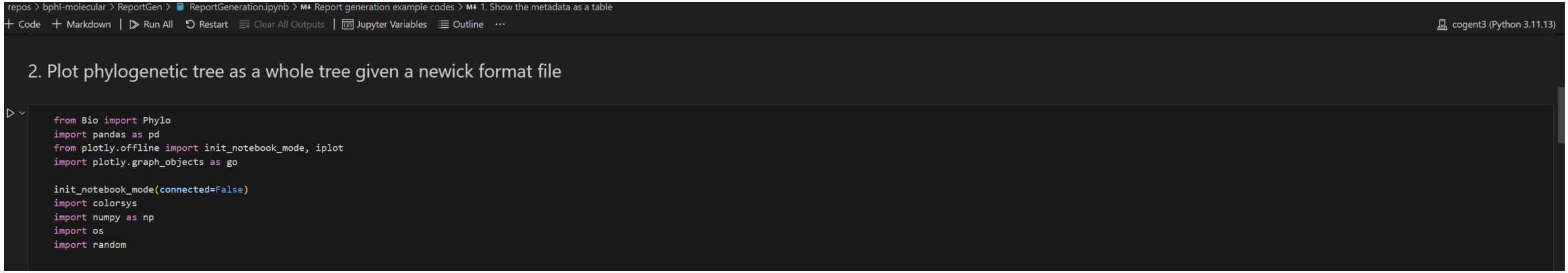


Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

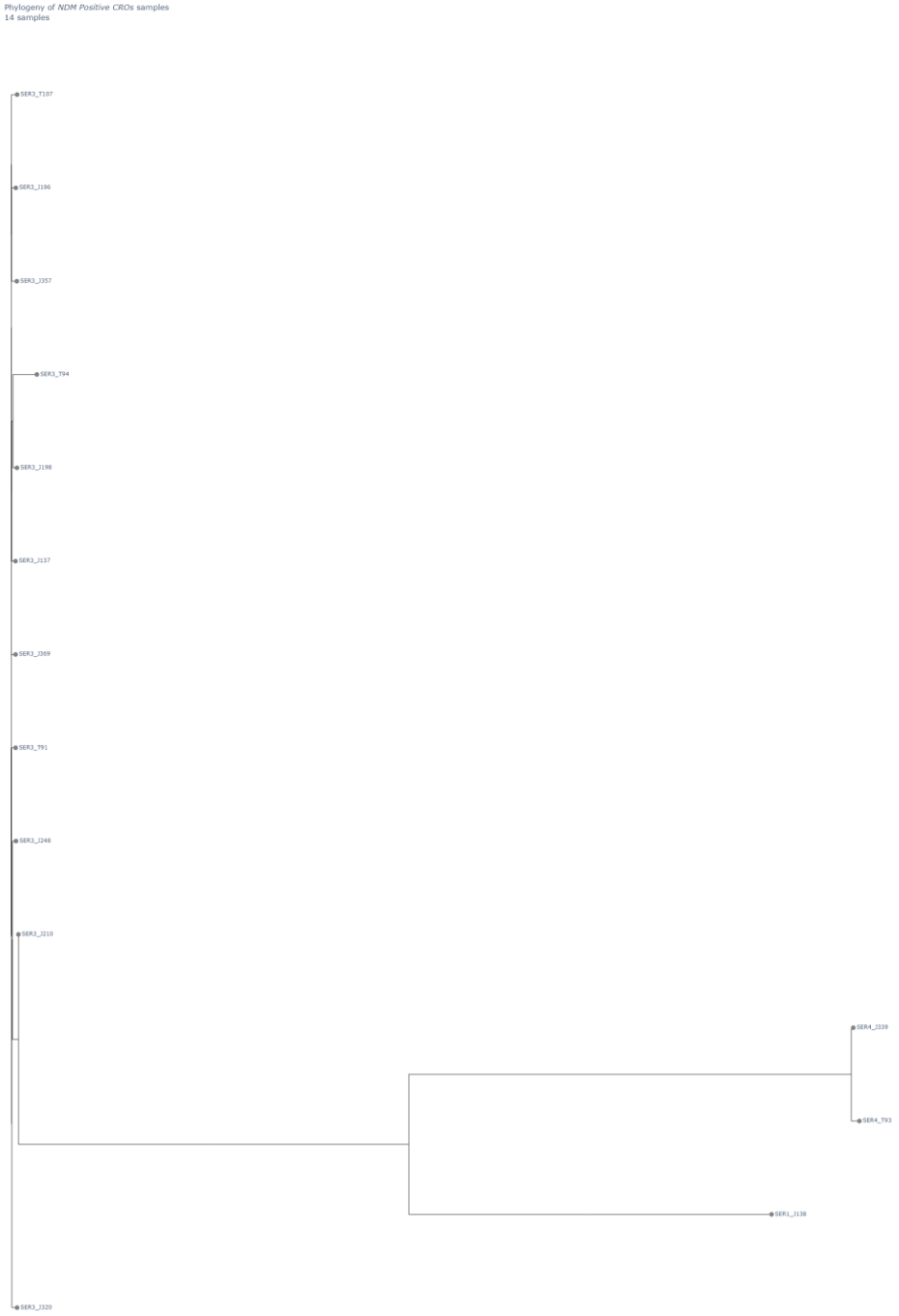
Sample ID	Collection Date	County of Origin	speciesID_kraken	speciesID_mash	NDM
J196	2023-02-02	Pinellas	Enterobacter cloacae	Enterobacter cloacae	blaNDM-7
J248	2023-02-10	Martin	Enterobacter hormaechei	Enterobacter cloacae	blaNDM-7
J369	2023-03-01	Alachua	Enterobacter cloacae	Enterobacter cloacae	blaNDM-7
T94	2023-03-06	Pinellas	Enterobacter hormaechei	Enterobacter cloacae	blaNDM-7
J138	2023-01-18	Orange	Escherichia coli	Escherichia coli	blaNDM-5
J357	2023-02-24	Martin	Escherichia coli	Escherichia coli	blaNDM-5
J320	2023-02-17	Orange	Klebsiella aerogenes	Klebsiella aerogenes	blaNDM-7
J314	2023-02-14	Alachua	Klebsiella michiganensis	Klebsiella michiganensis	blaNDM-1
J137	2023-01-15	Alachua	Klebsiella pneumoniae	Klebsiella pneumoniae	blaNDM-1
J210	2023-02-08	Hillsborough	Klebsiella pneumoniae	Klebsiella pneumoniae	blaNDM-1
J339	2023-02-20	Duval	Klebsiella pneumoniae	Klebsiella pneumoniae	blaNDM-1
T91	2023-03-04	Orange	Klebsiella pneumoniae	Klebsiella pneumoniae	blaNDM-1
T93	2023-03-05	Duval	Klebsiella pneumoniae	Klebsiella pneumoniae	blaNDM-1
J198	2023-02-06	Duval	Klebsiella variicola	Klebsiella variicola	blaNDM-1
T107	2023-03-10	Alachua	Klebsiella variicola	Klebsiella variicola	blaNDM-1

Application Cont.



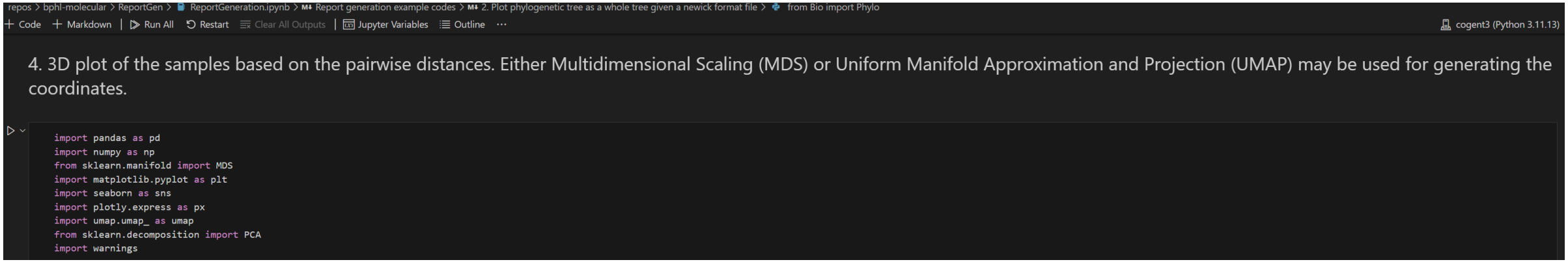
Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

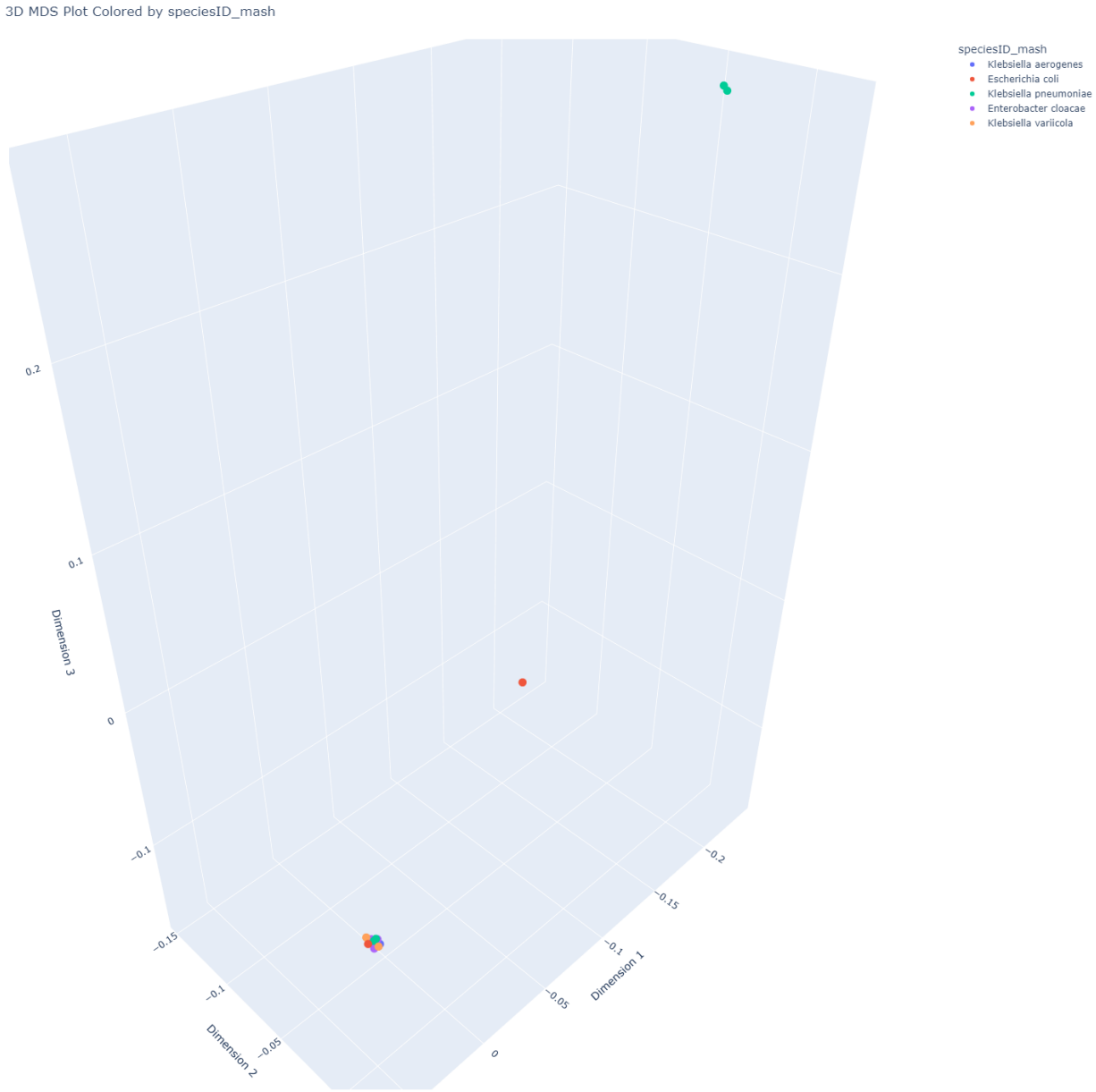


Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.



Application Cont.



Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

repos > bphl-molecular > ReportGen > ReportGeneration.ipynb > Report generation example codes > 2. Plot phylogenetic tree as a whole tree given a newick format file > from Bio import Phylo

+ Code + Markdown | Run All Restart Clear All Outputs | Jupyter Variables Outline ··· cogent3 (Python 3.11.13)

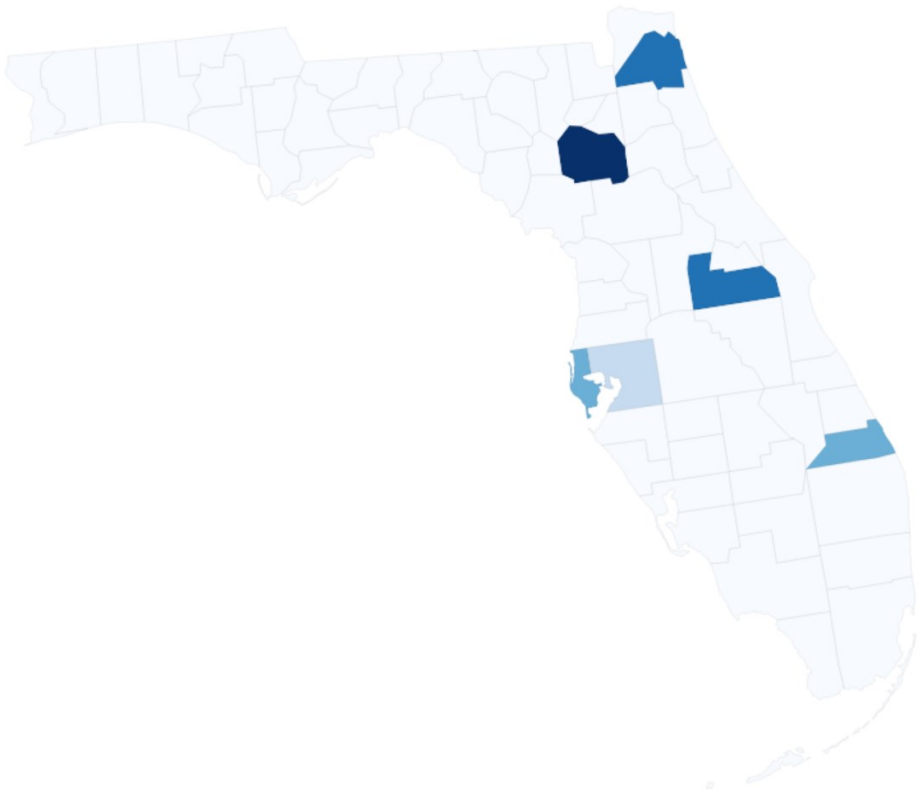
6. Geo map by county

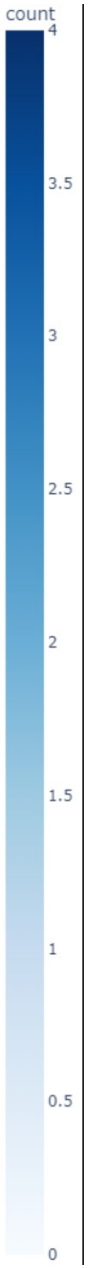
```
import plotly.graph_objects as go
import pandas as pd
import plotly.express as px
import os
import json

# Load the geojson file
cwd = os.getcwd()
with open("../Data/meta/fl-counties-fips.json", "r") as f:  # update the path to the geojson file
    geojson = json.load(f)

# Load the CSV file
meta = pd.read_csv("../Data/meta/meta_data.csv") # example meta data file; you may change the meta data file here
renamed_columns = ["Sample ID", "Collection Date", "County"]  # only show data for these columns; you may change the column names here
meta.columns = renamed_columns
```

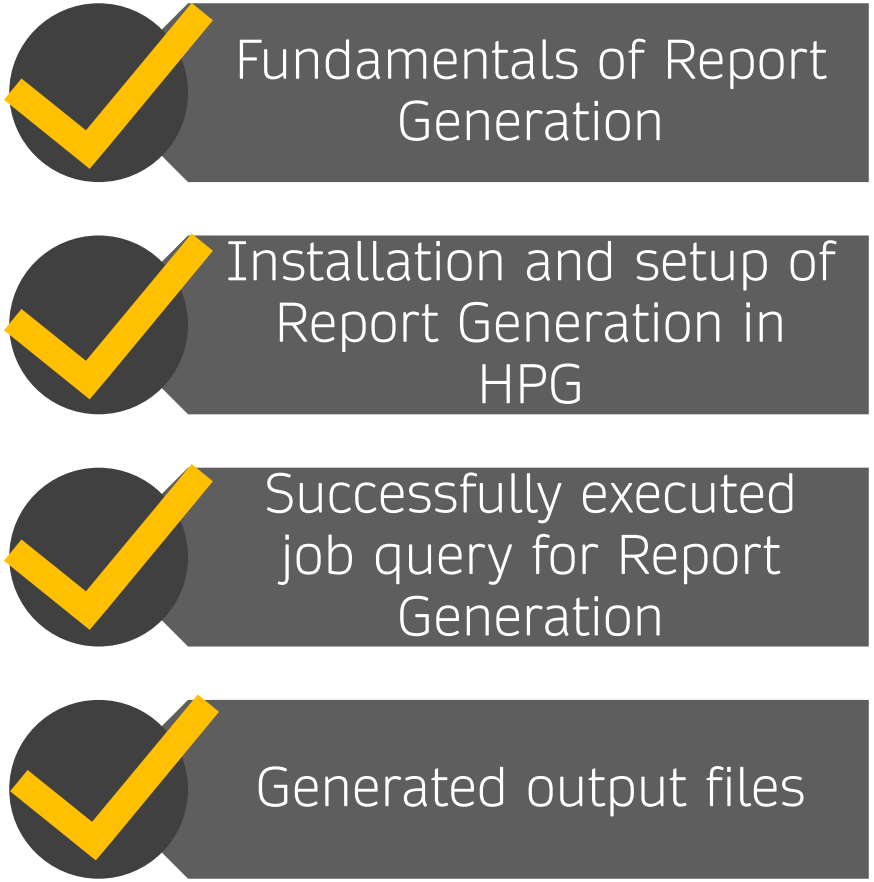
Application Cont.

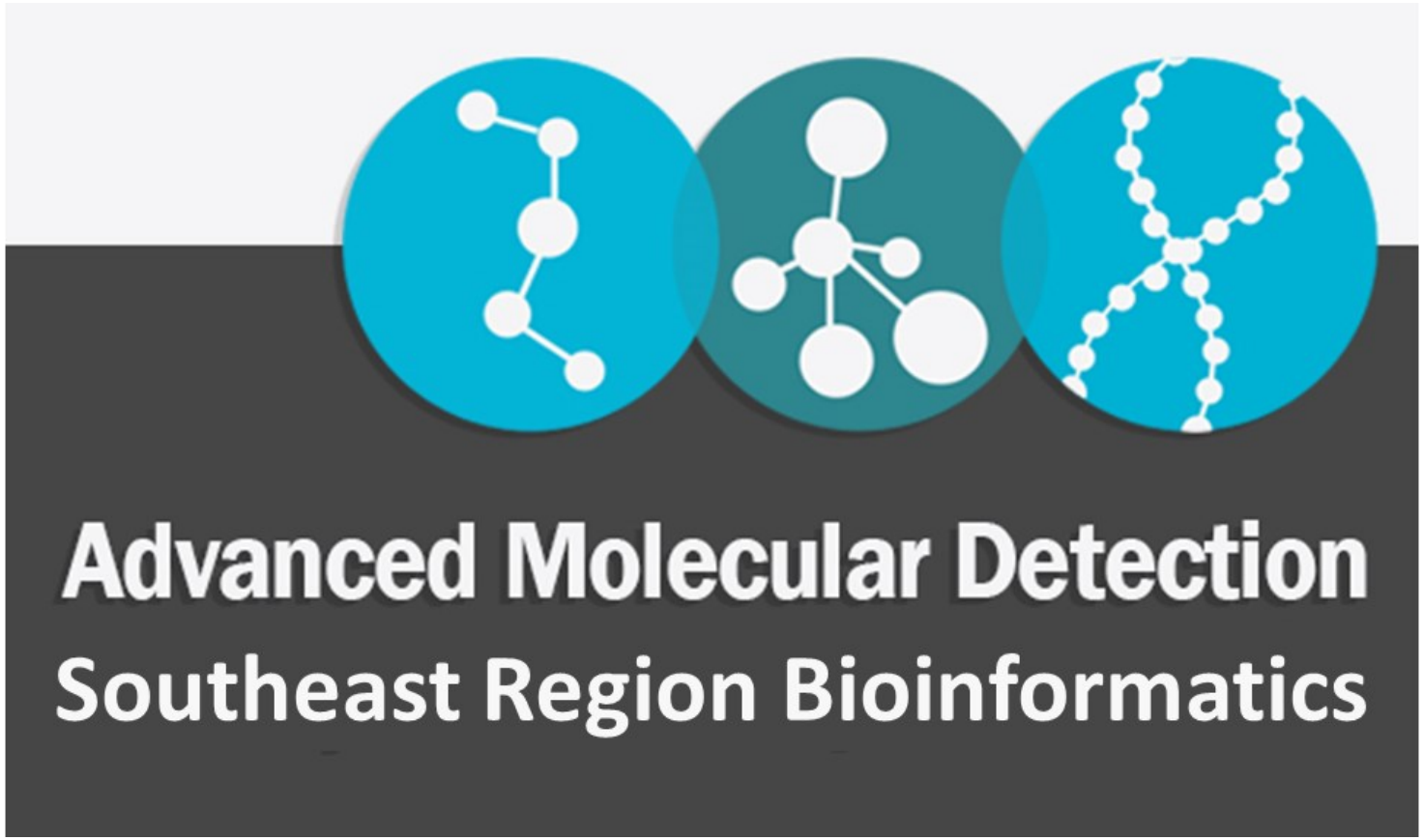




Advanced Molecular Detection
Southeast Region Bioinformatics

Conclusion


Fundamentals of Report Generation
Installation and setup of Report Generation in HPG
Successfully executed job query for Report Generation
Generated output files



Questions?

bphl-sebioinformatics@flhealth.gov

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

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

Sam Bernhoft, MPH
Bioinformatician
Samantha.bernhoft@flhealth.gov