

<p>28 October 2024</p>		
	<p><b>Sanibel</b></p> <p>Advanced Molecular Detection Southeast Region Bioinformatics</p>	

# Updates

## HiPerGator

- November 6, 2024  
(Starting at 7:00 AM EST) – HPG Software  
Update to the /blue Storage System  
(Expect Periods of Delay and  
Unavailability)

## Office Hours

- November 18, 2024 – FLAQ-SARS-CoV-2
- November 25, 2024 – Daytona



## Purpose

- To analyze NGS bacterial genome data with a goal on finding clonal complexes, serotypes and antimicrobial resistance genes

## Used For

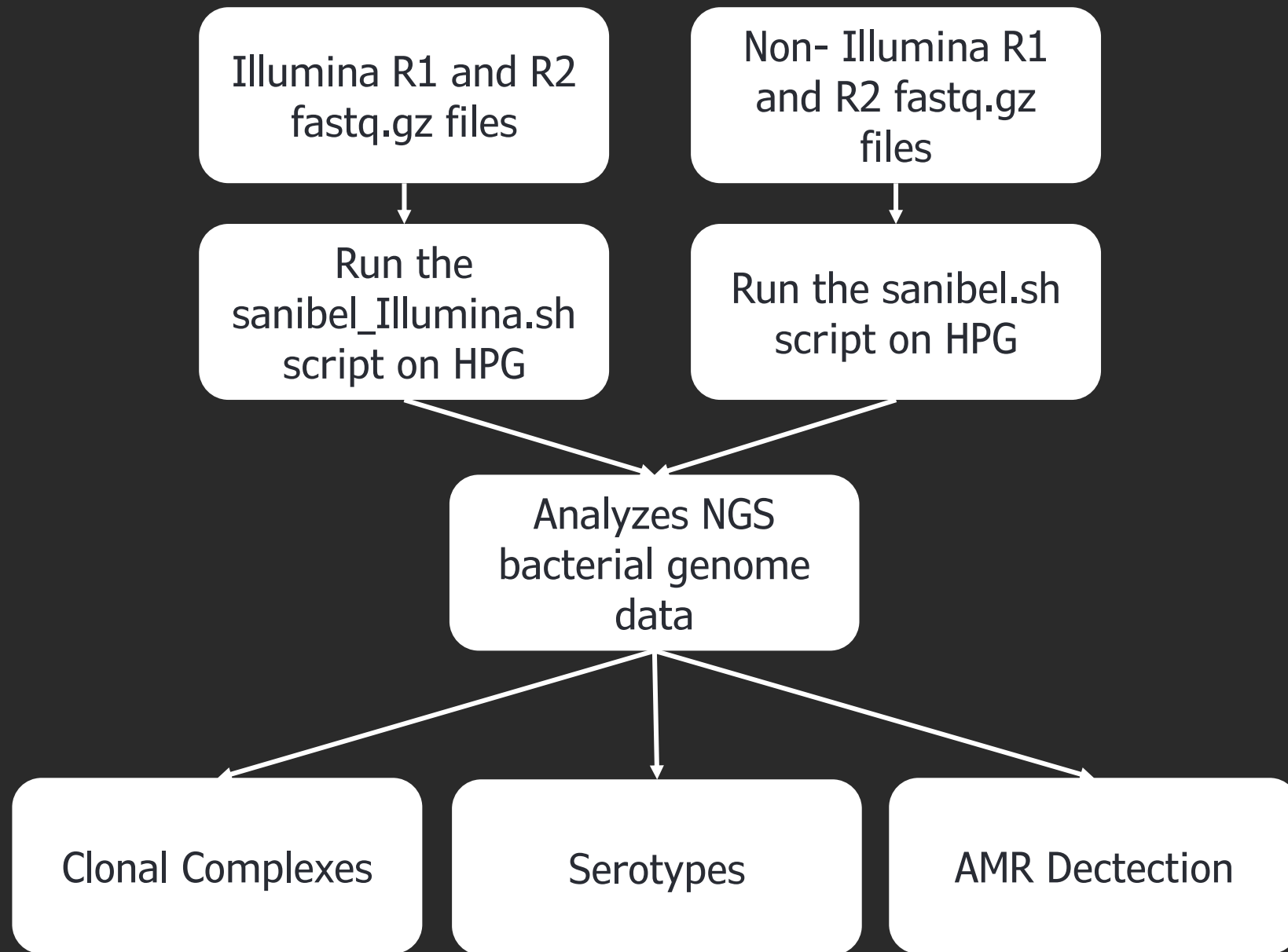
- To study bacterial strain diversity, monitor outbreaks, and detect resistance

## Dependencies

- Nextflow
- Python3 (need Pandas package)
- Singularity/ APPTAINER
- SLURM

# Overview





# Workflow



# Application

## Objective

Use an unknown dataset and analyze the dataset using Sanibel pipeline.



# Application Cont.

```
cd /blue/bphl-<state>/<user>/repos/bphl-molecular/
```

```
git clone https://github.com/BPHL-Molecular/Sanibel
```

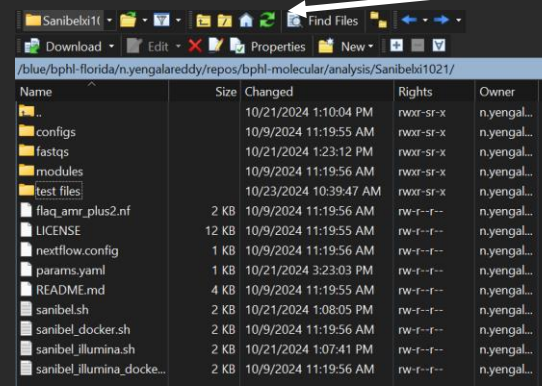
```
mkdir analysis/
```

```
cd analysis/
```

```
cp /blue/bphl-<state>/<user>/repos/bphl-molecular/Sanibel/*
```

```
Mkdir fastqs/
```

```
cp /path/to/fastqs/*.fastq.gz fastqs/
```



Name	Size	Changed	Rights	Owner
.		10/21/2024 1:10:04 PM	rw-r--r--	n.yengal...
configs		10/9/2024 11:19:55 AM	rw-r--r--	n.yengal...
fastqs		10/21/2024 1:23:12 PM	rw-r--r--	n.yengal...
modules		10/9/2024 11:19:56 AM	rw-r--r--	n.yengal...
test files		10/23/2024 10:39:47 AM	rw-r--r--	n.yengal...
flaq_amr_plus2.nf	2 KB	10/9/2024 11:19:56 AM	rw-r--r--	n.yengal...
LICENSE	12 KB	10/9/2024 11:19:55 AM	rw-r--r--	n.yengal...
nextflow.config	1 KB	10/9/2024 11:19:56 AM	rw-r--r--	n.yengal...
params.yaml	1 KB	10/21/2024 3:23:03 PM	rw-r--r--	n.yengal...
README.md	4 KB	10/9/2024 11:19:55 AM	rw-r--r--	n.yengal...
sanibel.sh	2 KB	10/21/2024 1:08:05 PM	rw-r--r--	n.yengal...
sanibel_docker.sh	2 KB	10/9/2024 11:19:56 AM	rw-r--r--	n.yengal...
sanibel_illumina.sh	2 KB	10/21/2024 1:07:41 PM	rw-r--r--	n.yengal...
sanibel_illumina_docke...	2 KB	10/9/2024 11:19:56 AM	rw-r--r--	n.yengal...

```
[n.yengalareddy@login8 Sanibel1021]$ ls
configs          nextflow.config      sanibel_illumina_docker.sh
fastqs           output-20241021184337 sanibel_illumina.sh
flaq_amr_plus2.nf params.yaml          sanibel.sh
LICENSE          README.md            'test files'
modules          sanibel_docker.sh
```



# Application Cont.



# Application Cont.

`nano params.yaml`

```
GNU nano 2.9.8 params.yaml
input      : "/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Sanibeli1021/fastqs"
output     : "/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Sanibeli1021/output"

# Note: The parameter "input" and "output" are the absolute path. Do not include the "/" at the end of the paths.
```





# Application Cont.

nano sanibel\_illumina.sh



```
#!/usr/bin/bash
#SBATCH --account=bphl-umbrella
#SBATCH --qos=bphl-umbrella
#SBATCH --job-name=sanibel
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=20          #This parameter should be equal to the number of samples if you want fastest results
#SBATCH --mem=200gb
#SBATCH --time=48:00:00
#SBATCH --output=sanibel.%j.out
#SBATCH --error=sanibel.%i.err
#SBATCH --mail-user=nikhil.vengala@flhealth.gov
#SBATCH --mail-type=FAIL,END

module load nextflow

APPTAINER_CACHEDIR=./
export APPTAINER_CACHEDIR

for f in ./fastqs/*R1*.gz
do
cp "$f" "${echo $f | cut -d - -f 1}_1.fastq.gz"
done
for f in ./fastqs/*R2*.gz
do
cp "$f" "${echo $f | cut -d - -f 1}_2.fastq.gz"
done

mkdir ./fastqs/original
mv ./fastqs/*_R1*.gz ./fastqs/original
mv ./fastqs/*_R2*.gz ./fastqs/original

singularity exec docker://staphb/mlst:2.23.0 cp /mlst-2.23.0/db/pubmlst/neisseria/neisseria.txt ./
singularity exec docker://staphb/mlst:2.23.0 cp /mlst-2.23.0/db/pubmlst/hinfluenzae/hinfluenzae.txt ./

nextflow run fla_q_amr_plus2.nf -params-file params.yaml

sort ./output/*/report.txt | uniq > ./output/sum_report.txt
sed -i 's/sampleID/tspeciesID/d' ./output/sum_report.txt
sed -i '11 sampleID\tspeciesID_mash\tnearest_neighb_mash\tmash_distance\tspeciesID_kraken\tkraken_percent\tmlst_scheme\tmlst_scheme\t' ./neisseria.txt
rm ./neisseria.txt
rm ./hinfluenzae.txt
```

nano sanibel.sh



```
#!/usr/bin/bash
#SBATCH --account=bphl-umbrella
#SBATCH --qos=bphl-umbrella
#SBATCH --job-name=sanibel
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=20          #This parameter should be equal to the number of samples if you want fastest results
#SBATCH --mem=200gb
#SBATCH --time=48:00:00
#SBATCH --output=sanibel.%j.out
#SBATCH --error=sanibel.err
#SBATCH --mail-user=nikhil.vengala@flhealth.gov
#SBATCH --mail-type=FAIL,END

module load nextflow
APPTAINER_CACHEDIR=./
export APPTAINER_CACHEDIR

singularity exec docker://staphb/mlst:2.23.0 cp /mlst-2.23.0/db/pubmlst/neisseria/neisseria.txt ./
singularity exec docker://staphb/mlst:2.23.0 cp /mlst-2.23.0/db/pubmlst/hinfluenzae/hinfluenzae.txt ./
nextflow run fla_q_amr_plus2.nf -params-file params.yaml
sort ./output/*/report.txt | uniq > ./output/sum_report.txt
sed -i 's/sampleID/tspeciesID/d' ./output/sum_report.txt
sed -i '11 sampleID\tspeciesID_mash\tnearest_neighb_mash\tmash_distance\tspeciesID_kraken\tkraken_percent\tmlst_scheme\tmlst_scheme\t' ./neisseria.txt
rm ./neisseria.txt
rm ./hinfluenzae.txt


mv ./*.out ./output
mv ./^err ./output

dt=$(date +%Y%m%d%H%M%S")
mv ./output ./output-$dt
#mv ./work ./work-$dt
rm -r ./work
rm -r ./cache
```



# Application Cont.

**sbatch** sanibel\_illumina.sh (or) sanibel.sh



Name	Size	Changed	Rights	Owner
..		10/21/2024 1:10:04 PM	rw-r--r--	n.yengal...
configs		10/9/2024 11:19:55 AM	rw-r--r--	n.yengal...
fastqs		10/21/2024 1:23:12 PM	rw-r--r--	n.yengal...
modules		10/9/2024 11:19:56 AM	rw-r--r--	n.yengal...
output-20241021184...		10/23/2024 10:39:47 AM	rw-r--r--	n.yengal...
test files		10/23/2024 1:36:19 PM	rw-r--r--	n.yengal...
flaq_amr_plus2.nf	2 KB	10/9/2024 11:19:56 AM	rw-r--r--	n.yengal...
LICENSE	12 KB	10/9/2024 11:19:55 AM	rw-r--r--	n.yengal...
nextflow.config	1 KB	10/9/2024 11:19:56 AM	rw-r--r--	n.yengal...
params.yaml	1 KB	10/21/2024 3:23:03 PM	rw-r--r--	n.yengal...
README.md	4 KB	10/9/2024 11:19:55 AM	rw-r--r--	n.yengal...
sanibel.sh	2 KB	10/21/2024 1:08:05 PM	rw-r--r--	n.yengal...
sanibel_docker.sh	2 KB	10/9/2024 11:19:56 AM	rw-r--r--	n.yengal...
sanibel_illumina.sh	2 KB	10/21/2024 1:07:41 PM	rw-r--r--	n.yengal...
sanibel_illumina_docke...	2 KB	10/9/2024 11:19:56 AM	rw-r--r--	n.yengal...

0 B of 23.0 KB in 0 of 14

9 hidden

🔒 SFTP-3 0:11:20



# Output

Name	Size	Changed	Rights	Owner
..		10/21/2024 1:10:04 PM	rw-r--r--	n.yengal...
configs		10/9/2024 11:19:55 AM	rw-r--r--	n.yengal...
fastqs		10/21/2024 1:23:12 PM	rw-r--r--	n.yengal...
modules		10/9/2024 11:19:56 AM	rw-r--r--	n.yengal...
output-20241021184...		10/23/2024 10:39:47 AM	rw-r--r--	n.yengal...
test files		10/23/2024 1:36:19 PM	rw-r--r--	n.yengal...
flaq_amr_plus2.nf	2 KB	10/9/2024 11:19:56 AM	rw-r--r--	n.yengal...
LICENSE	12 KB	10/9/2024 11:19:55 AM	rw-r--r--	n.yengal...
nextflow.config	1 KB	10/9/2024 11:19:56 AM	rw-r--r--	n.yengal...
params.yaml	1 KB	10/21/2024 3:23:03 PM	rw-r--r--	n.yengal...
README.md	4 KB	10/9/2024 11:19:55 AM	rw-r--r--	n.yengal...
sanibel.sh	2 KB	10/21/2024 1:08:05 PM	rw-r--r--	n.yengal...
sanibel_docker.sh	2 KB	10/9/2024 11:19:56 AM	rw-r--r--	n.yengal...
sanibel_illumina.sh	2 KB	10/21/2024 1:07:41 PM	rw-r--r--	n.yengal...
sanibel_illumina_docke...	2 KB	10/9/2024 11:19:56 AM	rw-r--r--	n.yengal...

0 B of 23.0 KB in 0 of 14







9 hidden

SFTP-3 0:11:20

Name	Size	Changed	Rights	Owner
SAMPLE1		10/23/2024 1:36:19 PM	rw-r--r--	n.yengal...
SAMPLE2		10/21/2024 3:41:50 PM	rw-r--r--	n.yengal...
sanibel.48189219.out	57 KB	10/21/2024 3:41:44 PM	rw-r--r--	n.yengal...
sanibel.err	2 KB	10/21/2024 3:42:19 PM	rw-r--r--	n.yengal...
sum_report.txt	1 KB	10/21/2024 3:42:19 PM	rw-r--r--	n.yengal...



# Output Cont.

Name	Size	Changed	Rights	Owner
 [empty]		10/23/2024 1:36:19 PM	rw-r--r--	n.yengal...
 SAMPLE1		10/21/2024 3:41:50 PM	rw-r--r--	n.yengal...
 SAMPLE2		10/21/2024 3:41:44 PM	rw-r--r--	n.yengal...
 sanibel.48189219.out	57 KB	10/21/2024 3:42:19 PM	rw-r--r--	n.yengal...
 sanibel.err	2 KB	10/21/2024 3:23:37 PM	rw-r--r--	n.yengal...
 <b>sum_report.txt</b>	1 KB	10/21/2024 3:42:19 PM	rw-r--r--	n.yengal...



	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	speciesID_mash	nearest_neighb_mash	mash_distance	speciesID_kraken	kraken_percent	mlst_scheme	mlst_st	mlst_cc	pmga_species	serotype	num_clean_reads	avg_readlength	avg_read_qual	est_coverage	num_contigs	longest_contig	N50	L50	total_length	gc_content	annotated_cds
2	Salmonella_enterica	NZ_AOXY	0.000483446	Salmonella enterica	97.61	senterica	50				449310	146.69	34.27	14.53	313	67899	24035	64	4534706	52	4196
3	Salmonella_enterica	NZ_AHUG	0.00357808	Salmonella enterica	96.85	senterica	46				651662	146.92	34.37	21.07	424	93501	17253	82	4542934	52	4274



Fundamentals of Sanibel



Installation and setup of  
Sanibel in HPG



Successfully executed  
job query for Sanibel



Generated Output Files

# Conclusion





# Advanced Molecular Detection

## Southeast Region Bioinformatics

# Questions?

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

**Molly Mitchell, PhD**

Bioinformatics Supervisor

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

**Nikhil Reddy, MS**

Bioinformatician

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

**Sam Marcellus, MPH**

Bioinformatician

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