FLAQ- SARS- COV- 2

Advanced Molecular Detection
Southeast Region Bioinformatics

November 18, 2024

Bioinformatics Updates From ARLN

Candida auris

- MycoSNP v 1.6 (Coming Soon-End of Year):
  - Clade Assignment Functionality
  - Genome Coverage and Depth Metrics
  - Additional Validated Hotspot Regions
  - Downloadable Clade References
- Exploring tools for Long-Sequencing Analysis

PHoeNIx

- Phlyophoenix:
  - Seamlessly integrates with Phoenix Workflow
  - Output supports direct upload to Microreact
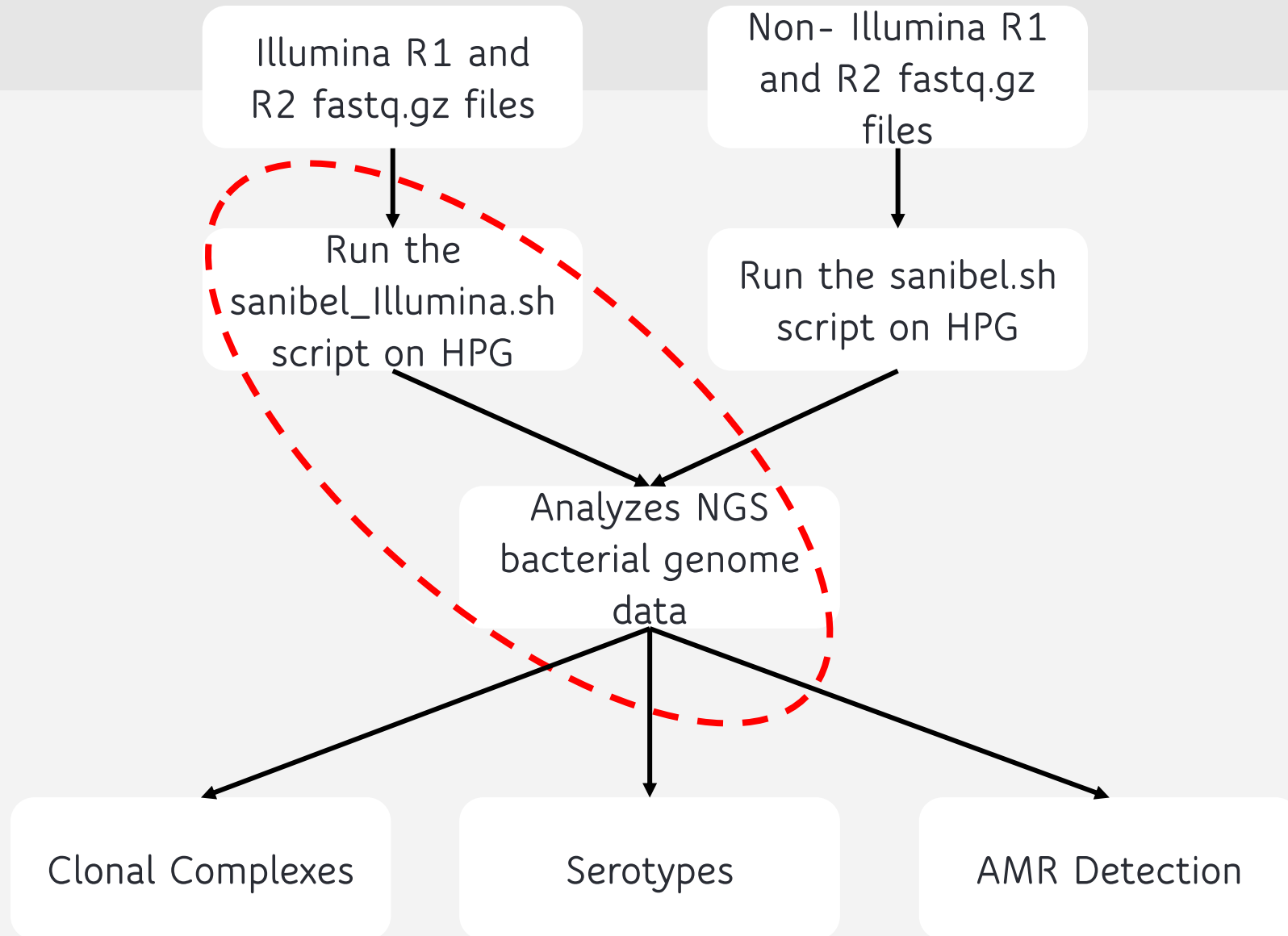 - New TSV output includes Epidemiology-Related Metadata

Office Hours

- November 25, 2024 – Daytona
- December 09, 2024 – FLAQ_SC2_Meta

Office Hours Updates

Recap of Sanibel

Sanibel Cont.

Sanibel Cont.

Purpose

- Analyzes SARS-COV-2 Genomic Data

- Generates Accurate Consensus Assemblies

Usage

- To support public health and researchers to monitor the spread and evolution of COVID-19

Dependencies

- Python3

- Singularity/Apptainer

- iVar

- Git

FLAQ-SC2 Overview

Advanced Molecular Detection
Southeast Region Bioinformatics

Workflow
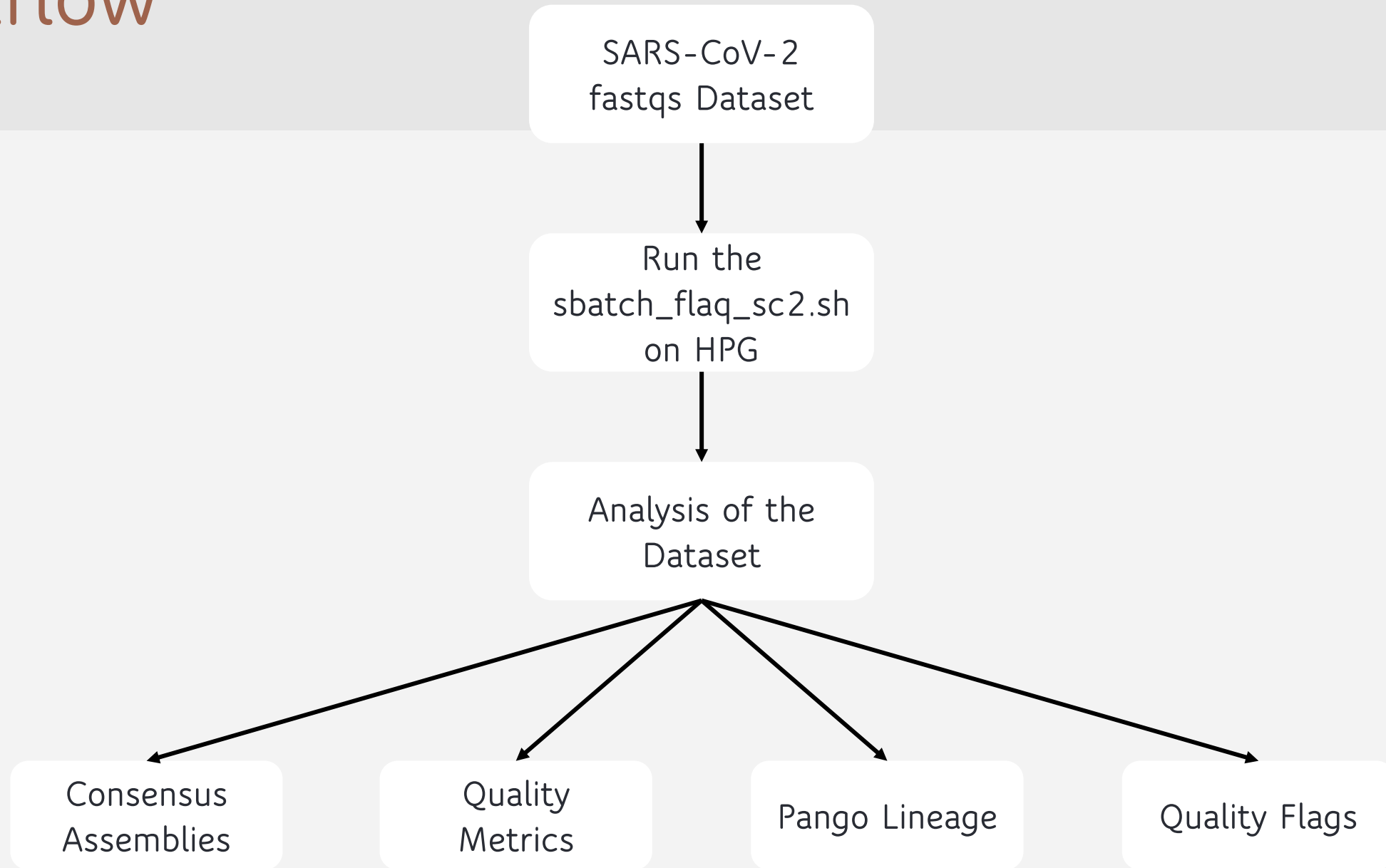
Application

Objective

- Use SC2-positive clinical datasets and analyze the dataset using FLAQ-SC2 pipeline.

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

git clone https://github.com/BPHL-Molecular/flaq_sc2

mkdir analysis/

cd analysis/

cp /blue/bphl-<state>/<user>/repos/bphl-molecular/flaq_sc2/*

mkdir fastqs/

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

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/flaq_sc2/*.*

Name	Size	Changed
		10/31/2024 12:51:14 PM
2024-10-30_flaq_run		10/30/2024 1:23:46 PM
assemblies		10/30/2024 1:24:57 PM
fastqs		10/30/2024 10:36:27 AM
primers		8/21/2024 8:16:31 AM
reference		8/21/2024 8:16:31 AM
vadr_error_reports		10/30/2024 1:21:39 PM
variants		10/30/2024 1:24:57 PM
assemblies.fasta	321 KB	10/30/2024 1:25:15 PM
flaq_sc2.48888753.err	63 KB	10/30/2024 1:25:13 PM
flaq_sc2.48888753.out	9 KB	10/30/2024 1:25:17 PM
flaq_sc2.py	19 KB	8/21/2024 8:16:31 AM
LICENSE	2 KB	8/21/2024 8:16:31 AM
nextclade_report.csv	5 KB	10/30/2024 1:25:17 PM
README.md	6 KB	8/21/2024 8:16:31 AM
sbatch_flaq_sc2.sh	1 KB	8/23/2024 9:10:21 AM

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

nano sbatch_flaq_sc2.sh

```bash
#!/bin/bash
#SBATCH --account=bphl-umbrella
#SBATCH --qos=bphl-umbrella
#SBATCH --job-name=flaq_sc2
#SBATCH --mail-type=END,FAIL
#SBATCH --mail-user=nikhil.yengala@flhealth.gov
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=15
#SBATCH --mem=100gb
#SBATCH --time=24:00:00
#SBATCH --output=flaq_sc2.%j.out
#SBATCH --error=flaq_sc2.%j.err

#Run script/command and use $SLURM_CPUS_ON_NODE
module load apptainer

module load conda

conda activate pyivar

primer_version="4.1"

python flaq_sc2.py fastqs/ --primer_bed primers/ARTIC-V${primer_version}.bed --lib_frag frag --threads $SLURM_CPUS_ON_NODE --r$
```

activate conda environment containing iVar
&
sbatch sbatch_flaq_sc2.sh

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/flaq_sc2/

Name	Size	Changed
..		10/31/2024 12:51:14 PM
2024-10-30_flaq_run		10/30/2024 1:23:46 PM
assemblies		10/30/2024 1:24:57 PM
fastqs		10/30/2024 10:36:27 AM
primers		8/21/2024 8:16:31 AM
reference		8/21/2024 8:16:31 AM
vadr_error_reports		10/30/2024 1:21:39 PM
variants		10/30/2024 1:24:57 PM
assemblies.fasta	321 KB	10/30/2024 1:25:15 PM
flaq_sc2.48888753.err	63 KB	10/30/2024 1:25:13 PM
flaq_sc2.48888753.out	9 KB	10/30/2024 1:25:17 PM
flaq_sc2.py	19 KB	8/21/2024 8:16:31 AM
LICENSE	2 KB	8/21/2024 8:16:31 AM
nextclade_report.csv	5 KB	10/30/2024 1:25:17 PM
README.md	6 KB	8/21/2024 8:16:31 AM
sbatch_flaq_sc2.sh	1 KB	8/23/2024 9:10:21 AM

Application Cont.

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

`ls variants`

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/flaq_sc2/variants/

Name	Size	Changed	Rights
📁		10/30/2024 1:25:17 PM	rwxr-sr-x
📄 J84.variants.tsv	1 KB	10/30/2024 1:10:27 PM	rw-r--r--
📄 J682.variants.tsv	2 KB	10/30/2024 1:06:39 PM	rw-r--r--
📄 J711.variants.tsv	1 KB	10/30/2024 1:08:06 PM	rw-r--r--
📄 J886.variants.tsv	1 KB	10/30/2024 1:11:49 PM	rw-r--r--
📄 M109.variants.tsv	1 KB	10/30/2024 1:14:46 PM	rw-r--r--
📄 M300.variants.tsv	2 KB	10/30/2024 1:16:08 PM	rw-r--r--
📄 M301.variants.tsv	2 KB	10/30/2024 1:17:31 PM	rw-r--r--
📄 M330.variants.tsv	1 KB	10/30/2024 1:19:11 PM	rw-r--r--
📄 M653.variants.tsv	2 KB	10/30/2024 1:21:33 PM	rw-r--r--
📄 M908.variants.tsv	1 KB	10/30/2024 1:23:28 PM	rw-r--r--
📄 T45.variants.tsv	1 KB	10/30/2024 1:24:57 PM	rw-r--r--

REGION	POS	REF	ALT	REF_DP	REF_RV	REF_QUAL	ALT_DP	ALT_RV	ALT_QUAL	ALT_FREQ	TOTAL_DP	PVAL	PASS	GFF_FEATURE	REF_CODON	REF_AA	ALT_CODON	ALT_AA	POS_AA
MN908947.3	241	C	T	0	0	0	400	86	48	1	400	1.93E-272	TRUE	NA	NA	NA	NA	NA	NA
MN908947.3	514	T	#NAME?	954	306	48	1180	0	20	0.907692	1300	5.94E-296	TRUE	NA	NA	NA	NA	NA	NA
MN908947.3	3037	C	T	0	0	0	1144	433	50	1	1144	0	TRUE	ORF1ab:cds-YP_009724389.1	TTC	F	TTT	F	924
MN908947.3	3037	C	T	0	0	0	1144	433	50	1	1144	0	TRUE	ORF1ab:cds-YP_009725295.1	TTC	F	TTT	F	924
MN908947.3	14408	C	T	1	1	76	5611	3615	47	0.999822	5612	0	TRUE	ORF1ab:cds-YP_009724389.1	CCT	P	CTT	L	314
MN908947.3	18877	C	T	2	1	37	2918	2036	47	0.999315	2920	0	TRUE	ORF1ab:cds-YP_009724389.1	CTA	L	TTA	L	1804
MN908947.3	23403	A	G	3	2	36	5722	3527	44	0.999476	5725	0	TRUE	S:cds-YP_009724390.1	GAT	D	GGT	G	614
MN908947.3	25563	G	T	0	0	0	1774	882	44	0.999437	1775	0	TRUE	ORF3a:cds-YP_009724391.1	CAG	Q	CAT	H	57

Application Cont.

Application Cont.

ls assemblies

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/flaq_sc2/assemblies/*.*

Name	Size	Changed	Rights
..		10/30/2024 1:25:17 PM	rwxr-sr-x
J84.consensus.fa	30 KB	10/30/2024 1:10:27 PM	rw-r--r--
J682.consensus.fa	30 KB	10/30/2024 1:06:39 PM	rw-r--r--
J711.consensus.fa	30 KB	10/30/2024 1:08:06 PM	rw-r--r--
J886.consensus.fa	30 KB	10/30/2024 1:11:49 PM	rw-r--r--
M109.consensus.fa	30 KB	10/30/2024 1:14:46 PM	rw-r--r--
M300.consensus.fa	30 KB	10/30/2024 1:16:08 PM	rw-r--r--
M301.consensus.fa	30 KB	10/30/2024 1:17:31 PM	rw-r--r--
M330.consensus.fa	30 KB	10/30/2024 1:19:11 PM	rw-r--r--
M653.consensus.fa	30 KB	10/30/2024 1:21:33 PM	rw-r--r--
M908.consensus.fa	30 KB	10/30/2024 1:23:28 PM	rw-r--r--
T45.consensus.fa	30 KB	10/30/2024 1:24:57 PM	rw-r--r--

Advanced Molecular Detection
Southeast Region Bioinformatics

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/flaq_sc2/2024-10-30_flaq_run/*.*

Name	Size	Changed	Rights
..		10/30/2024 1:25:17 PM	rwxr-sr-x
J84		10/30/2024 1:10:19 PM	rwxr-sr-x
J682		10/30/2024 1:06:38 PM	rwxr-sr-x
J711		10/30/2024 1:08:04 PM	rwxr-sr-x
J886		10/30/2024 1:11:48 PM	rwxr-sr-x
M109		10/30/2024 1:14:37 PM	rwxr-sr-x
M300		10/30/2024 1:16:06 PM	rwxr-sr-x
M301		10/30/2024 1:17:30 PM	rwxr-sr-x
M330		10/30/2024 1:19:08 PM	rwxr-sr-x
M653		10/30/2024 1:21:28 PM	rwxr-sr-x
M908		10/30/2024 1:23:24 PM	rwxr-sr-x
T45		10/30/2024 1:24:55 PM	rwxr-sr-x
report.txt	2 KB	10/30/2024 1:25:15 PM	rw-r--r--

sampleID	reference	start	end	num_raw_reads	num_clean_reads	num_mapped_reads	percent_mapped_clean_reads	cov_bases_mapped	percent_genome_cov_map	mean_depth	mean_base_qual	mean_map_qual	assembly_length	numN	percent_ref_genome_cov	VADR_flag	QC_flag	pangolin_version	lineage	SOTC
J682	MN908947.3	1	29903	176476	125270	71227	56.8588	29497	98.6423	503.937	37.8	60	29777	1310	95.1978	PASS	PASS	v4.3.1_pdata-v1.29	B.1	
J711	MN908947.3	1	29903	442888	279008	167235	59.9391	29818	99.7157	776.595	35.6	60	29829	650	97.5788	PASS	PASS	v4.3.1_pdata-v1.29	B.1	
J84	MN908947.3	1	29903	733132	606216	458864	75.6932	29868	99.883	3602.92	37.7	60	29862	50	99.6957	PASS	PASS	v4.3.1_pdata-v1.29	B.1	
J886	MN908947.3	1	29903	467264	295070	143787	48.7298	29634	99.1004	673.339	35.6	60	29777	1125	95.8165	PASS	PASS	v4.3.1_pdata-v1.29	B.1	
M109	MN908947.3	1	29903	1062220	898804	550019	61.1945	29858	99.8495	4225.17	37.7	60	29858	115	99.4649	PASS	PASS	v4.3.1_pdata-v1.29	B.1	
M300	MN908947.3	1	29903	289620	252704	121936	48.2525	29718	99.3813	878.313	37.8	60	29773	1439	94.753	REVIEW	PASS	v4.3.1_pdata-v1.29	B.1	
M301	MN908947.3	1	29903	447296	349038	67383	19.3053	29585	98.9366	494.466	37.8	60	29754	3041	89.3322	REVIEW	PASS	v4.3.1_pdata-v1.29	B.1	
M330	MN908947.3	1	29903	407520	363380	169527	46.6528	29756	99.5084	1226.17	37.8	60	29774	726	97.1408	PASS	PASS	v4.3.1_pdata-v1.29	B.1	
M653	MN908947.3	1	29903	872474	758820	333828	43.993	29825	99.7392	2292.56	37.8	60	29813	664	97.4785	REVIEW	PASS	v4.3.1_pdata-v1.29	B.1	
M908	MN908947.3	1	29903	530524	467348	253716	54.2885	29807	99.679	1836.15	37.8	60	29807	90	99.378	PASS	PASS	v4.3.1_pdata-v1.29	B.1	
T45	MN908947.3	1	29903	442412	340932	130766	38.3554	29777	99.5786	757.439	37.7	60	29777	1370	94.9972	PASS	PASS	v4.3.1_pdata-v1.29	B.1	

Application Cont.

Advanced Molecular Detection
Southeast Region Bioinformatics

Conclusion

Recap of Sanibel

Fundamentals of FLAQ-SC2

Installation and setup of FLAQ-SC2 in HPG

Successfully executed job query for FLAQ-SC2

Generated Output Files

Advanced Molecular Detection
Southeast Region Bioinformatics

Advanced Molecular Detection Southeast Region Bioinformatics

Questions?

bphl-sebioinformatics@flhealth.gov

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

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

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