

Daytona

Advanced Molecular Detection
Southeast Region Bioinformatics

November 25, 2024

U P D A T E S

Office Hours

- December 09, 2024 – FLAQ_SC2_Meta
- December 23, 2024 – Daytona_Dengue



Overview

Purpose

Nextflow version of the FLAQ_SC2 pipeline + Added function of human read removal is added

Usage

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

Dependencies

Nextflow

Python3

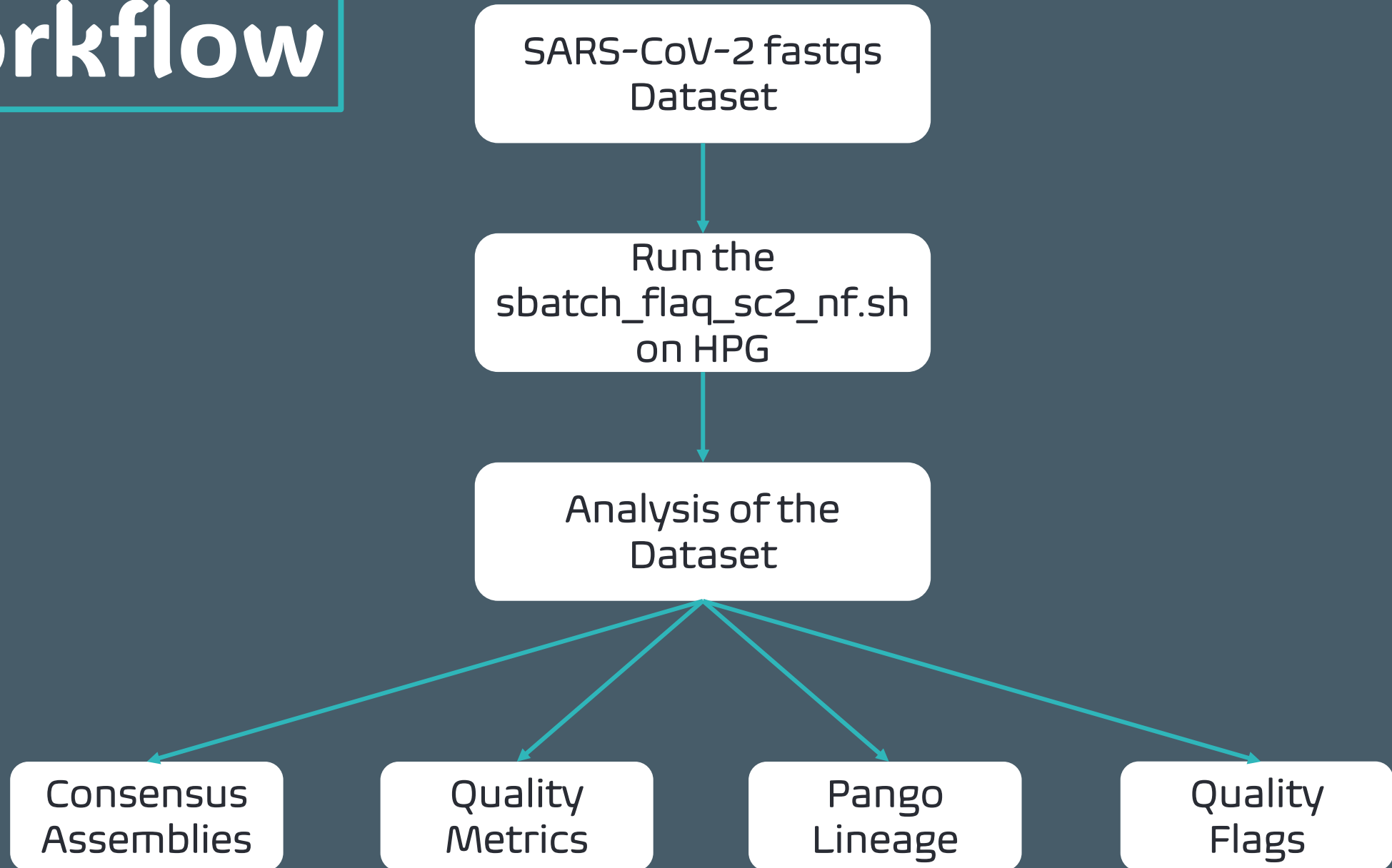
Singularity

Docker Container Images

- | | |
|---|-------------------------|
| 1. fastqc_0.11.9.sif | 2. trimmomatic_0.39.sif |
| 3. bbtools_38.76.sif | 4. multiqc_1.8.sif |
| 5. bwa_0.7.17.sif | 6. samtools_1.12.sif |
| 7. vadr_1.3.sif | |
| 8. pangolin_4.1.2-pdata-1.13.sif | |
| 9. nextclade_2021-03-15.sif | |
| 10. sra-human-scrubber_1.1.2021-05-05.sif | |



Workflow



Application

Objective

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



Application Cont.

```
cd /blue/bphl-<state>/<user>/repos/bphl-molecular/  
git clone https://github.com/BPHL-Molecular/Daytona  
mkdir analysis/  
cd analysis/  
cp /blue/bphl-<state>/<user>/repos/bphl-  
molecular/Daytona/*  
mkdir fastqs/  
cp /path/to/fastqs/*.fastq.gz fastqs/
```

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Daytona1107/*.*

Name	Size	Changed
..		11/7/2024 9:18:47 AM
apps		11/7/2024 11:06:53 AM
fastqs		11/7/2024 9:19:57 AM
output		11/12/2024 9:16:17 AM
primers		11/6/2024 1:01:00 PM
reference		11/6/2024 1:01:00 PM
work		11/12/2024 9:15:01 AM
flaq_sc2_humandclean.nf	18 KB	11/6/2024 1:01:00 PM
flaq_sc2_humandclean2.nf	19 KB	11/6/2024 1:01:00 PM
flaq_sc2_nf.49628631.err	0 KB	11/12/2024 9:09:43 AM
flaq_sc2_nf.49628631.out	52 KB	11/12/2024 9:16:17 AM
flaq_sc2_nf.nf	18 KB	11/6/2024 1:01:00 PM
LICENSE	2 KB	11/6/2024 1:00:58 PM
nextflow	18 KB	11/8/2024 3:01:12 PM
params.yaml	1 KB	11/7/2024 9:37:31 AM
README.md	2 KB	11/6/2024 1:00:58 PM
sbatch_flaq_sc2_nf.sh	2 KB	11/12/2024 9:08:50 AM



Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Daytona1107/apps/staphb-toolkit/containers/		
Name	Size	Changed
bbtools_38.76.sif	506,204 KB	11/8/2024 2:50:09 PM
bwa_0.7.17.sif	98,040 KB	10/19/2020 8:29:15 AM
fastqc_0.11.9.sif	167,016 KB	12/16/2021 12:58:43 PM
multiqc_1.8.sif	1,266,688 KB	10/18/2020 9:06:46 AM
nextclade_2021-03-15.sif	323,820 KB	3/15/2021 11:30:42 AM
pangolin_4.1.2-pdata-1.13.sif	710,076 KB	8/23/2022 7:29:35 AM
samtools_1.12.sif	117,248 KB	4/14/2021 6:55:50 AM
sra-human-scrubber_1.1.2021-05-05.sif	1,519,872 KB	8/30/2022 11:12:28 AM
trimmomatic_0.39.sif	197,320 KB	10/18/2020 9:03:45 AM
vadr_1.3.sif	727,704 KB	9/6/2021 7:19:16 AM

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Daytona1107/*.*		
Name	Size	Changed
..		11/7/2024 9:18:47 AM
apps		11/7/2024 11:06:53 AM
fastqs		11/7/2024 9:19:57 AM
output		11/12/2024 9:16:17 AM
primers		11/6/2024 1:01:00 PM
reference		11/6/2024 1:01:00 PM
work		11/12/2024 9:15:01 AM
flaq_sc2_humanclean.nf	18 KB	11/6/2024 1:01:00 PM
flaq_sc2_humanclean2.nf	19 KB	11/6/2024 1:01:00 PM
flaq_sc2_nf.49628631.err	0 KB	11/12/2024 9:09:43 AM
flaq_sc2_nf.49628631.out	52 KB	11/12/2024 9:16:17 AM
flaq_sc2_nf.nf	18 KB	11/6/2024 1:01:00 PM
LICENSE	2 KB	11/6/2024 1:00:58 PM
nextflow	18 KB	11/8/2024 3:01:12 PM
params.yaml	1 KB	11/7/2024 9:37:31 AM
README.md	2 KB	11/6/2024 1:00:58 PM
sbatch_flaq_sc2_nf.sh	2 KB	11/12/2024 9:08:50 AM

Copy Apps From:

/apps/staphb-toolkit/containers

Or

ncbi/sra-human-scrubber &
StaPH-B/docker-builds



Application Cont.

nano params.yaml



GNU nano 2.9.8


params.yaml

```
# The parameters "input", "output", "reference" and "primer" are the absolute paths. Do not include the "/" at the end of the paths.
input      : "/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Daytona1107/fastqs"
output     : "/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Daytona1107/output"
reference   : "/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Daytona1107/reference"
primer     : "/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Daytona1107/primers"
# specify if input amplicons were fragmented. If yes, the parameter setting is "frag". If no, the parameter setting is "no_frag". (default: frag)
frag : "no_frag"
# comma separated list of SOTCs to screen, default as "S:L452R,S:E484K"
sotc : "S:L452R,S:E484K"
```



Application Cont.

`nano sbatch_flgq_sc2_nf.sh`



```
GNU nano 2.9.8 sbatch_flgq_sc2_nf.sh

#!/bin/bash
#SBATCH --account=bph1-umbrella
#SBATCH --gos=bph1-umbrella
#SBATCH --job-name=flaq_sc2_nf
#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_nf.%j.out
#SBATCH --error=flaq_sc2_nf.%j.err

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

nextflow run flaq_sc2_humanclean2.nf -params-file params.yaml

sort ./output/*/report.txt | uniq > ./output/sum_report.txt
sed -i 's/sampleID\treference/d' ./output/sum_report.txt
sed -i 'i sampleID\treference\tstart\tend\tnum_raw_reads\tnum_clean_reads\tnum_mapped_reads\tpercent_mapped_clean_reads\tcov_bases_mapped\tpercent_genome_cov_map\tmean_depth\tmean_base_qual' ./output/sum_report.txt

cat ./output/assemblies/*.fa > ./output/assemblies.fasta
singularity exec /apps/staphb-toolkit/containers/nextclade_2021-03-15.sif nextclade --input-fasta ./output/assemblies.fasta --output-csv ./output/nextclade_report.csv
```



Application Cont.

activate conda environment containing Nextflow
&
sbatch sbatch_flgq_sc2_nf.sh

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Daytona1107/*.*

Name	Size	Changed
..		11/7/2024 9:18:47 AM
apps		11/7/2024 11:06:53 AM
fastqs		11/7/2024 9:19:57 AM
output		11/12/2024 9:16:17 AM
primers		11/6/2024 1:01:00 PM
reference		11/6/2024 1:01:00 PM
work		11/12/2024 9:15:01 AM
flaq_sc2_humanclean.nf	18 KB	11/6/2024 1:01:00 PM
flaq_sc2_humanclean2.nf	19 KB	11/6/2024 1:01:00 PM
flaq_sc2_nf.49628631.err	0 KB	11/12/2024 9:09:43 AM
flaq_sc2_nf.49628631.out	52 KB	11/12/2024 9:16:17 AM
flaq_sc2_nf.nf	18 KB	11/6/2024 1:01:00 PM
LICENSE	2 KB	11/6/2024 1:00:58 PM
nextflow	18 KB	11/8/2024 3:01:12 PM
params.yaml	1 KB	11/7/2024 9:37:31 AM
README.md	2 KB	11/6/2024 1:00:58 PM
sbatch_flgq_sc2_nf.sh	2 KB	11/12/2024 9:08:50 AM



Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Daytona1107/output/

Name	Size	Changed
..		11/14/2024 10:08:36 AM
assemblies		11/12/2024 9:15:48 AM
J84		11/12/2024 9:14:55 AM
J682		11/12/2024 9:12:25 AM
J711		11/12/2024 9:13:03 AM
J886		11/12/2024 9:13:01 AM
M109		11/12/2024 9:16:11 AM
M300		11/12/2024 9:13:06 AM
M301		11/12/2024 9:13:37 AM
M330		11/12/2024 9:13:38 AM
M653		11/12/2024 9:15:14 AM
M908		11/12/2024 9:14:05 AM
T45		11/12/2024 9:13:28 AM
vadr_error_reports		11/12/2024 9:14:56 AM
variants		11/12/2024 9:15:48 AM
assemblies.fasta	321 KB	11/12/2024 9:16:11 AM
nextclade_report.csv	5 KB	11/12/2024 9:16:17 AM
sum_report.txt	2 KB	11/12/2024 9:16:11 AM

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Daytona1107/**

Name	Size	Changed
..		11/7/2024 9:18:47 AM
apps		11/7/2024 11:06:53 AM
fastqs		11/7/2024 9:19:57 AM
output		11/12/2024 9:16:17 AM
primers		11/6/2024 1:01:00 PM
reference		11/6/2024 1:01:00 PM
work		11/12/2024 9:15:01 AM
flaq_sc2_humandclean.nf	18 KB	11/6/2024 1:01:00 PM
flaq_sc2_humandclean2.nf	19 KB	11/6/2024 1:01:00 PM
flaq_sc2_nf.49628631.err	0 KB	11/12/2024 9:09:43 AM
flaq_sc2_nf.49628631.out	52 KB	11/12/2024 9:16:17 AM
flaq_sc2_nf.nf	18 KB	11/6/2024 1:01:00 PM
LICENSE	2 KB	11/6/2024 1:00:58 PM
nextflow	18 KB	11/8/2024 3:01:12 PM
params.yaml	1 KB	11/7/2024 9:37:31 AM
README.md	2 KB	11/6/2024 1:00:58 PM
sbatch_flaq_sc2_nf.sh	2 KB	11/12/2024 9:08:50 AM

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	105571	84.2748	29526	98.7393	756.939	37.8	60	29801	1317	95.2547	PASS	PASS	v4.1.2_pdata-v1.13	B.1	
J711	MN908947.3	1	29903	442888	279008	277991	99.6355	29830	99.7559	1276.76	35.5	60	29841	673	97.5421	PASS	PASS	v4.1.2_pdata-v1.13	B.1	
J84	MN908947.3	1	29903	733132	606216	595444	98.2231	29849	99.8194	4674.64	37.7	60	29843	18	99.7392	PASS	PASS	v4.1.2_pdata-v1.13	B.1	
J886	MN908947.3	1	29903	467264	295070	285981	96.9197	29664	99.2007	1319.11	35.5	60	29802	1126	95.8967	PASS	PASS	v4.1.2_pdata-v1.13	B.1	
M109	MN908947.3	1	29903	1062220	898804	878602	97.7523	29858	99.8495	6765.29	37.7	60	29858	105	99.4984	PASS	PASS	v4.1.2_pdata-v1.13	B.1	
M300	MN908947.3	1	29903	289620	252704	240562	95.1952	29703	99.3312	1784.04	37.7	60	29758	1392	94.86	REVIEW	PASS	v4.1.2_pdata-v1.13	B.1	
M301	MN908947.3	1	29903	447296	349038	119960	34.3688	29606	99.0068	900.226	37.7	60	29759	2959	89.6231	REVIEW	PASS	v4.1.2_pdata-v1.13	B.1.124	
M330	MN908947.3	1	29903	407520	363380	359808	99.017	29741	99.4582	2692.15	37.8	60	29759	688	97.2177	PASS	PASS	v4.1.2_pdata-v1.13	B.1	
M653	MN908947.3	1	29903	872474	758820	754273	99.4008	29838	99.7826	5429.88	37.7	60	29825	634	97.619	REVIEW	PASS	v4.1.2_pdata-v1.13	B.1	
M908	MN908947.3	1	29903	530524	467348	467739	100.0837	29802	99.6622	3472.95	37.7	60	29802	91	99.3579	PASS	PASS	v4.1.2_pdata-v1.13	B.1	
T45	MN908947.3	1	29903	442412	340932	246111	72.1877	29801	99.6589	1443.61	37.7	60	29801	1427	94.8868	PASS	PASS	v4.1.2_pdata-v1.13	B.1	



Application Cont.

Name	Size	Changed
assemblies		11/14/2024 10:08:36 AM
J84		11/12/2024 9:15:48 AM
J682		11/12/2024 9:14:55 AM
J711		11/12/2024 9:12:25 AM
J886		11/12/2024 9:13:03 AM
M109		11/12/2024 9:13:01 AM
M300		11/12/2024 9:16:11 AM
M301		11/12/2024 9:13:06 AM
M330		11/12/2024 9:13:37 AM
M653		11/12/2024 9:13:38 AM
M908		11/12/2024 9:15:14 AM
T45		11/12/2024 9:14:05 AM
vadr_error_reports		11/12/2024 9:13:28 AM
variants		11/12/2024 9:14:56 AM
assemblies.fasta	321 KB	11/12/2024 9:15:48 AM
nextclade_report.csv	5 KB	11/12/2024 9:16:11 AM
sum_report.txt	2 KB	11/12/2024 9:16:11 AM

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM
seqName;clade;qc.overallScore;qc.overallStatus,totalGaps;totalInsertions;totalMissing,totalMutations;totalNonACGTNs;totalPrCPrimerChanges;substitutions;deletions;insertions;missing;nonACGTNs;prCPrimerChanges;aaSubstitutions;totalAminoacidSubstitutions;aaDeletions;totalAminoacidDeletions;alignmentEnd;alignmentScore;alignmentStart;qc.missingData.missingDataThreshold;qc.missingData.score;qc.missingData.percentMissing																																						
J682;20A;;T3011C	C3037T	C9438T	C14408T	A23403G	C25904T;	6803-7325	9866-9874	12953-125	1986-195	20516-205	21449-214	221256-221	22198-222	22298-225	27259-273	29832-298	ORF1b:P3: ORF3a:S1:	S:D614G;4,,0;	29839;89373;39;3000;37.666666666666664	mediocre;1317;10;0;good;0;24;-5;0;good;3,,0;good;0;																		
J711;20A;;C3037T	C14408T	A23403G;	6862-7043	19290-195	22334-225	29841-298	S:D614G;2,,0;	29878;89505;38;3000;13.814814814814815	good;673;10;0;good;0;24;-8;0;good;0,,0;good;0;																													
J84;20A;;C3037T	C14408T	C18877T	A23403G	G25563T;	29841-298	ORF3a:Q5 S:D614G;3;ORF1a:V86;-2;	29849;89499;1;3000;0;good;18;10;0;good;0;24;-7;0;good;1,,0;good;0;																															
J86;20A;;T3011C	C3037T	C14408T	A23403G;	6885-6935	6936-6964	7062-7072	7074-7312	19291-195	20516-205	21163-211	21373-214	21439-214	22334-225	27528-275	27590-277	27718-277	S:D614G;2,,0;	29839;89384;38;3000;30.592592592592595	mediocre;1126;10;0;good;0;24;-7;0;good;1,,0;good;0;																			
M109;20A C3037T	C14408T	C18877T	A19170T	A23403G	G25563T	T26717T;	22340-224	29832-298	ORF1a:Q5 S:D614G;3,,0;	29862;89540;5;3000;0;good;105;10;0;good;0;24;-6;0;good;2,,0;good;0;																												
M300;20A C3037T	C14408T	C18877T	A23403G	G25563T;	12585-2618	6865-7318	8656-8895	19291-195	22160-221	22182-225	29832-298	ORF1b:S1; ORF3a:Q5 S:D614G;4;ORF1b:T1202;-1,	29839;89244;78;3000;40.444444444444444	mediocre;1392;10;0;good;0;24;-7;0;good;1,,0;good;0;																								
M301;20A C14408T	C18877T	A23403G	G25563T;;	1916	2133-2185	2585-284C	7050-7321	8651-8897	9572-9595	14921-145	15148-151	16202-164	19292-195	20197	20205-202	20226-203	20362-204	22085-221	22141-225	24164-241	25923-262	27528-277	29810-298	ORF3a:Q5 S:D614G;3,,0;	29835;89248;78;3000;98.48148148148148	mediocre;2959;10;0;good;0;24;-7;0;good;1,,0;good;0;												
M330;20A C3037T	C14408T	A23403G;	7068-7312	20218-202	20472-204	22311-225	29762-298	ORF1b:P3: S:D614G;3;ORF1a:F1928;-1;	29839;89255;78;3000;14.37037037037037	good;688;10;0;good;1;24;-7;0;good;1,,0;good;0;																												
M653;20A C3037T	G10841T	C14408T	G15406T	C18877T	C21359T	A23403G	G25563T;	27794-277	6866-7215	7232-7241	7271-7273	7304-7312	22140-224	ORF1b:P3: ORF1b:A6 ORF1b:T2l ORF3a:Q5 ORF7b:F1: S:D294X	S:D614G;8,S:P295-S:L296-	S:S297-4;	29839;89429;2;3000;12.37037037037037	good;634;10;0;good;0;24;-3;0;good;5;0;good;0;																				
N908;20A C3037T	A13922G	C14408T	C18877T	C19170T	A23403G	G25563T;;	19314-193	22327-222	22273-222	ORF1b:P3: ORF3a:Q5 S:D614G;4,,0;	29839;89372;38;3000;0;good;91;10;0;good;0;24;-6;0;good;2,,0;good;0;																											
T45;20A;1,C3037T	C14408T	C18877T	C19170T	A23403G	G25563T;	7155-7196	7200-7253	7260-7323	8334-8393	8461-8637	8641	19287-195	21162-211	21434-216	22335-225	27545-278	29837-298	ORF3a:Q5 S:D614G;3,,0;	29839;89373;39;3000;41.74074074074074	mediocre;1427;10;0;good;0;24;-7;0;good;1,,0;good;0;																		

Application Cont.

Name	Size	Changed
assemblies		11/14/2024 10:08:36 AM
J84		11/12/2024 9:15:48 AM
J682		11/12/2024 9:14:55 AM
J711		11/12/2024 9:12:25 AM
J886		11/12/2024 9:13:03 AM
M109		11/12/2024 9:13:01 AM
M300		11/12/2024 9:16:11 AM
M301		11/12/2024 9:13:06 AM
M330		11/12/2024 9:13:37 AM
M653		11/12/2024 9:13:38 AM
M908		11/12/2024 9:15:14 AM
T45		11/12/2024 9:14:05 AM
vadr_error_reports		11/12/2024 9:13:28 AM
variants		11/12/2024 9:14:56 AM
assemblies.fasta	321 KB	11/12/2024 9:15:48 AM
nextclade_report.csv	5 KB	11/12/2024 9:16:11 AM
sum_report.txt	2 KB	11/12/2024 9:16:11 AM

[illegible]

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Daytona1107/output/variants/		
Name	Size	Changed
..		11/12/2024 9:16:17 AM
J84.variants.tsv	2 KB	11/12/2024 9:14:32 AM
J682.variants.tsv	2 KB	11/12/2024 9:11:47 AM
J711.variants.tsv	1 KB	11/12/2024 9:12:25 AM
J886.variants.tsv	1 KB	11/12/2024 9:12:25 AM
M109.variants.tsv	1 KB	11/12/2024 9:15:48 AM
M300.variants.tsv	2 KB	11/12/2024 9:12:25 AM
M301.variants.tsv	2 KB	11/12/2024 9:13:02 AM
M330.variants.tsv	1 KB	11/12/2024 9:13:02 AM
M653.variants.tsv	2 KB	11/12/2024 9:14:47 AM
M908.variants.tsv	1 KB	11/12/2024 9:13:39 AM
T45.variants.tsv	1 KB	11/12/2024 9:12:55 AM

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Daytona1107/output/		
Name	Size	Changed
assemblies		11/14/2024 10:08:36 AM
J84		11/12/2024 9:14:55 AM
J682		11/12/2024 9:12:25 AM
J711		11/12/2024 9:13:03 AM
J886		11/12/2024 9:13:01 AM
M109		11/12/2024 9:16:11 AM
M300		11/12/2024 9:13:06 AM
M301		11/12/2024 9:13:37 AM
M330		11/12/2024 9:13:38 AM
M653		11/12/2024 9:15:14 AM
M908		11/12/2024 9:14:05 AM
T45		11/12/2024 9:13:28 AM
vadr_error_reports		11/12/2024 9:14:56 AM
variants		11/12/2024 9:15:48 AM
assemblies.fasta	321 KB	11/12/2024 9:16:11 AM
nextclade_report.csv	5 KB	11/12/2024 9:16:17 AM
sum_report.txt	2 KB	11/12/2024 9:16:11 AM

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	436	29	48	1	436	1.36E-296	TRUE	NA	NA	NA	NA	NA	NA
MN908947.3	514	T	#NAME?	1070	185	48	1332	0	20	0.909215	1465	0	TRUE	NA	NA	NA	NA	NA	NA
MN908947.3	520	G	T	2	0	76	8	3	34	0.8	10	8.60E-19	TRUE	ORF1ab:cds-YP_009724389.1	ATG	M	ATT	I	85
MN908947.3	520	G	T	2	0	76	8	3	34	0.8	10	8.60E-19	TRUE	ORF1ab:cds-YP_009725295.1	ATG	M	ATT	I	85
MN908947.3	3037	C	T	0	0	0	1252	265	49	1	1252	0	TRUE	ORF1ab:cds-YP_009724389.1	TTC	F	TTT	F	924
MN908947.3	3037	C	T	0	0	0	1252	265	49	1	1252	0	TRUE	ORF1ab:cds-YP_009725295.1	TTC	F	TTT	F	924
MN908947.3	14408	C	T	2	0	57	5933	2609	45	0.999495	5936	0	TRUE	ORF1ab:cds-YP_009724389.1	CCT	P	CTT	L	314
MN908947.3	18877	C	T	3	1	37	3607	2082	47	0.999169	3610	0	TRUE	ORF1ab:cds-YP_009724389.1	CTA	L	TTA	L	1804
MN908947.3	23403	A	G	5	3	44	6298	3566	42	0.999207	6303	0	TRUE	S:cds-YP_009724390.1	GAT	D	GGT	G	614
MN908947.3	25563	G	T	0	0	0	2166	945	44	0.999539	2167	0	TRUE	ORF3a:cds-YP_009724391.1	CAG	Q	CAT	H	57



Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Daytona1107/output/vadr_error_reports/			
Name	Size	Changed	
..		11/12/2024 9:16:17 AM	
x M300.vadr.alt.list	4 KB	11/12/2024 9:12:37 AM	
x M301.vadr.alt.list	4 KB	11/12/2024 9:13:12 AM	
x M653.vadr.alt.list	1 KB	11/12/2024 9:14:56 AM	

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Daytona1107/output/		
Name	Size	Changed
..		11/14/2024 10:08:36 AM
assemblies		11/12/2024 9:15:48 AM
84		11/12/2024 9:14:55 AM
682		11/12/2024 9:12:25 AM
711		11/12/2024 9:13:03 AM
886		11/12/2024 9:13:01 AM
M109		11/12/2024 9:16:11 AM
M300		11/12/2024 9:13:06 AM
M301		11/12/2024 9:13:37 AM
M330		11/12/2024 9:13:38 AM
M653		11/12/2024 9:15:14 AM
M908		11/12/2024 9:14:05 AM
T45		11/12/2024 9:13:28 AM
vadr_error_reports		11/12/2024 9:14:56 AM
variants		11/12/2024 9:15:48 AM
assemblies.fasta	321 KB	11/12/2024 9:16:11 AM
nextclade_report.csv	5 KB	11/12/2024 9:16:17 AM
sum_report.txt	2 KB	11/12/2024 9:16:11 AM

#sequence	model	feature-type	feature-name	error	seq-coords	mdl-coords	error-description
M300	NC_045512	CDS	ORF1ab polyprotein	CDS_HAS_STOP_CODON	16990..16992:+	17120..17122:+	in-frame stop codon exists 5' of stop position predicted by homology to reference [TAG, shifted S:4433,M:4433]
M300	NC_045512	CDS	ORF1ab polyprotein	POSSIBLE_FRAMESHIFT	16944..21425:+	17070..21555:+	possible frameshift at 3' end of CDS [length:4482; inserts:none; deletes:S:16943,M:17070..17073(4); shifted_frame:3; dominant_frame:1;]
M300	NC_045512	CDS	ORF1ab polyprotein	INDEFINITE_ANNOTATION_END	16948..21425:+	21555..21555:+	protein-based alignment does not extend close enough to nucleotide-based alignment 3' endpoint [4478>8]
M300	NC_045512	CDS	ORF1ab polyprotein	UNEXPECTED_LENGTH	140..13342:+,13342..21425:+	266..13468:+,13468..21555:+	length of complete coding (CDS or mat_peptide) feature is not a multiple of 3 [21287]
M300	NC_045512	mat_peptide	leader protein	PEPTIDE_TRANSLATION_PROBLEM	-	-	mat_peptide may not be translated because its parent CDS has a problem [-]
M300	NC_045512	mat_peptide	nsp2	PEPTIDE_TRANSLATION_PROBLEM	-	-	mat_peptide may not be translated because its parent CDS has a problem [-]
M300	NC_045512	mat_peptide	nsp3	PEPTIDE_TRANSLATION_PROBLEM	-	-	mat_peptide may not be translated because its parent CDS has a problem [-]
M300	NC_045512	mat_peptide	nsp4	PEPTIDE_TRANSLATION_PROBLEM	-	-	mat_peptide may not be translated because its parent CDS has a problem [-]
M300	NC_045512	mat_peptide	3C-like proteinase	PEPTIDE_TRANSLATION_PROBLEM	-	-	mat_peptide may not be translated because its parent CDS has a problem [-]
M300	NC_045512	mat_peptide	nsp6	PEPTIDE_TRANSLATION_PROBLEM	-	-	mat_peptide may not be translated because its parent CDS has a problem [-]
M300	NC_045512	mat_peptide	nsp7	PEPTIDE_TRANSLATION_PROBLEM	-	-	mat_peptide may not be translated because its parent CDS has a problem [-]
M300	NC_045512	mat_peptide	nsp8	PEPTIDE_TRANSLATION_PROBLEM	-	-	mat_peptide may not be translated because its parent CDS has a problem [-]
M300	NC_045512	mat_peptide	nsp9	PEPTIDE_TRANSLATION_PROBLEM	-	-	mat_peptide may not be translated because its parent CDS has a problem [-]
M300	NC_045512	mat_peptide	nsp10	PEPTIDE_TRANSLATION_PROBLEM	-	-	mat_peptide may not be translated because its parent CDS has a problem [-]
M300	NC_045512	mat_peptide	RNA-dependent RNA polymerase	PEPTIDE_TRANSLATION_PROBLEM	-	-	mat_peptide may not be translated because its parent CDS has a problem [-]
M300	NC_045512	mat_peptide	helicase	PEPTIDE_TRANSLATION_PROBLEM	-	-	mat_peptide may not be translated because its parent CDS has a problem [-]
M300	NC_045512	mat_peptide	helicase	UNEXPECTED_LENGTH	16111..17909:+	16237..18039:+	length of complete coding (CDS or mat_peptide) feature is not a multiple of 3 [1799]
M300	NC_045512	mat_peptide	3'-to-5' exonuclease	PEPTIDE_TRANSLATION_PROBLEM	-	-	mat_peptide may not be translated because its parent CDS has a problem [-]
M300	NC_045512	mat_peptide	endoRNase	PEPTIDE_TRANSLATION_PROBLEM	-	-	mat_peptide may not be translated because its parent CDS has a problem [-]
M300	NC_045512	mat_peptide	2'-O-ribose methyltransferase	PEPTIDE_TRANSLATION_PROBLEM	-	-	mat_peptide may not be translated because its parent CDS has a problem [-]



Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Daytona1107/output/assemblies/		
Name	Size	Changed
..		11/12/2024 9:16:17 AM
J84.consensus.fa	30 KB	11/12/2024 9:14:32 AM
J682.consensus.fa	30 KB	11/12/2024 9:11:47 AM
J711.consensus.fa	30 KB	11/12/2024 9:12:25 AM
J886.consensus.fa	30 KB	11/12/2024 9:12:25 AM
M109.consensus.fa	30 KB	11/12/2024 9:15:48 AM
M300.consensus.fa	30 KB	11/12/2024 9:12:25 AM
M301.consensus.fa	30 KB	11/12/2024 9:13:02 AM
M330.consensus.fa	30 KB	11/12/2024 9:13:02 AM
M653.consensus.fa	30 KB	11/12/2024 9:14:47 AM
M908.consensus.fa	30 KB	11/12/2024 9:13:39 AM
T45.consensus.fa	30 KB	11/12/2024 9:12:55 AM

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Daytona1107/output/		
Name	Size	Changed
assemblies		11/12/2024 9:15:48 AM
J84		11/12/2024 9:14:55 AM
J682		11/12/2024 9:12:25 AM
J711		11/12/2024 9:13:03 AM
J886		11/12/2024 9:13:01 AM
M109		11/12/2024 9:16:11 AM
M300		11/12/2024 9:13:06 AM
M301		11/12/2024 9:13:37 AM
M330		11/12/2024 9:13:38 AM
M653		11/12/2024 9:15:14 AM
M908		11/12/2024 9:14:05 AM
T45		11/12/2024 9:13:28 AM
vadr_error_reports		11/12/2024 9:14:56 AM
variants		11/12/2024 9:15:48 AM
assemblies.fasta	321 KB	11/12/2024 9:16:11 AM
nextclade_report.csv	5 KB	11/12/2024 9:16:17 AM
sum_report.txt	2 KB	11/12/2024 9:16:11 AM

>J84

```
NNNNNNNTTATACCTTCCCAGGTAACAAACCAACCAACTTTTCGATCTCTTGTAGATCTGTTCTCTAAACGAACTTTAAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCACGCAGTATAATTAATAACTAA
TTACTGTCGTTGACAGGACACGAGTAACCTCGTCTATCTTCTGCAGGCTGCTTACGGTTTCGTCGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGG
TTTCAACGAGAAAAACACACGTCCAACCTCAGTTTGCCTGTTTTACAGGTTCCGCGACGTGCTCGTACGTGGCTTTGGAGACTCCGTGGAGGAGGTCTTATCAGAGGCACGTCAACATCTTAAAGATGGCACTTGTGGCTTAGTA
GAAGTTGAAAAAGGCGTTTTGCCTCAACTTGAACAGCCCTATGTGTTCAACACGTTCCGGATGCTCGAAGTGCACCTCATGGTCATGTTGAGCTGGTAGCAGAAGTCAAGGCATTAGTACGGTCGTAGTGGTGAGACAC
TTGGTGTCTTGTCCCTCATGTGGGCGAAATACCAAGTGGCTTACCAGCAAGGTTCTTCTTCTGTAAGAACGGAATAAAGGAGCTGGTGGCCATAGTTACGGCGCCGATCTAAAGTCATTTGACTTAGGCGACGAGCTTGGCAC
TGATCCTTATGAAGATTTTCAAGAAAAGTGAACACTAAACATAGCAGTGGTGTACCCGTGAAGTCAAGGAGGCGATACACTCGCTATGTGCATAACAACTTCTGTGGCCCTGATGGCTACCCCTCTT
GAGTGCATTAAGACCTTCTAGCACGTGCTGGTAAAGCTTCATGCACTTTGTCCGAACAAGTGGACTTTATTGACACTAAGAGGGGTGTATACTGCTGCCGTGAACATGAGCATGAAATTGCTTGGTACACGGAACGTTCTG
AAAAGAGCTATGAATTGCAGACACCTTTGAAATTAATTTGGCAAAGAAATTTGACACCTTCAATGGGGAATGTCCAAATTTGTATTTCCCTTAAATTTCCATAATCAAGACTATTCAACCAAGGGTTGAAAAGAAAAAGCT
TGATGGCTTTATGGGTAGAATTCGATCTGTCTATCCAGTTGCGTCACCAAATGAATGCAACCAAATGTGCCTTTCAACTCTCATGAAGTGTGATCATTGTGGTGAACTTCATGGCAGACGGGCGATTTTGTAAAGCCACT
TGCGAATTTTGTGGCACTGAGAATTTGACTAAGAAAGGTGCCACTACTTGTGGTTACTTACCCAAAATGCTGTTGTTAAATTTATGTCCAGCATGTCACAATTGAGAAGTAGGACCTGAGCATAAGTCTTGCCGAATACC
ATAATGAATCTGGCTTGAAGAACCTTCTCGTAAGGGTGGTGCACATTTGCCTTTGGAGGCTGTGTGTTCTTATGTTGGTTGCCATAACAAGTGTGCCTATTGGGTTCCACGTGCTAGCGCTAACATAGGTTGTAACCA
TACAGGTGTTGTTGGAGAAGGTTCCGAAGGTCTTAATGACAACCTTCTTGAAATACTCCAAAAGAGAAAGTCAACATCAATATTGTTGGTGACTTTAACTTAATGAAGAGATCGCCATTATTTTGGCATCTTTTTCTGCT
TCCACAAGTGCTTTTGTGGAAGCTGTGAAAGGTTTGGATTATAAGCATTCAACAAATTTGTAATCCTGTGGTAATTTTAAAGTTACAAAAGGAAAAGCTAAAAAGGTCCTGGAATATTGGTGAACAGAAATCAATAC
TGAGTCCTCTTATGCATTTGCATCAGAGGCTGCTCGTGTGTACGATCAATTTTCTCCGCACTCTTGAACTGCTCAAAATTTCTGTGCGTGTTTTACAGAAGGCCGCTATAACAATACTAGATGGAATTTACAGTATTC
```



Conclusion



Fundamentals of
Daytona



Installation and setup
of Daytona in HPG



Successfully executed
Job query for
Daytona



Generated Output
Files





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