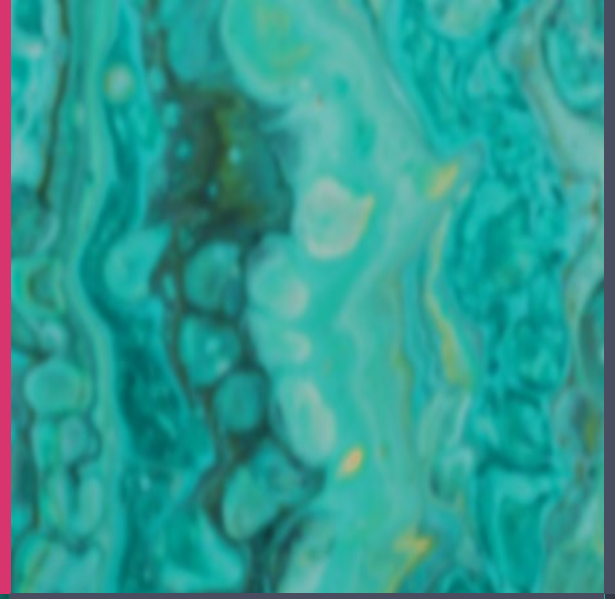


13 Oct 2025

Aquascope

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.

Overview

Purpose:

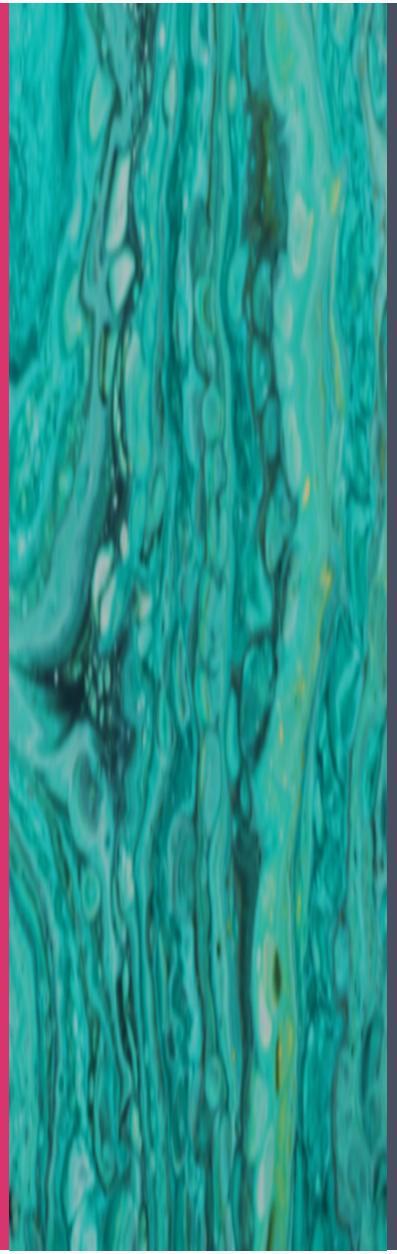
Pipeline is developed for *SARS-CoV-2* wastewater sequencing analysis, enabling standardized qc, read trimming, alignment, primer trimming and variant deconvolution to estimate lineage abundance in wastewater samples

Usage:

Can be used by public health labs and researchers for surveillance of *SARS-CoV-2* variants through wastewater for molecular typing, outbreak investigation, surveillance, and comparative genomic studies

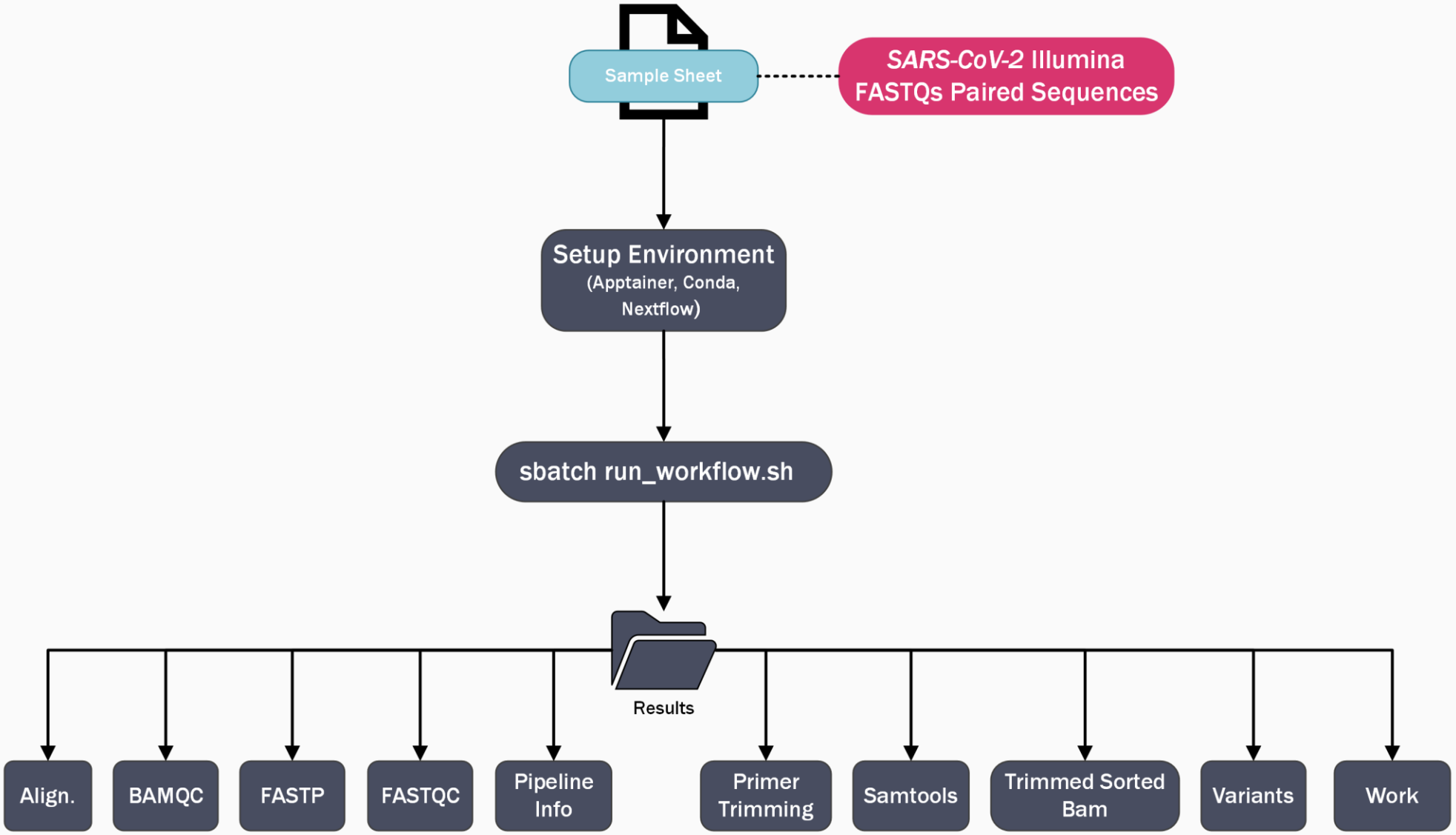
Dependencies:

Nextflow
Apptainer
Conda
Primer Scheme (ARTIC V3) – Required for Illumina Runs



Advanced Molecular Detection
Southeast Region Bioinformatics

Workflow



Advanced Molecular Detection
Southeast Region Bioinformatics

Application

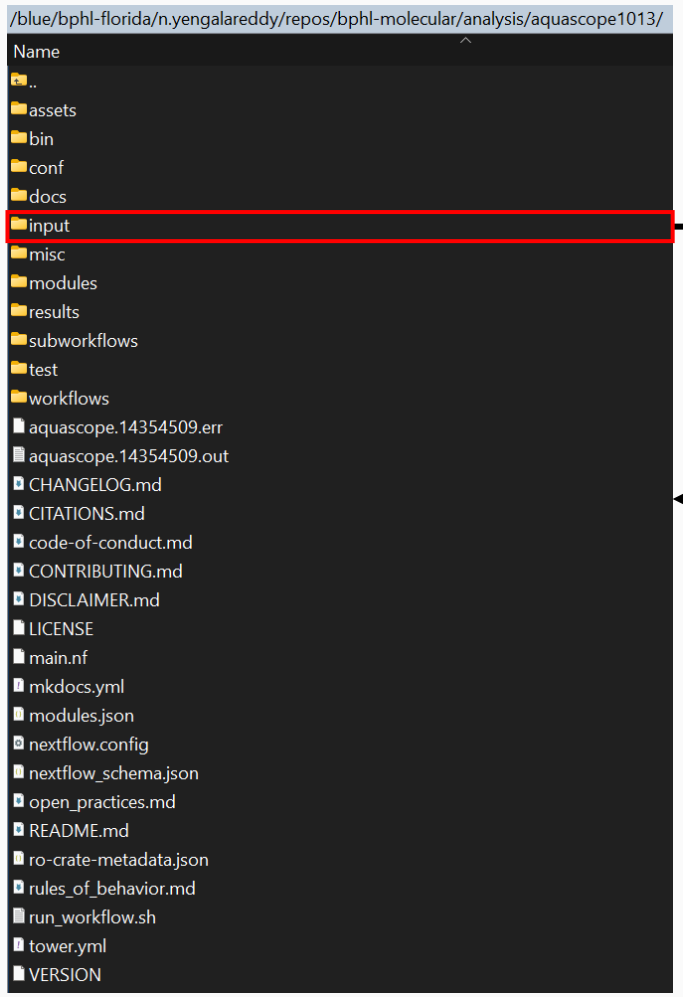
Objective

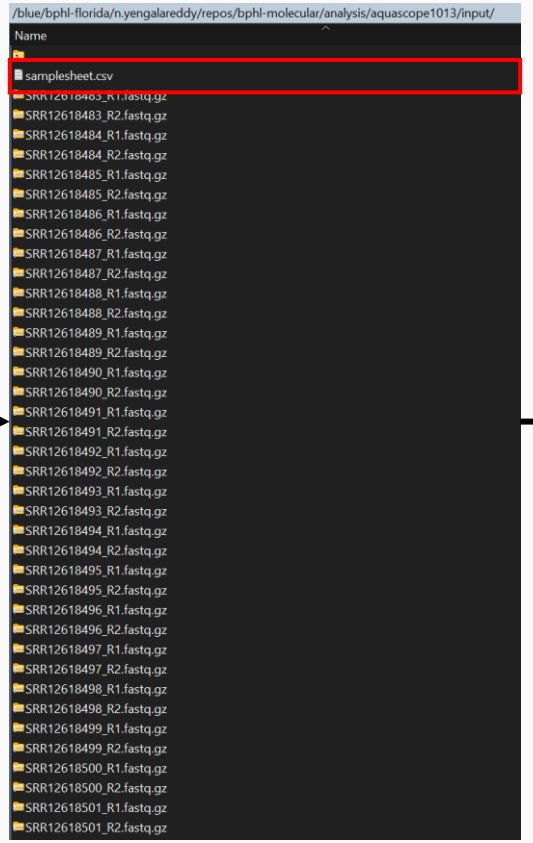
Use SARS-CoV-2 Illumina FASTQs paired sequences (acquired from wastewater) from NCBI database, and use Aquascope to analyze the samples

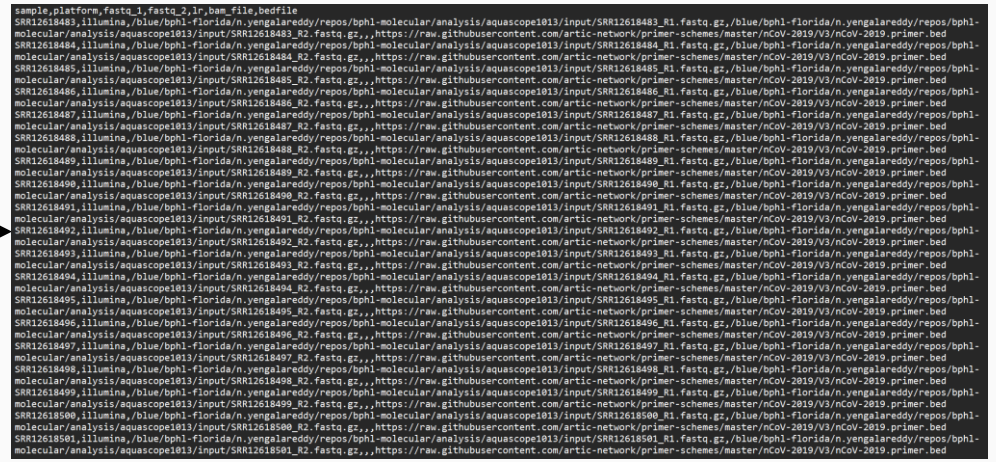
Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

```
cd blue/bphl-<state>/<user>/repos/bphl-molecular/
git clone https://github.com/CDCgov/aquascope
mkdir analysis/
cd analysis/
cp /blue/bphl-<state>/<user>/repos/bphl-molecular/aquascope/
```



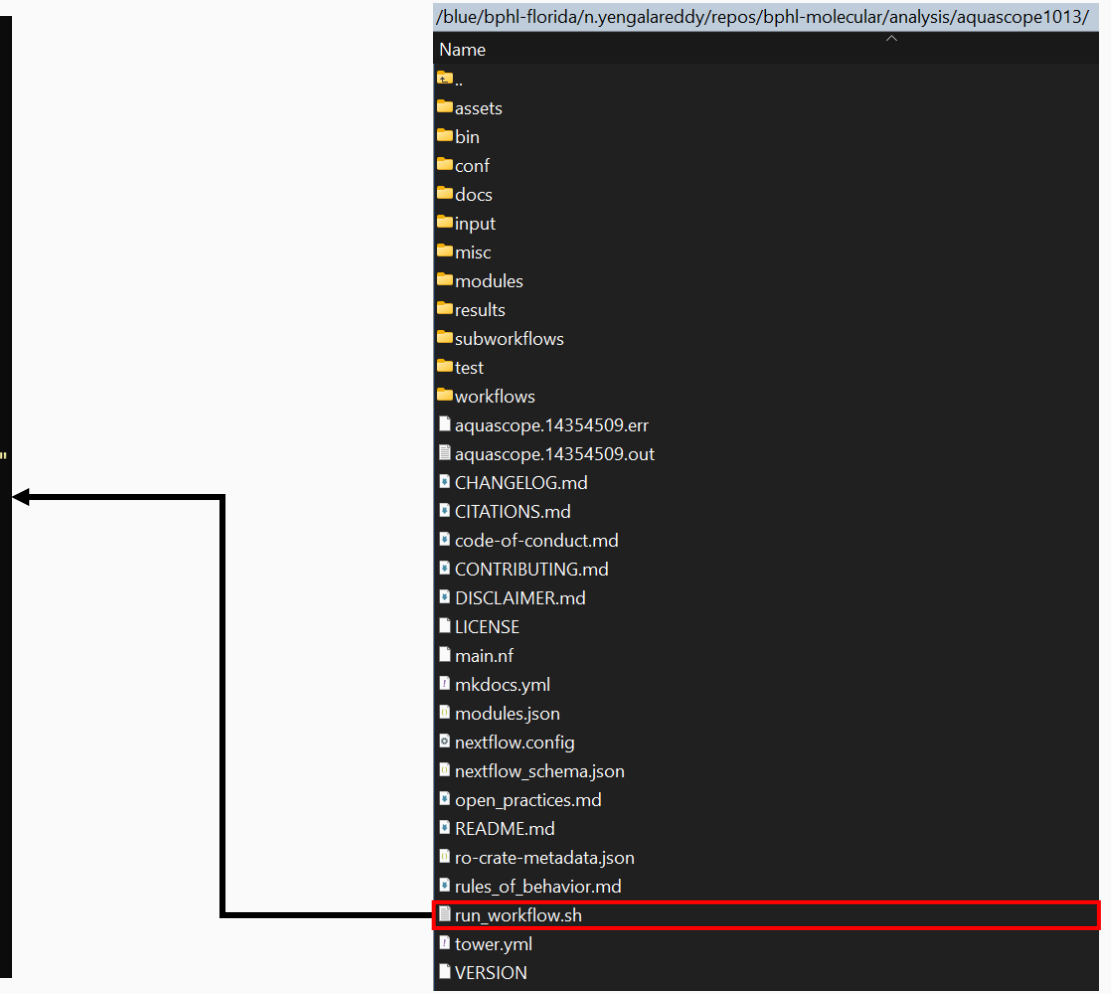




Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

```
#!/bin/bash
#SBATCH --account=bphl-umbrella
#SBATCH --qos=bphl-umbrella
#SBATCH --job-name=aquascope
#SBATCH --cpus-per-task=8
#SBATCH --mem=64gb
#SBATCH --time=48:00:00
#SBATCH --output=aquascope.%j.out
#SBATCH --error=aquascope.%j.err
#SBATCH --mail-user=nikhil.yengala@flhealth.gov
#SBATCH --mail-type=FAIL,END

module purge
module load apptainer
module load conda

# Activate conda env with Nextflow
conda activate Nextflow

BASE="/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/aquascope1013"
samplesheet="${BASE}/input/samplesheet.csv"
outdir="${BASE}/results"
workdir="${outdir}/work"

if [[ ! -f "$samplesheet" ]]; then
  echo "ERROR: samplesheet not found: $samplesheet"
  exit 1
fi

mkdir -p "$outdir" "$workdir"

export APPTAINER_CACHEDIR="${BASE}/.apptainer_cache"
export NXF_APPTAINER_CACHEDIR="${BASE}/.apptainer_cache"

echo "Launching Aquascope..."
nextflow run main.nf \
  -entry AQUASCOPE \
  -profile singularity \
  --input "$samplesheet" \
  --outdir "$outdir" \
  -work-dir "$workdir" \
  -resume
```



Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

sbatch run_workflow.sh

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/aquascope1013/

Name
- ..
- assets
- bin
- conf
- docs
- input
- misc
- modules
- results
- subworkflows
- test
- workflows
- aquascope.14354509.err
- aquascope.14354509.out
- CHANGELOG.md
- CITATIONS.md
- code-of-conduct.md
- CONTRIBUTING.md
- DISCLAIMER.md
- LICENSE
- main.nf
- mkdocs.yml
- modules.json
- nextflow.config
- nextflow_schema.json
- open_practices.md
- README.md
- ro-crate-metadata.json
- rules_of_behavior.md
- run_workflow.sh
- tower.yml
- VERSION

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/aquascope1013/results/

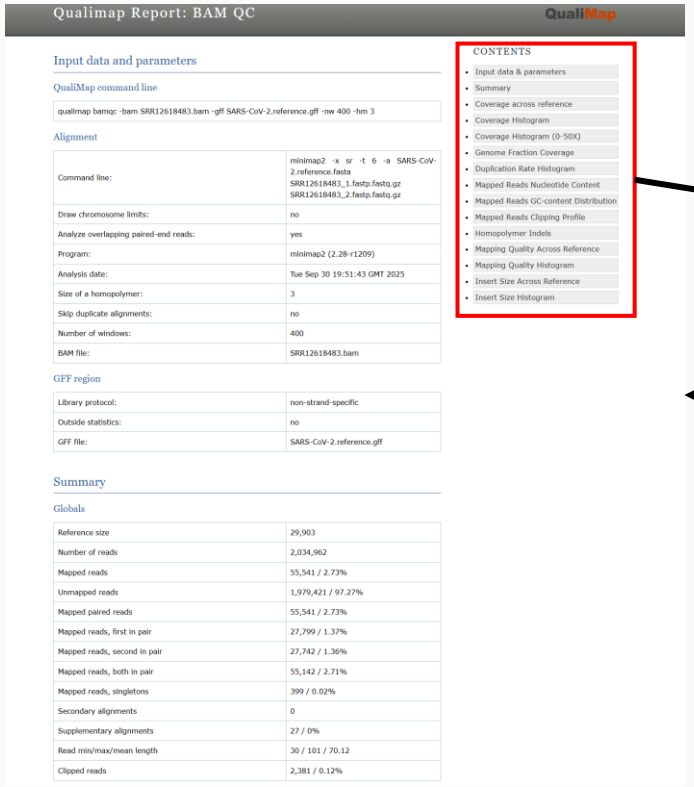
Name
- ..
- ALIGNMENT
- BAMQC
- FASTP
- FASTQC
- MULTIQC
- pipeline_info
- PRIMER_TRIMMING
- SAMTOOLS
- TRIMMED_SORTED_BAM
- VARIANTS
- work

/.../n.yengalareddy/repos/bphl-molecular/analysis/aquascope1013/results/ALIGNMENT/

Name
- ..
- SRR12618483.bam
- SRR12618484.bam
- SRR12618485.bam
- SRR12618486.bam
- SRR12618487.bam
- SRR12618488.bam
- SRR12618489.bam
- SRR12618490.bam
- SRR12618491.bam
- SRR12618492.bam
- SRR12618493.bam
- SRR12618494.bam
- SRR12618495.bam
- SRR12618496.bam
- SRR12618497.bam
- SRR12618498.bam
- SRR12618499.bam
- SRR12618500.bam
- SRR12618501.bam

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.


Qualimap Report: BAM QC

QualiMap

Input data and parameters

QualiMap command line

qualimap bamqc -bam SRR12618483.bam -gff SARS-CoV-2.reference.gff -nw 400 -hm 3

Alignment

Command line:	minimap2 -x sr -t 6 -a SARS-CoV-2.reference.fasta SRR12618483_1.fastp.fastq.gz SRR12618483_2.fastp.fastq.gz
Draw chromosome limits:	no
Analyze overlapping paired-end reads:	yes
Program:	minimap2 (2.28-r1209)
Analysis date:	Tue Sep 30 19:51:43 GMT 2025
Size of a homopolymer:	3
Skip duplicate alignments:	no
Number of windows:	400
BAM file:	SRR12618483.bam

GFF region

Library protocol:	non-strand-specific
Outside statistics:	no
GFF file:	SARS-CoV-2.reference.gff

Summary

Globals

Reference size	29,903
Number of reads	2,034,962
Mapped reads	55,541 / 2.73%
Unmapped reads	1,979,421 / 97.27%
Mapped paired reads	55,541 / 2.73%
Mapped reads, first in pair	27,799 / 1.37%
Mapped reads, second in pair	27,742 / 1.36%
Mapped reads, both in pair	55,142 / 2.71%
Mapped reads, singletons	399 / 0.02%
Secondary alignments	0
Supplementary alignments	27 / 0%
Read min/max/mean length	30 / 101 / 70.12
Clipped reads	2,381 / 0.12%

CONTENTS

- Input data & parameters
- Summary
- Coverage across reference
- Coverage Histogram
- Coverage Histogram (0-50X)
- Genome Fraction Coverage
- Duplication Rate Histogram
- Mapped Reads Nucleotide Content
- Mapped Reads GC-content Distribution
- Mapped Reads Clipping Profile
- Homopolymer Indels
- Mapping Quality Across Reference
- Mapping Quality Histogram
- Insert Size Across Reference
- Insert Size Histogram

/.../bphl-molecular/analysis/aquascope1013/results/BAMQC/Qualimap/SRR12618483/

Name

..

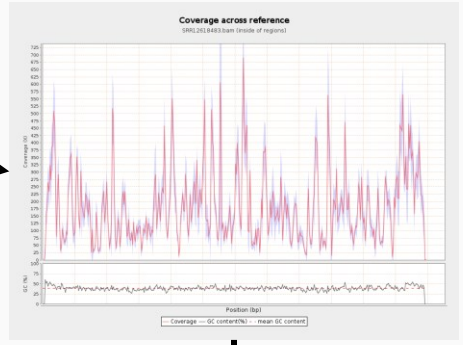
css

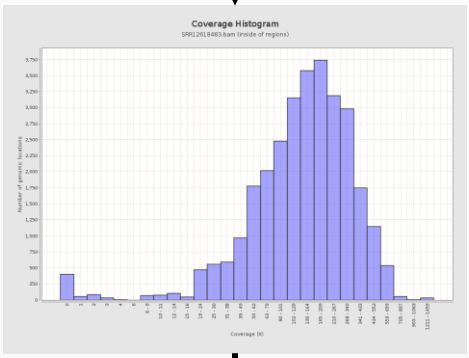
images_qualimapReport

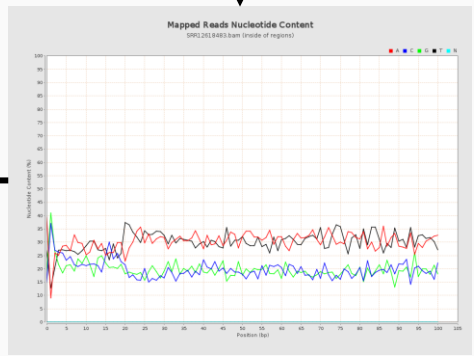
raw_data_qualimapReport

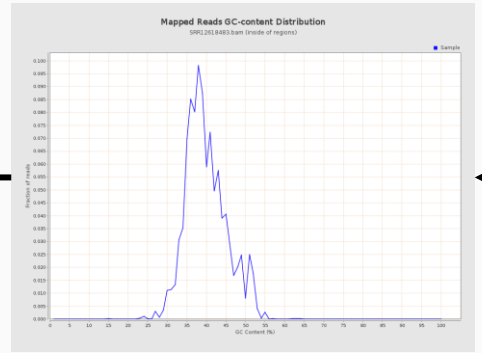
genome_results.txt

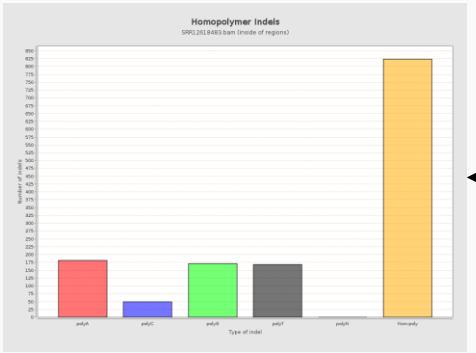
qualimapReport.html











/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/aquascope1013/results/

Name

..

ALIGNMENT

BAMQC

FASTP

FASTQC

MULTIQC

pipeline_info

PRIMER_TRIMMING

SAMTOOLS

TRIMMED_SORTED_BAM

VARIANTS

work

/.../n.yengalareddy/repos/bphl-molecular/analysis/aquascope1013/results/BAMQC/

Name

..

Qualimap

/.../repos/bphl-molecular/analysis/aquascope1013/results/BAMQC/Qualimap/

Name

..

SRR12618483

SRR12618484

SRR12618485

SRR12618486

SRR12618487

SRR12618488

SRR12618489

SRR12618490

SRR12618491

SRR12618492

SRR12618493

SRR12618494

SRR12618495

SRR12618496

SRR12618497

SRR12618498

SRR12618499

SRR12618500

SRR12618501

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

fastp report

Summary

General

fastp version:	0.23.4 (https://github.com/OpenGene/fastp)
sequencing:	paired end (101 cycles + 101 cycles)
mean length before filtering:	101bp, 101bp
mean length after filtering:	70bp, 70bp
duplication rate:	68.342228%
Insert size peak:	32
Detected read1 adapter:	AGATCGGAAGAGCACACGTCTGAACTCCAGTCA
Detected read2 adapter:	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

Before filtering

total reads:	2.443002 M
total bases:	246.743202 M
Q20 bases:	235.073983 M (95.270703%)
Q30 bases:	226.720291 M (91.885122%)
GC content:	52.929515%

After filtering

Filtering result

reads passed filters:	2.034962 M (83.297599%)
reads with low quality:	1.248000 K (0.051085%)
reads with too many N:	274 (0.011216%)
reads too short:	406.518000 K (16.640101%)

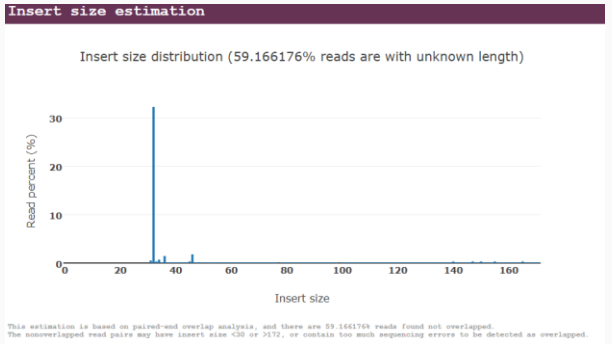
Adapters

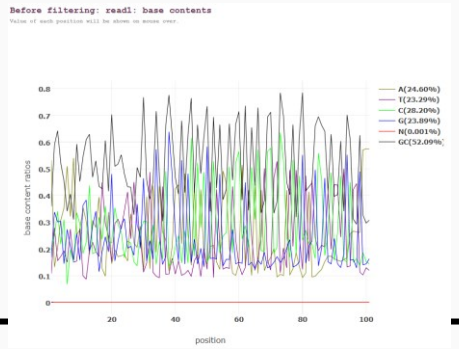
Adapter or bad ligation of read1

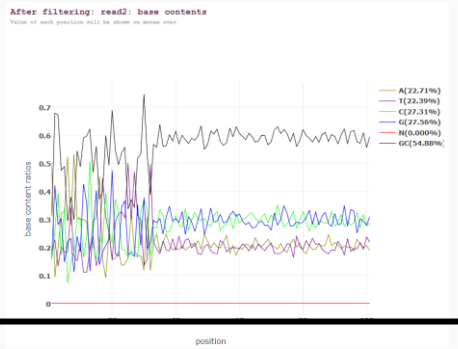
Sequence	Occurrences
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACATCGTTCCATGTGTATGCCG	19700
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACATCGTTCCATCTGTATGCCGTCTTCTGCTT	16245
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACATCGTTCCATCTGTATGCCGTCTTCTGCTTG	7657
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACATCGTTCCATCTGTATGCCGTCTTCTGCTTGAAA	360931
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACATCGTTCCATCTGTATGCCGTCTTCTGCTTGAAAAAAA	100945
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACATCGTTCCATCTGTATGCCGTCTTCTGCTTGAAAAAAAA	11100
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACATCGTTCCATCTGTATGCCGTCTTCTGCTTGAAAAAAAAAAAAAAAA	19919
other adapter sequences	128640

Adapter or bad ligation of read2

Sequence	Occurrences
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTTCAGAGCCGTGTAGATCTCGGTGGTCGCCGGATCAT	21806
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTTCAGAGCCGTGTAGATCTCGGTGGTCGCCGTATCAA	10403
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTTCAGAGCCGTGTAGATCTCGGTGGTCGCCGTATCAT	121425
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTTCAGAGCCGTGTAGATCTCGGTGGTCGCCGTGTCAT	9384
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTTCAGAGCCGTGTAGATCTCGGTGGTCGCCGGTCATAAA	8872
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTTCAGAGCCGTGTAGATCTCGGTGGTCGCCGTATCATTAAA	14454
other adapter sequences	479150







/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/aquascope1013/results/

Name
..
ALIGNMENT
BAMQC
FASTP
FASTQC
MULTIQC
pipeline_info
PRIMER_TRIMMING
SAMTOOLS
TRIMMED_SORTED_BAM
VARIANTS
work

/.../n.yengalareddy/repos/bphl-molecular/analysis/aquascope1013/results/FASTP/

Name
..
Short_Reads

/.../repos/bphl-molecular/analysis/aquascope1013/results/FASTP/Short_Reads/

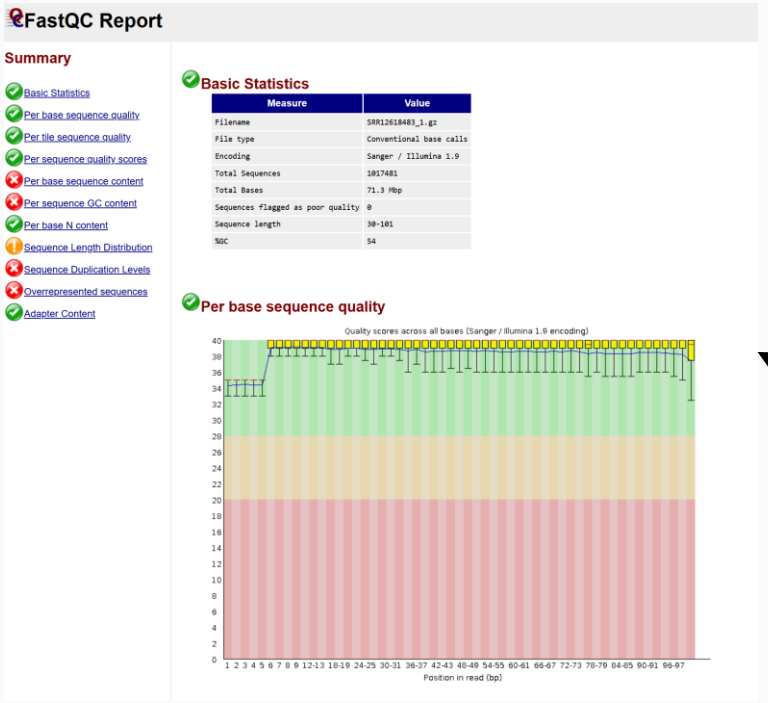
Name
..
SRR12618483
SRR12618484
SRR12618485
SRR12618486
SRR12618487
SRR12618488

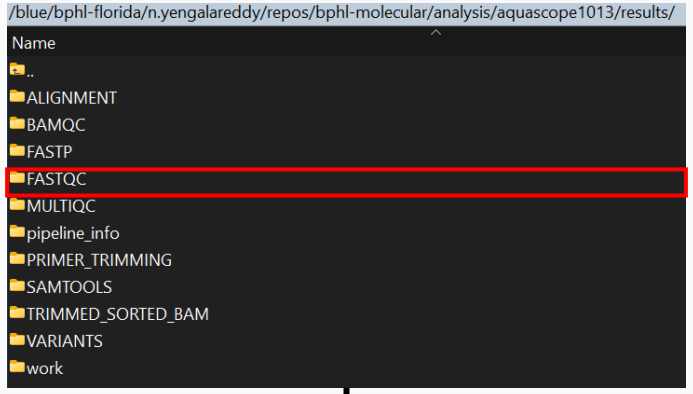
/.../bphl-molecular/analysis/aquascope1013/results/FASTP/Short_Reads/SRR12618483/

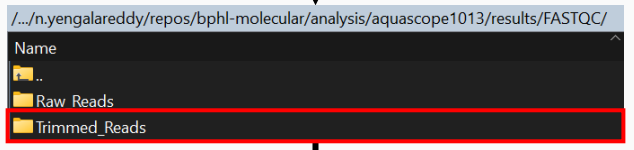
Name
..
SRR12618483.fastp.html
SRR12618483.fastp.json
SRR12618483_1.fastp.fastq.gz
SRR12618483_2.fastp.fastq.gz

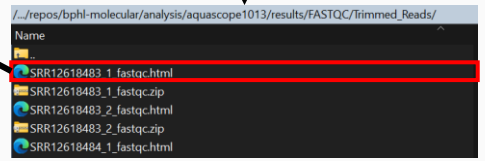
Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

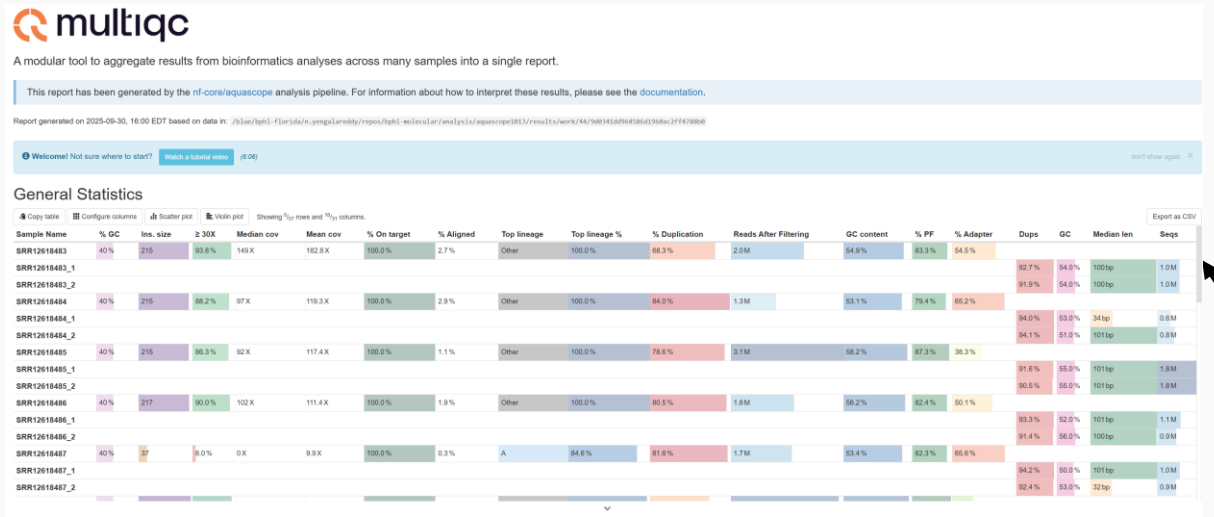


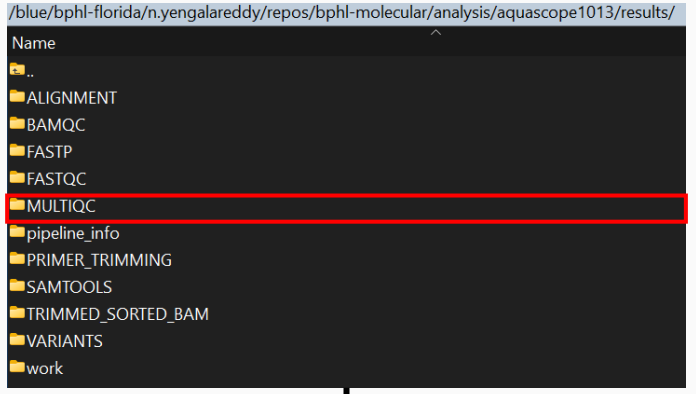


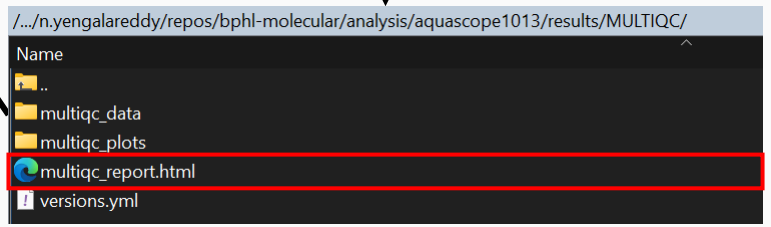


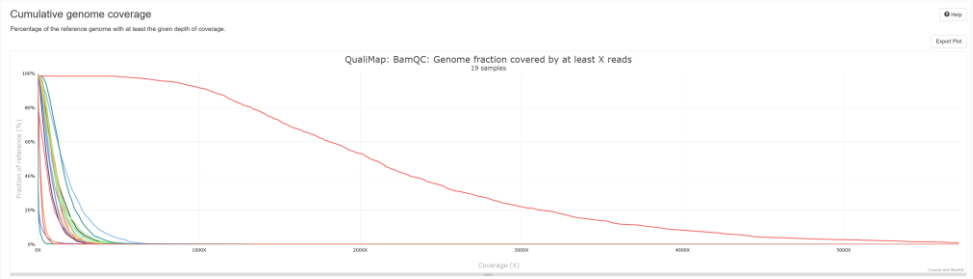


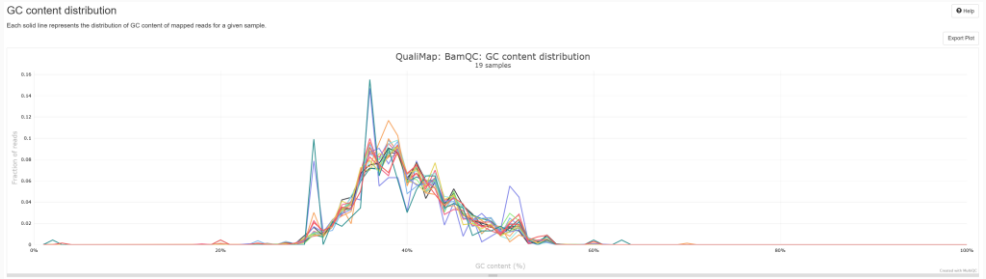
Application Cont.

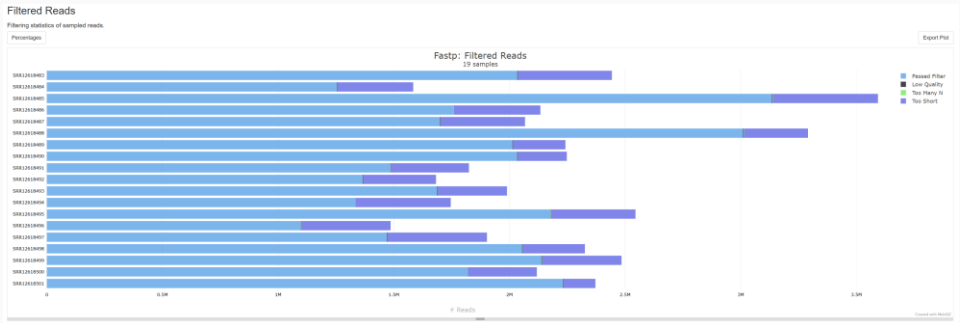












Advanced Molecular Detection
Southeast Region Bioinformatics

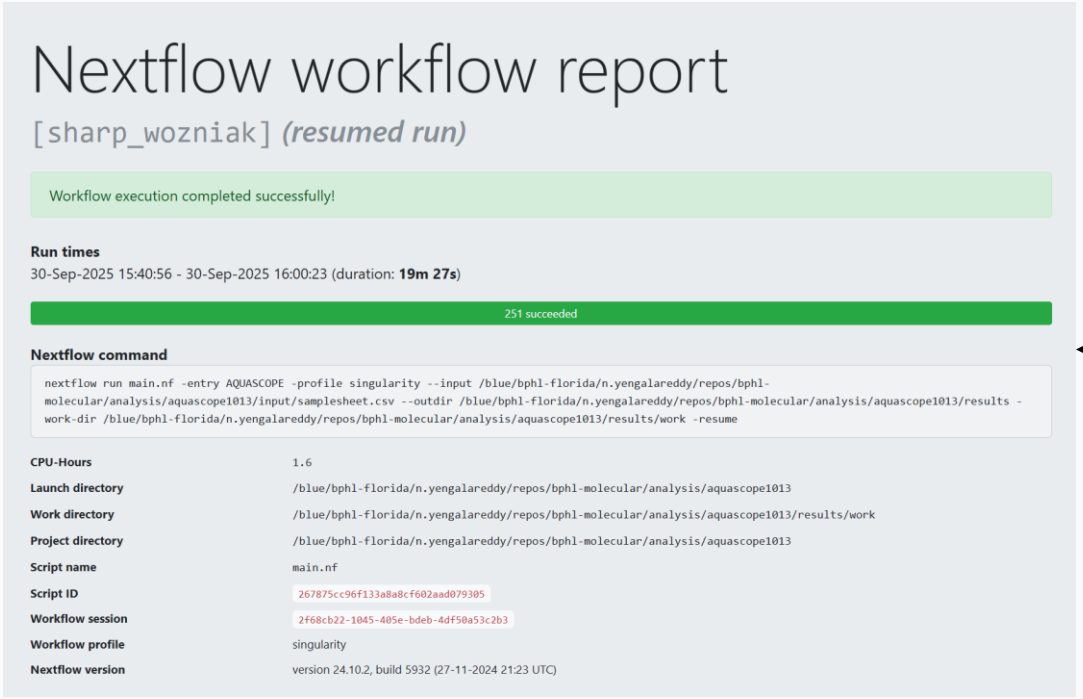
Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/aquascope1013/results/

Name
..
ALIGNMENT
BAMQC
FASTP
FASTQC
MULTIQC
pipeline_info
PRIMER_TRIMMING
SAMTOOLS
TRIMMED_SORTED_BAM
VARIANTS
work

/.../n.yengalareddy/repos/bphl-molecular/analysis/aquascope1013/results/pipeline_info/

Name
..
execution_report_2025-09-30_15-40-55.html
execution_timeline_2025-09-30_15-40-55.html
execution_trace_2025-09-30_15-40-55.txt
params_2025-09-30_15-40-59.json
pipeline_dag_2025-09-30_15-40-55.html
samplesheet.valid.csv
versions.yml



Nextflow workflow report

[sharp_wozniak] *(resumed run)*

Workflow execution completed successfully!

Run times

30-Sep-2025 15:40:56 - 30-Sep-2025 16:00:23 (duration: **19m 27s**)

251 succeeded

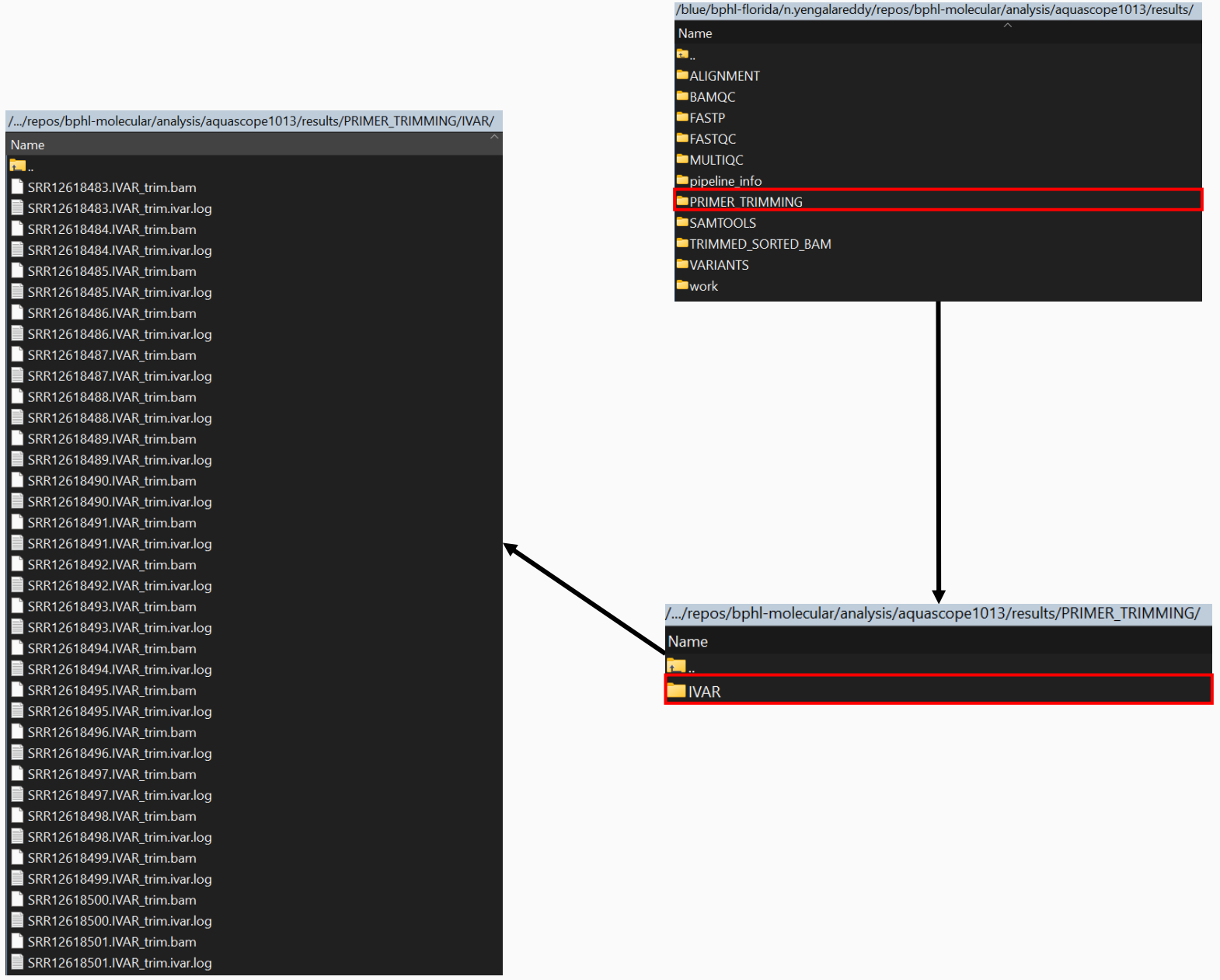
Nextflow command

```
nextflow run main.nf -entry AQUASCOPE -profile singularity --input /blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/aquascope1013/input/samplesheet.csv --outdir /blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/aquascope1013/results -work-dir /blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/aquascope1013/results/work -resume
```

CPU-Hours	1.6
Launch directory	/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/aquascope1013
Work directory	/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/aquascope1013/results/work
Project directory	/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/aquascope1013
Script name	main.nf
Script ID	267875cc96f133a8a8cf602aad079305
Workflow session	2f68cb22-1045-405e-bdeb-4df50a53c2b3
Workflow profile	singularity
Nextflow version	version 24.10.2, build 5932 (27-11-2024 21:23 UTC)

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.



Application Cont.

```
54907 + 0 in total (QC-passed reads + QC-failed reads)
54880 + 0 primary
0 + 0 secondary
27 + 0 supplementary
0 + 0 duplicates
0 + 0 primary duplicates
54907 + 0 mapped (100.00% : N/A)
54880 + 0 primary mapped (100.00% : N/A)
54880 + 0 paired in sequencing
27511 + 0 read1
27369 + 0 read2
54697 + 0 properly paired (99.67% : N/A)
54745 + 0 with itself and mate mapped
135 + 0 singletons (0.25% : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)
```

```
# This file was produced by samtools stats (1.21+htslib-1.21) and can be plotted using plot-bamstats
# This file contains statistics for all reads.
# The command line was:  stats --threads 1 --reference SARS-CoV-2.reference.fasta SRR12618483.trimmed.sorted.bam
# CHK, Checksum [2]Read Names   [3]Sequences    [4]Qualities
# CHK, CRC32 of reads which passed filtering followed by addition (32bit overflow)
CHK     1102a2d3        171eb6fb        228185ca
# Summary Numbers. Use `grep ^SN | cut -f 2-` to extract this part.
SN      raw total sequences:    54880   # excluding supplementary and secondary reads
SN      filtered sequences:     0
SN      sequences:      54880
SN      is sorted:      1
SN      1st fragments:  27511
SN      last fragments: 27369
SN      reads mapped:   54880
SN      reads mapped and paired:        54745   # paired-end technology bit set + both mates mapped
SN      reads unmapped: 0
SN      reads properly paired:  54697   # proper-pair bit set
SN      reads paired:   54880   # paired-end technology bit set
SN      reads duplicated:       0       # PCR or optical duplicate bit set
SN      reads MQ0:      1       # mapped and MQ=0
SN      reads QC failed:        0
SN      non-primary alignments: 0
SN      supplementary alignments:       27
SN      total length:   5462200 # ignores clipping
SN      total first fragment length:    2738133 # ignores clipping
SN      total last fragment length:     2724067 # ignores clipping
SN      bases mapped:   5462200 # ignores clipping
SN      bases mapped (cigar):   5312914 # more accurate
SN      bases trimmed:  0
SN      bases duplicated:       0
SN      mismatches:     9836    # from NM fields
SN      error rate:     1.851338e-03    # mismatches / bases mapped (cigar)
SN      average length: 100
SN      average first fragment length:  100
SN      average last fragment length:   100
SN      maximum length: 101
SN      maximum first fragment length:  101
SN      maximum last fragment length:   101
SN      average quality:        39.0
SN      insert size average:    202.0
SN      insert size standard deviation: 69.2
SN      inward oriented pairs:  26891
SN      outward oriented pairs: 432
SN      pairs with other orientation:   5
SN      pairs on different chromosomes: 0
SN      percentage of properly paired reads (%):        99.7
```

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/aquascope1013/results/

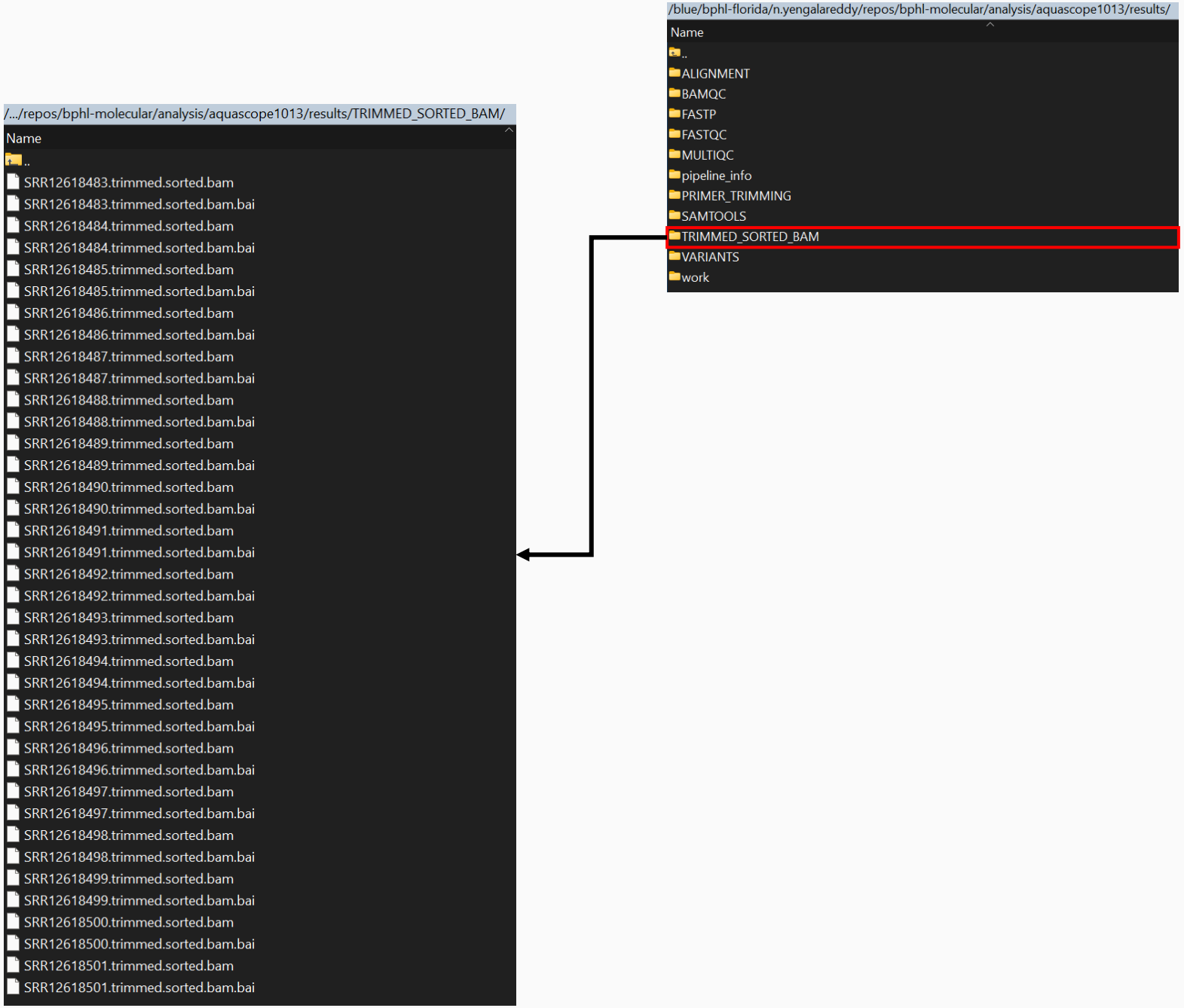
Name
- ..
- ALIGNMENT
- BAMQC
- FASTP
- FASTQC
- MULTIQC
- pipeline_info
- PRIMER_TRIMMING
- SAMTOOLS
- TRIMMED_SORTED_BAM
- VARIANTS
- work

/.../n.yengalareddy/repos/bphl-molecular/analysis/aquascope1013/results/SAMTOOLS/

Name
- ..
- SRR12618483.flagstat
- SRR12618483.idxstats
- SRR12618483.stats
- SRR12618484.flagstat
- SRR12618484.idxstats
- SRR12618484.stats

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.



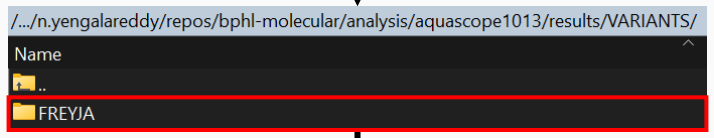
Advanced Molecular Detection
Southeast Region Bioinformatics

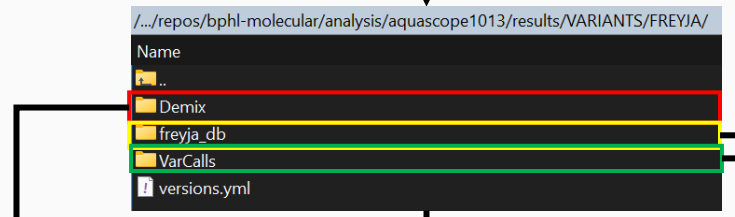
Application Cont.

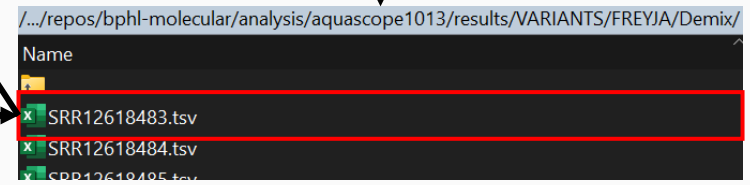
	SRR12618483.variants.tsv
summarized	[('Other', 1.0000000012495296)]
lineages	B.6 B.45 B.1.301 B.1.369 B.1.1.228 B.1.578
abundances	0.28440400 0.20719400 0.20563400 0.13082300 0.12865500 0.04329000
resid	4.423475409
coverage	97.40828679

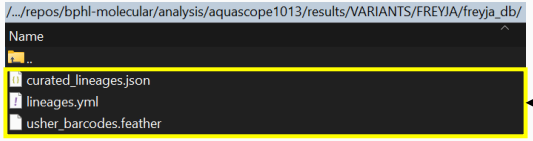
REGION	POS	REF	ALT	REF_DP	REF_RV	REF_QUAL	ALT_DP	ALT_RV	ALT_QUAL	ALT_FREQ	TOTAL_DP	PVAL	PASS	GFF_FEATURE	REF_CODON	REF_AA	ALT_CODON	ALT_AA	POS_AA
MN908947.3	213	G	A	53	0	39	1	0	40	0.0185185	54	0.504673	FALSE	NA	NA	NA	NA	NA	NA
MN908947.3	241	C	T	80	0	39	54	24	69	0.402985	134	8.20E-20	TRUE	NA	NA	NA	NA	NA	NA
MN908947.3	329	C	T	79	6	43	1	1	78	0.0125	80	0.503145	FALSE	NA	NA	NA	NA	NA	NA
MN908947.3	437	A	C	165	123	39	9	0	39	0.0517241	174	0.00180324	TRUE	NA	NA	NA	NA	NA	NA
MN908947.3	442	C	A	135	85	39	38	38	39	0.219653	173	4.16E-13	TRUE	NA	NA	NA	NA	NA	NA
MN908947.3	662	G	A	181	56	39	2	1	34	0.010929	183	0.250685	FALSE	NA	NA	NA	NA	NA	NA
MN908947.3	746	T	#NAME?	349	166	47	19	0	20	0.0544413	349	0.000568383	TRUE	NA	NA	NA	NA	NA	NA
MN908947.3	748	G	A	329	147	48	21	20	37	0.06	350	3.61E-07	TRUE	NA	NA	NA	NA	NA	NA
MN908947.3	752	A	-C	346	166	47	52	0	20	0.149856	347	4.90E-12	TRUE	NA	NA	NA	NA	NA	NA
MN908947.3	906	C	T	532	420	47	17	9	39	0.0309654	549	6.83E-06	TRUE	NA	NA	NA	NA	NA	NA
MN908947.3	942	G	A	296	180	52	24	22	45	0.075	320	3.96E-08	TRUE	NA	NA	NA	NA	NA	NA
MN908947.3	1018	G	C	114	8	42	9	0	39	0.0731707	123	0.0017421	TRUE	NA	NA	NA	NA	NA	NA
MN908947.3	1059	C	T	33	0	39	14	14	38	0.297872	47	2.30E-05	TRUE	NA	NA	NA	NA	NA	NA
MN908947.3	1212	G	A	254	137	44	31	31	39	0.108772	285	2.08E-10	TRUE	NA	NA	NA	NA	NA	NA

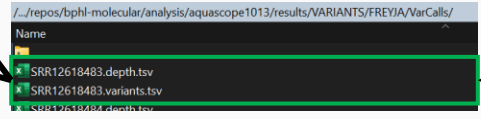
MN908947.3	1	A	0
MN908947.3	2	T	0
MN908947.3	3	T	0
MN908947.3	4	A	0
MN908947.3	5	A	0
MN908947.3	6	A	0
MN908947.3	7	G	0
MN908947.3	8	G	0
MN908947.3	9	T	0
MN908947.3	10	T	0
MN908947.3	11	T	0
MN908947.3	12	A	0
MN908947.3	13	T	0
MN908947.3	14	A	0











Advanced Molecular Detection
Southeast Region Bioinformatics

Conclusion

- Fundamentals of Aquascope
- Installation and setup of Aquascope in HPG
- Successfully executed job query for Aquascope
- 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 Bernhoft, MPH
Bioinformatician
Samantha.bernhoft@flhealth.gov