MycoSNP

Advanced Molecular Detection

Southeast Region Bioinformatics

Overview

Purpose:
This nextflow pipeline is developed for fungal whole-genome sequencing (e.g., *Candida auris*). This pipeline performs QC, read alignment to a masked reference, and SNP calling to generate phylogenetic outputs.

Usage:
Can be used by public health labs and researchers for molecular typing, outbreak investigation, surveillance, and comparative genomic studies

Dependencies:
Nextflow
Apptainer
Java
Reference Genome

Advanced Molecular Detection
Southeast Region Bioinformatics

Workflow

Sample Sheet

C. auris Illumina FASTQs paired sequences

Setup Environment
(Java, Apptainer, Conda, Nextflow)

sbatch run_mycosnp.sh

Results

Merged SNP Matrix

Combined Variant Calls

Samples

Lane

MultiQC

Pipeline Info

Stats
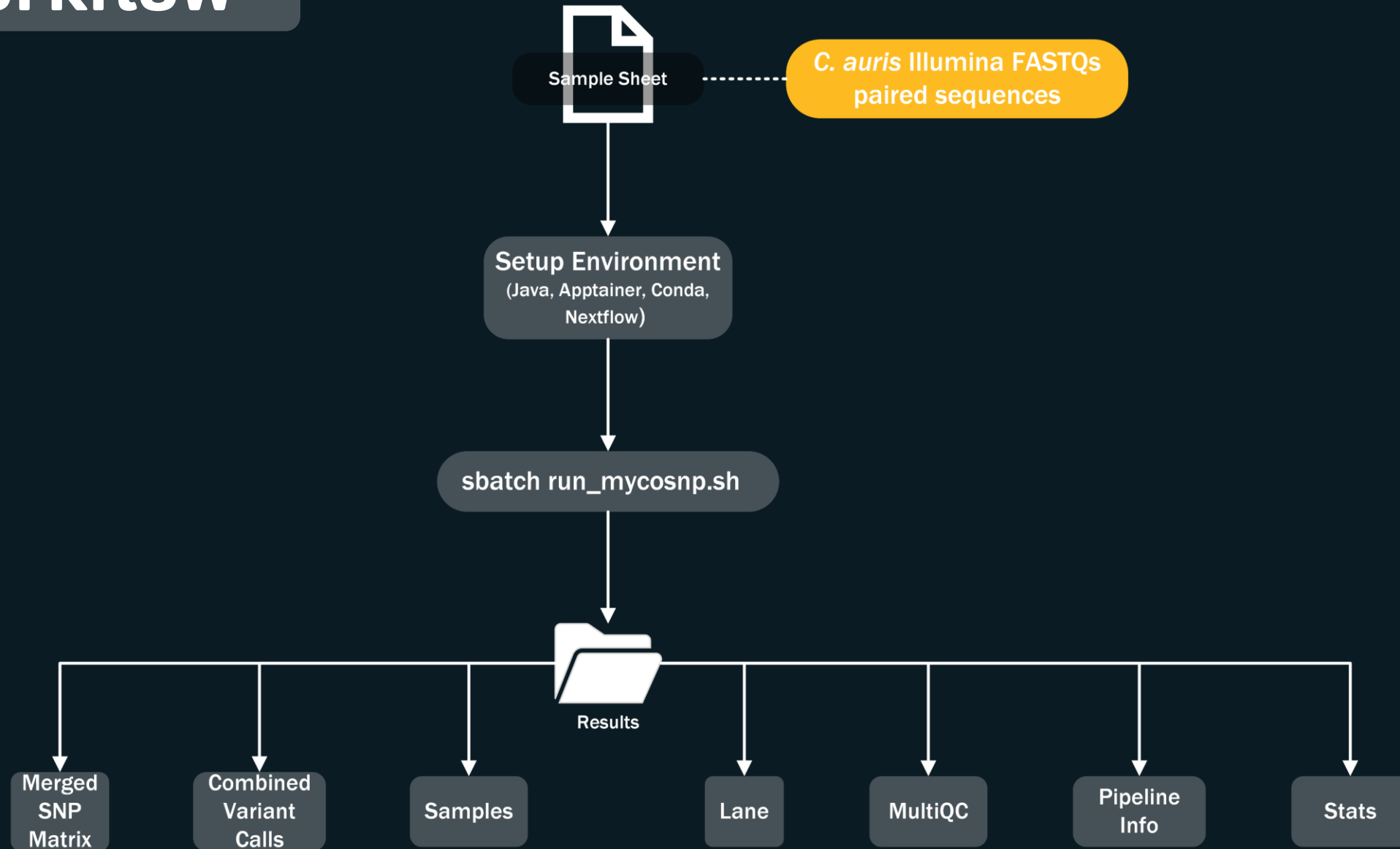
Application

Objective

Use *C. auris* Illumina FASTQs paired sequences from NCBI database, and use Mycosnp to analyze the samples

Application Cont.

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

git clone https://github.com/CDCgov/mycosnp-nf

mkdir analysis/

cd analysis/

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

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/

Name
- ..
- assets
- bin
- conf
- docs
- input
- lib
- Misc
- modules
- mycosnp_results
- subworkflows
- work
- workflows
- CHANGELOG.md
- CITATIONS.md
- CODE_OF_CONDUCT.md
- code-of-conduct.md
- CONTRIBUTING.md
- DISCLAIMER.md
- LICENSE
- main.nf
- modules.json
- mycosnp.15654644.err
- mycosnp.15654644.out
- nextflow.config
- nextflow_schema.json
- open_practices.md
- README.md
- rules_of_behavior.md
- run_mycosnp.sh
- tower.yml

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/input/

Name
- ..
- samplesheet.csv
- SRR35732983_R1.fastq.gz
- SRR35732983_R2.fastq.gz
- SRR35732984_R1.fastq.gz
- SRR35732984_R2.fastq.gz
- SRR35732985_R1.fastq.gz
- SRR35732985_R2.fastq.gz
- SRR35732986_R1.fastq.gz
- SRR35732986_R2.fastq.gz
- SRR35732987_R1.fastq.gz
- SRR35732987_R2.fastq.gz
- SRR35732988_R1.fastq.gz
- SRR35732988_R2.fastq.gz
- SRR35732989_R1.fastq.gz
- SRR35732989_R2.fastq.gz
- SRR35732990_R1.fastq.gz
- SRR35732990_R2.fastq.gz
- SRR35732991_R1.fastq.gz
- SRR35732991_R2.fastq.gz
- SRR35732992_R1.fastq.gz
- SRR35732992_R2.fastq.gz
- SRR35732993_R1.fastq.gz
- SRR35732993_R2.fastq.gz
- SRR35732994_R1.fastq.gz
- SRR35732994_R2.fastq.gz
- SRR35732995_R1.fastq.gz
- SRR35732995_R2.fastq.gz

```
sample,fastq_1,fastq_2
SRR35732983,/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/input/SRR35732983_R1.fastq.gz,/blue/bphl-florida/n.yengalareddy/
SRR35732984,/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/input/SRR35732984_R1.fastq.gz,/blue/bphl-florida/n.yengalareddy/
SRR35732985,/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/input/SRR35732985_R1.fastq.gz,/blue/bphl-florida/n.yengalareddy/
SRR35732986,/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/input/SRR35732986_R1.fastq.gz,/blue/bphl-florida/n.yengalareddy/
SRR35732987,/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/input/SRR35732987_R1.fastq.gz,/blue/bphl-florida/n.yengalareddy/
SRR35732988,/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/input/SRR35732988_R1.fastq.gz,/blue/bphl-florida/n.yengalareddy/
SRR35732989,/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/input/SRR35732989_R1.fastq.gz,/blue/bphl-florida/n.yengalareddy/
SRR35732990,/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/input/SRR35732990_R1.fastq.gz,/blue/bphl-florida/n.yengalareddy/
SRR35732991,/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/input/SRR35732991_R1.fastq.gz,/blue/bphl-florida/n.yengalareddy/
SRR35732992,/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/input/SRR35732992_R1.fastq.gz,/blue/bphl-florida/n.yengalareddy/
SRR35732993,/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/input/SRR35732993_R1.fastq.gz,/blue/bphl-florida/n.yengalareddy/
SRR35732995,/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/input/SRR35732995_R1.fastq.gz,/blue/bphl-florida/n.yengalareddy/
```

Application Cont.

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

# Environment Setup
module load conda
conda activate Nextflow
module load apptainer

# Paths
WORKDIR="${PWD}"
SAMPLESHEET="${WORKDIR}/input/samplesheet.csv"
FASTA="/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/assets/precomputed/reference/masked/GCA_016772135.1_ASM1677213v1_genomic/masked/reference.fa"
OUTDIR="${WORKDIR}/mycosnp_results"
NF_WORK="${WORKDIR}/work"

mkdir -p "${OUTDIR}" "${NF_WORK}"

# Run
nextflow run CDCgov/mycosnp-nf \
    --input "${SAMPLESHEET}" \
    --fasta "${FASTA}" \
    --outdir "${OUTDIR}" \
    -profile singularity \
    -work-dir "${NF_WORK}" \
    -resume \
    -with-report "${OUTDIR}/report.html" \
    -with-trace  "${OUTDIR}/trace.txt" \
    -with-timeline "${OUTDIR}/timeline.html"

echo "[DONE] Results in ${OUTDIR}"
```
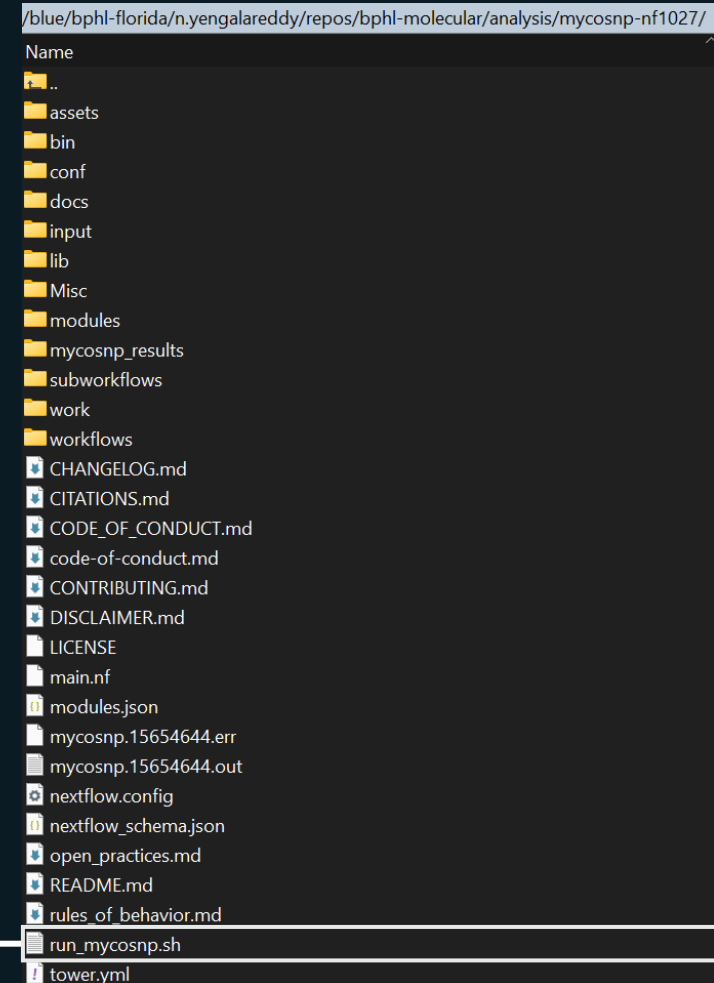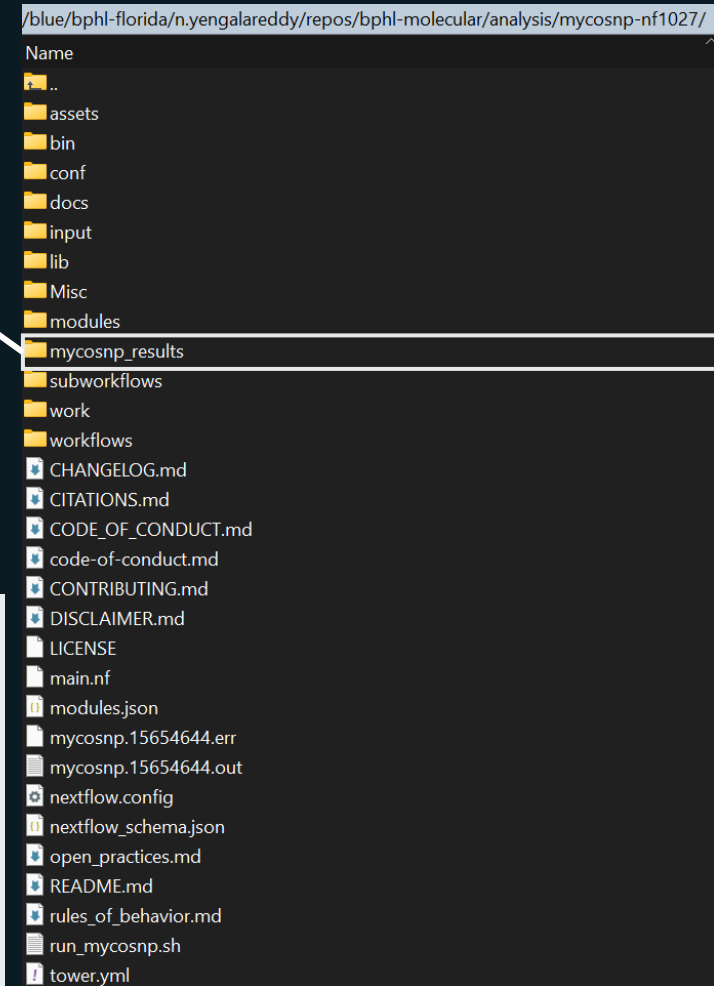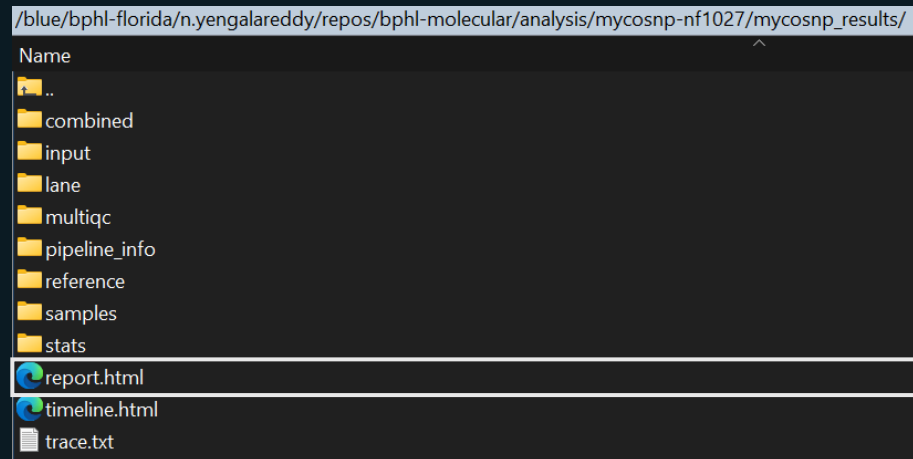
Name
- ..
- assets
- bin
- conf
- docs
- input
- lib
- Misc
- modules
- mycosnp_results
- subworkflows
- work
- workflows
- CHANGELOG.md
- CITATIONS.md
- CODE_OF_CONDUCT.md
- code-of-conduct.md
- CONTRIBUTING.md
- DISCLAIMER.md
- LICENSE
- main.nf
- modules.json
- mycosnp.15654644.err
- mycosnp.15654644.out
- nextflow.config
- nextflow_schema.json
- open_practices.md
- README.md
- rules_of_behavior.md
- run_mycosnp.sh
- tower.yml

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/

Name
..
combined
input
lane
multiqc
pipeline_info
reference
samples
stats
report.html
timeline.html
trace.txt

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/

Name
..
assets
bin
conf
docs
input
lib
Misc
modules
mycosnp_results
subworkflows
work
workflows
CHANGELOG.md
CITATIONS.md
CODE_OF_CONDUCT.md
code-of-conduct.md
CONTRIBUTING.md
DISCLAIMER.md
LICENSE
main.nf
modules.json
mycosnp.15654644.err
mycosnp.15654644.out
nextflow.config
nextflow_schema.json
open_practices.md
README.md
rules_of_behavior.md
run_mycosnp.sh
tower.yml

Nextflow workflow report

[disturbed_ampere] *(resumed run)*

Workflow execution completed successfully!

Run times
13-Oct-2025 23:30:41 - 14-Oct-2025 03:33:28 (duration: **4h 2m 46s**)

247 succeeded

Nextflow command

```
nextflow run CDCgov/mycosnp-nf --input /blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-
nf1027/input/samplesheet.csv --fasta
/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-
nf1027/assets/precomputed/reference/masked/GCA_016772135.1_ASM1677213v1_genomic/masked/reference.fa --outdir /blue/bphl-
florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results -profile singularity -work-dir /blue/bphl-
florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/work -resume -with-report /blue/bphl-florida/n.yengalareddy/repos/bphl-
molecular/analysis/mycosnp-nf1027/mycosnp_results/report.html -with-trace /blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-
nf1027/mycosnp_results/trace.txt -with-timeline /blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-
nf1027/mycosnp_results/timeline.html
```

CPU-Hours	30.4
Launch directory	/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027
Work directory	/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/work
Project directory	/home/n.yengalareddy/.nextflow/assets/CDCgov/mycosnp-nf
Script name	main.nf
Script ID	83e5fd41d6d046e50cd967695e74d11e
Workflow session	64d84f72-296c-44bd-822a-b748eae7b99b
Workflow repository	https://github.com/CDCgov/mycosnp-nf , revision master (commit hash c2d8698a3fb82c54d015f36580f3dafd0ef69669)
Workflow profile	singularity
Nextflow version	version 24.10.2, build 5932 (27-11-2024 21:23 UTC)

sbatch run_mycosnp.sh

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/

Name
- ..
- combined
- input
- lane
- multiqc
- pipeline_info
- reference
- samples
- stats
- report.html
- timeline.html
- trace.txt

/.../bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/combined/gvcf/

Name
- ..
- combined.combined.g.vcf.gz
- combined.combined.g.vcf.gz.tbi

/.../n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/combined/selectedsnps/

Name
- ..
- combined_genotype_filtered_snps.selectvariants.vcf.gz
- combined_genotype_filtered_snps.selectvariants.vcf.gz.tbi

/.../repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/combined/selectedsnpsfiltered/

Name
- ..
- finalfiltered.vcf.gz.csi

/.../n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/combined/splitvcf/

Name
- ..
- samplelist.txt
- SRR35732983.vcf.gz
- SRR35732984.vcf.gz
- SRR35732985.vcf.gz
- SRR35732986.vcf.gz
- SRR35732987.vcf.gz
- SRR35732988.vcf.gz
- SRR35732989.vcf.gz
- SRR35732990.vcf.gz
- SRR35732991.vcf.gz
- SRR35732992.vcf.gz
- SRR35732993.vcf.gz
- SRR35732994.vcf.gz
- SRR35732995.vcf.gz

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/combined/

Name
- ..
- filteredgvcfs
- finalfiltered
- genotypegvcfs
- gvcf
- phylogeny
- selectedsnps
- selectedsnpsfiltered
- snpdists
- splitvcf
- vcf-qc-report
- vcf-to-fasta

/.../n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/combined/filteredgvcfs/

Name
- ..
- combined_genotype_filtered.vcf.gz
- combined_genotype_filtered.vcf.gz.tbi

/.../n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/combined/vcf-qc-report/
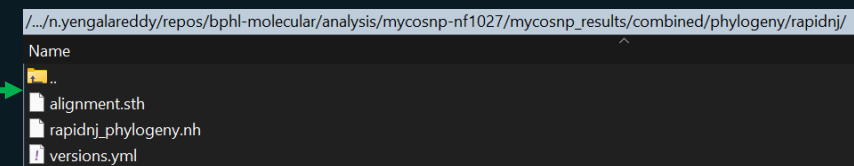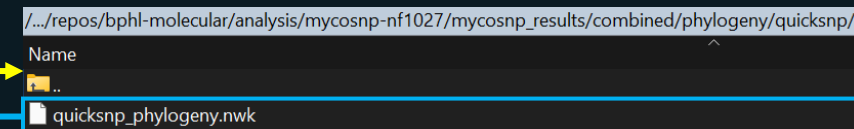
Name
- ..
- vcf-qc-report.txt

/.../n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/combined/finalfiltered/
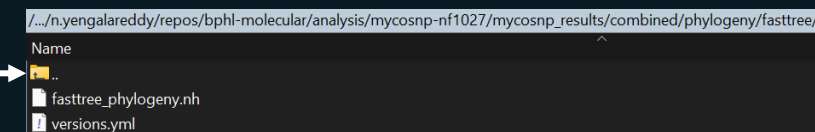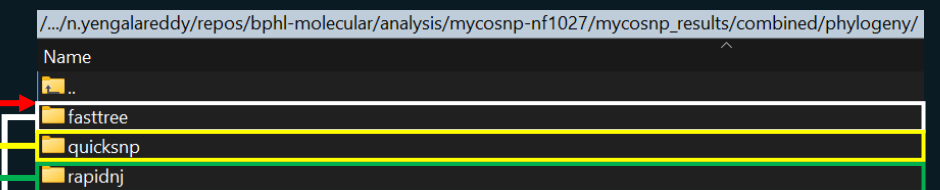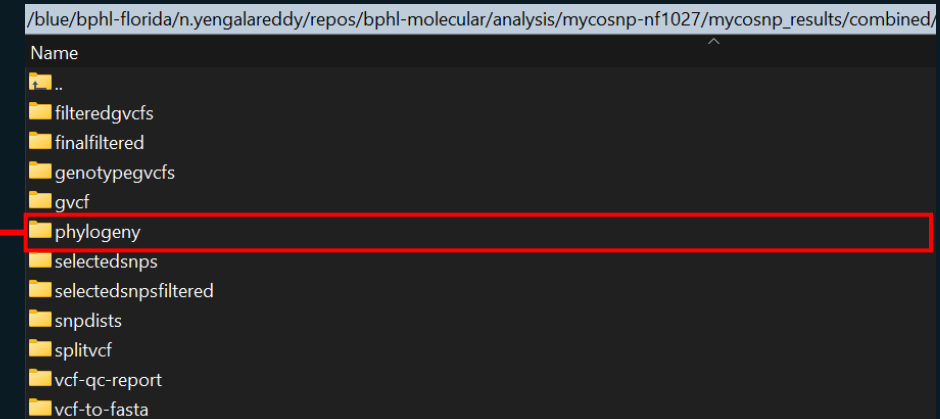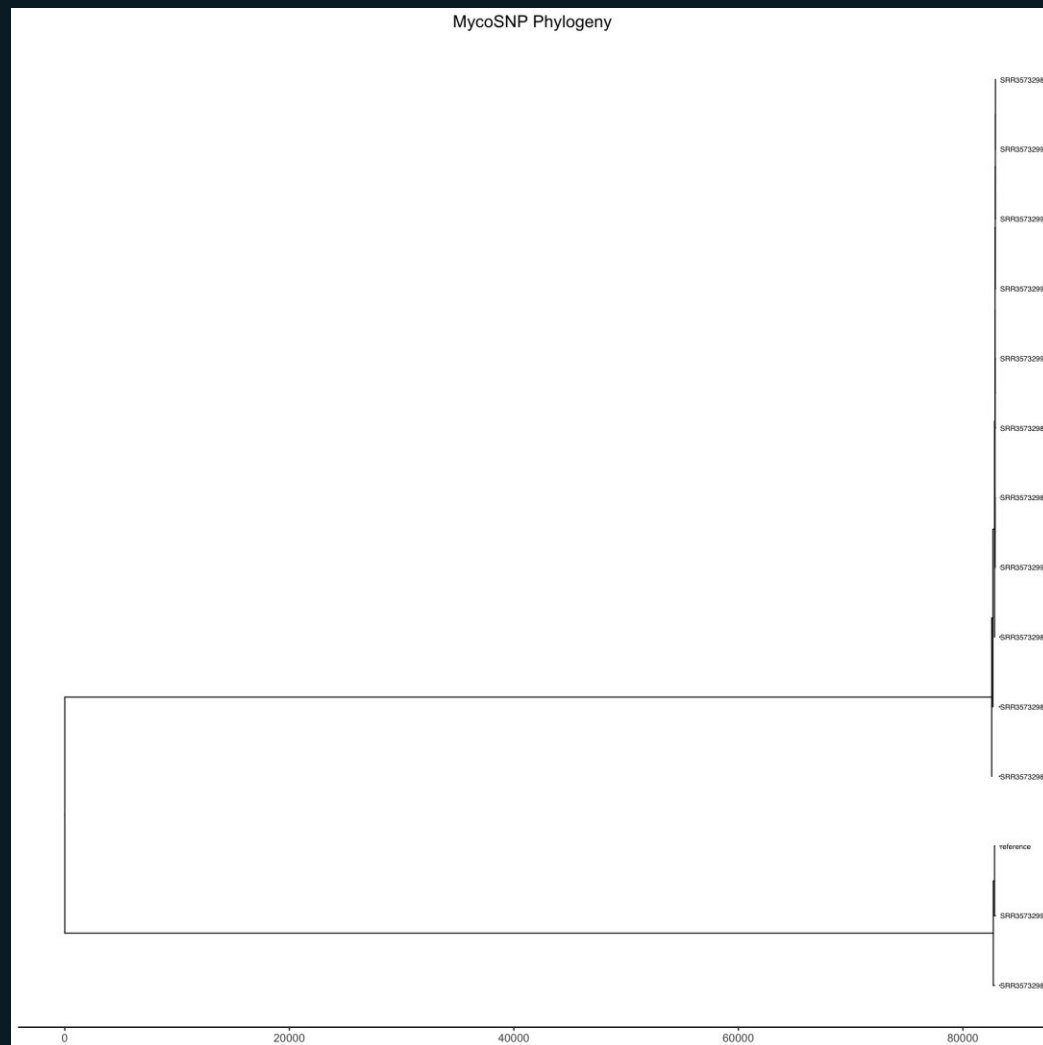
Name
- ..
- finalfiltered.vcf.gz

/.../n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/combined/vcf-to-fasta/

Name
- ..
- vcf-to-fasta.fasta

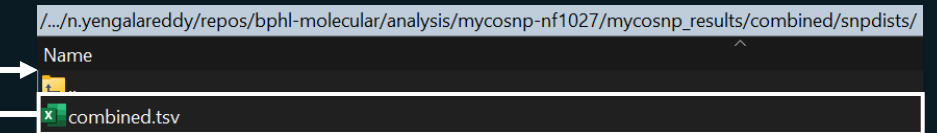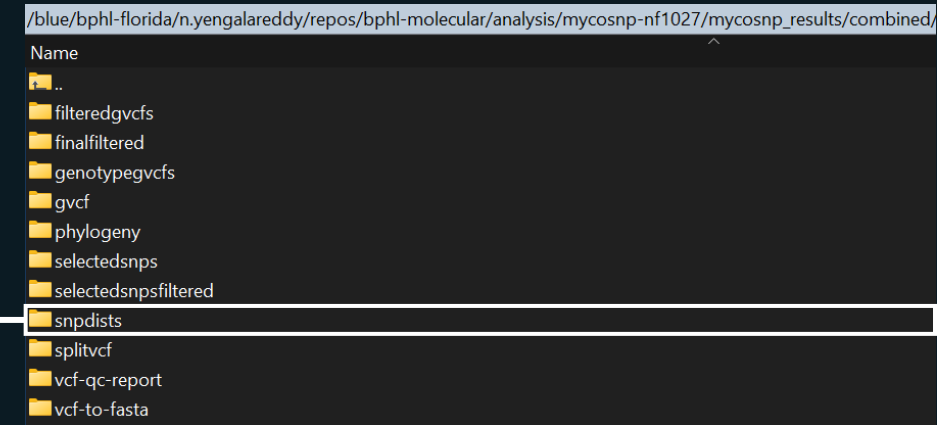/.../n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/combined/genotypegvcfs/

Name
- ..
- combined_genotype.vcf.gz
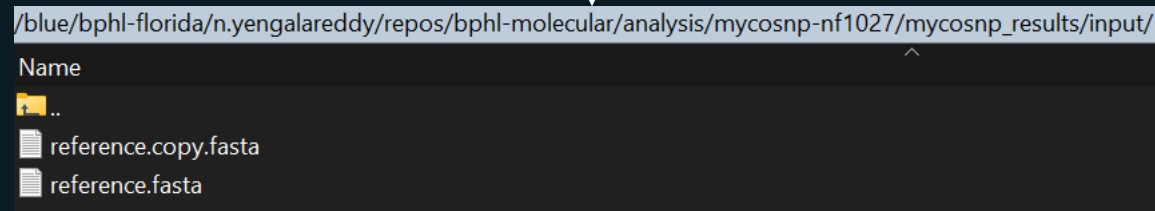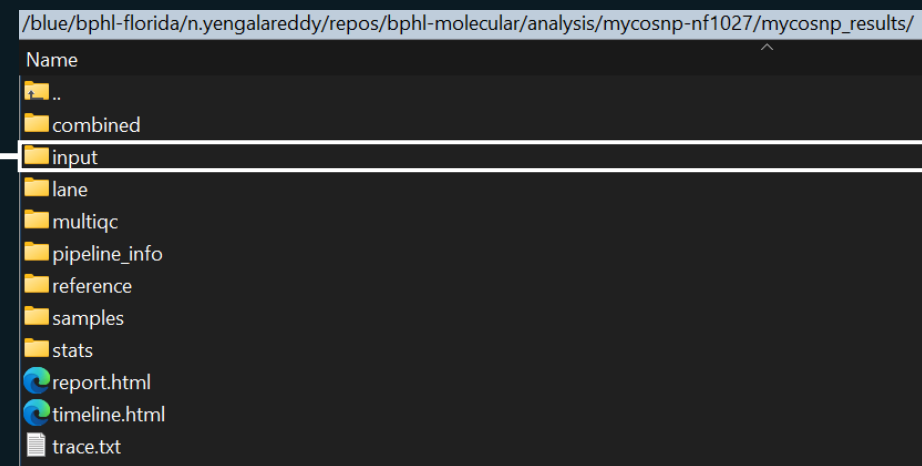- combined_genotype.vcf.gz.tbi

Application Cont.

MycoSNP Phylogeny

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/combined/

Name
- ..
- filteredgvcfs
- finalfiltered
- genotypegvcfs
- gvcf
- phylogeny
- selectedsnps
- selectedsnpsfiltered
- snpdists
- splitvcf
- vcf-qc-report
- vcf-to-fasta

/.../n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/combined/phylogeny/

Name
- ..
- fasttree
- quicksnp
- rapidnj

/.../n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/combined/phylogeny/fasttree/

Name
- ..
- fasttree_phylogeny.nh
- versions.yml

/.../repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/combined/phylogeny/quicksnp/

Name
- ..
- quicksnp_phylogeny.nwk

/.../n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/combined/phylogeny/rapidnj/

Name
- ..
- alignment.sth
- rapidnj_phylogeny.nh
- versions.yml

Application Cont.

Name

..
filteredgvcfs
finalfiltered
genotypegvcfs
gvcf
phylogeny
selectedsnps
selectedsnpsfiltered
snpdists
splitvcf
vcf-qc-report
vcf-to-fasta

/.../n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/combined/snpdists/

Name

..
combined.tsv

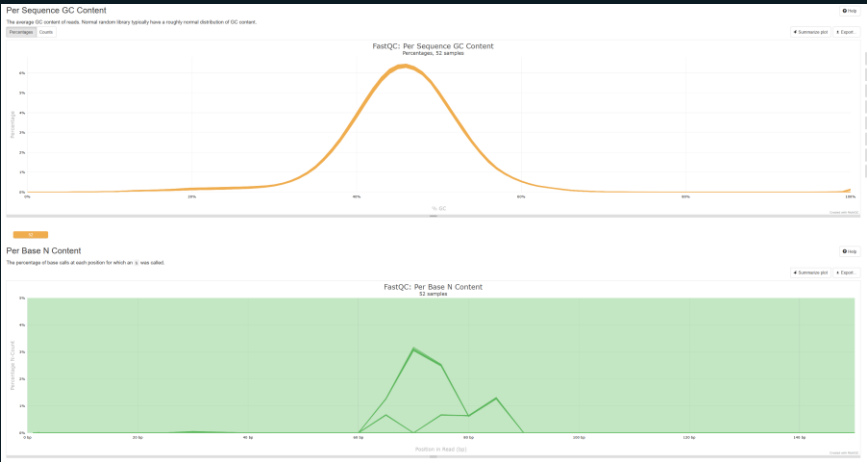| snp-dists 0.8.2	reference	SRR35732983	SRR35732984	SRR35732985	SRR35732986	SRR35732987	SRR35732988	SRR35732989	SRR35732990	SRR35732991	SRR35732992	SRR35732993	SRR35732994	SRR35732995
reference	0	238	166464	166429	165527	166411	165416	166006	166488	124	166460	166144	166420	166480
SRR35732983	238	0	166428	166394	165493	166374	165381	165972	166453	316	166425	166109	166385	166445
SRR35732984	166464	166428	0	73	74	63	120	67	21	166522	61	46	77	36
SRR35732985	166429	166394	73	0	60	51	102	55	77	166487	48	83	60	80
SRR35732986	165527	165493	74	60	0	53	108	56	79	165585	46	85	62	79
SRR35732987	166411	166374	63	51	53	0	97	12	63	166467	38	64	18	65
SRR35732988	165416	165381	120	102	108	97	0	96	124	165475	90	126	104	122
SRR35732989	166006	165972	67	55	56	12	96	0	66	166065	38	67	8	64
SRR35732990	166488	166453	21	77	79	63	124	66	0	166546	57	48	73	33
SRR35732991	124	316	166522	166487	165585	166467	165475	166065	166546	0	166518	166202	166478	166538
SRR35732992	166460	166425	61	48	46	38	90	38	57	166518	0	69	40	51
SRR35732993	166144	166109	46	83	85	64	126	67	48	166202	69	0	80	46
SRR35732994	166420	166385	77	60	62	18	104	8	73	166478	40	80	0	73
SRR35732995	166480	166445	36	80	79	65	122	64	33	166538	51	46	73	0

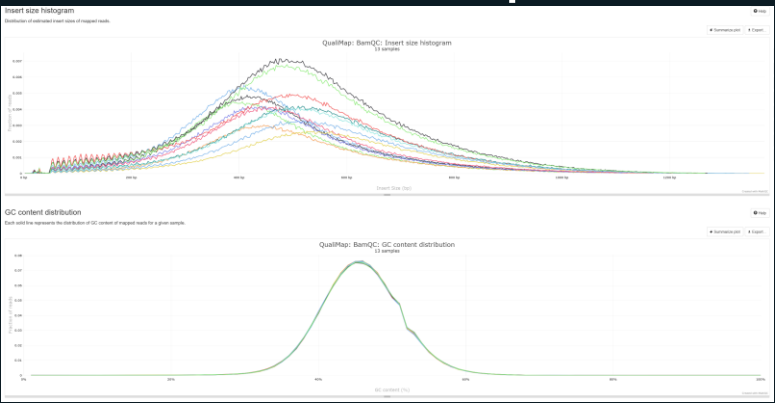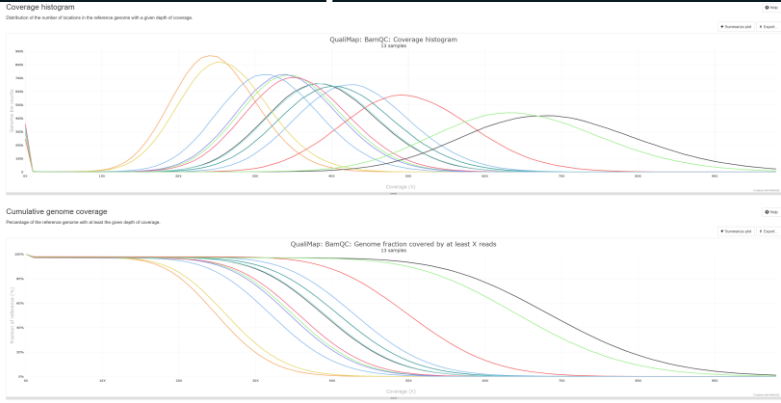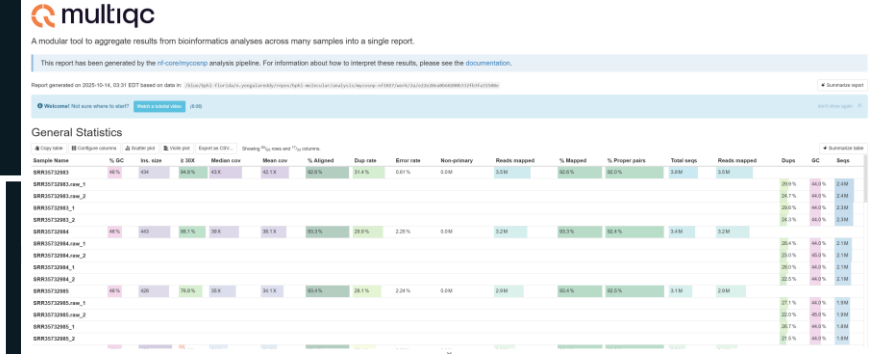# Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/

Name
- 📁 ..
- 📁 combined
- 📁 input
- 📁 lane
- 📁 multiqc
- 📁 pipeline_info
- 📁 reference
- 📁 samples
- 📁 stats
- 🌐 report.html
- 🌐 timeline.html
- 📄 trace.txt

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/input/

Name
- 📁 ..
- 📄 reference.copy.fasta
- 📄 reference.fasta

Application Cont.

Application Cont.

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/

Name
- ..
- combined
- input
- lane
- multiqc
- pipeline_info
- reference
- samples
- stats
- report.html
- timeline.html
- trace.txt

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/samples/

Name
- ..
- SRR35732983
- SRR35732984
- SRR35732985
- SRR35732986
- SRR35732987
- SRR35732988
- SRR35732989
- SRR35732990
- SRR35732991
- SRR35732992
- SRR35732993
- SRR35732994
- SRR35732995

/.../n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/samples/SRR35732983/

Name
- ..
- faqcs
- fastqc_post
- fastqc_raw
- finalbam
- qc_report
- variant_calling

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/

Name

📁 ..
📁 combined
📁 input
📁 lane
📁 multiqc
📁 pipeline_info
📁 reference
📁 samples
📁 stats
🌐 report.html
🌐 timeline.html
📄 trace.txt

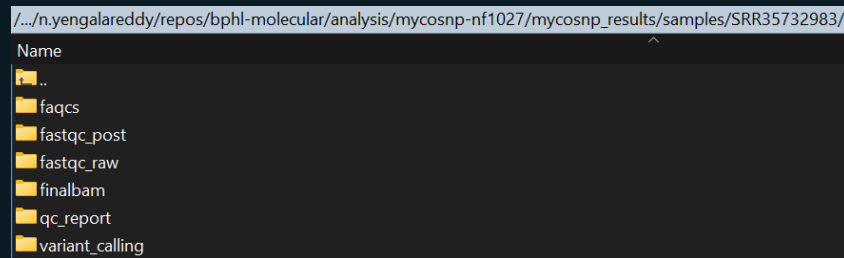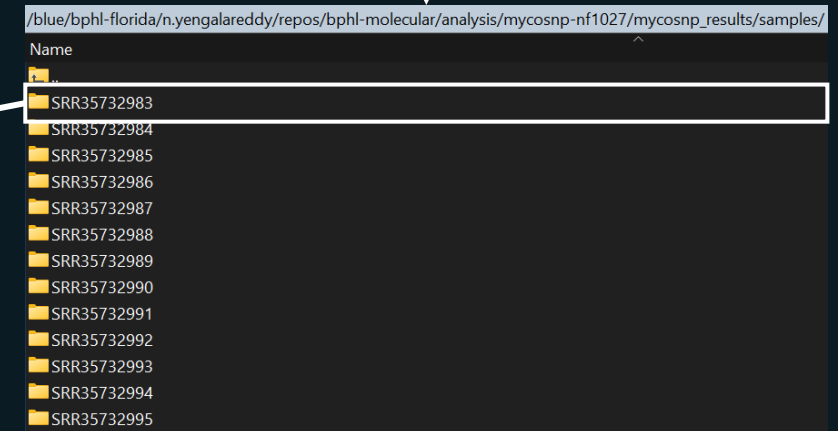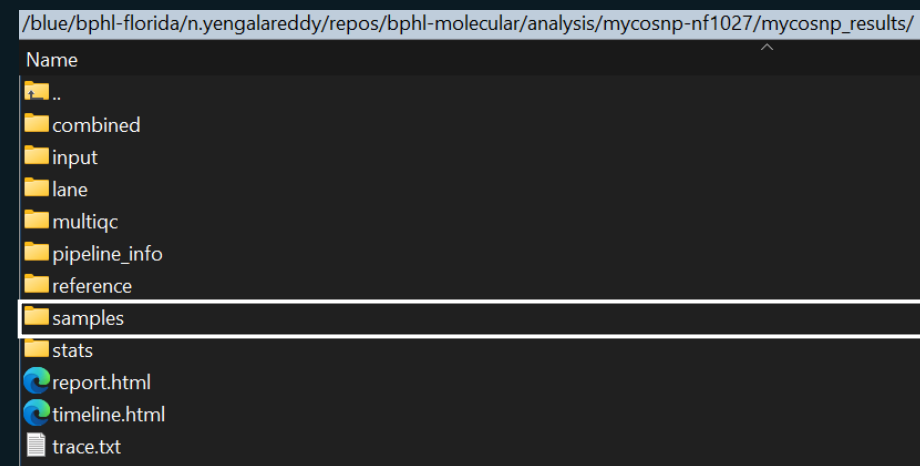/.../n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/stats/samtools_flagstat/

Name

📁 ..
📄 SRR35732983.flagstat
📄 SRR35732984.flagstat
📄 SRR35732985.flagstat
📄 SRR35732986.flagstat
📄 SRR35732987.flagstat
📄 SRR35732988.flagstat
📄 SRR35732989.flagstat
📄 SRR35732990.flagstat
📄 SRR35732991.flagstat
📄 SRR35732992.flagstat
📄 SRR35732993.flagstat
📄 SRR35732994.flagstat
📄 SRR35732995.flagstat

/.../n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/stats/samtools_idxstats/

Name

📁 ..
📄 SRR35732983.idxstats
📄 SRR35732984.idxstats
📄 SRR35732985.idxstats
📄 SRR35732986.idxstats
📄 SRR35732987.idxstats
📄 SRR35732988.idxstats
📄 SRR35732989.idxstats
📄 SRR35732990.idxstats
📄 SRR35732991.idxstats
📄 SRR35732992.idxstats
📄 SRR35732993.idxstats
📄 SRR35732994.idxstats
📄 SRR35732995.idxstats

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/stats/

Name

📁 ..
📁 qc_report
📁 qualimap
📁 samtools_flagstat
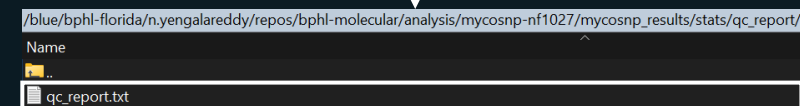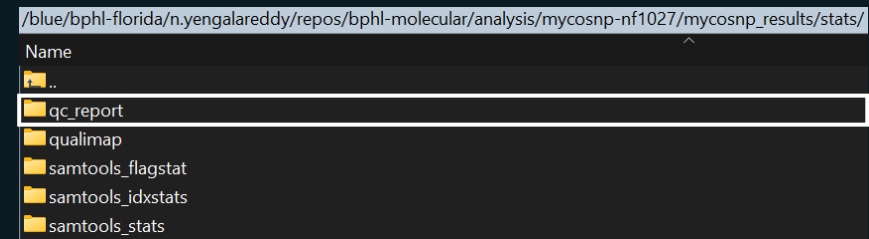📁 samtools_idxstats
📁 samtools_stats

/.../n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/stats/samtools_stats/

Name

📁 ..
📄 SRR35732983.stats
📄 SRR35732984.stats
📄 SRR35732985.stats
📄 SRR35732986.stats
📄 SRR35732987.stats
📄 SRR35732988.stats
📄 SRR35732989.stats
📄 SRR35732990.stats
📄 SRR35732991.stats
📄 SRR35732992.stats
📄 SRR35732993.stats
📄 SRR35732994.stats
📄 SRR35732995.stats

```
# This file was produced by samtools stats (1.15+htslib-1.15) and can be plotted using plot-bamstats
# This file contains statistics for all reads.
# The command line was:  stats --threads 1 --reference reference.fa SRR35732983.bam
# CHK, Checksum [2]Read Names    [3]Sequences      [4]Qualities
# CHK, CRC32 of reads which passed filtering followed by addition (32bit overflow)
CHK      1a8b9139      a55f966b      dd437f18
# Summary Numbers. Use `grep ^SN | cut -f 2-` to extract this part.
SN      raw total sequences:     3809318 # excluding supplementary and secondary reads
SN      filtered sequences:      0
SN      sequences:      3809318
SN      is sorted:      1
SN      1st fragments:  1903031
SN      last fragments: 1906287
SN      reads mapped:   3528272
SN      reads mapped and paired:        3521326 # paired-end technology bit set + both mates mapped
SN      reads unmapped: 281046
SN      reads properly paired:  3503676 # proper-pair bit set
SN      reads paired:   3809318 # paired-end technology bit set
SN      reads duplicated:       0       # PCR or optical duplicate bit set
SN      reads MQ0:      3077    # mapped and MQ=0
SN      reads QC failed:        0
SN      non-primary alignments: 0
SN      supplementary alignments:       12802
SN      total length:   569049190       # ignores clipping
SN      total first fragment length:    284286648       # ignores clipping
SN      total last fragment length:     284762542       # ignores clipping
SN      bases mapped:   527003664       # ignores clipping
SN      bases mapped (cigar):   522874252       # more accurate
SN      bases trimmed:  0
SN      bases duplicated:       0
SN      mismatches:     3210487 # from NM fields
SN      error rate:     6.140075e-03    # mismatches / bases mapped (cigar)
SN      average length: 149
SN      average first fragment length:  149
SN      average last fragment length:   149
SN      maximum length: 151
SN      maximum first fragment length:  151
SN      maximum last fragment length:   151
SN      average quality:        32.2
SN      insert size average:    447.1
SN      insert size standard deviation: 161.6
SN      inward oriented pairs:  1731216
SN      outward oriented pairs: 25224
SN      pairs with other orientation:   719
SN      pairs on different chromosomes: 3504
SN      percentage of properly paired reads (%):        92.0
```

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/stats/

Name
..
qc_report
qualimap
samtools_flagstat
samtools_idxstats
samtools_stats

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/stats/qc_report/

Name
..
qc_report.txt

Sample Name	Reads Before Trimming	GC Before Trimming	Average Q Score Before Trimming	Reference Length Coverage Before Trimming	Reads After Trimming	Paired Reads After Trimming	Unpaired Reads After Trimming	GC After Trimming	Average Q Score After Trimming	Reference Length Coverage After Trimming	Mean Coverage Depth	Reads Mapped	Genome Fraction at 10X
SRR35732983	4778776	44.68%	31.53	57.41	4685273 (98.04 %)	4607102 (98.33 %)	78171 (1.67 %)	44.68%	31.56	56.41	42.15	3528272 (92.62%)	97.61%
SRR35732984	4268062	44.85%	31.44	51.35	4186243 (98.08 %)	4116364 (98.33 %)	69879 (1.67 %)	44.83%	31.47	50.46	38.08	3202599 (93.29%)	97.34%
SRR35732985	3811464	44.86%	31.43	45.92	3739364 (98.11 %)	3675820 (98.30 %)	63544 (1.70 %)	44.84%	31.46	45.12	34.13	2869029 (93.42%)	97.33%
SRR35732986	2682508	44.84%	31.35	32.38	2631790 (98.11 %)	2585454 (98.24 %)	46336 (1.76 %)	44.81%	31.38	31.8	24.32	2041642 (93.15%)	97.12%
SRR35732987	3774968	44.76%	31.47	45.34	3699867 (98.01 %)	3637598 (98.32 %)	62269 (1.68 %)	44.75%	31.5	44.55	33.66	2832649 (93.23%)	97.29%
SRR35732988	3851766	44.91%	31.39	46.31	3776454 (98.04 %)	3713004 (98.32 %)	63450 (1.68 %)	44.89%	31.41	45.51	34.75	2923616 (93.67%)	96.93%
SRR35732989	2675970	44.87%	31.35	32.34	2668864 (99.73 %)	2664722 (99.84 %)	4142 (0.16 %)	44.79%	31.35	32.28	25.56	2142557 (93.02%)	97.17%
SRR35732990	4212104	44.89%	31.44	50.81	4199411 (99.70 %)	4193910 (99.87 %)	5501 (0.13 %)	44.83%	31.45	50.72	39.96	3351841 (93.2%)	97.32%
SRR35732991	5215094	44.66%	31.56	62.28	5183399 (99.39 %)	5177730 (99.89 %)	5669 (0.11 %)	44.60%	31.56	62.13	49.12	4132507 (92.15%)	97.98%
SRR35732992	4017390	44.89%	31.4	48.49	4005727 (99.71 %)	4000514 (99.87 %)	5213 (0.13 %)	44.83%	31.4	48.4	38.26	3208558 (93.36%)	97.34%
SRR35732993	3236024	44.97%	31.27	39.1	3227251 (99.73 %)	3222376 (99.85 %)	4875 (0.15 %)	44.89%	31.28	39.03	31.25	2619957 (93.83%)	97.15%
SRR35732994	7225246	44.79%	31.46	86.87	7195576 (99.59 %)	7186828 (99.88 %)	8748 (0.12 %)	44.72%	31.47	86.7	67.3	5659495 (92.45%)	97.36%
SRR35732995	6815548	44.75%	31.57	81.94	6788658 (99.61 %)	6781714 (99.90 %)	6944 (0.10 %)	44.70%	31.57	81.78	63.2	5310384 (92.37%)	97.35%

Application Cont.

Qualimap Report: BAM QC — QualiMap

Input data and parameters

QualiMap command line

`qualimap bamqc -bam SRR35732983.bam -nw 400 -hm 3`

Alignment

Command line:	bwa mem -t 4 ./bwa/reference SRR35732983_1.trimmed.fastq.gz SRR35732983_2.trimmed.fastq.gz
Draw chromosome limits:	no
Analyze overlapping paired-end reads:	yes
Program:	bwa (0.7.17-r1188)
Analysis date:	Tue Oct 14 04:24:05 GMT 2025
Size of a homopolymer:	3
Skip duplicate alignments:	no
Number of windows:	400
BAM file:	SRR35732983.bam

Summary

Globals

| Reference size | 12,406,106 |
Number of reads	3,809,318
Mapped reads	3,528,272 / 92.62%
Unmapped reads	281,046 / 7.38%
Mapped paired reads	3,528,272 / 92.62%
Mapped reads, first in pair	1,765,117 / 46.34%
Mapped reads, second in pair	1,763,155 / 46.29%
Mapped reads, both in pair	3,521,326 / 92.44%
Mapped reads, singletons	6,946 / 0.18%
Secondary alignments	0
Supplementary alignments	12,802 / 0.34%
Read min/max/mean length	30 / 151 / 149.53
Overlapping read pairs	270,783 / 14.22%
Duplicated reads (estimated)	1,220,151 / 32.03%
Duplication rate	31.43%
Clipped reads	138,274 / 3.63%

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
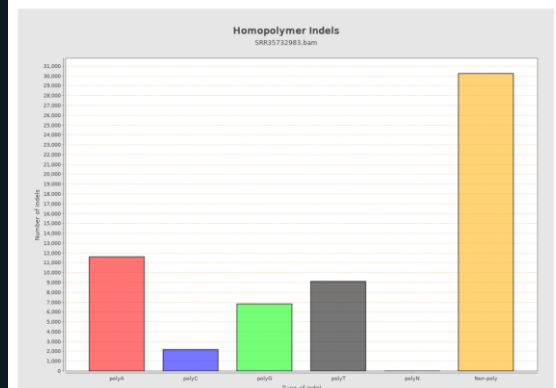- Insert Size Histogram

Coverage Histogram
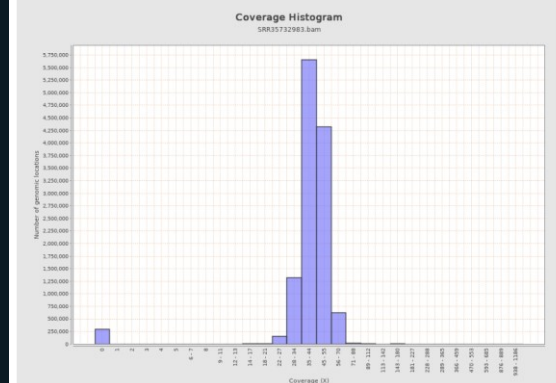
Mapped Reads GC-content Distribution
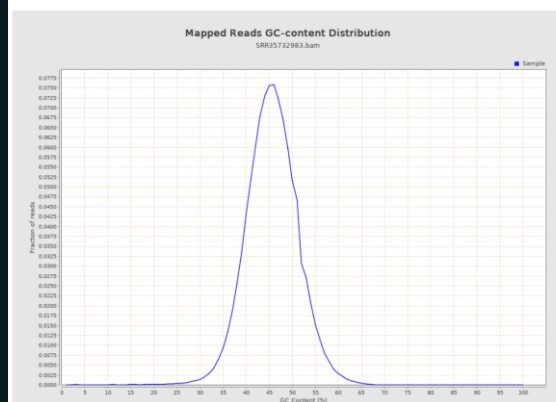
Homopolymer Indels

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/stats/

Name
- ..
- qc_report
- **qualimap**
- samtools_flagstat
- samtools_idxstats
- samtools_stats

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/stats/qualimap/

Name
- ..
- **SRR35732983**
- SRR35732984
- SRR35732985
- SRR35732986
- SRR35732987
- SRR35732988
- SRR35732989
- SRR35732990
- SRR35732991
- SRR35732992
- SRR35732993
- SRR35732994
- SRR35732995
- versions.yml

/.../n.yengalareddy/repos/bphl-molecular/analysis/mycosnp-nf1027/mycosnp_results/stats/qualimap/SRR35732983/

Name
- ..
- css
- images_qualimapReport
- raw_data_qualimapReport
- genome_results.txt
- **qualimapReport.html**

Advanced Molecular Detection
Southeast Region Bioinformatics

Conclusion

Fundamentals of MycoSNP

Installation and setup of MycoSNP in HPG

Successfully executed job query for MycoSNP

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