03 Feb 2025

Pensacola

Advanced Molecular Detection

Southeast Region Bioinformatics

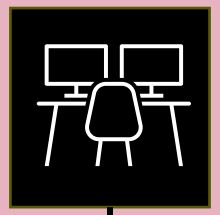
This resource was made possible through funding provided under the Epidemiology and Laboratory Capacity for Prevention and Control of Emerging Infectious Diseases (ELC) Cooperative Agreement (CK24-0002), Project D: Advanced Molecular Detection to the Florida Department of Health. The conclusions, findings, 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.

Advanced Molecular Detection
Southeast Region Bioinformatics

Updates

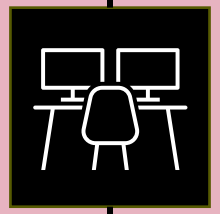
Office Hours

2025



February 17
Talbot

March 03
Sarek_mic



2025

Advanced Molecular Detection
Southeast Region Bioinformatics

Overview

Purpose

- Analyze *Candida auris* long-read sequences data for Quality Reports, Genome Assembly, Species Identification and Abundance, SNP analysis, Antifungal Resistance reports, and Phylogenetic analysis.

Usage

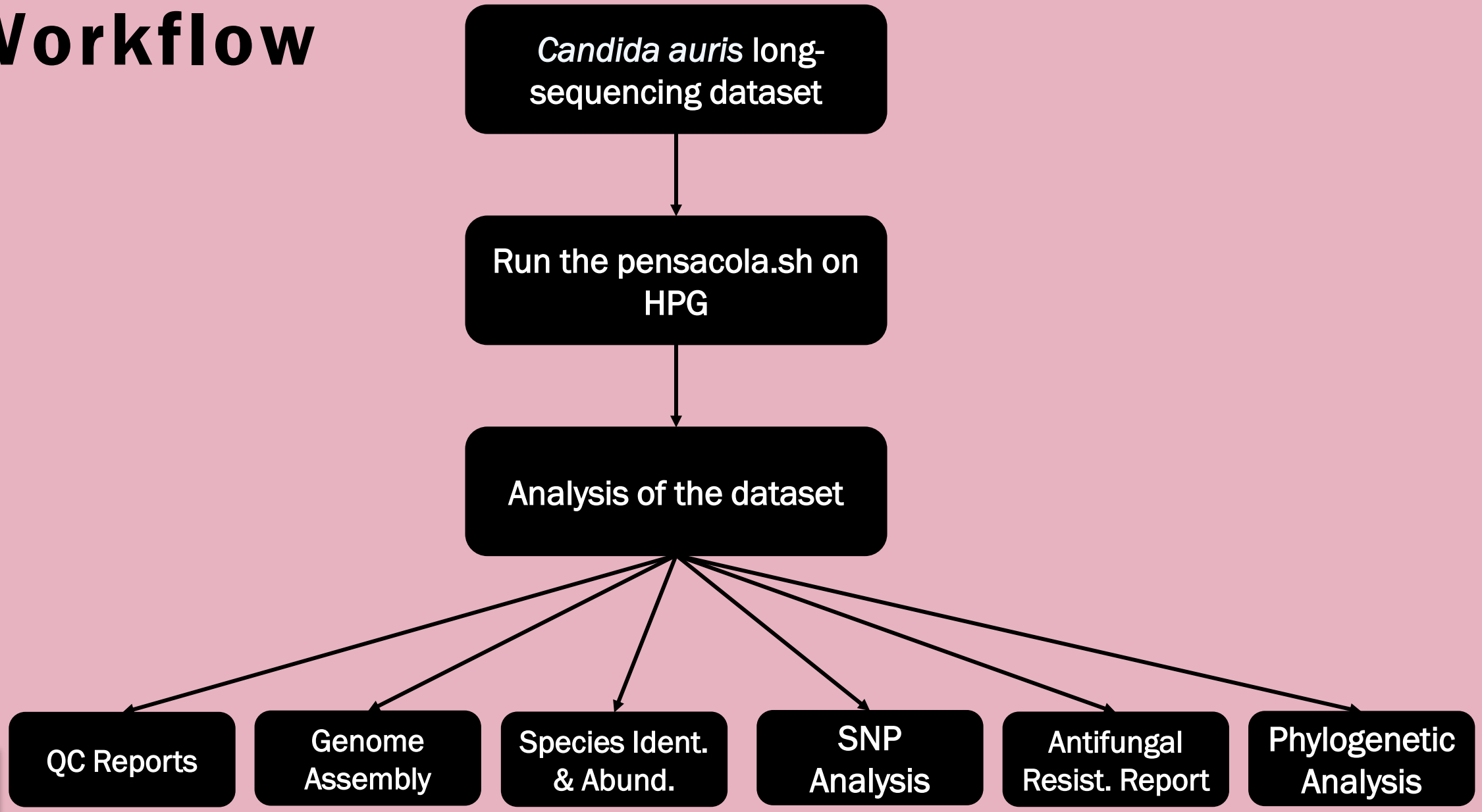
- To support public health and researchers by providing detailed reports and analyses of the data which enables insights into drug resistance monitoring, genomic research, outbreak surveillance and epidemiological studies.

Dependencies

- Nextflow
- Singularity/Apptainer
- SLURM
- Python3
- LongQC
- PacBio SMRTLINK
- Kraken2/Bracken

Advanced Molecular Detection
Southeast Region Bioinformatics

Workflow



Advanced Molecular Detection
Southeast Region Bioinformatics

Application

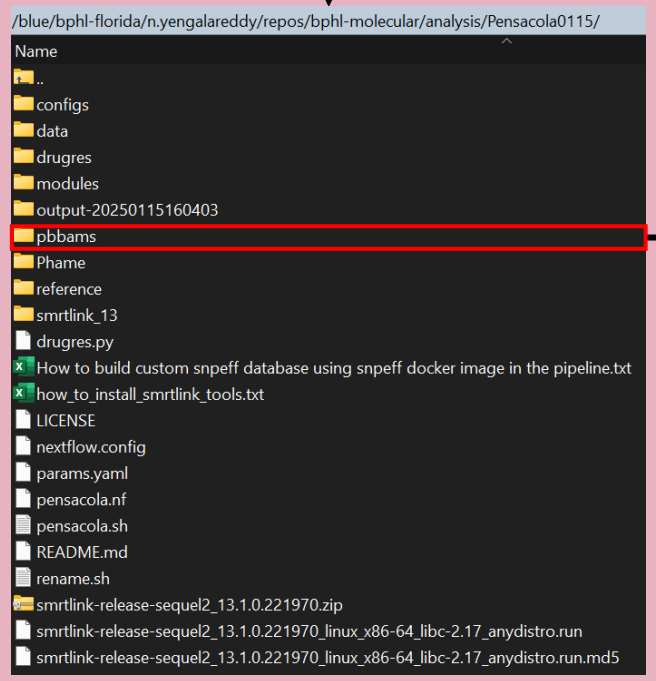
Objective

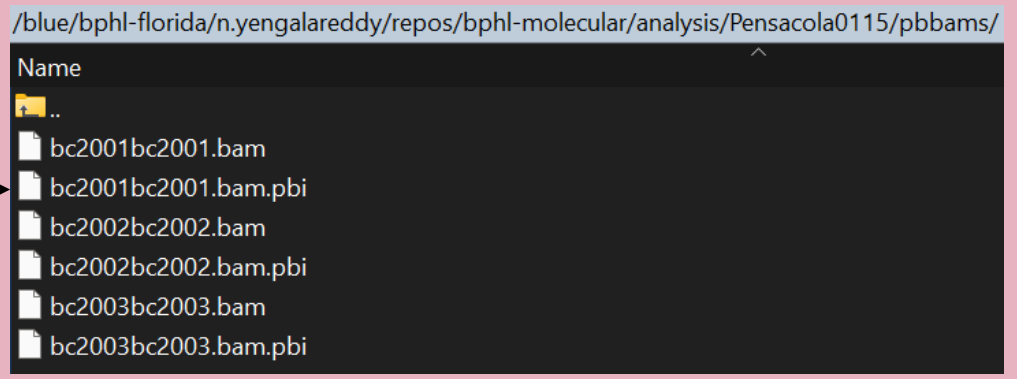
Use *Candida auris* dataset and analyze using **Pensacola** pipeline

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

```
cd /blue/bphl-<state>/<user>/repos/bphl-molecular/
git clone https://github.com/BPHL-Molecular/Pensacola
mkdir analysis/
cd analysis/
cp /blue/bphl-<state>/<user>/repos/bphl-molecular/ Pensacola/
copy .bam and .bam.pbi files to pbbams directory
```

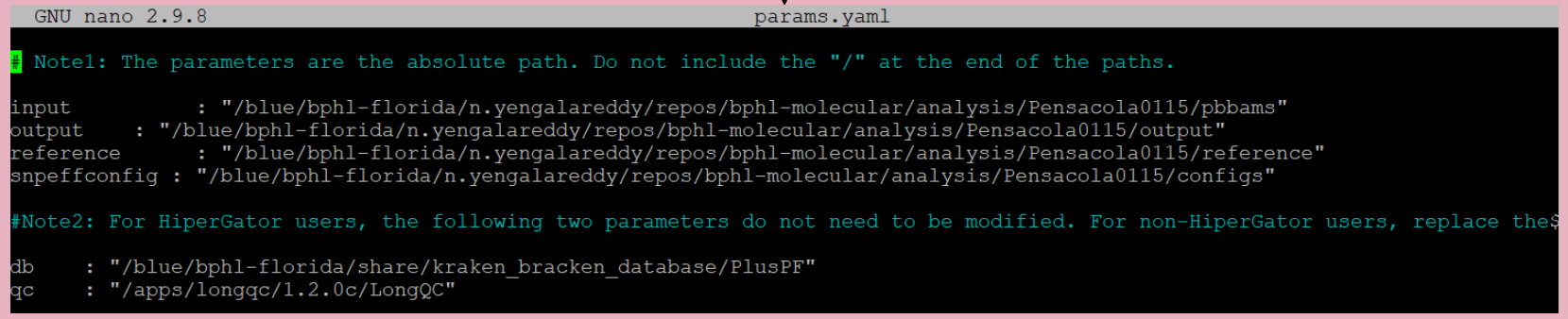




Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

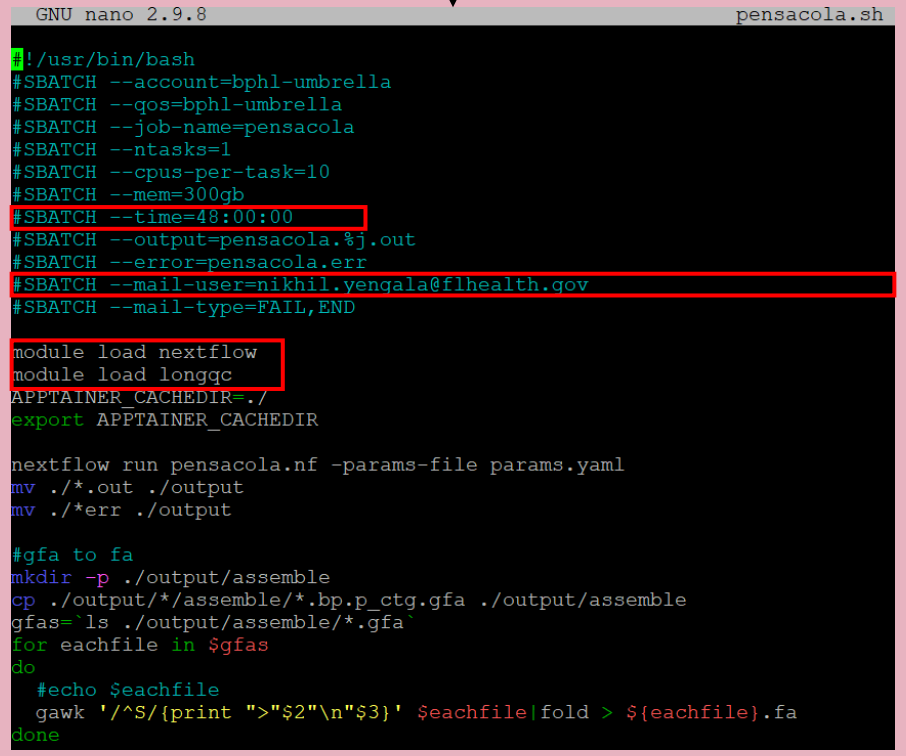
nano params.yaml



Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

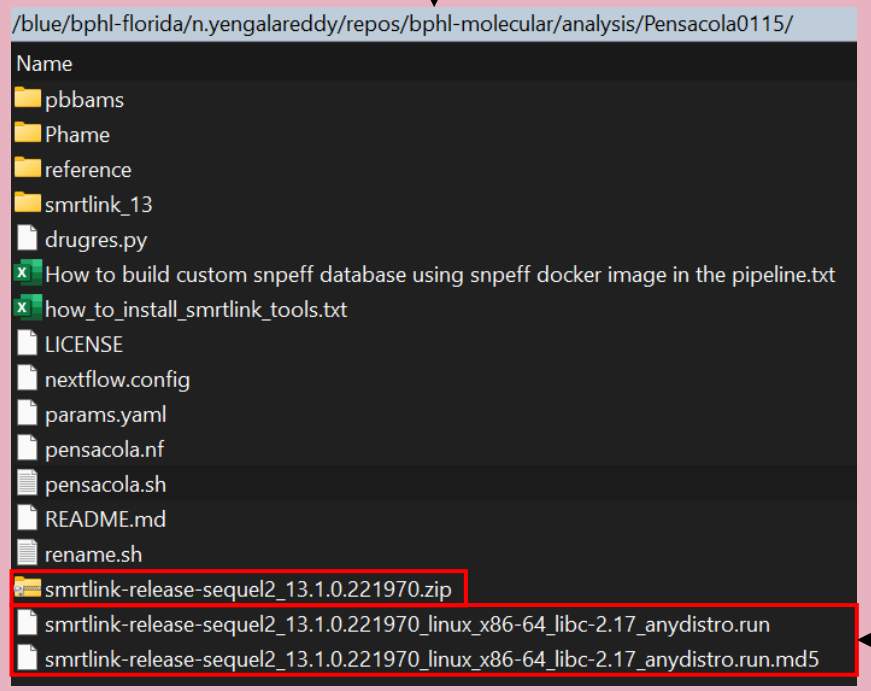
nano pensacola.sh



Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

1) wget https://downloads.pacbcloud.com/public/software/installers/smrtlink-release-sequel2_13.1.0.221970.zip



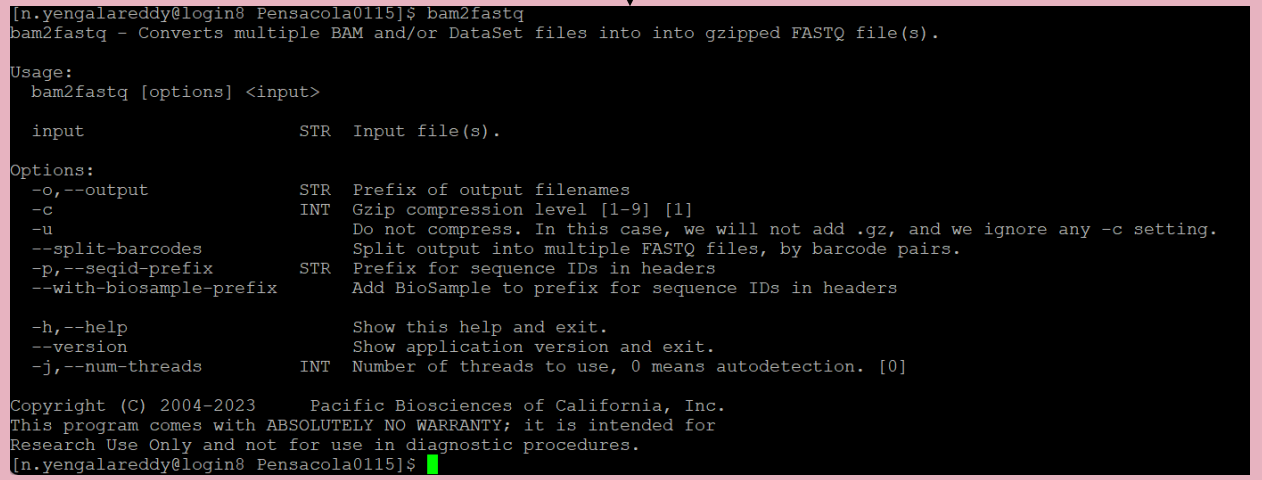
2) ./smrtlink-release-sequel2_13.1.0.221970.run –rootdir ./smrtlink_13 –smrttools-only

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

Note: Double check that smartlink tools works by entering

"bam2fastq" into the command line

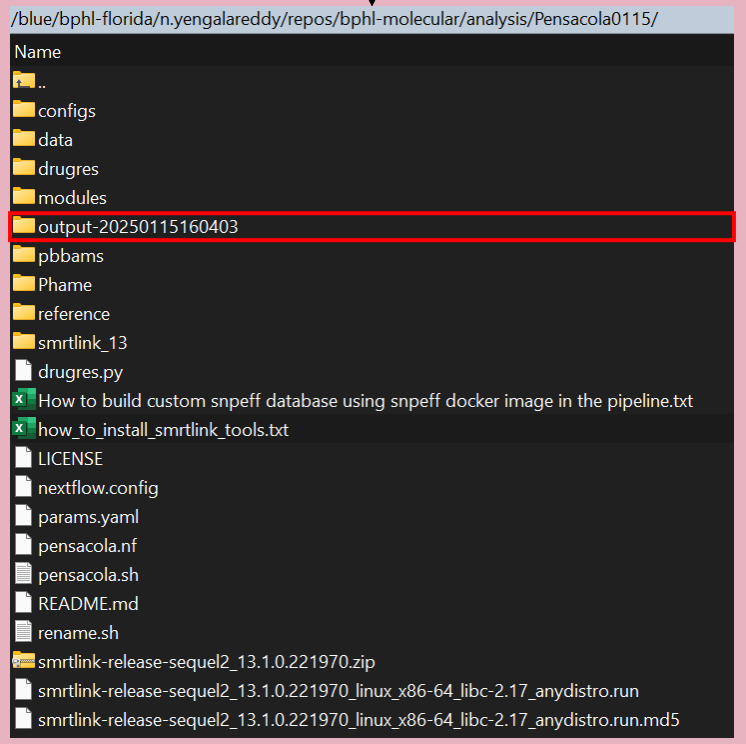


Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

activate conda environment containing Nextflow, Python3

sbatch pensacola.sh



Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Pensacola0115/

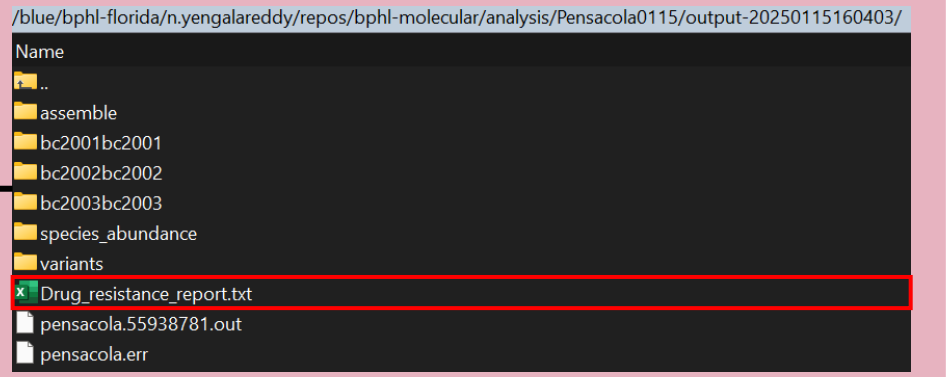
Name
- ..
- configs
- data
- drugres
- modules
- output-20250115160403
- pbbams
- Phame
- reference
- smrtlink_13
- drugres.py
- How to build custom snpeff database using snpeff docker image in the pipeline.txt
- how_to_install_smrtlink_tools.txt
- LICENSE
- nextflow.config
- params.yaml
- pensacola.nf
- pensacola.sh
- README.md
- rename.sh
- smrtlink-release-sequel2_13.1.0.221970.zip
- smrtlink-release-sequel2_13.1.0.221970_linux_x86-64_libc-2.17_anydistro.run
- smrtlink-release-sequel2_13.1.0.221970_linux_x86-64_libc-2.17_anydistro.run.md5

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Pensacola0115/output-20250115160403/

Name
- ..
- assemble
- bc2001bc2001
- bc2002bc2002
- bc2003bc2003
- species_abundance
- variants
- Drug_resistance_report.txt
- pensacola.55938781.out
- pensacola.err

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.



########## Detected drug resistance ##########

1. bc2001bc2001.variants_snpeff.ann.vcf:

ID,Drug,Gene,Mutation
B9J08_001448,FLC,ERG11,Y132F;K143R;Phe126Leu

B9J08_005576,FLC,,Cys110Phe

2. bc2003bc2003.variants_snpeff.ann.vcf:

ID,Drug,Gene,Mutation
B9J08_001448,FLC,ERG11,Y132F;K143R;Phe126Leu

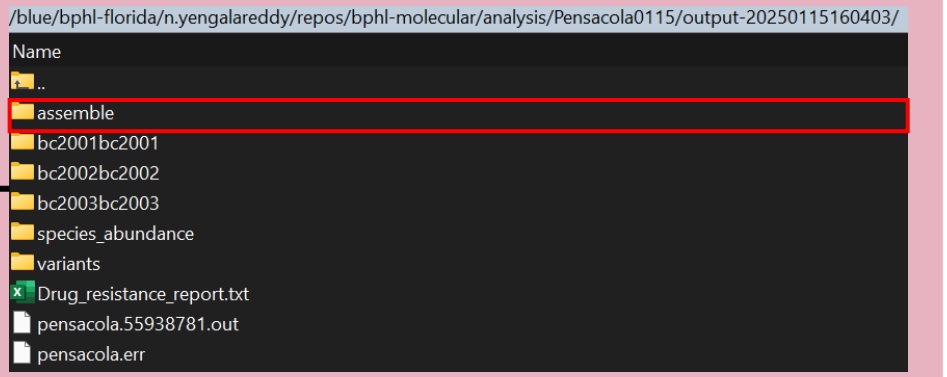
B9J08_005576,FLC,,Cys110Phe

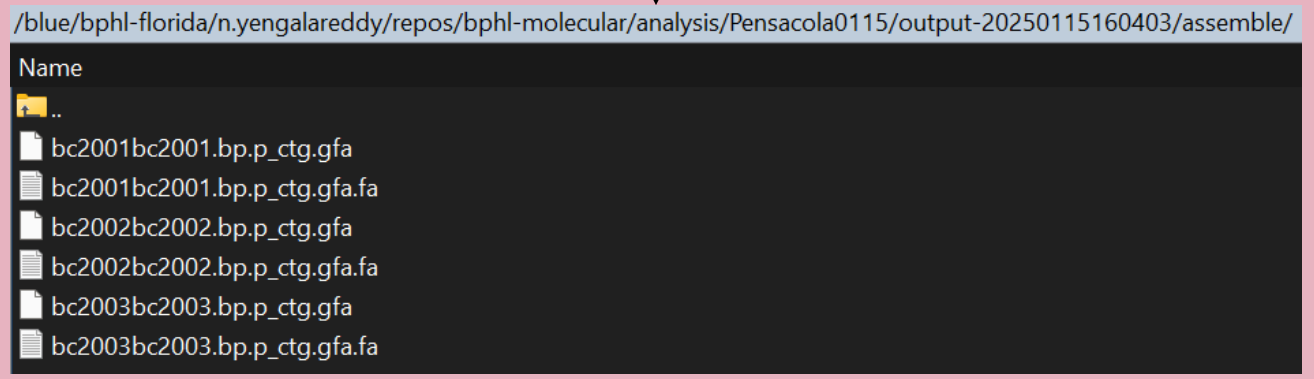
3. bc2002bc2002.variants_snpeff.ann.vcf:

ID,Drug,Gene,Mutation
B9J08_001448,FLC,ERG11,Y132F;K143R;Phe126Leu

B9J08_005576,FLC,,Cys110Phe

Application Cont.





Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Pensacola0115/output-20250115160403/

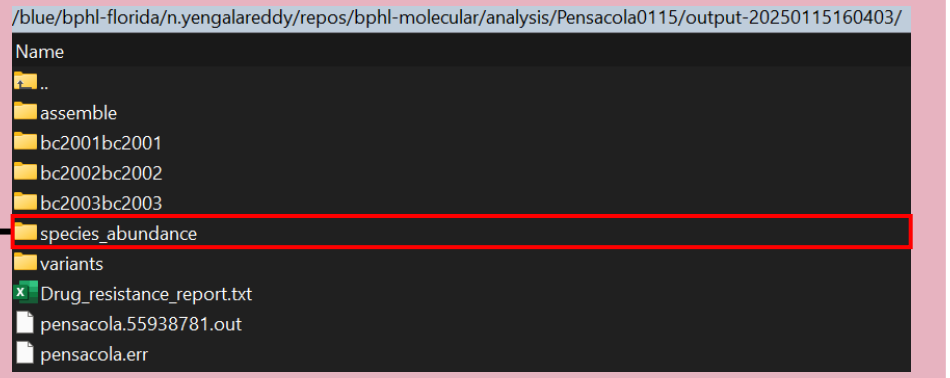
Name
- ..
- assemble
- bc2001bc2001
- bc2002bc2002
- bc2003bc2003
- species_abundance
- variants
- Drug_resistance_report.txt
- pensacola.55938781.out
- pensacola.err

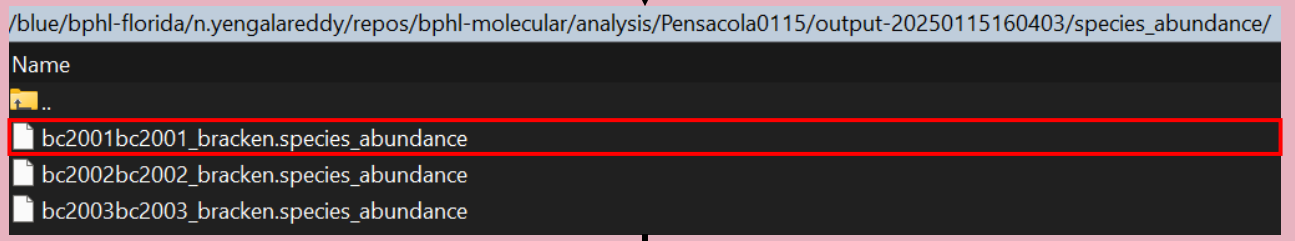
/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Pensacola0115/output-20250115160403/bc2001bc2001/

Name
- ..
- assemble
- kraken_out
- qc
- bc2001bc2001.bam
- bc2001bc2001.bam.pbi
- bc2001bc2001.coverage.txt
- bc2001bc2001.fastq.gz
- bc2001bc2001.mapped.bam
- bc2001bc2001.mapped.bam.bai
- bc2001bc2001.mapped.bam.depth
- bc2001bc2001.mapped.bam.mpileup
- bc2001bc2001.variants_bcftools.vcf
- bc2001bc2001.variants_snpeff.ann.vcf
- bc2001bc2001_snpeff.genes.txt
- bc2001bc2001_snpeff.html

#rname	startpos	endpos	numreads	covbases	coverage	meandepth	meanbaseq	meanmapq
PEKT02000001_C_auris_B8441	1	1083522	9352	1083188	99.9692	69.597	86.1	60
PEKT02000002_C_auris_B8441	1	1280737	13641	1279564	99.9084	85.9919	85.8	53.4
PEKT02000003_C_auris_B8441	1	1047474	11412	1033718	98.6867	86.1431	86	60
PEKT02000004_C_auris_B8441	1	887381	9538	886816	99.9363	86.9957	85.9	60
PEKT02000005_C_auris_B8441	1	639401	7825	638481	99.8561	94.9978	86.1	59.8
PEKT02000006_C_auris_B8441	1	776876	10316	776387	99.9371	107.672	86.1	60
PEKT02000007_C_auris_B8441	1	3195935	28493	3191857	99.8724	71.366	86.1	59.9
PEKT02000008_C_auris_B8441	1	898131	10741	894257	99.5687	95.8412	85.7	59.6
PEKT02000009_C_auris_B8441	1	1007143	12759	998219	99.1139	102.062	86	59.7
PEKT02000010_C_auris_B8441	1	1402902	15384	1399198	99.736	88.2547	86	60
PEKT02000012_C_auris_B8441	1	65067	978	63688	97.8806	110.579	85.9	59.8
PEKT02000013_C_auris_B8441	1	38216	931	38159	99.8508	166.528	85.9	59.8
PEKT02000014_C_auris_B8441	1	11792	159	11765	99.771	73.7636	85.3	60
PEKT02000011_C_auris_B8441	1	20765	0	0	0	0	0	0
PEKT02000015_C_auris_B8441	1	10617	0	0	0	0	0	0

Application Cont.





name	taxonomy_id	taxonomy_lvl	kraken_assigned_reads	added_reads	new_est_reads	fraction_total_reads
[Candida] auris	498019	S	127942	68	128010	0.94999
Debaryomyces hansenii	4959	S	4961	15	4976	0.03693
Brettanomyces nanus	13502	S	40	1	41	0.0003
Brettanomyces bruxellensis	5007	S	1	0	1	0.00001
Aspergillus nidulans	162425	S	3	0	3	0.00002
Homo sapiens	9606	S	1681	9	1690	0.01254
Chryseobacterium aureum	2497456	S	13	7	20	0.00015
Ferrimonas sp. YFM	3028878	S	1	1	2	0.00001
Finegoldia magna	1260	S	1	2	3	0.00002

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Pensacola0115/output-20250115160403/

Name
- ..
- assemble
- bc2001bc2001
- bc2002bc2002
- bc2003bc2003
- species_abundance
- variants
- Drug_resistance_report.txt
- pensacola.55938781.out
- pensacola.err

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Pensacola0115/output-20250115160403/variants/

Name
- ..
- bc2001bc2001.variants_bcftools.vcf
- bc2001bc2001.variants_snpeff.ann.vcf
- bc2001bc2001_snpeff.genes.txt
- bc2001bc2001_snpeff.html
- bc2002bc2002.variants_bcftools.vcf
- bc2002bc2002.variants_snpeff.ann.vcf
- bc2002bc2002_snpeff.genes.txt
- bc2002bc2002_snpeff.html
- bc2003bc2003.variants_bcftools.vcf
- bc2003bc2003.variants_snpeff.ann.vcf
- bc2003bc2003_snpeff.genes.txt
- bc2003bc2003_snpeff.html

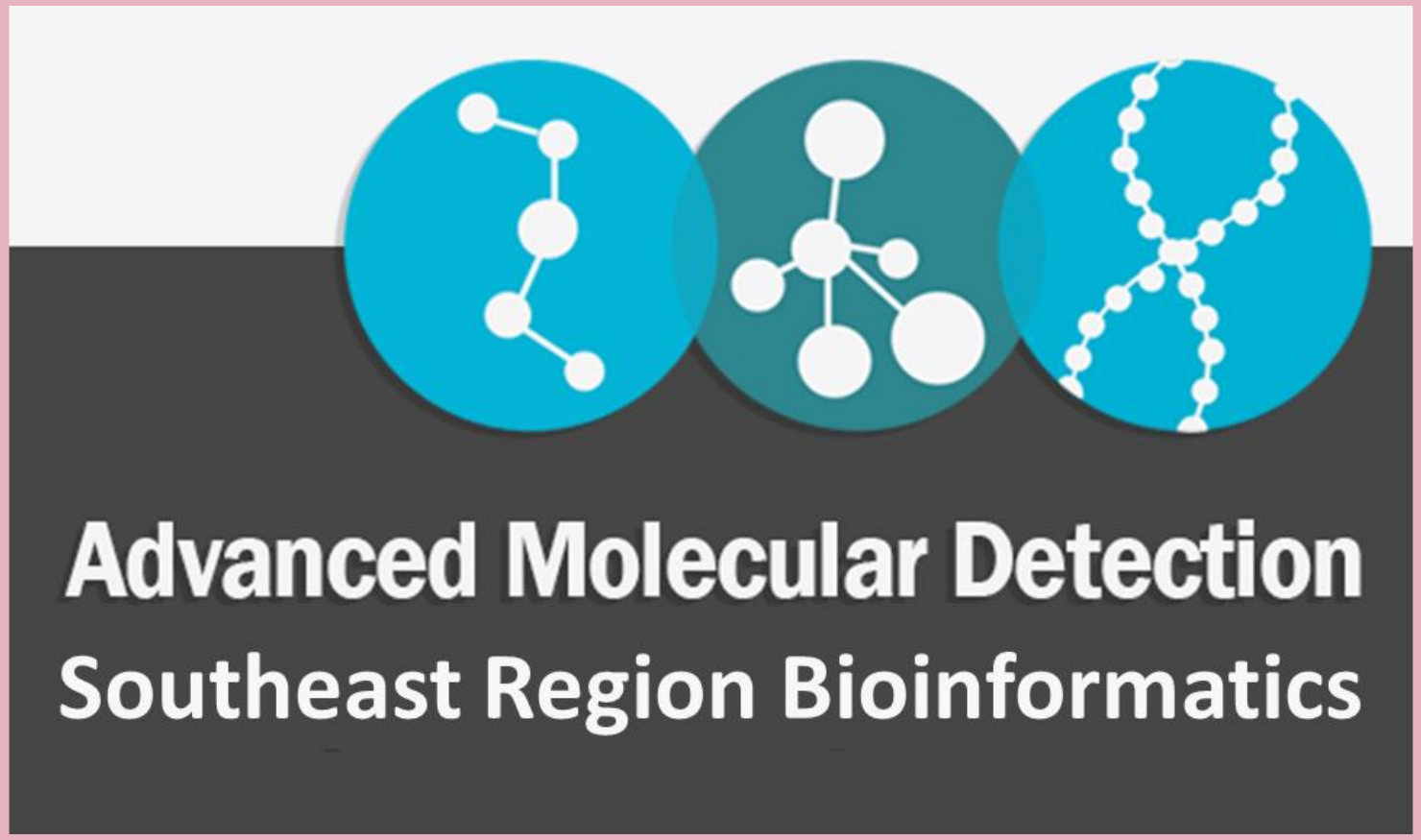
The following table is formatted as tab separated values.

#GeneName	GeneId	TranscriptId	BioType	variants_impact_HIGH	variants_impact_LOW	variants_impact_MODERATE	variants_impact_MODIFIER	variants_effect_conservative_inframe_deletion	variants_effect_conservative_inframe_insertion	variants_effect_disruptive_inframe_deletion	variants_effect_disruptive_inframe_insertion	variants_effect_downstream_gene_variant	variants_effect_frameshift_variant	variants_effect_intron_variant
B9J08_000001	B9J08_000001	B9J08_000001-T	protein_coding	0	1	0	4	0	0	0	0	2	0	0
B9J08_000002	B9J08_000002	B9J08_000002-T	protein_coding	0	0	0	4	0	0	0	0	2	0	0
B9J08_000003	B9J08_000003	B9J08_000003-T	protein_coding	0	0	0	5	0	0	0	0	2	0	0
B9J08_000004	B9J08_000004	B9J08_000004-T	protein_coding	0	0	0	7	0	0	0	0	5	0	0
B9J08_000005	B9J08_000005	B9J08_000005-T	protein_coding	0	0	1	21	0	0	0	0	19	0	0
B9J08_000006	B9J08_000006	B9J08_000006-T	protein_coding	0	0	1	24	0	0	0	0	23	0	0
B9J08_000007	B9J08_000007	B9J08_000007-T	protein_coding	0	1	0	39	0	0	0	0	3	0	0
B9J08_000008	B9J08_000008	B9J08_000008-T	protein_coding	0	11	7	57	0	0	0	0	8	0	0
B9J08_000009	B9J08_000009	B9J08_000009-T	protein_coding	0	10	2	62	0	0	0	0	26	0	0
B9J08_000010	B9J08_000010	B9J08_000010-T	protein_coding	0	0	0	64	0	0	0	0	10	0	0
B9J08_000011	B9J08_000011	B9J08_000011-T	protein_coding	0	3	4	56	0	0	0	0	15	0	0
B9J08_000012	B9J08_000012	B9J08_000012-T	protein_coding	0	2	0	22	0	0	0	0	8	0	0
B9J08_000013	B9J08_000013	B9J08_000013-T	protein_coding	0	1	1	25	0	0	0	0	10	0	0
B9J08_000014	B9J08_000014	B9J08_000014-T	protein_coding	0	0	1	24	0	0	0	0	9	0	0
B9J08_000015	B9J08_000015	B9J08_000015-T	protein_coding	0	4	1	22	0	0	0	0	10	0	0
B9J08_000016	B9J08_000016	B9J08_000016-T	protein_coding	0	4	0	24	0	0	0	0	8	0	0
B9J08_000017	B9J08_000017	B9J08_000017-T	protein_coding	0	1	4	27	0	0	0	0	12	0	0
B9J08_000018	B9J08_000018	B9J08_000018-T	protein_coding	0	4	0	100	0	0	0	0	82	0	0
B9J08_000019	B9J08_000019	B9J08_000019-T	protein_coding	0	21	10	57	0	0	0	0	46	0	0
B9J08_000020	B9J08_000020	B9J08_000020-T	protein_coding	0	1	0	90	0	0	0	0	9	0	0
B9J08_000021	B9J08_000021	B9J08_000021-T	protein_coding	0	0	1	80	0	0	0	0	12	0	0
B9J08_000022	B9J08_000022	B9J08_000022-T	protein_coding	0	0	1	22	0	0	0	0	14	0	0
B9J08_000023	B9J08_000023	B9J08_000023-T	protein_coding	0	5	2	22	0	0	0	0	5	0	0
B9J08_000024	B9J08_000024	B9J08_000024-T	protein_coding	0	0	0	31	0	0	0	0	12	0	0
B9J08_000025	B9J08_000025	B9J08_000025-T	protein_coding	0	0	2	28	0	0	0	0	14	0	0
B9J08_000026	B9J08_000026	B9J08_000026-T	protein_coding	0	2	1	30	0	0	0	0	16	0	0

Advanced Molecular Detection
Southeast Region Bioinformatics

Conclusion

- Fundamentals of Pensacola
- Installation and setup of Pensacola in HPG
- Successfully executed job query for Pensacola
- Generated output files



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