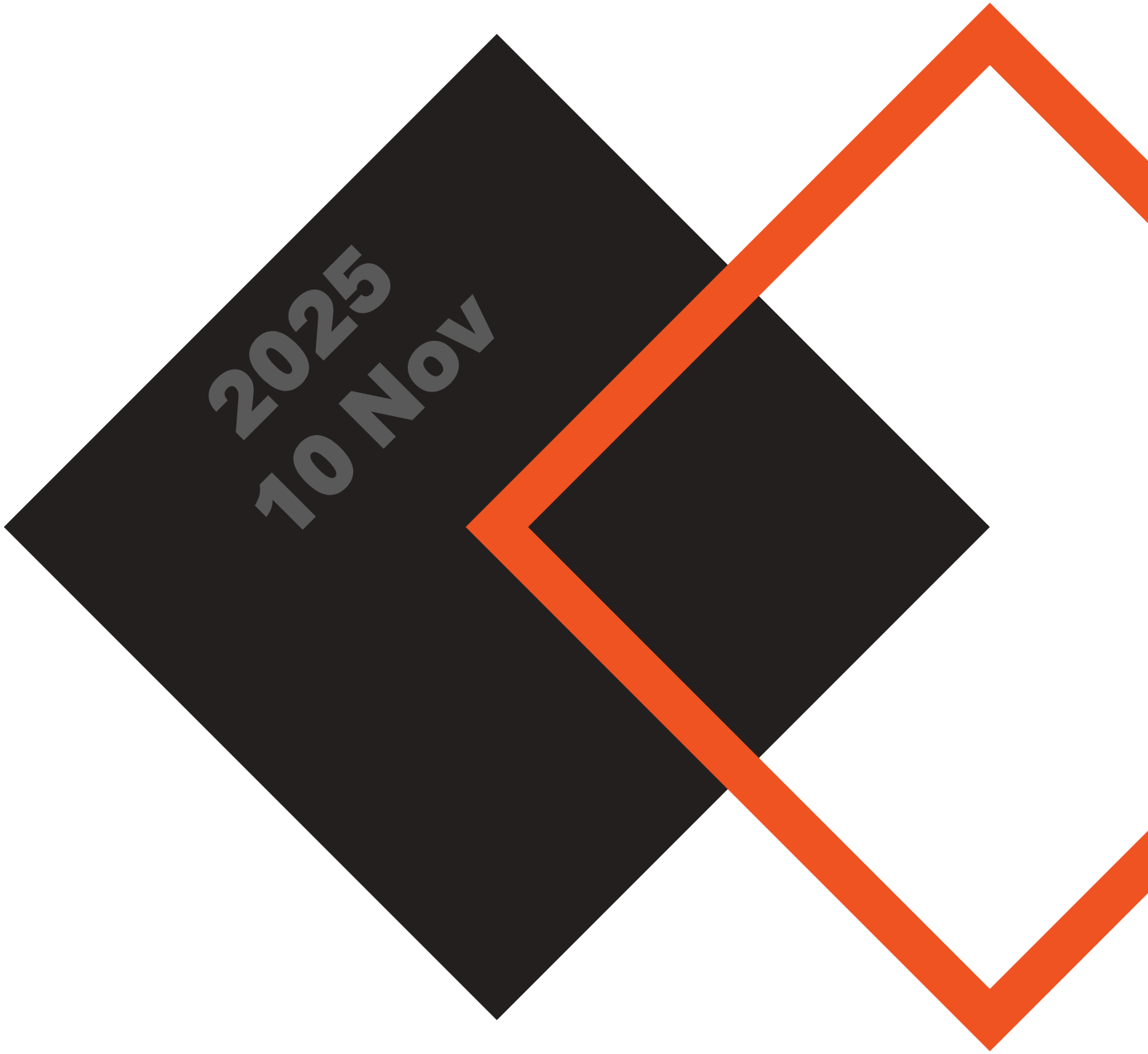
VARPIPE

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.

Updates

Nov 28, 2025-

- 2026 Needs Assessment Surveys

Varpipe
Overview

Purpose:

- This nextflow pipeline is developed for *Mycobacterium tuberculosis* (extendable to other references). Performs raw fastq QC, read alignment to a reference, variant calling and produces per sample output folder.

Usage:

- Can be used by public health labs and researchers to process WGS samples for surveillance, outbreak investigation and bacterial typing.

Dependencies:

- Apptainer/Singularity
- Reference genome FASTA file
- Java
- Nextflow

Advanced Molecular Detection
Southeast Region Bioinformatics

TB Extension

Overview

Purpose:

- This pipeline directory provides extensions to varpipe for *Mycobacterium tuberculosis* analysis. It performs QC summarization, contamination checking, and drug resistance and lineage profiling using TBProfiler.

Usage:

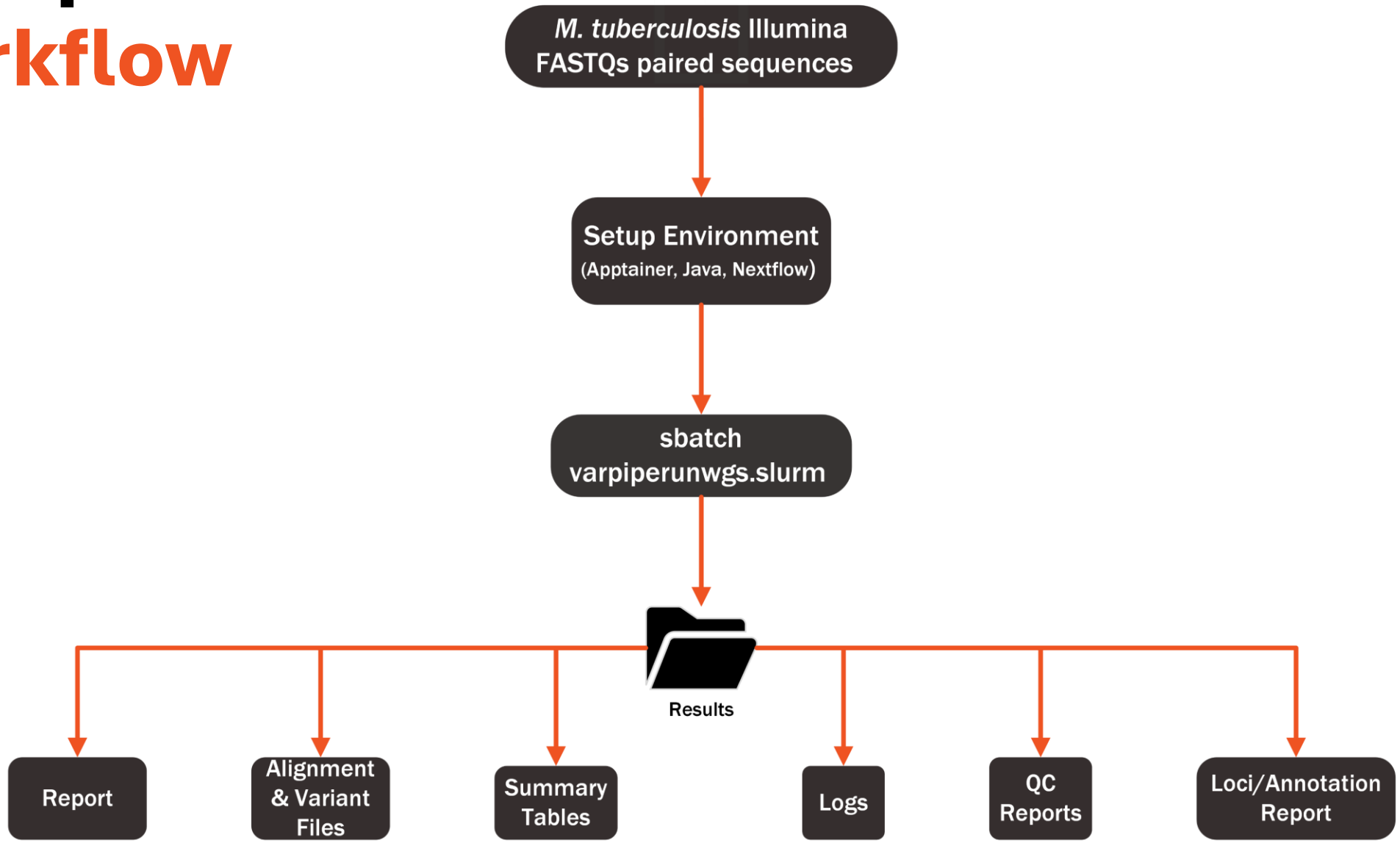
- Can be used by public health labs for *M. tuberculosis* WGS surveillance, outbreak investigation and molecular typing.

Dependencies:

- Python
- Nextflow
- Conda

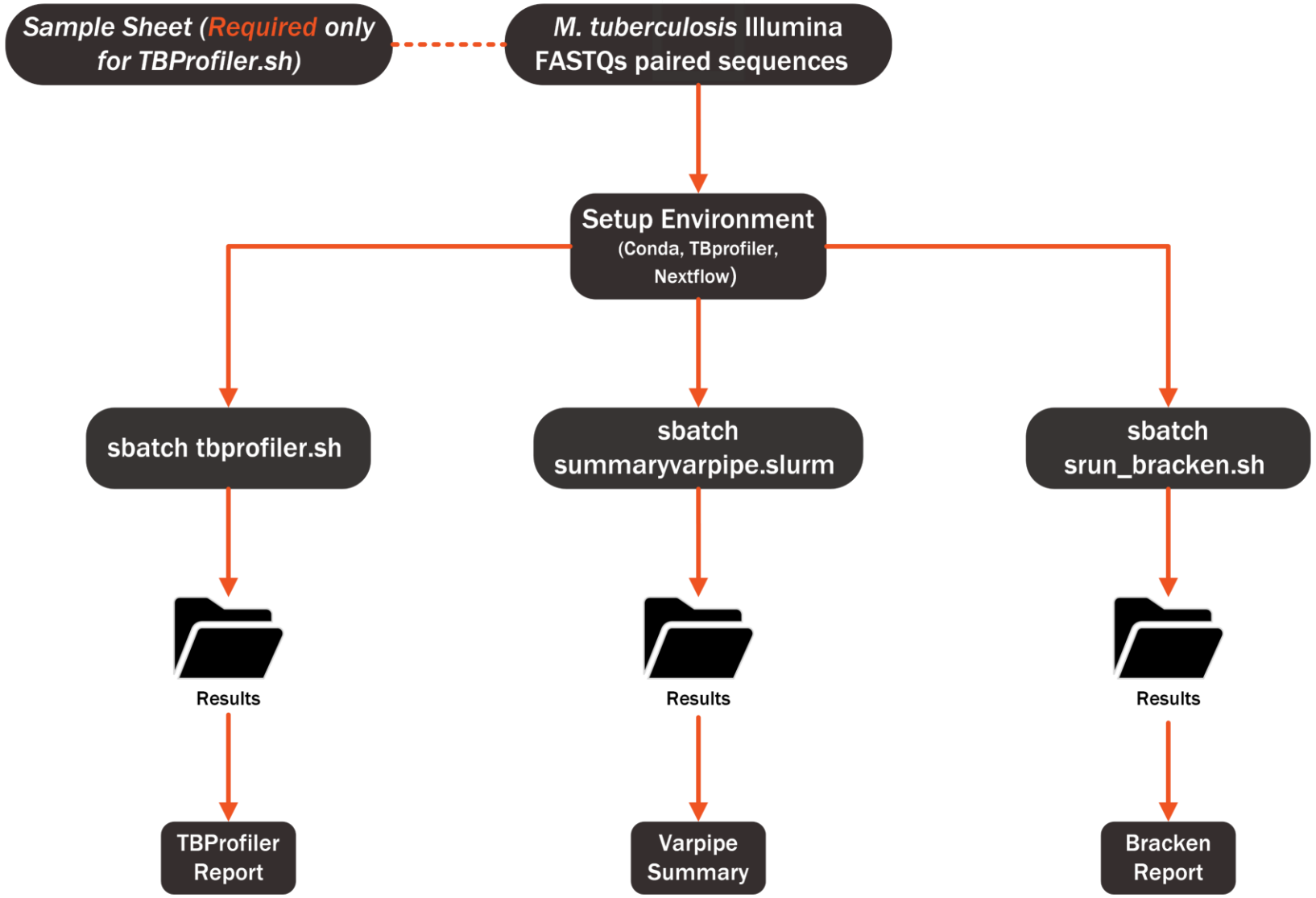
Advanced Molecular Detection
Southeast Region Bioinformatics

Varpipe
Workflow



Advanced Molecular Detection
Southeast Region Bioinformatics

TB Extension
Workflow



Advanced Molecular Detection
Southeast Region Bioinformatics

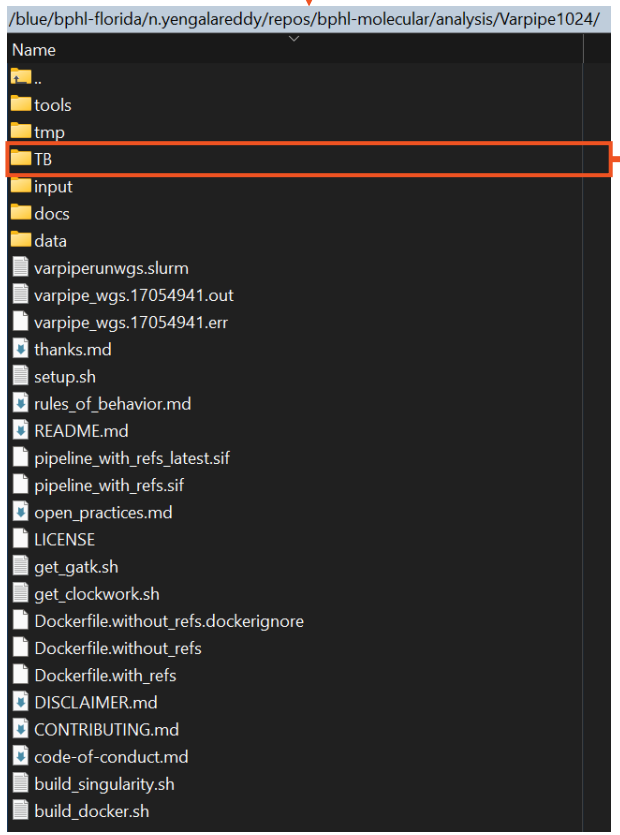
Application

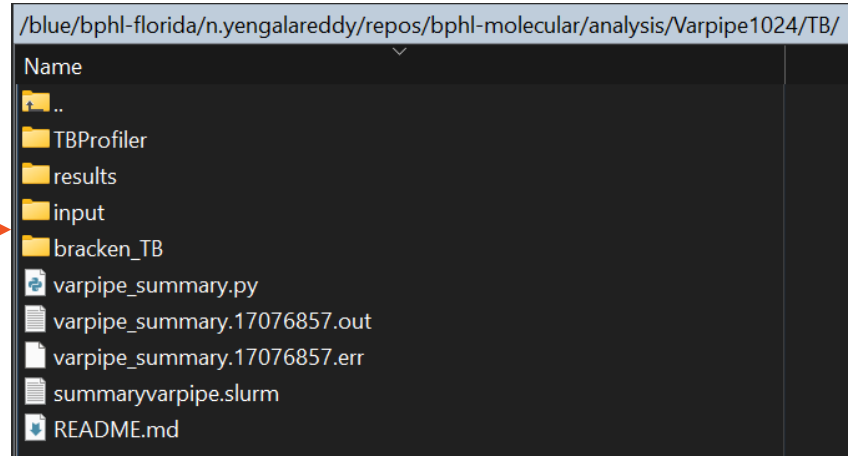
Use M. *tuberculosis* Illumina FASTQs paired sequences and use Varpipe and TB Extension to analyze the samples.

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

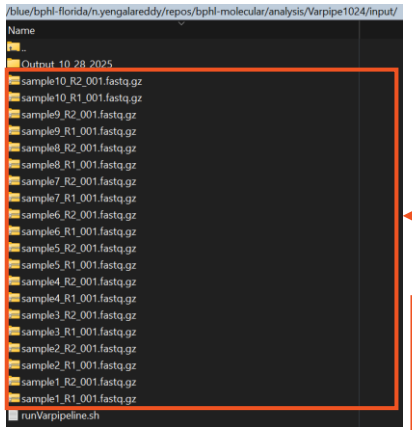
```
cd /blue/bphl-<state>/<user>/repos/bphl-molecular/

git clone https://github.com/CDCgov/NCHHSTP-DTBE-Varpipe-WGS

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

mkdir analysis/

cd analysis/

cp /blue/bphl-<state>/<user>/repos/bphl-molecular/NCHHSTP-DTBE-Varpipe-WGS/

cd NCHHSTP-DTBE-Varpipe-WGS/

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

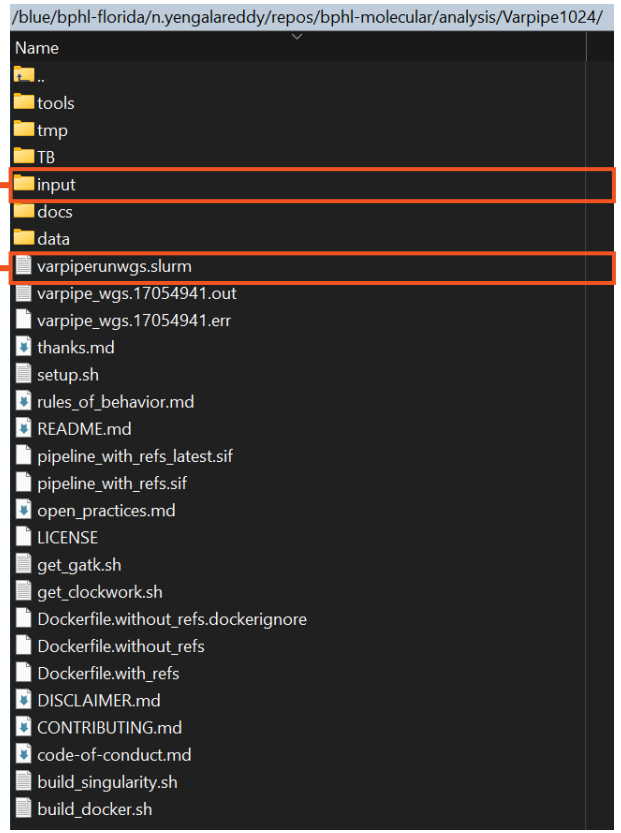




Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.






```
#!/bin/bash
#SBATCH --account=bphl-umbrella
#SBATCH --qos=bphl-umbrella
#SBATCH --job-name=varpipe_wgs
#SBATCH --mail-type=END,FAIL
#SBATCH --mail-user=nikhil.yengala@flhealth.gov
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=24
#SBATCH --mem=128G
#SBATCH --time=24:00:00
#SBATCH --output=varpipe_wgs.%j.out
#SBATCH --error=varpipe_wgs.%j.err

#Load packages
module purge
module load apptainer

#Pathways
BASE="/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Varpipe1024"
INPUT_DIR="${BASE}/input"
IMAGE="${BASE}/pipeline_with_refs_latest.sif"
THREADS="${SLURM_CPUS_PER_TASK}"

#Export Cache Directory
export APPTAINER_TMPDIR="${BASE}/.tmp"
export APPTAINER_CACHEDIR="${BASE}/.cache"
mkdir -p "${APPTAINER_TMPDIR}" "${APPTAINER_CACHEDIR}"

#Print statements for verification
echo "[info] Using image:   ${IMAGE}"
echo "[info] Input folder:   ${INPUT_DIR}"
echo "[info] Threads:        ${THREADS}"

#Run Code
apptainer exec \
  --bind "${INPUT_DIR}:/varpipe_wgs/data" \
  "${IMAGE}" \
  bash -lc '
cd /varpipe_wgs/data

echo "[info] Completed Varpipe_wgs run. Please check output directory in ${INPUT_DIR}"
```

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Varpipe1024/input/

Name

Output_10_28_2025
sample10_R2_001.fastq.gz
sample10_R1_001.fastq.gz
sample9_R2_001.fastq.gz
sample9_R1_001.fastq.gz
sample8_R2_001.fastq.gz
sample8_R1_001.fastq.gz
sample7_R2_001.fastq.gz
sample7_R1_001.fastq.gz
sample6_R2_001.fastq.gz
sample6_R1_001.fastq.gz
sample5_R2_001.fastq.gz
sample5_R1_001.fastq.gz
sample4_R2_001.fastq.gz
sample4_R1_001.fastq.gz
sample3_R2_001.fastq.gz
sample3_R1_001.fastq.gz
sample2_R2_001.fastq.gz
sample2_R1_001.fastq.gz
sample1_R2_001.fastq.gz
sample1_R1_001.fastq.gz
runVarpipeline.sh

sbatch `varpiperunwgs.slurm`

/.../n.yengalareddy/repos/bphl-molecular/analysis/Varpipe1024/input/Output_10_28_2025/

Name

..
sample10
sample9
sample8
sample7
sample6
sample5
sample3
sample1
QC

/.../n.yengalareddy/repos/bphl-molecular/analysis/Varpipe1024/input/Output_10_28_2025/sample1/

Name

..
sample1_target_region_coverage.txt
sample1_summary.txt
sample1_structural_variants.txt
sample1_stats.txt
sample1_report.pdf
sample1_Lineage.txt
sample1_interpretation.txt
sample1_genome_region_coverage.txt
sample1_full_raw_annotation.txt
sample1_full_Final_annotation.txt
sample1_full_annotation.txt
sample1_DR_loci_raw_annotation.txt
sample1_DR_loci_Final_annotation.txt
sample1_DR_loci_annotation.txt
sample1.log
sample1.lineage_report.txt

/.../n.yengalareddy/repos/bphl-molecular/analysis/Varpipe1024/input/Output_10_28_2025/QC/

Name

..
sample4
sample2
QC.log

/.../repos/bphl-molecular/analysis/Varpipe1024/input/Output_10_28_2025/QC/sample2/*.*

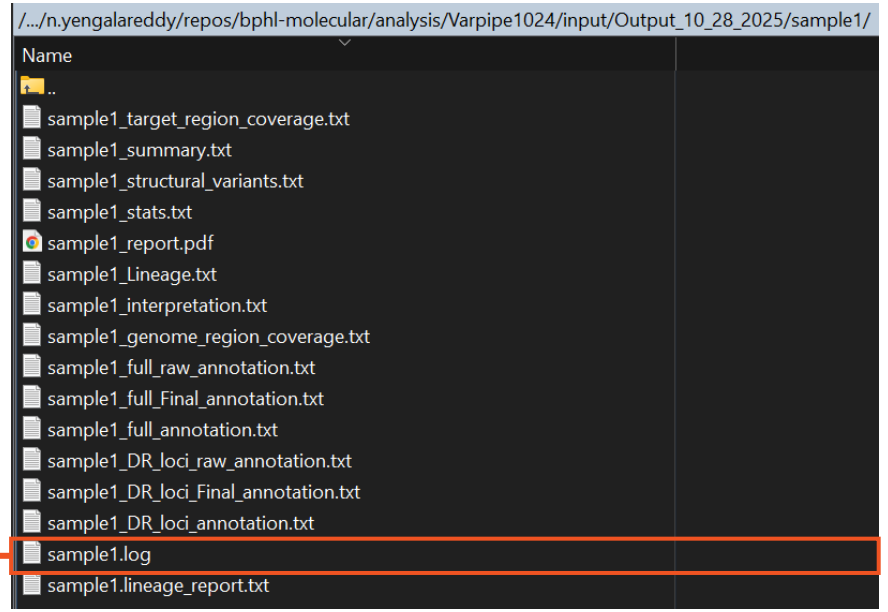
Name

..
sample2_target_region_coverage.txt
sample2_structural_variants.txt
sample2_stats.txt
sample2_sdrcsm.bai
sample2_genome_region_coverage.txt
sample2.log

Advanced Molecular Detection
Southeast Region Bioinformatics

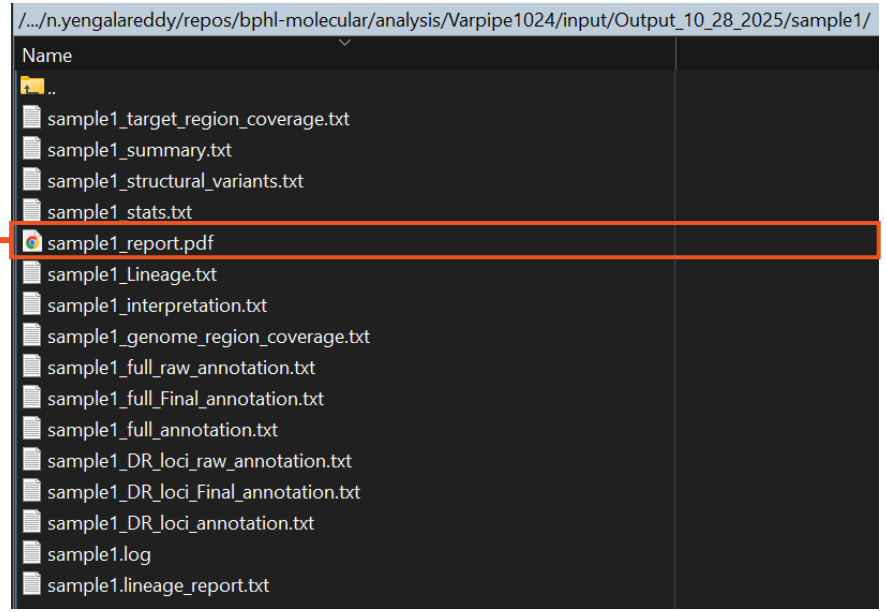
Application Cont.

```
../tools/Varpipeline -q sample1_R1_001.fastq.gz -q2 sample1_R2_001.fastq.gz -r ../tools/ref2.fa -n sample1 -a -v -t 24

---[ nextflow remove contamination ]---
Command:
/varpipe_wgs/tools/clockwork-0.11.3/results/nextflow run /varpipe_wgs/tools/clockwork-0.11.3/nextflow/remove_contam.nf --ref_fasta /varpipe_wgs/tools/clockwork-0.11.3/OUT/ref.fa --ref_metadata_tsv
/varpipe_wgs/tools/clockwork-0.11.3/OUT/remove_contam_metadata.tsv --reads_in1 sample1_R1_001.fastq.gz --reads_in2 sample1_R2_001.fastq.gz --outprefix Output_10_28_2025/sample1/clockwork/sample1 --
mapping_threads 4

Standard Output:
N E X T F L O W  ~  version 20.07.1
Launching `/varpipe_wgs/tools/clockwork-0.11.3/nextflow/remove_contam.nf` [cheeky_fourier] - revision: b5b8e4e89e
[-        ] process > make_jobs_tsv      -
[-        ] process > map_reads          -
[-        ] process > sam_to_fastq_files -

executor >  local (1)
[d3/3403da] process > make_jobs_tsv      [  0%] 0 of 1
[-        ] process > map_reads          -
[-        ] process > sam_to_fastq_files -

executor >  local (2)
[d3/3403da] process > make_jobs_tsv      [100%] 1 of 1 ✓
[3d/81cd6e] process > map_reads (1)      [  0%] 0 of 1
[-        ] process > sam_to_fastq_files -

executor >  local (3)
[d3/3403da] process > make_jobs_tsv          [100%] 1 of 1 ✓
[3d/81cd6e] process > map_reads (1)          [100%] 1 of 1 ✓
[c3/8b999d] process > sam_to_fastq_files (1) [  0%] 0 of 1

executor >  local (3)
[d3/3403da] process > make_jobs_tsv          [100%] 1 of 1 ✓
[3d/81cd6e] process > map_reads (1)          [100%] 1 of 1 ✓
[c3/8b999d] process > sam_to_fastq_files (1) [100%] 1 of 1 ✓
Completed at: 28-Oct-2025 19:25:06
Duration    : 3m 9s
CPU hours   : 0.1
Succeeded   : 3
```



Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.



Sample ID: sample1
Sample Name: sample1
Coverage Drop: 0
Pipeline Version: Varpipeline: Varpipe_wgs_1.0.2
Date: 2025/10/28 15:27:47

Target Coverage Summary:

Gene Name	Start	End	Flag
gyrB	6571	6762	No deletion
gyrA	7360	7583	No deletion
rpoB	760307	761286	No deletion
mmpR	778989	779487	No deletion
rplC1	800808	801462	No deletion
atpE	1461044	1461290	No deletion
rrs	1473245	1473331	No deletion
rrl	1473657	1476796	No deletion
fabG1	1673409	1674052	No deletion
inhA	1674201	1675012	No deletion
tlyA	1917939	1918747	No deletion
katG	2153888	2156112	No deletion
pncA	2288676	2289272	No deletion
eis	2714123	2715372	No deletion
ahpC	2726093	2726194	No deletion
pepQ	2859299	2860418	No deletion
embB	4246586	4249653	No deletion
ethA	4326003	4327474	No deletion

Application Cont.

/.../n.yengalareddy/repos/bphl-molecular/analysis/Varpipe1024/input/Output_10_28_2025/sample1/

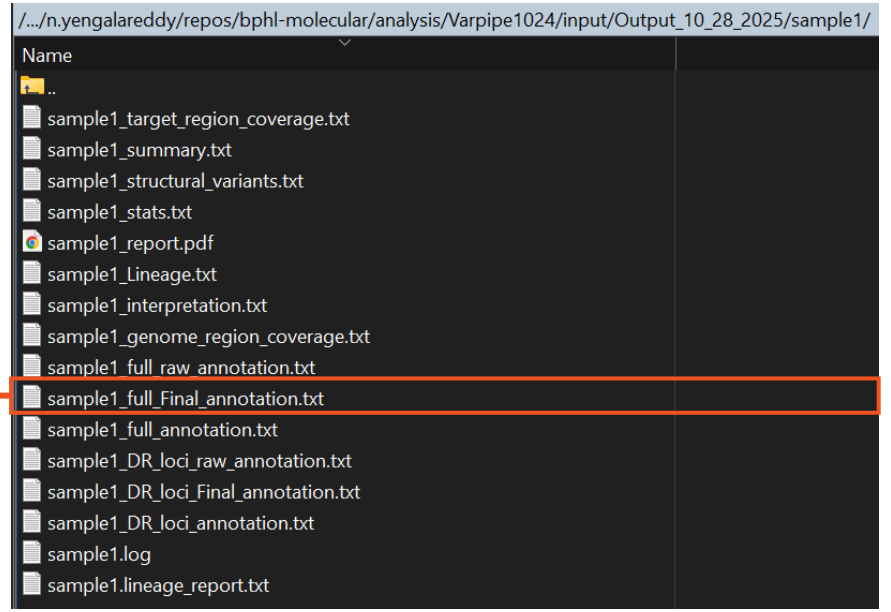
Name

..
sample1_target_region_coverage.txt
sample1_summary.txt
sample1_structural_variants.txt
sample1_stats.txt
sample1_report.pdf
sample1_Lineage.txt
sample1_interpretation.txt
sample1_genome_region_coverage.txt
sample1_full_raw_annotation.txt
sample1_full_Final_annotation.txt
sample1_full_annotation.txt
sample1_DR_loci_raw_annotation.txt
sample1_DR_loci_Final_annotation.txt
sample1_DR_loci_annotation.txt
sample1.log
sample1.lineage_report.txt

```
Sample Summary:
Sample ID:      sample1
Sample Name:    sample1
Coverage Drop:  0
Pipeline Version:       Varpipeline: Varpipe_wgs_1.0.2
Date:   2025/10/28 15:27:47


Target Coverage Summary:
Gene Name       Start   End     Flag
gyrB    6571    6762    No deletion
gyrA    7360    7583    No deletion
rpoB    760307  761286  No deletion
mmpR    778989  779487  No deletion
rplC1   800808  801462  No deletion
atpE    1461044 1461290 No deletion
rrs     1473245 1473331 No deletion
rrl     1473657 1476796 No deletion
fabG1   1673409 1674052 No deletion
inhA    1674201 1675012 No deletion
tlyA    1917939 1918747 No deletion
katG    2153888 2156112 No deletion
pncA    2288676 2289272 No deletion
eis     2714123 2715372 No deletion
ahpC    2726093 2726194 No deletion
pepQ    2859299 2860418 No deletion
embB    4246586 4249653 No deletion
ethA    4326003 4327474 No deletion


Variant Summary:
POS     Gene Name       Nucleotide Change       Amino acid Change       Read Depth      Percent Alt Allele      Annotation
761152  rpoB    c.1346T>A       p.Leu449Gln     70      5.10    Non-synonymous


Interpretations Summary:
Drug    Variant Interpretation
INH     No reportable variant detected  INH-S
RIF     rpoB_p.Leu449Gln        RIF-S
PZA     No reportable variant detected  PZA-S
FQ      No reportable variant detected  FQ-S
EMB     No reportable variant detected  EMB-S
```

Advanced Molecular Detection
Southeast Region Bioinformatics

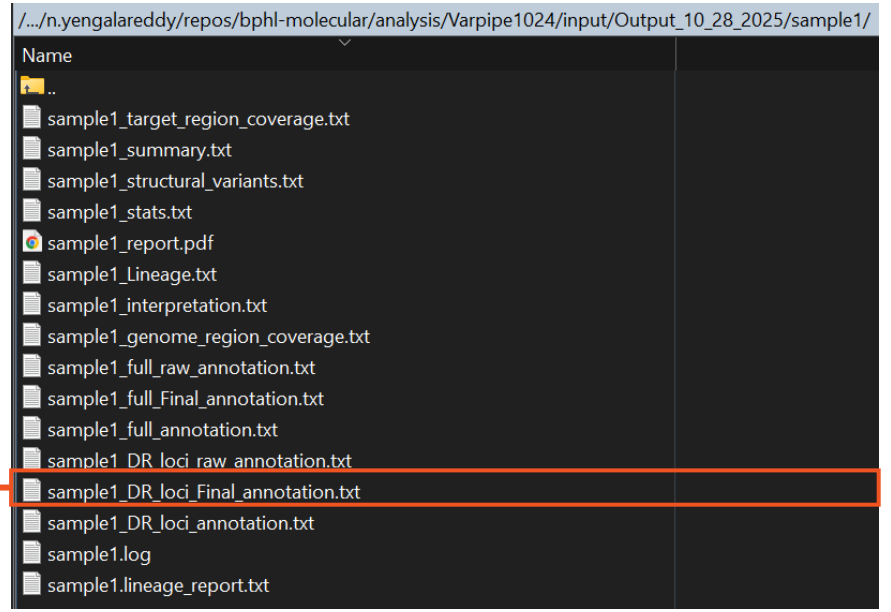
Application Cont.



Sample ID	CHROM	POS	REF	ALT	Read Depth	Percent Alt Allele	Annotation	Variant Type	Nucleotide Change	Position within CDS	Amino acid Change	REF Amino acid	ALT Amino acid	Codon Position	Gene Name	Gene ID
sample1	NC_000962	14785	T	C	128	99.1	Non-synonymous	SNP	c.697T>C	697	p.Cys233Arg	Cys	Arg	233	Rv0012	Rv0012
sample1	NC_000962	23137	A	G	59	5.9	Synonymous	SNP	c.45T>C	45	p.Asp15Asp	Asp	Asp	15	pstP	Rv0018c
sample1	NC_000962	32725	A	T	103	5	Synonymous	SNP	c.669A>T	669	p.Arg223Arg	Arg	Arg	223	Rv0029	Rv0029
sample1	NC_000962	42518	G	GGGACTA	71	8.9	Non-synonymous	Deletion	c.847_848insTAGTCC	847-848	p.Ala283delinsValValPro	Ala	NA	283	mtc28	Rv0040c
sample1	NC_000962	55533	TGCC	T	48	97.7	Non-synonymous	Deletion	c.1891_1893delCCG	1891-1893	p.Pro631del	Pro	NA	631	ponA1	Rv0050
sample1	NC_000962	55553	C	T	48	97.7	Non-synonymous	SNP	c.1891C>T	1891	p.Pro631Ser	Pro	Ser	631	ponA1	Rv0050
sample1	NC_000962	69989	G	A	81	98.6	Non-synonymous	SNP	c.1370G>A	1370	p.Gly457Asp	Gly	Asp	457	Rv0064	Rv0064
sample1	NC_000962	91967	GGCGCCC	G	46	5.2	Non-synonymous	Deletion	c.1570_1575delCGCCCG	1570-1575	p.Arg524_Pro525del	Arg	NA	524-525	Rv0083	Rv0083
sample1	NC_000962	94910	C	CAAT	48	6.8	Non-synonymous	Insertion	c.960_961insAAT	960-961	p.Ala320_Ala321insAsn	Ala	NA	320-321	hycQ	Rv0086
sample1	NC_000962	94911	G	GT	48	6.8	Non-synonymous	Insertion	c.961_962insT	961-962	p.Ala321fs	Ala	NA	321	hycQ	Rv0086
sample1	NC_000962	94913	G	GT	48	6.8	Non-synonymous	Insertion	c.963_964insT	963-964	p.Ala322fs	Ala	NA	322	hycQ	Rv0086
sample1	NC_000962	116000	T	G	82	98.4	Synonymous	SNP	c.6000T>G	6000	p.Val2000Val	Val	Val	2000	nrp	Rv0101
sample1	NC_000962	131174	T	TG	64	98.3	Non-Coding	Insertion	c.-71_-70insC	NA	NA	NA	NA	NA	Rv0108c upstream	Rv0108c upstream
sample1	NC_000962	132417	C	G	5	83.3	Non-synonymous	SNP	c.1036C>G	1036	p.Arg346Gly	Arg	Gly	346	PE_PGRS1	Rv0109
sample1	NC_000962	183862	A	T	72	5.2	Non-synonymous	SNP	c.241A>T	241	p.Thr81Ser	Thr	Ser	81	pntAa	Rv0155
sample1	NC_000962	234477	T	G	47	97.2	Non-synonymous	SNP	c.2247T>G	2247	p.Tyr749*	Tyr	NA	749	Rv0197	Rv0197
sample1	NC_000962	234496	C	CGT	46	97.2	Non-synonymous	Insertion	c.2266_2267insGT	2266-2267	p.Pro756fs	Pro	NA	756	Rv0197	Rv0197
sample1	NC_000962	236275	T	A	37	14.3	Non-synonymous	SNP	c.233A>T	233	p.Gln78Leu	Gln	Leu	78	zmp1	Rv0198c
sample1	NC_000962	239150	AG	CT	54	6.4	Non-synonymous	MNP	c.2142_2143delCTinsAG	2142	p.AspPhe714GluVal	AspPhe	GluVal	714-715	mmpL11	Rv0202c
sample1	NC_000962	267552	CT	AG	50	7.3	Non-synonymous	MNP	c.212_213delAGinsCT	212	p.Glu71Ala	Glu	Ala	71	Rv0223c	Rv0223c
sample1	NC_000962	268377	A	G	64	5.4	Non-synonymous	SNP	c.251T>C	251	p.Ile84Thr	Ile	Thr	84	Rv0224c	Rv0224c
sample1	NC_000962	268381	AGCCG	CGGCT	68	5.2	Non-synonymous	MNP	c.243_247delCGGCTinsAGCCG	243	p.GlyTyr82AlaAsp	GlyTyr	AlaAsp	82-83	Rv0224c	Rv0224c
sample1	NC_000962	268389	C	T	66	5.4	Non-synonymous	SNP	c.239G>A	239	p.Gly80Asp	Gly	Asp	80	Rv0224c	Rv0224c
sample1	NC_000962	283714	C	G	52	7.7	Synonymous	SNP	c.3138G>C	3138	p.Thr1046Thr	Thr	Thr	1046	aftD	Rv0236c
sample1	NC_000962	323550	TCTTGG	CCAAGA	58	5.7	Non-synonymous	MNP	c.977_982delCCAAGAinsTCTTGG	977	p.AlaLysThr326ValLeuAla	AlaLysThr	ValLeuAla	326-327	Rv0269c	Rv0269c
sample1	NC_000962	333548	GTGCCGCCGGCACCGGGCGTGGCCCCGGCTT	G	39	7	Non-synonymous	Deletion	c.2733_2762delAAGCCGGGGCCACGCCCGGTGC	2733-2762	p.Gln911_Arg920del	Gln	NA	911-920	Rv0278c	Rv0278c
sample1	NC_000962	333637	A	G	26	15	Non-synonymous	SNP	c.2674T>C	2674	p.Trp892Arg	Trp	Arg	892	Rv0278c	Rv0278c
sample1	NC_000962	333640	GC	AT	26	15	Non-synonymous	MNP	c.2670_2671delGCinsAT	2670	p.Arg891Trp	Arg	Trp	891	Rv0278c	Rv0278c
sample1	NC_000962	336680	GGGGA	G	16	94	Non-synonymous	Deletion	c.2390_2393delTCCC	2390-2393	p.Val797fs	Val	NA	797	PE_PGRS4	Rv0279c

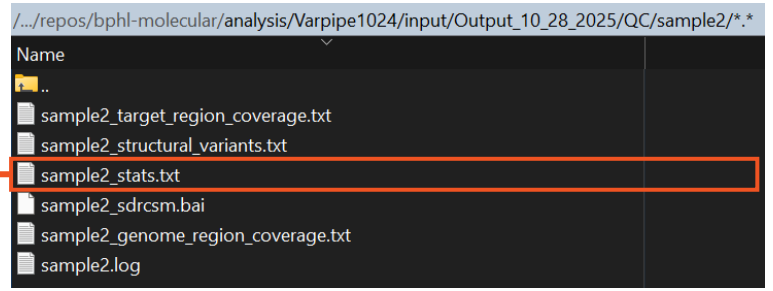
Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.



Sample ID	CHROM	POS	REF	ALT	Read Depth	Percent Alt Allele	Annotation	Variant Type	Nucleotide Change	Position within CDS	Amino acid Change	REF Amino acid	ALT Amino acid	Codon Position	Gene Name	Gene ID
sample1	NC_000962	761152	T	A	70	5.1	Non-synonymous	SNP	c.1346T>A	1346	p.Leu449Gln	Leu	Gln	449	rpoB	Rv0667

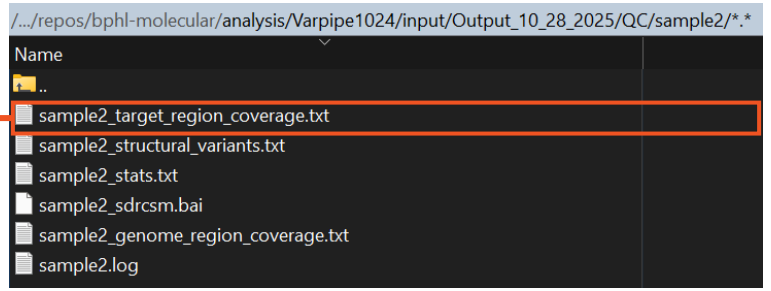
Application Cont.



Sample ID	Sample Name	Percent Reads Mapped	Average Genome Coverage Depth	Percent Reference Genome Covered	Coverage Drop	Pipeline Version	Date
sample2	sample2	4.8	3	28.88	18	Varpipeline: Varpipe_wgs_1.0.2	10/28/2025 15:32

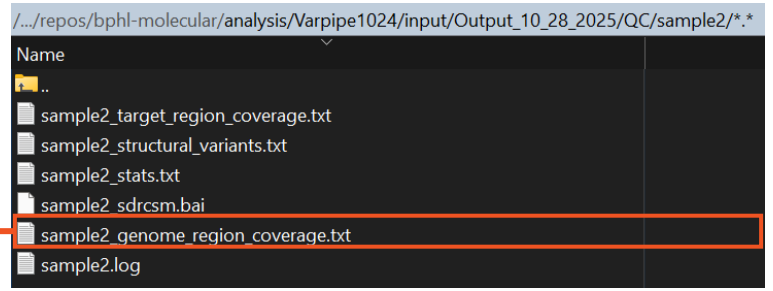
Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.



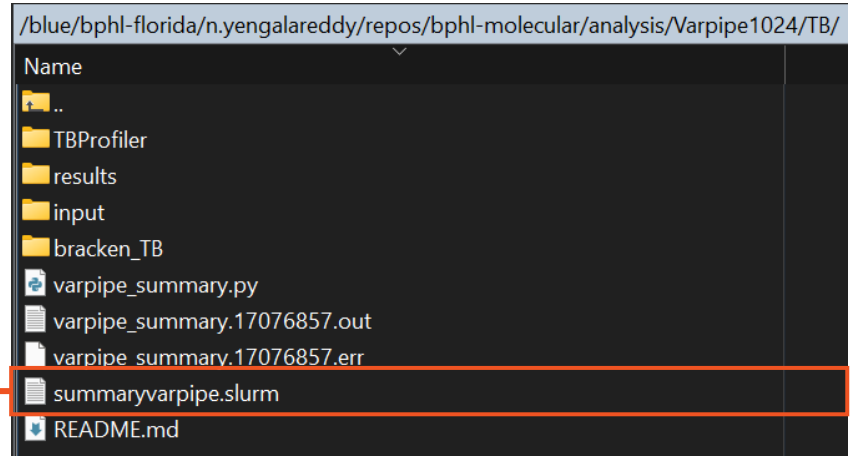
SAMPLE_ID	CHROM	Start	End	Gene Name	Gene ID	Flag
sample2	NC_000962.3	6571	6762	gyrB	Rv0005	Review
sample2	NC_000962.3	7360	7583	gyrA	Rv0006	Review
sample2	NC_000962.3	760307	761286	rpoB	Rv0667	Review
sample2	NC_000962.3	778989	779487	mmpR	Rv0678	Review
sample2	NC_000962.3	800808	801462	rplC1	Rv0701	Review
sample2	NC_000962.3	1461044	1461290	atpE	Rv1305	Review
sample2	NC_000962.3	1473245	1473331	rrs	MTB000019	Review
sample2	NC_000962.3	1473657	1476796	rrl	MTB000020	Review
sample2	NC_000962.3	1673409	1674052	fabG1	Rv1483	Review
sample2	NC_000962.3	1674201	1675012	inhA	Rv1484	Review
sample2	NC_000962.3	1917939	1918747	tlyA	Rv1694	Review
sample2	NC_000962.3	2153888	2156112	katG	Rv1908c	Review
sample2	NC_000962.3	2288676	2289272	pncA	Rv2043c	Review
sample2	NC_000962.3	2714123	2715372	eis	Rv2416c	Review
sample2	NC_000962.3	2726093	2726194	ahpC	Rv2428	Review
sample2	NC_000962.3	2859299	2860418	pepQ	Rv2535c	Review
sample2	NC_000962.3	4246586	4249653	embB	Rv3795	Review
sample2	NC_000962.3	4326003	4327474	ethA	Rv3854c	Review

Application Cont.



SAMPLE_ID	CHROM	START	END	GENE ID	GENE NAME	AVERAGE DEPTH	PERCENT REGION COVERAGE
sample2	NC_000962.3	1	1524	Rv0001	dnaA	1.19	60
sample2	NC_000962.3	1525	2051	Intragenic(Rv0001..Rv0002)	Intragenic(Rv0001..Rv0002)	1.07	46
sample2	NC_000962.3	2052	3260	Rv0002	dnaN	0.87	22
sample2	NC_000962.3	3261	3279	Intragenic(Rv0002..Rv0003)	Intragenic(Rv0002..Rv0003)	0.94	100
sample2	NC_000962.3	3280	4437	Rv0003	recF	0.96	78
sample2	NC_000962.3	4434	4997	Rv0004	Rv0004	0.44	26
sample2	NC_000962.3	4998	5239	Intragenic(Rv0004..Rv0005)	Intragenic(Rv0004..Rv0005)	0	0
sample2	NC_000962.3	5240	7267	Rv0005	gyrB	1.74	19
sample2	NC_000962.3	7268	7301	Intragenic(Rv0005..Rv0006)	Intragenic(Rv0005..Rv0006)	0	0
sample2	NC_000962.3	7302	9818	Rv0006	gyrA	0.66	22
sample2	NC_000962.3	9819	9913	Intragenic(Rv0006..Rv0007)	Intragenic(Rv0006..Rv0007)	0	0
sample2	NC_000962.3	9914	10828	Rv0007	Rv0007	1.05	23
sample2	NC_000962.3	10829	10886	Intragenic(Rv0007..Rvnt01)	Intragenic(Rv0007..Rvnt01)	0	0
sample2	NC_000962.3	10887	10960	Rvnt01	ileT	0	0
sample2	NC_000962.3	10961	11111	Intragenic(Rvnt01..Rvnt02)	Intragenic(Rvnt01..Rvnt02)	0.03	4
sample2	NC_000962.3	11112	11184	Rvnt02	alaT	1.39	100
sample2	NC_000962.3	11185	11873	Intragenic(Rvnt02..Rv0008c)	Intragenic(Rvnt02..Rv0008c)	0.71	45
sample2	NC_000962.3	11874	12311	Rv0008c	Rv0008c	0.62	42
sample2	NC_000962.3	12312	12467	Intragenic(Rv0008c..Rv0009)	Intragenic(Rv0008c..Rv0009)	1.5	94
sample2	NC_000962.3	12468	13016	Rv0009	ppiA	0.98	42
sample2	NC_000962.3	13017	13132	Intragenic(Rv0009..Rv0010c)	Intragenic(Rv0009..Rv0010c)	0	0
sample2	NC_000962.3	13133	13558	Rv0010c	Rv0010c	0	0
sample2	NC_000962.3	13559	13713	Intragenic(Rv0010c..Rv0011c)	Intragenic(Rv0010c..Rv0011c)	1.3	75
sample2	NC_000962.3	13714	13995	Rv0011c	Rv0011c	0.34	22
sample2	NC_000962.3	13996	14088	Intragenic(Rv0011c..Rv0012)	Intragenic(Rv0011c..Rv0012)	0	0
sample2	NC_000962.3	14089	14877	Rv0012	Rv0012	0.15	15
sample2	NC_000962.3	14878	14913	Intragenic(Rv0012..Rv0013)	Intragenic(Rv0012..Rv0013)	0.74	74
sample2	NC_000962.3	14914	15612	Rv0013	trpG	1.78	74
sample2	NC_000962.3	15590	17470	Rv0014c	pknB	0.64	22
sample2	NC_000962.3	17467	18762	Rv0015c	pknA	0.26	25
sample2	NC_000962.3	18759	20234	Rv0016c	pbpA	1.82	65
sample2	NC_000962.3	20231	21640	Rv0017c	rodA	2.48	29
sample2	NC_000962.3	21637	23181	Rv0018c	pstP	0.99	26
sample2	NC_000962.3	23182	23269	Intragenic(Rv0018c..Rv0019c)	Intragenic(Rv0018c..Rv0019c)	0.61	62

Application Cont.

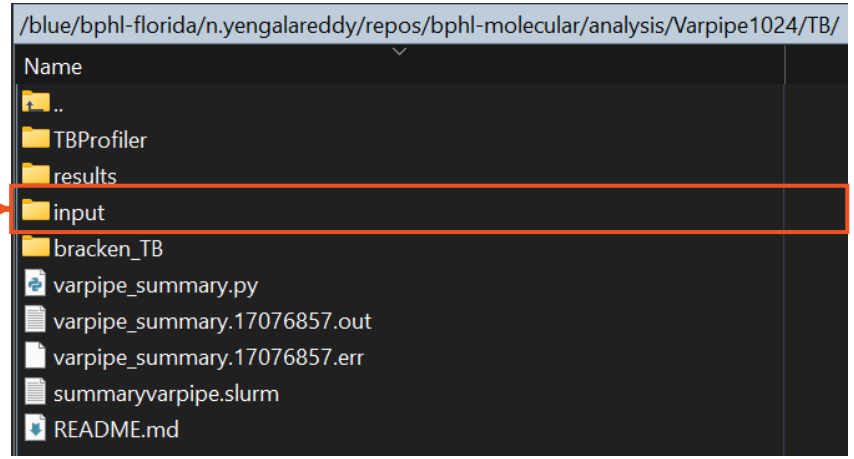



```
#!/bin/bash
#SBATCH --account=bphl-umbrella
#SBATCH --qos=bphl-umbrella
#SBATCH --job-name=varpipe_summary
#SBATCH --mail-type=END,FAIL
#SBATCH --mail-user=nikhil.yengala@flhealth.gov
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=4
#SBATCH --mem=64G
#SBATCH --time=12:00:00
#SBATCH --output=varpipe_summary.%j.out
#SBATCH --error=varpipe_summary.%j.err

#Load modules & conda environment
module purge
module load conda
conda activate TBprofiler

#Paths
BASE="/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Varpipe1024/TB"
INPUT_ROOT="${BASE}/input"
SUMMARY_PY="${BASE}/varpipe_summary.py"

# Find the latest Varpipe output folder
VARPIPE_OUT="$(ls -1dt "${INPUT_ROOT}"/Output_* 2>/dev/null | head -n1)"

if [[ -z "${VARPIPE_OUT}" ]]; then
  echo "[error] No Output_* directory found in ${INPUT_ROOT}" >&2
  exit 2
fi

echo "[info] Found Varpipe output directory: ${VARPIPE_OUT}"
echo "[info] Contents of this folder:"

ls -lah "${VARPIPE_OUT}"

#Run summary
OUT_SUMMARY="${INPUT_ROOT}/varpipe_summary.tsv"
echo "[info] Running summary â†’ ${OUT_SUMMARY}"
cd "${VARPIPE_OUT}"
python "${SUMMARY_PY}" -o "${OUT_SUMMARY}"

echo "[info] Summary step complete. Summary file: ${OUT_SUMMARY}"
```

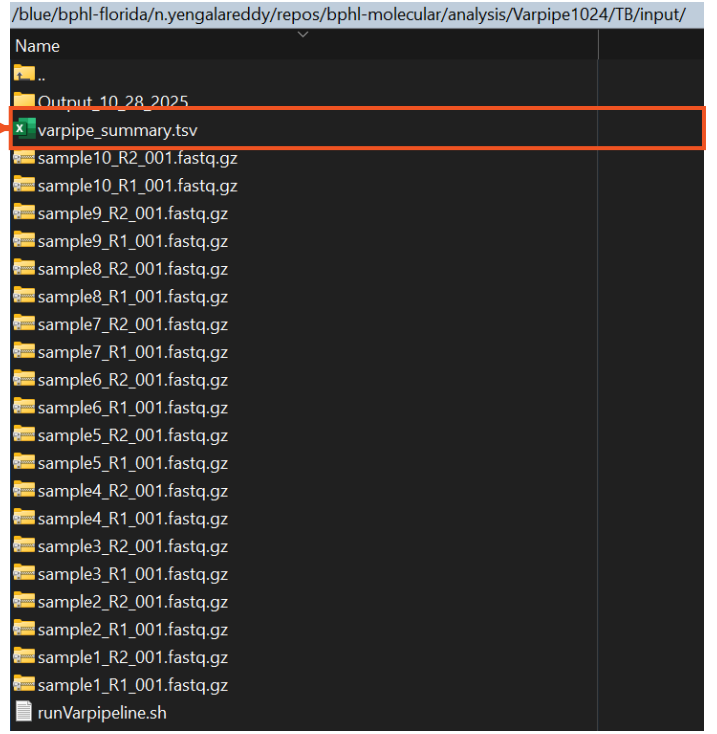
Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

```
sbatch
summaryvarpipe.slurm
```

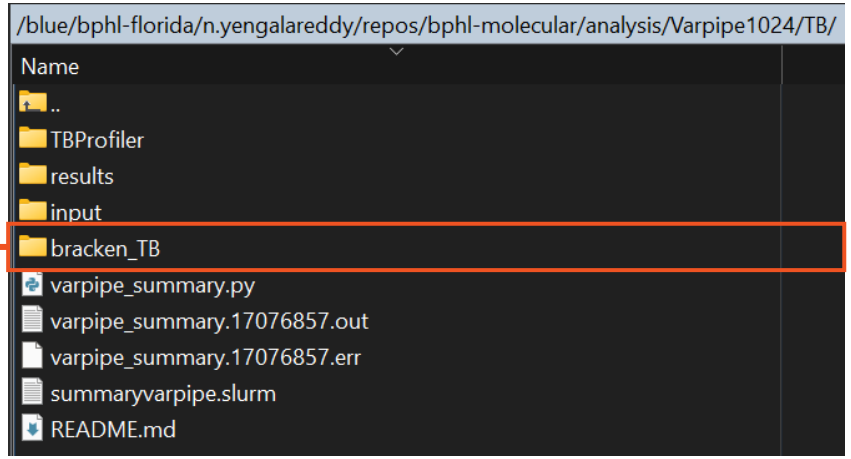


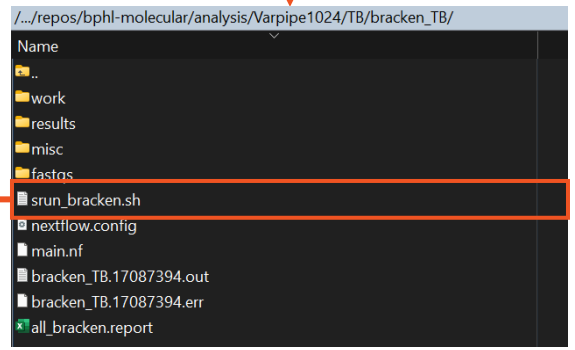
Sample ID	Sample Name	Percent Reads Mapped	Average Genome Coverage Depth	Percent Reference Genome Covered	Coverage Drop	QC_Status	Drug_Resistance_Summary
sample8	sample8	99.66	101	99.4	0	Pass	INH-S; RIF-U (rpoB_p.Leu443Phe); PZA-S; EMB-S
sample3	sample3	99.35	88	99.08	0	Pass	INH-S (katG_p.Arg463Leu); RIF-S (rpoB_p.Leu449Gln); PZA-S; EMB-S (embB_p.Glu378Ala)
sample5	sample5	99.5	77	99.39	0	Pass	INH-S; RIF-S; PZA-S; EMB-S
sample6	sample6	99.4	80	99	0	Pass	INH-S (katG_p.Arg463Leu); RIF-S (rpoB_p.Leu449Gln); PZA-S; EMB-S (embB_p.Glu378Ala)
sample10	sample10	98.87	85	99.38	0	Pass	INH-S; RIF-S; PZA-S; EMB-S
sample7	sample7	99.51	82	99.66	0	Pass	INH-S; RIF-S; PZA-S; EMB-S
sample9	sample9	99.81	82	99.28	0	Pass	INH-S; RIF-S (rpoB_p.Leu449Gln); PZA-S; EMB-S
sample1	sample1	99.91	82	99.9	0	Pass	INH-S; RIF-S (rpoB_p.Leu449Gln); PZA-S; EMB-S
sample4	sample4	4.19	8	6.96	18	Failed	QC Failed - No interpretation
sample2	sample2	4.8	3	28.88	18	Failed	QC Failed - No interpretation



Application Cont.

```
#!/usr/bin/bash
#SBATCH --account=bphl-umbrella
#SBATCH --qos=bphl-umbrella
#SBATCH --job-name=brackenTB
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=20
#SBATCH --mem=200gb
#SBATCH --time=48:00:00
#SBATCH --output=bracken.%j.out
#SBATCH --error=bracken.err
#SBATCH --mail-user=nikhil.yengala@flhealth.gov
#SBATCH --mail-type=FAIL,END

module load nextflow

APPTAINER_CACHEDIR=./
export APPTAINER_CACHEDIR

nextflow run main.nf --input "./fastqs/*_{R1_001,R2_001}.fastq.gz"

more results/bracken_out/*bracken >all_bracken.report
```

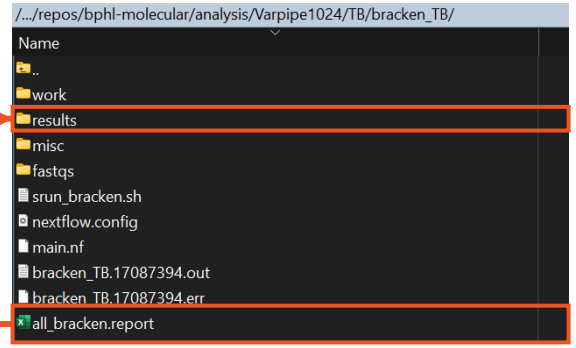




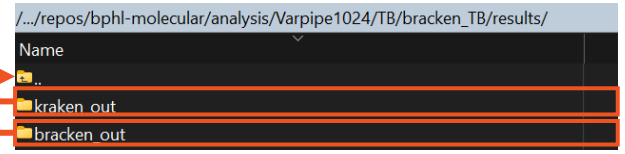
Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

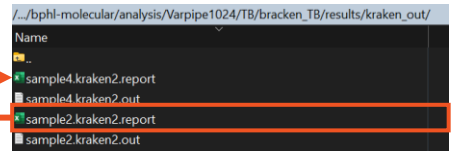
```
sbatch
srun_bracken.sh
```



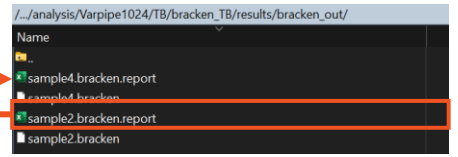
:::::::::::::::						
results/bracken_out/sample2.bracken						
:::::::::::::::						
name	taxonomy_id	taxonomy_lvl	kraken_assigned_reads	added_reads	new_est_reads	fraction_total_reads
Mycobacteroides abscessus	36809	S	1406210	15919	1422129	0.99891
Mycobacteroides chelonae	1774	S	1517	29	1546	0.00109
:::::::::::::::						
results/bracken_out/sample4.bracken						
:::::::::::::::						
name	taxonomy_id	taxonomy_lvl	kraken_assigned_reads	added_reads	new_est_reads	fraction_total_reads
Mycobacteroides abscessus	36809	S	1541247	62002	1603249	0.99213
Mycobacteroides chelonae	1774	S	12360	350	12710	0.00787



100	1423676	0	R	1	root
100	1423676	0	R1	131567	cellular organisms
100	1423676	0	D	2	Bacteria
100	1423676	0	D1	1783272	Terrabacteria group
100	1423676	0	P	201174	Actinomycetota
100	1423676	0	C	1760	Actinomycetes
100	1423676	0	O	85007	Mycobacteriales
100	1423676	0	F	1762	Mycobacteriaceae
100	1423676	0	G	670516	Mycobacteroides
99.89	1422129	1422129	S	36809	Mycobacteroides abscessus
0.11	1546	1546	S	1774	Mycobacteroides chelonae



4.07	61027	61027	U	0	unclassified
95.93	1438869	38	R	1	root
95.92	1438761	153	R1	131567	cellular organisms
95.87	1437902	460	D	2	Bacteria
95.75	1436200	62	D1	1783272	Terrabacteria group
95.74	1436066	40	P	201174	Actinomycetota
95.74	1436025	1251	C	1760	Actinomycetes
95.6	1433971	945	O	85007	Mycobacteriales
95.51	1432590	2131	F	1762	Mycobacteriaceae
94.72	1420765	11584	G	670516	Mycobacteroides
93.75	1406210	1351826	S	36809	Mycobacteroides abscessus
3	45048	45048	S1	1185650	Mycobacteroides abscessus subsp. abscessus
0.44	6546	3716	S1	319705	Mycobacteroides abscessus subsp. bolletii
0.19	2825	2825	S2	1091046	Mycobacteroides abscessus subsp. bolletii BD
0	5	5	S2	1303024	Mycobacteroides abscessus subsp. bolletii 50594
0.12	1823	1823	S1	1132508	Mycobacteroides abscessus UC22
0.06	897	518	S1	1962118	Mycobacteroides abscessus subsp. massiliense
0.02	314	314	S2	1001714	Mycobacteroides abscessus subsp. massiliense CCUG 48898 = JCM 15300
0	65	65	S2	1198627	Mycobacteroides abscessus subsp. massiliense str. GO 06
0	70	70	S1	561007	Mycobacteroides abscessus ATCC 19977

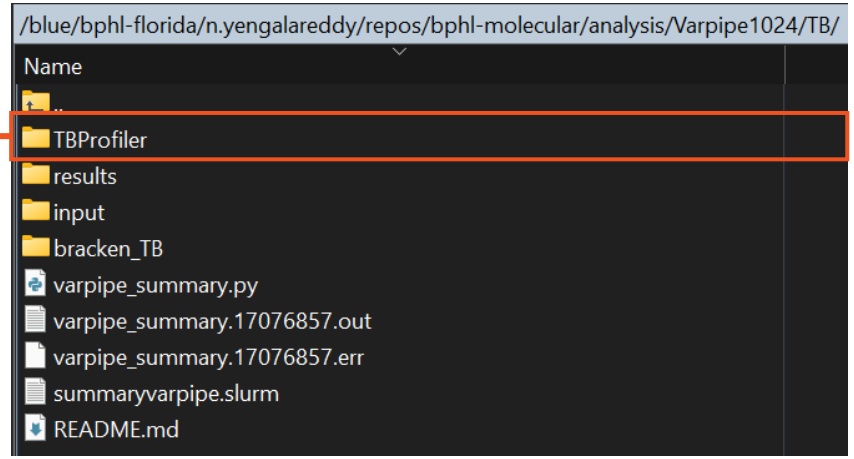


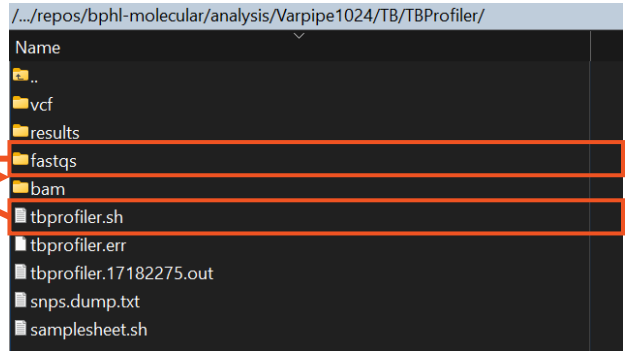
Advanced Molecular Detection
Southeast Region Bioinformatics

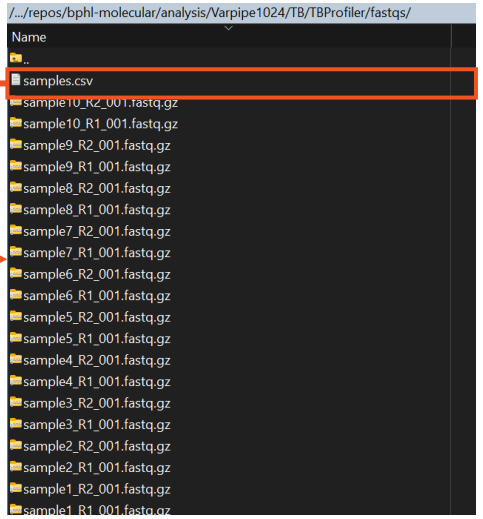
Application Cont.

```
#!/usr/bin/bash
#SBATCH --account=bphl-umbrella
#SBATCH --qos=bphl-umbrella
#SBATCH --job-name=TB
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=20
#SBATCH --mem=256gb
#SBATCH --time=48:00:00
#SBATCH --output=tbprofiler.%j.out
#SBATCH --error=tbprofiler.err
#SBATCH --mail-user=nikhil.yengala@flhealth.gov
#SBATCH --mail-type=FAIL,END

module load conda
conda activate TBprofiler

SAMPLESHEET="/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Varpipe1024/TB/TBProfiler/fastqs/samples.csv"

tb-profiler batch --csv "${SAMPLESHEET}" --args "--csv --af 0.1 --depth 10 --caller gatk"
```

```
id,read1,read2
sample10,./fastqs/sample10_R1_001.fastq.gz,./fastqs/sample10_R2_001.fastq.gz
sample1,./fastqs/sample1_R1_001.fastq.gz,./fastqs/sample1_R2_001.fastq.gz
sample2,./fastqs/sample2_R1_001.fastq.gz,./fastqs/sample2_R2_001.fastq.gz
sample3,./fastqs/sample3_R1_001.fastq.gz,./fastqs/sample3_R2_001.fastq.gz
sample4,./fastqs/sample4_R1_001.fastq.gz,./fastqs/sample4_R2_001.fastq.gz
sample5,./fastqs/sample5_R1_001.fastq.gz,./fastqs/sample5_R2_001.fastq.gz
sample6,./fastqs/sample6_R1_001.fastq.gz,./fastqs/sample6_R2_001.fastq.gz
sample7,./fastqs/sample7_R1_001.fastq.gz,./fastqs/sample7_R2_001.fastq.gz
sample8,./fastqs/sample8_R1_001.fastq.gz,./fastqs/sample8_R2_001.fastq.gz
sample9,./fastqs/sample9_R1_001.fastq.gz,./fastqs/sample9_R2_001.fastq.gz
```







Application Cont.

```
sbatch
tbprofiler.sh
```

/.../repos/bphl-molecular/analysis/Varpipe1024/TB/TBProfiler/

Name
- ..
- vcf
- results
- fastqs
- bam
- tbprofiler.sh
- tbprofiler.err
- tbprofiler.17182275.out
- snps.dump.txt
- samplesheet.sh

/.../repos/bphl-molecular/analysis/Varpipe1024/TB/TBProfiler/results/

Name
- ..
- sample10.results.json
- sample10.results.csv
- sample9.results.json
- sample9.results.csv
- sample8.results.json
- sample8.results.csv
- sample7.results.json
- sample7.results.csv
- sample6.results.json
- sample6.results.csv
- sample5.results.json
- sample5.results.csv
- sample4.results.json
- sample4.results.csv
- sample3.results.json
- sample3.results.csv
- sample2.results.json
- sample2.results.csv
- sample1.results.json
- sample1.results.csv

TBProfiler report
==========

The following report has been generated by TBProfiler.

Summary

ID	sample1
Date	03:26.9
Strain	lineage4.9
Drug-resis	Sensitive
Median De	88

Notes

Lineage report

Lineage	Fraction	Family	Rd
lineage4	1	Euro-Amer	None
lineage4.9	1	Euro-Amer	None

Resistance report

Drug	Genotypic	Mechanisms
Rifampicin		
Isoniazid		
Ethambutol		
Pyrazinamide		
Moxifloxacin		
Levofloxacin		
Bedaquiline		
Delamanid		
Pretomanid		
Linezolid		
Streptomycin		
Amikacin		
Kanamycin		
Capreomycin		
Clofazimine		
Ethionamide		
Para-aminosalicylic_acid		
Cycloserine		

Resistance variants report

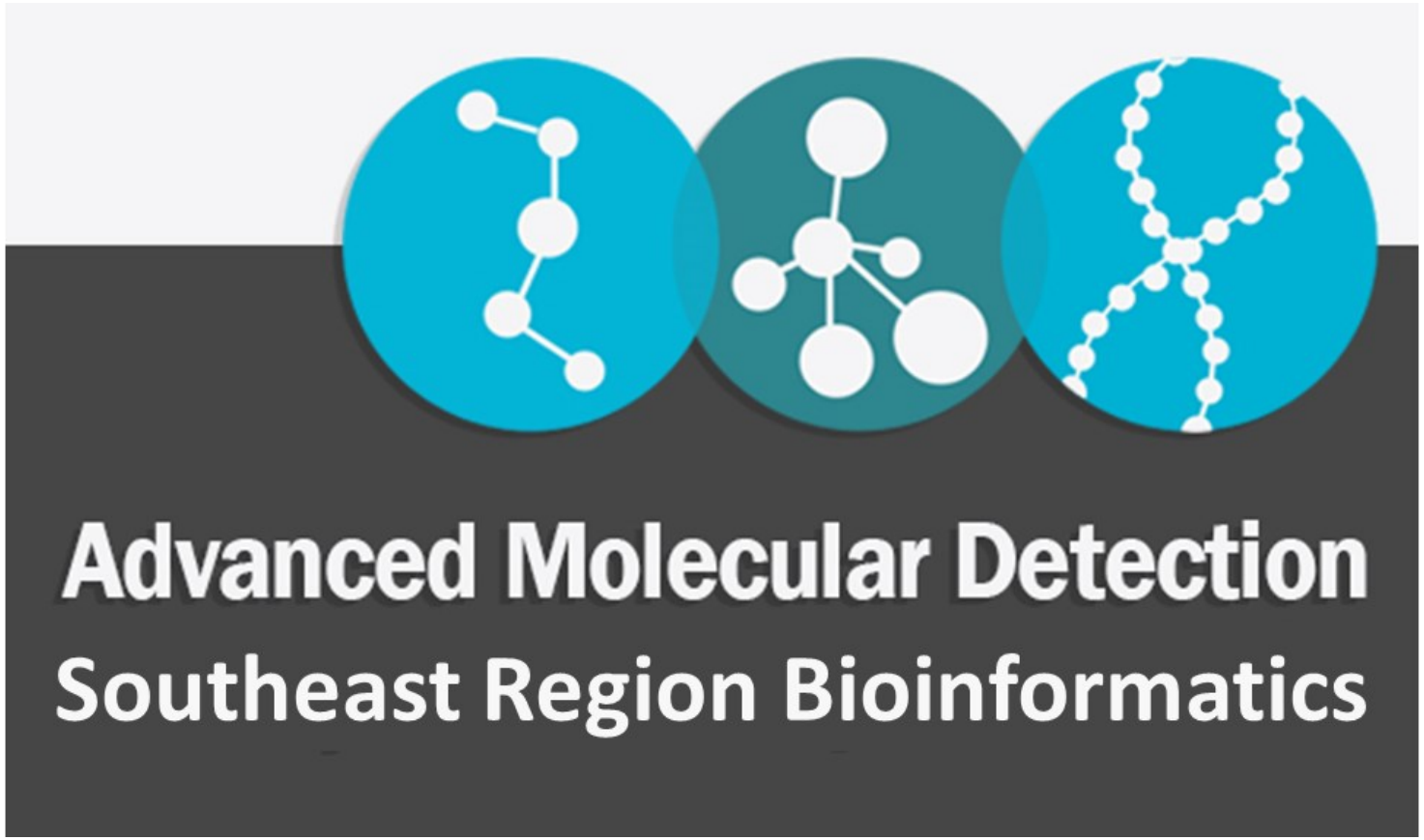
Genome P	Locus Tag	Gene Nam	Variant Ty	Change	Estimated Drug	Confidenc	Comment

Other variants report

Advanced Molecular Detection
Southeast Region Bioinformatics

Conclusion

- Fundamentals of Varpipe and TB Extension
- Installation and setup of Varpipe and TB Extension in HPG
- Successfully executed job query for Varpipe and TB Extension
- 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