

14 October 2024

Florida Assembly Quality – AMR Detection (FLAQ-AMR)

**Advanced Molecular Detection
Southeast Region Bioinformatics**

Updates

Office Hours

- October 28, 2024 – Sanibel
- November 18, 2024 – FLAQ-SARS-CoV-2



Advanced Molecular Detection
Southeast Region Bioinformatics

Overview

Purpose:

- Standard Bacterial Assembly

Used for:

- Taxonomic Identification
- Antimicrobial Resistance (AMR) Detection

Dependencies:

- Python3
- Singularity
- Git



Workflow

R1 and R2
fastq.gz files

Run the pipeline
on HPG

Analyzes
Illumina Paired-
End Sequencing

Outputs

AMR
Detection

Quality
Metrics
Report

Multi Locus
Sequencing
Typing (MLST)

De Novo
Assembly

Annotated
Assembly

Taxonomic
Classification



Application

Objective:

Use a training dataset (*Salmonella enterica* dataset) and analyze the dataset using the FLAQ-AMR pipeline.



Application Cont.

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

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

mkdir analysis/

cd analysis/

cp /blue/bphl-<state>/<user>/repos/bphl-molecular/flaq_amr/*

Mkdir fastqs/

cp /path/to/fastqs/*.fastq.gz fastqs/

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/flaq_amr1010/fastqs/*

Name	Size	Changed	Rights	Owner
📁		10/10/2024 11:45:46 AM	rw-r--r--	n.yeng...
📁 SRR2407650_1.fastq.gz	128,220 KB	9/20/2024 3:31:04 PM	rw-r--r--	n.yeng...
📁 SRR2407650_2.fastq.gz	164,480 KB	9/20/2024 3:31:04 PM	rw-r--r--	n.yeng...
📁 SRR2533501_1.fastq.gz	89,025 KB	9/20/2024 3:36:36 PM	rw-r--r--	n.yeng...
📁 SRR2533501_2.fastq.gz	106,571 KB	9/20/2024 3:36:36 PM	rw-r--r--	n.yeng...
📁 SRR2584382_1.fastq.gz	131,115 KB	9/20/2024 3:26:36 PM	rw-r--r--	n.yeng...
📁 SRR2584382_2.fastq.gz	152,329 KB	9/20/2024 3:26:36 PM	rw-r--r--	n.yeng...
📁 SRR6077140_1.fastq.gz	128,438 KB	9/20/2024 3:22:00 PM	rw-r--r--	n.yeng...
📁 SRR6077140_2.fastq.gz	155,578 KB	9/20/2024 3:22:00 PM	rw-r--r--	n.yeng...
📁 SRR6077143_1.fastq.gz	222,522 KB	9/20/2024 3:17:00 PM	rw-r--r--	n.yeng...
📁 SRR6077143_2.fastq.gz	238,799 KB	9/20/2024 3:17:00 PM	rw-r--r--	n.yeng...
📁 SRR6351069_1.fastq.gz	167,423 KB	9/20/2024 2:56:10 PM	rw-r--r--	n.yeng...
📁 SRR6351069_2.fastq.gz	178,768 KB	9/20/2024 2:56:10 PM	rw-r--r--	n.yeng...
📁 SRR6351075_1.fastq.gz	189,745 KB	9/20/2024 2:45:42 PM	rw-r--r--	n.yeng...
📁 SRR6351075_2.fastq.gz	218,433 KB	9/20/2024 2:45:42 PM	rw-r--r--	n.yeng...

n.yengalareddy@login10:/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis...

```
[n.yengalareddy@login10 flaq_amr1010]$ ls
2024-10-10_flaq_run  fastqs          LICENSE
amrfinder_results   flaq_amr.46901570.err  README.md
annotations          flaq_amr.46901570.out  sbatch_flaq_amr.sh
assemblies           flaq_amr.py
[n.yengalareddy@login10 flaq_amr1010]$
```



Application Cont.

nano sbatch_flgq_amr.sh

```
n.yengalareddy@login8:/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/fl...
GNU nano 2.9.8      sbatch_flgq_amr.sh

#!/bin/bash
#SBATCH --account=bphl-umbrella
#SBATCH --qos=bphl-umbrella
#SBATCH --job-name=flaq_amr
#SBATCH --mail-type=END,FAIL
#SBATCH --mail-user=nikhil.yengala@flhealth.gov
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=8
#SBATCH --mem=64gb
#SBATCH --time=1-00
#SBATCH --output=flaq_amr.%j.out
#SBATCH --error=flaq_amr.%j.err

module load apptainer

module load conda

conda activate base

#Run script/command and use $SLURM_CPUS_ON_NODE
python flaq_amr.py fastqs/

^G Get Help    ^O Write Out  ^W Where Is   ^K Cut Text   ^J Justify
^X Exit        ^R Read File  ^\ Replace    ^U Uncut Text ^T To Linter
```



```
n.yengalareddy@login8:/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/fla...  
[n.yengalareddy@login8 fla..._amr]$ nano sbatch_flg..._amr.sh  
[n.yengalareddy@login8 fla..._amr]$ sbatch sbatch_flg..._amr.sh  
Submitted batch job 46609146  
[n.yengalareddy@login8 fla..._amr]$
```

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/fla..._amr/

Name	Size	Changed	Rights	Owner
..		9/23/2024 8:04:39 AM	rw-r--r--	n.yeng...
2024-10-07_flg..._run		10/7/2024 3:09:00 PM	rw-r--r--	n.yeng...
amrfinder_results		10/7/2024 3:29:21 PM	rw-r--r--	n.yeng...
annotations		10/7/2024 3:29:21 PM	rw-r--r--	n.yeng...
assemblies		10/7/2024 3:29:21 PM	rw-r--r--	n.yeng...
fastqs		9/20/2024 3:38:42 PM	rw-r--r--	n.yeng...
fla..._amr.46609146.err	0 KB	10/7/2024 1:04:26 PM	rw-r--r--	n.yeng...
fla..._amr.46609146.out	0 KB	10/7/2024 1:04:26 PM	rw-r--r--	n.yeng...
fla..._amr.py	14 KB	9/19/2024 2:01:32 PM	rw-r--r--	n.yeng...
LICENSE	2 KB	9/19/2024 2:01:32 PM	rw-r--r--	n.yeng...
README.md	5 KB	9/19/2024 2:01:32 PM	rw-r--r--	n.yeng...
sbatch_flg..._amr.sh	1 KB	9/19/2024 2:06:38 PM	rw-r--r--	n.yeng...

Application Cont.



Output

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/flaq_amr/2024-10-07_flaq_run/*.*				
Name	Size	Changed	Rights	Owner
..		10/7/2024 1:04:29 PM	rwxr-sr-x	n.yeng...
SRR2407650		10/7/2024 1:23:49 PM	rwxr-sr-x	n.yeng...
SRR2533501		10/7/2024 1:43:52 PM	rwxr-sr-x	n.yeng...
SRR2584382		10/7/2024 2:06:03 PM	rwxr-sr-x	n.yeng...
SRR6077140		10/7/2024 2:24:04 PM	rwxr-sr-x	n.yeng...
SRR6077143		10/7/2024 2:49:38 PM	rwxr-sr-x	n.yeng...
SRR6351069		10/7/2024 3:08:24 PM	rwxr-sr-x	n.yeng...
SRR6351075		10/7/2024 3:28:45 PM	rwxr-sr-x	n.yeng...
report.txt	2 KB	10/7/2024 3:29:21 PM	rw-r--r--	n.yeng...

sampleID	speciesID_mash	nearest_neighbor_mash	mash_distance	speciesID_kraken	kraken_percent	mlst_scheme	mlst_st	num_clean_reads	avg_readlength	avg_read_qual	est_coverage	num_contigs	longest_contig	N50	L50	total_length	gc_content	annotated_cds
SRR2407650	Salmonella_enterica	GCF_000623775.1	0.0011856	Salmonella enterica	97.05	senterica	11	1027736	225.85	36.03	49.14	32	1508692	400610	3	4722943	52	4433
SRR2533501	Salmonella_enterica	GCF_000623775.1	0.00060662	Salmonella enterica	97.74	senterica	11	937312	221.7	36.44	44.04	26	1508692	405853	3	4717664	52	4420
SRR2584382	Salmonella_enterica	GCF_000623775.1	0.000931334	Salmonella enterica	97.29	senterica	11	1460314	222.61	36.64	69.75	55	634220	197759	7	4660154	52	4356
SRR6077140	Salmonella_enterica	GCF_000623775.1	0.00213483	Salmonella enterica	96.82	senterica	11	1131246	210.83	36.38	50.66	36	1215647	405952	4	4707886	52	4411
SRR6077143	Salmonella_enterica	GCF_000623775.1	0.00343111	Salmonella enterica	96.83	senterica	11	2255536	211.83	36.39	101.3	27	1508692	478604	3	4716521	52	4421
SRR6351069	Salmonella_enterica	GCF_000623775.1	0.00213483	Salmonella enterica	96.27	senterica	11	1506344	200.02	36.15	63.89	32	1024859	440575	4	4715769	52	4422
SRR6351075	Salmonella_enterica	GCF_000623775.1	0.00157453	Salmonella enterica	96.24	senterica	11	1124130	183.18	35.64	44.13	29	1024859	406018	4	4665490	52	4365











Output Cont.

/.../bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/flaq_amr1010/amrfinder_results/				
Name	Size	Changed	Rights	Owner
..		10/10/2024 11:45:46 AM	rw-r--r--	n.yeng...
SRR2407650_amrfinde...	2 KB	10/10/2024 12:05:26 PM	rw-r--r--	n.yeng...
SRR2533501_amrfinde...	2 KB	10/10/2024 12:25:25 PM	rw-r--r--	n.yeng...
SRR2584382_amrfinde...	2 KB	10/10/2024 12:47:56 PM	rw-r--r--	n.yeng...
SRR6077140_amrfinde...	2 KB	10/10/2024 1:06:52 PM	rw-r--r--	n.yeng...
SRR6077143_amrfinde...	2 KB	10/10/2024 1:33:21 PM	rw-r--r--	n.yeng...
SRR6351069_amrfinde...	2 KB	10/10/2024 1:52:26 PM	rw-r--r--	n.yeng...
SRR6351075_amrfinde...	2 KB	10/10/2024 2:14:26 PM	rw-r--r--	n.yeng...

0 B of 10.5 KB in 0 of 7



/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/flaq_amr1010/annotations/









Name	Size	Changed	Rights	Owner
		10/10/2024 11:45:46 AM	rw-r--r--	n.yeng...
 SRR2407650.gff	6,682 KB	10/10/2024 12:05:26 PM	rw-r--r--	n.yeng...
 SRR2533501.gff	6,673 KB	10/10/2024 12:25:25 PM	rw-r--r--	n.yeng...
 SRR2584382.gff	6,586 KB	10/10/2024 12:47:56 PM	rw-r--r--	n.yeng...
 SRR6077140.gff	6,657 KB	10/10/2024 1:06:52 PM	rw-r--r--	n.yeng...
 SRR6077143.gff	6,671 KB	10/10/2024 1:33:21 PM	rw-r--r--	n.yeng...
 SRR6351069.gff	6,670 KB	10/10/2024 1:52:26 PM	rw-r--r--	n.yeng...
 SRR6351075.gff	6,595 KB	10/10/2024 2:14:26 PM	rw-r--r--	n.yeng...

0 B of 45.4 MB in 0 of 7

Output Cont.



Output Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/flaq_amr1010/assemblies/					
Name	Size	Changed	Rights	Owner	
 ..		10/10/2024 11:45:46 AM	rw-r--r--	n.yeng...	
 SRR2407650.fasta	4,680 KB	10/10/2024 12:05:26 PM	rw-r--r--	n.yeng...	
 SRR2533501.fasta	4,675 KB	10/10/2024 12:25:25 PM	rw-r--r--	n.yeng...	
 SRR2584382.fasta	4,619 KB	10/10/2024 12:47:56 PM	rw-r--r--	n.yeng...	
 SRR6077140.fasta	4,666 KB	10/10/2024 1:06:51 PM	rw-r--r--	n.yeng...	
 SRR6077143.fasta	4,674 KB	10/10/2024 1:33:21 PM	rw-r--r--	n.yeng...	
 SRR6351069.fasta	4,674 KB	10/10/2024 1:52:26 PM	rw-r--r--	n.yeng...	
 SRR6351075.fasta	4,623 KB	10/10/2024 2:14:26 PM	rw-r--r--	n.yeng...	
0 B of 31.8 MB in 0 of 7					



Conclusion



Fundamentals of FLAQ-AMR



Installation and setup of FLAQ-AMR in HPG



Successfully executed job query for FLAQ-AMR



Generated Output Files





Advanced Molecular Detection

Southeast Region Bioinformatics

Questions?

bphl-sebioinformatics@flhealth.gov

Molly Mitchell, PhD

Bioinformatics Supervisor

Molly.Mitchell@flhealth.gov

Nikhil Reddy, MS

Bioinformatician

Nikhil.Yengala@flhealth.gov

Sam Marcellus, MPH

Bioinformatician

Samantha.marcellus@flhealth.gov