

27 Jan 2025

Amelia

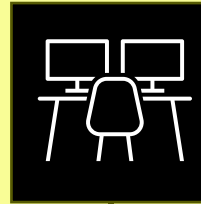
Advanced Molecular Detection
Southeast Region Bioinformatics

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

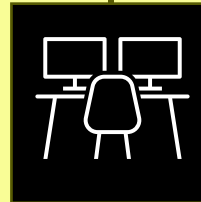
Office Hours

2025



February 03
Pensacola

February 17
Talbot



2025

Overview

Purpose

- Analyze targeted NGS data of *Mycoplasma genitalium* to produce outputs such as QC, amplicon statistics, SNP calling, Coverage, and AA Variation detection.

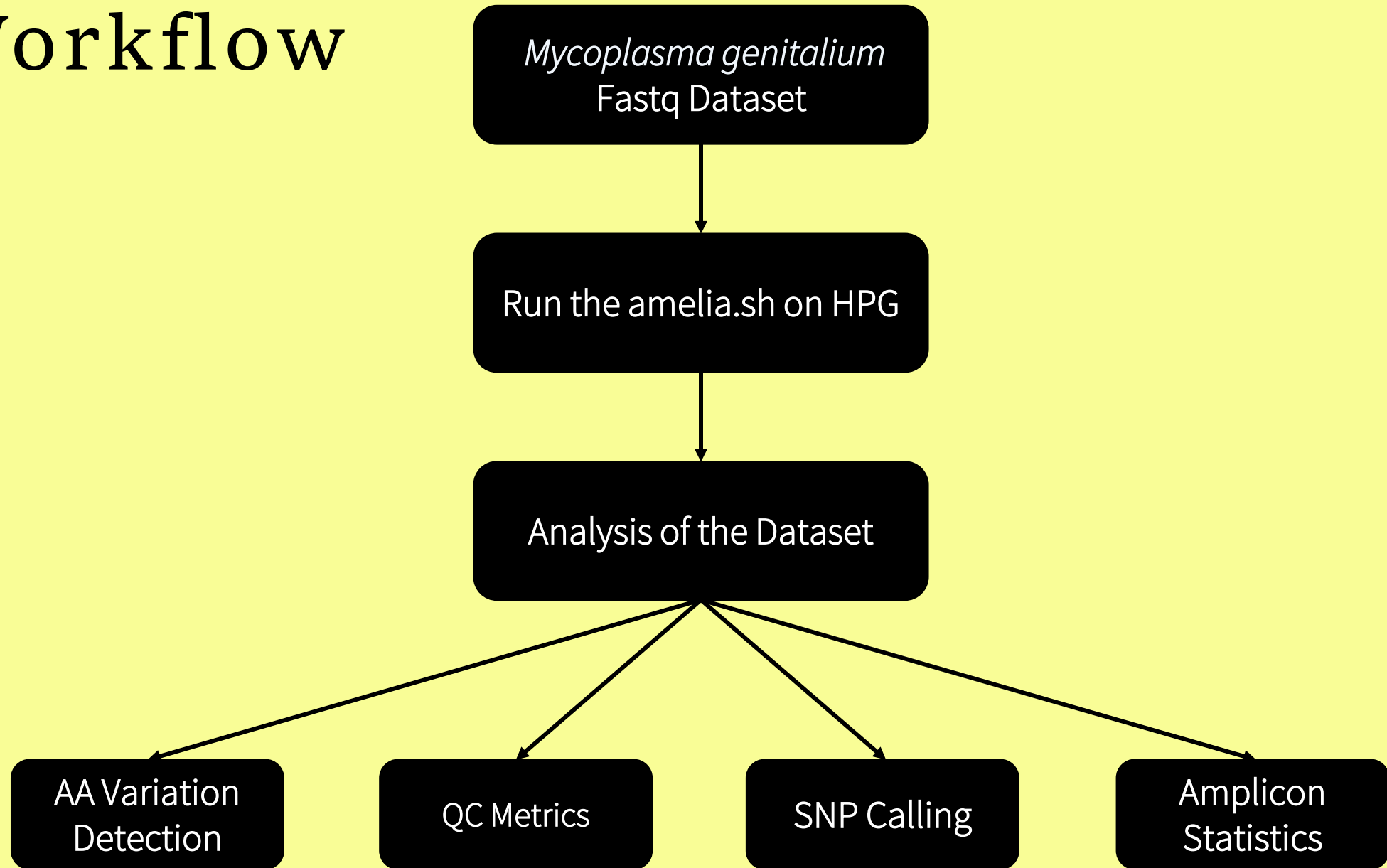
Usage

- To support public health and researchers by providing detailed reports and analyses of the NGS data which enables insights into genomic variations and quality metrics

Dependencies

- Nextflow
- Python3
- Singularity/Apptainer
- Biopython

Workflow



Application

Objective

Use *Mycoplasma genitalium* dataset and analyze
using **Amelia** pipeline

Application Cont.

```
cd /blue/bphl-<state>/<user>/repos/bphl-molecular/  
git clone https://github.com/BPHL-Molecular/Amelia  
mkdir analysis/  
cd analysis/  
cp /blue/bphl-<state>/<user>/repos/bphl-molecular/ Amelia/  
mkdir fastqs/  
cp /path/to/fastqs/*.fastq.gz fastqs/
```

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelia1220/

Name	Size
..	
reference	
primers	
output-20241220163752	
modules	
fastqs	
configs	
README.md	3 KB
params.yaml	1 KB
nextflow.config	1 KB
myco_target.nf	2 KB
LICENSE	2 KB
amelia.sh	1 KB

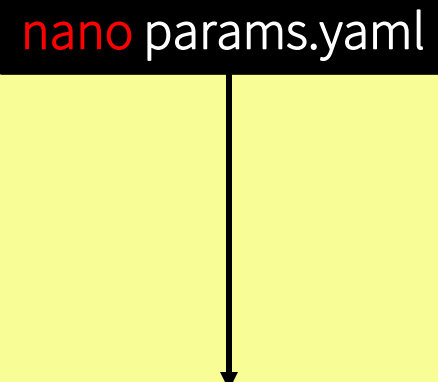
/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelia1220/fastqs/

Name
..
J877_2.fastq.gz
J877_1.fastq.gz
J852_2.fastq.gz
J852_1.fastq.gz
J839_2.fastq.gz
J839_1.fastq.gz
J750_2.fastq.gz
J750_1.fastq.gz
J627_2.fastq.gz
J627_1.fastq.gz
J620_2.fastq.gz
J620_1.fastq.gz
J614_2.fastq.gz
J614_1.fastq.gz
J464_2.fastq.gz
J464_1.fastq.gz
J418_2.fastq.gz
J418_1.fastq.gz
J414_2.fastq.gz
J414_1.fastq.gz
J407_2.fastq.gz
J407_1.fastq.gz
J197_2.fastq.gz
J197_1.fastq.gz
J179_2.fastq.gz
J179_1.fastq.gz
J148_2.fastq.gz
J148_1.fastq.gz
J147_2.fastq.gz



Application Cont.

nano params.yaml



```
GNU nano 2.9.8 params.yaml

# The parameters "input", "output", "reference" and "primer" are the absolute paths. Do not include the "/" $
input      : "/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelial220/fastqs"
output     : "/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelial220/output"
reference   : "/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelial220/reference"
primer     : "/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelial220/primers"
```



Application Cont.

nano amelia.sh

```
GNU nano 2.9.8 amelia.sh

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

module load apptainer
module load nextflow
APPTAINER_CACHEDIR=./
export APPTAINER_CACHEDIR
```



Application Cont.

activate conda environment containing Nextflow, Biopython
&
sbatch amelia.sh

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelia1220/

Name	Size
..	
reference	
primers	
output-20241220163752	
modules	
fastqs	
configs	
README.md	3 KB
params.yaml	1 KB
nextflow.config	1 KB
myco_target.nf	2 KB
LICENSE	2 KB
amelia.sh	1 KB

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelia1220/

Name	Size
..	
reference	
primers	
output-20241220163752	
modules	
fastqs	
configs	
README.md	3 KB
params.yaml	1 KB
nextflow.config	1 KB
myco_target.nf	2 KB
LICENSE	2 KB
amelia.sh	1 KB

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelia1220/output-20241220163752/

Name	Size
..	
variants	
J877	
J852	
J839	
J750	
J627	
J620	
J614	
J464	
J418	
J414	
J407	
J197	
J179	
J148	
J147	
sum_report.txt	6 KB
amelia.53426025.out	67 KB
amelia.53426025.err	1 KB

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelia1220/output-20241220163752/

Name	Size
..	
variants	
J877	
J852	
J839	
J750	
J627	
J620	
J614	
J464	
J418	
J414	
J407	
J197	
J179	
J148	
J147	
sum_report.txt	6 KB
amelia.53426025.out	6 / KB
amelia.53426025.err	1 KB

reference	start	end	num_raw_reads	num_clean_reads	num_mapped_reads	percent_mapped_clean_reads	mean_base_qual	mean_map_qual	23S_rRNA
L43967.2	1	580076	287052	178714	86344	48.3141	36.9	60	4280,872.6,55.42,A173799G;T173926G
L43967.2	1	580076	319242	189298	80275	42.4067	36.9	60	3563,720.7,55.42,A173799G
L43967.2	1	580076	322228	191366	82030	42.8655	36.9	60	4608,939.1,55.42,A173798G
L43967.2	1	580076	286856	179500	59476	33.1343	36.9	60	81,14.4,45.82,G173737A;G173738A;A173741G;A173759T;A1737
L43967.2	1	580076	249572	157344	74078	47.0803	36.9	60	4276,862.7,55.42,A173799G;G173980A;T174006C;G174102C;T17
L43967.2	1	580076	266844	161298	70563	43.747	36.9	60	3732,754.8,55.42,A173799G;G174188T
L43967.2	1	580076	415754	234506	94882	40.4604	36.9	60	3816,773.2,55.42,A173799G
L43967.2	1	580076	380106	213214	88602	41.5554	36.9	60	3977,805.1,55.26,A173799G
L43967.2	1	580076	384450	225350	90186	40.0204	36.9	60	4271,873.4,55.42,A173799G
L43967.2	1	580076	323576	183196	77618	42.3688	36.9	60	4328,878.9,55.42,A173799G;T173926G
L43967.2	1	580076	311558	181118	80960	44.7001	36.9	60	4564,926.9,55.42,A173799G
L43967.2	1	580076	244694	155484	74942	48.1992	36.9	60	4126,847.7,55.42,A173798G;G174405C
L43967.2	1	580076	285376	177854	82757	46.5309	36.9	60	4217,863.8,55.42,T173926G
L43967.2	1	580076	360978	191414	65514	34.2263	36.9	60	9,1.5,45.82,A173799G
L43967.2	1	580076	301984	178424	69138	38.7493	36.9	60	3558,723.8,55.42,C173967T

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelia1220/output-20241220163752/variants/		
Name		Size
J877.variants.vcf		4 KB
J852.variants.vcf		5 KB
J839.variants.vcf		4 KB
J750.variants.vcf		5 KB
J627.variants.vcf		6 KB
J620.variants.vcf		4 KB
J614.variants.vcf		6 KB
J464.variants.vcf		5 KB
J418.variants.vcf		4 KB
J414.variants.vcf		4 KB
J407.variants.vcf		7 KB
J197.variants.vcf		10 KB
J179.variants.vcf		4 KB
J148.variants.vcf		4 KB
J147.variants.vcf		4 KB

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelia1220/output-20241220163752/		
Name		Size
variants		
J877		
J852		
J839		
J750		
J627		
J620		
J614		
J464		
J418		
J414		
J407		
J197		
J179		
J148		
J147		
sum_report.txt		6 KB
amelia.53426025.out		67 KB
amelia.53426025.err		1 KB

```
##fileformat=VCFv4.2
##FILTER=ID=PASS,Description="All filters passed">
##bcftoolsVersion=1.12+htslib-1.12
##bcftoolsCommand=pileup -d 8000 -f /blue/bphl-florida/n.yengalareddy/repos/bphl-
molecular/analysis/Amelia1220/reference/L43967.2.fasta -B -a FORMAT/AD /blue/bphl-florida/n.yengalareddy/repos/bphl-
molecular/analysis/Amelia1220/output/3877/alignment/3877.sorted.bam
##reference=file:///blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelia1220/reference/L43967.2.fasta
##contig=ID=L43967.2,length=580076>
##ALT=<ID=*,Description="Represents allele(s) other than observed.">
##INFO=ID=INDEL,Number=0,Type=Flag,Description="Indicates that the variant is an INDEL.">
##INFO=ID=IDV,Number=1,Type=Integer,Description="Maximum number of raw reads supporting an indel">
##INFO=ID=IMF,Number=1,Type=Float,Description="Maximum fraction of raw reads supporting an indel">
##INFO=ID=DP,Number=1,Type=Integer,Description="Raw read depth">
##INFO=ID=VDB,Number=1,Type=Float,Description="Variant Distance Bias for filtering splice-site artefacts in RNA-seq data (bigger is
better)",Version="3">
##INFO=ID=RPB,Number=1,Type=Float,Description="Mann-Whitney U test of Read Position Bias (bigger is better)">
##INFO=ID=MQB,Number=1,Type=Float,Description="Mann-Whitney U test of Mapping Quality Bias (bigger is better)">
##INFO=ID=BQB,Number=1,Type=Float,Description="Mann-Whitney U test of Base Quality Bias (bigger is better)">
##INFO=ID=MQSB,Number=1,Type=Float,Description="Mann-Whitney U test of Mapping Quality vs Strand Bias (bigger is better)">
##INFO=ID=SQB,Number=1,Type=Float,Description="Segregation based metric.">
##INFO=ID=MQ0F,Number=1,Type=Float,Description="Fraction of MQ0 reads (smaller is better)">
##FORMAT=ID=PL,Number=6,Type=Integer,Description="List of Phred-scaled genotype likelihoods">
##FORMAT=ID=AD,Number=8,Type=Integer,Description="Allelic depths (high-quality bases)">
##FORMAT=ID=GT,Number=1,Type=String,Description="Genotype">
##INFO=ID=AC,Number=A,Type=Integer,Description="Allele count in genotypes for each ALT allele, in the same order as listed">
##INFO=ID=AN,Number=1,Type=Integer,Description="Total number of alleles in called genotypes">
##INFO=ID=DP4,Number=4,Type=Integer,Description="Number of high-quality ref-forward, ref-reverse, alt-forward and alt-reverse
bases">
##INFO=ID=MQ,Number=1,Type=Integer,Description="Average mapping quality">
##bcftools_callVersion=1.12+htslib-1.12
##bcftools_callCommand=call -mv -Ov; Date=Fri Dec 20 16:37:40 2024
##bcftools_normVersion=1.12+htslib-1.12
##bcftools_normCommand=norm -f /blue/bphl-florida/n.yengalareddy/repos/bphl-
molecular/analysis/Amelia1220/reference/L43967.2.fasta -; Date=Fri Dec 20 16:37:40 2024
##bcftools_filterVersion=1.12+htslib-1.12
##bcftools_filterCommand=filter -e 'QUAL < 20' -; Date=Fri Dec 20 16:37:40 2024
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT /blue/bphl-florida/n.yengalareddy/repos/bphl-
molecular/analysis/Amelia1220/output/3877/alignment/3877.sorted.bam
L43967.2 A 36790 T A 99.3668 PASS DP=9;VDB=6.71664e-05;SGB=-0.636426;RPB=0.5;MQB=0;BQB=0.5;MQ0F=0;AC=
1;AN=2;DP4=0,2,0,7;MQ=55 GT:PL:AD 0/1:134,0,32,2,7
L43967.2 173967 . C T 228.425 PASS DP=4208;VDB=0.999503;SGB=-0.693147;RPB=1;MQB=1;MQSB=1;BQB=1;MQ0F=
0;AC=2;AN=2;DP4=1,0,2077,890;MQ=60 GT:PL:AD 1/1:255,255,0,1,2967
L43967.2 193158 . T C 30.4183 PASS DP=2;SGB=-0.379885;MQ0F=0;AC=2;AN=2;DP4=0,0,1,0;MQ=60 GT:PL:AD
1/1:60,3,0,0,1
L43967.2 242310 . C T 225.417 PASS DP=4060;VDB=0.941952;SGB=-0.693147;MQSB=1;MQ0F=0;AC=2;AN=2;DP4=
0,0,1495,1396;MQ=60 GT:PL:AD 1/1:255,255,0,0,2891
```

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelia1220/output-20241220163752/J877/

Name	Size
..	
variants	
multiqc_data	
alignment	
report.txt	1 KB
output.txt	1 KB
multiqc_report.html	1,123 KB
J877_trimstats.txt	0 KB
J877_phixstats.txt	1 KB
J877_matchedphix.fq	0 KB
J877_2_original_fastqc.zip	375 KB
J877_2_original_fastqc.html	620 KB
J877_2_clean_fastqc.zip	350 KB
J877_2_clean_fastqc.html	570 KB
J877_2.fq.gz	4,772 KB
J877_2.fastq.gz	7,057 KB
J877_1_original_fastqc.zip	374 KB
J877_1_original_fastqc.html	619 KB
J877_1_clean_fastqc.zip	357 KB
J877_1_clean_fastqc.html	574 KB
J877_1.fq.gz	4,743 KB
J877_1.fastq.gz	6,936 KB
J877.log	23,476 KB
J877_adapters.stats.txt	1 KB
intersect	3 KB
astats	21 KB
aln-se.sam	64,659 KB
aln-se.bam	41,046 KB

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelia1220/output-20241220163752/

Name	Size
..	
variants	
J877	
J852	
J839	
J750	
J627	
J620	
J614	
J464	
J418	
J414	
J407	
J197	
J179	
J148	
J147	
sum_report.txt	6 KB
amelia.53426025.out	67 KB
amelia.53426025.err	1 KB

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelia1220/output-20241220163752/J877/

Name	Size
..	
variants	
multiqc_data	
alignment	
report.txt	1 KB
output.txt	1 KB
multiqc_report.html	1,123 KB
J877_trimstats.txt	0 KB
J877_phixstats.txt	1 KB
J877_matchedphix.fq	0 KB
J877_2_original_fastqc.zip	375 KB
J877_2_original_fastqc.html	620 KB
J877_2_clean_fastqc.zip	350 KB
J877_2_clean_fastqc.html	570 KB
J877_2.fq.gz	4,772 KB
J877_2.fastq.gz	7,057 KB
J877_1_original_fastqc.zip	374 KB
J877_1_original_fastqc.html	619 KB
J877_1_clean_fastqc.zip	357 KB
J877_1_clean_fastqc.html	574 KB
J877_1.fq.gz	4,743 KB
J877_1.fastq.gz	6,936 KB
J877.log	23,476 KB
J877_adapters.stats.txt	1 KB
intersect	3 KB
astats	21 KB
aln-se.sam	64,659 KB
aln-se.bam	41,046 KB

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelia1220/output-20241220163752/J877/variants/

Name	Size	Ch
..		12
J877.variants.vcf	4 KB	12

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelia1220/output-20241220163752/J877/multiqc_data/

Name	Size
..	
multiqc_sources.txt	1 KB
multiqc_qualimap_bamqc_genome_results.txt	1 KB
multiqc_general_stats.txt	1 KB
multiqc_fastqc.txt	1 KB
multiqc_data.json	132 KB
multiqc.log	16 KB

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelia1220/output-20241220163752/J877/

Name	Size
..	
variants	
multiqc_data	
alignment	
report.txt	1 KB
output.txt	1 KB
multiqc_report.html	1,123 KB
J877_trimstats.txt	0 KB
J877_phixstats.txt	1 KB
J877_matchedphix.fq	0 KB
J877_2_original_fastqc.zip	375 KB
J877_2_original_fastqc.html	620 KB
J877_2_clean_fastqc.zip	350 KB
J877_2_clean_fastqc.html	570 KB
J877_2.fq.gz	4,772 KB
J877_2.fastq.gz	7,057 KB
J877_1_original_fastqc.zip	374 KB
J877_1_original_fastqc.html	619 KB
J877_1_clean_fastqc.zip	357 KB
J877_1_clean_fastqc.html	574 KB
J877_1.fq.gz	4,743 KB
J877_1.fastq.gz	6,936 KB
J877.log	23,476 KB
J877_adapters.stats.txt	1 KB
intersect	3 KB
astats	21 KB
aln-se.sam	64,659 KB
aln-se.bam	41,046 KB

Sample	FastQC_mqc-generalstats-fastqc-total_sequences	FastQC_mqc-generalstats-fastqc-percent_fails	FastQC_mqc-generalstats-fastqc-avg_sequence_length	FastQC_mqc-generalstats-fastqc-percent_duplicates	FastQC_mqc-generalstats-fastqc-percent_gc
J877_1	150992	36.363636363637	105.0353661121119	73.98074070149411	44
J877_2	150992	36.363636363637	105.22481985800572	73.82179188301367	44

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelia1220/output-20241220163752/J877/alignment/

Name	Size
..	
J877.sorted.bam.mpileup	17,854 KB
J877.sorted.bam.bai	1 KB
J877.sorted.bam	1,887 KB
J877.positionsort.bam	3,332 KB
J877.namesorted.bam	5,246 KB
J877.markdup.bam	3,385 KB
J877.fixmate.bam	5,524 KB
J877.dedup.bam	1,887 KB
J877.coverage.txt	1 KB

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelia1220/output-20241220163752/J877/

Name	Size
..	
variants	
multiqc_data	
alignment	
report.txt	1 KB
output.txt	1 KB
multiqc_report.html	1,123 KB
J877_trimstats.txt	0 KB
J877_phixstats.txt	1 KB
J877_matchedphix.fq	0 KB
J877_2_original_fastqc.zip	375 KB
J877_2_original_fastqc.html	620 KB
J877_2_clean_fastqc.zip	350 KB
J877_2_clean_fastqc.html	570 KB
J877_2.fq.gz	4,772 KB
J877_2.fastq.gz	7,057 KB
J877_1_original_fastqc.zip	374 KB
J877_1_original_fastqc.html	619 KB
J877_1_clean_fastqc.zip	357 KB
J877_1_clean_fastqc.html	574 KB
J877_1.fq.gz	4,743 KB
J877_1.fastq.gz	6,936 KB
J877.log	23,476 KB
J877_adapters.stats.txt	1 KB
intersect	3 KB
astats	21 KB
aln-se.sam	64,659 KB
aln-se.bam	41,046 KB

#rname	startpos	endpos	numreads	covbases	coverage	meandepth	meanbase	meanmapq
L43967.2	1	580076	69138	3561	0.613885	15.5046	36.9	60

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelia1220/output-20241220163752/J877/

Name	Size
..	
variants	
multiqc_data	
alignment	
report.txt	1 KB
output.txt	1 KB
multiqc_report.html	1,123 KB
J877_trimstats.txt	0 KB
J877_phixstats.txt	1 KB
J877_matchedphix.fq	0 KB
J877_2_original_fastqc.zip	375 KB
J877_2_original_fastqc.html	620 KB
J877_2_clean_fastqc.zip	350 KB
J877_2_clean_fastqc.html	570 KB
J877_2.fq.gz	4,772 KB
J877_2.fastq.gz	7,057 KB
J877_1_original_fastqc.zip	374 KB
J877_1_original_fastqc.html	619 KB
J877_1_clean_fastqc.zip	357 KB
J877_1_clean_fastqc.html	574 KB
J877_1.fq.gz	4,743 KB
J877_1.fastq.gz	6,936 KB
J877.log	23,476 KB
J877_adapters.stats.txt	1 KB
intersect	3 KB
astats	21 KB
aln-se.sam	64,659 KB
aln-se.bam	41,046 KB

sampleID	reference	start	end	num_raw_reads	num_clean_reads	num_mapped_reads	percent_mapped_clean_reads	mean_base_qual	mean_map_qual	23S_rRNA	L4	L22	gyrA	gyrB	parC	parE
J877	L43967.2	1	580076	301984	178424	69138	38.7493	36.9	60	3558,723.8,55.42,C173967T	2233,447.8,58.21,None	2001,493.5,70.89,T193158C	GTA:GCA V:A 918,454.8,100.00,None	926,339.4,100.00,None	2066,399.0,53.28,C242310T ATC:ATT I:I	10,0.1,12.23,None

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Amelia1220/output-20241220163752/J877/

Name	Size
..	
variants	
multiqc_data	
alignment	
report.txt	1 KB
output.txt	1 KB
multiqc_report.html	1,123 KB
J877_trimstats.txt	0 KB
J877_phixstats.txt	1 KB
J877_matchedphix.fq	0 KB
J877_2_original_fastqc.zip	375 KB
J877_2_original_fastqc.html	620 KB
J877_2_clean_fastqc.zip	350 KB
J877_2_clean_fastqc.html	570 KB
J877_2.fq.gz	4,772 KB
J877_2.fastq.gz	7,057 KB
J877_1_original_fastqc.zip	374 KB
J877_1_original_fastqc.html	619 KB
J877_1_clean_fastqc.zip	357 KB
J877_1_clean_fastqc.html	574 KB
J877_1.fq.gz	4,743 KB
J877_1.fastq.gz	6,936 KB
J877.log	23,476 KB
J877_adapters.stats.txt	1 KB
intersect	3 KB
astats	21 KB
aln-se.sam	64,659 KB
aln-se.bam	41,046 KB

23S_rRNA	L4	L22	gyrA	gyrB	parC	parE	
3558,723.8,55.42,C173967T	2233,447.8	2001,493.5	918,454.8,	926,339.4,	2066,399.6	10,0.1,12.23,	None

Conclusion



Fundamentals of Amelia



Installation and setup of
Amelia in HPG



Successfully executed
Job query for Amelia



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