Advanced Molecular Detection
Southeast Region Bioinformatics

Analysis Options
January 5, 2023

Outline

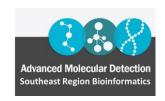- Current Analysis Options
- Pipelines
- Scripts & Individual tools
- Analysis support

Current Analysis Options

Scripts and pipelines available for state PHL use on HiPerGator

Pipelines

- FLAQ-AMR (FLAQ - Antimicrobial Resistance)
  - Generates annotated, de novo assemblies and quality metrics for WGS of bacterial species
  - Determines species ID
  - Determines ST using MLST schemes from PubMLST
  - Identifies AMR genes, virulence genes, and plasmids
  - Performs serotyping of *Salmonella* and *E. coli*, if applicable (more species to come)

Pipelines for Comparative Genomics

- FL-cgSNP (Core-Genome SNP analysis) – reference-free method
    - Pan-genome analysis (identifies core genes shared by all isolates) and generates a multiple sequence alignment
    - Identifies pair-wise SNPs between isolates and outputs a pairwise SNP matrix
    - Generates a maximum-likelihood phylogenetic tree
- hqSNP (High-Quality SNP analysis) – reference-based method
    - Uses CDC's Lyveset pipeline to identify hqSNPs
    - Generates a pairwise SNP matrix and maximum-likelihood phylogenetic tree
- Custom script to output annotated tree (based on metadata input)

Scripts and Individual Tools

- Quick Species ID (screen against RefSeq database)

- Species ID, contamination check, and metagenomic classification

- Run CDC's Lp Species ID Tool and Lp Serotyping Tool locally

- Pull out gene sequence of interest from assembly (e.g., pull out AR genes of interest)

- Merge fastqs from multiple lanes on a NextSeq into one R1 and one R2 file (compatible with BioNumerics)

Scripts and Individual Tools

- Download fastqs from NCBI's Sequence Read Archive

- Run any tool individually
  - SeqSero2, mlst, abricate, etc.

- Run any tool or workflow in the Staph-B Toolkit (https://staph-b.github.io/staphb_toolkit/)

- Batch runs (i.e., run >2 analysis scripts at one time)

- **Custom scripts and pipelines as requested or needed**

Viral Pipelines and scripts

- FLAQ - SC2 (FLAQ – SARS-CoV-2)
  - Generates SARS-CoV-2 consensus assemblies from ARTIC V1, V2, or V3, V4 and V4.1 targeted amplicon sequencing using Illumina (e.g., Nextera XT or Flex) and non-Illumina (e.g., PrimalSeq or MN Tailed) library prep
  - Outputs variant file and final report with quality metrics (including a PASS/FAIL quality flag based on public repository submission criteria)
  - Automatically runs NCBI's VADR tool for viral annotation and generates PASS/REVIEW flags to facilitate sample batch and review prior to submission.
- SC2 associated scripts to prepare/format assemblies for batch submissions to GISAID and NCBI's Genbank
- SC2-Correct-Assembly
  - Removes indels/SNPs that are likely PCR or sequencing artifacts/errors (requires prior manual review of mapped reads in IGV (or other program)

Viral Pipelines and scripts

- Targeted Amplicon Variant Calling and Consensus Sequence Generation
    1. For example, identify variants and generate a consensus sequence for HIV genotyping.
    2. Will work for any target/amplicon (just need a reference sequence)

- Flaq_sc2_meta

    1. SARS-CoV-2 metagenomics pipeline developed primarily for wastewater samples.

    2. Automatically runs the Freyza tool to estimate lineage populations.

Analysis Request and/or support

- Email **bphl16bioinformatics@flhealth.gov**

- One-on-one video conference sessions to set up your HiPerGator environment and walk-through the use of each pipeline/script needed

- Custom pipeline/script development as requested

- All pipelines & scripts will be shared via github or the public- share directory in HiPerGator via /blue/bphl-<state>/public-share/ directory. BPHL-Molecular · GitHub

- All pipelines are currently being converted to Nextflow workflows in the near future.

bphl16bioinformatics@flhealth.gov

Lakshmi Thsaliki, MS

Bioinformatician

Lakshmi.Thsaliki@flhealth.gov

Sarah Schmedes, PhD

Lead Bioinformatician

Sarah.Schmedes@flhealth.gov

Questions???