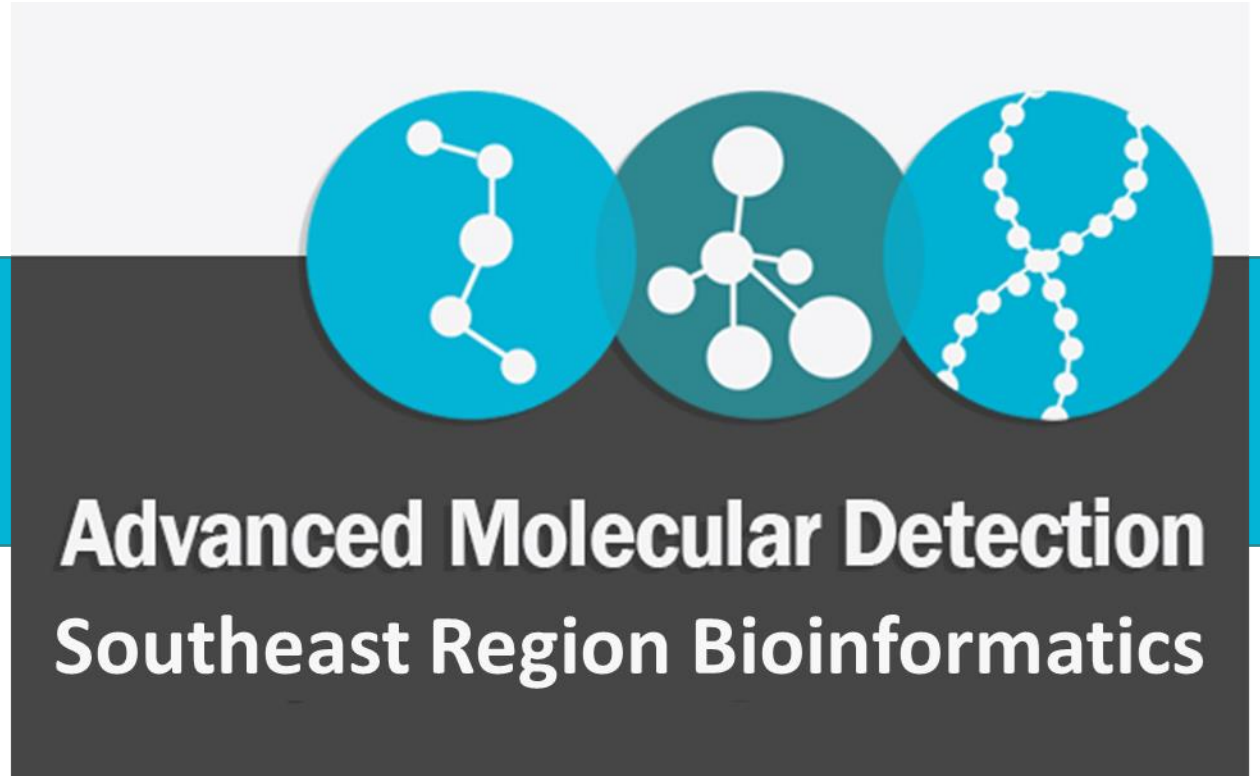


Bactopia tools

12/11/2023

Outline

Agenda

Updates

Shigatyper

Shigeifinder

Questions

Agenda

January 08 – Bactopia Tools: spatyper and ssuissero
January 22 – Bactopia Pipeline

Future Trainings

- ONT & FL's Flisochar pipeline
- StaPH-B Toolkit Programs/Pipelines
- GISAID flagged SARS-CoV-2
- R Training Series
- Dryad pipeline
- …and more

Advanced Molecular Detection
Southeast Region Bioinformatics

Updates

- AMD Southeast Region Annual Needs Assessment Survey is sent out last Tuesday. Please make sure you finish the survey before **December 22nd**
- Feel free to forward the survey to any new staff member not included in this email
- Also, please have at least one staff member respond back to make any updates to your lab AMD contact list

Shigatyper

- ShigaTyper is a quick and easy tool designed to determine *Shigella* serotypes using Illumina (single or paired-end), or Oxford Nanopore reads with low computation requirement
- ShigaTyper supports compressed FASTQs as inputs
  - FASTQs can be single-end or paired-end Illumina reads, or reads from Oxford Nanopore

[CFSAN-Biostatistics/shigatyper: CFSAN Shigella Typing Pipeline (github.com)]

Advanced Molecular Detection
Southeast Region Bioinformatics

Installation

Can be installed through conda

```
conda create –yp /blue/bphl-<state>/<user>/conda_envs/shigatyper/
conda activate /blue/bphl-<state>/<user>/conda_envs/shigatyper/
conda install –c conda-forge –c bioconda shigatyper
```

Advanced Molecular Detection
Southeast Region Bioinformatics

Usage

```
usage: shigatyper [-h] [--R1 FASTA] [--R2 FASTA] [--SE FASTA] [--ont] [-n SAMPLE_NAME] [-o OUTDIR] [--verbose] [--version]

ShigaTyper v. 2.0.5, 2022

A WGS-based genoserotyping pipeline for Shigella spp.

Yun Wu, Henry K Lau, Teresa Lee, David K Lau, Justin Payne

    The bacteria Shigella spp., consisting of 4 species and >50
serotypes, cause shigellosis, a foodborne disease of significant
morbidity, mortality, and economic loss worldwide. Classical Shigella
identification based on selective media and serology is tedious,
time-consuming, expensive, and not always accurate. Molecular diagnostic
assay does not distinguish Shigella at species level or from
enteroinvasive Escherichia coli (EIEC). We inspected the whole genome
sequencing (WGS) data from 219 Shigella isolates and observed low
concordance rate between conventional designation and molecular
serotyping, 86.8% and 78.9% at species and serotype level, respectively.
Serotype determinants for 6 additional serotypes were identified.
Examination of differentiation gene markers commonly perceived as
characteristic hallmarks in Shigella showed high variability among
different serotypes. Using this information, we developed ShigaTyper, an
automated workflow that utilizes limited computational resources to
accurately and rapidly determine 58 Shigella serotypes using Illumina
paired end WGS reads. Shigella serotype determinants and species-specific
```

Advanced Molecular Detection
Southeast Region Bioinformatics

Input

Takes paired-end reads, single-end reads, or Oxford Nanopore reads as input

```
shigatyper --SE JBI22000647.fasta > shigatyper_results
```

Advanced Molecular Detection
Southeast Region Bioinformatics

Results

	Hit	Number of reads	Length Covered	reference length	% covered	Number of variants	% accuracy
0	ipaH_c	1	717	780	91.9	0	100
1	Sf_wzx	1	0	1257	0	0	
2	Sf_wzy	1	0	1149	0	0	
3	gtrI	1	0	1521	0	0	
4	Oac1b	1	0	1002	0	0	

Advanced Molecular Detection
Southeast Region Bioinformatics

ShigEifinder

- This tool differentiates *Shigella*/EIEC using cluster-specific genes and identify the serotype using O-antigen/H-antigen genes
- This pipeline can serotype over 59 *Shigella* and 22 EIEC using either assembled genomes or WGS reads

[LanLab/ShigEiFinder: Cluster informed Shigella and EIEC serotyping tool from Illumina reads and assemblies (github.com)]

Advanced Molecular Detection
Southeast Region Bioinformatics

Installation

Can be installed through conda

```
conda create -yp /blue/bphl-<state>/<user>/training/conda_envs/shigeifinder/
conda activate /blue/bphl-<state>/<user>/training/conda_envs/shigeifinder/
conda install -c conda-forge -c bioconda shigeifinder
```

Advanced Molecular Detection
Southeast Region Bioinformatics

Usage

```
usage:
Assembly fasta input/s:
 ShigeiFinder.py -i <input_data1> <input_data2> ... OR
 ShigeiFinder.py -i <directory/*>
Paired end raw read fastq(.gz) input/s:
 ShigeiFinder.py -r -i <Read1> <Read2> OR
 ShigeiFinder.py -r -i <directory/*>
Single end raw read fastq(.gz) input/s:
 ShigeiFinder.py -r --single_end -i <Reads> OR
 ShigeiFinder.py -r --single_end -i <directory/*>

options:
  -h, --help            show this help message and exit
  -i I [I ...]          <string>: path/to/input_data
  -r                    Add flag if file is raw reads.
  -t T                  number of threads. Default 4.
  --single_end          Add flag if raw reads are single end rather than paired.
  --hits                To show the blast/alignment hits
  --dratio              To show the depth ratios of cluster-specific genes to House Keeping genes
  --update_db           Add flag if you added new sequences to genes database.
  --output OUTPUT       output file to write to (if not used writes to stdout)
  --check               To show the blast/alignment hits
  --o_depth O_DEPTH     When using reads as input the minimum depth percentage relative to genome average for positive O antigen gene call (default 1.0).
  --ipaH_depth IPAH_DEPTH
                        When using reads as input the minimum depth percentage relative to genome average for positive ipaH gene call (default 1.0).
  --depth DEPTH         When using reads as input the minimum read depth for non ipaH/Oantigen gene to be called (default 10.0).
  --tmpdir TMPDIR       temporary folder to use for intermediate files
  --noheader            do not print output header
  -v, --version         Print version information.
```

Advanced Molecular Detection
Southeast Region Bioinformatics

Input

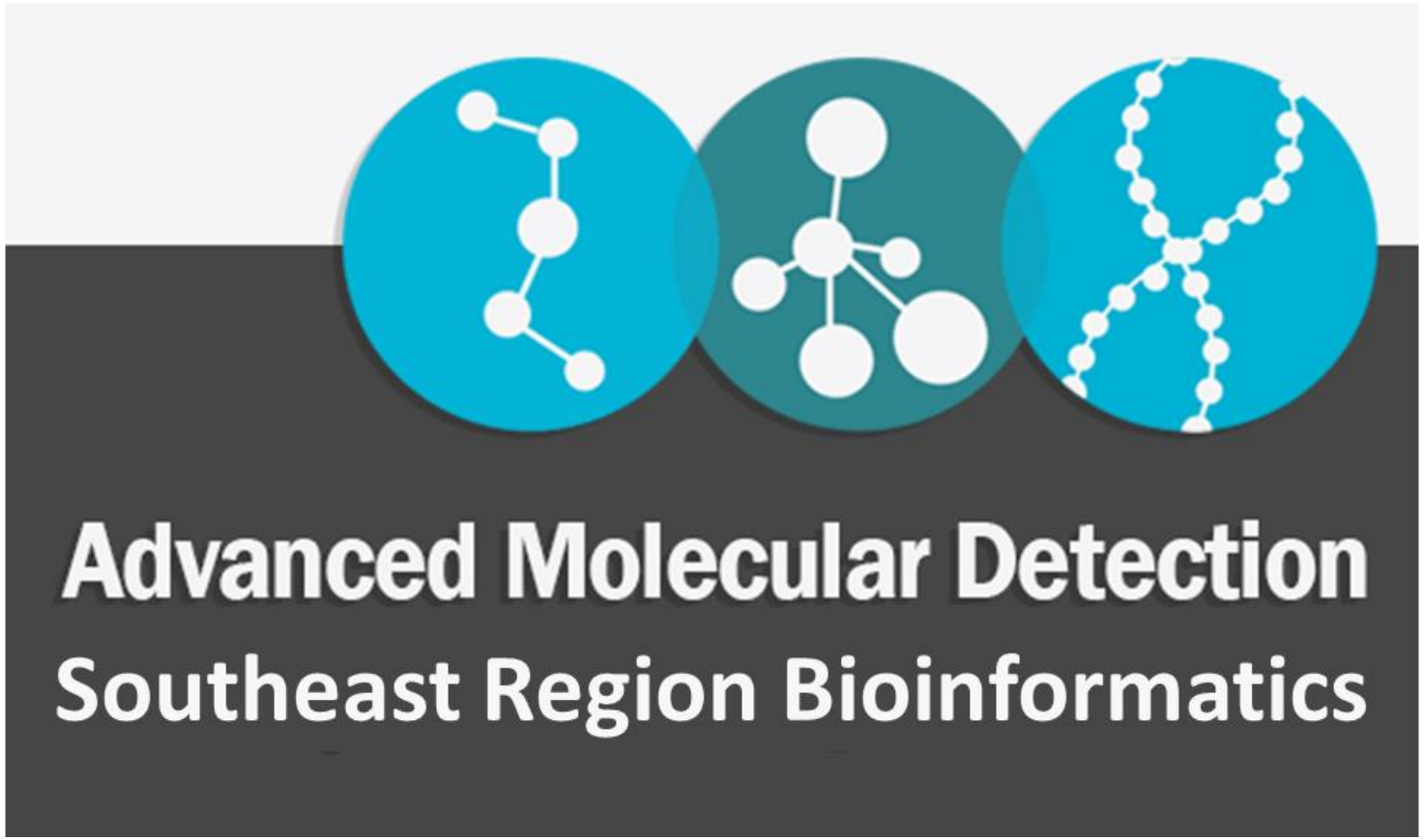
Takes fasta files as input

```
$ shigeifinder -i JBI22000648.fasta > shigeifinder_results
```

Advanced Molecular Detection
Southeast Region Bioinformatics

Results

1	#SAMPLE	ipaH	VIRULENCE	CLUSTER	SEROTYPE	O_ANTIGEN	H_ANTIGEN	NOTES
2	JBI22000648	+	38	C3	SF1b	SF1-5		



Questions?

bphl-sebioinformatics@flhealth.gov

Lakshmi Thsaliki, MS
Bioinformatician
Lakshmi.Thsaliki@flhealth.gov

Molly Mitchell, PhD
Bioinformatician
Molly.Mitchell@flhealth.gov