



Advanced Molecular Detection

Southeast Region Bioinformatics

Outline



Updates



Agenda



Bactopia



Hicap



HpsuisSEro



Questions

Updates –ABiL Trainings

ABiL in-person courses

- The ABiL in-person (advanced!) bioinformatics course will be held **Tuesday Oct. 24th - Friday Oct. 27th** at the Georgia Institute of Technology (GT)
- About 2 participants may attend from each state or jurisdiction
- Tanya has sent an email about the courses
- Email us ASAP if you're interested to guarantee a spot!

ABiL online courses

- If additional attendees decide to sign up later, that is not a problem, as the courses will still be available, and they can attend on a rolling basis under the contract
- If you did not receive the invitation for joining the courses, please let us know, we can help to get you enrolled

Updates – PHoeNix v2.0.2!

- New changes
 - <https://github.com/CDCgov/phoenix>
 - Run the newest version [here](#)
- Access the webinar reviewing these changes using [this link](#)
- View the change log
 - [CHANGELOG.md](#)

Agenda

September 18 – Bactopia Tools: Kleborate and Legsta

October 2 – Bactopia Tools: LisSero and Meningotype

Future Trainings

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

Hicap

- Identify cap locus serotype and structure in *Haemophilus influenzae* assemblies
- The cap locus of *H. influenzae* are categorized into 6 different groups based on serology (a-f). Cap locus has 3 functionally distinct regions namely region 1, region 2, and region 3
- Genes with region 1 (bexABCD) & 3 (hcsAB) are associated with transport and post-translation modification
- Region 2 genes encode serotype specific proteins, with each serotype (a-f) having a distinct set of genes
- hicap automates identification of the cap locus, describes the structural layout, and performs in-silico serotyping

[scwatts/hicap](https://github.com/scwatts/hicap): 🤝🎓 in silico typing of the *H. influenzae* cap locus ([github.com](https://github.com/scwatts/hicap))



Installation

Can be installed through conda

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

Usage

```
(/blue/bph1-florida/thsalikilakshmi/training/conda_envs/hicap) [thsalikilakshmi@login2 assemblies]$ hicap --help
usage: hicap -q QUERY_FP -o OUTPUT_DIR [-v] [-h] [--help_all]

File input and output:
  -q QUERY_FP, --query_fp QUERY_FP      Input FASTA query
  -o OUTPUT_DIR, --output_dir OUTPUT_DIR Output directory

Other:
  -v, --version                        Show version number and exit
  -h, --help                          Show this help message and exit
  --help_all                          Display extended help
```



Input

Requires an input genome assembly and an output directory for result files

```
# Create output directory and run serotype prediction
mkdir -p output/
hicap --query_fp input_genome.fasta --output_dir output/
```

Input

```
(/blue/bphl-florida/thsalikilakshmi/training/conda_envs/hicap) [thsalikilakshmi@login2 assemblies]$ hicap --query_fp /blue/bphl-florida/thsalikilakshmi/data/HAI/20230510_jax_230214_PLN_WAT_JD/assemblies/JBI23000022.fasta --output hicap_out/
15/06/2023 15:13:19 INFO: Checking dependencies
15/06/2023 15:13:19 WARNING: One or more contig names exceed the genbank spec limit of 20 characters. These will be truncated in the genbank output file
15/06/2023 15:13:19 INFO: Collecting ORFs from input assembly
15/06/2023 15:13:20 INFO: Searching for ORF hits in database
15/06/2023 15:13:47 INFO: Finding complete ORF hits
15/06/2023 15:13:48 INFO: Finding truncated ORFs
15/06/2023 15:13:48 INFO: Predicting serotype
15/06/2023 15:13:48 INFO: Searching for ORF fragments
15/06/2023 15:13:48 INFO: Searching for IS1016 sequences
15/06/2023 15:13:48 INFO: Collecting ORFs proximal to capsule locus
15/06/2023 15:13:48 INFO: Writing outputs
```



Output

- Analysis using hicap will generate three result files in the specified output directory:
 - Summary: a somewhat machine parsed file with detailed summary information
 - Genbank: a genbank file with the sequence marked up with cap locus annotations
 - Graphic: a visual representation of the annotated cap locus

Result

	A	B	C	D	E	F	G	H	I
1	#isolate	predicted_serotype	attributes	genes_identified	locus_location	region_I_genes	region_II_genes	region_III_genes	IS1016_hits
2	JBI23000022	type_a	full_gene_complement	hcsB,hcsA,acs4,acs1	length=538791	4-Apr	4-Apr	2-Feb	2

HpsuisSero

- Rapid *Haemophilus parasuis* serotyping pipeline for Nanopore data
- This pipeline is designed to rapidly infer *Haemophilus parasuis* serotype from Oxford Nanopore data by first assembling a draft genome using Flye followed by genome polishing with Medaka
- The processed assembly is subsequently queried against the Cps BLAST database to determine isolate serotype

[jimmyliu1326/HpsuisSero: Rapid Haemophilus parasuis serotyping pipeline for Nanpore data \(github.com\)](https://github.com/jimmyliu1326/HpsuisSero)

Installation

Can be installed through conda

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

Usage

```
(/blue/bphl-florida/thsalikilakshmi/training/conda_envs/hpsuissero) [thsalikilakshmi@login2 as  
semblies]$ HpsuisSero.sh --help  
  
Usage: /blue/bphl-florida/thsalikilakshmi/training/conda_envs/hpsuissero/bin/HpsuisSero.sh  
  
Required arguments:  
-i input file  
-o path to output directory  
-s sample name  
-x input type [fasta or fastq]  
  
Optional arguments:  
-h|--help display help message  
-t|--threads number of threads [Default: 4]
```



Input

Takes *.fastq* or *.fasta* as input

```
(/blue/bphl-florida/thsalikilakshmi/training/conda_envs/hpsuissero) [thsalikilakshmi@login2 as  
semblies]$ HpsuisSero.sh -s JBI22001252.fasta -i /blue/bphl-florida/thsalikilakshmi/data/HAI/2  
0230510_jax_230214_PLN_WAT_JD/assemblies/JBI22001252.fasta -o hpsuissero_out -x fasta  
Pipeline Finished!
```


Results



The screenshot shows a terminal window with a file path in the title bar and a table of results below a toolbar. The file path is `/blue/bphl-florida/thsalikilakshmi/data/HAI/20230510_jax_230`. The table has two columns: `Sample_Name` and `Serotype`. The first row of data shows `JB122001252.fasta` and `No Hits`.

Sample_Name	Serotype
JB122001252.fasta	No Hits



Advanced Molecular Detection

Southeast Region Bioinformatics

Questions?

bphl-sebioinformatics@flhealth.gov

Lakshmi Thsaliki, MS

Bioinformatician

Lakshmi.Thsaliki@flhealth.gov

Molly Mitchell, PhD

Bioinformatician

Molly.Mitchell@flhealth.gov