



Advanced Molecular Detection

Southeast Region Bioinformatics

Outline



Agenda



Pbptyper



Prokka



Questions

Agenda

November 27 – Bactopia Tools: rgi (resistance gene identifier) and Seqsero2

December 11 – Bactopia Tools: shigatyper and shigeifinder

Future Trainings

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

Pbptyper

- pbptyper is a tool to identify the Penicillin Binding Protein (PBP) of *Streptococcus pneumoniae* assemblies
- Using an input assembly (uncompressed or gzip-compressed), the PBP 1A, 2B, and 2X proteins are blasted against the assembly from which a PBP type is predicted

[rpetit3/pbptyper: In silico Penicillin Binding Protein \(PBP\) typer for Streptococcus pneumoniae assemblies \(github.com\)](https://github.com/rpetit3/pbptyper)

Installation

Can be installed through conda

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

Usage

```
(/blue/bph1-florida/thsalikilakshmi/training/conda_envs/pbtyper) [thsalikilakshmi@login6 assemblies]$ pbtyper --help
```

Usage: pbtyper [OPTIONS]

In silico Penicillin Binding Protein (PBP) typer for Streptococcus pneumoniae assemblies

```
â Options â
â --version          Show the version and exit.
â * --assembly       TEXT      Input assembly in FASTA format (gzip is OK) [required]
â --db               TEXT      Path to the input database [default: /blue/bphl-florida/thsalikilakshmi/training/conda_envs/pbptyper/share/pbptyper-1.0.4]
â --prefix           TEXT      Prefix to use for output files [default: basename of input]
â --outdir           TEXT      Directory to save output files [default: ./]
â --min_pident       INTEGER    Minimum percent identity to count a hit [default: 95]
â --min_coverage     INTEGER    Minimum percent coverage to count a hit [default: 95]
â --min_ani          INTEGER    Minimum S. pneumoniae ANI to predict PBP Type [default: 90]
â --check            Check dependencies are installed, then exit
â --quiet            Suppress all output
â --help             Show this message and exit.
```

Input

```
$ pbptyper --assembly JBI23001197.fasta
```

Results

1	sample	pbptype	ani	1A_coverage	1A_pident	1A_bitsco	2B_coverage	2B_pident	2B_bitsco	2X_coverage	2X_pident	2X_bitsco	comment
2	JB12300119	2:00:00	98.79	100	100	556	100	100	567	100	100	741	

This file will contain the final predicted PBP type based on highest coverage, percent identity, and bitscore

Prokka

- Whole genome annotation is the process of identifying features of interest in a set of genomic DNA sequences, and labelling them with useful information
- Prokka is a software tool to annotate bacterial, archaeal and viral genomes quickly and produce standards-compliant output files

[tseemann/prokka: :zap: Rapid prokaryotic genome annotation \(github.com\)](https://github.com/tseemann/prokka)

Installation

- Available as module on HPG

```
module load prokka/1.14.6
```

- Can be installed through conda

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

Usage

```
[thsalikilakshmi@login6 data]$ prokka --help
Name:
  Prokka 1.14.6 by Torsten Seemann <torsten.seemann@gmail.com>
Synopsis:
  rapid bacterial genome annotation
Usage:
  prokka [options] <contigs.fasta>
General:
  --help          This help
  --version       Print version and exit
  --citation      Print citation for referencing Prokka
  --quiet         No screen output (default OFF)
  --debug         Debug mode: keep all temporary files (default OFF)
Setup:
  --dbdir [X]     Prokka database root folders (default '/data/reference/prokka/1.14.6/db')
  --listdb        List all configured databases
  --setupdb       Index all installed databases
  --cleandb       Remove all database indices
  --depends        List all software dependencies
Outputs:
  --outdir [X]    Output folder [auto] (default '')
  --force         Force overwriting existing output folder (default OFF)
  --prefix [X]    Filename output prefix [auto] (default '')
  --addgenes      Add 'gene' features for each 'CDS' feature (default OFF)
  --addmrna       Add 'mRNA' features for each 'CDS' feature (default OFF)
  --locustag [X]  Locus tag prefix [auto] (default '')
  --increment [N] Locus tag counter increment (default '1')
  --gffver [N]    GFF version (default '3')
```



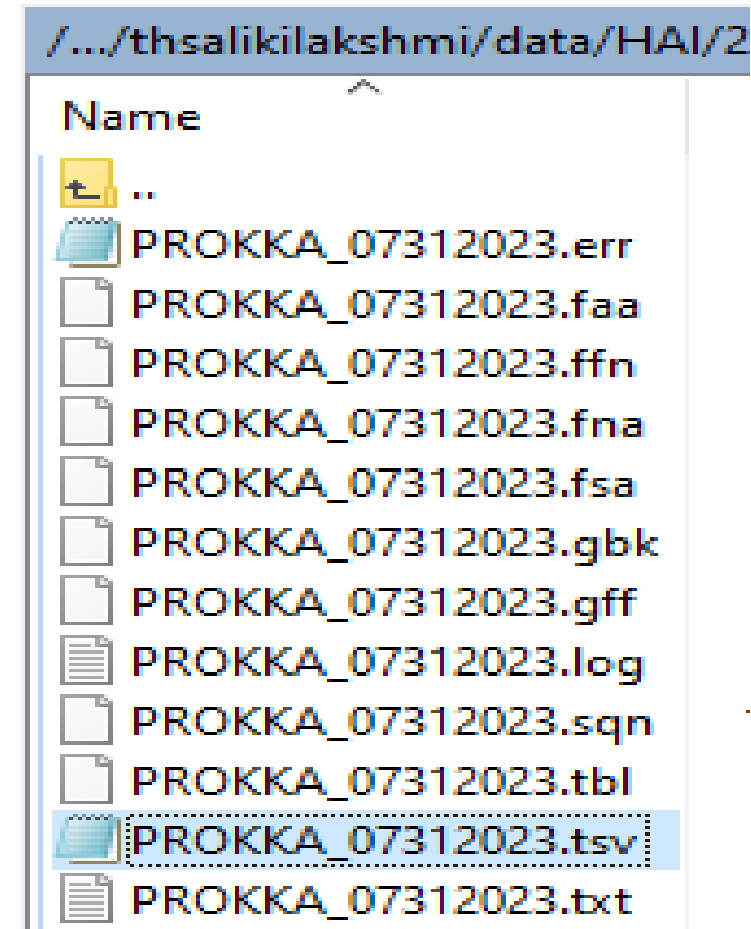
Input

Takes .fasta as input

```
$ prokka JBI22000793.fasta > prokka_results
```

Output files

It outputs the following files



Results

This is how prokka .tsv file looks:

1	locus_tag	ftype	length_bp	gene	EC_number	COG	product						
2	GOPHEEIJ_00001	CDS	513	yrdA_1		COG0663	Protein YrdA						
3	GOPHEEIJ_00002	CDS	792	ytpA	3.1.1.-	COG2267	Phospholipase YtpA						
4	GOPHEEIJ_00003	CDS	684	purC_1	6.3.2.6	COG0152	Phosphoribosylaminoimidazole-succinocarboxamide synthase						
5	GOPHEEIJ_00004	CDS	1314		1.1.1.38		NAD-dependent malic enzyme						
6	GOPHEEIJ_00005	CDS	543				hypothetical protein						
7	GOPHEEIJ_00006	CDS	1083		2.3.1.-	COG3424	Alpha-pyrone synthesis polyketide synthase-like Pks11						
8	GOPHEEIJ_00007	CDS	570				hypothetical protein						
9	GOPHEEIJ_00008	CDS	960				hypothetical protein						
10	GOPHEEIJ_00009	CDS	1170	mdrP_1			Na(+), Li(+), K(+)/H(+) antiporter						
11	GOPHEEIJ_00010	CDS	1344	norR_1			Anaerobic nitric oxide reductase transcription regulator NorR						
12	GOPHEEIJ_00011	CDS	1329				hypothetical protein						
13	GOPHEEIJ_00012	CDS	768	bdhA_1	1.1.1.30		D-beta-hydroxybutyrate dehydrogenase						
14	GOPHEEIJ_00013	CDS	690	scoA_1	2.8.3.5	COG1788	putative succinyl-CoA:3-ketoacid coenzyme A transferase subunit A						
15	GOPHEEIJ_00014	CDS	654	scoB_1	2.8.3.5	COG2057	Succinyl-CoA:3-ketoacid coenzyme A transferase subunit B						
16	GOPHEEIJ_00015	CDS	2412	leuS	6.1.1.4		Leucine--tRNA ligase						
17	GOPHEEIJ_00016	CDS	573	pabA_1	2.6.1.85	COG0512	Aminodeoxychorismate/anthranilate synthase component 2						
18	GOPHEEIJ_00017	CDS	1389	trpE	4.1.3.27	COG0147	Anthranilate synthase component 1						



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