Laboratory Update

Cody Glickman

cody.glickman@ucdenver.edu

May 11th 2018

Contents

May 2nd, 2018

1 Genomic Retrieval and BLAST Database Creation

Updated databases to include Strain level searches. Waiting on ISME Travel Grant Status (Mid-May)
BioRxiv: Resupply manuscript to MS ID 246553
Comments: Include data about functionality compared to other tools or show with an example dataset.
To-Do: Application for Mac, Windows, Linux using Tkinter in Python
Found Github Repository for Something Similar

2 Viral Contig Identification Study

Environments have been set-up and are ready to roll. The binner (MaxBin2) requires the raw reads to calculate abundances needed for binning. The CAMI data website contains dead links to retrieve reads. Will need to reach out to authors for data.

3 Identifying Virulence Factors in Phages

Have established baseline prevalence of virulence genes in phage genomes (See attached HTML). Focused on phages with more than 30 complete genomes in reference database (13 Genus types). Completed BLAST and HMM searches of Virulence factors against metagenomic contigs inferred from vHMM study.
Next Steps

- Identify taxonomy of contigs with virulence factors

- Gather study metadata

- Compare metadata to significance of VFs

4 Creation of rpoB, HSP65, and 16S Databases

Data gathered from locations below, next steps are to make this into a virtualbox with instructions to allow for technician usage.
Downloading rpoB data from NCBI Nucleotide (Copy and Paste into Search Window)

1780 Sequences as of May 1st, 2018

May 2nd, 2018

```
(("Mycobacterium"[Organism] OR ("Mycobacterium"[Organism]
OR Mycobacterium[All Fields])) AND rpoB[Title]) AND
(bacteria[filter] AND biomol_genomic[PROP] AND
ddbj_embl_genbank[filter] AND ("500"[SLEN] : "5000"[SLEN]))
```

Downloading HSP65 data from NCBI Nucleotide (Copy and Paste into Search Window)

1710 Sequences as of May 1st, 2018

```
(("Mycobacterium"[Organism] OR ("Mycobacterium"[Organism]
OR Mycobacterium[All Fields])) AND hsp65[Title]) AND
(bacteria[filter] AND biomol_genomic[PROP] AND
ddbj_embl_genbank[filter] AND ("400"[SLEN] : "5000"[SLEN]))
```

Downloading Silva-ARB Database: download SilvaSSUParctaxsilvatrunc.fasta.gz

See Silva Release information

5 Classifying predicted prophages as active or degraded

Concluded three types of genes indicate lysogenic life cycle (see below). Protein HMMs for these three types have been obtained through PFAM. Additional HMMs may need to be developed with sequenced viruses.

- Integrases
  Involved in phage insertion into host genome

- ParA - ParB - ParS
  Involved in extra chromosomal arrangement and replication

- Repressors
  Inhibits replication until stimulus

2

May 11th, 2018

1 Genomic Retrieval and BLAST Database Creation

Updated databases to include Strain level searches. ISME Travel Grant Status: Abstract Accepted (Word on Travel Grant Unclear)

2 Identifying Virulence Factors in Phages

Methodology Questions:
Gene count or percentage of genes
Virome Sampling Steps

- Identify taxonomy of contigs

- Predict genes and identify virulence factors

- Compare percentage in niche to baseline

3 Creation of rpoB, HSP65, and 16S Databases

SequenceServer is Awesome!
Data does comprise of identical reads with overlapping labels:
i.e. Mycobacterium massiliense strains containing the same sequence as M. abscessus

June 1st, 2018

1 Genomic Retrieval and BLAST Database Creation

Updated databases to include Strain level searches.
ISME Travel Grant Status: Abstract Accepted Progress and Publication Goals:

Hold (Incorporate with GUI and SequenceServer)
Winter 2018

2 Identifying Virulence Factors in Phages

NLM Presentation

Progress:
Baseline model completed, pipeline for clinical datasets completed, searching for clinical dataset with longer viral contigs

Publication Goals: July 2018

3 Creation of rpoB, HSP65, and 16S Databases

Creating a database of rpoB, hsp65, and reference mycobacteria with a user interface
Progress: Completed Publication Goals: June 2018

September 21st, 2018

1 Identifying Virulence Factors in Phages

Progress and Publication Goals:
Winter 2018 (Virology or Bioinformatics)
To-Do: - Remove general pfams - Need to complete NULL Model - Have performed Network Analysis (Move into Cytoscape)

2 GRAB Package + Sequenceserver

Progress and Publication Goals:
Spring 2019 (Software Announcements)
To-Do: - Add to Python Package - Integrate Package with SequenceServer

3 18S/16S Pipeline

Progress and Publication Goals:
Spring 2019 (Microbiome?)
To-Do: - Integrate Visualizations / Stats into Pipeline - Test with Mock Communities

4 Viral Contig Discovery Hybrid

Progress and Publication Goals:
Summer 2019 (Bioinformatics)
To-Do: - Train Machine Learning Models - Integrate VirSorter HMMs or Earth Virome Methdology

5 Creation of rpoB, HSP65, and 16S Databases

Creating a database of rpoB, hsp65, and reference mycobacteria with a user interface

Progress and Publication Goals:
Winter 2018 (ASM?)
To-Do: - Integrate Sequence Server Online