

---

# Laboratory Journal

---

Cody Glickman

[cody.glickman@ucdenver.edu](mailto:cody.glickman@ucdenver.edu)

Beginning 27 February 2018



# Contents

<b>Week of 26 February 2018</b>	<b>1</b>
1 Identifying Virulence Factors in Phages . . . . .	1
<b>Week of 5 March 2018</b>	<b>2</b>
1 Identifying Virulence Factors in Phages . . . . .	2
2 Lysogenic Pan Genome . . . . .	2
<b>Week of 12 March 2018</b>	<b>4</b>
1 Genomic Retrieval and BLAST Database Creation . . . . .	4
<b>Week of 19 March 2018</b>	<b>5</b>
1 Classifying predicted prophages as active or degraded . . . . .	5
2 Creation of rpoB, HSP65, and 16S Databases . . . . .	6
<b>Week of 26 March 2018</b>	<b>7</b>
1 Classifying predicted prophages as active or degraded . . . . .	7



# Week of 26 February 2018

## 1 Identifying Virulence Factors in Phages

Downloaded phage protein data searching by taxonomy from Uniprot/TrEMBL. Downloaded three datasets Claudioviruses, Ligamenvirales, and Unclassified. Downloaded viral contigs from vHMM [Earth Virome Project](#).

Metagenomic gene prediction through prodigal downloaded from conda "conda install prodigal"

```
prodigal -i mVGs_sequences_v2.fna -o my_genes -a my_proteins.faa -p meta
```

The myproteins.faa file contains the translated predicted genes. This set is applied to the VF HMM and similar to the other datasets, returned no hits with hmmsearch.

Creating virulence factor blast database to blast against viral contigs. Choosing a E value threshold:

**Goals** Finish written GRAB and Abstract for NLM Done (3/2/18)

```
makeblastdb -in Combined_VF.faa -dbtype prot -out Combined_VF -title "Combined_VF"
```

```
blastp -db /Users/stronglab2/blastdb/Combined_VF/Combined_VF -out results.txt  
-outfmt 6 -query phage_proteins.faa
```

---

```
## Hello World  
print(x)
```

---

# Week of 5 March 2018

## Goals

- Prophage Annotation and VF/ARG Pipeline (Wednesday)
- ML Pipeline Active Prophage
- CAMI Data With Conda for Reproducibility

## 1 Identifying Virulence Factors in Phages

Establishing the BASH pipeline

### 1. Prophage Prediction

Input: Contigs.fasta

Output: Prophage Zip Folder

### 2. Gene Prediction

Input: Sequences of Identified Prophages

Output: Protein Fasta

### 3. Virulence Factor Identification

Input: Protein Fasta

Output: Proteins called Virulence Factors

## 2 Lysogenic Pan Genome

Downloaded phage table from [PhageDB](#). Parse temperate phages from Graham Hatful's List and those that infect Mycobacterium. Save Genbank ID numbers as Numbers.txt in script PhageDBProcessing.Rmd.

Calling GBK Files from nuccore.

---

```
## Load GenBankIds and Remove Whitespace
with open("Numbers.txt") as f:
    content = f.readlines()
content = [x.strip() for x in content]

## Call Entrez for Genbank_Ids
```

```
for i in range(0,len(content)):
    handle = Entrez.efetch(db="nucleotide", id=content[i], rettype="gb")
    filename = 'genbank_files/' + content[i] + '.gbk'
    out_handle = open(filename, "w")
    out_handle.write(handle.read())
    out_handle.close()
    handle.close()
    print("Saved " + filename)
```

---

## Run Core Genome Analysis

Convert to GFF3

```
bp_genbank2gff3.pl --dir pathtofiles
```

Run Roary

```
roary -e --mafft -p 8 *.gff
```

# **Week of 12 March 2018**

## **1 Genomic Retrieval and BLAST Database Creation**

Submitted GRAB document to bioRxiv: received resubmission request

Resupply manuscript to MS ID 246553

Include data about functionality compared to other tools or show with an example dataset



# Week of 19 March 2018

## 1 Classifying predicted prophages as active or degraded

Met Monday with Graham Hatfull and members of his lab. Confirmed genes to predict lysogenic life cycle. Holins are hard to predict as they are transmembrane proteins and may have small amount of conservation. Holins may be better predicted by k-mer protein groups.

- Integrases  
Involved in phage insertion into host genome
- ParA - ParB - ParS  
Involved in extra chromosomal arrangement and replication
- Repressors  
Inhibits replication until stimulus

Additional Take-Aways: Portal genes may be targets of bacterial resistance to phages and Abscessus excised phages can be engineered to become lytic. Excised phages resemble cluster P or N, Abscessus infected by cluster A3

Review Kclust and MMseq2 for clustering sequences

Check for non synonymous mutations in presence of clustered gene

Installed MMSeqs2 for sequence clustering

Downloaded viral HMMs from [EggNog 4.5](#) — [VOGDB](#) — [pVOG](#)

### Process to Filter Viral Lysogeny HMMs from Pfam

Split the complete Pfam database and rename each file to the Pfam ID

```
python -c "import sys
for i, c in enumerate(sys.stdin.read().split('//')):
    open('out' + str(i), 'w').write(c)" < Pfam.hmm
```

Removed empty first space from Split:

```
for f in *; do grep . $f > $f.hmm; done;
```

Rename HMMs by Families Identifiers (3rd line of HMMs):

Week of 19 March 2018

```
for f in *.hmm;
do
output=$(awk 'NR == 3 {print $2}' "$f" | cut -f1 -d '.')
mv "$f" "$output".hmm
done
```

Filter HMMs by those present in list:

```
while read file
do
mv -v -i "$file".hmm /matched/
done < File_with_names.txt
```

Format filtered hmms for search against combined lysogenic profiles

*Note: Needed to add ending '/' to hmm files to hmmpress effectively*

Perform HMM Search and Create Table Separated

```
hmmsearch --tblout [table].txt [model].hmm [sequences].fasta > /dev/null

## Convert tblout to table
chmod +x convert_hmm_tblout_to_tab_seperated_table.sh

./convert_hmm_tblout_to_tab_seperated_table.sh -t [table].txt
```

## 2 Creation of rpoB, HSP65, and 16S Databases

Query:

```
((("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]))
```

Muscle Sequence aligner downloaded via bioconda

usage: More usage examples at [Drive5](#)

```
muscle -msf -in seqs.fa -out seqs.msf
```

# Week of 26 March 2018

## 1 Classifying predicted prophages as active or degraded

Thought: determine if focused kmer profile for mycobacteriophage would be ammendable.

```
for f in Pfam_Viral_HMMs/*.hmm; do filename=$(echo "${f%\%.*}"); hmmsearch --tblout $filename
```

Move output to new folder titled SearchOutput and move into active directory

```
for f in Search_Output/*.txt; do ./convert_hmm_tblout_to_tab_seperated_table.sh -t $f; done;
```

Testing Lytic Phage Density of HMMs

---

## Formulae and Media Recipes

---

# Media

## Media 1

<b>Compound</b>	<b>1L</b>	<b>0.5L</b>
Compound 1	10g	5g
Compound 2	20g	10g

Table 1: Ingredients in Media 1.

# Formulae

**Formula 1 - Pythagorean theorem**

$$a^2 + b^2 = c^2$$