

Clinical NTM Gene Databases
ooooooooo

Duobiome
ooooooo

Virulence Factors in Phages
oooooooooooooo

Hybrid Viral Contig Prediction
oooooo

Metagenomic Exploration the Sequel: Development of novel tools for viral and bacterial sequence analysis

Cody Glickman
CPBS Update Talk



Nov 12th, 2018

Clinical NTM Gene Databases
ooooooooo

Duobiome
ooooooo

Virulence Factors in Phages
oooooooooooooo

Hybrid Viral Contig Prediction
oooooo

Research Update

Clinical NTM Gene Databases

Submitted ... <https://mra.asm.org/latest>

Duobiome: 18S/16S Parallel Analysis

In progress

Hybrid Viral Contig Prediction

In progress

Virulence Factors in Bacteriophages

Submitted ...

Clinical NTM Gene Databases
ooooooooo

Duobiome
ooooooo

Virulence Factors in Phages
oooooooooooooo

Hybrid Viral Contig Prediction
oooooo

Progress of Other Projects

Asthma Environmental Microbiome

Submitted abstract to ATS

Building Up Domains: Lysogenic Host Discovery

Incorporated into large collaborative NCBI initiative

Genomic Retrieval and Blast Database Creation

Accepted Poster ISME 2018

Hawaiian Soil Chemistry and Culture

Submitted ...

Nontuberculous Mycobacterial (NTM) Infections

Number of Cases

The number of NTM cases is estimated over 100K

Increasing Case

The rate of cases is estimated to grow at 8% every year

Populations at risk of developing NTM

- Immunocompromised individuals
- Patients with lung damage or malfunction
- Residents of warm costal areas especially Hawaii

Strollo SE, et al. Ann Am Thorac Soc. 2015
Adjemian J, et al. Am J Respir Crit Care Med. 2012

Viral Focus

Bacteriophages (Phages)

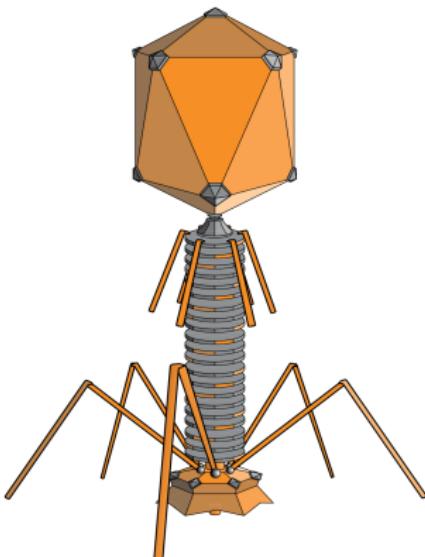
Phages are DNA viruses that infect prokaryotes

Phage Diversity

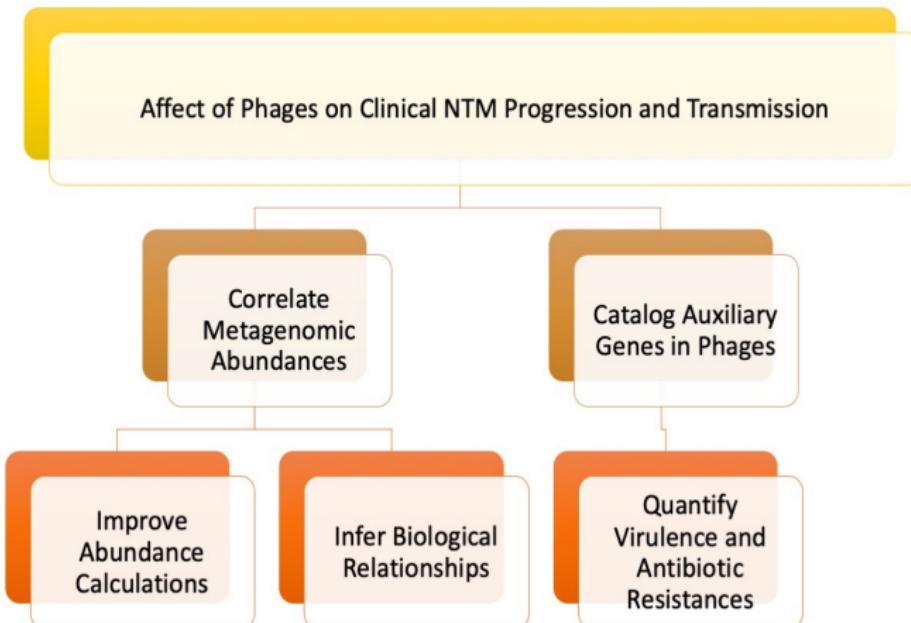
Investigating how phage abundance and diversity affect susceptibility to NTM lung infection

Phage Vectors

Researching how phages act as carriers of bacterial genes within clinical NTM infections



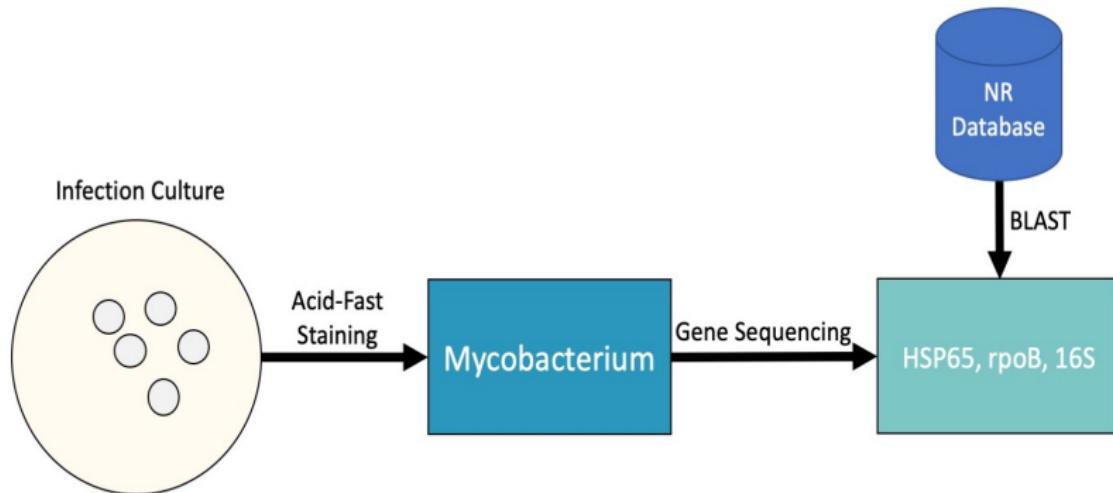
Objective



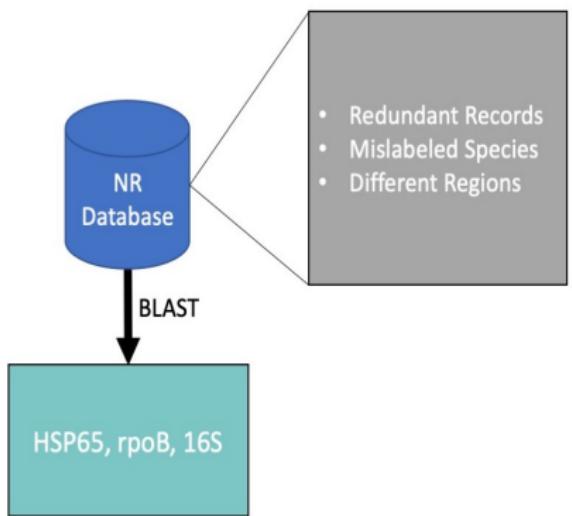
Species Identification of NTM at NJH

Clinical NTM Gene Database

Developed updated database to characterize clinical NTM



Limitations of Current Methods



Redundant Records

Sequences between species are indistinguishable at the gene level

Mislabelled Species

Naming conventions are constantly updated

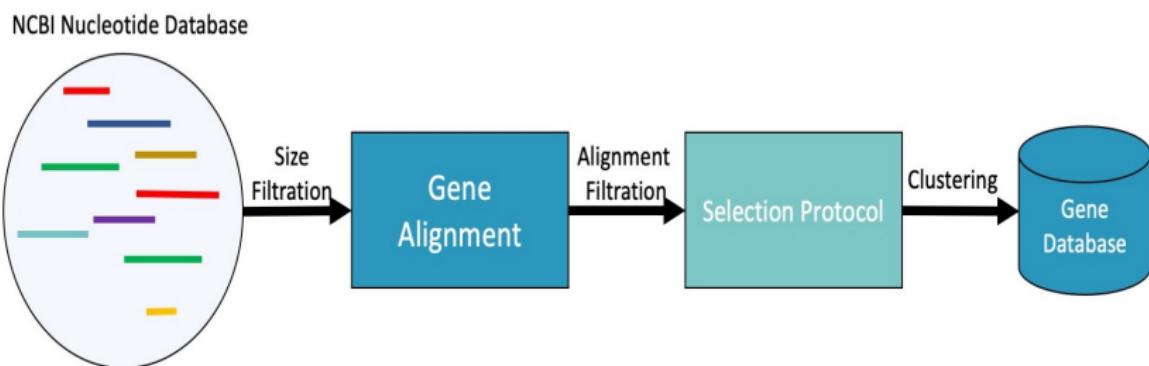
Different Regions

Current protocols amplify a specific region of gene

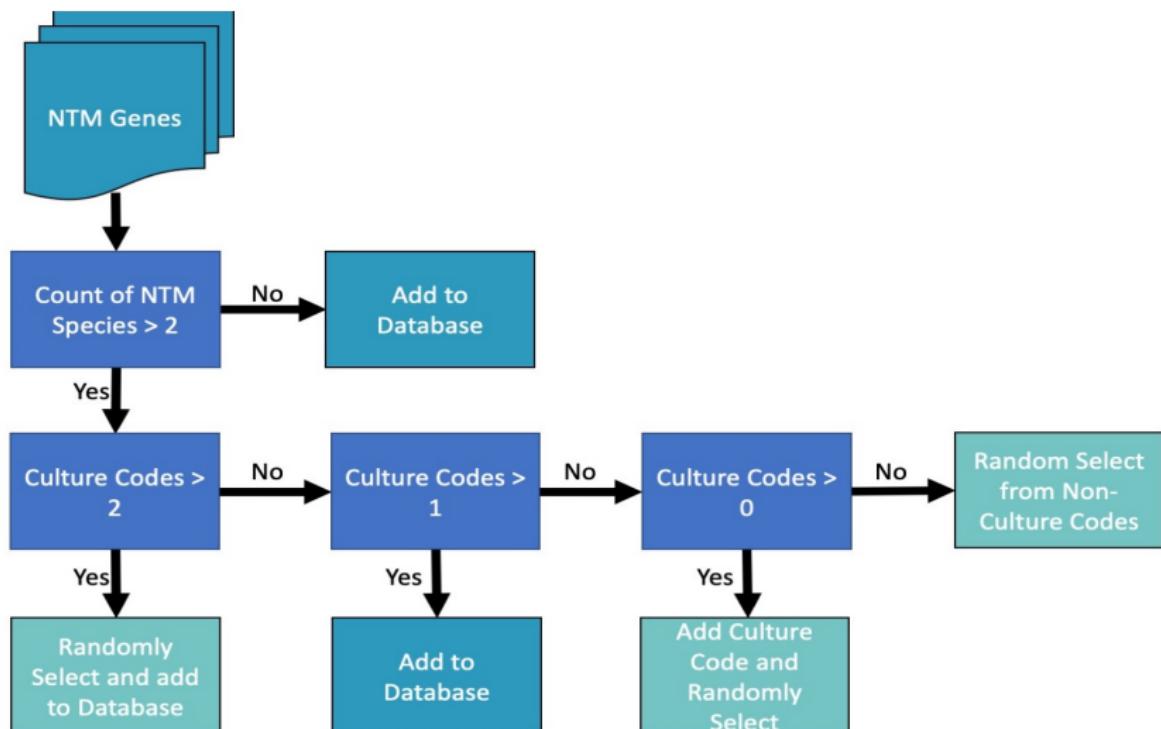
Curated Gene Databases

Number of Sequences per Species

Currently only two sequences per species are deposited into curated database



Selection Protocol



Clinical Gene Databases

<u>Gene</u>	<u>Region Size</u>	<u>Unique Species</u>
hsp65	382 bases	185
rpoB	657 bases	134
16s rRNA	1470 bases	184

Table: Table 1 highlights the regions lengths and size of the respective databases

Species Overlap

107 species overlap between all three databases

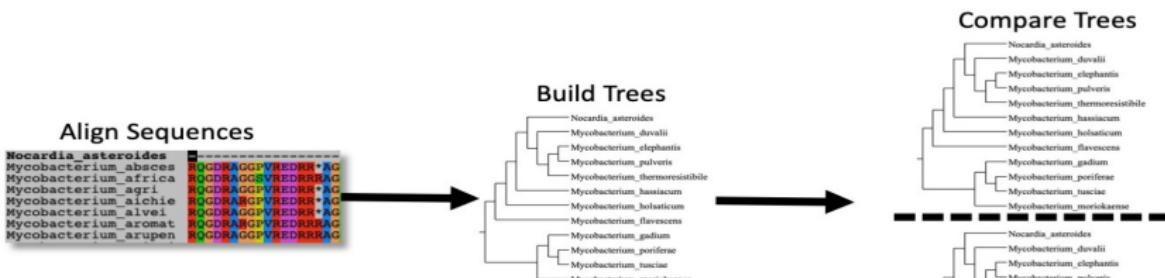
Database Validation

hsp65

156 full length hsp65 genes compared against the hsp65 database

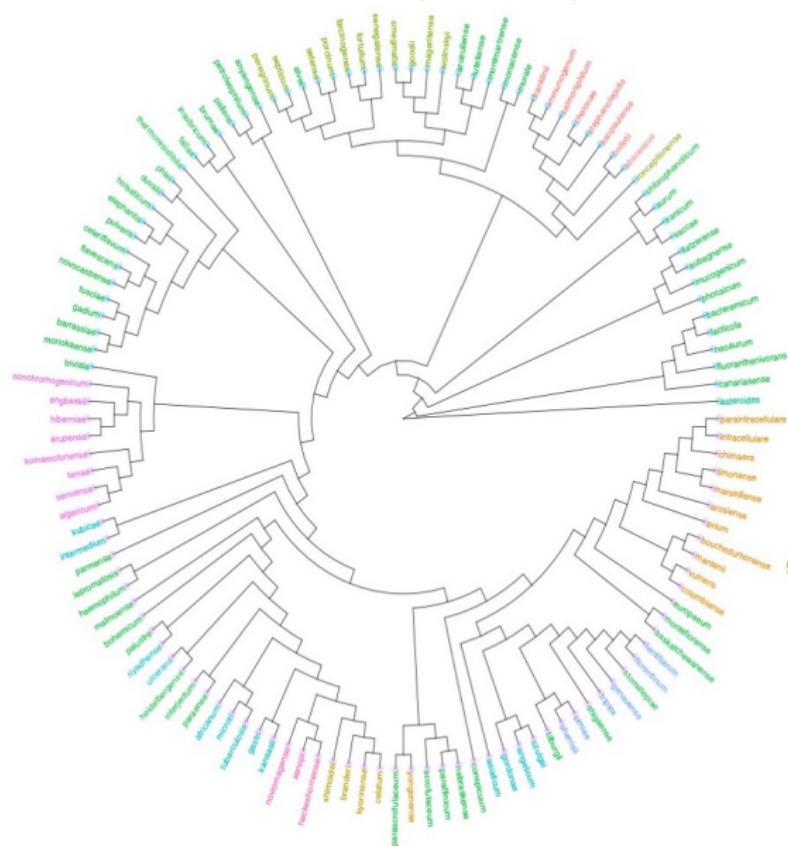
- 151/156 (96.73%) returned the exact species
- 2/5 are in the top 5 hits & 2/5 missing from database

Tree Comparison



E.Size	nRF	RF	maxRF	scr-br	ref-br
153	0.90	220	244	0.59	0.61

rpoB-hsp65 Tree



Growth Rate

- red: rapid
- purple: slow

Group

- abscessus-chelonae
- avium
- celatum
- fotuitum-smegmatis
- Other
- Outgroup
- pathogens
- simiae
- terrae
- xenopi

Conclusions and Future Directions

Representation

The subsetted databases are highly representative of prior published works

Benefits of Curated Database

- Aligned sequences to shared region
- Preferentially selected established culture codes
- Condensed and explicitly labeled ambiguous sequences

Limitations

Size of the gene regions in databases may not differentiate between species or subspecies in this version

Dai, J, et al. J Clin Microbiol. 2011

Tortoli, E, et al. Infection, Genetics, and Evolution 2017

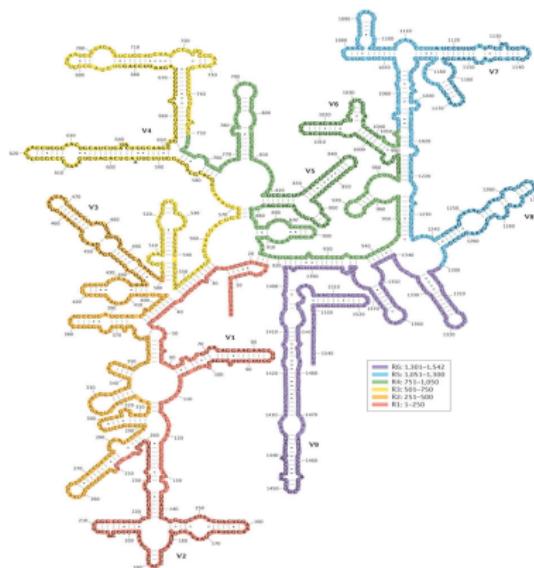
Microbiome

16S Ribosomal RNA Sequencing

- Amplifies a region of gene
- Community level analysis

Traditional Limitations

- Multiple copies of 16S
- Prokaryotic specific

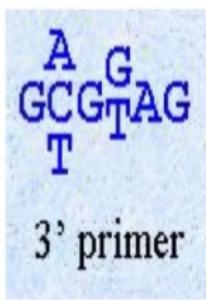
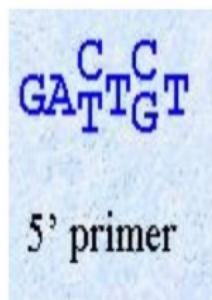


Nature Reviews | Microbiology

Yarza, P., et al. Nature Reviews Microbiology 2014

Degenerate Primers

Degenerate Primer Example



Feature of Degenerate Primers

Dual amplification of eukaryotic (18S) and prokaryote (16S)

Universal 16S/18S Primer
515F - 806R primer

Caporaso, J.G., et al. PNAS 2011
Wang, Y., et al. PLOS One 2014

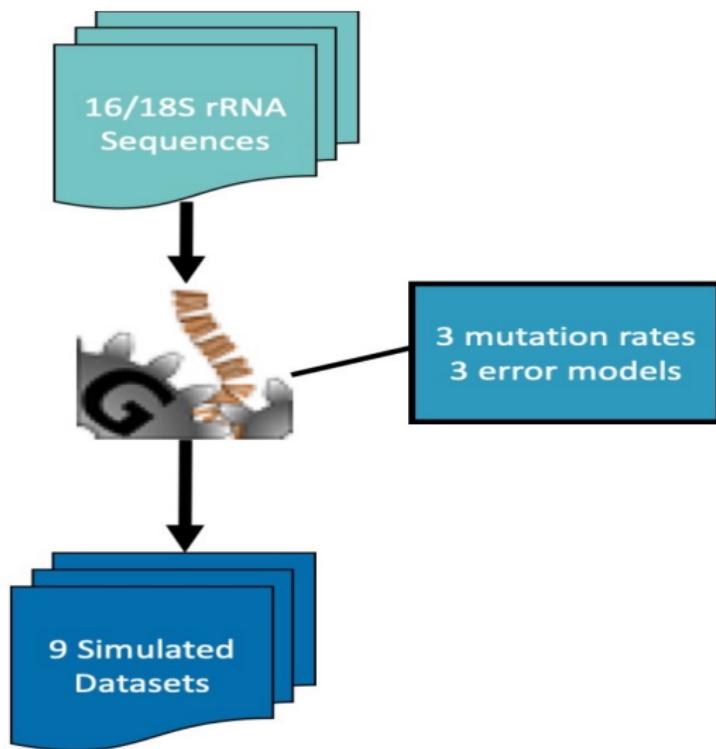
Objective

Develop an optimized pipeline to accurately describe composition of an environmental sample.

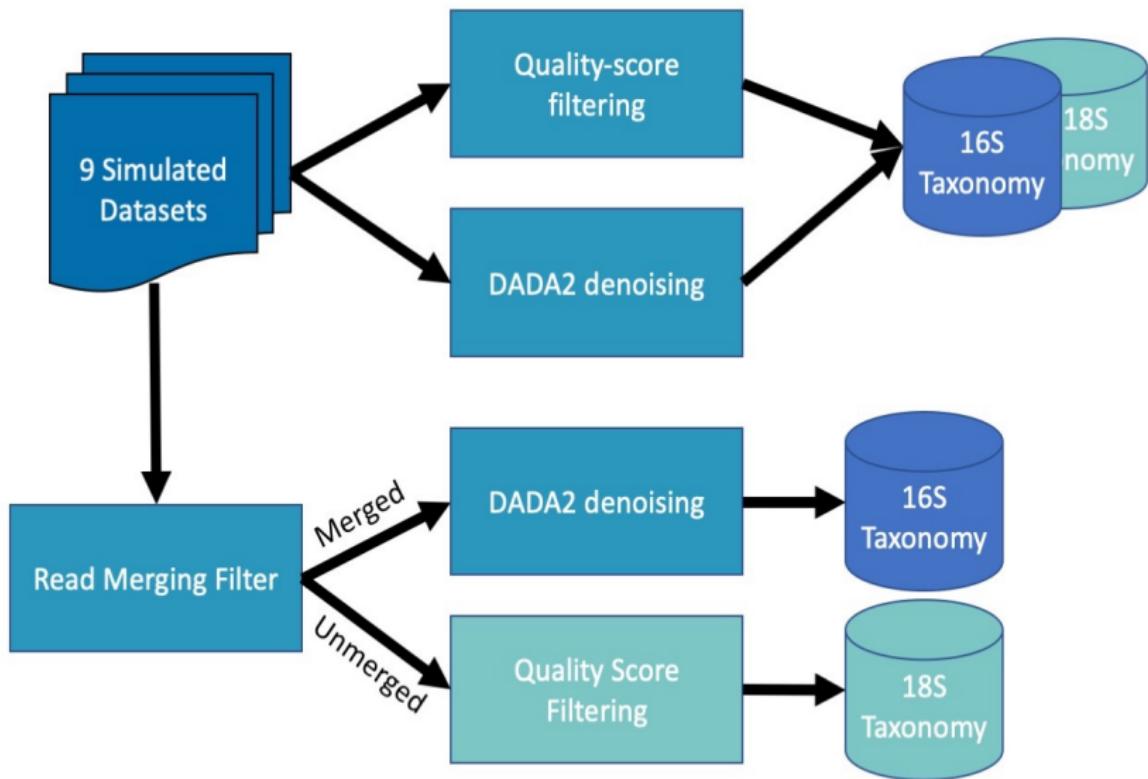
Methods Testing

- Standard OTU Picking with expanded database
- Error correction with expanded database
- Filtering 18S by merging status and parallel processing

Simulated Metagenome



Testing pipelines



Clinical NTM Gene Databases
ooooooooo

Duobiome
oooooo●o

Virulence Factors in Phages
oooooooooooo

Hybrid Viral Contig Prediction
oooooo

Results

Coming Soon

Clinical NTM Gene Databases
ooooooooo

Duobiome
ooooooo●

Virulence Factors in Phages
oooooooooooooo

Hybrid Viral Contig Prediction
oooooo

Conclusions and Future Directions

Discussion

Qiime2 Add-on

Virulence

Virulence Defined

The capacity of a microorganism to proliferate despite host defenses

Influences on Virulence

- Number of microorganisms
- Composition of the mobile genetic reservoir
- Location of niche
- Host immune capabilities

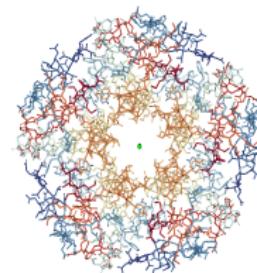
Bacterial Virulence Factors Increase Pathogenesis

Examples of Virulence Factors

- Increased fitness for nutrients
- Host immunity resistance
- Toxin secretion

Diseases from Virulence Factors

Cholera, dysentery, botulism, and food poisoning

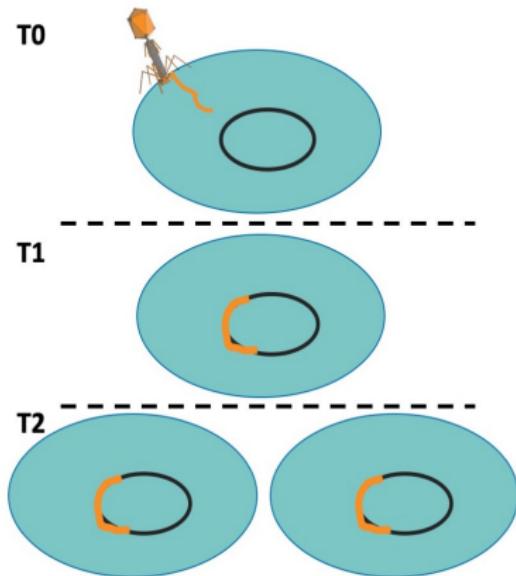


PDB Structure of Cholera Toxin

Phages as a Genetic Reservoir of Virulence Factors Genes

Phages and Pathology

Virulence Factors that cause cholera, dysentery, botulism, and food poisoning are carried on phage elements.



Objective

Characterize the abundance of bacterial virulence factors in phages

Clinical NTM Gene Databases
ooooooooo

Duobiome
ooooooo

Virulence Factors in Phages
oooo●oooooooooooo

Hybrid Viral Contig Prediction
oooooo

Data

Virulence Protein Databases

- VFDB
Chen, Lihong, et al. Nucleic Acids Research (2005)
- PatricVF
Wattam, AR, et al. Nucleic Acids Research (2017)

Virulence HMMs

- pFam
Bateman, Alex, et al. Nucleic Acids Research (2004)
- pVOG
Graziotin, AL, et al. Nucleic Acids Research (2016)

Phage Protein Database



Methods

Sequence Annotation Methods
BLAST and HMM

Normalizing By Gene Count

$$\text{Hit Percentage} = P$$

$$\text{Hit Count} = HC$$

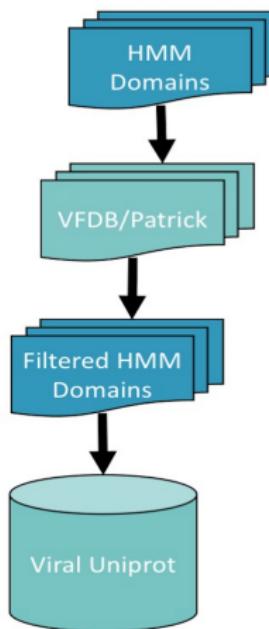
$$\text{Gene Count} = GC$$

$$P = HC/GC$$

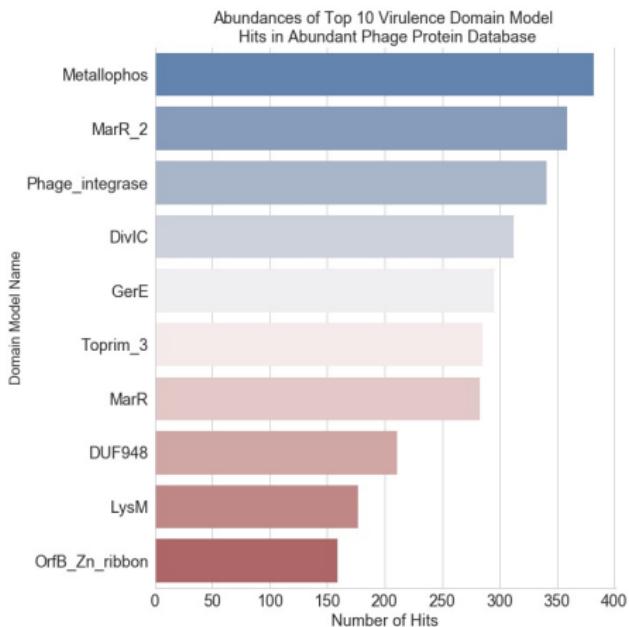
Filtering By Phage Abundance

Streptococcus phage:

Genera abundance greater than 30



HMM Hit Distribution



MarR

Domain involved in antibiotic resistance

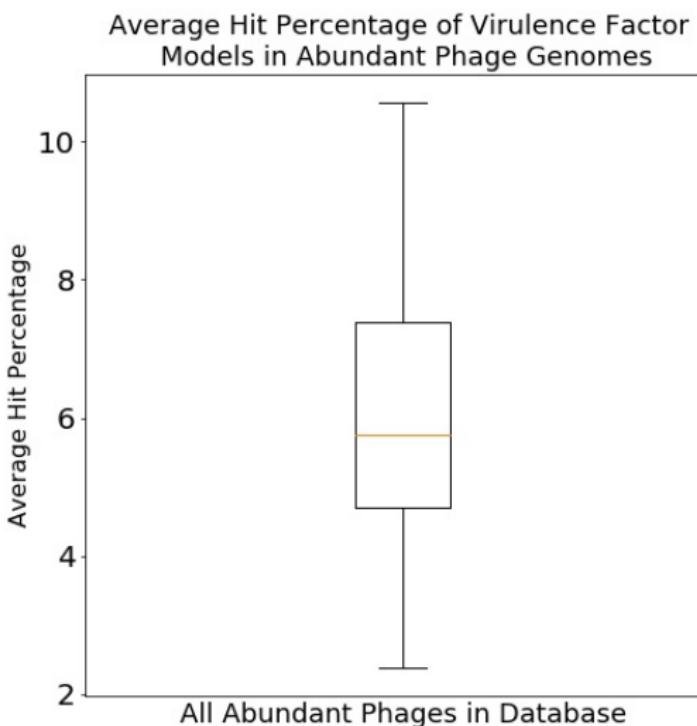
DivIC

Part of sporulation process

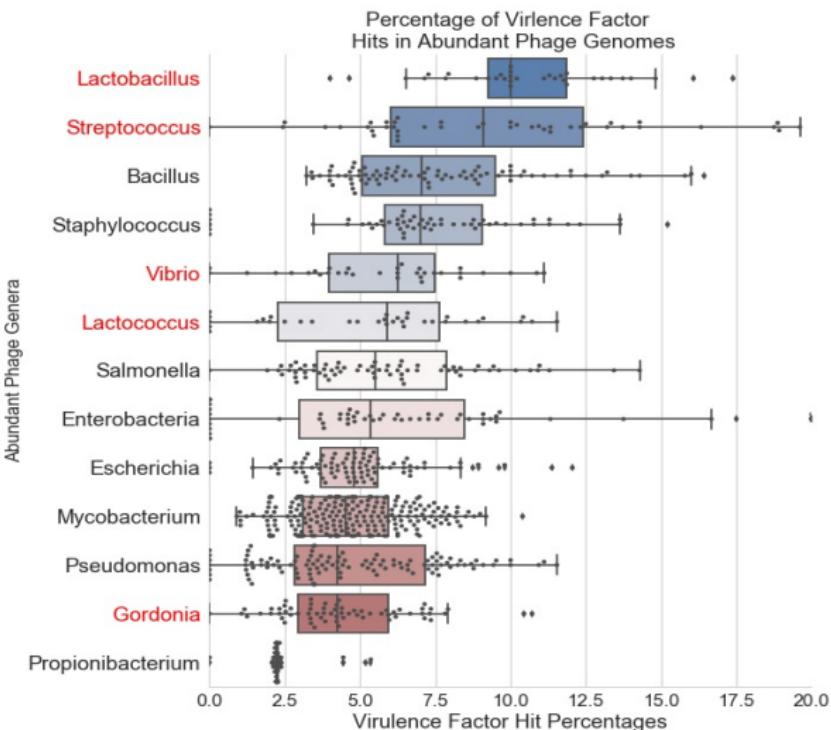
LysM

General peptidoglycan function

Distribution of Hit Percentage in All Phages



Abundant Phage Distributions by Genera Name



Clinical NTM Gene Databases
ooooooooo

Duobiome
ooooooo

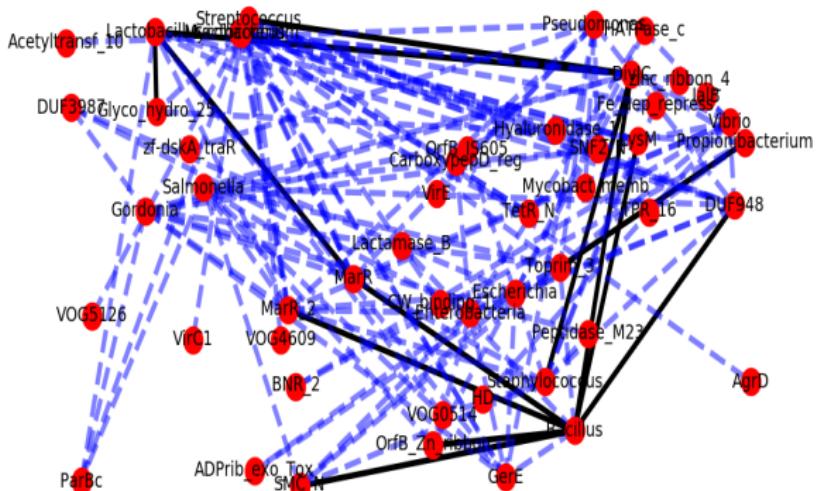
Virulence Factors in Phages
oooooooo●oooo

Hybrid Viral Contig Prediction
oooooo

Random Set of PFAMs

Insert comparable image here

Network



Clinical NTM Gene Databases
ooooooooo

Duobiome
ooooooo

Virulence Factors in Phages
oooooooooooo●○

Hybrid Viral Contig Prediction
oooooo

Comparison to Integrated Phages from Clinical NTM

Clinical NTM Gene Databases
ooooooooo

Duobiome
ooooooo

Virulence Factors in Phages
oooooooooooo●

Hybrid Viral Contig Prediction
oooooo

Conclusions and Future Directions

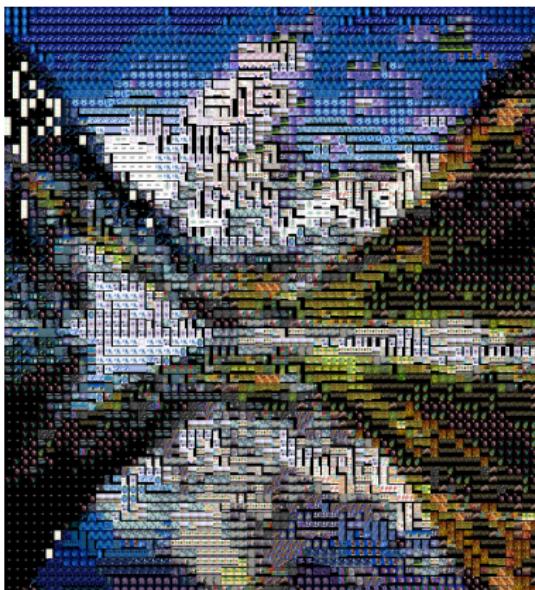
Metagenomics

What is Metagenomics?

Unbiased study of all genetic material in a sample

Importance of Metagenomics

- Functional capabilities of a sample
- Species level distinctions
- Due to lack of a universal gene marker, phages are studied by metagenomics



Clinical NTM Gene Databases
ooooooooo

Duobiome
ooooooo

Virulence Factors in Phages
oooooooooooooo

Hybrid Viral Contig Prediction
o●oooo

Methods to Isolate Phages in Metagenomics

Biological Isolations

Filtrations and density gradients to collect small particles

Sequence Similarity

Mapping to genomes, BLAST, and Hidden Markov Models

Machine Learning Methods

Linear discriminant analysis classifier on sequence k-mer profiles

Clinical NTM Gene Databases
ooooooooo

Duobiome
ooooooo

Virulence Factors in Phages
oooooooooooooo

Hybrid Viral Contig Prediction
oo●ooo

Two-Step Hybrid Model

Insert Pipeline

Clinical NTM Gene Databases
ooooooooo

Duobiome
ooooooo

Virulence Factors in Phages
oooooooooooooo

Hybrid Viral Contig Prediction
oooo●ooo

Methods

HMMs from Earth Virome

Python developed model with standalone operability

Clinical NTM Gene Databases
ooooooooo

Duobiome
ooooooo

Virulence Factors in Phages
oooooooooooooo

Hybrid Viral Contig Prediction
oooo●○

Performance Comparison Using CAMI

Clinical NTM Gene Databases
ooooooooo

Duobiome
ooooooo

Virulence Factors in Phages
oooooooooooooo

Hybrid Viral Contig Prediction
ooooo●

Conclusions and Future Directions

Concluding Remarks

Improve
Abundance
Calculations

Infer Biological
Relationships

Quantify
Virulence and
Antibiotic
Resistances

DuoBiome

Optimized methods
to simultaneously
explore eukaryotic
and prokaryotic
communities

Hybrid Viral Contig Prediction

A hybrid model to
identify phage
elements in
metagenomics and
connect them with
bacteria

Virulence Factors in Phages

First quantification
of bacterial
virulence factors
within phage
genomes



Computational Bioscience Program



Elaine Epperson

Chris Miller

Pamela Russell

Cathy Lozupone

Nabeeh Hasan

James Costello

Josephina Hendrix

Kirk Harris

Michael Strong

Funding



NLM: 2 T15 LM 9451-11

Clinical NTM Gene Databases
ooooooooo

Duobiome
ooooooo

Virulence Factors in Phages
oooooooooooooo

Hybrid Viral Contig Prediction
oooooo

Questions?

Cody Glickman



cody.glickman@ucdenver.edu

www.github.com/glickmac

www.codyglickman.com