

# Backups

Cody Glickman  
CPBS Update Talk



Nov 12th, 2018

## Research Update Outline

### Virulence Factors in Bacteriophages

In preperation

### Building Up Domains: Lysogenic Host Discovery

Incorporated into large collaborative NCBI initative

### Hybrid Viral Contig Prediction

In Preperation

## Progress of Other Projects

### Asthma Environmental Microbiome

Submitted abstract to ATS

### Clinical NTM Gene Databases

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

### Duobiome: 18S/16S Parallel Analysis

In progress

### 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

# Viral Focus

## Bacteriophages (Phages)

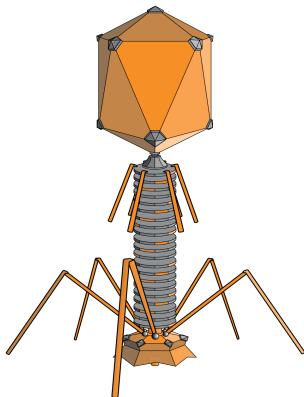
Phages are DNA viruses that infect prokaryotes

## Phage Diversity

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

## Phage Vectors

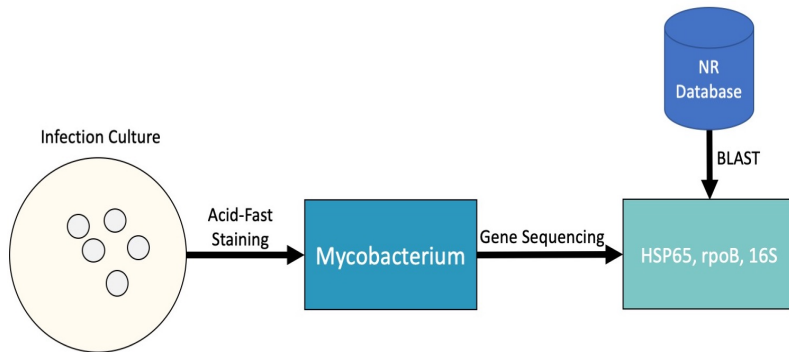
Researching how phages act as carriers of bacterial genes within clinical NTM infections



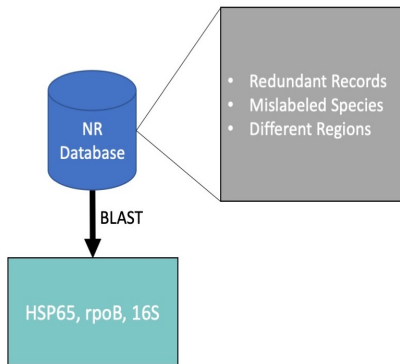
# Species Identification of NTM at NJH

## 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

## Mislabeled Species

Naming conventions are constantly updated

## Different Regions

Current protocols amplify a specific region of gene

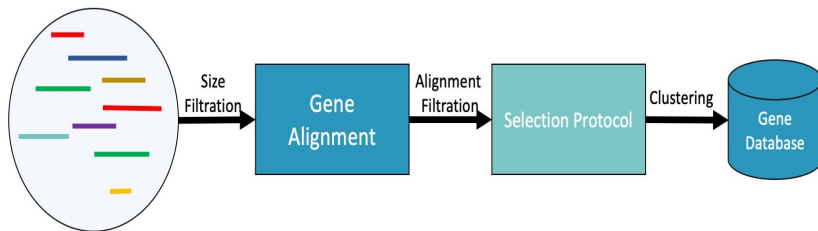
## Curated Gene Databases

Line probe to determine Mycobacterial status (rpoB amplification)  
Mycobacteriology lab depends on targets for sequencing Creating a more efficient database

### Number of Sequences per Species

Currently only two sequences per species are deposited into curated database

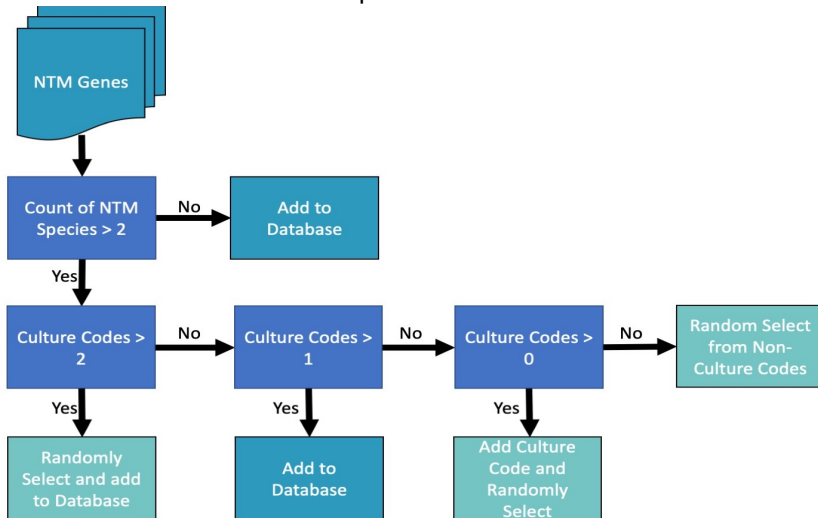
NCBI Nucleotide Database





## Selection Protocol

Culture Collection, internationally recognized strains, deposited in multiple locations



## 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

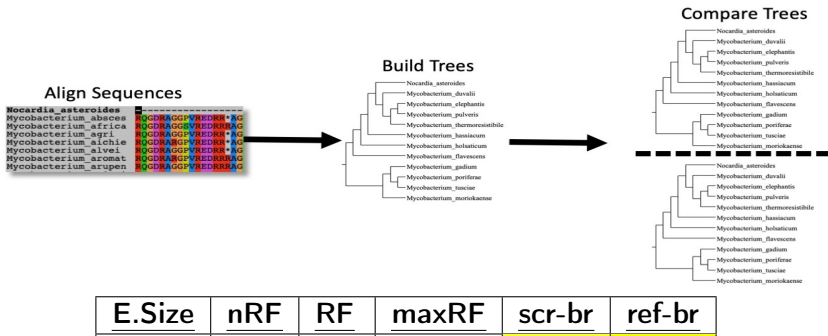
## Database Validation

### hsp65

how many species: 197 documented species: Actual species counts  
remove clusters 156 full length hsp65 genes compared against the  
hsp65 database

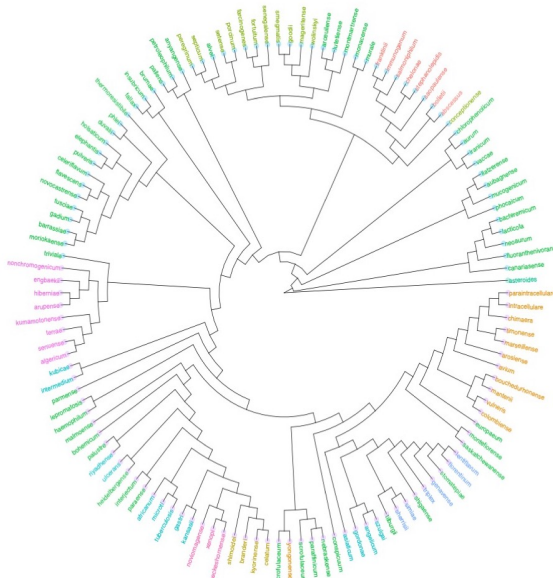
- 151/156 (96.73%) returned the exact species
- 2/5 are in the top 5 hits & 2/5 missing from database

### Tree Comparison



# rpoB-hsp65 Tree

rings around edges and lines for colors



## Growth Rate

**a** rapid

**a** slow

## Group

**a** abscessus-chelonae

**a** avium

**a** celatum

**a** fortuitum-smegmatis

**a** Other

**a** Outgroup

**a** pathogens

**a** simiae

**a** terrae

# 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. Infections, Genetics, and Evolution 2017

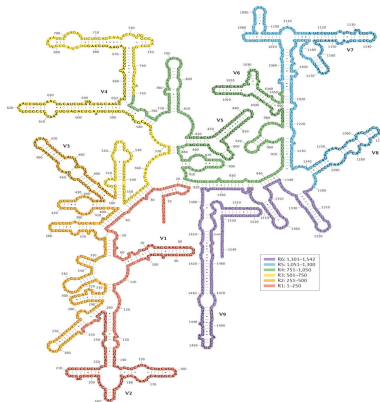
# Microbiome

## 16S Ribosomal RNA Sequencing

- Amplifies a region of gene
- Community level analysis

## Traditional Limitations

- Multiple copies of 16S
- Prokaryotic specific



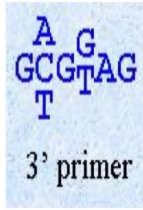
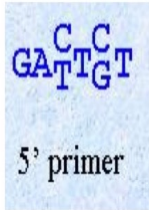
Nature Reviews | Microbiology

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

## Degenerate Primers

### Degenerate Primer Example

926 R



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

### Feature of Degenerate Primers

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

### Universal 16S/18S Primer

515F - 806R primer

# Objective

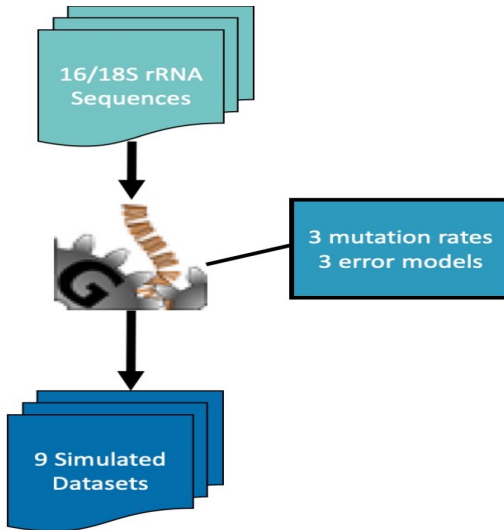
Develop an optimized pipeline to accurately describe composition of an environmental sample.

## Methods Testing

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



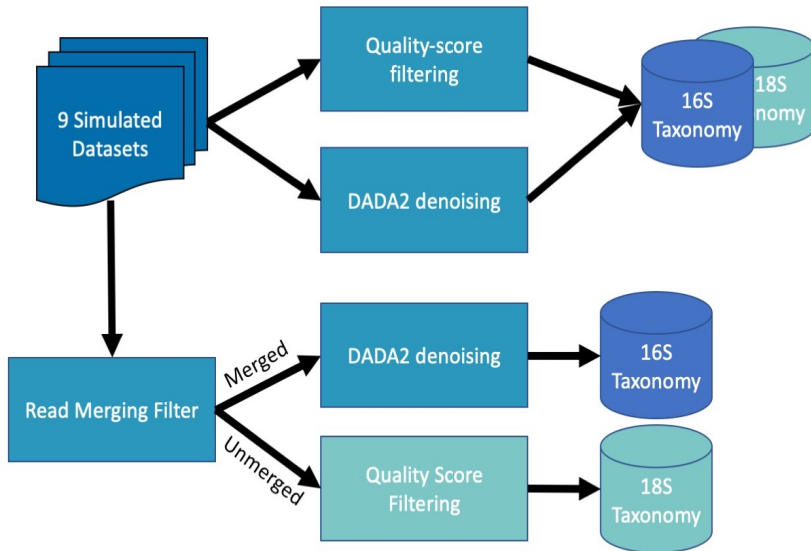
## Simulated Metagenome



not using mutation, used three different composition

Pamela Russell

## Testing pipelines



## Research Update Outline

Virulence Factors in Bacteriophages

In preperation

Building Up Domains: Lysogenic Host Discovery

Incorporated into large collaborative NCBI initative

Hybrid Viral Contig Prediction

In Preperation

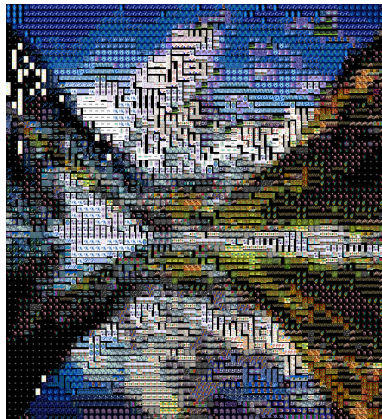
# Metagenomics

## 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



# 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

## Two-Step Hybrid Model

Insert Pipeline

## Methods

HMMs from Earth Virome

Python developed model with standalone operability

Mycobacterium, Pseudomonas, Lactobacillus

## Concluding Remarks

**Improve  
Abundance  
Calculations**

### DuoBiome

Optimized methods to simultaneously explore eukaryotic and prokaryotic communities

**Infer Biological  
Relationships**

### Hybrid Viral Contig Prediction

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

**Quantify  
Virulence and  
Antibiotic  
Resistances**

### Virulence Factors in Phages

First quantification of bacterial virulence factors within phage genomes





Elaine Epperson

Pamela Russell

Nabeeh Hasan

Josephina Hendrix

Michael Strong



## Computational Bioscience Program



Chris Miller

Cathy Lozupone

James Costello

Kirk Harris

Funding

NLM: 2 T15 LM 9451-11

## Questions?

Cody Glickman



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

[www.github.com/glickmac](https://www.github.com/glickmac)

[www.codyglickman.com](http://www.codyglickman.com)