

Virulence Factors in Phages
oooooooooooooo

Building Up Domains
oooooooo

Hybrid Viral Contig Prediction
ooooooo

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

Cody Glickman
CPBS Update Talk



Nov 12th, 2018

Virulence Factors in Phages
oooooooooooo

Building Up Domains
oooooooo

Hybrid Viral Contig Prediction
ooooooo

Research Update Outline

Virulence Factors in Bacteriophages

Glickman C., Hendrix J., Strong M. Computational identification and analysis of bacterial virulence factors embedded into bacteriophage genomes. Poster session accepted at: Rocky 18: 2018 Dec 8-10; Snowmass, CO

Building Up Domains: Lysogenic Host Discovery

Incorporated into NCBI's Virus Discovery Project

Hybrid Viral Metagenome Prediction

Glickman C., Strong M. Hybrid Viral Identification in Metagenomics. In preparation: Early 2019

Progress of Other Projects

Asthma Environmental Microbiome

Koon P., Glickman C., Epperson L.E., Strong M., Clemente J.C., Vicencio A., Diette G., Bose S.
Household determinants of the indoor environmental microbiome in an urban asthmatic population.
Poster submitted: ATS 2019: 2019 May 17-22; Dallas, TX

Clinical NTM Gene Databases

Glickman C., Epperson L.E., Hasan N., Strong M. Clinical NTM Gene Database. In preparation: Late 2018

DuoBiome: 18S/16S Parallel Analysis

Glickman C., Russell P., Epperson L.E., Strong M. DuoBiome: A workflow for mixed metagenomes. In preparation: Early 2019

Genomic Retrieval and Blast Database Creation

Glickman C., Strong M. Batch retrieval and BLAST database creation tool. Poster session accepted at: Modelling microbial communities and functions. ISME 17: 2018 Aug 12-17; Leipzig, Germany

Hawaiian Soil Chemistry and Culture

Glickman C., Virdi R., Epperson L.E., Strong M., Nelson S., Honda J. Relationship Between Soil Mineral Characteristics and Non-Tuberculosis Mycobacterium Growth. In preparation: Early 2019

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)

Virulence Factors in Phages
oooooooooooo

Building Up Domains
oooooooo

Hybrid Viral Contig Prediction
ooooooo

Viral Focus

Bacteriophages (Phages)

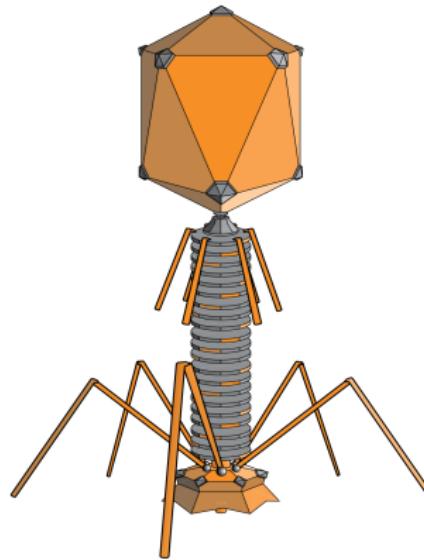
Phages are DNA viruses that infect prokaryotes

Phage Diversity

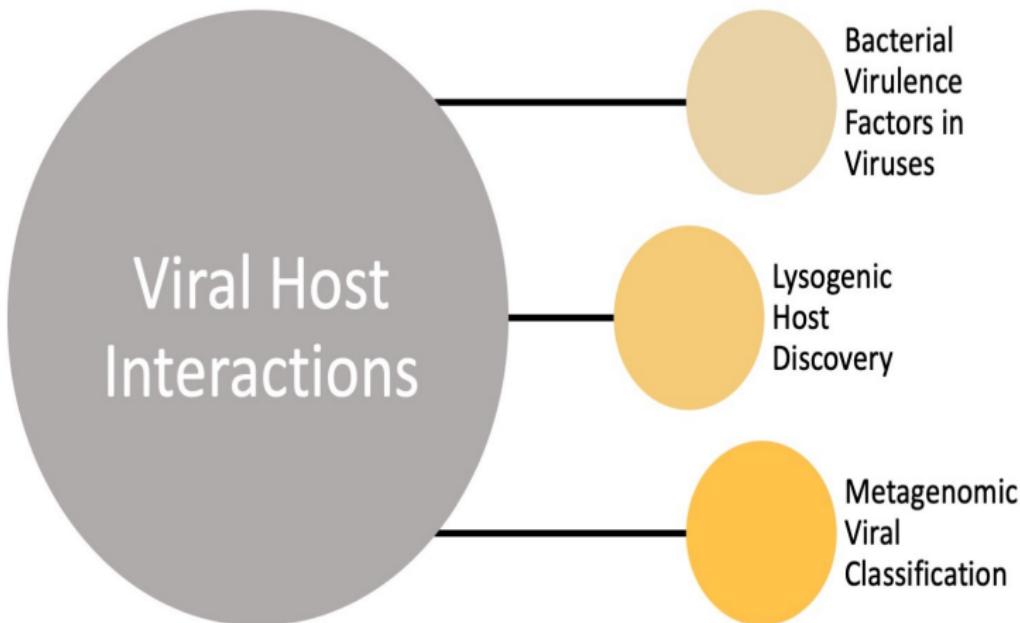
How phage abundance and diversity effects susceptibility to NTM lung infection

Phages as Vectors

How phages carry bacterial genes within clinical NTM infections



Objective



Virulence Factors in Phages
oooooooooooooo

Building Up Domains
oooooooo

Hybrid Viral Contig Prediction
ooooooo

Research Update Outline

Virulence Factors in Bacteriophages

Glickman C., Hendrix J., Strong M. Computational identification and analysis of bacterial virulence factors embedded into bacteriophage genomes. Poster session accepted at: Rocky 18: 2018 Dec 8-10; Snowmass, CO

Building Up Domains: Lysogenic Host Discovery
Incorporated into NCBI's Virus Discovery Project

Hybrid Viral Metagenome Prediction

Glickman C., Strong M. Hybrid Viral Identification in Metagenomics. In preparation: Early 2019

Virulence

Virulence Defined

The capacity of a microorganism to proliferate despite host defenses

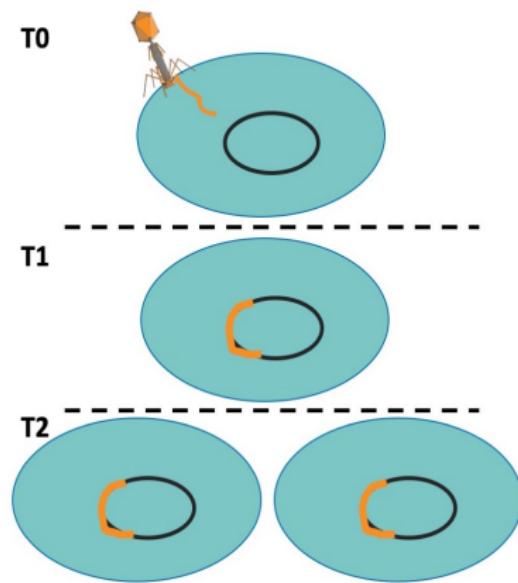
Influences on Virulence

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

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 within phages

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
Grazziotin, AL, et al. Nucleic Acids Research (2016)

Phage Protein Database



Methods

Sequence Annotation Methods

Used hidden markov models (HMM) due to variation from rapid mutation rate

Normalizing By Gene Count

Hit Percentage = P

Hit Count = HC

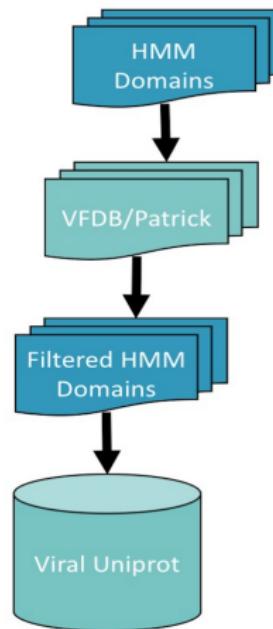
Gene Count = GC

$$P = HC/GC$$

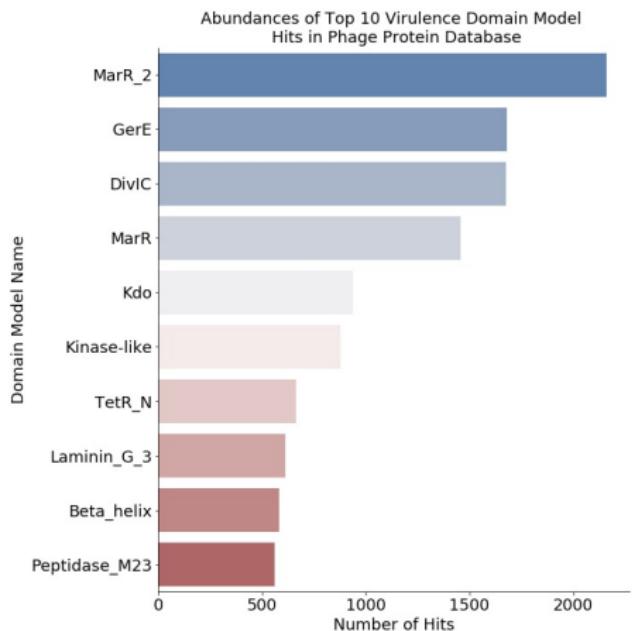
Filtering By Phage Abundance

Streptococcus phage:

Genera abundance greater than 45



HMM Hit Distribution



MarR_2/MarR

Domain involved in antibiotic resistance

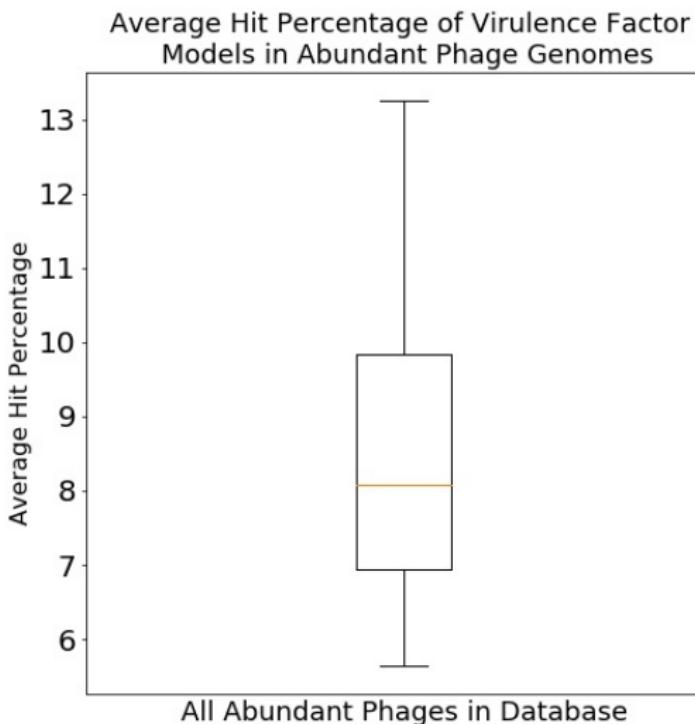
DivIC

Part of sporulation process

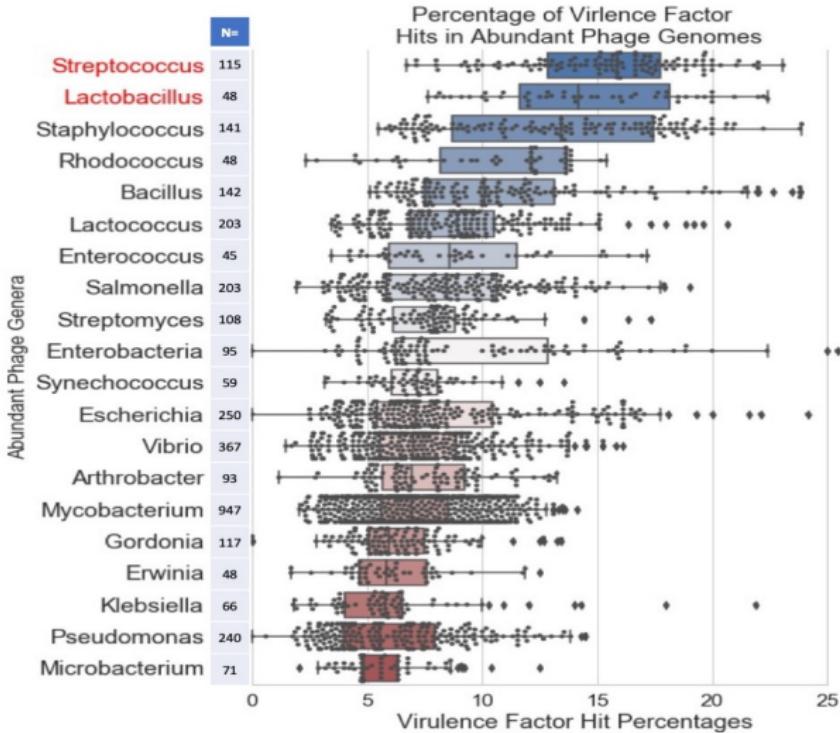
GerE

DNA binding domain found in virulence factors

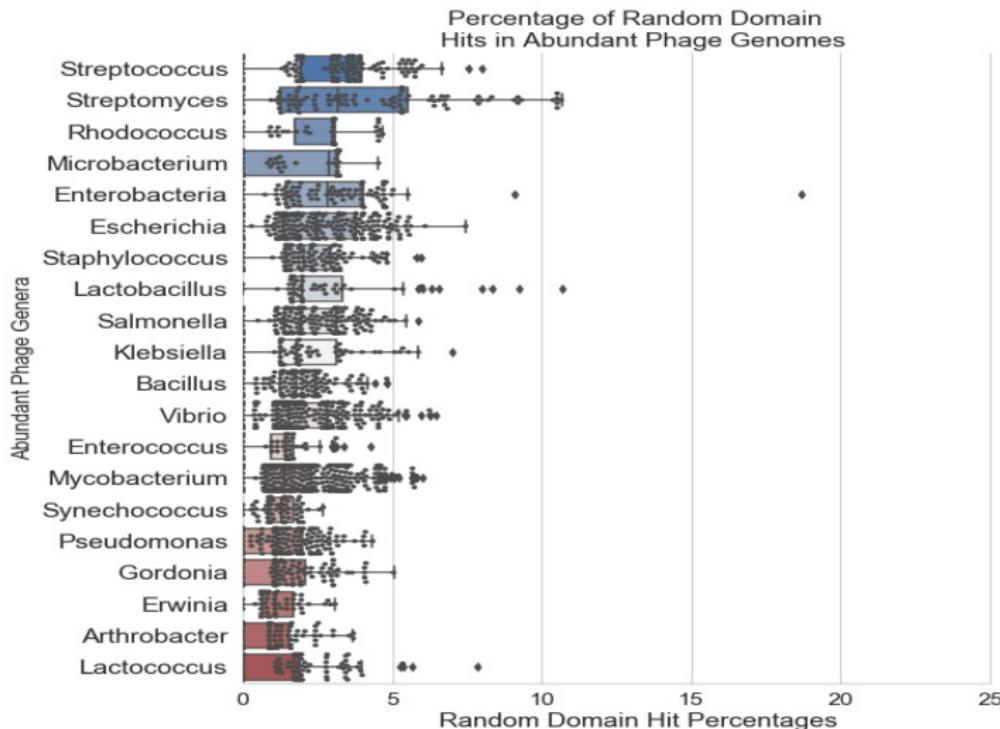
Distribution of Hit Percentage in All Phages



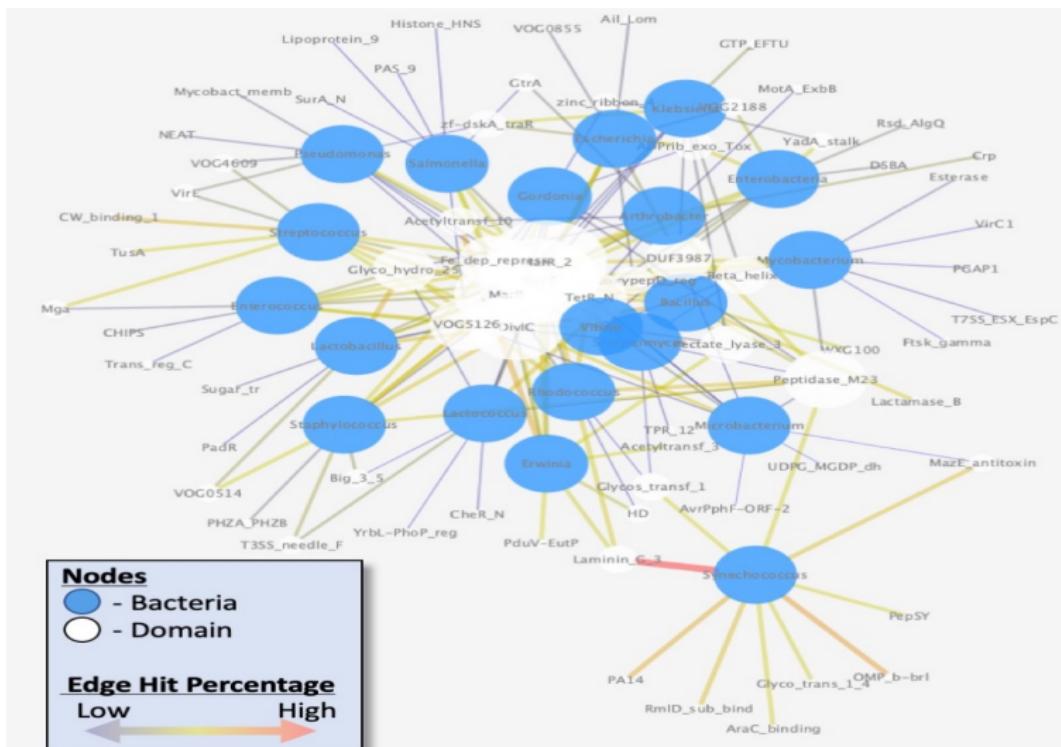
Abundant Phage Distributions by Genera Name



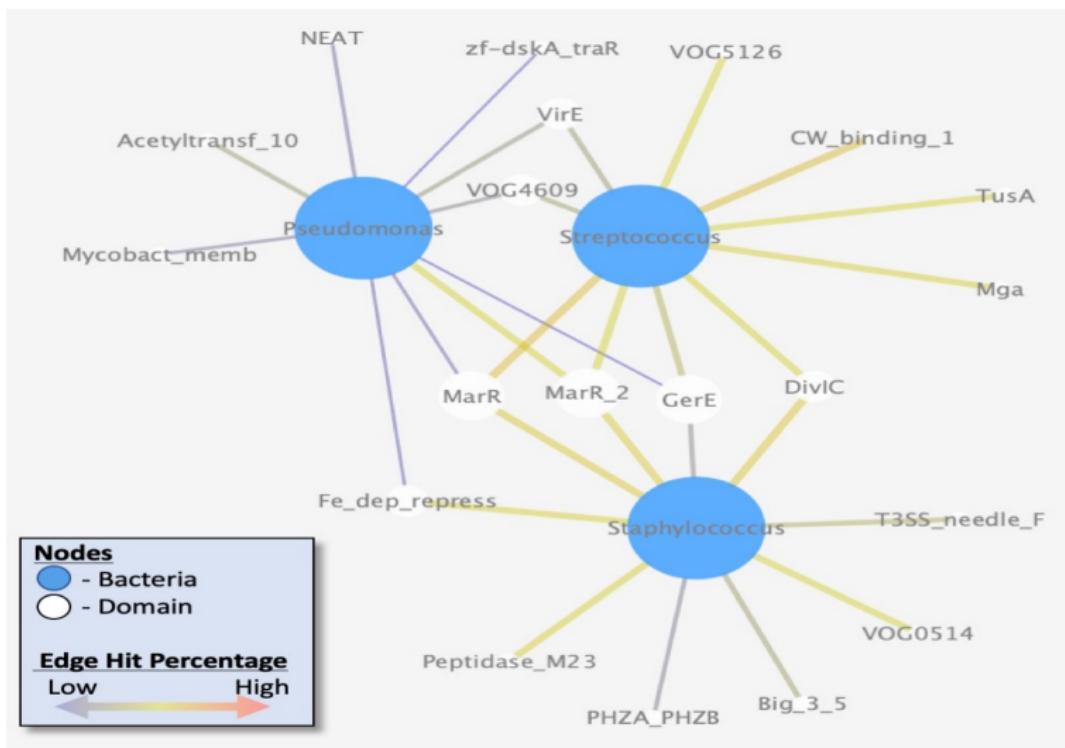
Random Set of PFAMs



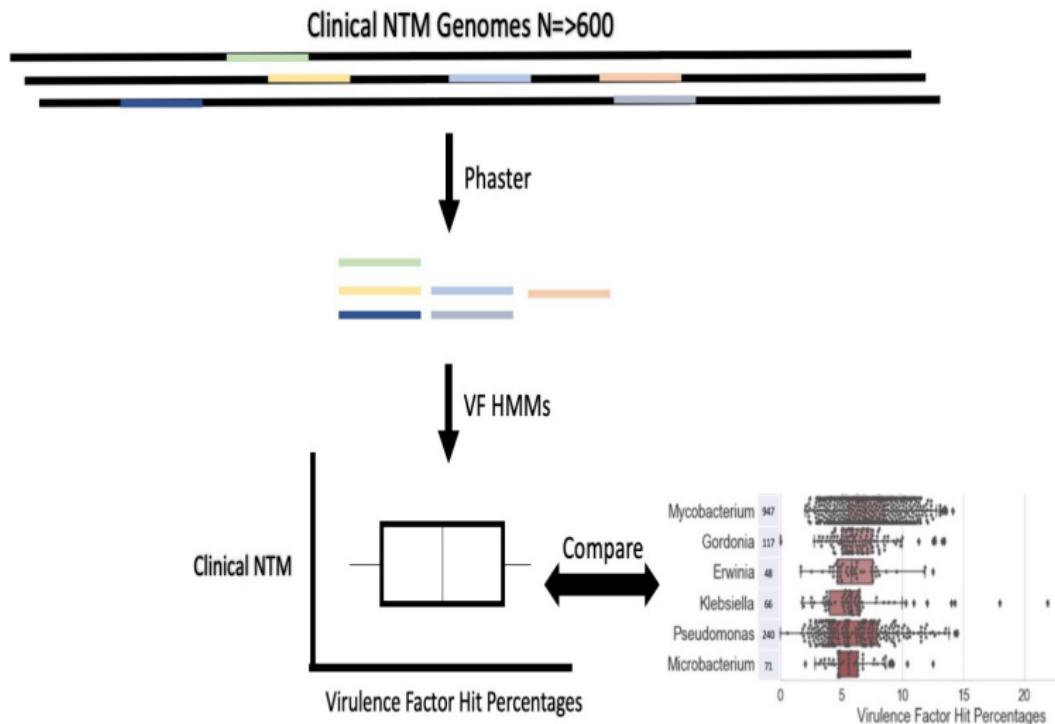
Full Top Domain Network



Pathogen Subset Domain Network



Comparison to Integrated Phages from Clinical NTM



Conclusions and Future Directions

Virulence Factors in Phages

Bacterial genera have different distributions of virulence factors in phage vectors

Clinical Hypothesis Generation

The distribution of virulence factors in integrated phages can provide mechanistic insight into clinical susceptibility and progression

Virulence Factors in Phages
oooooooooooooo

Building Up Domains
oooooooo

Hybrid Viral Contig Prediction
ooooooo

Research Update Outline

Virulence Factors in Bacteriophages

Glickman C., Hendrix J., Strong M. Computational identification and analysis of bacterial virulence factors embedded into bacteriophage genomes. Poster session accepted at: Rocky 18: 2018 Dec 8-10; Snowmass, CO

Building Up Domains: Lysogenic Host Discovery Incorporated into NCBI's Virus Discovery Project

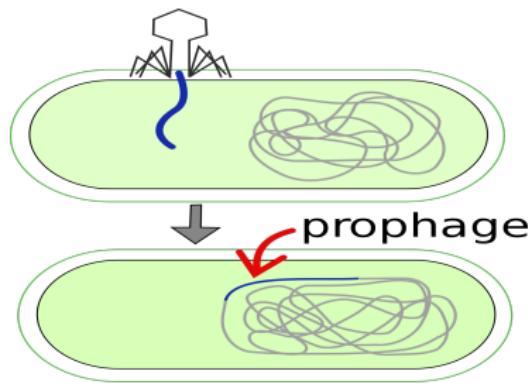
Hybrid Viral Metagenome Prediction

Glickman C., Strong M. Hybrid Viral Identification in Metagenomics. In preparation: Early 2019

Endogenous Viral Elements (Prophages)

Lysogenic Life Cycle

Viruses can integrate into host for an extended period of time



Importance of Prophages

Prophages can confer advantages to host improving survival

Prophages are important to the emergence of pathogenic bacteria

Canchaya C., et al. *Curr Opin Microbiol* (2003)
Wagner PL. & Waldor MK. *Infect Immune* (2002)

Finding Prophages

Prophage Discovery Problem

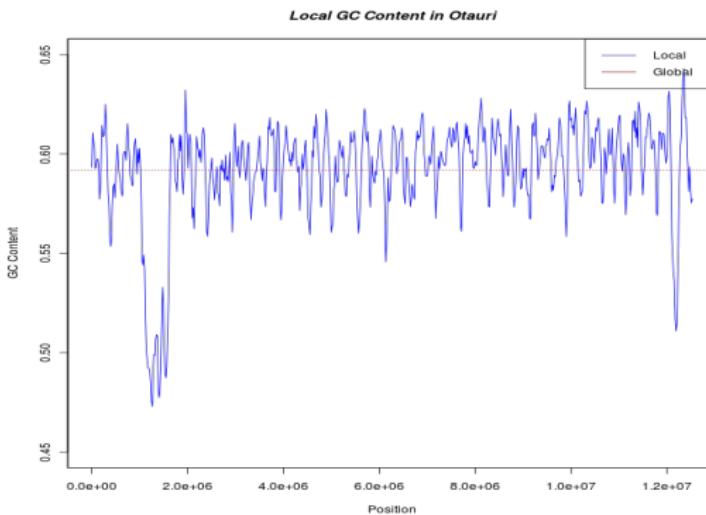
Same difficulties as gene prediction: finding signal in data



Prophage Discovery Tools

Current Methods Use

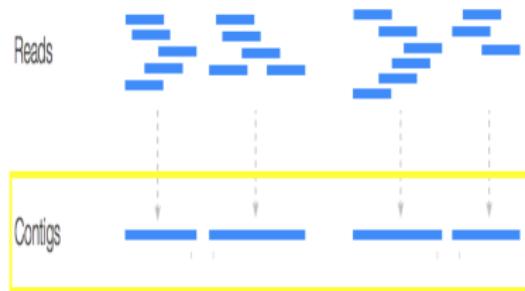
- Sequence similarity
- Hidden Markov models
- Transcription direction
- Protein length
- Sliding window GC content
- Phage specific kmer



Prophage Discovery Methods

Top Down Methods

All prophage discovery methods find prophages within contiguous sequences (contigs) or genomes



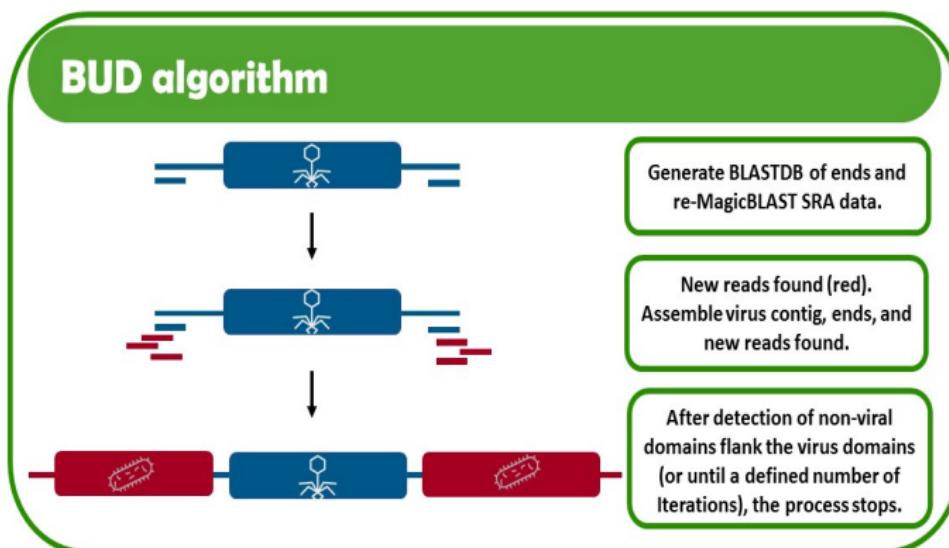
Top Down Pitfalls

Commonly require sequences longer than 1000 bases to make prediction with confidence

Building Up Domains (BUD) Algorithm

Initialization

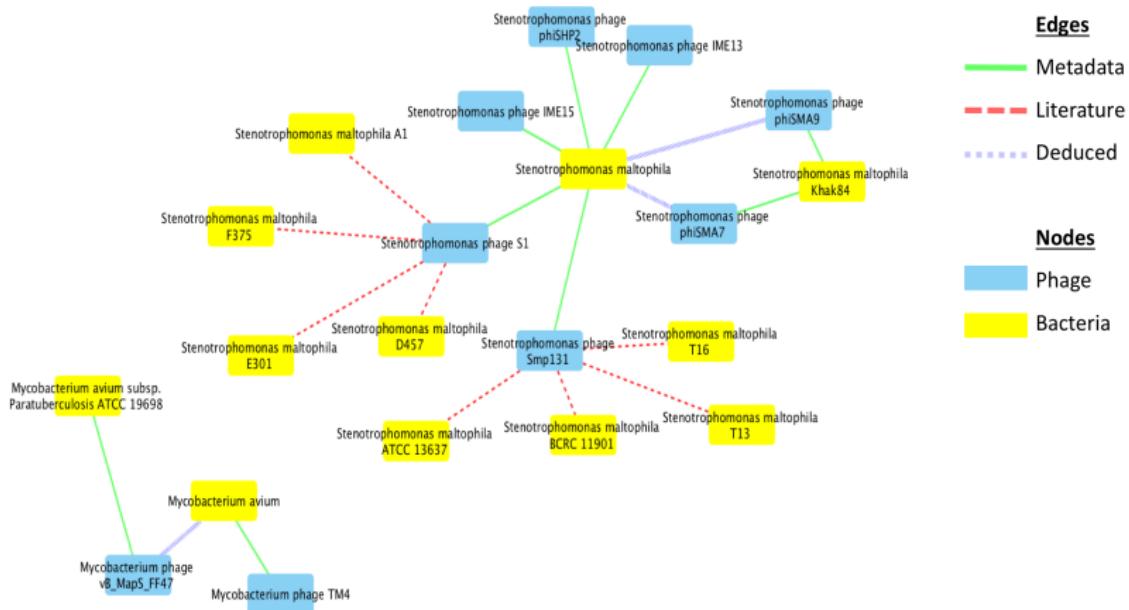
- Metagenomic reads are filtered by BLAST against Viral RefSeq
- BLAST hits are assembled into contigs



Potential Uses of BUD

Expanding Known Phage Host Range

BUD has the potential to identify novel hosts for prophages



Current Implementations of BUD

ViruSpy

- Originally written for NCBI Hackathon
- BUD Algorithm written in Perl and BASH
- Utilized Magic-BLAST for streaming of reads

EndoVir

NCBI Collaborators Jan Buchman and Ben Busby

- Written in Python
- Implementation of BUD with Magic-BLAST

Virulence Factors in Phages
oooooooooooo

Building Up Domains
oooooooo●

Hybrid Viral Contig Prediction
ooooooo

Future Directions

Overlap Consensus BUD

Create a version of BUD for local metagenomic sequences

Viral Discovery Project (VDP)

Integration into VDP a large collaboration aimed at renovating
Viral Refseq and provide viral research protocol standards

Virulence Factors in Phages
oooooooooooooo

Building Up Domains
oooooooo

Hybrid Viral Contig Prediction
ooooooo

Research Update Outline

Virulence Factors in Bacteriophages

Glickman C., Hendrix J., Strong M. Computational identification and analysis of bacterial virulence factors embedded into bacteriophage genomes. Poster session accepted at: Rocky 18: 2018 Dec 8-10; Snowmass, CO

Building Up Domains: Lysogenic Host Discovery Incorporated into NCBI's Virus Discovery Project

Hybrid Viral Contig Prediction

Glickman C., Strong M. Hybrid Viral Identification in Metagenomics. In preparation: Early 2019

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



Virulence Factors in Phages
oooooooooooo

Building Up Domains
oooooooo

Hybrid Viral Contig Prediction
o●ooooo

Methods to Isolate Phages in Metagenomics

Biological Isolations

Filtrations and density gradients to collect small particles

Lim, Y.W., et al. JoVE (2014)

Sequence Similarity

Mapping to genomes, BLAST, and Hidden Markov Models

Roux, S., et al. PeerJ (2015)

Machine Learning Methods

Linear discriminant analysis classifier on sequence k-mer profiles

Ren, Jie, et al. Microbiome (2017)

Current Tool Limitations

Sequence Similarity

- Accessibility of top performing tool is limited to integrated environment
- Requires sequences to be longer than 1000 nucleotides

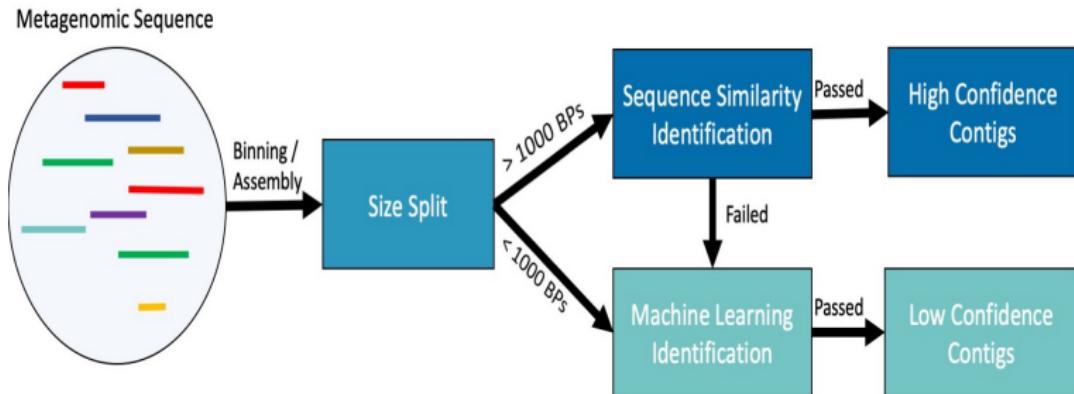
Roux, S., et al. PeerJ (2015)

Machine Learning Methods

- Isolated to R framework
- Performance is similar to sequence similarity with longer contigs

Ren, Jie, et al. Microbiome (2017)

Two-Step Hybrid Model



Virulence Factors in Phages
oooooooooooo

Building Up Domains
oooooooo

Hybrid Viral Contig Prediction
oooo●oo

Methods

HMMs from Earth Virome

Python developed model with standalone operability

Virulence Factors in Phages
oooooooooooo

Building Up Domains
oooooooo

Hybrid Viral Contig Prediction
ooooo●○

Comparison

Performance Comparison Using Critical Assessment of Metagenome Interpretation (CAMI) Data

Sczyrba, A., et al. Nature Methods (2017)

Virulence Factors in Phages
oooooooooooo

Building Up Domains
oooooooo

Hybrid Viral Contig Prediction
oooooo●

Future Directions

Concluding Remarks

Virulence Factors in Bacteriophages

Characterized bacterial virulence factors within phage vectors and compared distribution in database to clinical cohort

Building Up Domains: Lysogenic Host Discovery

A bottom up approach to identifying lysogenic phages within metagenomics

Hybrid Viral Contig Prediction An accessible two-step hybrid viral prediction pipeline for viral identification in metagenomics

Acknowledgements



Elaine Epperson

Jennifer Honda

Pamela Russell

Nabeeh Hasan

Josephina Hendrix

Michael Strong

Chris Miller

Cathy Lozupone

James Costello

Kirk Harris

Ben Busby

Jan Buchman

Paul Cantalupo

Funding

NLM: 5 T15
LM009451-12



Virulence Factors in Phages
oooooooooooooo

Building Up Domains
oooooooo

Hybrid Viral Contig Prediction
oooooooo

Questions?

Cody Glickman



cody.glickman@ucdenver.edu

www.github.com/glickmac

www.codyglickman.com