Bioinformatics discoveries brought to life: uncovering mycobacteriophage gene functions

Derek Drai

Abstract
The SEA-PHAGES program at Lehigh University is a collaborative undergraduate research enterprise focused on isolating and characterizing phages that infect Actinobacteria hosts to gain a better understanding of phage genome structure, gene function, and phage biology in general. In addition to uncovering new phages for comparative genome analysis, our program has focused on characterizing phage gene products that function in lytic infection to evaluate their potential as biocontrol agents to target pathogenic bacteria. Here we report on the genomic characterization of four of our most recently discovered novel phages (Derek [C1], Annyong [A4], James [F2], and Taptac [W]) that infect Mycobacterium smegmatis. We also present an update on the functional characterization of one of several phage gene products encoded by orphans found in the N cluster phage, Mycobacterium phage BuDers. Of particular interest among our newly discovered phages is Mycobacterium phage Taptac, a third member of the W cluster. With a genome size of 60,973 bp, Taptac has a 91 ORFs, all positioned on the forward strand – an organization common in genomes of only a few mycobacteriophages. Taptac plaques are minute (<1mm) and clear – the latter suggestive of a lytic phage. Annotation of a putative function for gp31 as a helix-turn-helix repression (99% probability in HHpred) for transcriptional regulation suggests several possibilities, including a putative lysogenic cycle for Taptac. To explore this possibility, lysogeny experiments were performed. The presence of Taptac meso and positive patch tests were suggestive of Taptac lysogens; however, stable lysogens were not formed as host cells were consistently lysed over time. The role of gp31 in the Taptac life cycle remains to be explored. In other experiments, the computational prediction that Busters gp31 is a transmembrane protein was tested by expressing a C-terminal tetracysteine-tagged gp31 and control ORFs within E. coli and imaging tagged proteins using fluorescence microscopy. Data show that Busters gp31 resides within the E. coli membrane coincident with a membrane marker. Further analysis of Busters gp31 function as a membrane protein is underway to determine how this membrane protein may function in a mycobacterial host. In summary, we present a comparative analysis of four novel mycobacteriophage genomes and provide an update on functional studies of a newly characterized membrane protein encoded by Busters gp31. Overall, our bioinformatics efforts continue to highlight genomic features of interest for laboratory exploration.

I. Genome Annotation of Novel Mycobacteriophages

Mycobacterium phage Annyong
- Founder: Audrey Ha
- Location: Avon, CT
- Characteristics: A4 cluster, 85 ORFs; 52,148 bp linear genome with 3’ sticky overhang
- Turbid plaques
- Temperate Phage
- Siphoviridae (1.1 head tail ratio)

Mycobacterium phage Derek
- Founder: Barbara Taouisits
- Location: Bethlehem, PA
- Characteristics: C cluster, 229 ORFs, 31 tRNAs, 1 trmRNA; 156,199 bp circularly permuted genome
- Turbid plaques (1mm after 24hr)
- Temperate Phage
- Myoviridae (1.1 head tail ratio)

Mycobacterium phage James
- Founder: Garrett Santini
- Location: Wayne, NJ
- Characteristics: F1 subcluster, 108 ORFs; 59,617 bp linear genome with 3’ sticky overhang
- Bullseye plaques (2mm after 24hr)
- Temperate Phage
- Siphoviridae (1.1 head tail ratio)

Mycobacterium phage Joseph
- Founder: Morgan Smith
- Location: Bethlehem, PA
- Characteristics: 3 cluster, 207 ORFs; 60,973 bp circularly permuted genome
- Clear plaques (pinprick)
- Temperate Phage
- Siphoviridae (1.1 head tail ratio)

Genomic Features of Interest: Annyong
- 87 ORFs; no orphans; 99% identical to MeeZee
- 230 ORFs, 31 tRNAs, 1 trmRNA; no orphans; 99% identical to SmallFry
- 108 ORFs; no orphans; 99% identical to Sall

Future Work:
- Determine if Derek lysogens can be formed
- Exploring the use of DNA fragmentation and Bacteriophage Recombineering using Electroporated DNA (BRID) to genetically manipulate large phages

II. Exploring the Taptac life cycle

Mycobacterium phage Taptac
- Founder: Emily Seier
- Location: Northampton, PA
- Characteristics: N cluster, 91 ORFs, 1 tRNAs; 60,973 bp circularly permuted genome
- Clear plaques (pinprick)
- Siphoviridae (1.1 head tail ratio)

When spotted onto a lawn of M. smegmatis, Taptac forms clearings with marras.

Phamrator map of Taptac Genome

Genomic Features of Interest: Taptac
- 93 ORFs, 1 tRNA-Gly; Repressor (gp31, gp32; HTH DNA binding protein; P1 ParB, plasmid partition; antitoxin component of toxin-antitoxin system; gp42; HTH DNA Binding protein; toxin-antitoxin system antibiotic; gp75; P1 ParB plasmid partition protein; 94% of gene function unknown)

Interesting gene functions with low probability scores (<90%) in HHpred:
- gp79; antitoxin of the type IV toxin-antitoxin system
- gp87; P1-like DOC (death on curing)

It may be temperate and model a mechanism of P1 phage lysogeny.

Extrachromosomal mechanism of P1 phage lysogeny: A genome is maintained as low-copy-number plasmid. Plasmid replication coupled to host genome replication is required for translocation of programmed nucleases. gp79 antitoxin systems encoded by P1 prevent plasmid

Integration-depedent mechanism of phage lysogeny:
- Unique-specific recombination of phage genome into bacterial chromosome

Experimental Plan:
- Evaluate bacterial cells from meso as possible lysogens (as described at phageDB.org) and test putative lysogens for phage releas and immunity response to Taptac.

Results:
- PATCH tests reveal putative Taptac lysogens

Putative Taptac lysogens from modified patch tests. Single bacterial colonies (A1, B1, C1, D1, D2) were re-streaked on 7H11 and an equal volume spotted onto each of two plates. Left plate shows bacterial spots grown in the absence of a top agar overlay. Right plate shows bacterial cells spotted onto M. smeg/TOP agar. Phage clearing are observed in putative lysogens A1, C1, C2, D2. Note minimal bacterial growth of lysogens A1 and C2, but the presence of large clearings.

Phamrator map with region of interest highlighted.

III. Investigating GP31 in Mycobacteriophage Butters

An introduction to Mycobacteriophage Butters
- Founder: Lena Ma SEA 2011-12
- Characteristics: Turbid plaques (0.5-3 mm 24-72 hours)
- Temperate phage
- N cluster, 49,420bp linear genome, defined physical ends
- 66 ORFs, includes 3 orphans of unknown function, and N-specific ORF (phamator map 6/2016)

Hypothesis
- Mycobacteriophage Butters GP31 is a membrane protein.

Experimental Plan:
- Use TMHMM to computationally model Butters GP31. Clone gp31-tetracysteine into pE3XPS. Using a molecular probe, FIAISH, determine Butters GP31 localization within BL21 E. coli.

Mycobacteriophage Butters GP31 has four predicted transmembrane domains

Experimental Approach for Imaging Butters GP31
- A) A snake model of Butters GP31
- B) The amino acid sequence of Butters GP31 was analyzed for transmembrane domains using TMHMM. GP31 is predicted to be a four (4) pass transmembrane protein.

Mycobacteriophage Butters GP31-tetracysteine localizes to the BL21 E. coli membranes primarily at the poles

Future Work:
- Create M. smegma GP31 strain(s) expressing Butters GP31 and GP30-31
- Determine how the incorporation of GP31 into the host membrane alters host membrane physiology, susceptibility to phage infection, susceptibility to antibiotics
- Investigate if Butters GP31 and GP31 interest using differentially tagged proteins and fluorescence microscopy

ACKNOWLEDGMENTS: HHMM and Lehigh Department of Biological Sciences for financial support; HHMM SEA for infrastructure assistance; The Hatfield lab including Wilkin Ropo for SC’ing the Busters genome; Maria Sepe for lab support and lab imaging; Bryan Berger for Butters GP31 snake model; Patricia Moreira for additional computational analysis.


Bioinformatics discoveries brought to life: uncovering mycobacteriophage gene functions

Angela V Albrecht, Lauren E Schlegel

Jennifer G Bateman, Jessica A Billir, Juliet S Chung, Lisa M Glover, Chelsea M Hipwell, Audrey K Hia, Allison B Kellinder, Nicole A Lando, Idoa D Olgabade, Garrett T Santini, Emily R Seier, Divya G Sirdeshpande, Barbara E Taouisits, Juliana N Young, Elise C Esposto, Emily L Heckman, Kenneth I Billir, Mikila A Kowal, Catherine M Magoneyn, Javier A Bucuta, Margaret A Kenna, Vasice C Ward

Department of Biological Sciences, Lehigh University, Bethlehem, PA 18015

Butters Plasmator map with region of interest highlighted.

The protein of interest (GP31) is tagged with the tetracysteine tag. The BL21 cells are allowed to amplify. The FIAISH-EDTA is added. Cells expressing GP31-tetracysteine will fluoresce, while those without will not.

Mycobacteriophage Butters GP31-tetracysteine localizes to the BL21 E. coli membranes primarily at the poles

A) Top: Control BL21 cells, FA4-64 labeled membrane only. Middle: Butters GP31-tetracysteine does not co-localize with FA4-64. Bottom: Butters GP38-tetracysteine (annotated holin) co-localizes with FA4-64 at the poles of BL21 cells.

B) Butters GP33-tetracysteine co-localizes with FA4-64 at the poles of BL21

The amino acid sequences of Butters GP21 and GP38 were analyzed for transmembrane regions using TMHMM. A) GP21 is a minor tail protein, predicted to have one transmembrane region. B) GP38 is a holin, with two transmembrane regions.