Spring 5-2017

Identifying Genes Essential for Lysogeny in Cluster E Mycobacteriophage Ukulele

Robert Levi Soohey
University of Maine

IDENTIFYING GENES ESSENTIAL FOR LYSOGENY IN CLUSTER E MYCOBACTERIOPHAGE UKULELE

by

Robert L. Soohey

A Thesis Submitted in Partial Fulfillment
of the Requirements for a Degree with Honors
(Microbiology)

The Honors College

University of Maine

May 2017

Advisory Committee:

Sally D. Molloy, Assistant Professor of Genomics, Advisor
Benjamin L. King, Assistant Professor of Bioinformatics
Keith W. Hutchison, Professor Emeritus of Molecular Biology
John T. Singer, Professor of Microbiology
Nico Jenkins, Assistant Professor College of Humanities, Husson University

ABSTRACT
Mycobacteriophage are viruses that infect *Mycobacterium*, including nonpathogenic *M. smegmatis* and pathogenic *M. tuberculosis* (1). Their genomes are diverse and the majority of their genes have no known function (1). Mycobacteriophage Ukulele was isolated from soil in Old Orchard Beach, Maine. Ukulele is temperate, carrying out both lysogenic and lytic lifestyles (8). Temperate phage typically encode integrase, excise, and repressor proteins, which regulate the change from lysogenic to lytic growth (10). Ukulele gp49 has been identified as the integrase but repressor and excise proteins have not been identified (8). Ukulele gp52 was previously proposed as a candidate repressor, predicted to encode a DNA binding domain and upstream of the integrase (8). Analyzing Ukulele-*M. smegmatis* lysogen gene expression profiles suggests that gp52 does not serve as the repressor, and is more likely excise or a Cro-like protein. Transcriptome analysis also revealed that the lytic growth is highly induced during lysogeny, many lytic genes present in the Ukulele lysogen expression profile; which is also evidenced by the high titer $1x10^{10}$ PFUs/ml for Ukulele *M. smegmatis* lysogens when incubated at 37 ^{o}C (8). To confirm the function of gp52, a strain of *M. smegmatis* will be created that over expresses Ukulele gp52. If gp52 has Cro or excise function, overexpression of gp52 in Ukulele-infected *M. smegmatis* cells will alter Ukulele plaque morphology. It is also recommended that RNA be isolated from Ukulele *M. smegmatis* lysogens at a lower temperature to increase lysogen stability (8), and reduce lytic induction.

ACKNOWLEDGMENTS

- Dr. Sally Molloy, for helping me become comfortable being uncomfortable.
- Dr. Ben King, for your guidance through something I knew little about (i.e. R)
- Dr. Keith Hutchison, for all of the questions you asked (and made me answer).
- Dr. Nico Jenkins, for helping me live up to the Honors motto "thinking hard about stuff that matters."
- Dr. John Singer, for graciously adopting our lab into your lab space last year.
- Gwendolyn Beacham, Emily Whitaker, Campbell Belisle Hailey, Emily Illingworth, and other past and present members of the Hutchison-Molloy laboratory.
- The University of Maine Honors College, for providing a space to challenge everything.
- The University of Maine Department of Molecular and Biomedical Sciences, for the incredible faculty and opportunities.
- SEA-PHAGES program, Dr. Graham Hatfull, and everyone else in support of the phage genomics course.
- The University of Maine Center for Undergraduate Research
- NIH-INBRE grant number P20GM103423.
- The INBRE Junior Year Research Award and Senior Year Research Award

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 Introduction

Mycobacteriophage (phage) are viruses that infect bacteria of the genus *Mycobacterium.* This genus includes non-pathogenic *M. smegmatis* and pathogenic species such as *M. tuberculosis.* The genomes of mycobacteriophage are extremely diverse and the majority of their genes have no known function (1). By studying mycobacteriophage and their gene function we can better understand how they interact with their mycobacterial hosts. This not only gives us insight into host physiology, but could lead to the creation of early detection methods (31), molecular tools to manipulate mycobacterial cells (2), and new treatment options such as phage therapy (38). The overall goal of this project is to learn about lysogenic regulation in phage by identifying which genes are essential for establishing and maintaining lysogeny in Cluster E mycobacteriophage Ukulele.

Currently, 8426 mycobacteriophage have been isolated and 1360 mycobacteriophage genomes sequenced (19). Due to their diversity, phages are divided into clusters and subclusters based on their genomes sharing at least 50% nucleotide identity (16). While some clusters are well-characterized, such as Clusters A, K, N, and O (5,6, 25, 7), Cluster E is not well characterized (8).

As of April, 2017, there are 84 members of Cluster E (24). Cluster E phages are temperate, although the mechanism with which they regulate their life cycles has yet to be determined (8). Temperate phage integrate their genome into the bacterial genome and replicate with the host genome, a state called lysogeny. Transcriptional silencing prevents the phage from expressing lytic genes; genes that promote replication of the phage genome, production of phage progeny and lysis of the cell. (9). Genomes of temperate

phage typically encode an integrase, excise, and regulator proteins that control the switch between lytic and lysogenic growth (10).

Cluster E mycobacteriophage Ukulele was isolated from soil at Old Orchard Beach, Maine in 2011 during the HON 150 course, and exhibits a temperate phenotype (8). Ukulele gp49 has been identified as the integrase, however, genes regulating lytic and lysogenic growth; excise, Cro, and the immunity repressor have not been identified (8).

Typically, excise and regulator proteins have DNA binding domains . In the Ukulele genome, four proteins were identified with a DNA binding domain: gp30, gp39, gp52, and gp87 (8). A phage deletion mutant of gp87 was generated but was not successfully purified indicating it is essential for growth and not likely the repressor (8). Repressor and excision genes are typically located in the integration cassette, an adjacent set of genes related in function to the integrase (10). Of these potential genes only gp52 is located near the integrase, gp49, and is divergently transcribed; a common characteristic among integration cassettes (10).

Ukulele gp52 could encode the repressor, excise, or a cro-like protein (8). The gene exists in a similar position of repressor genes in the cluster K genome, yet the theoretical protein product is predicted to share similarity to the mycobacteriophage Pukovnik excise (8). A Ukulele gp52 deletion mutant was also isolated but particles were not purified, again indicating that gp52 is essential for growth. It is possible that these deletion mutants exist as an integrated phage in a lysogen.

To gather further evidence for the function of gp52, as well as identify which genes are expressed during lysogeny, genome-wide gene expression profiles were

created for Ukulele early lytic, late lytic, and lysogenic growth, and transcriptome analysis conducted.

2.0 Literature Review

2.1 Mycobacteriophage relevance

Bacteriophages (phage), viruses that infect bacteria, are the most abundant biological entities on Earth with a population of approximately 10^{31} particles (17). Mycobacteriophages are a type of phage that infect bacteria of the genus *Mycobacterium*, including the pathogenic *M. tuberculosis* and *M. leprae*, causative agents of tuberculosis and leprosy, respectively, as well as nonpathogenic *M. smegmatis* (15).

Studying the diversity of mycobacteriophages has led to numerous advancements in molecular biology and microbiology. Tools for working on *M. tuberculosis* genetics, such as shuttle phasmids, have been important contributions (15), and the relative simplicity of their genomes have allowed these viruses to make excellent models for exploring the mechanisms of genomics. Mycobacteriophage are also important contributors to bacterial host fitness, expressing genes that benefit host bacterium survival (18).

2.2 Phage clustering and genome mosaicism

Mycobacteriophage are incredibly diverse (21). Currently, 1360 mycobacteriophage genomes have been sequenced (19). To represent genome diversity, phages are assigned to clusters and subclusters based on their genomes sharing at least

50% nucleotide identity (16). Phage genomes are highly mosaic, meaning that two genomes may contain regions that have high sequence similarity with adjacent regions having little to no sequence similarity; these similar genomic regions acquired through horizontal exchange between phages. Regions that are highly similar even appear in distantly related phage (20).

Genes are classified into "phamilies"(phams) based on sequence similarity, and can be present in related and unrelated phage (21). Genes belonging to the same pham are thought to have been passed between genomes through illegitimate recombination (22). While some clusters include phage with well-characterized genes and genomes, Clusters A, K, N, and O (5,6, 25, 7), other clusters are not well understood, including Cluster E (8).

2.3 Cluster E mycobacteriophage

As of April, 2017, there were 84 members of Cluster E (24). These phages have tails of approximately 300 nm and produce slightly turbid plaques on a lawn of *M. smegmatis* (8). Their genomes have structural genes and the lysis cassette located in the left arm of the genome, while genes involved in nucleic acid metabolism are located in the right arm, a typical characteristic of most phages (18). Cluster E phages are temperate, although the mechanism with which they regulate their life cycles has yet to be determined (8). Figuring out this process is crucial to understanding phage-host interactions.

2.4 Phage life cycles

Phage are considered either temperate or lytic based on their lifestyle. Lytic phage inject their genomes into the host cell, hijacking the host transcriptional and translational

machinery to generate new virion particles that lyse the host in order to escape into the environment to infect new cells (17). Temperate phage maintain a copy of their genome within their host, typically by integration into the bacterial genome or extrachromosomally similarly to a plasmid, both of which are replicated along with the host genome, a state called lysogeny (10).

One of the most well-characterized temperate phage is the *E. coli* phage, Lambda. Lambda regulates the switch between lytic and lysogenic growth via phage encoded proteins CI and Cro; two phage encoded transcriptional regulators. When host protease levels are high, CII is subject to proteolytic processing, and is degraded, preventing activation (10). Without transcriptional repressor CI, Cro is expressed, binding to operator site O_R within, blocking transcription of CI from promoter P_{RM}, and allowing lytic growth to proceed (12). If host protease levels are low, phage encoded CII is protected by CIII and accumulates, allowing transcription of the CI and integrase from promoters P_{RE} and promoter P_I respectively; establishing lysogeny (12). Lysogeny is maintained by CI's repression of lytic genes until the bacterial SOS response is initiated. The Lambda CI repressor utilizes the same initiation mechanism for auto-proteolytic cleavage as the *E. coli* lexA repressor (12). Both proteins are bound by an active RecA filament, which is activated by the bacterial SOS signal in response to DNA damage (12). The bound protein undergoes a conformational change and self cleavage, leaving the repressor unable to bind to its target sequence in the DNA groove. Inactivation of the repressor allows repair genes in *E.coli* to be expressed while allowing Cro expression, which initiates lytic growth (12). While this model is well-characterized, mycobacteriophage tend to regulate lysogeny via different mechanisms.

2.5 Mycobacteriophage lysogeny regulation

Mycobacteriophage can establish lysogeny through a variety of methods. Several phages in Clusters G, N, and P, including Cluster G phage BPs, utilize integration-dependent immunity, where phage attachment site (*attP*) is located within the repressor gene, and upon integration is truncated, producing a stable prophage repressor gene (12). The initial structure of the integration cassette is comparable to that of Lambda, with the repressor located upstream of the integrase, and divergently transcribed from Cro (12), until it is disrupted upon integration.

Cluster A2 phage L5, lysogeny is established and maintained primarily by expression of the gp71 repressor (9). The L5 repressor binds 'stoperator' sites within promoter P_{Left} to prevent transcription of lytic genes, and stops transcriptional expression at 28 other sites throughout the genome by protein-binding (9). Like the majority of prophages, L5 integrates into its host chromosome via expression of a phage encoded integrase, either a tyrosine- or serine-class of site-specific recombinases, that leads to recombination at a specific site within the bacterial chromosome (9). In Cluster A3 phage Ollie, a relative of L5 whose repressors share homology, the predicted repressor, gp80, structure does not share homology with the *M. smegmatis* lexA, like the Lambda CI repressor shares homology with the *E. coli* lexA (Fig.2). Ollie gp90 lacks a potential binding site for RecA, as well as the predicted protein structure not featuring a peptidase domain (Fig.3). Without the binding of RecA to cause a conformational change in the protein structure, and a peptidase to cleave the protein, Ollie cannot utilize the same cleavage mechanism as its host. Furthermore, mycobacteriophages Ollie was tested for a response to the SOS signal. A sample the *M. smegmatis* lysogen was treated with Mitomycin C and analyzed for changes in concentration of virus particles in culture

supernatant. Ollie *M. smegmatis* lysogen cultures treated with Mitomycin C had lower concentrations of free phage than untreated cultures, indicating that DNA damage, and the SOS response, does not promote induction of prophage (Fig.4). This suggests a different mechanism of regulating lysogeny for Ollie, and potentially many mycobacteriophage.

A subset of Cluster A mycobacteriophage establish lysogeny using a *parABS* system that allows for extrachromosomal replication of the prophage (23). Similar to partitioning systems of bacterial chromosomes and plasmids, expression of phage encoded *parA* and *parB* allows for segregation from the host genome (23).

Cluster E phage appear to establish lysogeny by integrating their genome into the host genome and maintain it through repression of lytic genes (8). Cluster E phage integration cassettes share a similar structure to that of Cluster G phage BPs (Fig.5). In model Cluster E phage Ukulele (Fig.1), g49 has been identified as the integrase, and is downstream of divergently transcribed genes (Fig.5). Excise and regulator proteins, such as the immunity repressor and Cro, require DNA binding domains in order to carry out their function. In the Ukulele genome four proteins were identified with a DNA binding domain: gp30, gp39, gp52, and gp87 (8). A phage deletion mutant of gp87 was generated but was not successfully purified indicating it is essential for growth and not likely the repressor (8). Repressor and excision genes are typically located in the integration cassette, an adjacent set of genes related in function to the integrase (10). Of these potential genes only gp52 is located near the integrase, gp49, and is divergently transcribed; a common characteristic among integration cassettes (10). It could encode the repressor, excise, or a cro-like protein (8). The gene exists in a similar position of

repressor genes in the Cluster K genome, yet the theoretical protein product is predicted to share similarity to the mycobacteriophage pukovnik excise (8). A Ukulele gp52 deletion mutant was also isolated but particles were not purified, again indicating that gp52 is essential for growth. It is possible that these deletion mutants exist as an integrated phage in a lysogen. Upstream of gp52 are several forward-transcribed genes, gp53, gp54, and gp55, which have unknown functions, and could play a role in integration and establishing or maintaining lysogeny. In Lambda, cro and cI are divergently transcribed, and if gp52 expresses a Cro-like protein then it is plausible that the divergently transcribed genes could express a repressor like CI (12).

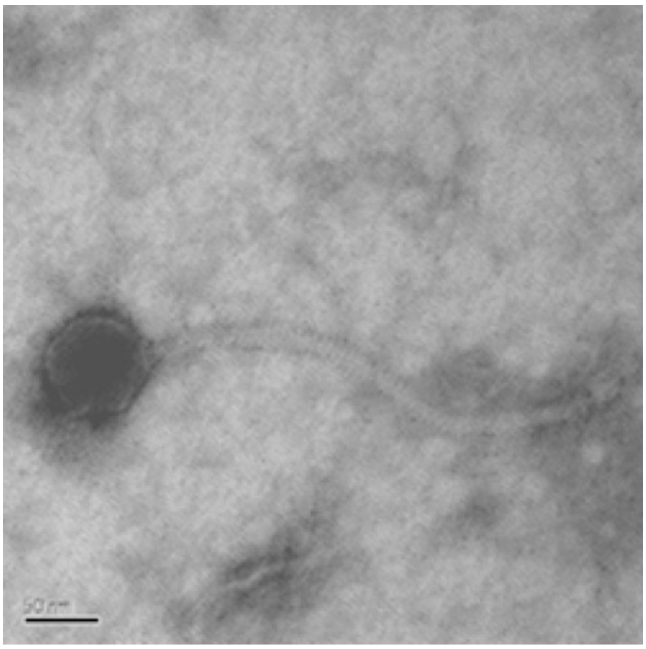

Figure 1. Ukulele particle morphology. Electron micrograph of mycobacteriophage Ukulele. Scale = 50 nm.

```
M. smegmatis    _34      STGGRRGLESALTERQRTILDVIRASVTSR  63
                ------GG-R-+E+---ER+--IL--IR-+--SR
Ollie           126  AQGGWRYVEATPEERESGIL--IRENEHSR  153

E. coli         93 HIEGHYQVDPSLFKPSADFLLRVSGMSMKDIGIMDGDLLAVHKSEDVRNGQVVVARI-DDEVTVKRLKKQGNKVELLPENSEFTPIVVD 180
                -----EG+----P+--KPS-----------------DG-L+-V---+-V--G---+AR+--DE-T-K+L-+---+V-L-P-N-+++-I--+
Lambda          145 EVEGNSMTAPTGSKPS---------------FPDGMLILVDPEQAVEPGDFCIARLGGDEFTFKKLIRDSGQVFLQPLNPQYPMIPCN 217
```

Figure 1. Comparison of amino acid sequences of studied repressors.

Figure 2. Comparison of amino acid sequence similarity between phage and host repressors. (Top) Alignment of Ollie gp80 to the *M. smegmatis* LexA amino acid sequence. (Bottom) Alignment of Lambda CI to the *E. coli* LexA. LexA and CI sequences found and aligned on BLAST (36)

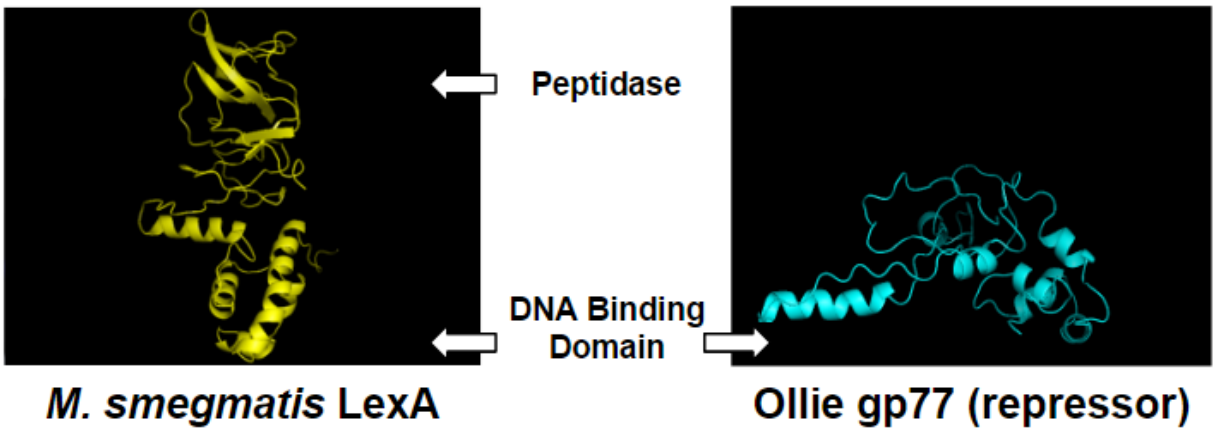

Figure 3. Comparison of the predicted protein structure of *M. smegmatis* LexA repressor and Ollie gp80.. Structures were constructed using Phyre2 (35) and visualized in PyMOL. Amino acid sequences were analysed with HHPred (34) to identify protein domains.

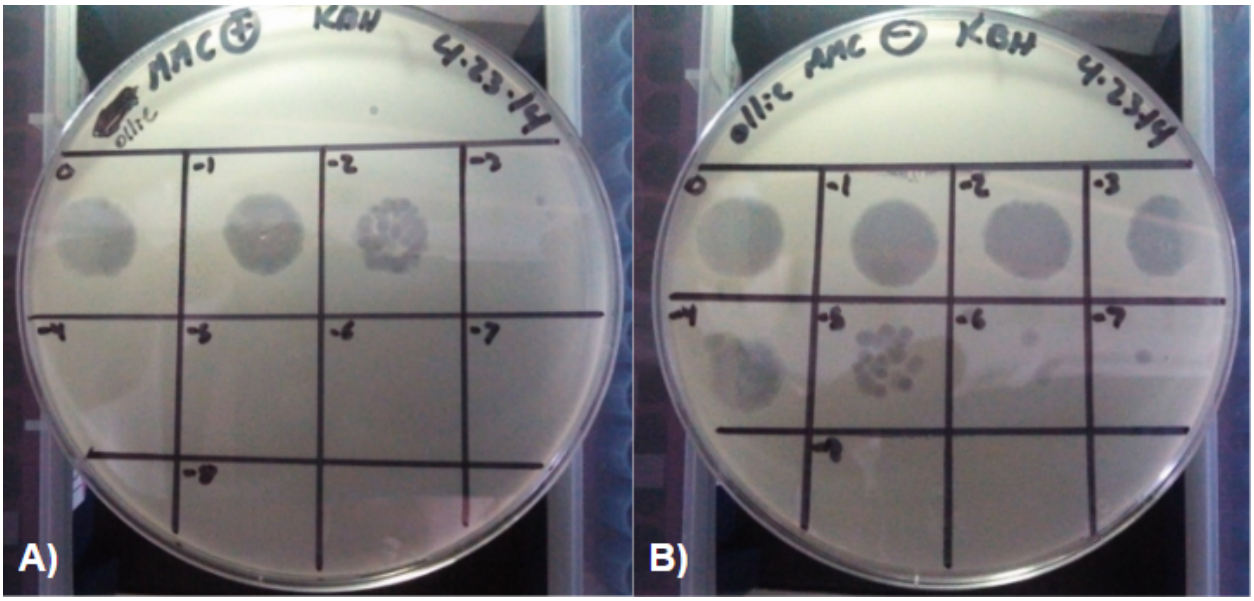

Figure 4. Mitomycin C treatment of Ollie lysogens. A) Ollie lysogen inoculated with Mitomycin C. Each cell contains different dilutions of the Ollie lysogen with Mitomycin C. B) Ollie lysogen negative control. Each cell contains different dilutions of the Ollie lysogen.

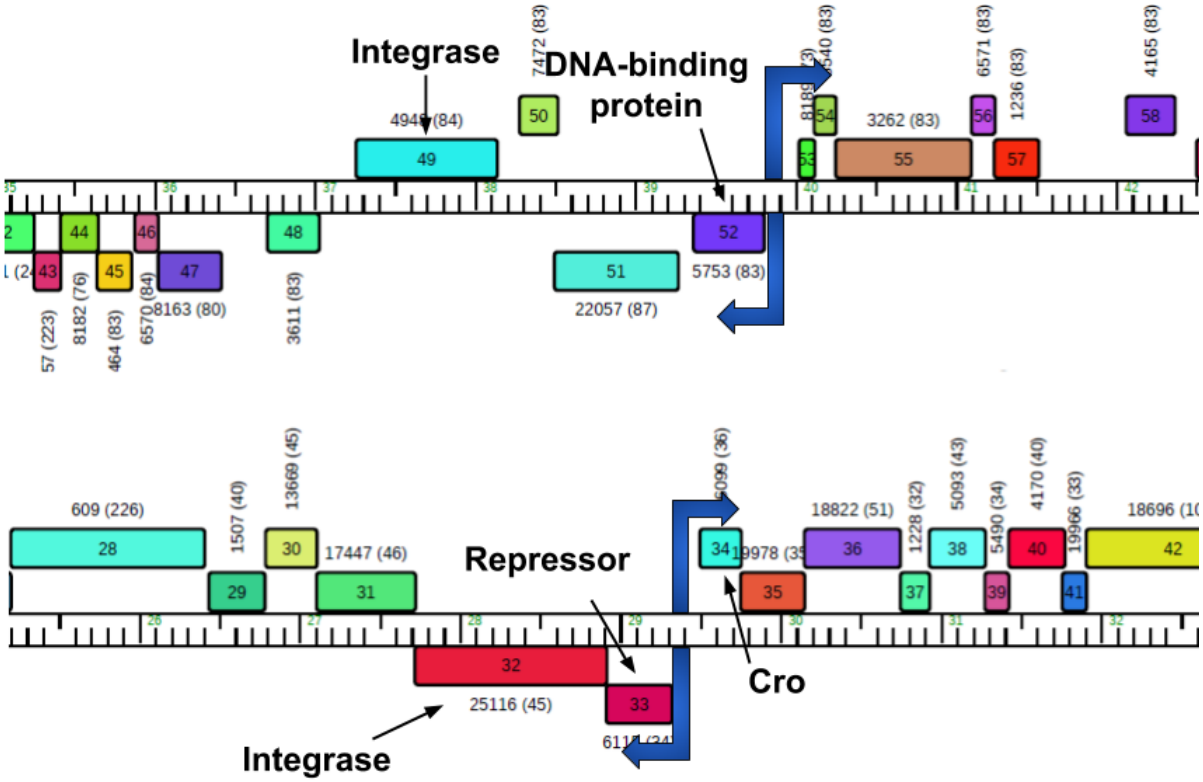

Figure 5. Comparison of BPs and Ukulele integration cassettes. Region mapped using phamerator (37). The Ukulele integration cassette is shown on top, and BPs on bottom. Genes are shown as boxes above or below the genome ruler to indicate rightward transcription (top) or leftward transcription (bottom). Phamily designations are shown above each gene with the number of members in parentheses. Genes are colored according to their pham designation. Blue arrows indicate predicted sites for promoters. Functions are listed for genes that are, or predicted to be, necessary to regulate lysogeny, with arrows indicating which gene they belong to.

2.6 Identifying gene function through transcriptome analysis

Many bacteriophage genomes have been sequenced. They all contain novel genes with unknown relatives or function; making it difficult to address which genes are required for lytic and lysogenic growth (14). Transcriptomic and functional genomic methods can be used to analyze transcription patterns and predict gene function based on when they are expressed (14). Using this approach, genes required for lytic growth and lysogenic growth were determined for mycobacteriophage Giles, including the identification of gene that encodes a repressor that lacks a DNA binding domain; an element common to most repressor proteins (14). Significant defects in lysogeny were detected in Δ47. Expression from cloned gp47 conferred host immunity to Giles superinfection, concluding that it is the phage repressor (14). The experimentally gathered evidence is consistent with transcriptome analysis, which reports gp47 having the highest level of expression during lysogeny, validating the legitimacy of this approach (14).

Transcriptome analysis can also provide evidence for novel gene function (25), and was performed for on several Cluster N lysogens (25). Using this approach, they observed expression of the repressor during lysogeny in model phage MichelleMyBell, with little to no expression of the integrase as anticipated (25). Unexpectedly, they observed expression from five genes located between the lysis and immunity cassettes, gp29-gp33, with homologues present in the many of the other Cluster N phages observed. They predicted that lysogenic gene expression in the lysis-immunity region suggests these genes could defend against viral attack via nom-repressor-mediated immunity (25). To support this claim, they determined efficiencies of plating for 80 phages against 10

Cluster N lysogens, including both lytic and temperate phages. They observed in over 75 instances that the efficiency of plating was reduced by four orders of magnitude or more (25). None of this is repressor mediated, as these phages plate efficiently on a *M. smegmatis* strain (GB203) containing just the Cluster N phage Charlie immunity cassette (25).

3.0 Materials and Methods

3.1 Mycobacteriophage and bacterial strains

Mycobacteriophage Ukulele was isolated using the enrichment method at the University of Maine, from soil collected at Old Orchard Beach, Maine. Lytic and lysogenic infections were carried out in *M. smegmatis* mc^2155 (ATCC: 700084; NC_008596.1). Total RNA was isolated from these samples in preparation for RNAseq. Transformations were performed in *E. coli* XL1-Blue (Bullock as cited in 30) or NEB 5-alpha Competent *E.coli* (New England Biolabs (NEB), Ipswich, MA).

3.2 Media and growth conditions

E. coli XL1-Blue was grown with shaking at 37 °C in L broth (1% tryptone, 0.5% yeast extract, and 0.5% NaCl). Solid media consisted of L broth containing 1.6% agar. Kanamycin was added to a final concentration of 50 ng/ml when required. *M. smegmatis* mc^2155 was grown on complete media (7H9 broth (BD), 10% AD supplement, 1 mM $CaCl_2$, 50 μg mL^{-1} carbenicillin (Sigma, St. Louis, MO), carboheximide 10 μg mL^{-1} (Sigma). Kanamycin was added to a final concentration of 25 ng/ml when required. Each organism was incubated at 37˚C with shaking at 220 rpm overnight for transformations and plasmid preparations, and until late-log stage of growth (4 d) prior to infection. Ukulele-*M. smegmatis* lysogens were incubated at room temperature with 0.005%

Tween-80 (Fisher Scientific, Fair Lawn NJ), then sub-cultured into complete media without Tween-80 and grown until late-log stage of growth (8).

3.3 Lysogen isolation

Ukulele *M. smegmatis* lysogens were isolated using a procedure modified from W. Pope, G. Sarkis, & G. Hatfull, and Greg Broussard (27). Dilutions of late-log stage *M. smegmatis* were plated on 7H10 agar and seeded with Ukulele lysate. Resulting colonies were tested by spotting Ukulele lysate on a bacterial lawn of the potential lysogen, with an absence of lysis suggesting true lysogen status (8).

3.4 Plasmid preparation

Small-scale plasmid preparations were performed on overnight 3-ml cultures of *E. coli* using a Wizard PureYield Plasmid Miniprep (Promega, Inc. USA, Madison, WI) according to the manufacturer's recommendations. Vector pST-KT is a mycobacterial expression plasmid that encodes a kanamycin resistance gene (26). Proteins are tagged with histidine and FLAG tags (26). Plasmid DNA was stored in TE buffer (10 mM Tris-HCl; 1 mM EDTA, pH 7.5) and quantified by spectrophotometry using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Rockland, DE).

3.5 Polymerase Chain Reaction (PCR)

PCR was performed using 1 ng of Ukulele genomic DNA, 0.5 μM of forward and reverse primers, Q5 Buffer, Q5 Enhancer, and Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA) according to the manufacturer’s recommendations in 25-μl total reaction volumes. Ukulele gp52 amplification reactions

were incubated at 98 °C for 30 sec, then cycled 35 times through 98°C for 10 s, 62°C for 30 s, and 72°C for 1 min. For the addition of homologous regions to gp52 for use in Gibson Assembly, reactions were incubated at 98 °C for 30 sec, then cycled 35 times through 98°C for 10 s, 72°C for 1:30 min.

Table 1 Primers used for amplification of Ukulele gp52

Primer name	Sequence (5'-3')
gp52ORF_F	GTGTGTCGCATGGGGTCA
gp52ORF_R	TCTGAACCGTTTAAGATTGGAA
gp52_Gibson_FWD	AGGAATCACTTCGCAATGCACCACCACCACCACCATATGGGTGTGT CGCATGGGGTCAAA
gp52_Gibson_RVS	GAATTCGGGCCCCAGCTGTGCGGCCGCTCTAGAGATATCGTCACAG CGGAGGAGCGTACT

3.6 Restriction endonuclease digestion

Reactions endonuclease digests were performed in 20-l total reaction volumes and contained CutSmart buffer, 0.1 mg of BSA/ml, 500–1000 ng of total DNA and 10 units of each restriction enzyme according to the manufacturer's recommendations (NEB, USA, Ipswich, MA).

3.7 Agarose gel electrophoresis

PCR products and DNA fragments were separated on a 2% Seakem LE agarose (Lonza, Rockland, ME) gel in TAE in buffer (40 mM Tris, 20 mM acetic acid, 2 mM EDTA). Gels were stained in ethidium bromide (0.5 µg/ml) and visualized via UV transillumination.

3.8 Construction of pST-KT recombinant plasmid

Plasmid construction will be performed using Gibson Assembly (SGI-DNA, Genomics Inc. USA, La Jolla, CA) according to manufacturer's recommendations. The pST-KT plasmid was linearized by *Bam*HI restriction endonuclease digestion. Ukulele gp52 was PCR amplified using primers specific to the gp52 coding region within the Ukulele genome (Table 1). Ukulele gp52 PCR product was amplified with overlapping primers designed to add regions of complementarity with pST-KT to gp52. Specific 40-nt overlapping sequences were added to the 5' ends of the forward and reverse primers, and were complementary to the sequences flanking the *Bam*HI restriction site in pST-KT (Table1).

3.9 Transformations

Transformations were performed in NEB 5-alpha competent *E. coli* (NEB) according to manufacturer's recommendations.

3.10 DNA sequencing

To confirm the recombination of Ukulele gp52 into the pST-KT vector, recombinant plasmid DNA will be sequenced at University of Maine DNA Sequencing Facility (Orono, ME).

3.11 Total RNA isolation and RNAseq

The *M. smegmatis*-Ukulele lysogen and *M. smegmatis* was grown at 37°C to an OD_{600} of 1.0. Cells were pelleted and resuspended in phage lysate at a multiplicity of infection (MOI) of 3.0. Control cells were resuspended in equal volume of phage buffer (10 mM Tris, pH 7.5; 10 mM $MgSO_4$; 68 mM NaCl; and 1 mM $CaCl_2$). Cells in triplicate 3.0-ml samples were harvested from the control flask at 0 min and from virus-treated

flask at 30 min and 2.5 h and treated with 6 ml of RNAProtect Bacteria Reagent (Qiagen Inc., Hilden, Germany) for 5 min. Cells were centrifuged 1 min at 5000 x *g*, and the supernatant removed. Pellets were resuspended in 100 µl of RNAse-free TE containing lysozyme. After a 10-min incubation at room temperature, 700 µl of RLT buffer (QIAGEN) was added. The samples were transferred to 2.0-mL tube Lysing Matrix B (MP Bio, Santa Ana, CA) and subjected to bead beating in ice-cold blocks of the TissueLyser (QIAGEN) for 10 pulses of 20 s (30 Hz). Samples were centrifuged for 1 min at 12,000 x *g*. Total RNA was isolated from each sample using the RNAeasy Mini Kit (Qiagen) according to the manufacturer's instructions. Samples were treated twice with DNAse. During RNA isolations RNA was treated with DNAseI on the column (Qiagen). After eluting RNA in 50 l of water, samples were treated with Turbo DNAseI (Ambion). RNA was analyzed for quality by agarose gel electrophoresis and and quantity by Nanodrop spectrophotometry (Nanodrop). Samples were sent to the Delaware Biotechnology Institute (Newark, Delaware) for ribo-depletion, quality control analysis and paired-end RNA-seq library preparations. , Libraries were sequenced by 50-bp read HiSeq Illumina sequencing (Illumina, Inc., USA, San Diego, CA).

3.12 Analysis of Ukulele and *M. smegmatis* transcriptomes

Diagnostic analysis of FASTQ data for each sample was performed using Galaxy (Penn State University). Reads were trimmed using Trimmomatic to remove specific adapters and low quality reads (Add Reference for Trimmomatic PMID: 24695404). Short reads were aligned to both the *M. smegmatis* and Ukulele reference genomes using Bowtie (Add reference for Bowtie PMID: 19261174), generating SAM files. SAM files were converted to BAM files using samtools (Add reference for samtools PMID:

19505943). The GenBank file for the Ukulele genome was converted into a GTF file using a custom Perl script (42). To perform differential analysis, the counts per read that map to each feature, or gene, was generated using htseq-count (28). Reads were normalized using edgeR (28), which utilizes the trimmed mean of M values, and estimates the dispersion based on a trended mean (28). This process created tab-delimited text files with annotated lists of differentially expressed genes and their statistical significance. Aligned reads (BAM files) were viewed in the Integrative Genomics Viewer (IGV) as a means of quality control for the genome annotation and expression levels across samples (29).

3.13 Computational analysis of Ukulele genes

Protein structures encoded by genes of interest were predicted by Homology detection & structure prediction by HMM-HMM comparison (HHPred) (34) and Protein Homology/analogY Recognition Engine V 2.0 (Phyre2) (35). Sequences were aligned with Basic Local Alignment Search Tool (BLAST) (36).

4.0 Results

4.1 "Marker" genes are not expressed when predicted

Specific phage genes encode proteins with strict functions and expression profiles (33). Examining when these genes, "marker" genes, are expressed can be used to measure the validity of gene expression levels. Genes necessary for DNA replication are required to be expressed during early lytic growth to prepare for the production of new viral progeny (33). Structural genes necessary for the production of new phage particles are required to be exclusively expressed during late lytic growth in order to prevent premature bacterial cell lysis (33).

Ukulele gp22 and gp108 should have strict expression profiles. Ukulele gp22 encodes the Tape Measure protein, which is necessary for construction of the phage tail, and should be expressed during late lytic growth, while gp108 encodes a DNA polymerase III subunit, and should be expressed during early and late lytic growth (33). Ukulele gp22 is expressed during both late lytic and lysogeny with no statistically significant difference between the two levels of gene expression, and gp108 is expressed during both late lytic and lysogeny with no statistically significant difference between the two levels of gene expression (Fig.6).

Ignoring normalization of the data, the raw read counts for each time point indicate that there is 3-fold increase in reads for Ukulele gp22 and a 2.5-fold increase in reads for Ukulele gp108 in the lysogenic samples compared to the late lytic (Table 2).

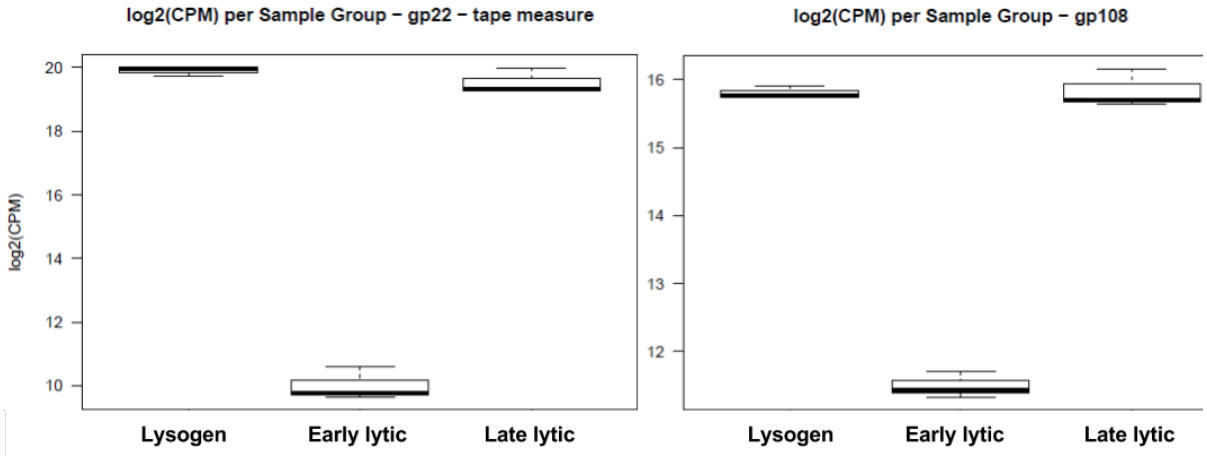

Figure 6. Boxplots of Ukulele genes with a strict expression profiles. X-axis indicates sample, and Y-axis shows log2 (counts per million) of RNAseq reads mapping to each gene.

Table 2. Averaged raw read counts for "marker" genes

Table 2. Averaged raw read counts for "marker" genes

Sample ID	gp22	gp108	Ukulele mapped reads	*M. smegmatis* mapped reads
Early lytic 1	222	698	83491	27593470
Early lytic 2	238	398	50538	29837804
Early lytic 3	117	485	57633	28042732
Late lytic 1	105852	8611	707072	33183771
Late lytic 2	109633	8765	741332	32477807
Late lytic 3	136134	9568	840161	31588050
Lysogen 1	413636	25052	1933007	26744448
Lysogen 2	368238	19672	1717051	21860136
Lysogen 3	369508	23616	1833770	30336911

4.2 None of the five genes predicted to encode DNA binding proteins are exclusively expressed during lysogeny.

Five genes were predicted to encode DNA binding proteins with helix-turn-helix domains commonly found in regulatory proteins (Fig.7) (8). Of these five genes, none are exclusively expressed during lysogeny (Fig.8). Ukulele gp30, gp39, and gp52 are all most highly expressed during early lytic growth (Fig.8). Ukulele gp100 is most highly expressed during late lytic growth (Fig.8). Ukulele gp87 is expressed during both early lytic and lysogeny with no statistically significant difference between the two levels of gene expression (Fig.8).

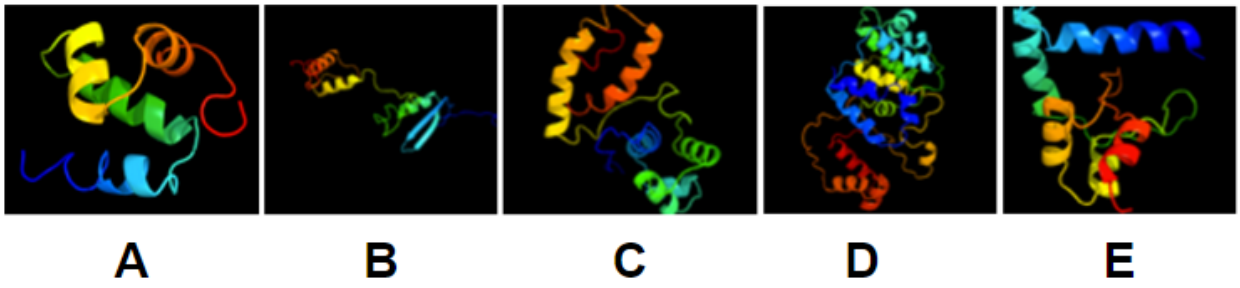

Figure 7. Predicted tertiary structure for Ukulele DNA binding proteins (8). Ukulele has five ORFs predicted to encode proteins with DNA binding domains: gp30 (A), gp39 (B), gp52 (C), gp87 (D) and gp100 (E). Folding predictions were made using the program Phyre2 (35).

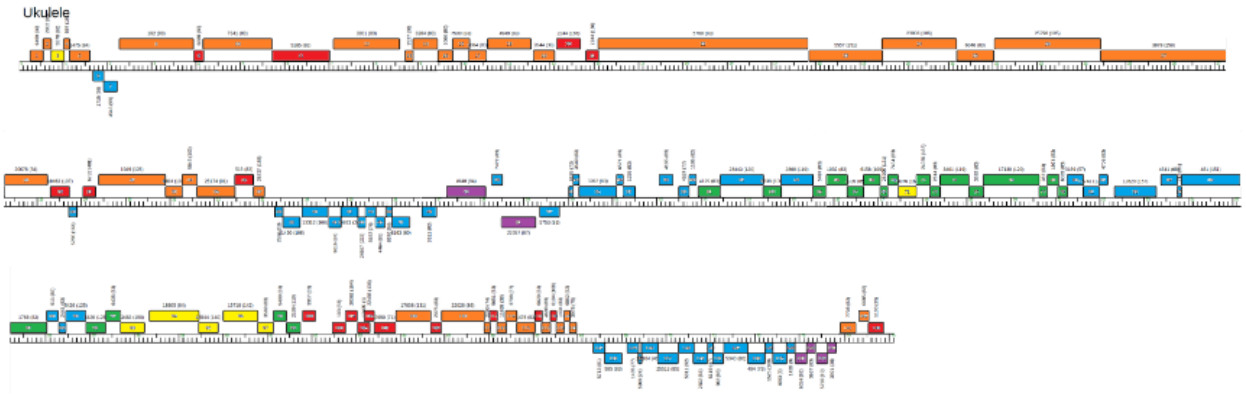

Figure 8. Genome map of Ukulele with indication of significant upregulation of genes. Blue, red, and yellow genes indicate genes upregulated during early lytic, late lytic, and lysogenic growth respectively. Purple, green and orange genes indicate upregulation during multiple time points; both early and late lytic, early lytic and lysogenic, and late lytic and lysogenic respectively.

4.3 Genes within the integration cassette are expressed during early lytic growth

To determine when genes were most highly expressed, log2(CPM) reads were compared across biological replicates and each sample. Within the integrase cassette, Ukulele gp49 is highly expressed during each time point, gp51 is highly expressed during both early and late lytic growth, and gp50, gp52, gp53, gp54, gp55, gp56, and gp57 are all most highly expressed during early lytic growth (Fig.9).

Ignoring normalization of the data, the averaged raw read counts of gp50, gp52, gp53, gp54, gp55, gp56, and gp57 are highest during early lytic growth, and highest during lysogeny for gp49 and gp51 (Table 3).

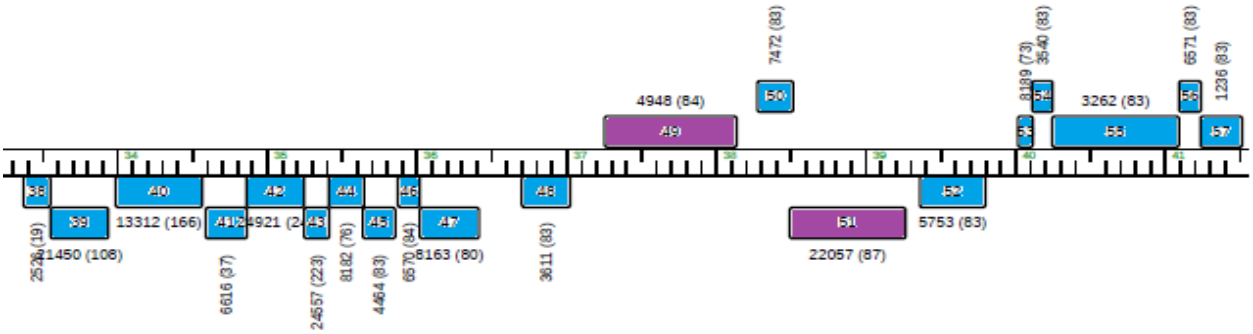

Figure 9. Ukulele integration cassette. Region mapped using phamerator (Cresawn et al., 2011). Genes are shown as boxes above or below the genome ruler to indicate rightward transcription (top) or leftward transcription (bottom). Phamily designations are shown above each gene with the number of members in parentheses. Genes colored blue are most upregulated during early lytic growth. Genes colored purple are upregulated during early and late lytic growth with no statistically significant difference.

Table 3. Averaged raw read counts for genes within the Ukulele integration cassette

Table 3. Averaged raw read counts for genes within the Ukulele integration cassette

Sample ID	gp49	gp50	gp51	gp52	gp53	gp54	gp55	gp56	gp57	Ukulele mapped reads	*M. smegmatis* mapped reads
Early lytic 1	473	593	315	1474	103	52	1410	56	584	83491	27593470
Early lytic 2	299	352	167	883	89	49	802	52	320	50538	29837804
Early lytic 3	367	524	218	1002	61	36	1043	28	475	57633	28042732
Late lytic 1	430	122	204	608	17	18	355	16	210	707072	33183771
Late lytic 2	366	131	205	528	42	22	426	29	192	741332	32477807
Late lytic 3	328	109	156	448	13	4	276	7	112	840161	31588050
Lysogen 1	752	252	294	450	30	1	204	2	27	1933007	26744448
Lysogen 2	644	251	294	495	12	5	175	1	18	1717051	21860136
Lysogen 3	719	288	467	729	11	3	256	4	35	1833770	30336911

4.4 Eight genes are consistently upregulated during lysogeny

Pairwise comparisons between each growth phase (Lytic 150 vs Lysogen and Lytic 30 vs Lysogen) were made to observe the log fold-change of gene expression across samples and measure the statistical significance. Due to background expression of early and late lytic genes in the lysogen samples, genes upregulated in lysogeny but not during either lytic growth phase were identified (Fig.8). Of these genes, only 8 genes in Ukulele that are consistently upregulated during lysogeny: gp3, gp71, gp93, gp94, gp95, gp96, gp97, and gp102. Predictions made by HHPred (34) for Ukulele gp3, gp71, gp93 included only low probability matches to protein structures in the databases (Fig.10). Genes encoding predicted protein structures with high probability matches include gp94, matching to an adenine-specific methyltransferase, gp95, matching to a ssDNA binding protein, and gp96, matching to an ATP dependent Clp protease (Fig.10).

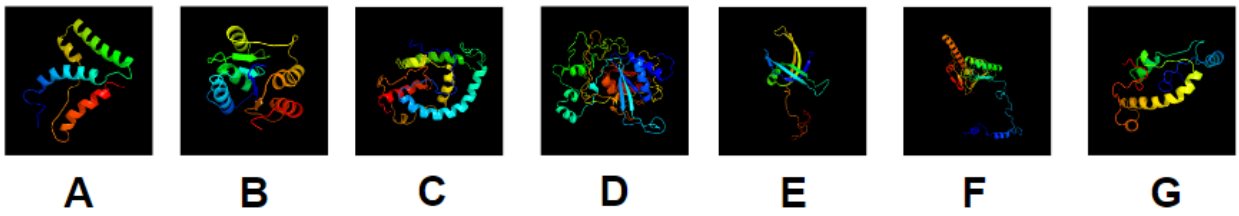

Figure 10. Protein structures/functions and probabilities for 8 consistently upregulated

Ukulele has eight genes consistently upregulated during lysogeny: gp3 (A), gp71 (B), gp93 (C), gp94 (D), gp95 (E), gp96 (F) and gp97 (G).

5.0 Discussion

Mycobacteriophage are diverse viruses that infect bacteria of the genus *Mycobacterium* (1). By studying mycobacteriophage and their gene function, we can better understand how they interact with their mycobacterial hosts; giving us insight into host physiology that could lead to the creation of early detection methods (31), molecular tools to manipulate mycobacterial cells (2), and new treatment options such as phage therapy (38). In order to understand the phage-host relationship, it important to understand how phage regulate their life cycles. While this process is understood in phage belonging to other clusters, including Clusters A, K, N, and O (5,6,25,7), Cluster E phage are not well characterized (8). This study aimed to identify what genes are essential for establishing and maintaining lysogeny in Cluster E phage, using Ukulele as a

model; determining the function of candidate regulator genes within the Ukulele integration cassette.

5.1 Strengths and weaknesses of RNA-seq study

RNA-seq is an incredible tool for transcriptome analysis; having both advantages and drawbacks. The primary advantage of this approach is the broadness of scope. RNA-seq allows for the observation of differential gene expression across the entirety of the Ukulele genome; across multiple time points. Looking at gene expression in a global way, and observing how gene expression changes between lytic and lysogenic growth, can direct the focus of future analysis and experiments to more specific regions of the genome. Perhaps the biggest issue with this study is only having three biological replicates, which might not adequately account for inherent biological variation across samples. The only way to account for this variation would be to reproduce this experiment with many biological replicates. Also, what makes this study invaluable is also a problem. Gene expression is broad and relative to the sensitivity of read depth across the entire genome, and the correlation between transcription and translational levels, that create phenotypic changes, is being assumed. To validate these results it would be advisable to perform qRT-PCR to increase the sensitivity for specific genes of interest, and validate differential expression across samples. It would also be advisable to perform mass spectrometry to detect the presence of translational products of genes of interests.

5.2 Lytic growth highly induced during lysogeny

Expression levels of both early and late lytic “marker” genes are equally as high during lysogeny (Fig.6). Late lytic gene gp22, the tapemeasure gene, should be expressed

exclusively during late lytic growth, however it is expressed during late lytic and lysogenic growth with no statistically significant difference in the level of gene expression (Fig.6). Early and late lytic gene gp108, the DNA polymerase III subunit, should be expressed during early and late lytic growth, however it is expressed during late lytic and lysogenic growth with no statistically significant difference in the level of gene expression (Fig.6).

It is likely that lytic growth is highly induced during lysogeny, and would explain the high level of lytic gene expression present in the lysogen samples (Fig.8). For many of these genes it would not make sense to be expressed during lysogeny, such as the capsid or tail proteins, because it would be a metabolic waste for the phage to begin creating viral progeny in a prophage state. Late lytic genes lysin A and lysin B are cytotoxic, and have to be expressed during late lytic growth to ensure that the virus can create new virion particles prior to lysing the cell. However, both of these genes are expressed during lysogeny and late lytic growth with no statistical significance, which if expressed during lysogeny would result in premature cell lysis, and Ukulele could not replicate (39).

When incubated at 37 °C, Ukulele *M. smegmatis* lysogens have a titer or $1x10^{10}$ PFUs/ml (8), which is much greater than the other lysogens, like an Ollie *M. smegmatis* lysogen which has a typical titer closer to$1x10^{6}$ PFUs/ml (41). The high concentration of particles in the culture supernatant suggests that Ukulele is a weak lysogen. If Ukulele is a weak lysogen then there would be a high incidence of lytic growth induction from Ukulele prophages. This would result in high levels of lytic gene expression, which was observed in the Ukulele expression profile created by RNAseq. This high level of lytic

gene expression present in the lysogenic expression profile makes it difficult to make definitive conclusions about which genes are essential for establishing and maintaining lysogeny. When incubated at room temperature, Ukulele *M. smegmatis* lysogens have a titer of up to $8x10^9$ PFUs/ml (8). Increased lytic induction may occur at a higher temperature due to a change in host metabolism or due to the change in temperature affecting phage repressor protein stability, although the actually cause is unknown. A less likely, but possible explanation of, lytic expression during lysogeny is that these genes are required during both phases, and therefore expressed during both.

5.3 Ukulele gp52 is not the immunity repressor

Ukulele gp52 is likely an excise or Cro-like protein. It is upregulated during early lytic growth compared to both the late lytic and lysogenic samples, with statistical significance (Fig.9), which is when excise or Cro should be expressed (11). If it functioned as the immunity repressor it would be expressed during lysogeny to silence lytic gene transcription. Also, deletion of gp52 hinders lytic growth (8). When gp52 is deleted from the phage genome, 52 mutants can only be detected in the presence of wild type Ukulele, and pure mutants cannot be isolated (8). Both Cro and excise are vital for lytic growth induction, and either being deleted would result in an inability to exit the host cell. Finally, gp52 is predicted to encode a DNA-binding protein (8) which is necessary to function as a transcriptional regulator, such as a Cro-like protein, and is share predicted structure with the winged helix DNA binding domain of the mycobacteriophage Pukovnik excise (8).

5.4 Characterized lysogenic genes and genes encoding predicted DNA binding proteins are nonessential for lysogeny

Of the genes consistently upregulated during lysogeny, none of them have predicted functions that are necessary for lysogeny (Fig.10), and all are located within predicted lytic operons (Fig.11). The raw read counts agree with the statistical significance, all of these genes having much higher read counts during lysogeny than either lytic growth phase (Table 4). While it is possible that these genes are lysogenic genes, it is also possible that they are present in the lysogenic expression profile due to lytic induction during lysogeny.

Of the five genes predicted to encode DNA binding proteins, none stand out as obvious candidates for the repressor (Fig. 10). Several of the genes have predicted functions that are unrelated to the immunity repressor, and all but gp87 are exclusively expressed during lytic growth (Fig. 8). Although gp87 is predicted to encode a protein with DNA binding domain that is structurally similar to a repressor/lambda repressor-like (Fig.8), previous experiments suggest that gp87 is necessary for lytic growth (8), and it is located within a predicted early lytic operon (Fig.11). Although it is upregulated during lysogeny and early lytic growth with no statistically significant difference (Fig.8), it is likely that its presence in the lysogen samples is due to high amounts of lytic induction.

Table 4. Averaged raw read counts for genes located outside of predicted lytic operons

Gene ID	Early lytic expression	Late lytic expression	Lysogenic gene expression
gp29	22	13461	25017
gp31	2235	10845	14669

5.5 Conclusions and future work

This research has contributed to the characterization of Cluster E lysogeny regulation and overall gene expression profile using Ukulele as a model phage. Using RNAseq and computational analysis, gene expression profiles were created for many of the genes in Ukulele giving insight into gene functions. Ukulele gp52 is not the immunity repressor, and none of the other previously identified genes predicted to encode DNA binding proteins are promising candidates. It is difficult to make definitive claims about the expression profile of the Ukulele lysogen because of high lytic induction, leading to high lytic gene expression detected.

Additional experiments are needed to confirm that Ukulele gp52 is necessary for lytic growth. To confirm that gp52 does not encode the repressor, gp52 will be overexpressed in *M. smegmatis* (pST-KT-gp52) and will be tested for superinfection immunity to Ukulele infection. Given that there are no obvious candidates for the repressor, it might be worthwhile to create a Ukulele29 mutant and observe its ability to form a lysogen, just to rule it out or gather further evidence to support the idea that it might serve as a nontraditional repressor. It would also be advisable that RNA be isolated from a Ukulele *M. smegmatis* lysogen grown at a lower temperature to increase stability and decrease lytic induction. Without reads mapping to lytic genes in the lysogenic expression profile, it would be much clearer which genes are required during lysogeny.

REFERENCES

1. Hatfull G.F. et al (2013). Cluster M Mycobacteriophage Bongo, PegLeg, and Rey with Unusually Large Repertoires of tRNA Isotypes. J.Virol Department of Biological Sciences, Pittsburgh Bacteriophage Institute, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
2. Jacobs WR, Tuckman M, Bloom BR. 1987. Introduction of foreign DNA into mycobacteria using a shuttle phasmid. Nature Publishing Group. Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, New York 10461, USA.
3. Beachem G.M. et al. Mycobacterium phage Ukulele, complete genome. GenBank. Nucleic Acids Res. 2015; http://www.ncbi.nlm.nih.gov/nuccore/918360239
4. Mycobacteriophage Database http://www.phagesdb.org/
5. Ford M.E. et al (1997). Genome structure of mycobacteriophage D29: implications for phage evolution. Journal of Molecular Biology, Volume 279, Issue 1, 29 May 1998, Pages 143–164.
6. Pope W.H. et al (2011). Cluster K Mycobacteriophages: Insights into the Evolutionary Origins of Mycobacteriophage TM4. PLoS ONE. DOI: 10.1371/journal.pone.0026750
7. Cresawn S.G. et al (2015). Comparative Genomics of Cluster O Mycobacteriophage. PLoS ONE 10(3): e0118725. doi:10.1371/journal.pone.0118725

8. Beacham G.M. Complete annotation of the cluster E mycobacteriophage Ukulele genome and characterization of cluster E lysogeny regulation. Honors Thesis Spring 2015.
9. Hatfull G.F. et al (1997). Transcriptional silencing by the mycobacteriophage L5 repressor. The EMBO Journal Vol.16 No.19 pp.5914–5921, 1997.
10. Broussard G.W. et al (2013). Integration-Dependent Bacteriophage Immunity Provides Insights into the Evolution of Genetic Switches. Molecular Cell 49, 237–248, January 24, 2013
11. Oppenheim B.A. et al (2005). Switches in Bacteriophage Lambda Development. Annual Review of Genetics Vol. 39: 409-429. DOI: 10.1146/annurev.genet.39.073003.113656.
12. Mustard JS, Little JW. 2000. Analysis of Escherichia coli RecA Interactions with LexA, λ CI, and UmuD by Site-Directed Mutagenesis of recA. J Bacteriol. Mar 2000; 182(6): 1659–1670.
13. Parikh A. et al (2013). New generation of vectors for mycobacteria. National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi - 110067, INDIA. Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore - 560012, INDIA
14. Derick et al (2013). Functional requirements for bacteriophage growth: Gene essentiality and expression in Mycobacteriophage Giles. Molecular Microbiology. ; 88(3): 577–589. doi:10.1111/mmi.12210.
15. Jacobs WR, Tuckman M, Bloom BR. 1987. Introduction of foreign DNA into mycobacteria using a shuttle phasmid. Nature Publishing Group (ed). Department

of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, New York 10461, USA.

16. Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, Lewis JA, Jacobs-Sera D, Falbo J, Gross J, Pannunzio NR, Brucker W, Kumar V, Kandasamy J, Keenan L, Bardarov S, Kriakov J, Lawrence JG, Jacobs Jr. WR, Hendrix RW, Hatfull GF. 2003. Cell Press. Origins of Highly Mosaic Mycobacteriophage Genomes. Volume 113, Issue 2, 18 April 2003, Pages 171–182.
17. Hendrix, R. W. (2002). Bacteriophages: evolution of the majority. *Theoretical Population Biology*, 61: 471−480.
18. Zhao, M., Gilbert, K., Danelishvili, L., Jeffrey, B. and Bermudez, L.E. (2016) Identification of Prophages within the Mycobacterium avium 104 Genome and the Link of Their Function Regarding to Environment Survival. Advances in Microbiology, 6, 927-941.
19. Mycobacteriophage Database http://www.phagesdb.org/
20. Hendrix, R. W. (2002). Bacteriophages: evolution of the majority. *Theoretical Population Biology*, 61: 471−480.
21. Hatfull, G. F. (2010). Mycobacteriophages: genes and genomes. *Annual Review of Microbiology*, 64: 331−356.
22. Sampson, T., Broussard, G. W., Marinelli, L. J., Jacobs-Sera, D., Ray, M., Ko, C. C., Hendrix, R. W. & Hatfull, G.F. (2009). Mycobacteriophages BPs, Angel, and Halo: comparative genomics reveals a novel class of ultra-small mobile genetic elements. *Microbiology*, 115(Pt9): 2962−2977

23. Dedrick, R. M., Mavrich, T. N., Ng, W. L., Cervantes Reyes, J. C., Olm, M. R., Rush, R. E., Jacobs-Sera, D., Russell, D. A. and Hatfull, G. F. (2016), Function, expression, specificity, diversity and incompatibility of actinobacteriophage parABS systems. Molecular Microbiology, 101: 625–644. doi:10.1111/mmi.13414
24. Actinobacteriophage database <www.phagesdb.org>
25. Hatfull G.F. et al. (2016). Temperate phage as kingmakers of microbial diversity. Unpublished
26. Parikh, A., Kumar, D., Chawla, Y., Kurthkoti, K., et al. (2013). Development of a new generation of vectors for gene expression, gene replacement, and protein- protein interaction studies in mycobacteria. *Applied Environmental Microbiology*, 79(5): 1718-1729.
27. Pope, W., Sarkis, G., Hatfull, G., & Broussard, G. (2013). "Lysogeny experiments."http://phagesdb.org/media/workflow/protocols/pdfs/LysogenyProtocol_3.19.13.pdf
28. Anders, S. et al. (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nature Protocols. Vol. 8: 1765-1786.
29. James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. Integrative Genomics Viewer. Nature Biotechnology 29, 24–26 (2011)
30. Singer, J. T. 1989. Molecular cloning of the recA analog from the marine fish pathogen Vibrio anguillarum 775. J. Bacteriol., 171(11): 6367-6371.

31. Schofield D.A., Sharp N.J., Westwater C. (2012). Phage based-platform for the clinical detection of human bacterial pathogens. Bacteriophage. Apr 1; 2(2): 105–283.

32. Haley C.B. Characterization of transcriptional control elements in Custer E mycobacteriophage Ukulele. Biochemistry and Spanish Honors Thesis Spring 2016.

33. Hatfull, G. F. (2012). The Secret Lives of Mycobacteriophages. *Adv. Virus Res.* 82: 179−288.

34. Söding J., Biegert A., Lupas AN. (2005) The HHpred interactive server for protein homology detection and structure prediction.Nucleic Acids Res 33 (Web Server issue), W244-W248. PMID: 15980461

35. Kelley LA et al. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols 10, 845-858.

36. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403−410.

37. Cresawn, S., Bogel, M., Day, N., Jacobs-Sera, D., Hendrix, R. W., & Hatfull, G. F. (2011). Phamerator: a bioinformatics tool for comparative bacteriophage genomes. *BMC bioinformatics*, 12(12):395. Doi: 10.1186/1471-2105-12-395.

38. Broxmeyer L. et al. (2002). Killing of Mycobacterium avium and Mycobacterium tuberculosis by a Mycobacteriophage Delivered by a Nonvirulent Mycobacterium: A Model for Phage Therapy of Intracellular Bacterial Pathogens. J Infect Dis (2002) 186 (8): 1155-1160.

39. Payne K.M., Hatfull G.F. (2012). Mycobacteriophage Endolysins: Diverse and Modular Enzymes with Multiple Catalytic Activities. PLOS ONE 7(3): e34052.https://doi.org/10.1371/journal.pone.0034052
40. Tarazona S., Garcia-Alcalde F., Dopazo J., Ferrer A., Conesa A. (2011). Differential expression in RNA-seq: A matter of depth. Genome Res. 2011. 21:2213-2223. doi:10.1101/gr.124321.111.
41. Soohey R., unpublished.
42. King B.L., personal communication.

AUTHOR'S BIOGRAPHY

Robert Soohey was born in Augusta, Maine on November 7, 1994. He was raised in Whitefield, Maine and graduated from Erskine Academy in 2013. He is majoring in microbiology, and has received a Radke Undergraduate Research Fellowship, Charlie Slavin Research Award, and Junior and Senior INBRE Research Fellowships. Upon graduation, Robert plans to attend medical school and practice medicine in Maine.