Spring 5-2017

Characterizing the Intact Prophage of Mycobacterium chelonae Bergey

Erica Sewell
University of Maine

CHARACTERIZING THE INTACT PROPHAGE OF MYCOBACTERIUM CHELONAE BERGEY

by

Erica J. Sewell

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:

Keith W. Hutchison, Professor Emeritus of Biochemistry and Molecular Biology
Sally D. Molloy, Assistant Professor of Genomics
John T. Singer, Professor of Microbiology
Melissa S. Maginnis, Assistant Professor of Microbiology
Stefano Tijerina, Adjunct Assistant Professor in Political Science

ABSTRACT

Mycobacteriophage (phage), are viruses that infect bacteria. All bacteria can be infected by phage, and each bacterial species has a unique set of phage that infect them, making phage prime candidates for studying viral diversity and evolution. Some phage integrate their genome into the host genome upon infection (prophage), where they may potentially remain indefinitely, coevolving with the host, and providing growth factors and other benefits to the host. The purpose of my research is to characterize a prophage within the genome of the bacterial host *Mycobacterium chelonae* Bergey to determine if it is still functional and potentially impacting the fitness of the host bacterium. Characterization of this prophage has revealed that multiple genes are conserved with regard to both the DNA and protein sequences. The integrase cassette is highly conserved, complete with integrase and two potential repressors, suggesting the phage may be capable of excising from the host genome. Multiple structural genes including capsid and tail proteins are also conserved, suggesting the prophage may be capable of producing intact virions. At least two prophage genes are transcriptionally active. These include a predicted repressor and transmembrane protein. Expression of these genes suggests that the prophage does indeed have some potential for affecting the biology of the host bacterium. Experiments are currently underway to determine if intact virion particles are being produced during bacterial growth.

ACKNOWLEGEMENTS

I would first like to thank my advisor, Keith Hutchison, for his guidance and support through the past four years. This thesis began as a much smaller project from my freshman year of college, and has grown with the help of Keith's encouragement and expertise. Weekly meetings with Keith have always resulted in new ideas and never fail to end without a smile.

Next I would like to thank Sally Molloy for her help during lab meetings and in the lab over the course of this project. I would like to thank my thesis committee members for their time and advice.

I would like to thank my sources of funding which include the Honors College Comparative Functional Genomics Thesis Fellowship and Junior Year Research Award under NIH-INBRE Grant 8P20GM1003423-12 through the Maine INBRE program, and the 2016-17 Charlie Slavin Undergraduate Research Grant.

Lastly, I would like to thank my family for their un-ending love and support. My parents, Jon and Kelley Sewell, have encouraged and inspired me from the beginning. My sister, Marissa Sewell, has been a life-long best friend and cheerleader. My grandparents, Dorothy and Keith Kinney, have provided wisdom and hugs especially during the tough times. I would not be where I am today without such wonderful people to surround myself with.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 Introduction

Phage are viruses that infect bacterial hosts. There are an estimated 10^{31} phage in the environment, making phage the most abundant biological entity on earth (18). The 10^{31} phage can be divided up into an estimated 100 million different species of phage, each with genomes ranging from 15,000 bp to 165,000 bp in length. This high level of genetic diversity makes them prime candidates for research in the areas of genomics and evolution. Mycobacteriophage are a group of phage that infect the genus *Mycobacterium*. These viruses are divided into groups known as clusters based on genome similarity. There are currently 1366 sequenced phage, which are divided up into 26 different clusters (A-Z) and subclusters (18). Mycobacteriophage replicate by injecting their DNA into a host cell, then follow one of two different life-cycles; lytic or lysogenic. In the lytic pathway the viral DNA circularizes and uses the host cell machinery to produce progeny, eventually lysing the cell and releasing new viruses into the environment. For the lysogenic pathway, the viral DNA is incorporated into the host genome, and replicates as part of the host genome when the cell divides. Integrated phage are called prophage, the infected host is known as a lysogen. When the host cell is under stress, the phage genome will excise from the host genome and enter into the lytic pathway. However, if the host bacterium does not encounter a stressful environment, the integrated prophage has the opportunity to remain and co-evolve along with the host. Selective pressure will dictate which phage genes are conserved and will retain their function, and which phage genes are silenced and eventually lost.

It is not uncommon for bacterial genomes to contain prophage or vestigial prophage. Many of the most highly pathogenic strains of bacteria, including *Escherichia*

coli 0157 and *Vibrio cholerae*, have retained and utilized certain phage genes such as toxins to increase their pathogenicity in hosts (32, 42). Over time after integration, prophage genomes may begin to decay, mutate, and rearrange. It has been hypothesized that by retaining key phage genes, such as repressors, transmembrane proteins, and toxin-antitoxin systems, the bacteria enters into a state of superinfection immunity where it is protected from further infection by other phage. This and other possible advantages have been proposed as an explanation for why the host bacteria would retain viral genes.

This line of research began with the discovery and characterization of two small prophage regions in *Mycobacterium smegmatis* MC2 155, both fewer than 11,000 bp in length. This strain is used in the SEA-PHAGES course to isolate novel mycobacteriophage from the environment (23). These prophage regions were characterized and primers were designed to determine transcriptional expression of the phage-encoded transposase gene, with no evidence of expression found. This thesis includes a summary of the characterization of the *M. smegmatis* MC2 155 prophage regions, as well as RNAseq analysis for the bacteria.

During the course of this study the genome of *M. chelonae* Bergey became available for analysis and focus shifted from the *M. smegmatis* prophage regions to this newly sequenced genome. *M. chelonae* Bergey contains an intact prophage consisting of 95 open reading frames, of which at least two are transcriptionally active. The prophage is organized as most mycobacteriophage appear within the host, with structural and assembly genes on the left arm and nonstructural and replication genes on the right arm (18). Investigation of this prophage has provided insights into the function and role of mycobacteriophages in the superinfection immunity and virulence of their bacterial hosts.

The discovery of *M. chelonae's* prophage represents yet another phage-host combination with which to better understand the effects of viral infection and lysogeny on individual bacterial hosts and on total bacterial populations.

2.0 Literature Review

2.1 History of Bacteriophage Research

Bacteriophages were co-discovered in 1915 and 1917 by Frederick Twort and Felix d'Herelle respectively. D'Herelle coined the name bacteriophage using the root Greek word "phagein" meaning to eat, with the name bacteriophage literally meaning bacteria-eater. The first mycobacteriophage was isolated with *M. smegmatis* in 1947. Mycobacteriophages were thought to be very limited in total effect on bacterial populations because of the comparatively late discovery of this type of virus (13). However soon after this initial discovery, mycobacteriophages that infected *M. tuberculosis* were discovered (12) and mycobacteriophage research began to greatly increase.

Mycobacteriophage have traditionally been involved in typing strains of bacteria, particularly clinical isolates of *Mycobacteria*. Early on in molecular biology, mycobacteriophage were used as tools to alter and manipulate gene expression in the host bacteria. In the 1980's mycobacteriophage were used as shuttle vectors which could be manipulated in *E. coli* and used to deliver novel DNA to mycobacteria. (19). This type of research holds great promise in better understanding the physiology and virulence of mycobacterial hosts.

Phage therapy is a particularly interesting area of bacteriophage research. It involves exploiting the ability of phages to lyse and destroy bacteria as a result of their lytic infectious cycle in order to fight off infectious bacterial diseases. This treatment could be a valuable tool in fighting off bacterial culprits such as tuberculosis, especially in light of the antibiotic resistance crisis. Most early experiments in this area of research were unsuccessful in treating tuberculous diseases. However in some parts of the world, particularly in eastern Europe, phage therapy is commonly used preventatively and in the treatment of purulent and enteric pathogens (26). More clinical trials are needed to expand knowledge and use of phage therapy as a legitimate treatment option in more countries.

In 2008 the SEA-PHAGES program introduced the use of mycobacteriophage as a learning tool in the college-level STEM curriculum (23). To date, 1366 unique mycobacteriophage genomes have been sequenced, offering valuable insight into evolution, particularly phage-host coevolution, and the effect these small DNA viruses have on modulating and controlling the total mycobacterial population. Powerful computational analysis combined with traditional molecular biology has revealed great diversity in mycobacteriophage (17). These tools have also allowed researchers to ask broader questions about mycobacteriophage evolution, and coevolution, with their mycobacterial hosts.

2.2 Significance of *Mycobacterium smegmatis* and *Mycobacterium chelonae*

M. smegmatis is a common largely non-pathogenic soil bacterium that may safely be used to isolate novel mycobacteriophage. It has been reported as a human pathogen in very rare instances including traumatic injury or surgery and is known as a rare pathogen

in immunocompromised patients (33). *M. smegmatis* is well studied and has been the subject of many research articles. This bacterium was therefore the first used in this research to study the coevolution between mycobacteriophages and their hosts.

M. chelonae is a fast growing non-tuberculous bacteria that commonly causes infection after intrusion or injury to the skin, such as plastic surgery or injections (43). It is also an opportunistic pathogen in individuals with HIV/AIDS and those with lung cancer and lung diseases (44). *M. chelonae* is a member of a group of mycobacteria that includes *M. abscessus* and *M. fortuitum*. Infections with these mycobacteria are notoriously difficult to treat as a majority are resistant to tetracyclines and most other antimicrobial agents (37). This bacterium is more commonly known for its role as a fish pathogen and is one causative agent of piscine mycobacteriosis, along with *M. marinum* and *M. fortuitum* (14).

M. chelonae was originally isolated in 1923 from a tortoise tubercule by David Hendricks Bergey, author of Bergey's Manual of Determinative Bacteriology. The draft *M. chelonae* Bergey genome sequence was published in GenBank in 2015 (16). A more complete sequence was published in 2016 (21).

2.3 Introduction to coevolution of phages and their bacterial hosts

Upon infection and entrance into the lysogenic lifestyle, phages are thought to lay dormant, simply replicating as part of the host genome. However further investigation has revealed that prophage can have a significant impact on the fitness and pathogenicity of their bacterial hosts.

As mentioned earlier, there is an absence of selective pressure in the bacterium to maintain functionality of phage DNA that does not confer any fitness factors to the host.

This results in defective or "cryptic" prophage regions, where genes coding for phage specific functions such as lysis and phage-particle formation are no longer complete and functional. Some prophage code for so-called fitness factors, such as toxins, that benefit the host with increased pathogenicity or superinfection immunity. These genes have been shown to be conserved and expressed within multiple bacterial hosts including *V. cholera, E. coli* and *Staphylococcus aureus* (15). It is interesting to note that the most virulent strains of bacteria often contain prophage remnants or phage encoded toxins. Toxin genes are the most well-known phage-encoded genes known to confer increased virulence to the host. Other examples of beneficial phage-encoded genes include toxin-antitoxin systems, which provide the benefit of an apoptosis-like response to cells. This acts to limit total viral infection in a bacterial population (11). Superinfection immunity can be provided to the host by phage-encoded transmembrane and repressor proteins (40, 9).

2.4 Well known and studied phage toxins conserved in bacteria

V. cholerae is a gram-negative bacteria that normally resides in brackish water. When ingested by humans, *V. cholera* causes the disease cholera which is characterized by severe dehydration, diarrhea, and vomiting. Cholera has two main virulence factors, the cholera toxin (CT) and toxin-corrugated pili (TCP) which lie on the phage CTXφ. This phage is known to code for at least 6 genes including cholera toxin subunits A and B, a core-encoded pilin, as well as an accessory cholera enterotoxin. *V. cholerae's* notable virulence is due to this phage-encoded pili, which allows establishment of the bacterium in the intestinal epithelium. The CTXφ element is capable of producing phage particles which present as curved bundled filaments under electron microscopy. These

phage particles have been shown to contain single stranded DNA corresponding to the CTXφ element (42).

Enterohemorrhagic *E. coli* 0157:H7 produces the phage encoded shiga toxin, which is the cause of a type of food poisoning in humans known as hemorrhagic colitis. This illness is characterized by painful cramping, dehydration, and diarrhea and can be deadly if left untreated (22). The shiga toxin is a type of exotoxin that originates from a group of phages known as stx (shiga toxin) converting phages (38). Lysogenization of novel *E. coli* strains with stx converting strains successfully converts these strains to produce functional shiga exotoxins (30).

Another significant example of phage encoded toxins in Gram positive bacteria is the exfoliative toxin in *S. aureus*. If infected with this pathogen, young children will develop skin splitting caused by exfoliative toxins A and B (ETA and ETB). ETA is located directly on the bacterial chromosome while ETB is located on a separate plasmid (45). These phage encoded exotoxins exhibit super antigenic capabilities, indicating the potential for dual functionality (6). In the ETA producing strain *S. aureus* E-1, mitomycin C was used to induce phage particle production of phage φETA. Phage particles complete with a hexagonal head and flexible tail fibers were visualized under electron microscopy. The 43081 bp phage genome was sequenced and found to contain several highly conserved genes including integrase, a repressor, and the gene encoding ETA which was located at the end of the phage genome near attP. Other strains of *S. aureus* were converted to ETA producers after becoming φETA lysogens (45).

2.5 Phage encoded genes in *Mycobacterium tuberculosis*

M. tuberculosis is often associated with the extreme morbidity and mortality that it causes in third world populations and in immunocompromised individuals. It is the most well-known member of the *Mycobacterium* of the tuberculosis complex, otherwise known as the causative agents of human and animal tuberculosis. This illness is characterized by symptoms including chronic cough, blood in the sputum, and weight loss. *M. tuberculosis* bacteria will make way to the lungs and become engulfed by immune cells such as macrophages. The phagocytosed *M. tuberculosis* bacterium possesses a thick mycolic acid capsule that protects it from destruction by reactive oxygen species and acids inside the macrophage. The bacteria is then capable of reproduction inside the immune cell and will go on to eventually kill the cell. This process continues while all types of immune cells migrate to the site of infection, but when new macrophages migrate and attempt to engulf and destroy the infected macrophage, the two fuse and form an aggregated multi-nucleated cell, otherwise known as a granuloma. Eventually immune cells build up around multiple granulomas and form the key characteristic of an active tuberculosis infection, a tubercule.

There are two *M. tuberculosis* strains of significance in explaining the role of phage genes in the biology of *Mycobacteria*, *M. tuberculosis* H37Rv, a well-studied laboratory strain, and CDC1551, a clinical isolate. *M. tuberculosis* H37Rv contains a 9427 bp region, RD3, that is not found in *Mycobacterium bovis*, which is an avirulent strain used to make a vaccine against *M. tuberculosis*. Within this RD3 region are five genes with similarity to phage genes encoding protein capsid, protease, terminase, primase, and integrase, together collectively known as phiRv1. This segment of phage-

like similarity is too stunted to be able to produce fully intact phage particles, but is interesting because it lies in a repetitive element REP13E12 that is repeated seven times throughout the genome. In *M. tuberculosis* CDC1551, a nearly perfect copy of phiRv1 is also present, but lies in a different REP13E12 repeat. Subsequent experiments showed that the phiRv1 is mobile element capable of integration and excision in and out of the REP13E12 repeats. *M. tuberculosis* H37Rv contains a second prophage element, phiRV2, integrated at a different REP13E12 repeat. It is not yet known what function these mobile phage elements have in the *M. tuberculosis* genome, but it is possible that these mobile elements are in some way beneficial to the bacterium due to the strong conservation of both sequence and function in this seemingly derelict prophage (4).

2.6 Significance of toxin-antitoxin systems

Toxin-antitoxin (TA) systems are found in bacterial genomes and are thought to function as an anti-bacteriophage defense, in addition to playing other cellular roles. TA systems promote cell death during infection to limit overall phage infection in the bacterial population. In RNA-RNA systems, the antitoxin will inhibit the translation of the toxin, thereby suppressing a cell death response. The toxin is translated only when the antitoxin has been degraded in response to some stressor. Free toxins target and lead to the inhibition of central cell processes such as DNA replication. Inhibition of these crucial mechanisms eventually leads to an “altruistic” cell death in the case of viral infection (10). Phage are known to carry these types of systems as a resistance mechanism against the host TA systems (8). Although it has not been shown experimentally, it can be hypothesized that a prophage encoded TA system would offer the host protection against subsequent phage infections.

2.7 Significance of phage-encoded transmembrane proteins and repressor proteins as contributors to superinfection immunity

Superinfection immunity is the state in which a lysogen is immune to infection by viruses that are genetically similar to the currently integrated prophage. To understand the mechanism behind superinfection immunity, the decision between lytic and lysogenic life cycles must first be understood. Phage Lambda lysogeny is one of the most highly characterized systems. The decision in Lambda between the two life cycles, lytic or lysogenic, depends on the amounts of two different phage-encoded proteins; Cro and CI. When CI prevails and binds to the operator regions O_R, the lysogenic lifecycle will initiate and the phage will integrate into the host genome. After integration, the CI repressor is produced to maintain lysogeny and also acts to recognize and bind to the DNA of genetically similar viruses, marking them for destruction (9). This hinders infection by other viruses and acts as a mutually beneficial defense for both phage and host by offering the host protection from other phage that may infect, enter the lytic cycle, and destroy the cell.

Phage encoded transmembrane proteins have been shown to provide superinfection immunity to the host bacteria by interfering with viral binding and injection of viral genomic material through altering the appearance of the host cell membrane. This has been shown in a *Streptococcus thermophilus* lysogen with temperate phage TP-J34 where the phage-encoded *ltp* gene, a lipoprotein, was associated with the cell membrane. The *ltp* protein provided limited superinfection immunity by presenting on the outside of the cell and interfering with phage DNA release and/or injection. Complete immunity was not provided, but plaque formation in lysogens was reduced by a

factor of 40 (40). Phage-encoded transmembrane proteins have also been found to provide superinfection immunity in *E. coli* where gp15 of the HK97 prophage was expressed and located on the inner bacterial membrane. This transmembrane protein was predicted to interact with the incoming phage tail to prevent the formation of a channel through the cell membrane. This particular example is interesting because unlike traditional repressor-mediated immunity, gp15 has been found to provide immunity against phage belonging to different immunity groups (7).

3.0 Materials and Methods

3.1 Bioinformatic Analysis

A number of different software packages and online databases were used over the course of this project. NCBI (31) was used to access nucleotide sequences and to utilize the BLAST functions blastn and blastp. This program was used to find gene functions during characterization of the prophage regions, and to compare attP and attB sequences. Phast (46) is a program that references two different protein databases; the NCBI phage database, and a separate database including other proteins from phage genomes not included in NCBI. Phast was used to identify potential prophage regions and to access nucleotide and amino acid sequences for predicted genes. Phyre2 (24) predicts and analyzes protein structure based on a given amino acid sequence. It was used to predict function and to visualize folding structures of prophage amino acid sequences. HHpred (39) references multiple protein structure databases to detect structural homology and subsequently predict potential protein function. It was used to assist in assigning potential functions to prophage genes. GenemarkS (3) was utilized to predict coding potential for

prophage open reading frames, and was used to generate nucleotide and amino acid sequence files for all potential *M. chelonae* Bergey prophage genes.

3.2 Growth of Bacterial Strains

M. chelonae strains, Bergey, and isolates from Fat Head Minnow, flounder (F5), and Bowfin were grown at 30°C with shaking in 7H9 media, supplemented with OAD supplement (oleic acid 0.05%, bovine serum albumin 5%, dextrose 2%) and 50 µg/ml carbenicillan (CB) and 10 µg/ml cyclohexamide (CX) or on OAD 7H10 agar plates.

3.3 DNA Isolation

Mycobacterial DNA was isolated by harvesting cells from exponential to late log bacterial cells. These cells were centrifuged at maximum speed for 1 min and pellets were resuspended in 0.5 mL of lysis buffer (50 mM Tris-Cl, 5 mM EDTA, pH 8, lysozyme 1 ng/µl). The solution was incubated for 1.5 hrs at 37 °C. Afterwards, 50 µl 10% SDS page were added, tubes were mixed by inversion, and 2.5 µl proteinase k (20 mg/ml) were added to each tube. Solutions were incubated at 65 °C for 10 min. DNA was extracted using equal volumes of TE (10mM Tris, 1mM EDTA, pH 8.0) saturated phenol and chloroform and tubes were vortexed and centrifuged for 3 min at maximum speed at room temperature. The aqueous layer was transferred to a clean 2 ml microcentrifuge tube and one half volume of 7.5 M ammonium acetate and 2 volumes of 100% ethanol were added. The solution was vortexed and incubated overnight at -20 °C, then centrifuged at 20,000 x g for 30 min at 4 °C. The supernatant was removed and the pellet was washed with ice cold 70% ethanol. The pellet was allowed to dry and then was resuspended in TE.

3.4 Primer Design

All primers were designed using the web program Primer3Plus (37). Two sets of primers were designed to amplify the left and right ends of the integrated prophage in *M. chelonae* strains. Each set of primers contained one primer flanking the prophage by ~250 bp away from the attL or attR sequence and one primer within the prophage sequence ~250 bp away from the attL or attR sequence. Gene specific primers were designed to amplify prophage genes of interest. Specific primer sequences and parameters can be found in Table 1.

Table 1. Primer sequences and parameters.

Primer Name	Oligo Sequence (5' – 3')	Melting Temperature	Primer Length	Expected cDNA length
MCProphageLF	TCCTCGTTCTCGTACACCTTC	56	21	-
MCProphageLR	CACTCGTACGGCTTGACG	55.9	18	-
MCProphageRF	GACATCAACACCCCCGACTA	56.5	20	-
MCProphageRR	ATAACTTTCGGCGGTTCCTT	54.5	20	-
McProphage_gp3F	ACACTCGGCAAGCTGTTCTT	57.2	20	162 bp
McProphage_gp3R	GGACGGCAGAGCATAGTTTC	55.8	20	
McProphage_gp7F	TGATCCACAAGGCATATCCA	53.6	20	150 bp
McProphage_gp7R	CATCCGAGCACTACCATCCT	56.4	20	
McProphage_gp52F	TATGTTGCGCTTGCGATTAG	53.5	20	198 bp
McProphage_gp52R	TGGGATTTCCGAGTTACCAG	54.4	20	
McProphage_gp54F	TCAACGTGAAGCTCGATGTC	55.0	20	202 bp
McProphage_gp54R	GGCACTTTCAGTGTCCCTGT	57.6	20	
McProphage_gp55F	TCGAGGTCAAGAACCTACCG	56.1	20	217 bp
McProphage_gp55R	GGGCTGAAAAAGATTCGACA	53.2	20	
McProphage_gp67F	CATCGACTACGACCAGCAGA	56.4	20	188 bp
McProphage_gp67R	TTCTCCAGCGACTTCTGGAT	55.7	20	

3.5 Polymerase Chain Reaction

PCR was performed in 25 µl volumes containing 1 ng mycobacterial DNA, 0.5µM of each primer, and Promega PCR mastermix with Taq polymerase (Promega, Madison, WI). Reactions were incubated as follows; 95 °C for 5 min, then cycled 35 times at 95 °C for 30 s, 55 °C for 40 s, 72 °C for 1 min, then finally 72 °C for 5 min and 4 °C indefinitely.

3.6 Agarose Gel Electrophoresis

Nucleic acid PCR products were fractionated on 1.5% (w/v) SeaKem Agarose (Lonza, Basel, Switzerland) gels, which were stained with ethidium bromide (0.5 mg/ml). The following were added to each well for a total volume of 10µl; 4µl nucleic acid sample, 4µl nuclease-free water, and 2µl loading dye. For ladders, the following were added to each well for a total of 10µl; 7µl nuclease-free water, 2µl loading dye, 1µl 100bp ladder. Gels were run in TAE buffer (40 mM Tris, 20 mM acetic acid, 2 mM EDTA) at a constant voltage of 90 to 110 V. Gels were visualized under UV light.

3.7 DNA Sequencing

Samples to be sequenced were prepared in sterile water at a concentration of 10ng/µL and sent to the UMaine DNA Sequencing Facility. The facility is equipped with an ABI 3730 DNA Sequencer with the XL Upgrade. Primers were provided at a concentration of 5µM.

3.8 RNA Isolation

Five ml cultures were grown to saturation and used to inoculate two 50 ml cultures. These were grown to OD_{600} of 1 overnight. Cells were centrifuged for 3 min at 5500xg at room temperature. Supernatent was saved and the pellet was resuspended in

phage lysate at an MOI of 3 and incubated for 10 min at room temperature. One cell pellet was resuspended in phage buffer for a control. The remaining supernatant was poured back into the flask after 10 min and incubated in a shaker at 37 °C. Cells were harvested in triplicate at 0 min from the control flask, and from virus-treated flasks at 30 and 150 min. In triplicate for each time point, 3 ml of cell culture were added to 6 ml of RNA Protect Bacteria Reagent (Qiagen, Valencia, CA) in conical tubes, vortexed for 5 sec, and incubated at room temperature for 5 min. Samples were centrifuged for 1 min at 5000xg and supernatant was removed. RNA was isolated from cell pellets and was resuspended in 100 µl RNAse-free TE containing 100 µg/ml lysozyme using the RNeasy Mini Kit according to the manufacturers recommendations (Qiagen) During isolation, RNA was treated with DNAse using the RNAse-free DNAse set (Qiagen). After RNA isolations were complete, RNA was treated a second time with DNAse using the Turbo DNA-free kit (Life Technologies, Thermo Fisher, Waltham, MA) according to the manufacturers recommendations.

3.9 cDNA Synthesis

Mycobacterium chelonae Bergey cDNA was made from RNA using the iScript Reverse Transcription Supermix kit according the manufacturer recommendations (Bio-Rad Laboratories) to conduct RT-PCR.

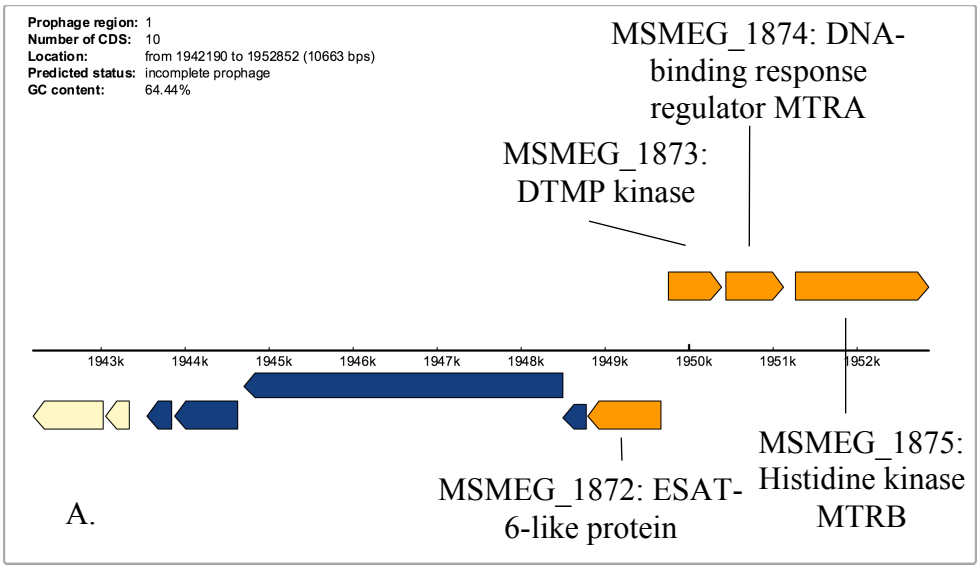
3.10 Spot Testing

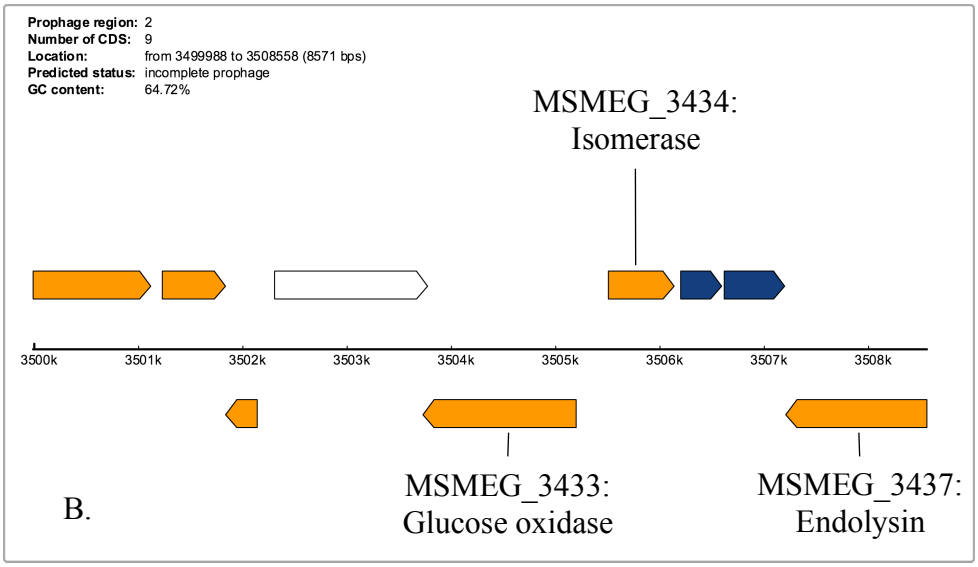
Spot testing was conducted by inoculating 4.5 ml of warm 7H10 agar with 0.5 ml of *M. chelonae* Bowfin culture and plating on agar plates. After plates had cooled, 10 µl *M. chelonae* Bowfin culture was pipetted onto the plates at 10^{0}, 10^{-1}, 10^{-2}, and 10^{-3} dilutions.

4.0 Results

4.1 *M. smegmatis* MC2 155 contains two prophage regions.

M. smegmatis MC2 155 contains two prophage regions each less than 11,000 bp in length. Prophage region 1 (Figure 1A) spans 10663 bp and consists of 10 genes of differing levels of sequence similarity, including predicted transposase genes, and a region of divergent transcription containing multiple transcriptional regulators and regulators of DNA synthesis. Genes MSMEG_1872 through MSMEG_1875 are related to corresponding genes in *M. tuberculosis*. MSMEG_1872 appears to be a component of a well-known gene cluster in *M. tuberculosis*, ESAT-6, that encodes a type VII secretion system which promotes cell to cell migration during infection (1, 21). MSMEG_1874 and MSMEG_1875 are homologous to MtrA and MtrB (28) which is a two-component system found in *Cornebacterium*, *Mycobacterium*, and *M. tuberculosis* in particular, that responds to environmental stress and regulates cell wall permeability (29). The MtrAB system has been shown to be especially important in regulating growth during *M. tuberculosis* infection of macrophages (11). It should be noted that the MtrAB system is conserved in both *M. tuberculosis* and *M. leprae* but are surrounded by different genes, indicating that the prophage either brought the MtrAB system to *M. smegmatis*, or the phage integrated adjacent to this two component system. Prophage region 2 (Figure 1B) spans 8571 bp. This region contains fewer readily identifiable phage genes, but does contain 9 total genes including conserved phage proteins such as endolysin and an isomerase.





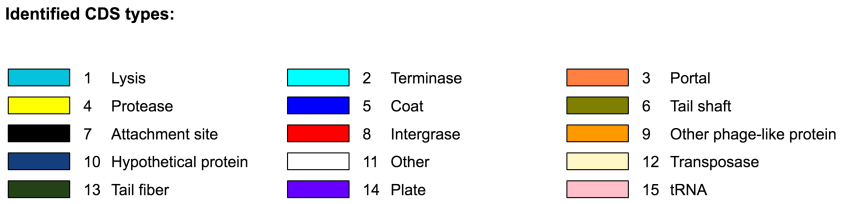


Figure 1. Prophage regions 1 (A) and 2 (B) gene maps. Forwardly transcribed genes are represented by forward facing arrow, genes transcribed in reverse are represented by

backward facing arrows. General predicted gene functions can be found using the color key at the bottom of the figures. Gene maps were generated using Phast (46).

4.2 *M. smegmatis* MC2 155 predicted prophage gene expression changes during lysogenic and lytic infection by phage Ukulele.

RNAseq data recently became available for *M. smegmatis* MC2 155 courtesy of the Molloy laboratory. This data includes the quantification of mRNA and tRNA from *M. smegmatis* lysogens of mycobacteriophage Ukulele, and RNA isolated from lytic infections of *M. smegmatis* with Ukulele at 0 (uninfected), 30 and 150 min post infection. This data was initially visualized using the Integrative Genomics Viewer (35) to compare expression levels (Figure 2).

The expression of multiple prophage encoded genes changes in the Ukulele lysogen compared to the uninfected control. Prophage genes MSMEG_1873, MSMEG_1874, and MSMEG_1875 are all downregulated in the Ukulele lysogen, while MSMEG_1872 is slightly upregulated. All four prophage genes are downregulated 30 min post infection with Ukulele. At 150 min post infection MSMEG_1872, MSMEG_1873, and MSMEG_1874 return to nearly the same expression levels as in the uninfected control while MSMEG_1875 is upregulated, but still falls below expression levels of the control (Figure 3).

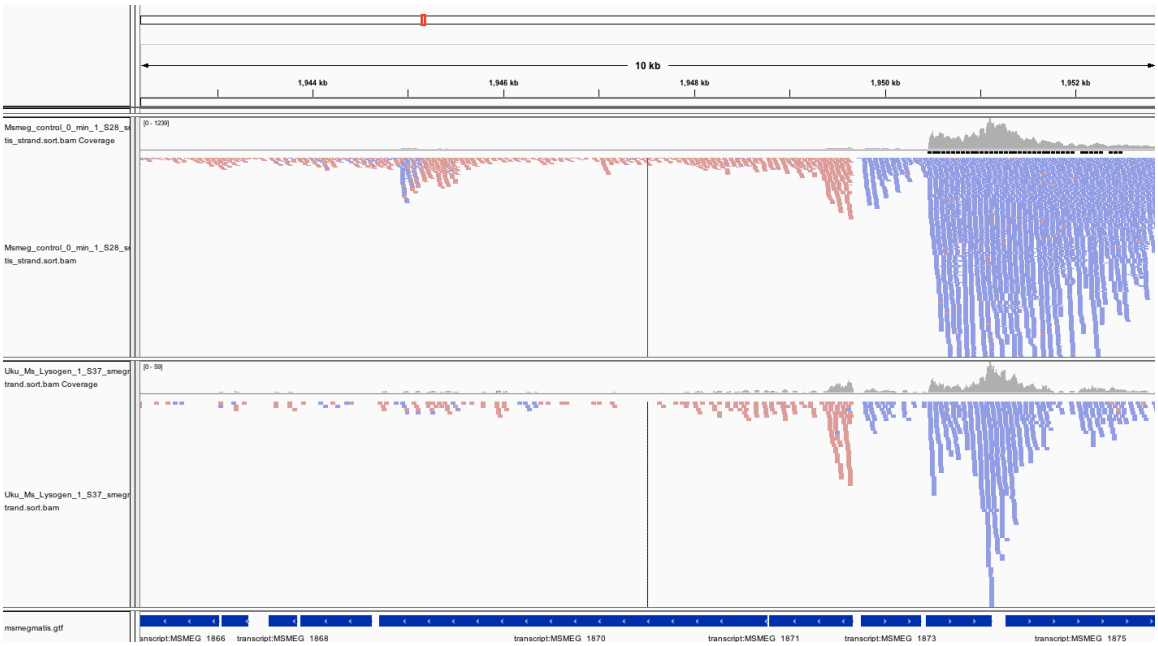


Figure 2. *M. smegmatis* prophage region 1 RNAseq data. The entire prophage region 1 is shown, with the uninfected control in the top panel compared to the Ukulele lysogen in the bottom panel. Forwardly transcribed reads are shown in purple. Visualization of the RNAseq data was done using the Integrative Genomics Viewer (35).

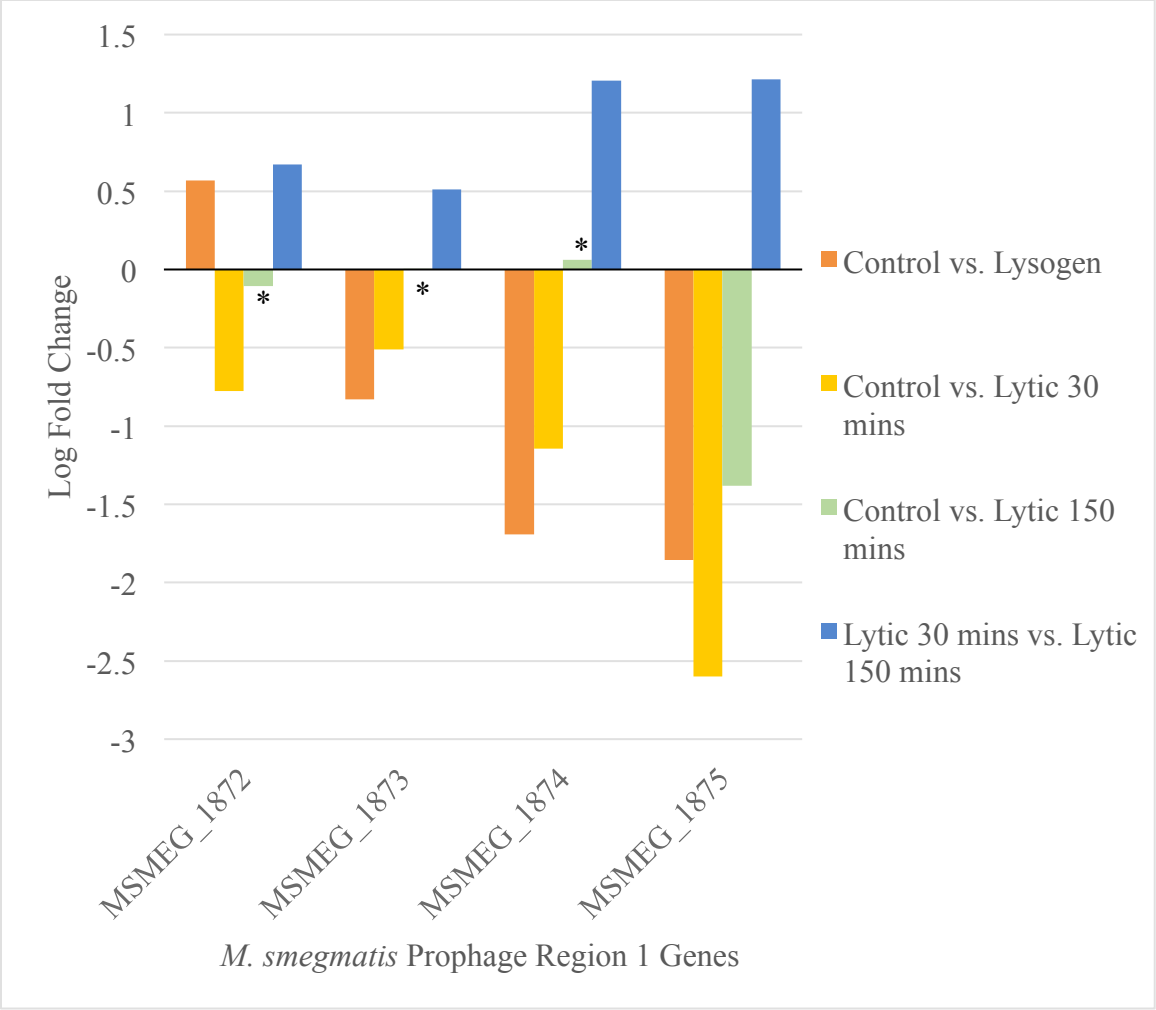


Figure 3. Prophage region 1 gene expression is modulated during lysogenic and lytic infection. Down-regulation is represented by a negative log fold change, up-regulation is represented by a positive log fold change. Insignificant values are denoted by an asterisk.

4.3 An intact prophage was found in *M. chelonae* Bergey.

Analysis of the newly sequenced *M. chelone* Bergey genome with Phast (46) revealed a region spanning 68, 224 nucleotides (Figure 4), containing 95 open reading frames with significant similarity to phage genes. The ORF's with readily identifiable functions were annotated using a variety of bioinformatics programs analyzing both nucleotide and amino acid sequences, and are described in the Appendix. The genes are numbered according to traditional phage genome organization and gene designations, using the predicted genome organization of the excised prophage. The prophage is organized with structural related genes on the canonical left arm, including terminase, portal protein, tail and capsid proteins, and lysin A and lysin B. The canonical right arm consists of replication related genes including primase and DNA polymerase. The integrase cassette appears mostly intact, with a highly conserved integrase gene and two potential repressors, gp54 and gp55. Multiple ORFs with potential transmembrane domains were identified, including gp3 and gp67 (Figure 4).

The prophage contains genes related to a variety of mycobacteriophages, the most common being singleton DS6A. DS6A is known to only infect *Mycobacterium* of the *M. tuberculosis* complex, including *M. bovis*, *M. africanum*, *M. canetti*, and *M. microti* (19).

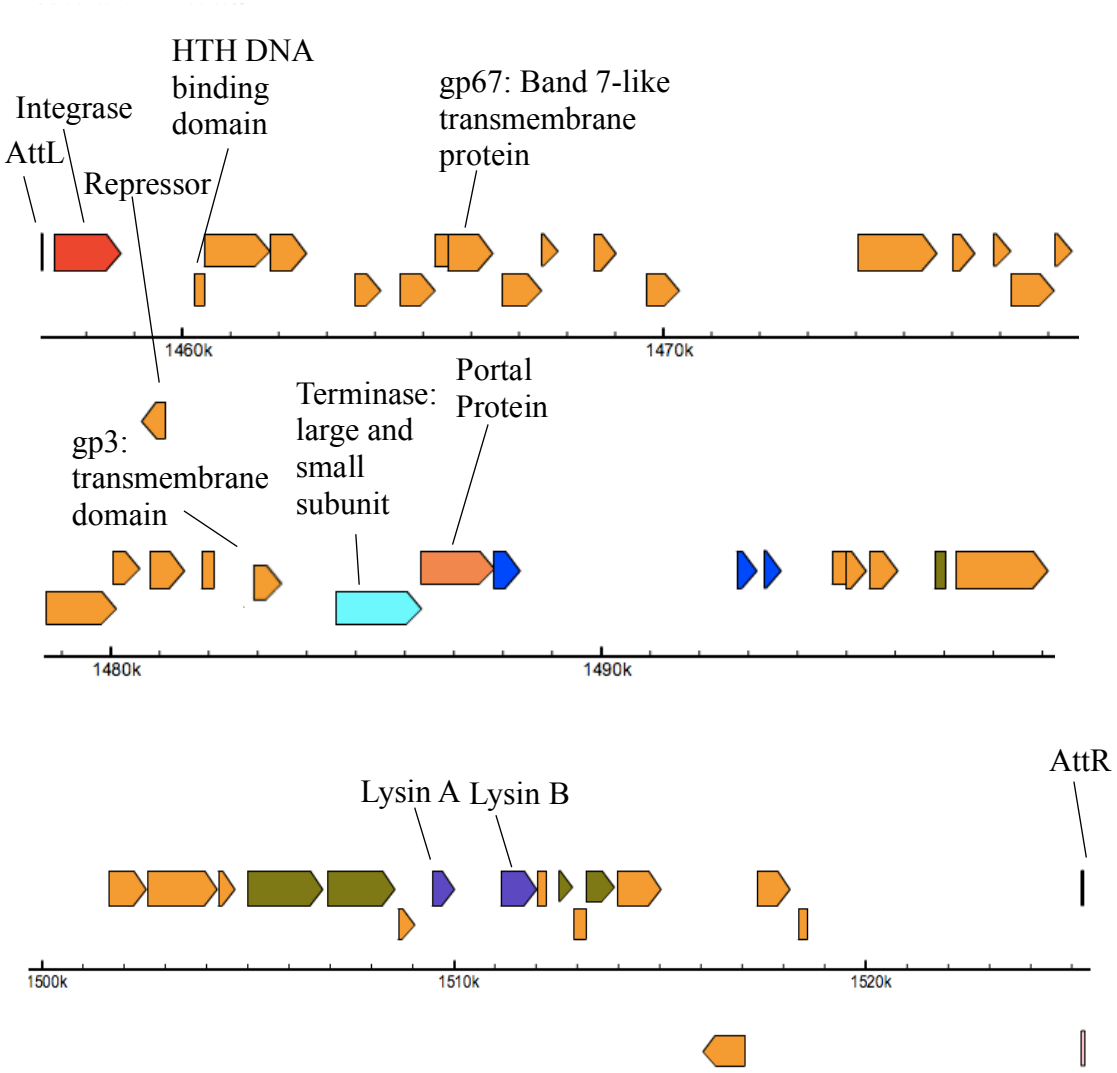


Figure 4. Gene map of the *M. chelonae* Bergey prophage. Not all predicted open reading frames are shown. Important genes are annotated. The integrase cassette consists of integrase, a repressor (gp54), and a short open reading frame containing a DNA-binding domain (gp55). Eight genes were predicted to contain membrane spanning domains, only two are denoted on this gene map: gp67 and gp3. The lysis cassette consists of Lysin A and Lysin B. Multiple structural genes were identified and are colored green on the gene map. This gene map was generated using Phast (46).

4.4 *M. chelonae* Bergey prophage contains two ORF's containing DNA-binding domains.

There are two ORF's within the prophage that have helix-turn-helix domains that are potential DNA-binding domains. These ORF's lay within the integrase cassette and are therefore strong candidates for regulators of the transition between lytic and lysogenic lifestyles.

Gp54 is transcribed in the reverse direction, has high coding potential according to GeneMark (3), and has strong sequence identity to a helix-turn-helix DNA binding domain protein in cluster K phage Marcoliusprime. Analysis of the folding structure of gp54 reveals that it folds similarly to transcriptional regulator EspR (Rv3849) in *M. tuberculosis* (Figure 5) which adopts a traditional helix-turn-helix conformation. Traditional helix-turn-helix domains are comprised of a core of three helices, with the core helix domains packed closely together when dimerized. Upon dimerization the 2nd and 3rd helices in the core domain will interact with the major groove of the target DNA (2). The helix-turn-helix domain is located at the N-terminus, with 15 out of 17 of the same amino acids involved in the helix-turn-helix as in the EspR known secondary structure (Figure 6). EspR positively controls the ESX-1 delivery system, which delivers important effector proteins to host cells during *M. tuberculosis* infection and is an important virulence factor (36).

The second ORF with a predicted helix-turn-helix domain, gp55, is oriented in the forward direction, has high coding potential according to GeneMark (3), and has significant similarities to lambda phage repressor-like DNA binding domains as well as a transcriptional regulator in *V. cholera* according to Phyre2 (24) (Figure 7). The HTH

domain is organized in a traditional manner (Figure 8). BLAST (31) analysis reveals similarities to multiple HTH DNA binding domains in *M. abscessus* and some mycobacteriophage. Gp55 also shows similarity to an antitoxin component in *Prevotella buccae* according to Phyre2 (24) and to multiple types of antitoxins according to HHpred (39). There is no prophage-encoded toxin component detected, but gp55 could be active in suppressing translation of a host-encoded toxin.

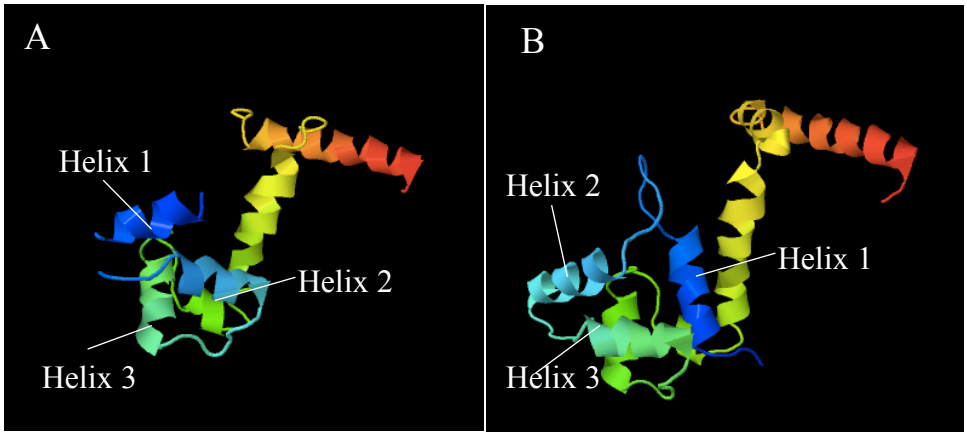


Figure 5. Predicted folding structure of *M. chelonae* prophage repressor gp54 (A) compared to known folding structure of *M. tuberculosis* EspR (B). *M. tuberculosis* EspR structure was solved using x-ray crystallography. Folding predictions were made using Phyre2 (24).

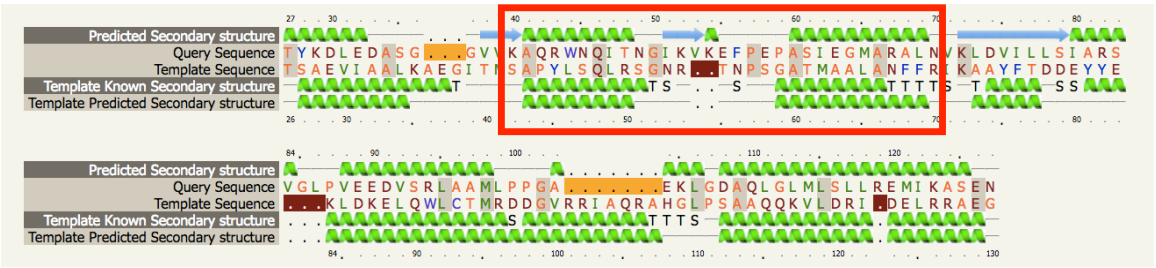


Figure 6. Predicted secondary structure of gp54 compared to known secondary structure of *M. tuberculosis* EspR. Amino acids 41 through 49 are involved in the second helix and amino acids 60 through 69 are involved in the third helix. This helix-turn-helix region is denoted by a red box. Folding predictions were made using Phyre2 (24).

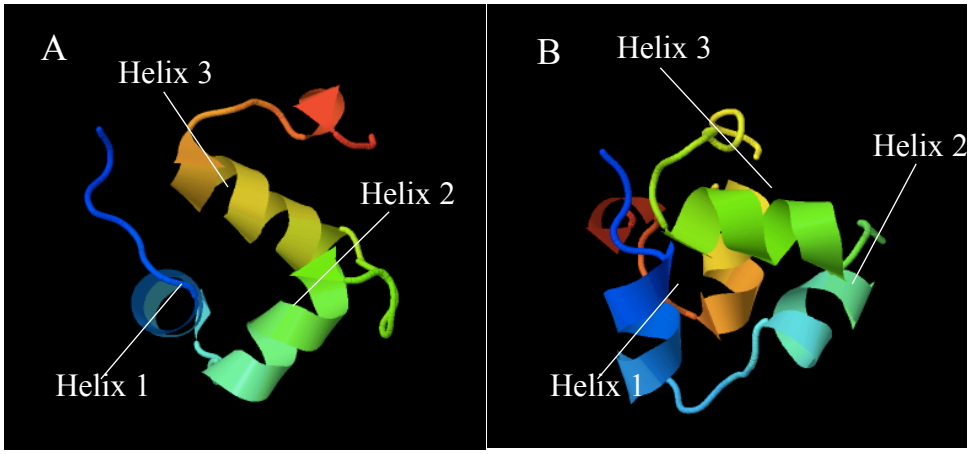


Figure 7. Predicted folding structure of *M. chelonae* prophage repressor gp55 (A) compared to known folding structure of Lambda phage repressor DNA binding domain (B). Lambda phage repressor structure was solved using x-ray crystallography. Folding predictions were made using Phyre2 (24).

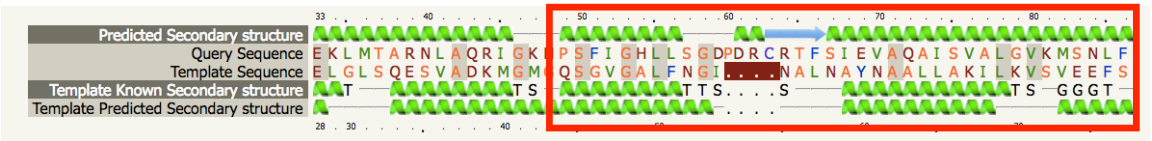


Figure 8. Predicted folding structure of gp55 compared to known secondary structure of Lambda repressor DNA binding domains. The predicted helix-turn-helix region is denoted by a red box with the second helix involving amino acids 49 through 56 and the third helix involving amino acids 67 through 86. Folding predictions were made using Phyre2 (24).

4.5 *M. chelonae* Bergey prophage contains multiple transmembrane proteins.

M. chelonae Bergey prophage contains 8 potential transmembrane proteins, predicted using the online program TMHMM (25). Of these 8 potential proteins, gp3 and gp67 are described below.

Gp3 contains one membrane spanning domain with high coding potential (Figure 9). It does not bear any resemblance to ORF's in any other phages sequenced thus far. At present, the function of this gene is unknown and is difficult to determine. There is a moderate level of similarity to cystatin-like protein containing a nuclear transport factor-2 (NTF2) domain predicted by Phyre2 (24) (Figure 10) as well as HHpred (39). NTF domains typically function as dimers and mediate the nuclear import of essential molecules (34).

Gp67 has two membrane spanning domains (Figure 11) with high coding potential and is highly similar to band-7-like membrane proteins in multiple phages from clusters M, R, and C. Band-7 proteins are characterized by a highly-conserved sequence motif known as the SPFH domain, named after membrane proteins stromatins, prohibatins, flotillins, and the HflK protein. These proteins are known to modulate membrane associated activities, including ion-channel function and vesicle trafficking. The SPFH domain is highly conserved across all of nature, but its roll and function in prokaryotes is poorly characterized (5). SPFH domains are generally characterized by an alpha helical subdomain folded over a beta sheet subdomain with the two subdomains connected by a hinge region. Although the common function of the SPFH domain has not been identified, it is known to often function as a scaffolding protein on lipid membranes

(27). Analysis of the predicted folding structure of gp67 (24) shows strong similarity to stromatin and SPFH domains (Figure 12).

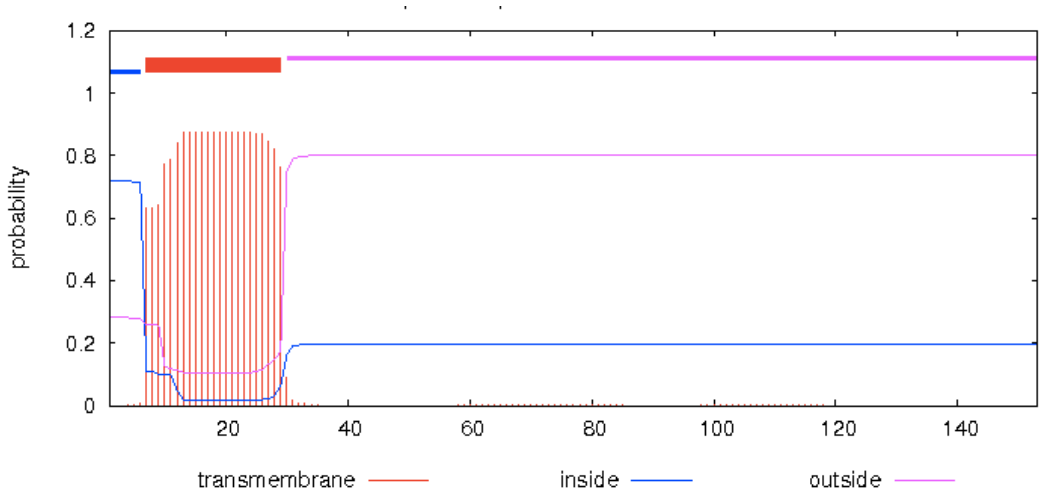


Figure 9. Predicted alpha helix membrane spanning domain in gp3. Probability is shown on the y-axis and amino acid numbers are shown on the x-axis. The membrane spanning helix is shown in red and involves amino acids 7 through 29. Regions predicted to lay inside the cell are denoted by a blue line and regions laying outside the cell are denoted by a pink line. Transmembrane domain analysis was done using TMHMM (25).

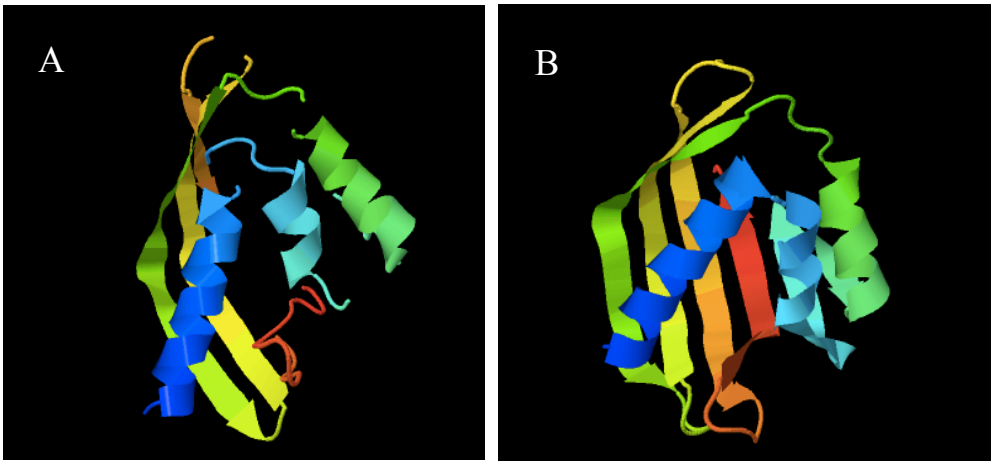


Figure 10. Predicted tertiary structure of gp3 (A) compared to known crystal structure of a putative dehydratase from the NTF2-like family from *Streptomyces avermitilis* (B). *S. avermitilis* structure was solved using x-ray crystallography. Folding predictions were made using Phyre2 (24).

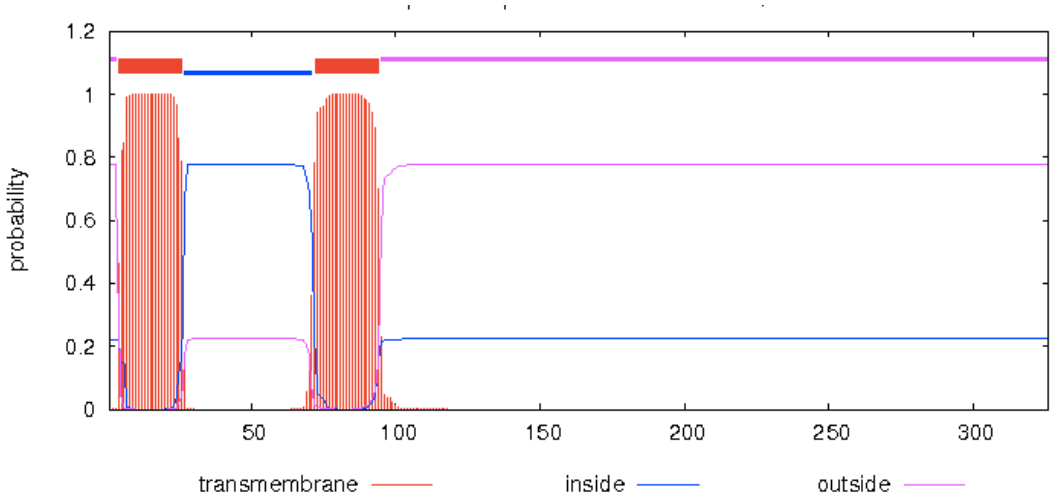


Figure 11. Predicted alpha helix membrane spanning domains in gp67. Probability is shown on the y-axis and amino acid numbers are shown on the x-axis. Membrane spanning helices are shown in red and involve amino acids 4 through 26 and amino acids 72 through 94. Regions predicted to lay inside the cell are denoted by a blue line and regions laying outside the cell are denoted by a pink line. Transmembrane domain analysis was done using the web program TMHMM (25).

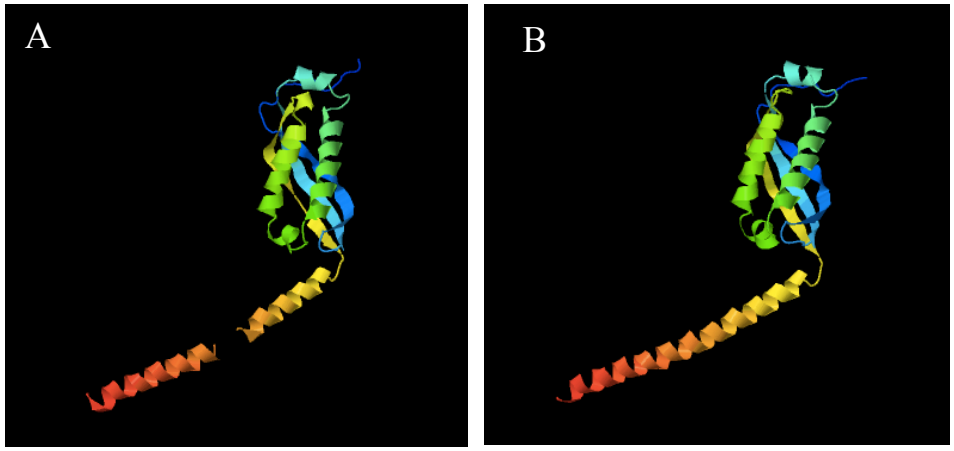


Figure 12. Predicted tertiary structure of gp67 (A) compared to known crystal structure of the core domain of stromatin from *Pyrococcus horikoshii* (B). *P. horikoshii* structure was solved using x-ray crystallography. The N-terminal

transmembrane region in both images is denoted by the red, orange, and yellow alpha helix. Folding predictions are made using Phyre2 (24).

4.6 *M. chelonae* Bowfin does not contain the prophage in the same position.

M. chelonae Bowfin was shown through PCR to be empty in the site where the prophage lies in *M. chelonae* Bergey. Primers were designed to bind outside the prophage region on the left and right. The PCR fragment produced when these primers were used in *M. chelonae* Bowfin was approximately 600 bp, and was the result of amplification across the empty attB site (Figure 13). The primers were confirmed to bind in *M.chelonae* Bergey and amplified the left and right ends of the prophage.

4.7 *M. chelonae* Bowfin attB site is identical to *M. chelonae* Bergey attB.

The *M. chelonae* Bowfin attB site does not contain the prophage sequence. The empty site was sequenced and shown to be 100% identical to *M. chelonae* Bergey attB. The region flanking the left side of the Bowfin attB site is more variable, compared to the *M. chelonae* Bergey attB site, than the rightward flanking region.

4.8 *M. chelonae* Bergey attB is highly conserved in singleton DS6A and in clusters K and F.

M. chelonae Bergey attB is 100% identical to the DS6A attP, and highly conserved in a number of K cluster phage, mostly K3 phage including TBond007, ShedlockHolmes, and Pixie. The attP site is also highly conserved in F cluster phage, particularly F1 cluster phage including Sparkdehlily, Zerg, and Whouxphf.

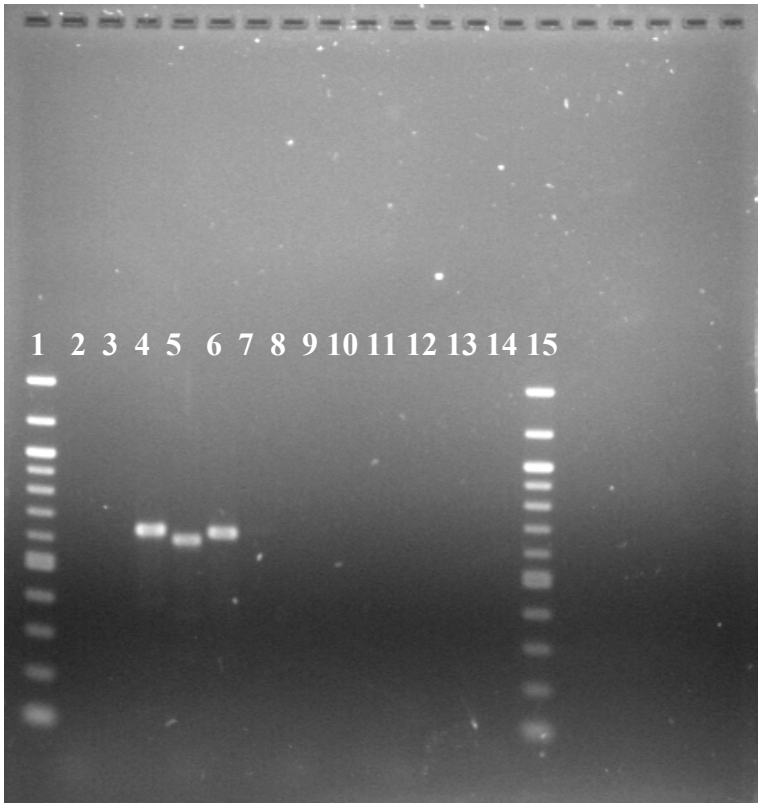


Figure 13. Agarose gel electrophoresis of PCR products produced from prophage PCR detection assay performed on *M. chelonae* strains. A combination of two sets of primers were used. Each set was designed to flank the left and right ends of the prophage with one primer laying within bacterial DNA and the other laying in prophage DNA, denoted LF (left forward) and LR (left reverse), and RF (right forward) and RR (right reverse). Lanes 2-4 contain *M. chelonae* Bowfin DNA with LF and LR used in lane 2, RF and RR in lane 3, and LF and RR in lane 4. Lanes 5-7 contain *M. chelonae* Bergey DNA with LF and LR in lane 5, RF and RR in lane 6, and LF and RR in lane 7. Lanes 8-10 contain *M. chelonae* isolated from a Fathead Minnow with LF and LR in lane 8, RF and RR in lane 9, and LF and RR in lane 10. Lanes 11-13 contain *M. chelonae* F5 with LF and LR in lane 11, RF and RR in lane 12, and LF and RR in lane 13. Lane 14 is a negative control. Lanes 1 and 15 contain a 100 bp ladder.

4.9 Prophage transmembrane protein and repressor are transcriptionally expressed in Bergey.

RT-PCR analysis of RNA isolated from *M. chelonae* Bergey detected expression of gp3, a predicted transmembrane protein, and gp54, a repressor (Figure 14). In addition to these, four more primer sets were used in this experiment. These included gp7, gp52, gp55, and gp67, which are predicted to be terminase, integrase, a repressor, and a transmembrane protein. PCR with these primers did not detect any transcriptional expression.

4.10 No infectious phage particles were detected using spot testing.

Because *M. chelonae* Bowfin has an empty attB sequence identical to *M. chelonae* Bergey, Bergey prophage particles may be capable of infecting Bowfin. Spot testing Bergey supernatant on Bowfin did not reveal any infectious prophage particles in Bergey supernatant.

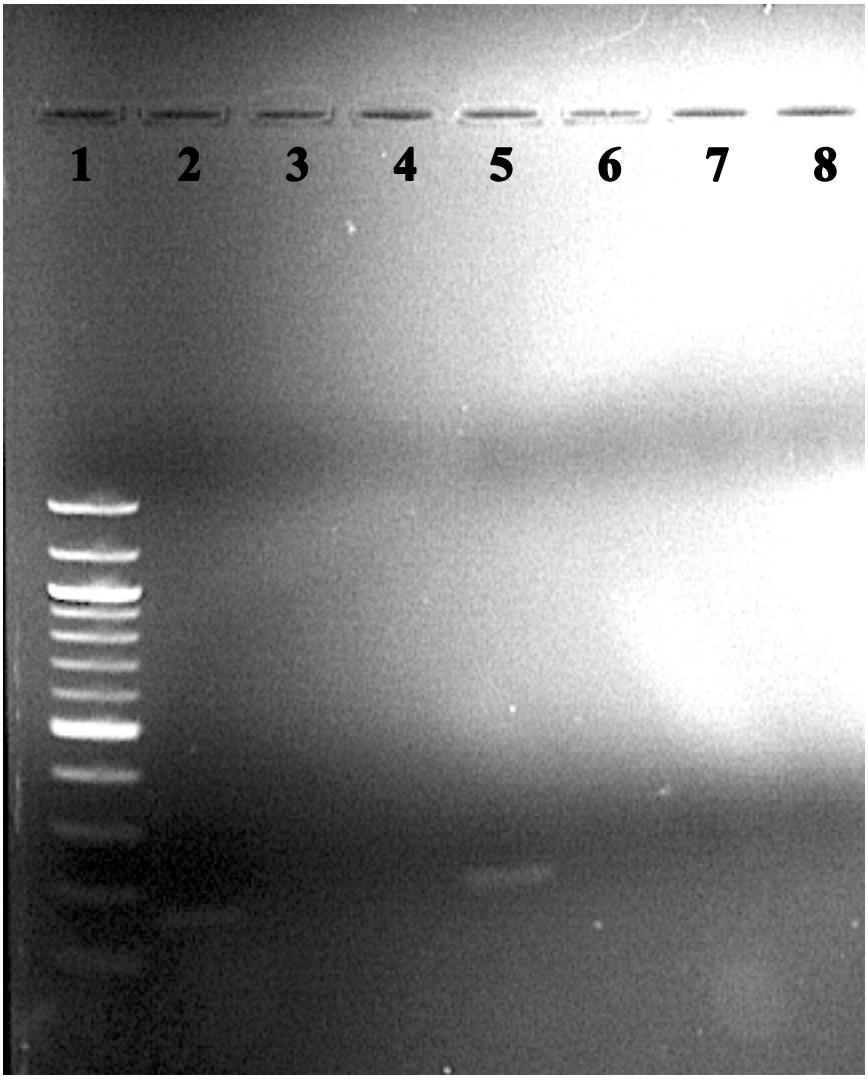


Figure 14. Agarose gel electrophoresis of RT-PCR products using *M. chelonae* Bergey RNA as template. Lane 2 contained gp3 primers and lane 5 contained gp54 primers. Lanes 3, 4, 6 and 7 contained primers for gp7, gp52, gp55 and gp67 respectively. Lane 8 is a negative control. A 100 bp ladder is shown in lane 1.

5.0 Discussion

5.1 *M. smegmatis* MC2 155 prophage region 1 has strong potential to have transcriptional regulatory action during and after infection.

Expression levels change drastically in prophage region 1 in the *M. smegmatis* genome depending on lysogen status and time after lytic infection. These results strongly suggest that the prophage encoded genes, including the MtrAB two-component regulatory system and the type VII secretion system component, are up and downregulated to modulate host functions during and after phage infections. These prophage-encoded genes may have some role in modulating the host cell response to changes in the environment or in modulating the host cell wall permeability (1). It should be noted that it is still unclear if the MtrAB system is prophage encoded, or if the phage simply integrated adjacent to this two component regulatory system. In either case, the prophage is likely to have an impact on the expression of the MtrAB system. In a *C. glutamicum* model, deletion mutants of MtrAB had noticeably different cell morphology and were more sensitive to penicillin, vancomycin, and lyosozyme. Introducing plasmid-encoded copies of mtrAB reversed these changes (29). The phage-encoded mtrAB system could be offering the benefit of resistance to these antimicrobial agents to *M. smegmatis*.

5.2 The *M. chelonae* Bergey prophage may provide superinfection immunity against DS6A, and clusters F and K.

Characterization of the *M. chelonae* prophage revealed that many open reading frames showed similarity to a common mycobacteriophage; DS6A. This phage is a singleton and is known to only infect *Mycobacterium* of the *M. tuberculosis* complex (19). A minority of genes were most related to DS6A, while the rest were related to many

different clusters and subclusters, most commonly clusters F and K. This is interesting in part because this multi-cluster origin is true of singletons in general, but also because it serves as a predictor of which clusters the prophage may offer some level of superinfection immunity to. In this case, the prophage may offer immunity against clusters F and K. While these are good predictions based on current information, the continual sequencing of more phage genomes will provide a better understanding of the prophage's origins and potential cluster designation.

5.3 The *M. chelonae* Bergey prophage integration is recent.

Characterization of the *M. chelonae* Bergey prophage revealed an intact mycobacteriophage with conserved genome structure, complete with notable cassettes such as the integrase and lysis cassettes. The prophage also contains highly conserved individual genes, including repressors, transmembrane proteins, structural proteins such as capsid and tail proteins, and important enzymes such as a HNH endonuclease and DNA methyltransferase. It was also shown that at least two prophage genes are transcriptionally active, indicating that those two open reading frames are conserved to such a high level that they are capable of being recognized by RNA polymerase. These analyses suggest that the prophage integration is recent, assuming there is not evolutionary pressure within the bacteria to conserve the entirety of the prophage.

5.4 The *M. chelonae* prophage has the potential to affect the biology of the host.

The high conservation of the prophage along with the discovery that at least two genes are transcriptionally active shows the prophage has retained the ability to be expressed on the transcriptional level. This suggests that the prophage might have a greater impact on the host than was previously thought. Expression of these two prophage

genes, transmembrane protein gp3 and predicted repressor protein gp54, offer an interesting area of study. If gp3 transcripts are translated and presented on the cell membrane, it could be offering superinfection immunity to the host. Alternatively, gp3 could be providing some unknown benefit to the host. If potential repressor gp54 is fully expressed, it could be repressing the prophage and preventing the phage from entering the lytic lifecycle and lysing the cell. It could also be offering superinfection immunity by binding foreign phage DNA and marking it for destruction.

It can be hypothesized that if gp3 and gp55 are transcriptionally active, it is likely that more of the 95 total genes in the prophage are also transcriptionally active because many of these genes are also highly conserved. For example, the putative antitoxin gp55 could function to suppress the translation of a host-encoded toxin component. The prophage therefore has the potential to be affecting the host fitness levels in some way, but it is too early in this research to better define the prophage's function and role in the bacterium.

5.5 Conclusions and Future work

Regarding the *M. smegmatis* prophage regions, more investigation is needed to understand why prophage region 1 genes MSMEG_1872 through MSMEG_1975 are significantly up and down regulated in response to viral infection. It could be that these genes are modulating the host cell's response to changes in the environment as well as changing the host cell wall permeability.

In the intact *M. chelonae* Bergey prophage, it was shown that gp3 and gp54 were transcriptionally active through RT-PCR. It is now necessary to determine the transcriptional status of the entire prophage region in *M. chelonae* Bergey through

RNAseq. This knowledge will provide deeper insight into the effect of the prophage on the biology of the host bacterium and will help determine the level of conservation of the prophage region.

REFERENCES

1. Abdallah, A. M., Van Pittius, N. C. G., Champion, P. A. D., Cox, J., Luirink, J., Vandenbroucke-Grauls, C. M., ... & Bitter, W. (2007). Type VII secretion—mycobacteria show the way. *Nature reviews microbiology*, *5*(11), 883-891.

2. Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M., & Iyer, L. M. (2005). The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS microbiology reviews*, *29*(2), 231-262.

3. Besemer, J., & Borodovsky, M. (2005). GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic acids research*, *33*(suppl 2), W451-W454.

4. Bibb, L.A. and Hatfull, G.F., 2002. Integration and excision of the Mycobacterium tuberculosis prophage-like element, φRv1. *Molecular microbiology*, *45*(6), pp.1515-1526.

5. Boehm, M., Nield, J., Zhang, P., Aro, E. M., Komenda, J., & Nixon, P. J. (2009). Structural and mutational analysis of band 7 proteins in the cyanobacterium Synechocystis sp. strain PCC 6803. *Journal of bacteriology*, *191*(20), 6425-6435.

6. Choi, Y. W., Kotzin, B. L. L. H., Herron, L., Callahan, J., Marrack, P., & Kappler, J. (1989). Interaction of Staphylococcus aureus toxin" superantigens" with human T cells. *Proceedings of the National Academy of Sciences*, *86*(22), 8941-8945.

7. Cumby, N., Edwards, A. M., Davidson, A. R., & Maxwell, K. L. (2012). The bacteriophage HK97 gp15 moron element encodes a novel superinfection exclusion protein. *Journal of bacteriology*, *194*(18), 5012-5019.

8. Dedrick, R. M., Jacobs-Sera, D., Bustamante, C. A. G., Garlena, R. A., Mavrich, T. N., Pope, W. H., ... & Bonilla, J. A. (2017). Prophage-mediated defence against viral attack and viral counter-defence. *Nature Microbiology*, *2*, 16251.

9. Donnelly-Wu, M. K., Jacobs, W. R., & Hatfull, G. F. (1993). Superinfection immunity of mycobacteriophage L5: applications for genetic transformation of mycobacteria. *Molecular microbiology*, *7*(3), 407-417.

10. Fineran, Peter C., et al. "The phage abortive infection system, ToxIN, functions as a protein–RNA toxin–antitoxin pair." *Proceedings of the National Academy of Sciences* 106.3 (2009): 894-899.

11. Fol, M., Chauhan, A., Nair, N. K., Maloney, E., Moomey, M., Jagannath, C., ... & Rajagopalan, M. (2006). Modulation of Mycobacterium tuberculosis proliferation

by MtrA, an essential two-component response regulator. *Molecular microbiology*, *60*(3), 643-657.

12. Froman, S., Will, D. W., & Bogen, E. (1954). Bacteriophage active against virulent Mycobacterium tuberculosis—I. Isolation and activity. *American Journal of Public Health and the Nations Health*, *44*(10), 1326-1333.

13. Gardner, Grace M., and Russell S. Weiser. "A Bacteriophage for Mycobacterium smegmatis." *Experimental Biology and Medicine* 66.1 (1947): 205-206.

14. Gauthier, David T., and Martha W. Rhodes. "Mycobacteriosis in fishes: a review." *The Veterinary Journal* 180.1 (2009): 33-47.

15. Golais, František, Jaroslav Hollý, and Jana Vítkovská. "Coevolution of bacteria and their viruses." *Folia microbiologica* 58.3 (2013): 177-186.

16. Hasan NA, Davidson RM, de Moura VC, Garcia BJ, Reynolds PR, Epperson LE, Farias-Hesson E, DeGroote MA, Jackson M, Strong M. Draft Genome Sequence of Mycobacterium chelonae Type Strain ATCC 35752.Genome Announc. 2015 May 28;3(3). pii: e00536-15. doi: 10.1128/genomeA.00536-15.

17. Hatfull, G. F., Jacobs-Sera, D., Lawrence, J. G., Pope, W. H., Russell, D. A., Ko, C. C., ... & Hoyte, N. N. (2010). Comparative genomic analysis of 60 mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. *Journal of molecular biology*, *397*(1), 119-143.

18. Hatfull, G. F. (2014). Molecular genetics of mycobacteriophages. *Microbiology spectrum*, *2*(2), 1.

19. Hatfull, G. F. (2010). Mycobacteriophages: genes and genomes. *Annual review of microbiology*, *64*, 331-356.

20. Houben, E. N., Korotkov, K. V., & Bitter, W. (2014). Take five—Type VII secretion systems of Mycobacteria. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1843*(8), 1707-1716.

21. Jaén-Luchoro D, Salvà-Serra F, Aliaga-Lozano F, Seguí C, Busquets A, Ramírez A, Ruíz M, Gomila M, Lalucat J, Bennasar-Figueras A. Complete Genome Sequence of Mycobacterium chelonae Type Strain CCUG 47445, a Rapidly Growing Species of Nontuberculous Mycobacteria. Genome Announc. 2016 Jun 9;4(3). pii: e00550-16. doi: 10.1128/genomeA.00550-16.

22. Johansen, B. K., Wasteson, Y., Granum, P. E., & Brynestad, S. (2001). Mosaic structure of Shiga-toxin-2-encoding phages isolated from Escherichia coli O157: H7 indicates frequent gene exchange between lambdoid phage genomes. *Microbiology*, *147*(7), 1929-1936.

23. Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, DeJong RJ, Dennehy JJ, Denver DR, Dunbar D, Elgin SCR, Findley AM, Gissendanner CR, Golebiewska UP, Guild N, Hartzog GA, Grillo WH, Hollowell GP, Hughes LE, Johnson A, King RA, Lewis LO, Li W, Rosenzweig F, Rubin MR, Saha MS, Sandoz J, Shaffer CD, Taylor B, Temple L, Vazquez E, Ware VC, Barker LP, Bradley KW, Jacobs-Sera D, Pope WH, Russell DA, Cresawn SG, Lopatto D, Bailey CP, Hatfull GF. 2014. A broadly implementable research course in phage discovery and genomics for first-year undergraduate students. mBio 5(1):e01051-13. doi:10.1128/mBio.01051-13.

24. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, *10*(6), 845-858.

25. Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology*, *305*(3), 567-580.

26. Kutter, Elizabeth, et al. "Phage therapy in clinical practice: treatment of human infections." *Current pharmaceutical biotechnology* 11.1 (2010): 69-86.

27. Kuwahara, Y., Unzai, S., Nagata, T., Hiroaki, Y., Yokoyama, H., Matsui, I., ... & Hiroaki, H. (2009). Unusual thermal disassembly of the SPFH domain oligomer from Pyrococcus horikoshii. *Biophysical journal*, *97*(7), 2034-2043.

28. Lew, J. M., Kapopoulou, A., Jones, L. M., & Cole, S. T. (2011). TubercuList–10 years after. *Tuberculosis*, *91*(1), 1-7.

29. Möker, Nina, et al. "In vitro analysis of the two-component system MtrB-MtrA from Corynebacterium glutamicum." *Journal of bacteriology* 189.9 (2007): 3645-3649.

30. Muniesa, M., Recktenwald, J., Bielaszewska, M., Karch, H., & Schmidt, H. (2000). Characterization of a Shiga toxin 2e-converting bacteriophage from an Escherichia coli strain of human origin. *Infection and immunity*, *68*(9), 4850-4855.

31. NCBI Resource Coordinators. (2016). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *44*(Database issue), D7–D19. http://doi.org/10.1093/nar/gkv1290

32. Perna, N. T., Plunkett, G., Burland, V., Mau, B., Glasner, J. D., Rose, D. J., ... & Pósfai, G. (2001). Genome sequence of enterohaemorrhagic Escherichia coli O157: H7. *Nature*, *409*(6819), 529-533.

33. Pierre-Audigier, C., Jouanguy, E., Lamhamedi, S., Altare, F., Rauzier, J., Vincent, V., ... & Gaillard, J. L. (1997). Fatal disseminated Mycobacterium smegmatis infection in a child with inherited interferon γ receptor deficiency. *Clinical infectious diseases*, *24*(5), 982-984.

34. Quimby, B. B., Lamitina, T., L'Hernault, S. W., & Corbett, A. H. (2000). The mechanism of ran import into the nucleus by nuclear transport factor 2. *Journal of Biological Chemistry*, *275*(37), 28575-28582.

35. Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature biotechnology*, *29*(1), 24-26.

36. Rosenberg, O. S., Dovey, C., Tempesta, M., Robbins, R. A., Finer-Moore, J. S., Stroud, R. M., & Cox, J. S. (2011). EspR, a key regulator of Mycobacterium tuberculosis virulence, adopts a unique dimeric structure among helix-turn-helix proteins. *Proceedings of the National Academy of Sciences*, *108*(33), 13450-13455.

37. Saluja, A., Peters, N. T., Lowe, L., & Johnson, T. M. (1997). A surgical wound infection due to Mycobacterium chelonae successfully treated with clarithromycin. *Dermatologic surgery*, *23*(7), 539-543.

38. Sato, T., Shimizu, T., Watarai, M., Kobayashi, M., Kano, S., Hamabata, T., ... & Yamasaki, S. (2003). Genome analysis of a novel Shiga toxin 1 (Stx1)-converting phage which is closely related to Stx2-converting phages but not to other Stx1-converting phages. *Journal of bacteriology*, *185*(13), 3966-3971.

39. Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic acids research*, *33*(suppl 2), W244-W248.

40. Sun, X., Göhler, A., Heller, K. J., & Neve, H. (2006). The ltp gene of temperate Streptococcus thermophilus phage TP-J34 confers superinfection exclusion to Streptococcus thermophilus and Lactococcus lactis. *Virology*, *350*(1), 146-157.

41. Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., & Leunissen, J. A. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic acids research*, *35*(suppl 2), W71-W74.

42. Waldor, M. K., & Mekalanos, J. J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science*, *272*(5270), 1910.

43. Wallace, R. J., Brown, B. A., & Onyi, G. O. (1992). Skin, soft tissue, and bone infections due to Mycobacterium chelonae chelonae: importance of prior corticosteroid therapy, frequency of disseminated infections, and resistance to oral

antimicrobials other than clarithromycin. *Journal of Infectious Diseases*, *166*(2), 405-412.

44. Wallace, R. J., Dukart, G., Brown-Elliott, B. A., Griffith, D. E., Scerpella, E. G., & Marshall, B. (2014). Clinical experience in 52 patients with tigecycline-containing regimens for salvage treatment of Mycobacterium abscessus and Mycobacterium chelonae infections. *Journal of Antimicrobial Chemotherapy*, *69*(7), 1945-1953.

45. Yamaguchi, T., Hayashi, T., Takami, H., Nakasone, K., Ohnishi, M., Nakayama, K., ... & Sugai, M. (2000). Phage conversion of exfoliative toxin A production in Staphylococcus aureus. *Molecular microbiology*, *38*(4), 694-705.

46. Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., & Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic acids research*, gkr485.

APPENDIX

Table 2. *M. chelonae* prophage coordinates, direction, and predicted functions.

ORF	Start	Stop	Function	Direction	Length
1	1	228	NKF	FWD	228
2	230	538	HNH endonuclease	FWD	309
3	691	1152	Transmembrane domain	RVS	462
4	1149	1343	NKF	RVS	195
5	1356	2348	NKF	RVS	993
6	2554	3033	Terminase, small subunit	FWD	195
7	3034	4776	Terminase, large subunit	FWD	1743
8	4773	6263	Portal protein	FWD	1491
9	6247	9804	Capsid maturation protease	FWD	3558
10	9824	10192	NKF	FWD	369
11	10249	10431	NKF	RVS	183
12	10560	11180	NKF	FWD	621
13	11213	12106	Major capsid protein	FWD	894
14	12123	12527	NKF	FWD	405
15	12524	12988	NKF	FWD	465
16	12988	13167	NKF	FWD	180
17	13157	13429	NKF	FWD	273
18	13422	13844	Assembly protein	FWD	423
19	13911	14492	Major tail subunit	FWD	582
20	14598	15056	NKF	FWD	459
21	15089	15475	NKF	FWD	387
22	15468	20024	Tapemeasure	FWD	4557
23	20021	20974	Minor tail protein	FWD	954
24	20971	22698	Minor tail protein	FWD	1728
25	22699	23130	Minor tail protein	FWD	432
26	23127	25262	Minor tail protein	FWD	2136
27	25262	27010	Minor tail protein	FWD	1749

28	27079	27498	NKF	FWD	420
29	27495	27761	NKF	FWD	267
30	27886	29376	Lysin A	FWD	1491
31	29373	30449	Lysin B	FWD	1077
32	30471	30683	NKF	FWD	213
33	30685	31011	NKF	FWD	327
34	31001	31321	NKF	FWD	321
35	31325	31645	NKF	FWD	321
36	31642	32337	NKF	FWD	696
37	32422	33600	DNA polymerase III	FWD	1179
38	33612	33860	NKF	RVS	249
39	34049	34483	NKF	RVS	1011
40	34483	35493	NKF	RVS	1011
41	35486	35686	NKF	RVS	201
42	35804	36601	Transcription factor	FWD	798
43	36659	37027	DNA binding domain, transcription factor	FWD	369
44	37009	37311	NKF	FWD	303
45	37397	38902	NKF	RVS	1506
46	39093	39284	NKF	RVS	192
47	39485	39682	NKF	FWD	198
48	39814	40167	NKF	RVS	354
49	40233	40880	NKF	RVS	648
50	40885	43086	HNH Endonuclease	RVS	2202
51	43094	43411	NKF	RVS	318
52	43910	45382	Integrase	FWD	1473
53	45461	45805	NKF	FWD	345
54	45815	46309	Repressor	RVS	495
55	46574	46885	Helix-turn-helix DNA binding domain	FWD	312
56	46885	47121	NKF	FWD	237
57	47121	48479	DNA methyl transferase	FWD	1359

58	48476	49243	NKF	FWD	768
59	49286	49393	NKF	FWD	108
60	49393	49716	NKF	FWD	324
61	49709	50257	NKF	FWD	549
62	50257	50784	DNA polymerase III subunit	FWD	417
63	50781	51197	NKF	FWD	417
64	51194	51913	Helix-turn-helix DNA binding domain	FWD	720
65	51910	52188	NKF	FWD	279
66	52185	53114	NKF	FWD	930
67	53149	54129	Band 7-like transmembrane protein	FWD	981
68	54126	54473	MazG nucelotide pyrophosphohydrolase	FWD	348
69	54470	54961	NKF	FWD	492
70	54958	55146	NKF	FWD	189
71	55213	55677	NKF	FWD	465
72	55852	56061	NKF	FWD	210
73	56058	56285	NKF	FWD	228
74	56285	56998	Primase	FWD	714
75	56991	57302	NKF	FWD	312
76	57362	57490	NKF	FWD	129
77	57487	57735	NKF	FWD	249
78	57738	58001	NKF	FWD	264
79	57998	58243	NKF	FWD	246
80	58448	58999	NKF	FWD	552
81	59187	59654	NKF	FWD	468
82	59657	60088	NKF	FWD	432
83	60088	60207	NKF	FWD	120
84	60261	60695	NKF	FWD	435
85	60708	62345	NKF	FWD	1638
86	62342	62683	NKF	FWD	342
87	62668	63135	Endodeoxyribonuclease	FWD	468

88	63129	63275	NKF	FWD	147
89	63265	63519	NKF	FWD	255
90	63519	63884	NKF	FWD	366
91	63877	64788	NKF	FWD	912
92	64785	65159	NKF	FWD	375
93	65358	66782	Glycosyltransferase	FWD	1425
94	66779	67456	NKF	FWD	678
95	67453	68166	NKF	FWD	714

BIOGRAPHY OF THE AUTHOR

Erica Joy Sewell was born in Portland, Maine on May 4, 1995. She grew up in Southern Maine and graduated from Marshwood High School in 2013. She decided to attend The University of Maine to study Microbiology, but has continued to nurture her love of music and guitar playing outside of school. Upon graduation she is planning on moving back to Southern Maine to spend more time with her family. She plans on backpacking across Europe in the near future, and would like to eventually pursue a graduate degree in plant pathology.