# PRELIMINARY INVESTIGATION OF CLUSTER M LYSOGENY AND PROPHAGE-MEDIATED DEFENSE MECHANISMS

A thesis presented to the faculty of the Graduate School of
Western Carolina University in partial fulfillment of the
requirements for the degree of Montana Jo Henson.
Master of Science in Biology

By

Montana Jo Henson

Director: Dr. Maria Gainey
Assistant Professor
Chemistry Department

Committee Members: Dr. Amanda Storm, Biology
Dr. Jeremy Hyman, Biology

Thesis Reader: Dr. Darby Harris, Biology

March 2023

# ACKNOWLEDGEMENTS

I would like to thank my director, Dr. Maria Gainey, and committee members, Dr. Jeremy Hyman and Dr. Amanda Storm, for their assistance and encouragement. Thank you, Dr. Gainey, for being a continual support and foundation through this process. I would like to thank Dr. Darby Harris for being my thesis reader and helping navigate the writing process, and his continual support overall during my time at Western Carolina University.

I would also like to extend sincere gratitude to the following people: undergraduates in the Gainey laboratory and previous students involved in this study including Erin Cafferty and Vance Reynolds. Lastly, I offer my warmest regards and appreciation to my parents, Jo Henson and Gary Henson, for their unconditional love and unwavering support through my academic journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF TERMS/ABBREVIATIONS

Gp1 – gene product 1
Gp2 – gene product 2
aa – amino acid
TA – Top Agar
Kan – Kanamycin
aTc – anhydrotetracycline
phages – bacteriophages
DNA – deoxyribonucleic acid
RNA – ribonucleic acid
dsDNA – double-stranded deoxyribonucleic acid
IHF – integration host factor
Int-Y – tyrosine integrase
Int-S – serine integrase
*attP* – phage attachment site
*attB* – bacterial attachment site
tRNA – transfer ribonucleic acid
RDF – recombination directionality factor
DEM(s) – defense escape mutant(s)
TSS – truncated signal sequence

# ABSTRACT

PRELIMINARY INVESTIGATION OF CLUSTER M LYSOGENY AND PROPHAGE-DEFENSE MECHANISMS

Montana Jo Henson, M.S.

Western Carolina University (May 2023)

Director: Dr. Maria Gainey

Bacteriophages are viruses that infect bacteria and are characteristically specific for their bacterial hosts with preferences rarely traversing genus boundaries. Mycobacteriophages – bacteriophages that infect Mycobacterial hosts – are the largest collection of genetically characterized phages known to infect a single host and are used to study phage biology and evolution. Mycobacteriophages have been divided into 32+ genetic clusters based on nucleotide sequence similarity and shared gene content. When a bacteriophage infects a bacterial cell, it usually enters the lytic cycle. Lytic replication results in the production of virus particles, and lysis of the host cell. Temperate bacteriophages may also enter the lysogenic cycle. During lysogeny, the bacteriophage integrates its genome into the host cell's chromosome. After integration, the viral genome is referred to as a prophage, and the bacterial cell is called a lysogen. To defend the lysogen cell against bacteriophages in the environment, some prophages have evolved to contain prophage defense genes. These genes can allow a prophage to prevent homotypic and/or the heterotypic bacteriophages from establishing a productive infection in the lysogen cell.

Mycobacteriophage treatment of antibiotic resistant human *Mycobacterium abscessus* and *Mycobacterium tuberculosis* infections has recently shown great promise. However, patient *Mycobacterium abscessus* isolates are only susceptible to a narrow range of mycobacteriophages due to the presence of prophages carrying uncharacterized defense systems. Cluster M-like prophages have been identified in patient isolates. Though cluster M-like prophages may play a significant role in resistance to bacteriophage therapy, little is known about how they establish lysogeny and what prophage defense mechanisms they may contain.

During this study, the integration site and defense profile of the IPhane7 (subcluster M1) prophage was investigated. The location of the bacterial integration site (*attB*) was determined by full genome Illumina sequencing of the IPhane7 lysogen. Prophage-host junction specific PCR, followed by Sanger sequencing was used to confirm the *attB/attP* sites. The defense profile of cluster M lysogens was determined by the creation of cluster M lysogens and subsequent challenge with a genetically diverse panel of Mycobacteriophages.

While the overall defense profile of cluster M prophages is unknown, previous work from the Gainey laboratory revealed that expression of only gene products (gp) 1 and 2 can prevent infection of cluster M bacteriophages. Gene 1 is unique to cluster M bacteriophages and is predicted to code for a 97 amino acid protein with a Sec-signal sequence. Gene 2 is found in other bacteriophages and sequenced Actinobacterial genomes. Select cluster M gp1 or gp1/gp2 were introduced into *M. smegmatis* cells using high and low plasmid expression vectors. The defense profile against cluster M and other heterotypic mycobacteriophages was then determined via viral challenges, and the results displayed high specificity amongst phage phenotypes. Additionally, we investigated how the synergistic actions of gp1/gp2 cluster M defense may be overcome by using comparative genomics to analyze gp1/2 defense escape mutants.

# INTRODUCTION

Bacteriophages, or phages, are viruses that infect bacteria. Bacteriophages are characteristically specific for their bacterial hosts with preferences rarely traversing genus boundaries. A bacteriophage host range is the taxonomic diversity of hosts that it may infect successfully [38]. Mycobacteriophages – bacteriophages that infect Mycobacterial hosts –are the largest collection of genetically characterized phages known to infect a single host and are used to study phage biology and evolution [74]. Mycobacteriophages have been divided into genetic clusters A-Z and AA-AE including 10+ singletons that do not share a close relative based on nucleotide sequence similarity and shared gene content [78,83]. These clusters may be further divided into subclusters (e.g. A1, A2, A3, etc.) [16]. As displayed in Figure 1, as of 2012 there were 220 actinobacteriophages identified and separated into clusters. That number has seen exponential growth over the past decade to include 22,513 total phages identified with 4,374 finished genomes, as of January 2023. Currently, of the 12,404 phages that have been identified to infect the host genus *Mycobacterium*, 2213 have been sequenced, as reported by The Actinobacteriophage Database at PhagesDB.org. Figure 1 displays a small proportion of the intricate and seemingly insurmountable web amongst actinobacteriophages. Bacteriophages may either contain DNA or RNA as the genetic material with both single and double-stranded forms identified [89]. Mycobacteriophages are all dsDNA. It is noted that there is little observed variation in GC% content between phages of the same cluster, but the GC% content varies considerably between clusters [37].

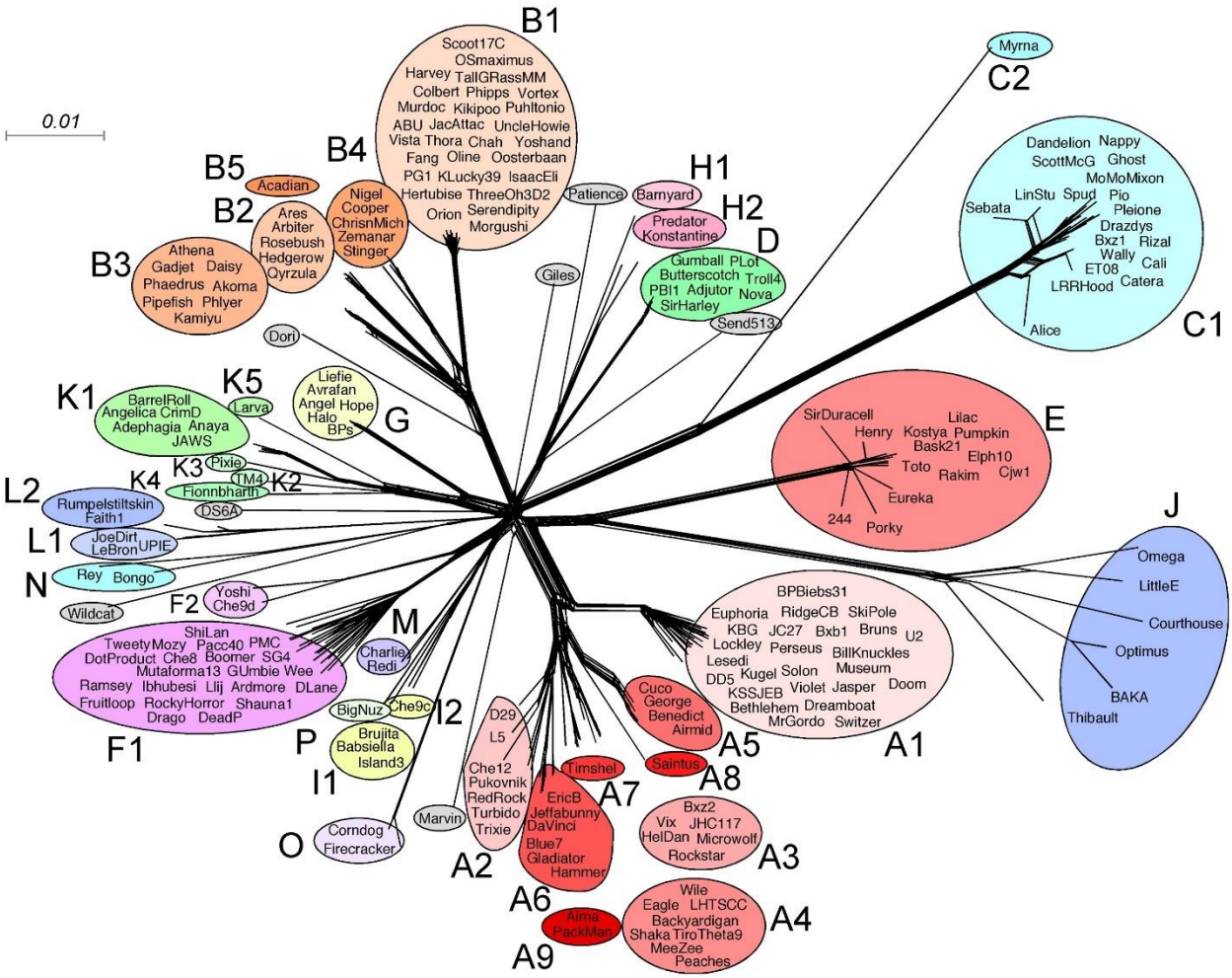



Figure 1. Genomic relationships amongst mycobacteriophages. The relationships among 220 mycobacteriophages are displayed using the NeighborNet function in Splitstree4 [34]. Data were generated from the database Mycobacteriophages_220 in the program Phamerator [13]. Phages within a cluster or subcluster are circled and labeled accordingly. Singletons are shown in gray circles. This figure was generated in 2012 and not fully inclusive of phages discovered post-2012. *Note figure and figure description adapted from* [37].

Phage genomic architecture displays common themes. Although there are three virion morphotypes of dsDNA tailed phages (*Myoviridae*, *Podoviridae*, and *Siphoviridae*)[29], all known mycobacteriophages isolated to date are tailed dsDNA phages with two morphotypes observed: the myoviridae that contain contractile tails and the siphoviridae with long flexible non-

contractile tails [1,33] as displayed in Figure 2.

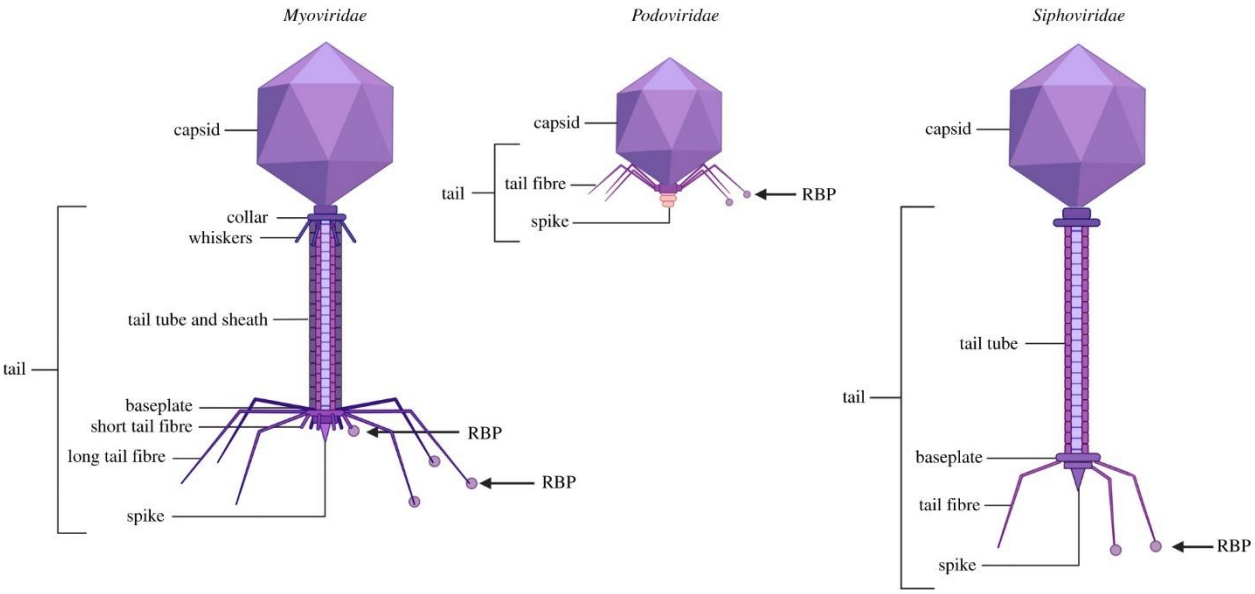



Figure 2. Tailed bacteriophages. The order *Caudoviraes* consists of three families: (a) *Myoviridae* with contractile tails, (b) *Podoviridae* that lack a baseplate and are short-tailed, and (c) *Siphoviridae* that possess long, non-contractile tail. *Figure generated by Ongenae V. et al, 2021* [57].

Siphoviral morphotypes, like cluster M phages, display clear synteny among genes that encode for virion structure and assembly functions [12], as seen in Figure 3 displaying an electron microscope image of IPhane7. Genes responsible for head and tail assembly are arranged with head genes 5' to the tail genes and tend to be located in the left arm of the genome [39]. Made up of >20 genes together, the head genes typically include one to two terminate subunits, a portal protein, a prohead protease, a scaffold protein, and the major capsid subunit while the tail genes include the major tail subunit, two overlapping open reading frames expressed via a programmed translational frameshift and the tapemeasure protein, followed by the minor tail proteins [28]. The tapemeasure protein gene length has been found to directly correspond with the length of the phage tail, and is typically the largest gene of the genome [41,59]. Virion morphotypes of

mycobacteriophages generally belong to the siphoviridae or myoviridae morphotypes, with no knowledge of a single mycobacteriophage of the podoviral morphotype [33]

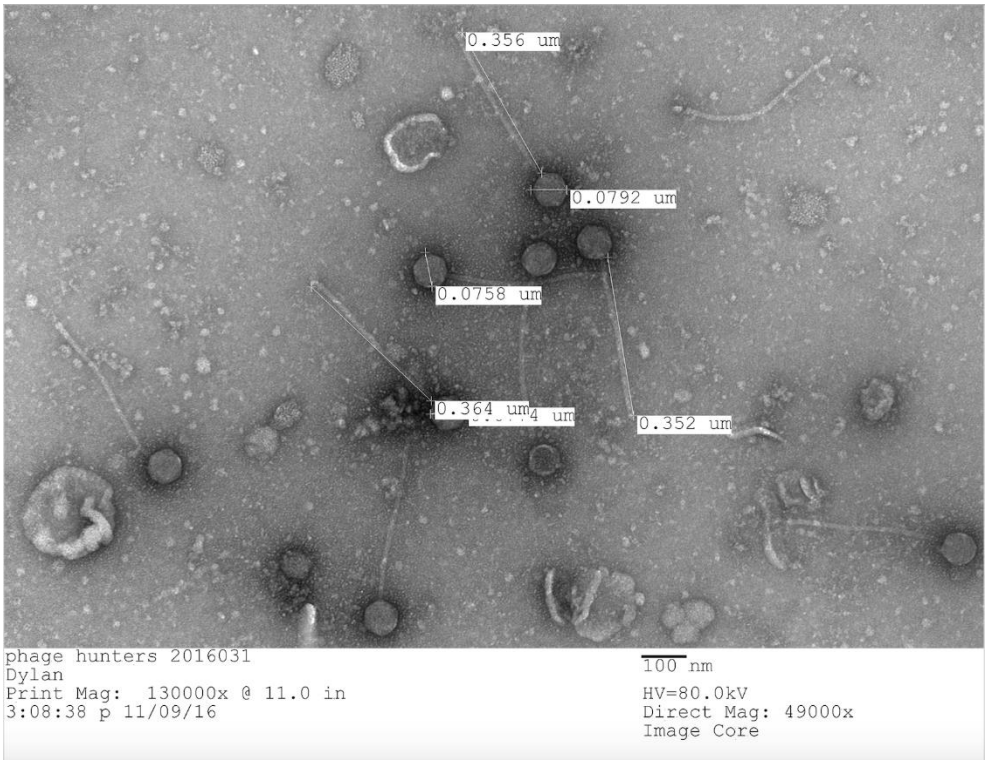



Figure 3. Electron Microscope Image of IPhane7. IPhane7 is observed to have a Siphoviridae morphotype with an icosahedral head. *Image from PhagesDB.org* [36]

Phage tails are responsible for host cell recognition and delivery of phage DNA into the cell's cytoplasm via cell wall perforation [3(p5),4]. Host cell recognition is achieved via interactions with specific cell surface receptors, like sugars and proteins, that the phage tail recognizes [57]. Receptor-binding proteins (RBPs), typically located at the tip of the phage's tail aid in recognizing surface-associated molecules on the host bacterium including enzymes, transporter proteins, substrate receptors, and other outer membrane structures [57].

When a bacteriophage infects a host bacterial cell, it may enter one of two pathways: the lytic cycle or the lysogenic cycle. Figure 4 displays the two pathways by which bacteriophages

infect a host bacterium. Typically, virulent bacteriophages enter lytic replication that results in viral particle production within the host cell and subsequent lysis of the bacteria and release of phage virions. Some phages may also enter the lysogenic cycle prior to lytic replication. Temperate bacteriophages (phages that may enter the lytic or lysogenic cycles) may switch between the dormant, lysogenic state and the active/productive, lytic state [15]. During lysogeny, the bacteriophage integrates its viral DNA into the host cell's chromosome. After integration, the viral genome is referred to as a prophage, and the bacterial cell is called a lysogen. The phage represses its viral genome by employing repressor genes to down-regulate desired lytic gene expression [8], and expresses prophage encoded genes that may increase bacterial fitness.

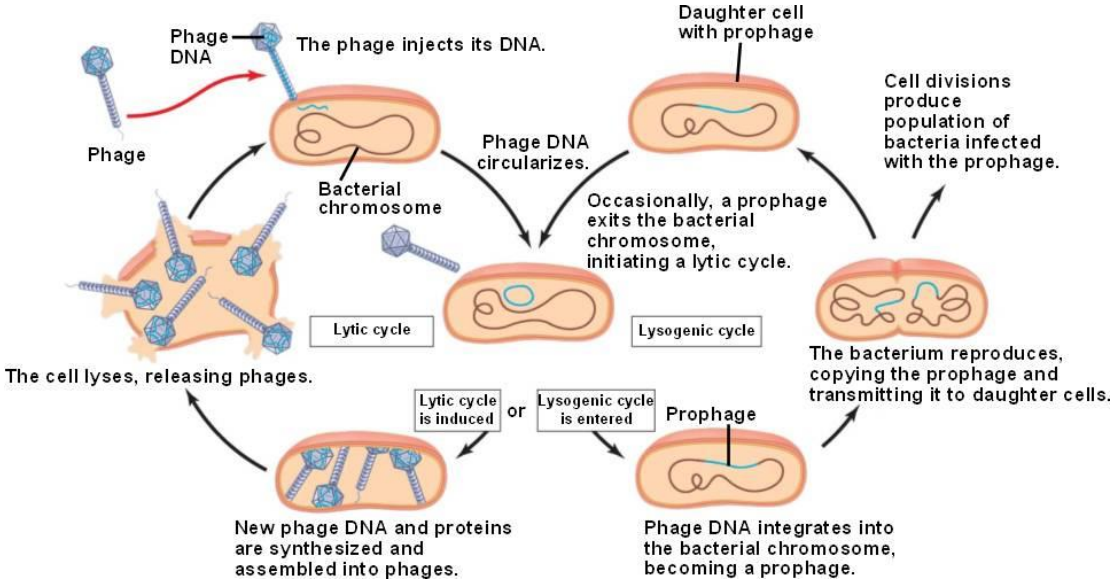



Figure 4. A comparison of the lytic and lysogenic cycles. *Figure generated by* [47]

Temperate phages encode an integrase gene (a recombinase) and a host-encoded integration host factor (IHF) that allows for the integration reaction whereby the integrase catalyzes strand cleavage and exchange [43]. Integration systems of most temperate mycobacteriophages tend to be site-specific for prophage integration into the host chromosome

[30]. Mediating this establishment of lysogeny, there have been two main types of integration systems observed, the tyrosine-integrases (Int-Y) and serine-integrases (Int-S) [30]. Integrases catalyze site-specific recombination between a bacterial attachment site (*attB*) and the phage attachment site (*attP*) that typically share a brief region of similarity in sequence between 3-45bp, referred to as the 'common core,' where strand exchange occurs [30]. Although serine and tyrosine integrases conduct essentially the same function, they do so by different modes of action. Typically, tyrosine-integrase systems' common cores are 25-45bp in length and may more easily be predicted by homology searching through the host chromosome. The *attP* sites in Int-Y systems are typically longer, thus allowing for multiple binding sites for integrase subunits and other accessory proteins required to stimulate recombination [81].

In contrast, Int-S systems have shorter common cores from 3-10bp long and do not require host-encoded proteins for recombination; the shortness of the common core causes the *attB* site not to be predictable bioinformatically [42,81]. Int-Y systems exclusively use *attB* sites that overlap tRNA genes, which is not the case for Int-S phages who do not integrate into tRNAs [31]. Phage-encoded serine integrases aide in directionality regulation of site-specific recombination between short common core sites without host factor requirements and without the recombination directionality factor (RDF) direct binding to the DNA [77]. DNA strand exchange mechanisms differ between the two systems. Int-Y systems make single-strand breaks in the DNA and exchange a single strand of each site to form an intermediate Holliday-junction-like DNA structure that is resolved to recombinants via cleavage and exchange of other strand pairs [81]. Int-S systems make double-stranded breaks in the DNA and exchange strands via rotational mechanisms, as displayed in Figure 5 [79].

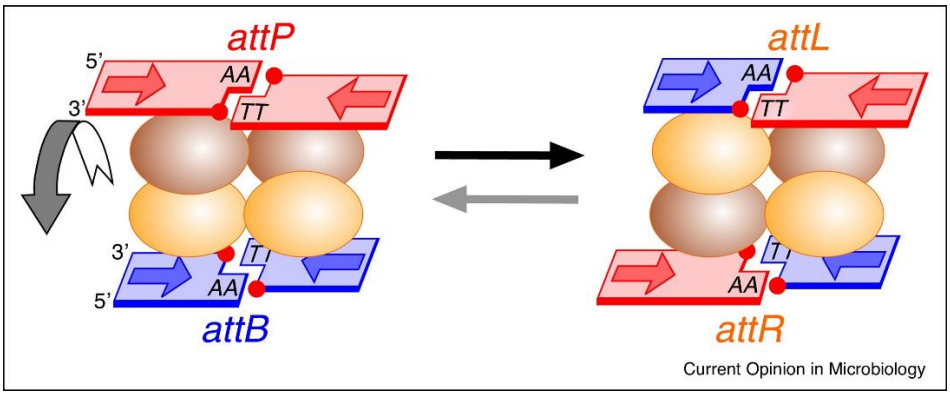



Figure 5. Mechanism of strand exchange by serine integrases. The *att*- sites are shown with double-strand breaks, after strand cleavage (left) and before ligation (right). The ends are exchanged by a rotational movement of DNA with attached integrase subunits (orange ovals). The recessed 5′ ends, attached to the integrase via a phosphodiester to the active site serine residue, are shown as red dots. The blue and red arrowheads indicate similar sequence motifs giving the sites imperfect two-fold symmetry. *Note. This model and description were produced by Stark in 2017, summarizing the rotational mechanisms of strand exchange used by serine integrases. From "Making serine integrases work for us," by* [81].

Based on the difficulty to locate Int-S systems due to short common cores sequences that make it unable to be identified bioinformatically or via tRNA gene locations, serine integrase systems must be determined experimentally, as appears to be the case with IPhane7's integration sites. The experimental determination of the *attP* and *attB* sites may lead to the identification of other cluster M phage integration sites and have influence on prophage gene expression. Of relative importance, cluster M-Like prophages have been observed in human *M. abscessus* infection isolates that are resistant to antibiotic treatment, identified as MabI prophages [16]. *M. abscessus* prophages have a vast range of diversity that may be assorted into 17 clusters and most of the prophages are generally not closely related to the multitude of gnomically variant *M. smegmatis* phages [16]. Interestingly, cluster MabI and MabJ that have been identified in the *M. abscessus* isolates share similarity to cluster M and A mycobacteriophages, respectively [16,48,62].

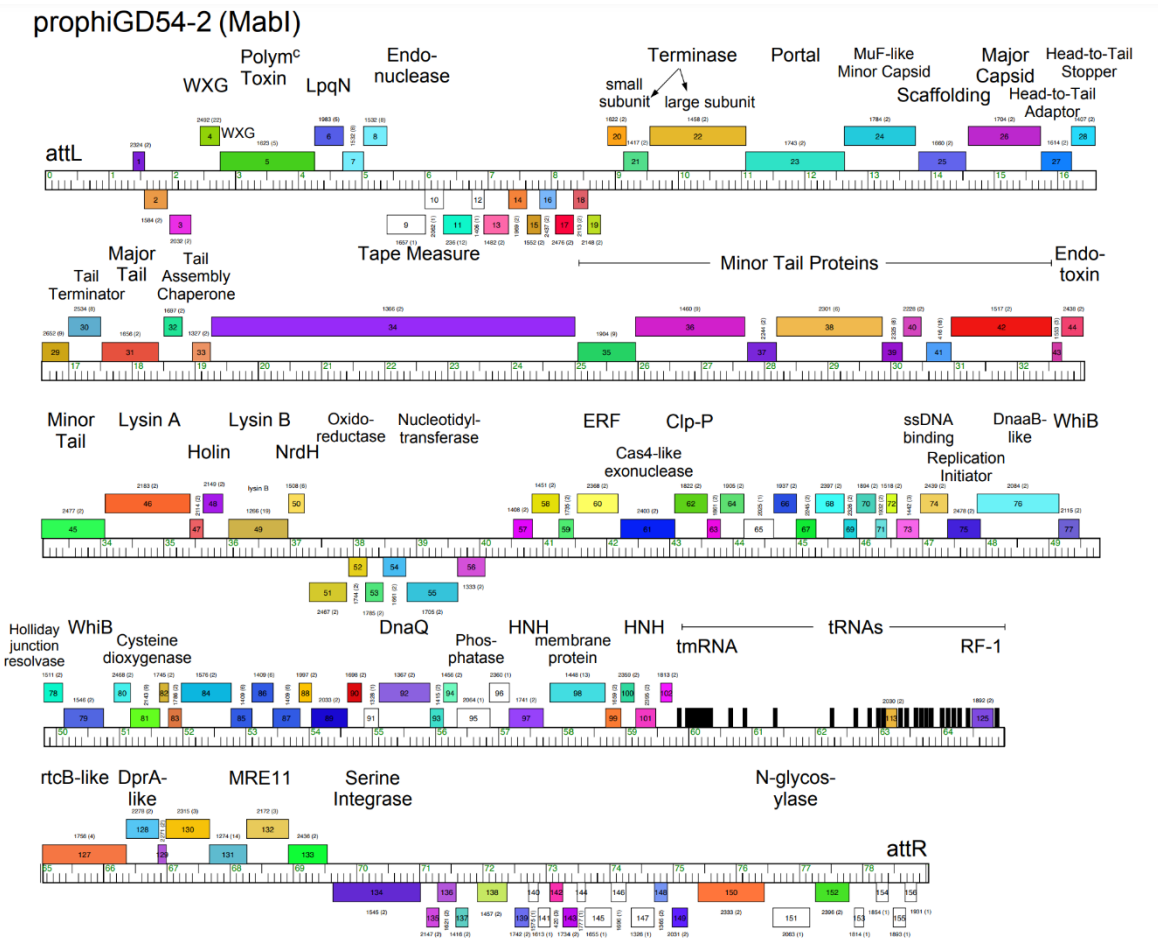



Figure S22. Genome organization of prophiGD54-2, Cluster MabI. See Figure S2 for details.

Figure 6. ProphiGD54-2 (cluster MabI), which is organized similarly to cluster M mycobacteriophages. Most phages are observed to use a tyrosine integrase (Int-Y) within the clusters, with the exceptions of clusters MabI and MabJ (cluster A similarity), which both use serine integrases (Int-S) [16]. *Note. Figure map of prophiGD54-2 from* [16].

A genomic map of MabI prophiGD54-2 in Figure 6 displays some unique features of the prophage and notable similarity in organization to cluster M mycobacteriophages including a serine-integrase in the right arm, an array of 21 tRNA genes, and the coding of a release factor [16]. Of the *attB* sites that were identified in the *M. abscessus* isolates, three utilized Int-S systems (attB-9, attB-7, and attB-17) that were noted to have characteristically short common core sequences of 5 to 8 bp [16,42]. Both cluster M-like MabI phages use the attB-9 and attB-17 and are located within genes MAB_3230, a SnoaL_4 domain related to an oxidoreductase

of *Streptomyces* [91], and MAB_3265, a dienelactone hydrolase family protein of unknown function, respectively [16]. Figure 7 displays these integration sites. Integration within open reading frames is not an uncommon phenomenon amongst temperate phages, such as observed by Bxb1 integration into the *groEL1* gene of *M. smegmatis* [42,56].

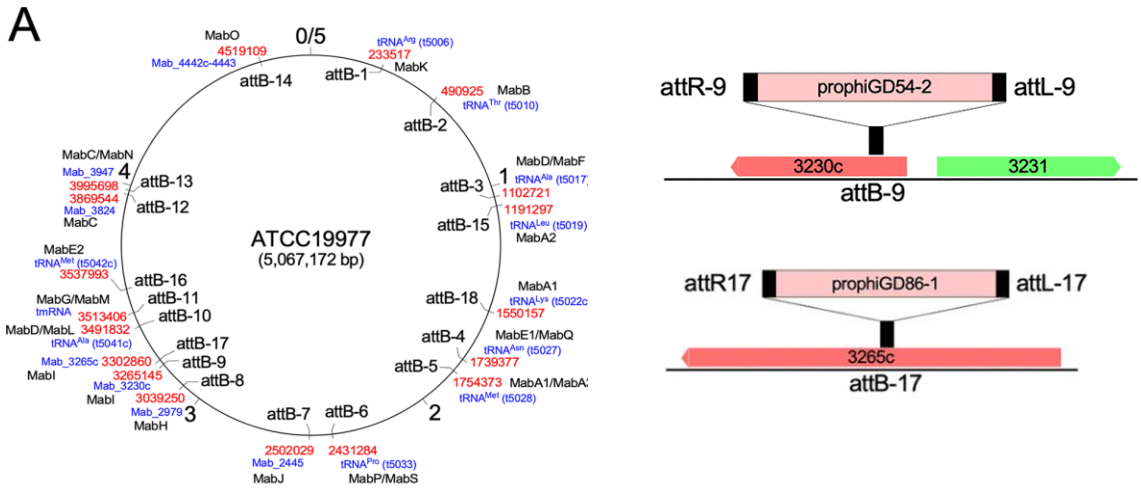



Figure 7. *M. abscessus* prophage integration and immunity. (A) Location of attB sites in the *M. abscessus* genome. The 5-Mbp *M. abscessus* ATCC 19977 circular genome is represented, with the location of the 18 *attB* sites (attB-1 to attB-18) indicated inside the circle. Outside the circle the coordinate of the site in ATCC 19977 is shown in red, the associated *M. abscessus* ATCC 19977 gene name is shown in blue, and the prophage clusters using each *attB* are shown in black. *attB* locations and consequences of integration are shown (black bars) relative to the *M. abscessus* ATCC 19977 genes for reference; rightward and leftward transcribed genes are shown as green and red boxes, respectively, with their ATCC 19977 gene number. An integrated prophage example is shown for the prophiGD54-2 *attB* site and prophiGD86-1, with the corresponding *attL* and *attR* sites shown to reflect the orientation of integration. *Note. That the figure and modified description are adapted from* [16]

It is likely that prophage and plasmid abundance and mobility play a large role in influencing gene virulence, antibiotic susceptibility, and defense against viral infections [16]. With advances in synthetic biology, the capacity of phages to mediate the transduction of genes which may increase bacterial virulence or promote antibiotic resistance, has presented the opportunity for temperate phages to be utilized as treatments against bacterial infections by removing 'lysogeny genes' from their genome [46]. The unusual properties of serine integrases including recombination directionality manipulation via simplistic site sequence requirements are leading

to development as efficient and versatile tools with applications in experimental biology, biotechnology, and gene therapy [81]

Vital for successful bacteriophage integration into a host bacterial cell is the ability to repress virally encoded phage genes and the expression of lysogenic genes to allow for a state of dormancy. Bacteriophage repressor systems allow for a quiescent state of being once integrated into its host cell. A repressor system, at the most basic ideology, is employed when a phage-encoded repressor gene that acts by binding at operators that control early promotors to ensure that functions deployed in lytic growth remain unexpressed [92]. The complex genetic switch between lytic and lysogenic growth has been thought to arise from simpler ancestral systems that, over time, has been refined and added to allowing it to 'work better' under certain selective pressures [7].

Of the studies conducted investigating the genetic switched of bacteriophages, the most prominent is that of phage λ (lambda) as chronicled by Mark Ptashne's 1986 book, *A Genetic Switch*. When Lambda infects its host, *Escherichia coli,* lysogeny is achieved via expression of a repressor gene (Cl) that binds to tripartite operators (*oR* and *oL*) at early lytic promotors *pL* and *pR* [67]. Lambda cl autoregulates its synthesis by activation of its own transcription from the promoter for lysogenic maintenance (*pRM*) at moderate Cl concentrations, and represses it when the Cl concentration is high [90]. During induction from lysogeny to a lytic state, Cl is inactivated while the two lytic promotors are de-repressed to allow for Cro expression from pR – Cro is known to prevent lysogeny by repressing Cl transcription from *pRM* [44]. As roughly shown in Figure 8, Lambda's 'genetic switch' is directly determined by the interplay of two regulatory proteins (CI and Cro), three promoters (*pR*, *pL*, and *pRM*), and six nearly homologous operators

(*oR1*, *oR2*, and *oR3* adjacent to *pR* and *oL1*, *oL2* and *oL3* adjacent to *pL*), all encoded in a 5.4-kb DNA segment of the phage chromosome called the immunity region [44].

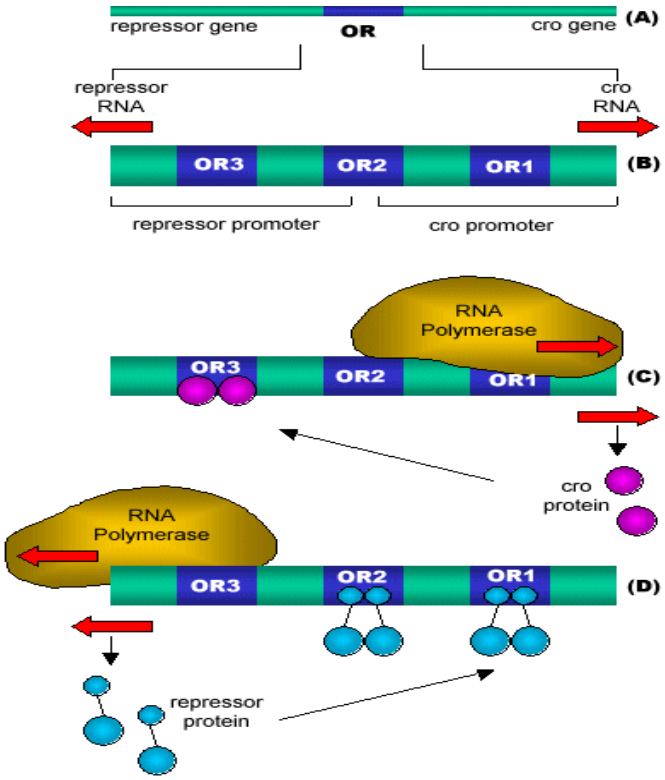



Figure 8. Genetic Switch Overview. (A) The genetic switch region of the temperate lambdoid phages. The repressor and cro genes lie on opposite sides of the operator (OR) region. (B) The repressor and cro genes are transcribed in opposite directions from their respective promoters, which overlap in the middle operator, OR2. (C) The situation in a lytic infection. Cro protein dimers occupy OR3, preventing RNA Polymerase from initiating transcription from the repressor promoter. RNA Polymerase transcribes the cro gene (and other genes), producing more of cro protein, which silences repressor transcription. (D) The situation in a lysogenic cycle. Repressor throws the genetic switch to repressor transcription and cro gene silencing. Repressor protein dimers bind OR2 and OR1, preventing RNA Polymerase from transcribing the cro gene, and promoting repressor transcription from the repressor promoter. Unlike cro, the repressor has an activation domain that promotes RNA Polymerase binding to the repressor promoter. The repressor and cro proteins have differential affinity for the operators. Repressor has lower affinity for OR3 than for OR2 and OR1, and cro has higher affinity for OR3 than OR1 and OR2. The structural bases for this differential affinity will not be addressed in this paper. *Figure and description adapted from (http://earth.callutheran.edu/Academic_Programs/Departments/BioDev/omm/cro/frames/cro_rep.htm)*

Repressor systems serve as vital components to any temperate phage favoring the lysogenic life cycle. They are carried out by bacteriophages such that repressor systems obtain the same end goal of lysogeny but may do so by various means. Studies of mycobacteriophage L5 display an atypical phage repressor system. L5's repressor gene, gp71, binds to multiple asymmetric DNA sites whereby regulating transcription initiation at an early lytic promotor, $P_{left}$, while also affecting downstream gene expression at 'stoperator' sites [8]. Binding of gp71 to these sites results in a strong orientation-dependent polar effect on downstream gene expression and global silencing of prophage gene expression [8]. Another case of unusual repressor systems may be seen in the temperate mycobacteriophage BP which utilizes an integration-dependent immunity system in which the phage attachment site (*attP*) is located within the repressor gene, gp33 [90]. The highly site-specific integration enables the synthesis of a prophage-encoded product, $gp33^{103}$, that is 33 residues shorter at the C-terminus rather that the virally-encoded protein, $gp33^{136}$ [90]. By shortening the virally-encoded gene by 33 residues, a shorter form of the repressor gene becomes an active and stable protein that represses an early lytic promotor, $P_R$, whereas the longer virally-encoded form ($gp33^{136}$) is inactive due to targeted degradation *via* a C-terminal ssrA-like tag [90]. Variations amongst encoded repressor systems have been overserved, including discoveries of a monomeric mycobacteriophage immunity repressor which uses two domains to recognize an asymmetric DNA sequence, and are continually being unearthed and studied [49,60,73].

Once the phage has established lysogeny within the host cell, it aims to protect the cell to allow the lysogen to undergo replication and pass on the integrated viral phage DNA to progeny. To defend the lysogen cell, some prophages have evolved to contain prophage defense genes that aim to mutually benefit the prophage and bacterial host. Temperate phages collude with their

hosts to confer defense against heterotypic, genomically distinct, bacteriophages from establishing a productive infection in the lysogen cell [24]. More than 50% of mycobacteriophage cluster groups contain temperate phages that likely encode, many as of yet undiscovered, prophage-mediated defense mechanisms that interfere with heterotypic, or unrelated, phage infection, often with high specificity [16,18,51]. Studies investigating the collusion between prophages and their hosts has been of ongoing research, as a plethora of bacterially-encoded phage defense systems have been described [22,25,55]. An example of a prophage-mediated defense mechanism, superinfection exclusion, is observed in prophages known to inhabit *E. coli*, whereby phages HK97 [14] and φ80 [87] express proteins that inhibit further phage infection at the cellular surface level. The protein mechanisms observed to block further phage infection include interactions with the cytoplasmic membrane by blocking phage genomic injection and interactions with phage receptors on the bacteria's outer membrane to inhibit phage binding [6].

A miniscule subset of mycobacteriophages that infect *M. smegmatis mc$^2$155,* subclusters/clusters (A2, A3, G, K), can infect *Mycobacterium tuberculosis*, suggesting some may possess a broader host range than the majority of others [37]. While the constraints on infection are not fully understood, they are known to involve a plethora of mechanisms including restriction modification and cell surface receptor availability [35], a variety of mediated abortive infection processes [23], and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) [20].

What is known is that prophages collude extensively with their bacterial hosts to provide defense against viral attack, often profoundly influencing host physiology, endowing pathogenic properties and additional metabolic pathways, and ultimately are capable of providing survival advantages in various environments [24]. Prophage mediated defense mechanisms often act with

exquisite specificity, as observed by a new heterotypic prophage-mediated defense system encoded by the cluster I2 mycobacteriophage Sbash [24]. In the defense system encoded by for Sbash, two lysogenically – expressed genes (*30* and *31*) target the cluster L2 phage Crossroads genes *132* and *142* with outstanding specificity by activating a loss in membrane potential, as shown in Figure 9 [24]; however, Sbash's genes *30* and *31* do not defend against any other mycobacteriophage tested, including other closely related subcluster L2 phages. Distantly related homologues of the Sbash genes *30*/*31* defense genes have been observed in *Gordonia* phage CarolAnn and have been observed to defend against Crossroads when expressed in *M. smegmatis* [24,52]. CarolAnn's genes *43/44* use a different targeting system than those observed by Sbash *30/31*, by expression with the repressor gene with necessity and sufficiency to confer defense against unrelated phages in cluster CZ, namely phage Kita and its close relatives [52]. Blatant similarity can be observed between the two systems employed by phages Sbash and CarolAnn in Figure 9, showing how evolutionary relationships between prophage-mediated defense mechanism acquisition exist between distant relatives.

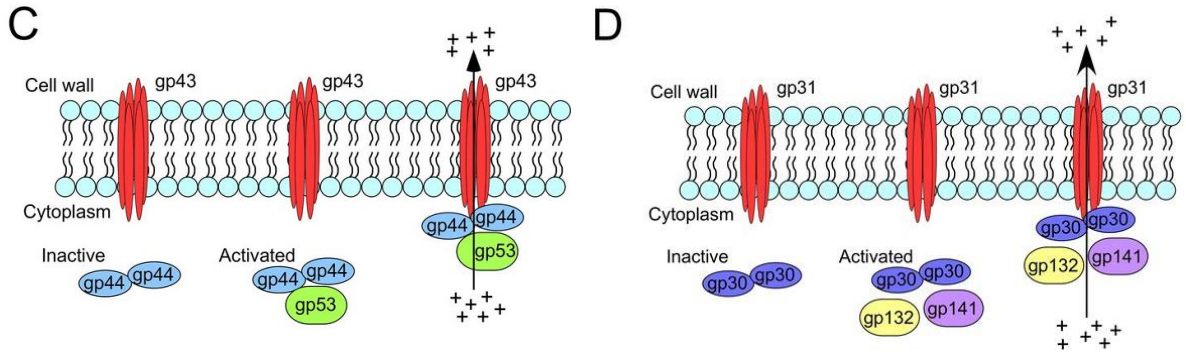



Figure 9. Prophage-mediated defense mechanisms of CarolAnn and Sbash. **(C)** A model for CarolAnn 43/44-mediated defense. CarolAnn gp43 is proposed to be membrane located but inactive as an ion channel until infection with phage Kita (or relatives). During early lytic growth of Kita, gp53 acts either directly or indirectly through CarolAnn gp44 to activate the gp43 ion channel, leading to loss of membrane potential and of intracellular ATP, interruption of

macromolecular synthesis, and loss of cell viability. **(D)** Mechanism of Sbash *30/31* mediated defense against Crossroads. Sbash gp31 is proposed to be membrane located but inactive as an ion channel until infection with phage Crossroads. During early lytic growth of phage Crossroads, gp132 and gp141 act either directly or indirectly through Sbash gp30 to activate the gp31 ion channel, leading to loss of membrane potential and of intracellular ATP, interruption of macromolecular synthesis, and loss of cell viability. It is not known if Sbash gp30 interacts directly with gp31 in either the inactive or the activated state, but substitutions in the central cytoplasmic loop of Sbash gp31 are inactive for defense, consistent with an interaction between Sbash gp30 and gp31. *Note figures and descriptions adapted from* [24,52].

Mycobacteriophages are of interest for a few reasons. Their large database of complete genome sequences expose their high levels of genomic variation that offer insight into overall phage genome diversity and evolutionary mechanisms that create this diversity [37]. Mycobacteriophages have also been used as tools for developing genetic systems for hosts – including the causative agent of human tuberculosis (TB), *Mycobacterium tuberculosis* – that contribute to novel diagnostic strategies, preventative, and therapeutic approaches for TB [32]. Due to the mobile nature of prophages and plasmids, they play a key role in phage infection variations among closely related bacteria [16]. Therapeutic use of phages to treat and control bacterial infections in an age where widespread antibiotic-resistant bacteria threaten our current cache of antibiotics is of fast-rising importance [16,46].

Phage infection profiles do not correlate with whole-genome phylogenies, and mobile elements like prophages and plasmids are likely major contributors [75]. Very few mycobacteriophages infect *Mycobacterium abscessus*, but a cocktail of three phages within the subset known to infect both *M. tuberculosis* and *M. smegmatis* previously discussed were used for therapy of a disseminated, drug-resistant infection in a cystic fibrosis patient with a bilateral lung transplant [17]. Mycobacteriophage treatment of antibiotic resistant human *Mycobacterium abscessus* and *Mycobacterium tuberculosis* infections has recently shown great promise [19].

However, patient *Mycobacterium abscessus* isolates are only susceptible to a narrow range of mycobacteriophages due to the presence of prophages and plasmids carrying uncharacterized defense systems. Constituting as much as 10-20% of a bacterium's genome, prophages are contributors to inter-strain variability [80]. Cluster M-like prophages have been identified in patient isolates [16]. Bacteriophage therapy for antibiotic resistant *M. abscessus* infections has become as beneficial avenue, and cluster M-like prophages therefore have become of peak interest. However, little is known about how cluster M phages establish lysogeny and what prophage-mediated defense genes they may express. Thus, the important of the identification of cluster M IPhane7's integration sites become of elevated importance, especially with the difficulty of identifying serine integrase systems.

There are seventeen members belonging to cluster M mycobacteriophages. Cluster M is divided into three subclusters – M1 (10 members), M2 (6 members), and M3 (1 member) – and are known to be temperate. Cluster M phages possess siphoviral morphologies with icosahedral heads and unusually long, non-contractile tails. They contain larger than average dsDNA genomes of approximately 80.2 to 83.7 kbp, and have an average of 61.2% GC content, 141.3 genes, and 18.7 tRNAs [82]. In contrast to other mycobacteriophages, cluster M's have noncanonical genome architectures and have been found to contain several unusual sets of conserved repeats, suggesting novel regulatory systems for transcription and translation. They are known to have a non-centrally located integrase and a set of reverse genes found before structural genes. Though cluster M-like prophages may play an important role in resistance to bacteriophage therapy, little is known about how they establish lysogeny and what prophage defense mechanisms they may contain [16]. Temperate phages possess a repressor gene that is key

in lysogenic establishment, however bioinformatic analysis failed to identify a repressor gene encoded by Cluster M bacteriophages.

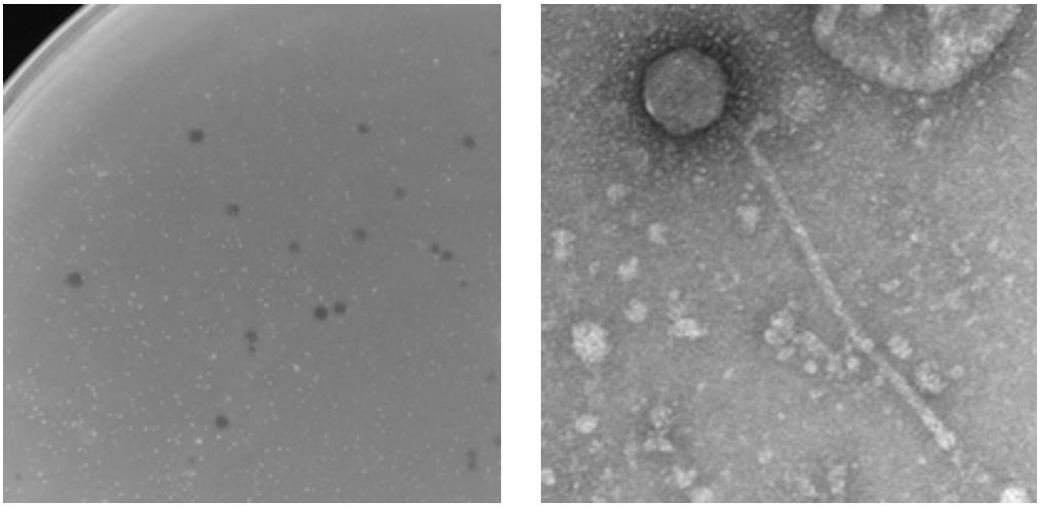


Figure 10. Electron microscope image of IPhane7 plaques and phage structure. *Images gathered from* (*The Actinobacteriophage Database | Phage IPhane7*) *on 12/11/2022.*

Cluster M bacteriophage IPhane7 was discovered in Cullowhee, North Carolina in 2016. Electron microscope images of IPhane7 plaques and phage structure are displayed in Figure 10 showing the long, non-contractile tails and siphoviridae morphology characteristics. Previous work in the Gainey laboratory by Erin Cafferty raised interest in IPhane7's gene product 1 (gp1) and gene product 2 (gp2) possible roles during lysogeny. IPhane7's gp1/2 were initially thought to play a role in lysogeny, possibly encoding for the phage's repressor system, as Cluster M repressor genes failed to be identified bioinformatically. Preliminary analysis of IPhane7 had revealed that gp1 is likely a secreted-signal peptide and likely acts to prevent infection at the level of phage entry into the cell. Further investigation in this study led to the identification of a likely prophage-mediated defense mechanism encoded for by IPhane7's gp1 that synergizes gp2.

Presented is a preliminary investigation into the novel prophage-mediated defense mechanisms encoded for by cluster M mycobacteriophages gp1 and gp1/2 together, with the support of bioinformatically analyzed IPhane7 defense escape mutants (DEMs) that have gained mutations to overcome defense mechanisms encountered. Experimental determination of the viral integration sites in the bacterial genome of an IPhane7 lysogen will be discussed, and defense profiles of a panel of cluster M lysogens will be investigated. Our results provide the presumed first identification of integration sites for cluster M IPhane7 lysogen and initial investigations into the novel and inherently complex prophage defense systems of cluster M phages in an effort to add to the knowledge of bacteriophage use as therapy against antibiotic-resistant bacterial infections in humans.

# CHAPTER ONE: MATERIALS AND METHODS

## Solution, Media, and Gel Preparations

### Solutions, Media, and Agar Plates

All protocols used to make solutions, mediums, and plates were followed from the SEA-GENES Instructor Guide 'Recipes Cards.' The following solutions were used: AD Supplement, Anhydrotetracycline Stock, Calcium Chloride (1 M), Glycerol (40%), Glycerol (10%), Kanamycin Stock, and Tween80 (20%). [68]

The following mediums and plates were used: 7H9 Liquid Medium (Neat), 7H9 Top Agar (2X), 7H9 Top Agar (1X), Agarose Gel (1%), LB Agar Plates, and LB Liquid Media. [68] All bacteriophages used in the study were accessed from the Gainey laboratory stock or the Hatfull laboratory with fully sequenced and annotated genomes. Their Genbank accession numbers can be found in Table A1.

## Lysogen Immunity Assay Preparation

### Lysogen Creation and Purification

Cluster M bacteriophages, Reindeer and PegLeg, were obtained from the Gainey laboratory for lysogen creation. Bacteriophages of interest were extracted from stock via aseptic technique. Creation of the lysogens followed the Phagehunting Program protocol for 'Lysogeny Experiments' (phagesdb.org). Two plates were prepared by pouring a lawn of host cells, *M. Smegmatis mc$^2$155*, on top of Middlebrook 7H9 agar plates. Then, serial dilutions to $10^{-4}$ of Reindeer and PegLeg phages were spotted on each plate to obtain mesas from each phage. Mesas samples were collected from each plate and streaked. Five colonies from each plate were chosen

to purify. Each lysogen candidate underwent three rounds of purification. Upon completion of triple purification, two lysogen candidates for each phage, Reindeer and PegLeg, were prepared for archive in -80℃ freezer.

### Lysogen Immunity Assay

Immunity assays were performed for the following lysogens: IPhane7 (#1), Reindeer (#1), Reindeer (#2), PegLeg (#1), PegLeg (#2), Rey (#5), Nanosmite (#1). The Immunity Assay protocol outlined in the Phagehunting Program protocol for 'Lysogeny Experiments' (phagesdb.org) was followed. Two lysogen immunity assays were performed utilizing two differing phages panels to test against. A lawn of each lysogen was prepared by adding 1mL bacteria and 4.0mL top agar to a plate and allowed to solidify. Cluster M lysogen immunity assay applied 2µL spots of 8-fold serial dilutions of cluster M phages IPhane7, Bongo, PegLeg, Reindeer, Rey, and Nanosmite to plates containing a lawn of lysogen cells for each lysogen listed above. Non-cluster M viral panel lysogen immunity assay applied 2µL spots of 8-fold serial dilutions of phages Charlie (N), Xeno (N), Phayonce (P5), Island3 (I1), and Che9C (I2) to plates containing a lawn of lysogen cells for each lysogen listed above. Spots were allowed to dry completely before moving. Cluster M Immunity Assay plates were incubated for 72 hours at 30℃ and then imaged. Non-cluster M viral panel plates were incubated for 72 hours at 37℃.

## Bioinformatic Characterization

### DEM Assembly and Analysis

IPhane7 Defense Escape Mutants (DEMs) were created and sequenced in the Gainey Laboratory. Following sequence data generation, Montana Henson assembled and analyzed the following samples: 1, 7-12. Assembly was performed via programs in the 2020 SEA VM

package made available by the HHMI SEA-PHAGES program. DEMs samples were assembled into contigs using GS De Novo Assembler (Newbler) v2.9, consed v29 (150417), AceUtil, and analyzed for the mutations using NCBI's BLASTn [2] compared to IPhane7 genomic data from the Actinobacteriophage Database at phagesbd.org. Mutation coordinates were investigated using PECAAN v20220930 to observe mutations within coding regions of the genome. Phyre2 v2.0 was used to analyze mutation affects to protein assembly within coding regions.

**Lysogen Integration**

The Gainey Laboratory had previous sequenced an IPhane7 (#1) lysogen via whole genome sequencing. Montana Henson assembled and analyzed the IPhane7 lysogen for *attP* and *attB* sites. The lysogen was assembled into contigs using GS De Novo Assembler (Newbler) v2.9, consed v29 (150417), and analyzed for areas of discrepancies in AceUtil. Partial assembly of the lysogen resulted in 586 contigs. Each contig was individually investigated for the presence of IPhane7's genome via NCBI's BLASTn [2]. Seven of the 586 contigs were found to contain IPhane7 genomic sequence (contigs 14, 20, 67, 79, 194, 212, and 445). The seven contigs identified were then mapped to the respective genome they aligned to – either IPhane7 or *Mycobacterium smegmatis mc$^2$ 155*, or both due to junction locations – via NCBI's BLASTn align two or more sequences for direct comparison. Each contig's coordinates were aligned to IPhane7 (Genbank: MH697587.1) and *M. smegmatis* (GenBank: CP009494.1) during the NCBI BLASTn alignment. Once coordinates were identified, all contig coordinate information was synthesized to locate the *attP* and *attB* sites within the lysogen. Primers were designed in SnapGene for PCR amplification of each junction and subsequent Sanger sequence to confirm integration sites.

Table 1. Primers for amplification of junctions in the IPhane7 lysogen.

| JUNCTION: | LENGTH (BP): | DIRECTION: | SEQUENCE: |
|---|---|---|---|
| **1** | ~684bp | Forward | 5' –GGCATCGGCCCCGTCTGA– 3' |
| **(PCR)** | Annealing: 64℃ | | |
| | | Reverse | 5' –GGCTGCGACGTGTCCTTAAACGT -3' |
| **2** | ~586bp | Forward | 5' -CAATGGCCTCACTCGGCTTAACGTC -3' |
| **(PCR)** | Annealing: 63℃-65℃ | | |
| | | Reverse | 5' –CGTGCAGGTCGACGACGCAGAA– 3' |
| **1 (Sequencing)** | ~684bp | Forward | 5' – GCATCGGCCCCGTC– 3' |
| | Annealing: 55℃ | | |
| | | Reverse | 5' – GGCTGCGACGTGTC-3' |
| **2 (Sequencing)** | ~586bp | Forward | 5' - GGCCTCACTCGGCTTAA-3' |
| | Annealing: 63℃-65℃ | | |
| | | Reverse | 5' – AGGTCGACGACGCAGA– 3' |

## IPhane7 gp1 Characterization

IPhane7 gp1 protein was run through AlphaFold [40,88] to generate a predicted protein folding model and displayed a long string of amino acids that then formed into a "beta ball." IPhane7's gp1 was run through HHpred [93] and identified to possess a predicted Sec-signal peptide with cleavage between amino acid 24 and 25. IPhane7 gp1 protein sequence was input into TOPCONS [86], which generated a unanimous conclusion for the inclusion of a Signal peptide sequence, as shown in Figure 11.

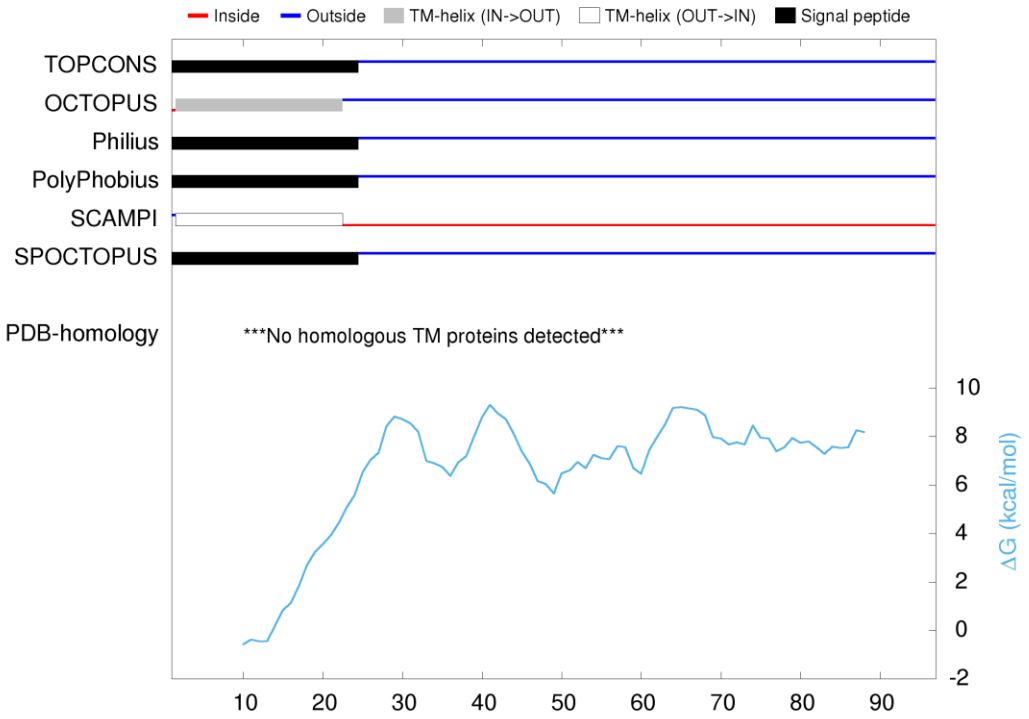



Figure 11. A TOPCONS generated for IPhane7 gp1 protein sequence displayed a unanimous opinion of a signal peptide at the beginning of the gene by all the programs: TOPCONS, OCTOPUS, Philius, PolyPhobius, SPOCTOPUS, and SCAMPI who was similar with a slightly different generation.

**Protein Sequence Alignment of Cluster M gp1's**

Protein sequences were extracted from the "Actino_Draft" Phamerator database (www.phamerator.org © Cresawn, Hatfull, Bogel, Mavrich, Gauthier, HHMI SEA-PHAGES) for all gp1 sequences used for alignment. ClustalW2 [85] was used as a multiple sequence alignment tool. ESprit v3.0 [69] was used to render sequence similarities and secondary structure information from the aligned sequences for analysis.

## Plasmid Vector Cloning

**pExTra Vector PCR, Ligation and Digestion**

PCR amplification was performed to linearize and amplify the pExTra vector using NEB's Q5 Hot Start 2x Master Mix [66]. Primer design for all PCR was performed in SnapGene. All primers were ordered from IDT. Primers were resuspended with $ddH_2O$ to a 100µM concentration, and then diluted with $ddH_2O$ to 10µM for working concentrations. Upon completion, primers were stored at -4℃. The PCR reaction for the pExTra vector was carried out using the following conditions:

Table 2. pExTra vector PCR reaction cycles.

| **STEP** | **TEMPERATURE** | **TIME** |
|---|---|---|
| Initial Denaturation | 98℃ | 1 minute |
| 29 cycles | 98℃ | 10 seconds |
| | 68℃ | 10 seconds |
| | 72℃ | 3 minutes |
| Final Extension | 72℃ | 5 minutes |
| Hold | 4℃ | ∞ |

Gel electrophoresis was performed on the pExTra PCR product using an Ethidium Bromide 1% agarose gel with 10µL/100mL Ethidium Bromide for 60 minutes at 100 V. The linearized pExTra PCR product was cut from the gel using a razor blade, and then purified following the instructions from the NEB's Monarch Gel Extraction Kit. The pExTra product was then digested for 2 hours at 37℃ using NEB's DpnI restriction enzyme digestion protocol [53]. After DpnI digestion, PCR cleanup was performed on the reaction mixture to get the linearized pExTra into water, and the final product was stored at -4℃.

**HiFi Assembly of Genes into pExTra**

Primers were designed in SnapGene to amplify genes of interest and contained 15-25 bases of gene-specific sequence with the addition of uniform pExTra forward homology sequence (ATGCGGAGGAATCACTTCCA) on the 5' end of the forward strand, and the addition of uniform pExTra reverse homology sequence (TGCAGGATCCGACTCGAGTGTCGAC) on the 5' end of the reverse strand. These added sequences will provide overlaps with the vector used for cloning and will allow for the use of a technique called HiFi assembly to ligate the PCR products into the vector.

Table 3. Primers for amplification of Cluster M Bacteriophages gp1. Highlighted sequences are the forward and reverse homologies added.

| **Cluster:** | **Phage:** | **Direction:** | **Sequence:** |
|---|---|---|---|
| **M1** | IPhane7 | Forward | 5' – ATGCGGAGGAATCACTTCCATATGAGAACCACACTCACCGCGC – 3' |
| | | Reverse | 5' – TGCAGGATCCGACTCGAGTGTCGACTCAGGCTGGCGTGAGCGT – 3' |
| **M1** | Reindeer | Forward | 5' – ATGCGGAGGAATCACTTCCATATGAGAACCAAACTCTTCGCACTCATCG -3' |
| | | Reverse | 5' – TGCAGGATCCGACTCGAGTGTCGACTCAGGCTGGTGTGAGCGTCAG – 3' |
| **M2** | Rey | Forward | 5' – ATGCGGAGGAATCACTTCCATATGGGACTCAAAACCGCTGTGC – 3' |
| | | Reverse | 5' –TGCAGGATCCGACTCGAGTGTCGACCTACTGGTGGTCGGCGCGA– 3' |
| **M3** | Nanosmite | Forward | 5' –ATGCGGAGGAATCACTTCCATATGAGAGCTGTCTTAGCCGTCTCA– 3' |
| | | Reverse | 5' –TGCAGGATCCGACTCGAGTGTCGACTCAGGCTGGCTGGAGCATC –3' |

Gene(s) of interest, gp1, were amplified via PCR from the following genomes - IPhane7, Reindeer, Rey, and Nanosmite – using NEB's Q5 Hot Start 2x Master Mix [66] to a volume of 25µL per phage. The PCR reaction was conducted using the following conditions:

Table 4. pExTra genes of interest PCR reaction cycles.

| STEP | TEMPERATURE | TIME |
|---|---|---|
| Initial Denaturation | 98℃ | 30 seconds |
| 29 cycles | 98℃ | 10 seconds |
| | 63℃ | 10 seconds |
| | 72℃ | 25 seconds |
| Final Extension | 72℃ | 2 minutes |
| Hold | 4℃ | ∞ |

PCR product verification was performed via gel electrophoresis using the following setup: 4 µL of each PCR product was mixed with 2µL 6x dye and loaded onto a 1% agarose gel containing a 10µL/100mL concentration of ethidium bromide alongside 10 µL NEB's 100bp standard DNA ladder. The gel was run for 60 minutes at 100 V. Upon verification, pExTra gp1 PCR products underwent PCR clean up using NEB's Monarch® PCR & DNA Cleanup Kit (5 µg) [65], were eluted to 20µL, and stored at -4℃.

**Ligation and Transformation of pExTra Vector and Products.**

Ligation of the pExTra vector and PCR product clean-ups was performed following NEB's HiFi DNA Assembly Protocol [54]. 2µL of 0.06pmol PCR DNA was added to 1µL of 0.03pmol pExTra vector DNA, 2µL of $ddH_2O$, and 5µL of HiFi Master Mix and placed in the thermocycler for 30 minutes at 50℃. Transformation of the pExTra vector and DNA construct into chemically competent *E. coli* cells was done by following the SeaGenes Protocol 2.9: Chemical Transformation of Bacteria [63]. We proceeded by adding 3µL of pExTra ligation mixture into 25µL of c2981 *E. coli* cells for 10 minutes, followed by heat shock at 42℃ for 30

seconds, then placed back on ice for 5 minutes. 100µL of SOC media was added to each transformation, and the tubes were placed in a shaker at 37° C for 60 minutes. 10µL and 90µL of each transformation was plated onto agar plates containing LB plus 50 µg/mL kanamycin. Colonies were allowed to incubate overnight at 37 degrees. No growth was observed on the 10µL, only 90µL plated plates.

**Ligation and Digestion of pMH94 Vector**

To linearize the pMH94 vector, 4µg of the pMH94 vector was digest for 2 hours at 37℃, with EcoRI and KpnI following NEB's standard 50µL digest protocol (*Restriction Digest Protocol | NEB*, n.d.). After digestion, 10µL of CIP was added to the mixture and incubated for 10 minutes at 37℃ followed by heat inactivation at 80℃ for 2 minutes. The digested vector was then purified using NEB's Monarch DNA Cleanup Kit protocol. The vector was eluted in a final volume of 25µL of ddH20

**Traditional Cloning of Genes into pMH94**

Primers were designed in SnapGene to include gp1 and the preceding intergenic region of the gene with the addition of an EcoRI site (GAATTC) on the 5' end of the forward strand plus several cytosines, and the addition of a KpnI site (GGTACC) on the 5' end of the reverse strand enabling directional control of gene ligation into the vector.

Table 5. Primers for amplification of Cluster M Bacteriophages gp1. Highlighted in blue are added cytosines at the 5' ends. Highlighted in yellow are the EcoRI (GAATTC) and KpnI (GGTACC) sites necessary for cloning.

| CLUSTER: | PHAGE: | DIRECTION: | SEQUENCE: |
|---|---|---|---|
| **M1** | IPhane7 | Forward | 5' – CCCGAATTCCGAACGGCTCCCGCCC – 3' |
| | | Reverse | 5' –CCCGGTACCTCAGGCTGGCGTGAGCGT – 3' |
| **M1** | Reindeer | Forward | 5'- CCC GAA TTCCGAAGGGCTCCCTGCCCA -3' |
| | | Reverse | 5' –CCCGGTACCTCAGGCTGGTGTGAGCGTCAG -3' |
| **M2** | Rey | Forward | 5' – CCCGAATTCCGAAGACAGCTTCTCGCTTGAGC – 3' |
| | | Reverse | 5' –CCCGGTACCCTACTGGTGGTCGGCGCGA - 3' |
| **M3** | Nanosmite | Forward | 5' – CCCGAATTCCAAACGGTGCTTCTCGCTTGAGC – 3' |
| | | Reverse | 5' –CCCGGTACCTCAGGCTGGCTGGAGCATC – 3' |

Gene(s) of interest, gp1, were then amplified via PCR from the following genomes - IPhane7, Reindeer, Rey, and Nanosmite – using NEB's Q5 Hot Start 2x Master Mix [66] to a volume of 25µL per phage. The PCR reaction was conducted using the following conditions:

Table 6. pMH94 genes of interest PCR reaction cycles.

| STEP | TEMPERATURE | TIME |
|---|---|---|
| Initial Denaturation | 98°C | 30 seconds |
| 29 cycles | 98°C | 10 seconds |
| | 63°C | 10 seconds |
| | 72°C | 35 seconds |
| Final Extension | 72°C | 2 minutes |
| Hold | 4°C | ∞ |

PCR product verification was performed via gel electrophoresis using the following setup: 4 µL of each PCR product was mixed with 2µL 6x dye and loaded onto a 1% agarose gel containing a 10µL/100mL concentration of ethidium bromide alongside 10 µL of NEB's 100bp

standard DNA ladder. The gel was run for 60 minutes at 100 V. Upon size verification, pMH94 gp1 PCR products underwent PCR clean up using NEB's Monarch® PCR & DNA Cleanup Kit [65] and were eluted in 20µL of sterile nanopure$H_20$. Then 20µL of the cleaned up pMH94 gp1 PCR products were digested for 2 hours at 37℃, with EcoRI and KpnI following NEB's standard 50µL digest protocol (*Restriction Digest Protocol | NEB*, n.d.). Upon completion of the digest, PCR products were purified using NEB's DNA clean-up kit [65], eluted in 20µL of nanopure$H_20$, and stored at -4℃.

**Ligation and Transformation of PCR products in pMH94**

Ligation of pMH94 vector and digested PCR products was performed by combining 5µL of DNA PCR product,1µL pMH94 vector DNA, 10µL of Quick Ligase buffer and 1µL of Quick Ligase and incubated at 37℃ for 5 minutes. Transformation of the pMH94 vector and DNA construct into chemically competent *E. coli* cells was done by adding 3µL of the ligation mixture to 50µL of C2981 *E. coli* cells for 10 minutes, on ice, followed by heat shock at 42°C for 30 seconds, and placed on ice for 3 minutes. 950µL of SOC media was added to each transformation, and the tubes were placed in a shaker at 37° C for 60 minutes. 50µL and 200µL of each transformation was plated onto agar plates containing LB plus 50 µg/mL kanamycin. Colonies were allowed to incubate overnight at 37 degrees.

**Colony Screening**

Colony PCR was performed to ensure correctly ligated clones in both pMH94 and pExTra constructs. In pMH94, forward and reverse primers were designed to sit approximately 100bps before the EcoRI site and after the KpnI sites. In pExTra, vector primers uni_F and

uni_R were used for the PCR reaction and designed to sit 16bp and 47bp before IPhane7 gp1 respectively.

PCR reaction mixtures were created by combining 6.25µL 2x OneTaq DNA polymerase to 0.25µL forward and 0.25µL reverse primers, selecting a colony from each transformed plate and adding it to the mixture tube, and adding 5.75µL dd$H_2O$ to a total volume of 12.5µL. PCR reaction weas carried out using the following conditions:

Table 7. PCR reaction cycles for vectors pMH94 and pExTra Colony PCRs.

| **STEP** | **TEMPERATURE** | **TIME** |
|---|---|---|
| Initial Denaturation | 94℃ | 5 minutes |
| 30 cycles | 94℃ | 30 seconds |
| | 63℃ | 45 seconds |
| | 68℃ | 75 seconds (in pMH94)<br>60 seconds (in pExTra) |
| Final Extension | 68℃ | 5 minutes |
| Hold | 4℃ | ∞ |

The correct size of each PCR product was confirmed by running 6µL of mixed with 4µL 6x dye on a 1% agarose gel containing a 10uL/100mL concentration of ethidium bromide alongside NEB's 100bp standard DNA ladder. Colonies that were screened were archived on an LB plate prior to being added to the PCR mixture. A pipette tip was touched to each colony then was gently scraped across a fresh LB agar plate containing 50ug/ml concentration of kanamycin, the pipette tip was then shaken into each PCR reaction. This was left to incubate at 37°C overnight. Positive clone confirmation was completed using Sanger sequencing. Positive

colonies were transferred in to 5mL LB plus 50µg/mL Kanamycin and grown to saturation at 37° C overnight in the shaking incubator. Plasmid DNA was purified using NEB's Monarch Plasmid DNA Miniprep Kit [50] prior to Sanger sequencing.

**Electroporation of Plasmids into *M. smegmatis***

pMH94 vector, pMH94 vector + genes of interest, pExTra vector, and pExTra vector + genes of interest were then electroporated into 50µL of electrocompetent *M. smegmatis* cells that had been thawed on ice. 1µL of plasmid from each of the samples listed was added to 50µL of electrocompetent cells and incubated on ice for 10 minutes. Cells from each microtube were then transferred to 1mm cuvettes that had been cooled on ice. Cells in the cuvette were then electroporated using standard *E. coli* 1mm settings, and then promptly placed back on ice. 1mL of 7H9 media – 9mL 7H9 neat, 1mL AD Supplement, 100µL $CaCl_2$ at 100mM – was then added to the cuvette containing the electroporated cells. The mixture was swooshed in the cuvette to ensure the cells were retrieved from the cuvette, and then transferred from the cuvette to an Eppendorf tube. This process was performed for all samples. Once all samples had been electroporated and 7H9 media added, they were placed at 37℃ for 2 hours in a shaking incubator. After two hours, 100µL was plated onto 7H9 agar plates containing 5ug/mL kanamycin and placed at 37℃ for 4 days. Colonies on the plates were picked with a pipette tip and transferred to a test tube with 5mL of 7H9 media with Tween80 (20%) (4.5 mL 7H9 neat, 0.5 mL AD supplement, 50uL CaCl2, 12.5uL tween 0.5uL kanamycin (50µg/mL). These reaction tubes incubated in a 37°C shaker for 48 hours. The samples were archived at -80℃ in a mixture of 0.5mL culture and 0.5mL 40% glycerol.

**Defense Assays**

**Viral Purification**

Prior to beginning viral challenge assays, our viral panel needed to be verified for purity and identity of the phage stocks received in the Gainey Laboratory for the following phages: Che9c, Xeno, Phayonce, Charlie, and Island3. Primers were designed in SnapGene for each phage that included unique features of each genome that only that respective phage possessed to test for identity. All primers were designed to encompass between 400bp – 800bp of each genome and have an average annealing temperature of 63℃ – Che9c (670bp), Xeno (799bp), Phayonce (617bp), Charlie (775bp), and Island3 (428bp). PCR reaction mixtures were set up as followed: 12.5µL OneTaq 2x DNA polymerase Master Mix, 0.5µL of 10µM forward primer and 0.5µL of 10µM reverse primer, 1µL of DNA template obtained from boiled phage cultures, and 10.5µL dd$H_2O$ to a total volume of 25µL.

Table 8. PCR reaction cycles for Viral Panel PCR Verification.

| **STEP** | **TEMPERATURE** | **TIME** |
|---|---|---|
| Initial Denaturation | 94℃ | 30 seconds |
| 30 cycles | 94℃ | 30 seconds |
| | 63℃ | 45 seconds |
| | 68℃ | 75 seconds |
| Final Extension | 68℃ | 5 minutes |
| Hold | 4℃ | ∞ |

Amplification of the correct size PCR products was confirmed by running 6µL of each PCR product from the 25µL total PCR volume mixed with 8µL 6x dye on a 1% agarose gel containing a 10uL/100mL concentration of ethidium bromide alongside NEB's 100bp standard

DNA ladder. Upon successful confirmation of viral stocks, viral stocks were then titered to the same concentrations.

**Viral Titer**

The following bacteriophages were obtained from stocks in the Gainey laboratory and titered to $1x10^9$ concentrations for experimental use: IPhane7, Bongo, Reindeer, Rey, PegLeg, Nanosmite, Che9c, Xeno, Phayonce, Charlie, and Island3. Each phage was serial diluted in phage buffer to $10^{-8}$. 2µL of each dilution was pipetted onto top agar mixed with *M. smegmatis* using a multi-channel micro-pipette. Plates were incubated at 37℃ for 48 hours, except for Rey who was incubated at 30℃ for 48 hours. At lower dilutions of virus, individual plaques can be counted. From the number of individual plaques formed in a single 2µL droplet of diluted viral stock, the concentration of the viral stock can be calculated in plaque forming units (PFU)/mL. The number of plaques formed in a 2µL droplet of serial diluted viral stock is multiplied by $10^n$, where n is the number of dilutions away from the stock the in the serial dilution the droplet came from, to calculate the number of plaque forming units in 2µL of the stock solution. This was then divided by 0.002mL to give the number of PFU's in 1mL of stock solution. This titer was done in triplicate and the average concentration was taken.

$$\frac{\#\ plaques}{0.002mL}\ x\ dilution\ factor = X\ PFU/mL$$

Viral stocks were then diluted to $1x10^9$ PFU/mL in phage buffer for experimental use.

**Defense Immunity Assays**

**In pMH94.**

After electroporation, colonies formed on electroporated plates were picked with a pipette tip and transferred to a test tube with 2mL of 7H9 media with Tween80 (20%) (9mL 7H9 neat, 1mL AD supplement, 100µL CaCl2, 25µL tween 1µL kanamycin (50µg/mL) at a total volume of 10mL to be used in 2mL increments for experimentation). These reaction tubes incubated in a 37°C shaker for 48 hours. 50µL of the culture was then transferred into the same mixture of 7H9 media containing no Tween80 (20%) and incubated in a 37° C shaker for 48 hours. After the 48 hours, top agar was prepared by combining 25mL warm 2x Top Agar, 25mL 7H9 neat, 0.5mL 100mM CaCl2, and 5µL kanamycin. While still warm (55°C) 4mL top agar was mixed with 1mL of the 7H9 media culture and pipetted onto 7H9 agar plates containing kanamycin (50µg/mL) and allowed to solidify. A control plate containing only *M. smegmatis* + vector electroporated cells was used. Serial dilutions to $10^{-8}$ in phage buffer were performed for each phage and 2µL of each dilution was spotted onto the plate using a multi-channel micropipette. Spots were allowed to dry completely and then plates were inverted and incubated at 37℃ for 72 hours, then imaged. Exceptions to incubating temperatures apply for bacteriophage Rey (M2), who is incubated at 30℃ in all experiments.

**In pExTra.**

All defense assays were performed following the SeaGenes Protocol 3.3: Defense Assay [64(p3)] involving pExTra experimentation. After electroporation, colonies formed on electroporated plates were picked with a pipette tip and transferred to a test tube with 2mL of 7H9 media with Tween80 (20%) (9mL 7H9 neat, 1mL AD supplement, 100µL CaCl2, 25µL tween 1µL kanamycin (50µg/mL) at a total volume of 10mL to be used in 2mL increments for experimentation). These reaction tubes incubated in a 37°C shaker for 48 hours. 50µL of the culture was then transferred into two tubes, one containing 5µL (50µg/mL) of aTc inducer and

one without aTc, with the same mixture of 7H9 media containing no Tween80 (20%) to a volume of 2mL per tube and incubated in a 37° C shaker for 48 hours. After the 48 hours, top agar was prepared by combining 25mL warm 2x Top Agar, 25mL 7H9 neat, 0.5mL 100mM CaCl2, and 5µL kanamycin. While still warm (55°C) 4mL top agar was mixed with 1mL of the 7H9 media culture containing no aTc and pipetted onto 7H9 agar plates containing kanamycin (50µg/mL) and allowed to solidify. Mixtures that contained aTc during the no tween incubation were plated on 7H9 agar plates containing kanamycin (50µg/mL) and aTc (50µg/mL) and allowed to solidify A control plate containing only *M. smegmatis* + vector electroporated cells was used, one that was incubated with aTc and one without aTc. Serial dilutions to $10^{-8}$ in phage buffer were performed for each phage and 2µL of each dilution was spotted onto the plate using a multi-channel micropipette. Spots were allowed to dry completely and then plates were inverted and incubated at 37°C for 72 hours, then imaged. Exceptions to incubating temperatures apply for bacteriophage Rey (M2), who is incubated at 30°C in all experiments.

CHAPTER TWO: RESULTS OF CLUSTER M LYSOGENY ANALYSIS

## Determining the Defense Profiles of Cluster M Lysogens

In an effort to investigate the defense profiles of Cluster M lysogens, two independent defense profile challenge experiments were conducted using an in-house panel of Cluster M bacteriophages to evaluate possible homoimmunity to fellow cluster M members, and then a viral panel consisting of phages not belonging to Cluster M to test for heteroimmunities.

A former graduate student in the Gainey lab, Erin Cafferty, had previously created IPhane7 (M1), Rey (M2), and Nanosmite (M3) lysogens. To get a holistic view of all M subclusters – M1, M2, and M3 – defense capabilities, Reindeer (M1) and PegLeg (M1) lysogens were created by Montana Henson. Reindeer and PegLeg were chosen to encompass the genomic difference amongst cluster M1 phages.

### Creating Lysogens

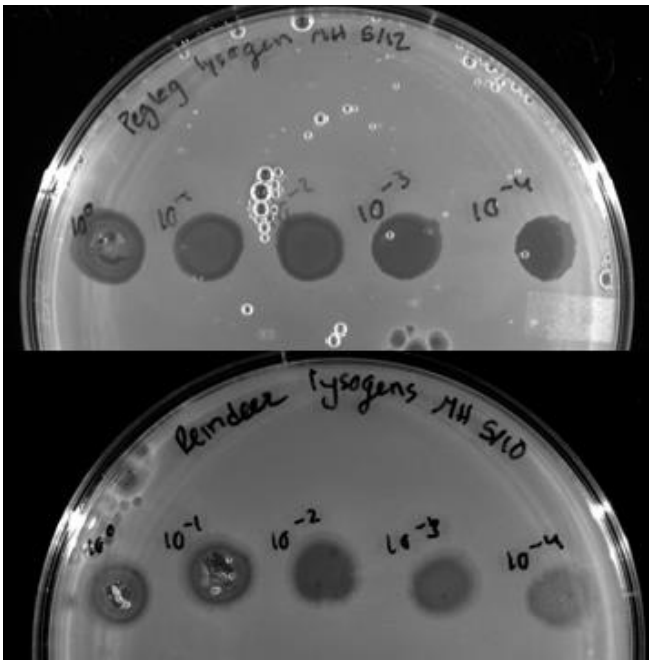



Figure 12. Growing bacteriophages Reindeer (M1) and PegLeg (M1) to obtain mesas. On a lawn of host *M. smegmatis* cells, serial dilutions of $10^{-4}$ were performed for each phage and 5µl spotted on each plate then incubated for several days at 37℃. A mesa is recognizable in

all PegLeg dilutions and in all Reindeer dilutions. Cells were obtained from the mesas in both PegLeg and Reindeer$10^0$ dilutions and streaked onto a new plate and incubated for 2-4 days at 37℃.

A lysogen is a bacterium that contains the bacteriophage genome. To begin, serial dilutions of phages Reindeer and PegLeg were spotted on a lawn of host cells and allowed to incubate for several days. The plates were checked daily for an overgrowth of bacterial cells in the zone of clearing. The turbid 'overgrowth' that appears as a very cloudy center spot surrounded by clearing is called a 'mesa;' this represents cell growth in the presence of a phage and will yield lysogens of each phage desired. Cells from each phage's mesa are streaked onto a new plate. From the streaked plate, five colonies were chosen to undergo triple purification via patch tests. A patch test utilized two plates, one 7H9 plate and one containing a lawn of host cells and allows for the determination of putative lysogens.

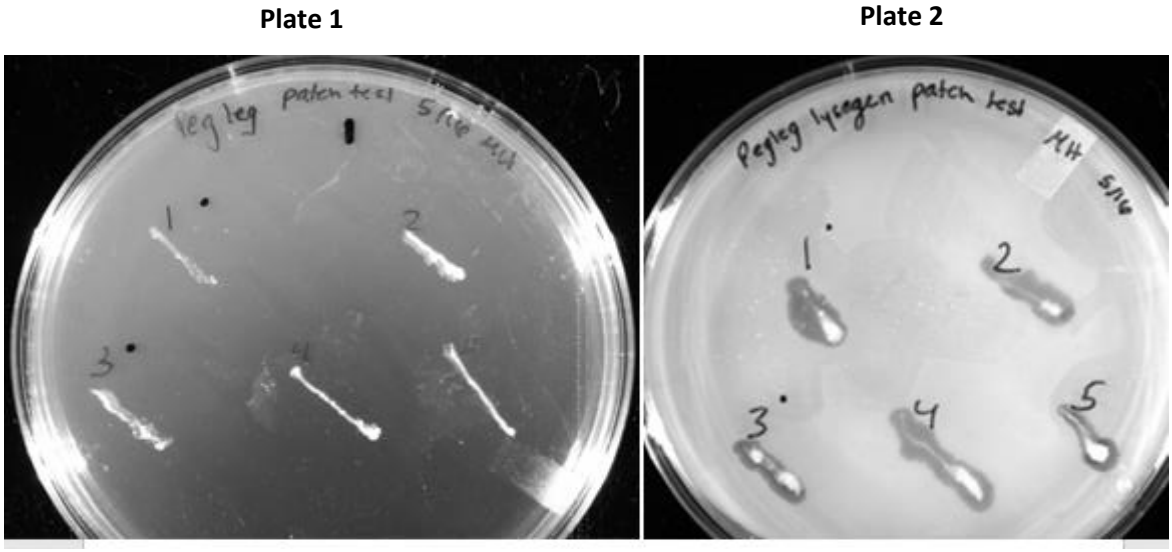



Figure 13. A plate containing PegLeg lysogen colonies obtained from the streaked plate using the mesa cells from the plate in Figure 12. Plate 1 is patches of 5 individual colonies streaked on a 7H9 plate (Plate 1) and a *M. smegmatis* lawn (plate 2). Patches 1-5 show phage release. Cells from Plate 1 were used for further purification.

Once cells from a streaked plate have undergone a patch test, two colonies from each patch test were carried forward for further rounds of purification. Three rounds of purification were performed for two candidate Reindeer and PegLeg lysogens, which are referred to as PegLeg #1 and #2, and Reindeer #1 and #2 in all Lysogen challenge experiments conducted.

**Verification of Viral Panel**

In order to begin viral challenges, verification of our viral panel was necessary. Our viral panel consisted of bacteriophages Charlie (N), Xeno (N), Phayonce (P), Island3 (I1), and Che9C (I2). These phages were selected based on certain criteria including the inclusion or lack of a minor tail protein (gene 30 in IPhane7, Pham 44634) and accessibility to the lab. Charlie (gene20) is the only phage that possesses the desired minor tail protein; Xeno, Phayonce, Island3, and Che9C do not possess the gene notated by Pham 44634. Once the viral panel was verified (Figure 14), each virus was serial diluted to titer the stock to $10^{-9}$ viral concentrations.

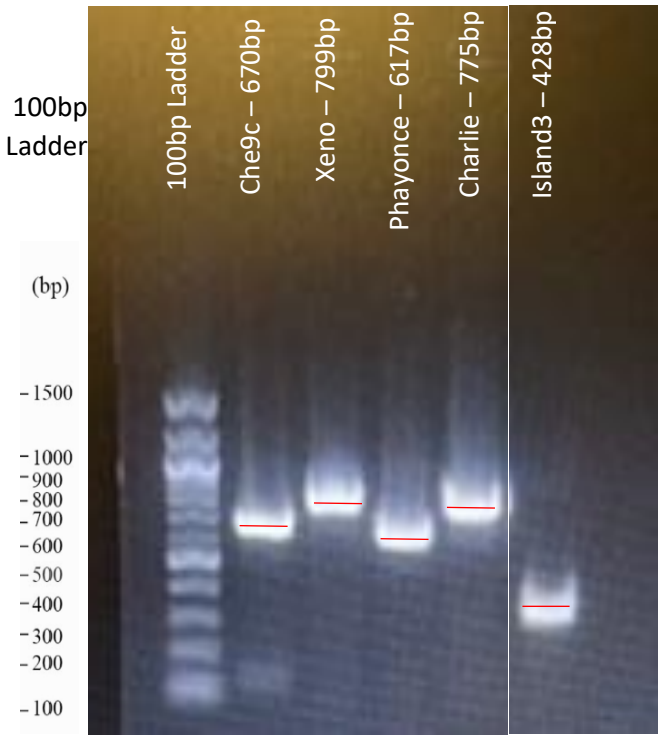


Figure 14. Ethidium Bromide 1% Agarose Gel image of viral panel verification using a 100bp DNA Ladder (5µl) in lane 1. Lane 2 is Che9C expected length 670bp. Lane 3 is Xeno with expected length 799bp. Lane 4 is Phayonce with expected length 617bp. Lane 5 is Charlie with expected length 775bp. Lane 6 is Island3 with expected length 428bp.

**Lysogen Challenges**

Two lysogen challenge experiments were conducted using a Cluster M panel (Figure 15) and Viral Challenge panel (Figure 16); each challenge experiment was performed in duplicate. During the lysogen challenges, two lysogen candidates of PegLeg and Reindeer were used to function as controls to ensure purity and consistence between the lysogens that were created *in vivo*. Both lysogens, #1 and #2 respectively for PegLeg and Reindeer, displayed equivalent results in each challenge.

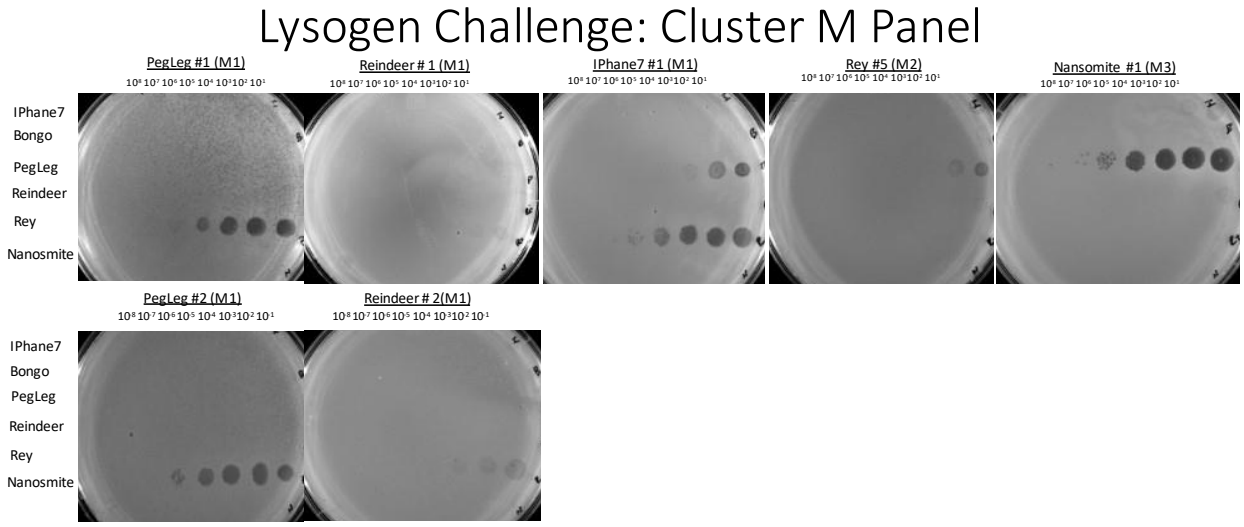



Figure 15. Lysogen Challenge results for Cluster M bacteriophages. Top agar layers containing the indicated lysogen cells were challenged with the indicated cluster M viruses – IPhane7, Bongo, PegLeg, and Reindeer (M1), Rey (M2), and Nanosmite (M3). Serial dilutions to $10^{-8}$ were performed for each phage. Incubated at 30°C for 72 hours, then imaged. These results are representative of two independent experiments that were performed in duplicate for the Cluster M Panel. The # next to each lysogen indicates the clone used.

The Cluster M Panel consisted of IPhane7, Bongo, PegLeg, Reindeer (M1), Rey (M2), and Nanosmite (M3). The cluster M lysogen challenge displayed selective inhibition between the subclusters and appeared to be phage specific. The experiment showed that all appeared to be homoimmune, except for PegLeg and Rey phage infections. Rey (M2) exhibited the ability to infect PegLeg lysogens #1 and #2, and IPhane7 lysogen with no log inhibition. Rey showed subtle infection of Reindeer #1 & #2 with an approximate $10^{-2}$ log inhibition. PegLeg (M1) exhibited the ability to infect IPhane7 lysogen with an approximate $10^{-4}$ log inhibition and Rey lysogen with a $10^{-5}$ log inhibition. PegLeg displayed the ability to infect the Nanosmite lysogens with no logs inhibited.

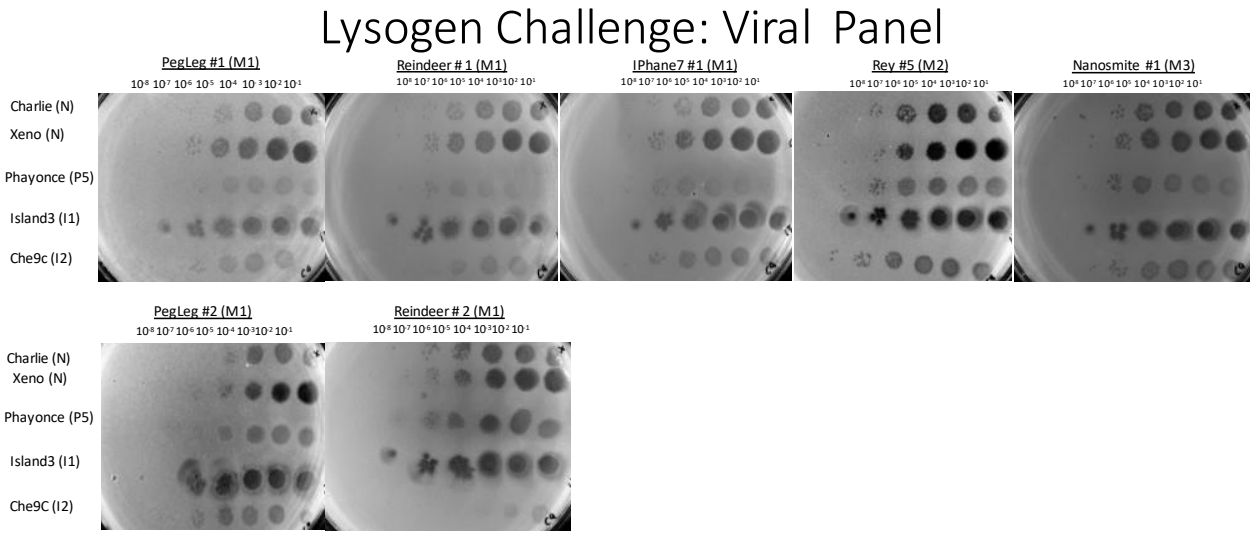



Figure 16. Lysogen Challenge results for Viral Panel bacteriophages. Top agar layers containing the indicated lysogen cells were challenged with the indicated viral panel viruses – Charlie (N), Xeno (N), Phayonce (P5), Island3 (I1), and Che9c (I2). Serial dilutions to $10^{-8}$ were performed for each phage. Incubated at 37°C for 72 hours, then imaged. These results are representative of two independent experiments that were performed in duplicate for the Viral Panel. The # next to each lysogen indicates the clone used.

No inhibition was observed from any Cluster M lysogen against the viral panel. Charlie, Xeno, Phayonce, Island3, and Che9C were all able to effectively infect each of the cluster M lysogens in Figure 16. These results indicate that no heteroimmunity is observed, and that phage infection is possible by the following cluster N, P5, I1, and I2 phages. Cluster M lysogens do not appear to confer immunity to infection from phages that are not members of Cluster M, based on this panel.

## Determining the Prophage Integration site of Bacteriophage IPhane7 in *Mycobacterium smegmatis mc²155*

Cluster M bacteriophage IPhane7's prophage integration site was determined via whole genome sequencing of an IPhane7 lysogen and the subsequent analysis of contigs using NCBI's BLASTn to locate IPhane7/*M. smegmatis* genome junctions within the lysogen.

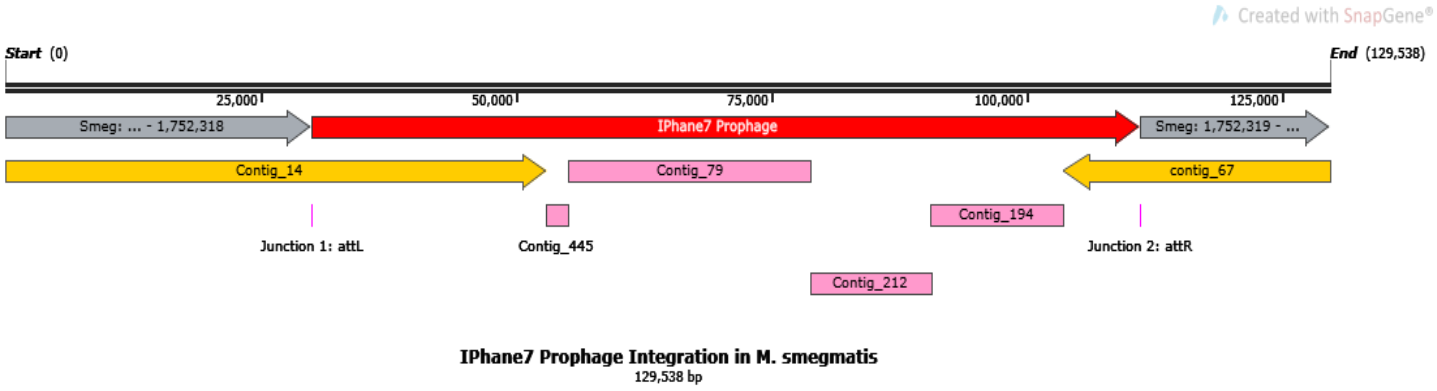



Figure 17. IPhane7 prophage integration in *Mycobacterium smegmatis mc² 155*. The location of the *attP* and *attB* sites are shown in this SnapGene map of IPhane7's prophage integration into *Mycobacterium smegmatis mc² 155*. *attL* will be referred to as Junction 1, and *attR* will be referred to as Junction 2 as displayed from left to right in the map. Arrows indicate the directionality of the genome, such as IPhane7's genome (in red) starts with 1bp at the blunt end and ends at its 81,036bp coordinates. *M. smeg*'s genome (in grey) is not complete but provides a zoomed in view of the genomic regions encompassed by the junctions. SnapGene software (www.snapgene.com)

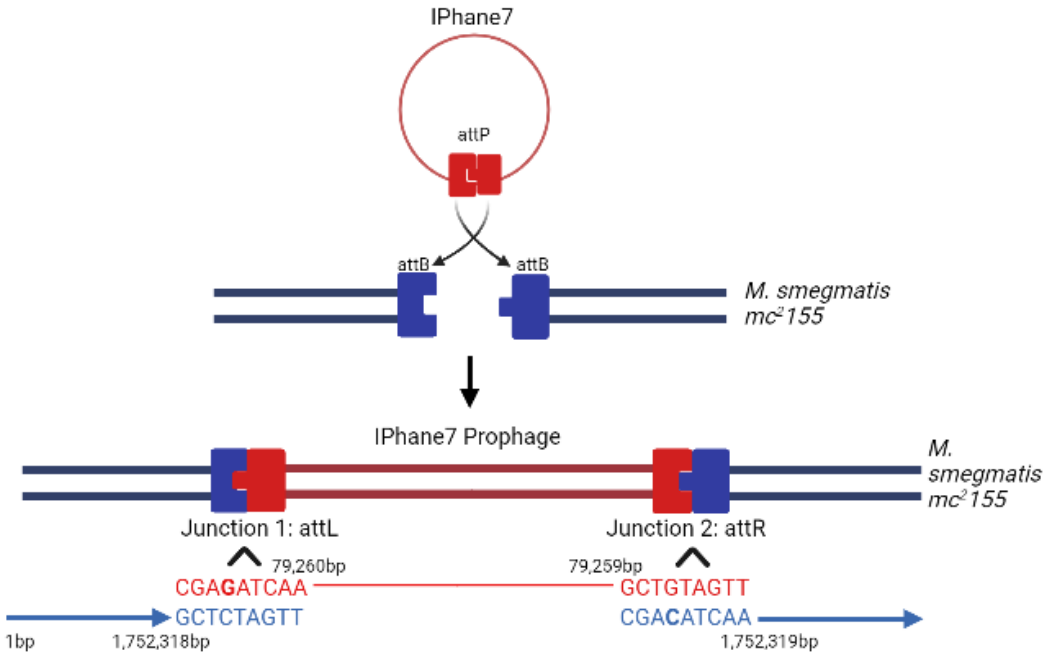



Figure 18. A figure displaying the process of IPhane7 viral integration into host cell *M. smegmatis mc²155*. *attP* and *attB* sites are located in both junctions because they are split during the process of integration. Zoomed in look at the specific nucleotide sequence of the junctions is provided to display the single nucleotide difference between the shared 9mers that appears to aide in integration directionality. Created with BioRender.com

IPhane7 lysogen was mapped to the following coordinates in *Mycobacterium smegmatis mc² 155*. Partial assembly resulted in 586 contigs. Each contig was individually investigated for presence of IPhane7's genome via NCBI's BLASTn. Seven of the 586 contigs were found to contain IPhane7 genomic sequence (contigs 14, 20, 67, 79, 194, 212, and 445). The seven contigs identified were then mapped to the respective genome they aligned to – either IPhane7 or *Mycobacterium smegmatis mc² 155*, or both due to junction locations – and a universal integration map was created as seen in Figure 17. Due to limitations in SnapGene software version used, the entire genome of *Mycobacterium smegmatis mc² 155* was unable to be included in the map; thus, only a portion of *Mycobacterium smegmatis mc² 155* where the *attB* site is shown. Contigs 79, 194, 212, and 445 mapped only to IPhane7's genome.

Contigs 14 and 67 mapped to both *M. smegmatis* and IPhane7, allowing us to determine the prophage junctions. The *attP* and *attB* sites of Junction 1 and Junction 2 are shown in Figure 18. The two areas of overlap that serve as the junctions for IPhane7 and *Mycobacterium smegmatis mc$^2$ 155* share a 9mer that differs only in one nucleotide respective to each junction. It is important to note that each Junction, 1 and 2, both contain *attP* and *attB* sites because both genomes are splitting at each Junction. However, there is a single nucleotide difference that appears to aide in directionality and integration for IPhane7's prophage.

The 9mer observed in Junction 1, CGA**G**ATCAA, was identified in IPhane7 and contig 67. The 9mer observed in Junction 2, CGA**C**ATCAA, was identified in *M. smegmatis* and contig 14. However, one can see that Junction 1 falls within contig 14, while Junction 2 is found within contig 67. This is to be expected. Junction 1in Figure 18 displays an overlap with the *attP* 9mer sequence found in contig 67 which originates from IPhane7's genome, due to the mechanism of how a prophage integrates via a loop-flip mechanism that reverses the ends of the prophage attachment sites, as observed in Figure 18. Similarly, *attB* 9mer sequence found in *M. smegmatis's* genome (contig 14) is located in Junction 2, confirming the 'flipping of the ends' that occurs during integration. Confirmation of *attP* and *attB* sites for Junction 1 and 2 was performed via PCR amplification of the junctions and subsequent Sanger sequencing.

## Discussion of Cluster M Lysogeny Analysis

### Discussion of Cluster M Lysogen Challenges

Temperate phages control lytic gene expression during lysogeny via encoded immunity systems that suppress viral genes using an encoded repressor gene, and express genes responsible for lysogeny. Once lysogeny is established by the infecting phage, the host remains susceptible to superinfection via a second round of infection by another phage [48]. A genetic spectrum of

various phages may attempt infection of the lysogen whose genomes are closely related (homotypic), moderately related (mesotypic), or unrelated (heterotypic) to the resident prophage [48]. As a result, evolutionary differences have risen in an effort to maintain lysogeny while simultaneously defending against superinfecting phages[48].

The results of the cluster M lysogen challenges experiments helped to confirm the homo-immunity and hetero-immunity properties of the cluster M lysogens. The inclusion of phages from each subcluster – M1, M2, and M3 – allowed for preliminary analysis of each subclusters ability to defend against one another. While two lysogen candidates of PegLeg (M1) and Reindeer (M1) were used throughout both experiments, each candidate was observed to have comparably identical results; this confirmed the purification and validity of the lysogens created. In this section, homo-immunity will be identified as the ability to defend against other genetically similar phage infection, such that IPhane7 (M1) may defend against a Reindeer (M1) infection, and as the ability to defend against infection by itself – such that IPhane7 may defend against infection by another IPhane7.

The cluster M lysogen panel used for the challenge specifically consisted of a broad panel of the cluster to allow for optimal results. When each lysogen was challenged with its phage form, such that bacteriophage IPhane7 was attempting to infect an IPhane7 lysogen, it was unsuccessful. The results showed that all phages are homo-immune to genetically identical phage infections – themselves – such that they may not infect a lysogen containing a prophage of itself. This observation confirms that each prophage expresses defense genes that inhibit viral infection of the same phage, thus ensuring that only one individual phage may infect and establish lysogeny in a bacterial genome for its own species. This is to be expected as lysogens are characteristically immune to superinfection of the same phage by the phenomena where the

prophage-expressed repressor will down regulate any newly introduced genomes of the same phage [21].

Differences amongst defense capabilities were observed among challenges between different phages. The results display specific homo-immunities amongst the cluster M bacteriophages. Notably, cluster M1 lysogens PegLeg and Reindeer displayed the ability to defend against all other cluster M1 and M3 members. However, cluster M1 IPhane7 lysogen was able to defend against all cluster M1 infections except for slightly inhibited infection of PegLeg. PegLeg (M1) has the capacity to slightly overcome IPhane7 prophage defense mechanisms and pursue hindered infection of the IPhane7 lysogen. When analyzing the genetic sequence similarities between IPhane7 and Pegleg, they fall into the category of mesotypic. The two phages of are relatively similar genomes but containing subtle nucleotide dissimilarities that may result in the observed incomplete superinfection results we see. Complete immunity of fellow cluster M1 and M3 phages is observed in PegLeg, Reindeer, and IPhane7 lysogens, except for phage PegLeg having the ability to infect the IPhane7 lysogen with approximately $10^{-4}$ log inhibition.

Interestingly, no cluster M1 was able to defend against infection by phage Rey (M2). Only Rey (M2) and Nanosmite (M3) lysogens were able to defend against Rey infections. When comparing notable areas of interest in relation to prophage-mediated defense genes under investigation expressed during lysogeny in cluster M phages, it is observed that while PegLeg, Rey, and Nanosmite have conserved areas of sequence around gene 2 in their genomes. Rey and Nanosmite possess dissimilar sequences for the genome around gene 1 of these phages, as seen in Figure 19; however, this does not appear to affect the ability to defend against superinfection of each other. Nanosmite lysogen is unable to defend against PegLeg phage infection, which may

be a result of their genomic differences, in relation to gene 1 and other regions of their genomes, that do not genetically interact, resulting in symmetric infection phenotypes [48]. PegLeg's ability to fully overcome Nanosmite prophage defense genes could correlate with the possession of different gene 1 sequences. Rey lysogen exhibits slight inhibition of PegLeg infection with an approximate $10^{-2}$ log inhibition, displaying that the two contain genetically related but distinct genetic elements leading to incomplete immunity phenotypes. It is also of interest to note the difference in possession of the minor tail protein (IPhane7 gene *30* for reference) amongst these four phages. Of the minor tail proteins, IPhane7 and PegLeg encode the same minor tail protein, gene *30* (pham 68831), while Rey and Nanosmite encode different minor tail proteins, gene *34* (pham 41088). Differences amongst these minor tail proteins may also play a role in the inhibitions and selectivity observed amongst the Cluster M's of interest.

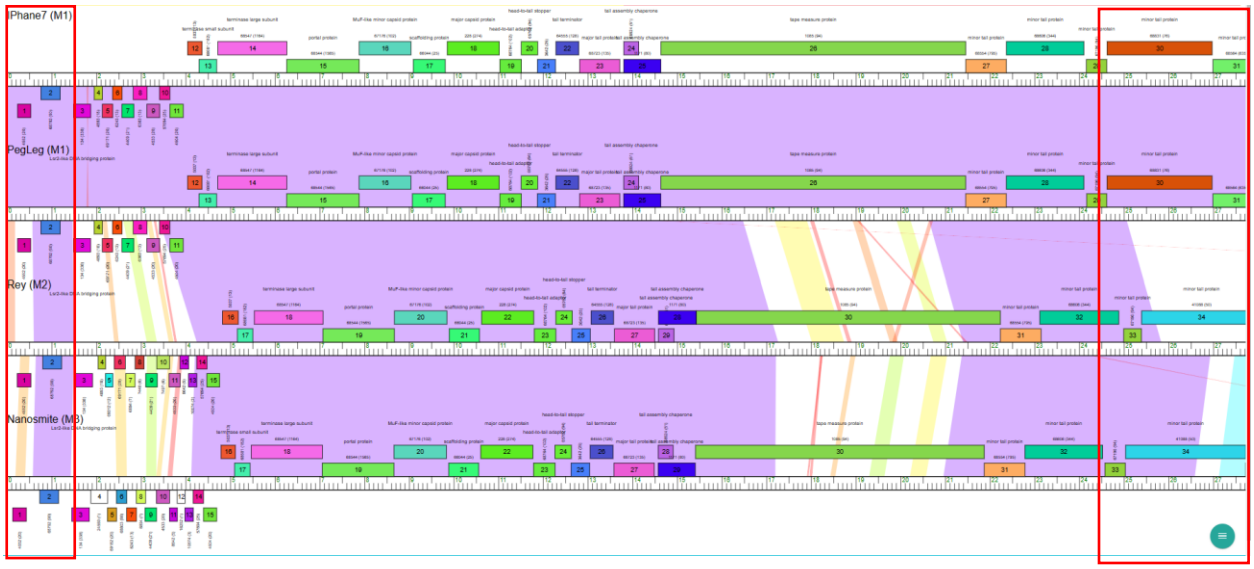



Figure 19. A snapshot view of the beginning of the genome of phages IPhane7, PegLeg, Rey, and Nanosmite displaying nucleotide sequence similarities and differences. Areas colored in purple signify high percentages of shared sequence similarities, while descending colors show areas of somewhat similarities and white signifies no sequence similarity. Image taken from phamerator.org. Red boxes are placed around genes of interest: genes *1*, *2*, *30*, and *34*.

While the cluster M lysogen challenge against a panel of cluster M bacteriophages displayed the specificity of homo-immune capabilities of the lysogens, experiments against a viral panel containing phages from genetically different clusters displayed different results. As seen in Figure 16, no inhibition was observed from any of the cluster M lysogens. All viral panel members, Charlie (N), Xeno (N), Phayonce (P5), Island3 (I1), and Che9c (I2), were able to overcome any prophage-mediated defense genes expressed by cluster M lysogens. The results of these experiments aide in the conclusion that cluster M prophage defense mechanisms are seemingly cluster M phage specific and may not confer any hetero-immunity to phages outside of the cluster M members. However, absolute conclusion of this is unable to drawn at this time based on the size of our viral panel. To truly assess the ability of cluster M lysogens to defend against members of other cluster, one would want to perform this challenge amongst a broad panel of phages from a multitude of clusters. While we challenged select cluster M lysogens against phages from clusters N, P5, I1, and I2, there are more than 32+ other genetic clusters that should be observed before drawing a definitive conclusion.

**Discussion of IPhane7 Prophage Integration**

Cluster M bacteriophage IPhane7's prophage integration site was determined via whole genome sequencing of an IPhane7 lysogen (#1) that was previously created in the Gainey laboratory. Subsequent analysis of assembled contigs using NCBI's BLASTn allowed for location of the *attP* and *attB* sites. IPhane7 has been identified to possess a serine integrase gene, gene 129 (Pham 54044), in the right hand of its genome. As expected with Int-S systems, the junction sites were observed to contain a short, unique 9bp sequence, the common core, that served as the overlapping sequencing for the two genomes to integrate. The two junctions contained an almost identical 9mer, only differing in 1 nucleotide. As observed in Figure 18, the

9mer CGA**G**ATCAA in Junction 1 aligns to IPhane7 genomic sequence, while Junction 2's 9mer CGA**C**ATCAA aligns to *M. smegmatis* genomic sequence. IPhane7's genome splits at 79,260bp, found roughly 112bp after the last gene (156). Of relative importance, cluster M-Like prophages have been observed in human *M. abscessus* infection isolates that are resistant to antibiotic treatment, identified as MabI prophages [16]. Interestingly, cluster MabI and MabJ that have been identified in the *M. abscessus* isolates share similarity to cluster M and A mycobacteriophages, respectively [16,48,62] (see Figure 6 for genomic similarities of MabI to Cluster M). The MabI prophages were identified to interrupt an ORF within genes MAB_3230 at the attB-9 site and MAB_3265 at the attB-17 site. Similarly, the IPhane7 prophage has been identified to integrate within gene LJ00_08275, an exodeoxyribonuclease III, as seen in Figure 20. The effects upon interruption of the exodeoxyribonuclease III have not been investigated, such as rendering the protein ineffective or enabling the transcription of a different protein based on the shortened gene.

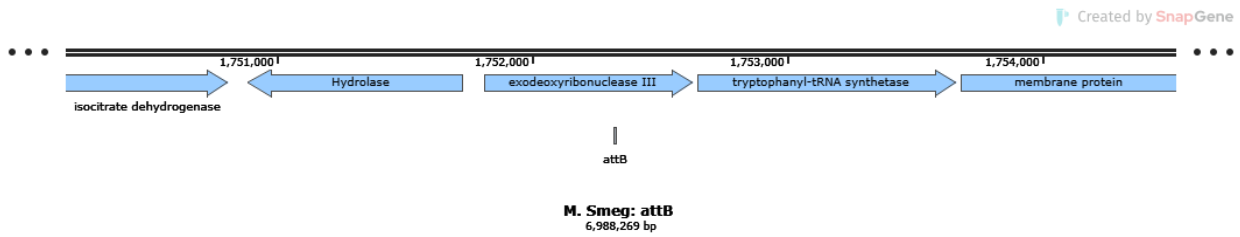



Figure 20. *M. smegmatis attB* site located within ORF encoding for an exodeoxyribonuclease III. *M. smegmatis* genome is signified by double-stranded black lines, and genes denoted in blue.

Bioinformatic analysis shows that the *attB* site 9mer, CGA**C**ATCAA, appears at 187 other sites in *M. smegmatis mc*$^2$*155*, and the *attP* site 9mer, CGA**G**ATCAA, appears in 3 sites in IPhane7. When the IPhane7 *attP* site is extended to a 12mer, *GAT*CGA**G**ATCAA, to include the three preceding nucleotides italicized, there is only one site at which this sequence occurs – our

*attP*. When the *M. smegmatis attB* site is extended to include a 12 mer, *GAA*CGA**C**ATCAAGAA, to include the three preceding nucleotides italicized, there is only two sites at which this sequence occurs. Consequently, when including the three preceding and three following nucleotides italicized, *GAA*CGA**C**ATCAA*GAA*, there is only one site at which this 15mer sequence occurs in the *M. smegmatis* genome – our *attB*. It is unknown at this time of the possible roles that the nucleotides immediately surrounding the identified common core sequence play in site specific integration, but there is suspicion of their possible roles in aiding of the site-specific recombination efforts.

A recently proposed model for serine integrase-mediated recombination displays that during viral integration, an integrase (Int) dimer binds to specific 'attachment site sequences' in the phage (*attP*) and host (*attB*) DNA [70]. The Int-*attP* and Int-*attB* complexes form tetrameric intermediates, whereby all four substrate DNA strands are cleaved and the Int subunits covalently bond to their corresponding DNA half-sites [26]. By rotating the two Int-DNA subunits 180° in respect to the other two subunits, strand exchange occurs and the re-aligned DNA half-sites are ligated to generate new attachment sites referred to as *attL* and *attR* [45] as displayed in Figure 21.

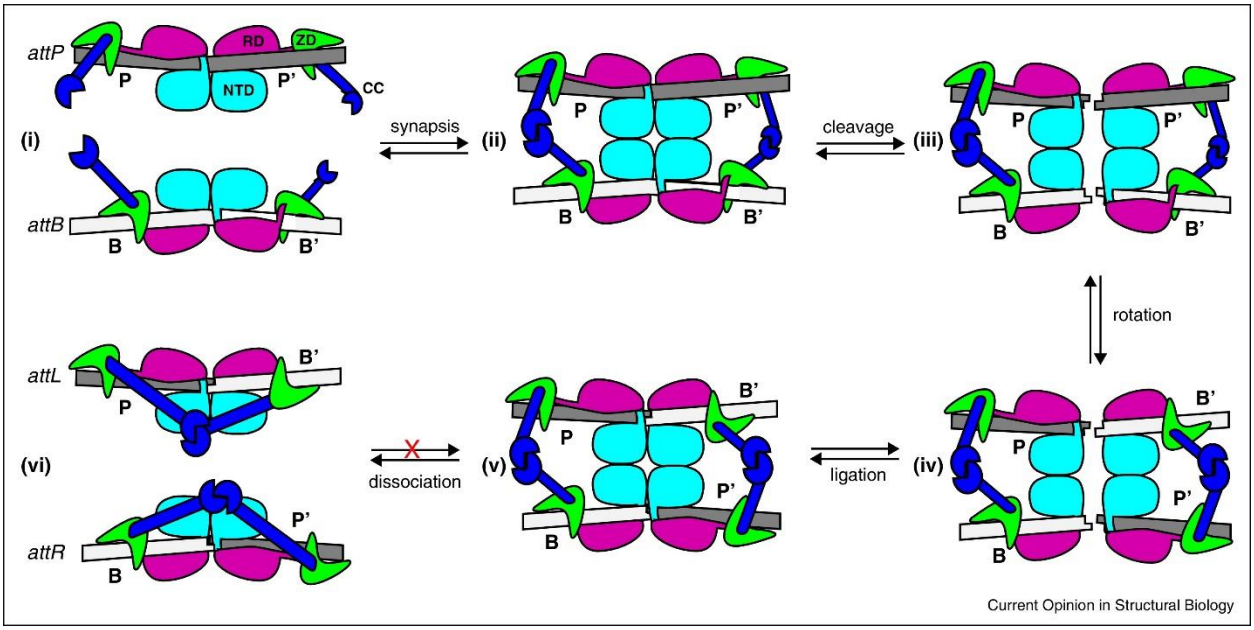



Figure 21. Recently proposed model of the integration reaction catalyzed by serine integrases. Firstly, integrase (Int) dimers bind to specific sequences in the phage (*attP*) and host (*attB*) DNA. The Int–*attP* and Int–*attB* complexes are conformationally distinct due to different positioning of a zinc ribbon domain (ZD). Secondly, Int–*attP* and Int–*attB* associate to form a synaptic complex that is stabilized by interactions between coiled-coil (CC) motifs. Thirdly, the Int subunits cleave all four DNA strands at the central dinucleotide, forming 5′-phosphoserine linkages between integrase subunits and DNA half-sites (not illustrated) and generating 3′-dinucleotide overhangs. Fourthly, The P′ and B′-linked subunits can exchange places by rotating 180° about a horizontal axis relative to the P and B-linked subunits. Fifthly, base-pairing between the central dinucleotides promotes ligation of the DNA strands, resulting in formation of two new attachment sites, *attL* and *attR*. Finally, the unique arrangement of ZDs in the *attL* and *attR* sites allows the CC motifs to form intra-molecular interactions that prevent the reaction from running efficiently in the reverse direction. NTD: N-terminal catalytic domain (cyan); RD: recombinase domain (magenta); ZD: zinc ribbon domain (green); CC: coiled-coil motif (blue). *Note figure and description adapted from* [70].

Studies investigating the structure of the C-terminal domain (CTD) bound to its *attP* half-site in the *Listeria innocua* prophage integrase (LI integrase) revealed how serine integrases bind to *attB* sites and how the *attP* and *attB* binding modes differ [71]. The CTD, comprised of three structural domains: a recombinase domain (RD), a zinc-ribbon domain (ZD), and an extended coiled-coil motif (CC) which is embedded in the ZD, contains a majority of the enzyme's DNA-binding functionality and is therefore primarily responsible for the attachment-site specificity [70].

Alignments of *attP* versus *attB* sequences for a number of serine integrase systems reveal that the innermost 13-bp do share considerable similarity, however, overall similarity of the short 40-50bp attachment site is not as significant, which allows for simple manipulations of substrates [70,71]. Sites share only limited sequence similarity in most systems, yet the same serine recombinase dimer binds to both *attP* and *attB* sites with high affinity, as seen in the bacteriophage φC31 and Bxb1 integrase systems [71]. It is suspected that the surrounding sequences to IPhane7's *attP* and *attB* common core sequences play a significant role in specific site recognition for the respective Int dimers at each site.

The differences between the integrase– *attP* and integrase– *attB* complexes are key to understanding how the serine recombinase enzymes function and is an important goal within the field. As serine integrase systems are further examined, a deeper understanding may be achieved on how these systems function in various phages. A deeper investigation on the serine integrase system particular to IPhane7 would shed light on the intricacies of integration and aide in discovering why might IPhane7 integrate at the identified site in this study, versus the 187 other *attB* common core sites or 3 other *attP* common core sites observed.

Bacteriophage therapy for antibiotic resistant *M. abscessus* infections has become a beneficial avenue, and cluster M-like prophages therefore have become of peak interest. However, little is known about how cluster M phages establish lysogeny and what prophage-mediated defense genes they may express. Thus, the important of the identification of cluster M IPhane7's integration sites become of elevated importance, especially with the inability of bioinformatically identifying serine integrase systems. The unusual properties of serine integrases including recombination directionality manipulation via simplistic site sequence requirements are leading to development as efficient and versatile tools with applications in

experimental biology, biotechnology, and gene therapy [81]. Further investigations focused on integration sites involving other cluster M members would be of interest to field to allow for a greater understanding of Int-S systems as well as cluster M integration site characteristics.

# CHAPTER THREE: RESULTS OF CLUSTER M PROPHAGE-MEDIATED DEFENSE GENES ANALYSIS

## Investigation of Defense Escape Mutants That May Overcome IPhane7's gp1 and gp2 Defense Mechanisms

Assembly of IPhane7 gp1/gp2 Defense Escape Mutants (DEMS), previously sequenced in the Gainey Laboratory, led to the discovery of insertions, deletions, point mutations, and truncations in Samples 7-12, with Sample 1 as the control IPhane7 genome. The profile of mutations, in Table 9, confer defense escape phenotypes that provide compelling evidence for specific genes that are required for the IPhane7 gp1/2 system to inhibit IPhane7 superinfection.

Table 9. A table containing the Mutations observed in the IPhane7 DEMs Samples. The table contains color coordinates that coincide with the colored asterisks in Figure 22.

| Color | Sample #: | Mutations Observed: |
|---|---|---|
| **DEM 1** | Sample 1 | 1.1: 25981: G/A: Point mutation: Changes a Valine to a Methionine<br>1.2: 37204-06: Query C --/ Sub GGG: Proline (CCT) to insert Glycine (GGG) (intergenic)<br>1.3: 37208: Query T/ Sub A<br>1.4: 37211: Query -/ Sub T<br>1.5: 37213-14: Query --/ GA |
| **DEM 2** | Sample 2 | 2.1: 26018: Query T/ Sub C; Amino Acid change: GTG calls Valine, GCG calls Alanine<br>2.2: Insertion at 75225/75226: Q --/ Sub CG; Truncation |
| **DEM 3** | Sample 3 | 3.1: 26018: Query T/ Sub C; Amino Acid change: GTG calls Valine, GCG calls Alanine<br>3.2: Insertion at 75198/75199: Q --/ Sub CT; Truncation |
| **DEM 4** | Sample 4 | 4.1: 26018: Query T/ Sub C; Amino Acid change: GTG calls Valine, GCG calls Alanine<br>4.2: 75928: Query G/ Sub T |
| **DEM 5** | Sample 5 | 5.1: Mutation at 26018: Query T/ Sub C; Amino Acid change: GTG calls Valine, GCG calls Alanine<br>5.2: Mutation at 75056: Query C/ Sub A; Amino acid change: GAG calls E, TAG: Stop |
| **DEM 6** | Sample 6 | 6.1: Mutation at 26018: Query T/ Sub C –; Amino Acid change: GTG calls Valine, GCG calls Alanine<br>6.2: Mutation at 37195: Query C/ Sub T; INTERGENIC Right before CR2-3 37219-37230 (-) |

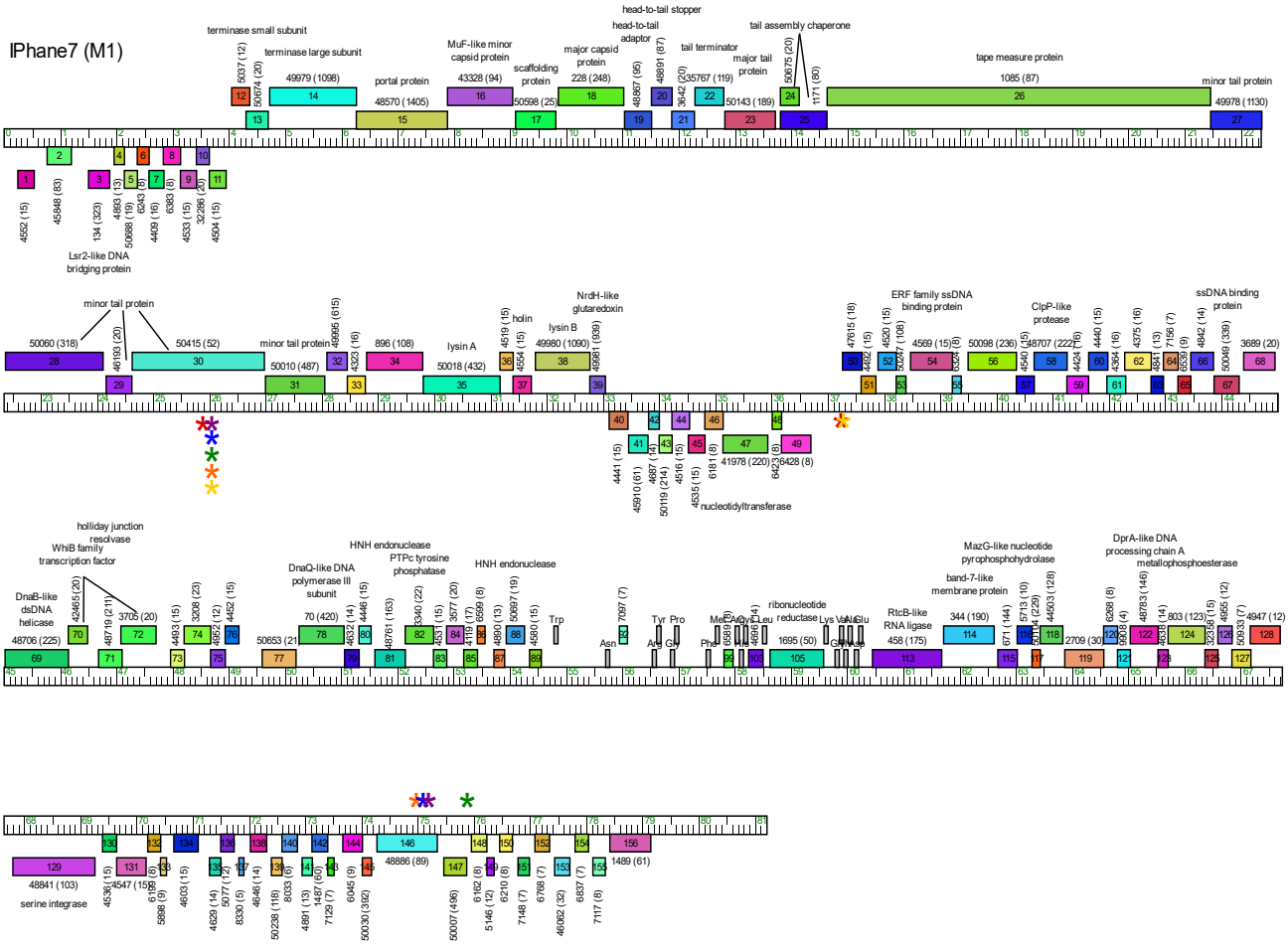


Figure 22. IPhane7 genome map indicating locations of IPhane7 DEMs. Colored asterisks denote the location of observed mutations and correlate with the color key found in Table 9.

Upon sequencing of the IPhane7 DEMs, a plethora of mutations were observed and characterized in Table 9 and identified on a genomic map of IPhane7 in Figure 22. Notably, sample 1 contained the highest number of mutations (5), while samples 2-6 contained two. While substitutions and insertions 1.3-1.5 did not result in amino acids changes, mutations 1.1 and 1.2 did; however, mutation 1.2 falls within a non-coding region and does not concern gene product structure. Mutation 1.1 is a point mutation results in a change from a valine to methionine and falls within gene 30 (pham 52895 as of 11/17/22) that encodes a minor tail protein. Interestingly, Samples 2-6 have identical mutations at coordinate 26018bp, which falls within gene 30 (pham 52895 as of 11/17/22) as well. Samples 2.2, 3.2, 4.2, and 5.2 observe their second mutations within the same region of the right arm around coordinate 75000bp. Sample 6.2 mutation is noted to match mutation locations of 1.3-1.5.

## Investigation into IPhane7's gp1 Function

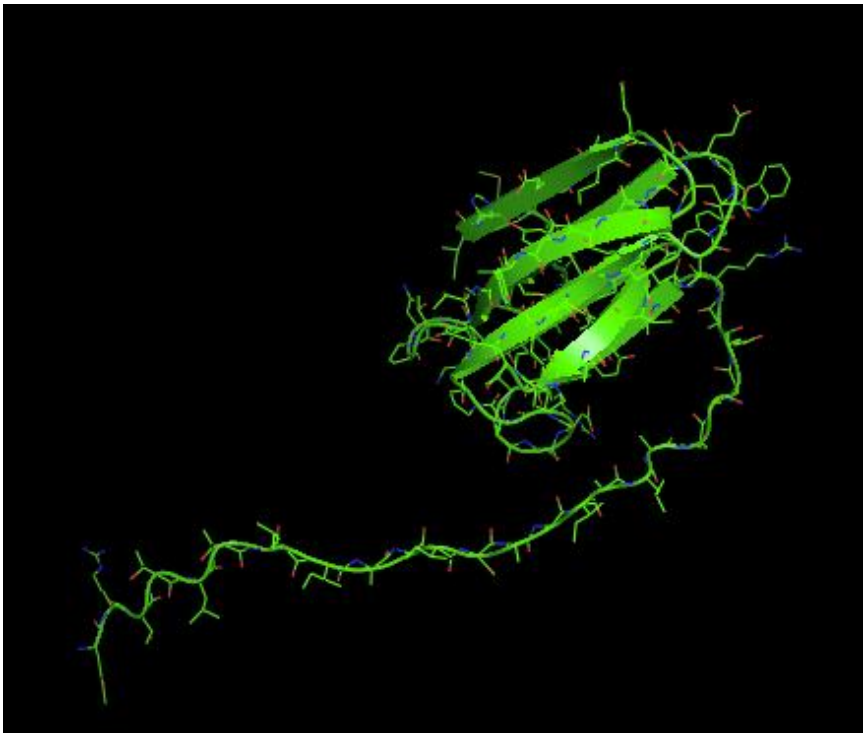


Figure 23. Predicted tertiary structure generated by AlphaFold of IPhane7's gp1 protein. A conglomeration of beta sheets with a predicted secreted signal sequence represented by the long singular strand of amino acids.

Only 97 aa in length, IPhane7's gp1 was predicted to poses a Sec-signal peptide with a cleavage between aa 24 and 25. A TOPCONS report was generated for IPhane7 gp1 protein sequence. The report displays a unanimous opinion of a signal peptide at the beginning of the gene by all the programs: TOPCONS, OCTOPUS, Philius, PolyPhobius, SPOCTOPUS, and SCAMPI who was similar with a slightly different generation. Based on the observed secreted signal peptide in cluster M gp1's, we hypothesized a link to gp1 as a defense mechanism to inhibit infection by other phages. To investigate the ability of gp1 to defend against cluster M phages, we conducted an experiment to study the effects of overexpression and endogenous expression of the gene. We elected to conduct a second experiment involving non-cluster M phages to determine the defense ability of cluster M gp1's against phages outside of the cluster.

## Characterization of the Defense Profile of Cluster M Bacteriophages' gp1

Based on protein sequence alignments of all Cluster M phages, the following phages were chosen to evaluate gp1 effects: IPhane7 (M1), Reindeer (M1), Rey (M2), Nanosmite (M3). Upon phage selection to test their gp1's ability to defend against infection, we proceeded to clone in IPhane7 (M1), Reindeer (M1), Rey (M2), and Nanosmite (M3) gp1's into two different expression level vectors: pExTra (overexpression) and pMH94 (endogenous). After successfully cloning each gp1, we then conducted experimental challenges for pExTra clones in triplicate, and pMH94 clones in duplicate against a panel of cluster M bacteriophages. The same experiments were conducted against a panel of non-cluster M 'Viral' panel bacteriophages to analyze the ability of gp1 to defend against genomically different phages.

### Protein Sequence Alignment of Cluster M gp1

Cluster M1, M2, and M3 protein sequences were analyzed for similarities to determine which phages to use for the study.

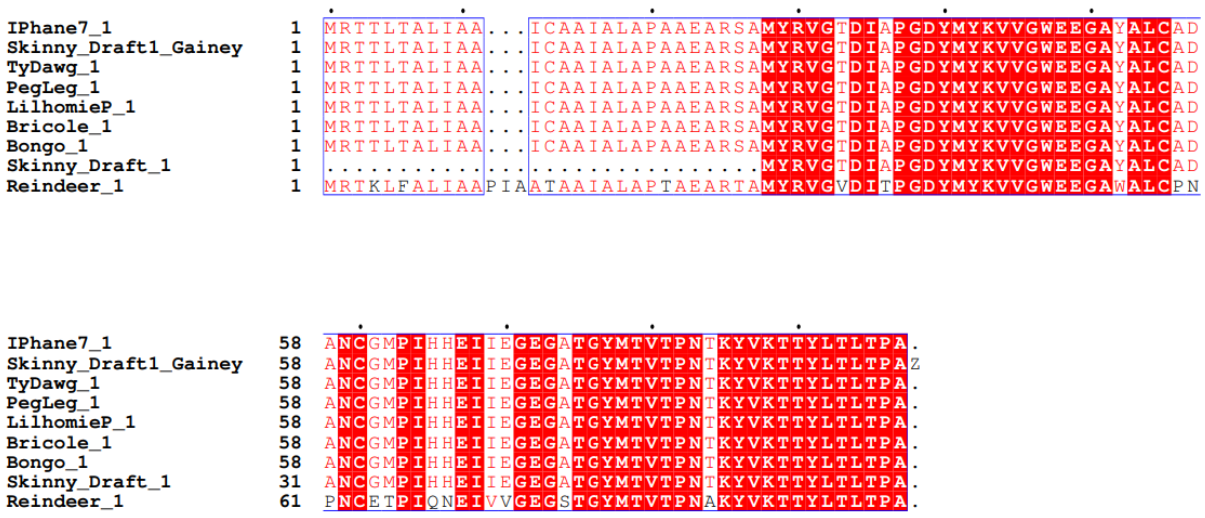



Figure 24. ESprit3 alignment of M1 phages gp1 protein sequence. All M1 phages have identical protein sequences excluding Reindeer. [69]

To determine the similarity of protein sequence of gp1 in the M1's, ClustalW and ESprit3 were used to align the protein sequences for comparison. As seen in Figure 24, all cluster M1's share identical protein sequences, except for Reindeer. Therefore, IPhane7 and Reindeer were selected for examining gp1 effects. All M1 phages, excluding Reindeer, and considered to display the same phenotype as IPhane7 since their protein sequences are identical to IPhane7's gp1 protein sequence.

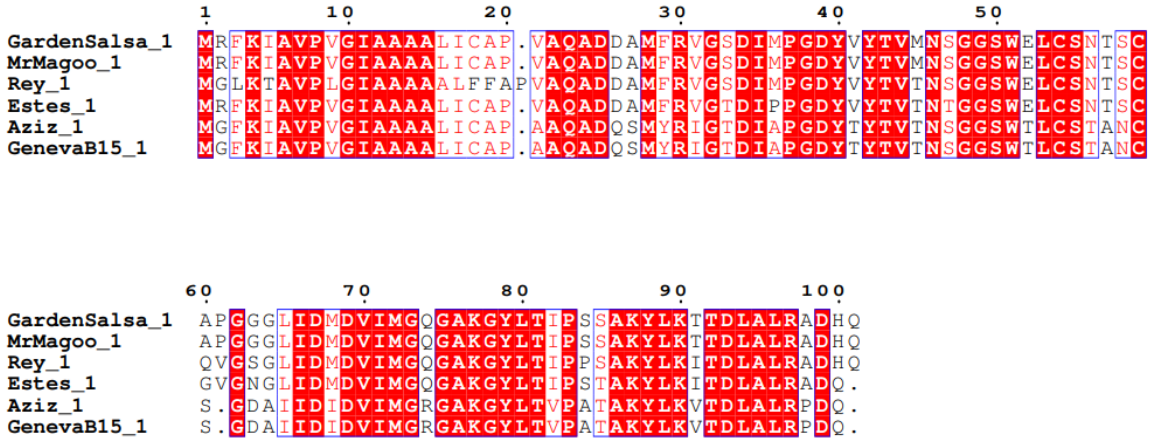



Figure 25. ESprit3 alignment of M2 gp1 protein sequences. GardenSalsa and MrMagoo have identical gp1 protein sequences. Aziz and GenevaB15 have identical gp1 protein sequences. Rey and Estes share no identical sequences and are stand alone. [69]

To determine the similarity of protein sequence of gp1 in the M2's, ClustalW and ESprit3 were used to align the protein sequences for comparison. As seen in Figure 25, the M2's do not share as much similarity as the M1's. GardenSalsa and MrMagoo share the same protein sequence. Rey and Estes are unique and share no identical protein sequences to any other M2. Aziz and GenevaB15 share identical protein sequences. At aa 5, Rey has a polar (P) Threonine, while all others have a nonpolar (NP) isoleucine. At aa 18, Rey contains a NP phenylalanine (F) while the others have a NP Cystine. At aa 19, Rey has a NP Alanine while all other hold a NP Proline. Upon analysis of the alignment, of the locations where Rey, or other M2's, do not share

similar aa's, there is an observed change of nonpolar or polar amino acids. These differences could impact protein structure and function. Based on accessibility and ability to grow *in vivo*, Rey was chosen as the only M2 for experimental investigation.

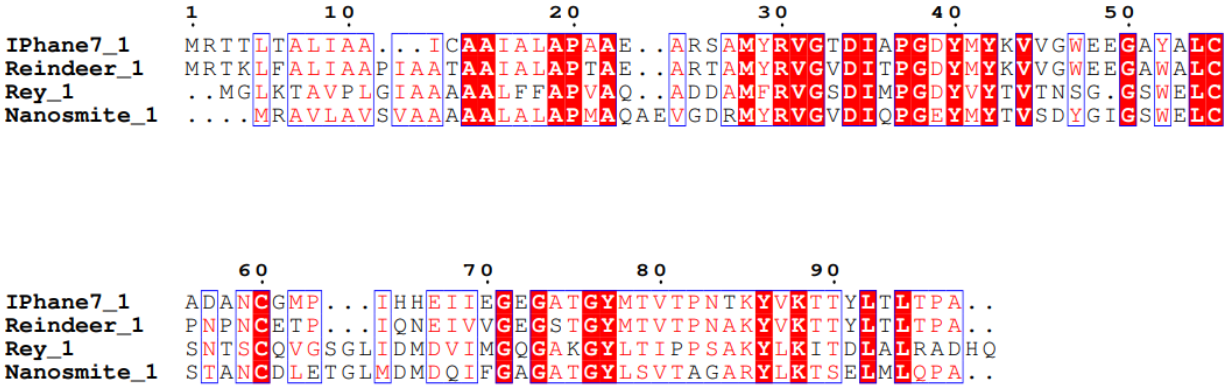



Figure 26. ESprit3 protein sequence alignment of selected Cluster M phages for gp1 experiments that possess different amino acids sequences for gp1. ESprit3 generates an alignment based on an algorithm that it deems acceptable and thus in this alignment, the first nucleotides are not aligned. This shift is not observed in Figures 24 and 25, and the first nucleotides are aligned. The first nucleotides should align here one-to-one, however, ESprit3 chose to align them slightly differently. [69]

An alignment of all Cluster M phages, Figure 26, displays the unique protein sequences between the subclusters and confirms the importance and relevance of including members from each of the subclusters for holistic analysis of gp1's phenotype. As Nanosmite is the only M3 identified, it was chosen to be included for investigation.

**High Expression Vector pExTra Containing Gene(s) of Interest**

The high expression vector used for gp1 testing, pExTra, was selected to test the effects of overexpression of gp1 on a Cluster M panel and a Viral panel. Unique features to the pExTra plasmid are the inclusion of origin of replication sites for *E. coli* and *M. smegmatis*, an mCherry reporter for gene expression validation, a kanamycin resistance gene for bacterial selection via resistance, a Tet promoter and Tet repressor to control gene expression. pExTra vector design

allows for transcriptional and translational control over a gene based on the use of a transcriptional inducer, anhydrotetracycline (aTc), and thus does not require the inclusion of the preceding gp1 intergenic region to begin transcription. Levels of gene expression were control by the amount of aTc, which allowed us to investigate how various levels of expression affected the phenotypic results.

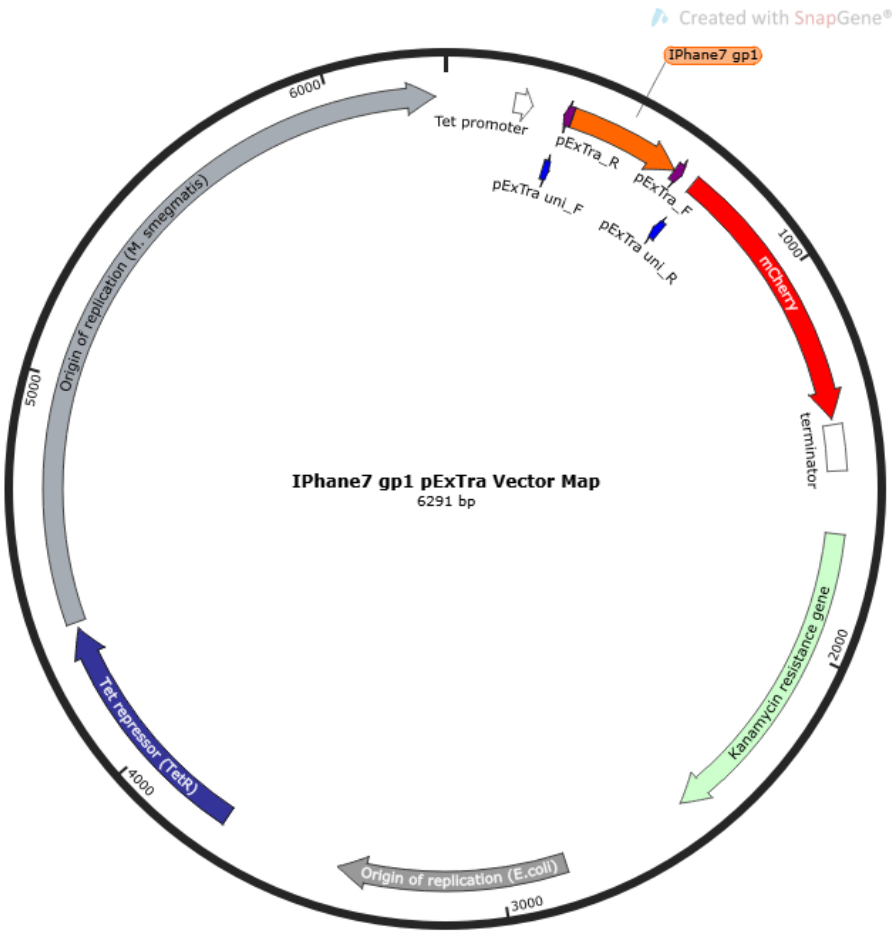



Figure 27. pExTra overexpression vector map including IPhane7 gp1. Map generated in SnapGene. IPhane7 gp1 is denoted in orange. SnapGene software (www.snapgene.com)

All bacteriophages selected for cloning – IPhane7, Reindeer, Rey, and Nanosmite – were cloned into the pExTra vector in the same location as shown in Figure 27 with IPhane7 gp1.

A separate experiment was conducted to investigate the importance of the signal peptide in IPhane7's gp1 protein function. IPhane7 gp1 protein sequence was truncated between aa 24 and 25, where the Sec-signal sequence was identified. To investigate if the same phenotype of IPhane7 would be displayed by truncating the signal sequence in the gene, IPhane7 TSS gp1 was cloned into pExTra in the same location as shown in Figure 27.

**Endogenous Expression Vector pMH94 Containing Gene(s) of Interest**

Endogenous expression vector pMH94 was used to evaluate the effects of a natural level of expression of gp1 and gp1/gp2 on a panel of Cluster M and Viral Panel challenge. The intergenic region that proceeds gp1 was included in the cloning process. Due to vector design, pMH94 will activate transcription of gp1 via the inclusion of the preceding intergenic region and does not utilize an inducer, like pExTra does.

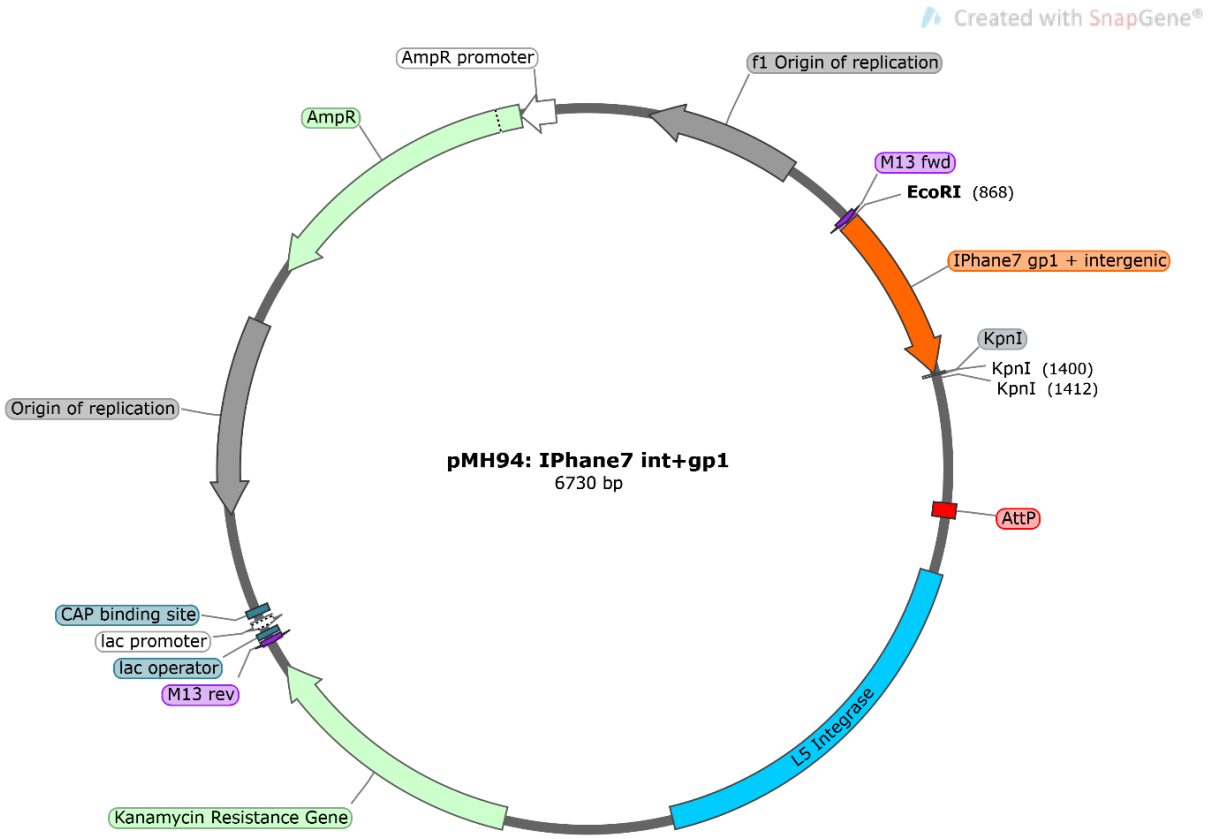


Figure 28. pMH94 endogenous expression vector map of IPhane7 intergenic (int) and gp1 created in SnapGene. IPhane7 int+gp1 is denoted in orange. Restriction enzymes EcoRI and KpnI were used to ensure proper directionality upon ligation into the vector. A kanamycin resistance gene is included to allow for bacterial selection via antibiotic resistance during experiments. SnapGene software (www.snapgene.com)

To investigate the effects that endogenous levels of gp1 have on the defense profiles of phages during the viral challenges, the gene of interest was cloned into pMH94, Figure 28, for all selected phages – IPhane7 (M1), Reindeer (M1), Rey (M2), and Nanosmite (M3) – via the use of restriction enzymes EcoRI and KpnI to allow for proper orientation in the vector during traditional cloning practices. Two experiments were conducted with the pMH94 gp1 clones; one aimed at investigating the defense abilities that endogenous levels of express would have on fellow cluster M members, while the other aimed to study the ability of gp1 to defend against a non-cluster M 'Viral' panel.

An additional round of experiments was conducted to study the effects that gp1 and gp2 have together when challenged against a Cluster M panel and a Viral Panel. Previous investigations in the Gainey laboratory that involved gp1/2 expressed together displayed a strong phenotype in phage inhibition. Synergistic traits are shown when the two genes are expressed together and appeared to confirm great inhibition when challenged against other phages. To further investigate the correlative effects of gp1/2, two experiments were conducted: a Cluster M Panel challenge to study how expression of gp1/2 together affect Cluster M infections, and a Viral panel challenge to observe the ability of gp1/2 together to inhibit infection for phages outside of cluster M.

**Viral Challenge Results**

**pExTra truncated signal sequence in IPhane7's gp1 results.**

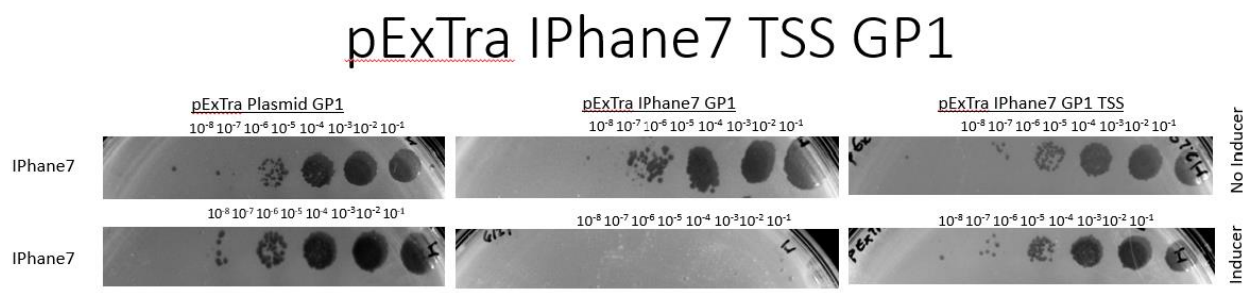



Figure 29. pExTra IPHane7 TSS gp1 Viral Panel challenge results. pExTra vector clone was incubated at 37℃ on a shaker for 48 hours in 20% tween and 7H9Neat mixture, and then incubated on a shaker with no 20% tween in a 7H9 mixture that either contained 10ng/mL aTc or no aTc for 48 hours. Upon completion of the 48 hours, the mixtures that contained 10ng/mL aTc were mixed with TA containing 10ng/mL aTc and plated on 7H9+Kan plates and allowed to gel. Similarly, the mixtures that did not contain any inducer were mixed with TA and plated on 7H9+Kan plates. The plates were spotted with serial dilutions of Viral Panel member phages and incubated at 37℃ for 72 hours and then imaged. *Note, full plate image in appendix.*

The pExTra IPhane7 TSS gp1 challenge, shown in Figure 29, displayed that by truncating the gene between aa 24 and 25, the gene lost its ability to inhibit phage infection against IPhane7. pExTra IPhane7 gp1, that contains the complete gene sequence, plates display that in the presence of the inducer total inhibition of IPhane7 infection was achieved. Contrastingly, pExTra IPhane7 TSS gp1 showed no inhibition of phage infection. These test results lead to the conclusion that the signal sequence coded for in IPhane7's gp1 plays a vital role in the gene's defense mechanism.

**pExTra gp1 results.**

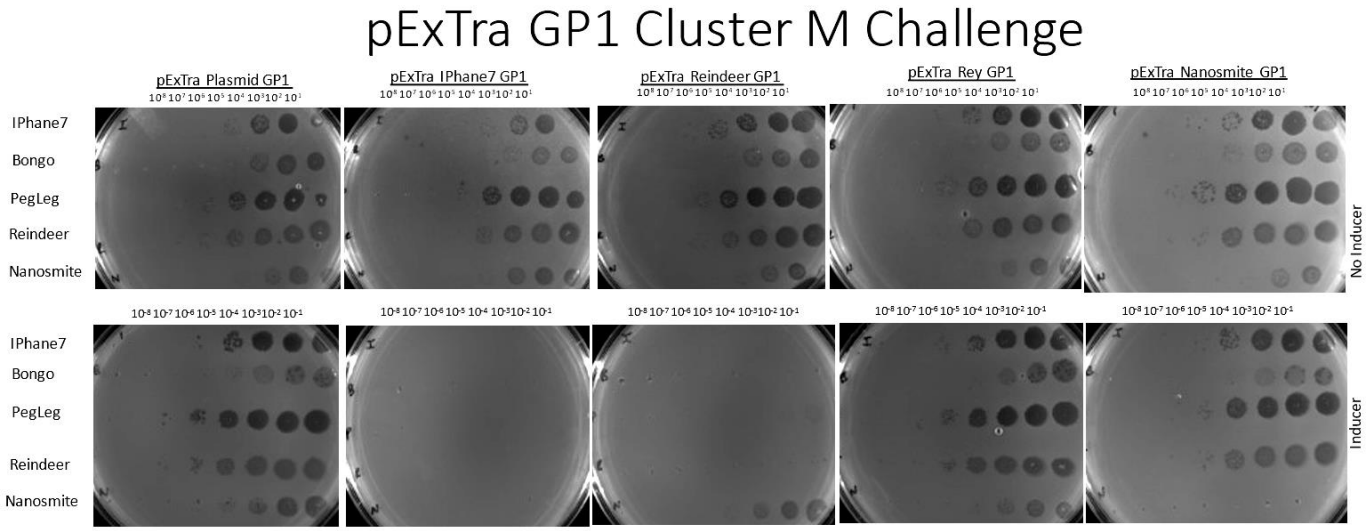



Figure 30. pExTra gp1 Cluster M challenge results. pExTra vector clones were incubated at 37℃ on a shaker for 48 hours in 20% tween and 7H9Neat mixture, and then incubated on a shaker with no 20% tween in a 7H9 mixture that either contained 10ng/mL aTc or no aTc for 48 hours. Upon completion of the 48 hours, the mixtures that contained 10ng/mL aTc were mixed with TA containing 10ng/mL aTc and plated on 7H9+Kan plates and allowed to gel. Similarly, the mixtures that did not contain any inducer were mixed with TA and plated on 7H9+Kan plates. The plates were spotted with serial dilutions of Cluster M member phages and incubated at 37℃ for 72 hours and then imaged. This figure is representative of four experiments conducted.

The pExTra gp1 results in Figure 30 showed phage specific inhibition using the lowest concentration of aTc (10ng/mL). Figure 30 displays how in the presence of the aTc inducer, IPhane7 gp1 completely inhibits infection of the panel of Cluster M phages. Reindeer gp1 is almost identical except possible slight inhibition of Nanosmite to $10^{-3}$ when compared to the Plasmid gp1 control. No inducer challenges were included for controls for all pExTra gp1 clones. pExTra Rey gp1 does not seem to display any effect on infection. pExTra Nanosmite gp1 only fully inhibits Nanosmite, while other phages are unaffected. To investigate whether increased

levels of aTc would affect the defense profiles, the same test in Figure 30 was conducted with 100ng/ml aTc during the no tween shake stages and when plated. The results of the 100ng/mL test were identical to Figure 30 (data not shown).

During this experiment four rounds of challenges were performed on the Cluster M Panel. Test 1 and Test 2 included bacteriophage Rey in the panel, where it displayed no inhibition to gp1 from any of the phages. Bacteriophage Rey was excluded from the Cluster M panel challenge experiment in Figure 30 due to experimental conditions required. Rey grows at 30℃, while all other cluster M phages grow at 37℃. The pExTra system appears to work best at 37℃ as displayed by favored growth conditions of the cells observed. As well, with Rey showing no significant phenotype when tested in duplicate, we elected to retire Rey from further experimentation due to resource conservation based on required laboratory growth conditions.

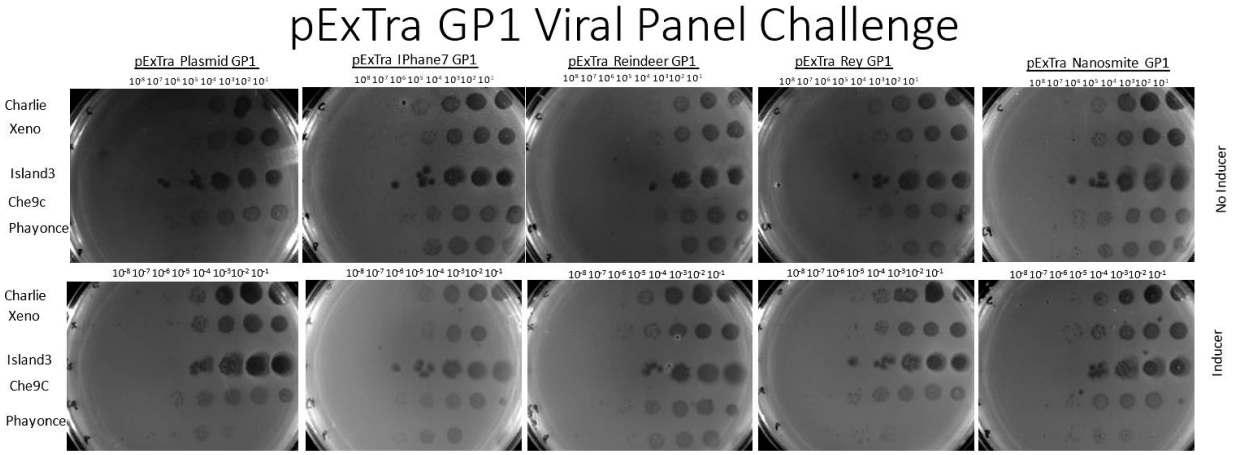



Figure 31. pExTra gp1 Viral Panel challenge results. pExTra vector clones were incubated at 37℃ on a shaker for 48 hours in 20% tween and 7H9Neat mixture, and then incubated on a shaker with no 20% tween in a 7H9 mixture that either contained 10ng/mL aTc or no aTc for 48 hours. Upon completion of the 48 hours, the mixtures that contained 10ng/mL aTc

were mixed with TA containing 10ng/mL aTc and plated on 7H9+Kan plates and allowed to gel. Similarly, the mixtures that did not contain any inducer were mixed with TA and plated on 7H9+Kan plates. The plates were spotted with serial dilutions of Viral Panel member phages and incubated at 37°C for 72 hours and then imaged. This figure is representative of the experiment performed in duplicate.

The results in Figure 31 display no phage specific inhibition. All experimental plates containing inducer match the control plates that possess no inducer, such that the same results are observed whether gene translation is turned on or off. This shows that the defense mechanisms observed in Figure 30 appear to be Cluster M specific and do not impact the other phages used in the viral experiment, none of which are Cluster M members. Final determination of non-cluster M inhibition capabilities cannot be confirmed without testing a larger panel of non-cluster M bacteriophages; however, conclusions can be drawn that clusters N, P5, I1, and I2 are able to overcome the defense mechanisms encoded for by cluster M gp1's.

**pMH94 intergenic+gp1 results.**

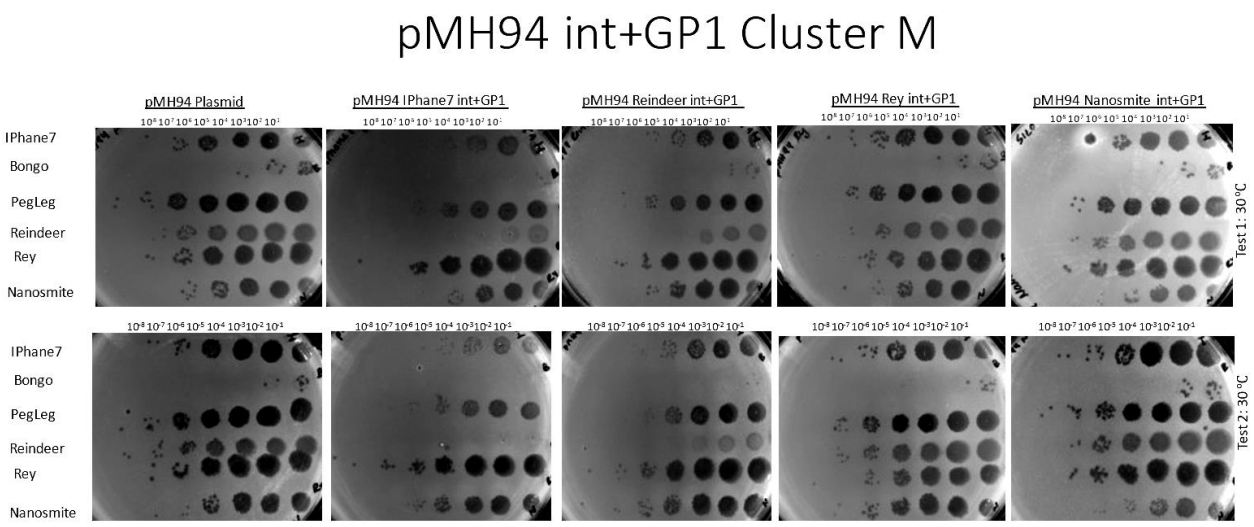



Figure 32. pMH94 int+gp1 Cluster M Panel challenge results. pMH94 vector clones were incubated at 37°C on a shaker for 48 hours in 20% tween and 7H9Neat mixture, and then

incubated on a shaker with no 20% tween in a 7H9 mixture. Upon completion of the 48 hours, the mixtures were mixed with TA and plated on 7H9+Kan plates. The plates were spotted with serial dilutions of Cluster M Panel member phages and incubated at 30℃ for 72 hours and then imaged.

The results of the pMH94 intergenic and gp1 cluster M challenge are shown in Figure 32. Rey will not grow at 37℃, and thus all plates were incubated at 30℃. Incubation temperatures do not affect other phage growth rates, as all other Cluster M phages displayed the same phenotype at 37℃ and 30℃. Note that Bongo displayed only $10^{-3}$ growth on the pMH94 Plasmid control plates. pMH94 IPhane7 test 1 and test 2 results display slight inhibition of IPhane7 to $10^{-3}$, and inhibition of Reindeer growth as well. Nanosmite growth was inhibited by a factor of 10. pMH94 Reindeer showed subtle inhibition of Reindeer and IPhane7 growth. pMH94 Rey and pMH94 Nanosmite displayed no effect on growth for any of the phages plated.

**pMH94 intergenic+gp1+gp2 results.**

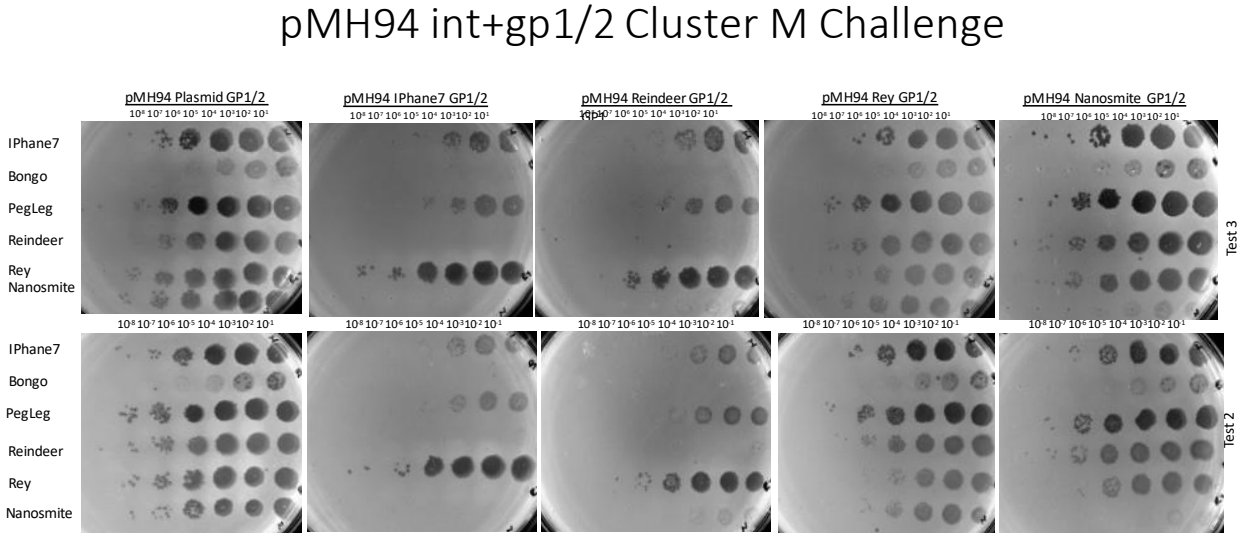


Figure 33. pMH94 int+gp1/2 Cluster M Panel challenge results. pMH94 vector clones were incubated at 37℃ on a shaker for 48 hours in 20% tween and 7H9Neat mixture, and then incubated on a shaker with no 20% tween in a 7H9 mixture. Upon completion of the 48 hours, the mixtures were mixed with TA and plated on 7H9+Kan plates. The plates were spotted with serial dilutions of Cluster M Panel member phages and incubated at 30℃ for 72 hours and then imaged. Imaged are Test 2 and Test 3 results for comparison. This figure is representative of the experiment performed in triplicate.

The pMH94 int+gp1/2 Cluster M challenge results in Figure 33 showed phage specific inhibitions. Gp1 and gp2 appear to have a stronger phenotype than the effects of gp1 alone. pMH94 IPhane7 gp1/2 displays a complete inhibition if Bongo, Reindeer, and Nanosmite while slight inhibition of PegLeg and IPHane7 is observed; Rey saw no inhibition. Comparable results are shown by pMH94 Reindeer gp1/2 expression. In the presence of Reindeer's gp1/2, Bongo and Reindeer see full inhibition, while Nanosmite, IPhane7 and PegLeg see mild inhibition. Again, Rey was not affected. pMH94 Rey gp1/2 appeared to have slight effect on Rey and Nanosmite, but it is very mild. pMH94 Nanosmite gp1/2 had strong effects on Nanosmite but did not appear to affect any other phages. These results confirm that gp1/2 defense mechanism is phage specific for Cluster M phages.

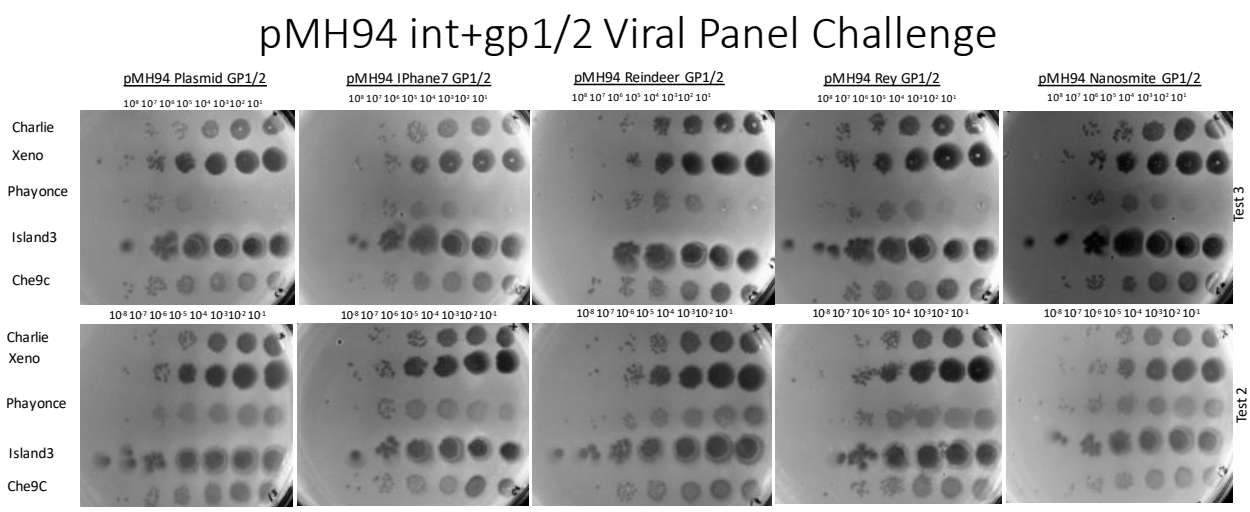



Figure 34. pMH94 int+gp1/2 Viral Panel challenge results. pMH94 vector clones were incubated at 37°C on a shaker for 48 hours in 20% tween and 7H9Neat mixture, and then incubated on a shaker with no 20% tween in a 7H9 mixture. Upon completion of the 48 hours, the mixtures were mixed with TA and plated on 7H9+Kan plates. The plates were spotted with serial dilutions of Cluster M Panel member phages and incubated at 37°C for 72 hours and then imaged. Imaged are Test 2 and Test 3 results for comparison. This figure is representative of the experiment performed in triplicate.

The pMH94 int+gp1/2 Viral Panel challenge in Figure 34 displayed no inhibition by gp1/2 from any of the non-cluster M bacteriophages. All bacteriophages from the viral panel were able to grow with no notable log inhibition observed when compared to the control plates.

## Discussion of Cluster M Prophage-Mediated Defense Genes Analysis

### Discussion of DEM mutations

Analysis of DEMs led to the observation of a possible connection between gp1 and a mutation in the Minor Tail Protein. Samples 1-6 saw identical mutations at 26018bp in gene 30 (pham52895 as of 11/4/2022), a minor tail protein of IPhane7. IPhane7 DEMs samples 1-6

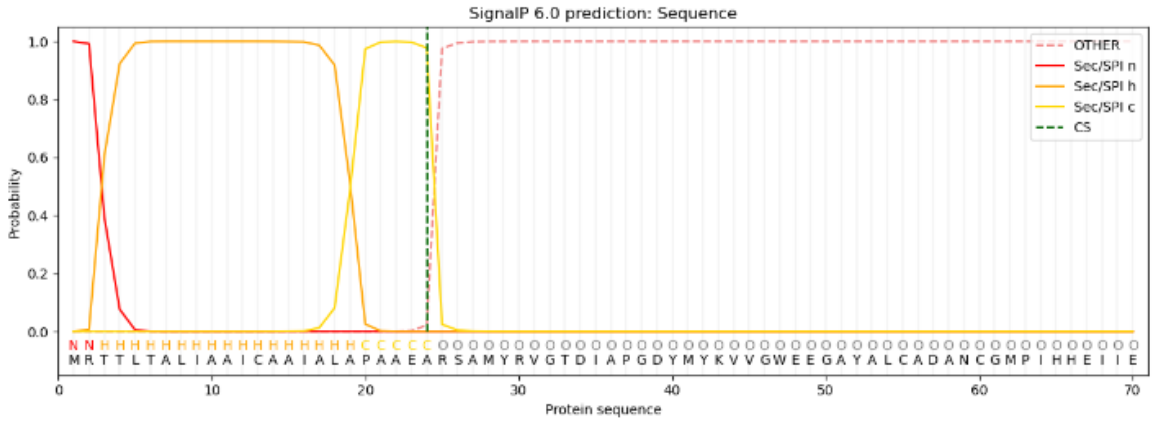
displayed a conserved point mutation at 26018bp from tyrosine (T) to cytosine (C) which results in an amino acid change from a Valine (GTG) to an alanine (GCG), both of which are neutral and non-polar. Valine possesses one more alkyl group than alanine, thus making valine more non-polar than alanine [58]. Further analysis revealed that sixty-eight other phages possess this gene including clusters: singletons, B2, F1, I1, M1, N, P1, W. Of interest to note in cluster M, only cluster M1 contains the minor tail protein, while cluster M2 and M3 have other gene variations for the minor tail proteins. Given the genetic mosaicism nature of temperate phages, it is not uncommon to see occurrences where there are various different sequences observed that provide the same or similar gene function [12]. When establishing lysogeny, a temperate phage's integrated DNA may not be the only phage DNA present in the host cell. Thus, prophages constitute a substantial portion of laterally acquired DNA in many bacteria [10]. At some point in evolutionary history, cluster M phages gained different minor tail protein genes that play a role in phage attachment at the cellular surface and ability to overcome some prophage-mediated defense mechanisms that inhibit infection at the stage of phage attachment.

Mutations within the minor tail protein appears to grant IPhane7 DEM mutants the ability to overcome superinfection exclusion, which generally targets homotypic particles [18]. Based on the results of our study and identified responsibilities of tail proteins, the observed mutation between the six samples in the minor tail protein, pham 52895, proposes correlation with cell-surface attachment inhibition. IPhane7 prophage-mediated defense genes are designed to inhibit infection from other IPhane7 phages it is exposed to in the environment. However, the IPhane7 DEMs that are no longer subject to IPhane7-mediated defense developed key mutations notably in the minor tail protein, which aids in responsibility of host bacterium cell surface recognition and attachment. The IPhane7 DEMs presumably escape IPhane7 prophage-mediated defenses

either due to a structural change in the minor tail protein gene product that may affect receptor recognition, receptor-binding, or some other stage of phage attachment, or a gain of function mutations that has granted a counter-defense ability. IPhane7 gp1 has been identified as a Sec-signal peptide and is believed to act as a defense gene at the level of phage attachment to the host cell. Conserved mutations in the minor tail protein of IPhane7 DEMS and known IPhane7's gp1 properties guide us to the conclusion of correlation between gp1's defense mechanism and the importance of the amino acid change in gene 30, pham 52895. Further investigations into the IPhane7 DEMs mutations in other locations may lead to connections between the other mutations observed amongst the samples.

**Discussion of IPhane7 gp1, Cluster M gp1 and Cluster M gp1/2 defense profiles**

Coming in at only 97 aa, IPhane7's gp1 was predicted to possess a Sec-signal peptide with cleavage between aa 24 (alanine) and 25 (arginine). Signal peptides are short amino acid sequences found at the amino terminus of many proteins that target proteins into or across membranes. They may be found in secreted or transmembrane (TM) proteins, whereby the secretory pathway (Sec) transports proteins in an unfolded state [39]. Figure 35 shows SignalP 6.0 results of IPhane7 gp1 and is predicted to be a Sec-signal peptide with a 0.977916 probability.




| Protein type | Other | Signal Peptide (Sec/SPI) | Lipoprotein signal peptide (Sec/SPII) | TAT signal peptide (Tat/SPI) | TAT Lipoprotein signal peptide (Tat/SPII) | Pilin-like signal peptide (Sec/SPIII) |
|---|---|---|---|---|---|---|
| Likelihood | 0.0003 | 0.9991 | 0.0002 | 0.0002 | 0.0002 | 0.0001 |

Figure 35. SignalP 6.0 results of IPhane7 gp1. IPhane7's gp1 is predicted to be a Sec-signal peptide with a 0.977916 probability.

To understand how gp1's encoded sec-signal peptide plays a role in cluster M prophage gene expression, a protein alignment was performed on all cluster M members as seen in Figures 24, 25, and 26. The results displayed significant sequence conservation amongst the cluster M1's with all but Reindeer sharing identical primary protein structures. The same pattern does not continue in the cluster M2s, where there is a lower percentage of identical sequences. GardenSalsa and MrMagoo share the same protein sequence. Rey and Estes are unique and share no identical protein sequences to any other M2. Aziz and GenevaB15 share identical protein sequences. As seen in Figure 36, protein sequence similarities of cluster M gp1 can be observed based on an evolutionary standpoint. Nanosmite's gp1 protein sequence is more similar to the M2's than the M1's sequence but is unique as it is the only M3. These trends align with their predicted evolutionary development observed in Figure 36.

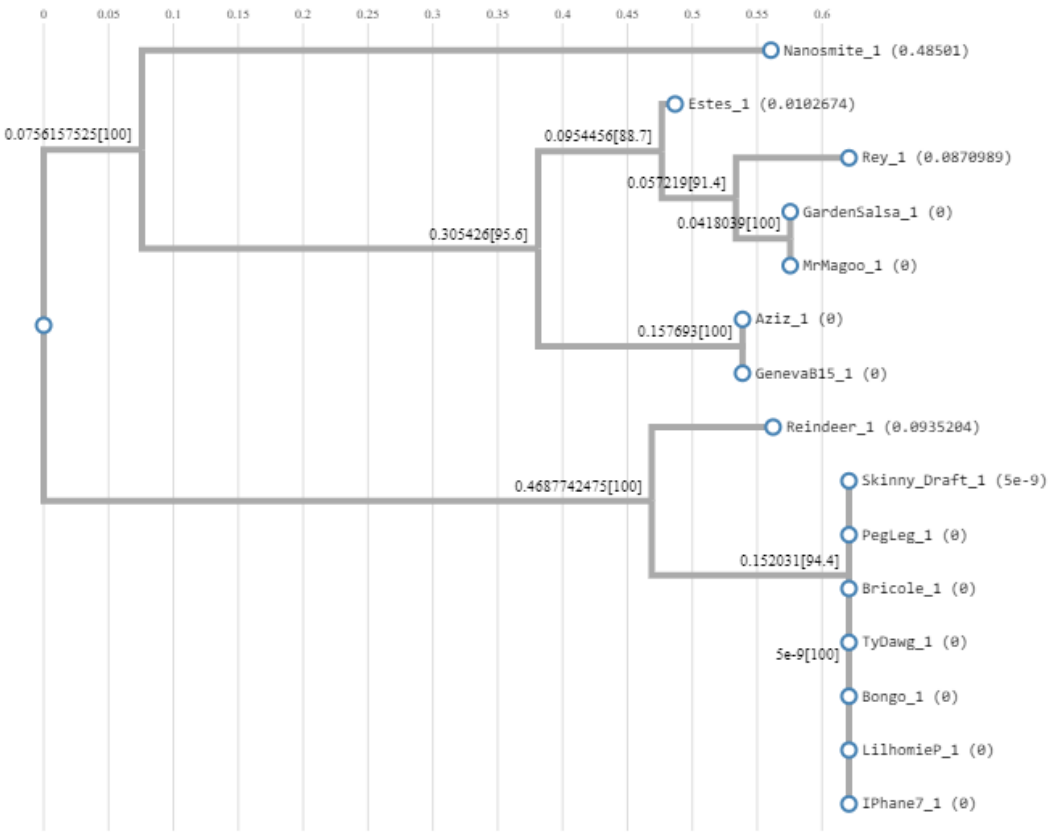



Figure 36. Phylogenetic Tree of Cluster M phages for gp1 acquisition. The phylogenetic tree was generated by Phylogenetic analysis pipeline by ETE3 from a ClustalW generation comparing the protein sequences of gp1. [61]

Imperative to our understanding of the possible defense mechanism encoded for by cluster M gp1s, we tested a truncated version of IPhane7's gp1, where the sec-signal peptide was removed from the gene at the cleavage site between aa 24 and 25 to test whether the peptide was secreted outside of the cell and affected the ability to establish a productive infection within a host cell containing the gene. The results showed that by removing the signal sequence, we eliminated the ability to defend against IPhane7 infection. IPhane7's gp1 sec-signal peptide is imperative for proper function of the gene to confer defense against superinfection of other phages. The exact effects on the protein's loss of function are unexamined when truncating the protein; it is possible that the protein still folded and/or remained stable but was insoluble and

thus unable to act outside of the cell, or the protein lost the ability to properly fold into a stable form. Further investigation into the effects on protein structure would be necessary to determine the exact effects on solubility, stability, and folding capabilities in the truncated form. For this study, it was important to investigate the loss of function phenotype observed when truncating IPhane7 gp1.

Observed differences of gp1 protein sequences amongst the cluster M members led to a panel selection to examine how they may impact prophage-mediated defense specificity between cluster M members as well as a broad panel of phages. Our results using a high-level expression vector displayed the power of cluster M1 gp1's to defend against cluster M phage infection. IPhane7 (M1) and Reindeer (M1) saw almost complete inhibition of all infection attempts, with the exception of Nanosmite to overcome Reindeer's gp1 (Figure 30). The same is not observed by Rey's gp1 and Nanosmite's gp1 who saw no defense capabilities provided except with homotypic inhibition of self-infections. Clearly, IPhane7 (M1) gp1 encodes for a powerful defense mechanism that no other cluster M was able to overcome when overexpressed. Our results using an endogenous-level expression vector, pMH94, displayed the natural defense profile of cluster M gp1's as highly phage specific defense capabilities that appear to be limited to those within the cluster M phamily, as seen in Figure 32. Less drastic inhibitions were observed in the pMH94 int+gp1 experiments confirming that endogenous levels of gp1 alone had a weaker phenotype and could not fully inhibit infection from any cluster M members. Mesotypic, and heterotypic immunities were observed directly correlating with genomic differences in gp1 protein sequences.

We propose that slight sequence differences amongst cluster M gp1s has led to varying degrees of inhibition between the subclusters based on the ability of the secreted signal peptide

encoded for by gp1's ability to confer superinfection immunity at the level of phage attachment. Interestingly, the minor tail protein identified to have conserved mutations in the IPhane7 DEMs is only present in cluster M1 phages. Clusters M2 and M3 possess genomically different minor tail proteins. Perhaps cluster M1 gp1 gene encodes for defense mechanisms that target the minor tail protein gene present in similar cluster M1s, just as cluster M2 gp1 would target minor tail proteins specifically encoded for within their subcluster genomes. Similar trends are observed with Nanosmite gp1 inhibiting Nanosmite infection but unable to defend against any other cluster M phage infections. Glaringly, the genomic interworking's of encoded defense genes between the cluster M subclusters is highly specific and displays evolutionary impacts on prophage-mediated defense mechanisms.

Knowing the highly intricate cluster M interacts based on our results, we looked towards a varying panel of phages to examine cluster M gp1's defense profile outside of cluster M phages. Members from clusters N, P5, I1, and I2 were challenged against high-level expression of cluster M gp1 and displayed uninhibited infections (Figure 31). Since no level of inhibition was observed using high-levels of expression of cluster gp1s, we chose to not examine the effect of endogenous-levels of expression on inhibition of the viral panel. The severe genomic differences between clusters N, P5, I1 and I2, as seen in Figure 37, against cluster M gp1 led to the conclusion that cluster M defense genes have evolved to prevent infection from similar cluster M phages, but not a broad genomic capability based on our panel.

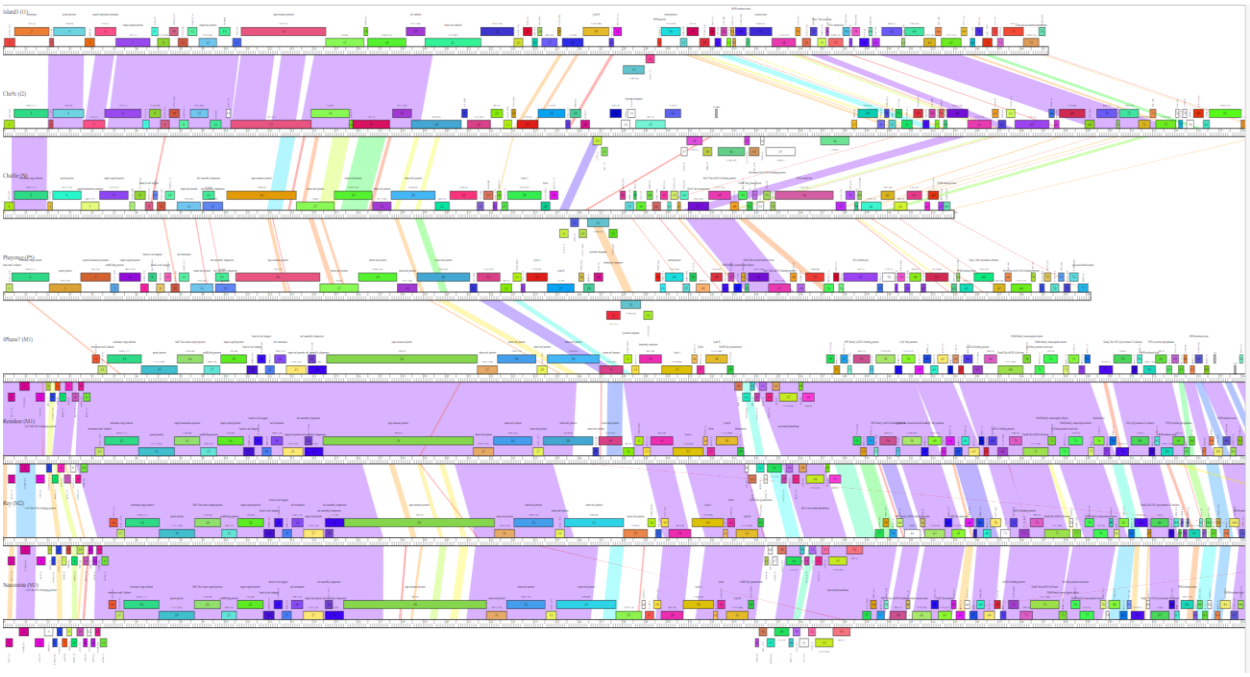


Figure 37. Phamerator map of cluster I1, I2, M1, M2, M3, N, and P5 phages used during experiments. Clear nucleotide sequence similarities and differences can be observed between clusters I1, I2, N, and P5 in relation to cluster M1, M2, and M3 genomes.

Further investigation is necessary to confirm the ability of cluster M defense genes to inhibit infection from all other clusters. By conducting a viral challenge with a broader, more inclusive panel of phages from multiple other clusters, a complete analysis of cluster M gp1 defense capabilities may be confirmed.

**Discussion of pMH94 gp1/2 defense profile**

IPhane7's gp2, as well as all other cluster M phages, is of unknown function. However, previous experiments in the Gainey laboratory displayed synergistic properties of IPhane7's gp1 and gp2 when expressed together. Upon bioinformatic analysis of gp2 protein sequence, HHPred shows inconclusive results and UniProt's BLAST identifies it as a DUF3846 domain-containing protein that shares sequence similarity with Mycobacteriophages Bongo (100%), Reindeer

(88.8%), Rey (62.9%), Estes (62.5%), Nanosmite (65.1%), and prophiGD54-2 (42.6%) as some of the top hits [5]. Knowing that gp1 alone displayed a strong phenotype, with some full knockout of fellow cluster M members, we analyzed the ability of cluster M gp1 and gp2 to defend against infection at the endogenous level using the pMH94 vector. Recalling that pMH94 int+gp1 alone displayed a very weak inhibitory phenotype against cluster M phages (Figure 32), we expected the combination of gp1/2 together would display a different phenotype. Our hypothesis was correct, as the pMH94 int+gp1/2 cluster M challenge (Figure 33) displayed a stronger phenotypic response of defense capabilities only against cluster M phages. The combination of both gp1 and gp2 clearly has a strong inhibitory effect on specific cluster M members, while Nanosmite (M3) gp1/2 can only inhibit Nanosmite infections and Rey (M2) gp1/2 is unable to defend against any infection – including itself, which is highly intriguing as it is uncharacteristic of a phage to be unable to prevent infection of itself, leading to the possibility that the defense mechanism has evolved differently in the M2s. The diversity of cluster M genomes clearly plays a role in the highly specific defense mechanism that is encoded for by gp1 and appears to intensify by expression of gp2. Further bioinformatic analysis between cluster M1 minor tails, as well as other cluster M phages, is necessary to understand the connection between the secreted signal peptide encoded for by gp1 and its interactions with phage attachment capabilities at the cellular surface level.

# CHAPTER 4: CONCLUSIONS AND FUTURE DIRECTIONS

The results gathered thus far present a multitude of avenues for future investigation. Cluster M phages may carry a wide range of superinfection resistance mechanisms that clearly correlates with phage infectivity on lysogenic hosts. Analysis of host ranges for the phages when infecting various Cluster M lysogens displayed phages belonging to the same lysogen groups having similar infectivity ranges, such that they were typically unable to infect members of the same lysogen group. For example, a majority of the cluster M1 phages were unable to plaque on lysogens of any cluster M1 lysogens, except for Pegleg (M1) that exhibited the ability to infect IPhane7 (M1) lysogen with an approximate $10^{-4}$ log inhibition (Figure 15). In contrast, heterotypic phages that did not belong to cluster M were all able to infect cluster M lysogens (Figure 16). The complete superinfection resistance abilities of cluster M lysogens cannot be concluded without a large-scale challenge between all cluster M lysogens against a broad panel including a multitude of clusters.

Knowing that bacterial genomes may harbor multiple prophages from various clusters within a host range [11], it is unclear how vast the effects of horizontal gene transfer may occur between prophages inhabiting the same host cell. One may observe how phages have evolved amongst the presence of one another due to the apparent mosaicism of their genomes, where each genome may be viewed as a unique combination of modules that are exchanged among the population [28]. To further extrapolate the defense capabilities of cluster M lysogens against other clusters by challenging against a broad panel of other phages, one may discover shared genomic elements of cluster M prophages. Upon completion of this study, new bacteriophages, Glaske16 (M1), Auspice (M1), Diminimus (M1), Dulcita (M1), SlimJimmy (M1), and TyDawg (M1) were

added to the cluster M database on PhagesDB.org. These phages were not included in this study, as they had yet to be added to the cluster Ms. Future experiments involving cluster M bacteriophages should include all known phages for holistic representation of the cluster.

Characterization of the IPhane7 lysogen integration sites required experimental identification, as serine integrases cannot be determined bioinformatically. This is unlike tyrosine integrases that are capable of bioinformatic identification due to certain biomarkers and longer common core sequences [27]. Int-S systems have notably shorter common core sequences that play a role in directionality during integration. As observed in the IPhane7 and *M. smegmatis* genomes, the identified common core sequence at the *attL* and *attR* sites was found in various areas of each genome, such that there is reason to question why the prophage integrated at these specific sites. Future studies examining the redundancy of integration at the identified sites in this study with other IPhane7 lysogens to see if the integration sites are conserved would aide in the consistency and confidence of one specific recombination site. Additional studies analyzing the integration sites of other cluster M lysogens would behoove the understanding of serine integrase recombination conserveness within the cluster.

IPhane7 defense escape mutants sequenced, assembled, and bioinformatically analyzed during this study were found to possess a conserved mutation in the minor tail protein gene 30, amongst other mutations in conserved regions of the genome. IPhane7 gp1 is known to encode for a Sec-signal peptide. Based on the mutation's observed in the IPhane7 DEMs, and the experimentally determined vitality of the signal sequence to the function of gp1, we conclude that gp1 acts to defend against phage infection at the level of cellular attachment of the phage tail to the bacterial cell surface. Further confirmation of gp1's secretion outside of the cell would aide to the confidence of this conclusion in a future study. Future experiments may be conducted

to examine the importance of the single nucleotide mutation that results in an amino acid change within the minor tail protein gene 30 and its ability to overcome the defense mechanism encoded for by gp1 or gp1/2 synergistically.

As observed in this study, cluster M gp1 and gp1/2 challenges displayed phage specific inhibitions which were likely due to the slight genomic differences amongst the subclusters, as well as within the same subcluster. Clearly displayed in the viral challenges, cluster M gp1 and gp1/2 defense mechanisms were ineffective in inhibiting infection by heterotypic phages belonging to other clusters. To confirm the abilities of gp1 and gp1/2 to defend against other clusters, a broad panel experiment should be conducted containing a vast array of phages. For instance, cluster M gp1 or gp1/2 may be challenged against a panel of 50 cluster A members, and only display a phenotype against one phage due to the nature the phage genetic mosaicism [12].

Of future interest and investigation into the Cluster M gp1/2 system, the WhiB family transcription factors genes appear to interact with gp2 by binding in what is believed a type of on/off switch for the system. Exclusively found in Actinobacteria, the WhiB-like (Wbl) family of proteins have been show to play key roles in virulence and antibiotic resistance in *Mycobacteria* and *Corynebacteria* [9]. WhiB-like proteins have been described in phages infecting *Mycobacteria* species, such as the WhiB of phage TM4 was shown to inhibit the expression of the host *whiB2* gene resulting in morphological changes upon expression [72,76]. WhiB of phage Tweety was suggested to be targeted by prophage-mediated defense mechanisms [18]. Thus far, the functions of WhiB-like proteins in the phage life cycle remain largely enigmatic.

Particular to our study, it is thought that the interaction between gp2 and the WhiB genes in an infecting phage may affect the ability of the defending phage to inhibit infection. For

instance, IPhane7's gp2 cannot bind to Nanosmite's WhiB genes, therefore the system is unable to "turn on" and is unable to inhibit infection. In contrast, IPhane7 gp2 binds with high affinity to IPhane7's WhiB genes and is able to turn on the system to inhibit infection. The highly specific interactions between specifically encoded gp2 and WhiB family transcription factor genes may unearth further intricacies of the prophage-mediated defense mechanism found here in the Cluster M gp1/2 system. The complex and intricate systems that comprise prophage-mediated defense mechanisms and systems is a new area of investigation with much to be discovered.

REFERENCES

1. Ackermann H-W. 5500 Phages examined in the electron microscope. Archives of Virology. 2007;152(2):227–243. doi:10.1007/s00705-006-0849-1

2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology. 1990;215(3):403–410. doi:10.1016/S0022-2836(05)80360-2

3. Arnaud C-A, Effantin G, Vivès C, Engilberge S, Bacia M, Boulanger P, Girard E, Schoehn G, Breyton C. Bacteriophage T5 tail tube structure suggests a trigger mechanism for Siphoviridae DNA ejection. Nature Communications. 2017;8(1):1953. doi:10.1038/s41467-017-02049-3

4. Auzat I, Dröge A, Weise F, Lurz R, Tavares P. Origin and function of the two major tail proteins of bacteriophage SPP1. Molecular Microbiology. 2008;70(3):557–569. doi:10.1111/j.1365-2958.2008.06435.x

5. BLAST results | UniProt. 2022 Nov 18 [accessed 2022 Nov 18]. https://www.uniprot.org/blast/uniprotkb/ncbiblast-R20221118-200839-0825-41896360-p1m/overview

6. Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR, Maxwell KL. Prophages mediate defense against phage infection through diverse mechanisms. The ISME Journal. 2016;10(12):2854–2866. doi:10.1038/ismej.2016.79

7. Broussard GW, Hatfull GF. Evolution of genetic switch complexity. Bacteriophage. 2013;3(1):e24186. doi:10.4161/bact.24186

8. Brown KL, Sarkis GJ, Wadsworth C, Hatfull GF. Transcriptional silencing by the mycobacteriophage L5 repressor. The EMBO Journal. 1997;16(19):5914–5921. doi:10.1093/emboj/16.19.5914

9. Bush MJ. The actinobacterial WhiB-like (Wbl) family of transcription factors. Molecular Microbiology. 2018;110(5):663–676. doi:10.1111/mmi.14117

10. Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann M-L, Brüssow H. Phage as agents of lateral gene transfer. Current Opinion in Microbiology. 2003;6(4):417–424. doi:10.1016/S1369-5274(03)00086-9

11. Canchaya C, Proux C, Fournous G, Bruttin A, Brüssow H. Prophage Genomics. Microbiology and Molecular Biology Reviews. 2003;67(2):238–276. doi:10.1128/MMBR.67.2.238-276.2003

12. Casjens SR. Comparative genomics and evolution of the tailed-bacteriophages. Current Opinion in Microbiology. 2005;8(4):451–458. (Host--microbe interactions: fungi / edited by Howard Bussey · Host--microbe interactions: parasites / edited by Artur Scherf · Host--microbe interactions: viruses / edited by Margaret CM Smith). doi:10.1016/j.mib.2005.06.014

13. Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. BMC Bioinformatics. 2011;12(1):395. doi:10.1186/1471-2105-12-395

14. Cumby N, Edwards AM, Davidson AR, Maxwell KL. The Bacteriophage HK97 gp15 Moron Element Encodes a Novel Superinfection Exclusion Protein. Journal of Bacteriology. 2012;194(18):5012–5019. doi:10.1128/JB.00843-12

15. Davies EV, Winstanley C, Fothergill JL, James CE. The role of temperate bacteriophages in bacterial infection. FEMS Microbiology Letters. 2016;363(5):fnw015. doi:10.1093/femsle/fnw015

16. Dedrick RM, Aull HG, Jacobs-Sera D, Garlena RA, Russell DA, Smith BE, Mahalingam V, Abad L, Gauthier CH, Hatfull GF. The Prophage and Plasmid Mobilome as a Likely Driver of Mycobacterium abscessus Diversity. mBio. 2021;12(2):e03441-20. doi:10.1128/mBio.03441-20

17. Dedrick RM, Guerrero-Bustamante CA, Garlena RA, Russell DA, Ford K, Harris K, Gilmour KC, Soothill J, Jacobs-Sera D, Schooley RT, et al. Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant Mycobacterium abscessus. Nature Medicine. 2019;25(5):730–733. doi:10.1038/s41591-019-0437-z

18. Dedrick RM, Jacobs-Sera D, Bustamante CAG, Garlena RA, Mavrich TN, Pope WH, Reyes JCC, Russell DA, Adair T, Alvey R, et al. Prophage-mediated defence against viral attack and viral counter-defence. Nature Microbiology. 2017;2(3):1–13. doi:10.1038/nmicrobiol.2016.251

19. Dedrick RM, Smith BE, Garlena RA, Russell DA, Aull HG, Mahalingam V, Divens AM, Guerrero-Bustamante CA, Zack KM, Abad L, et al. Mycobacterium abscessus Strain Morphotype Determines Phage Susceptibility, the Repertoire of Therapeutically Useful Phages, and Phage Resistance. mBio. 2021;12(2):e03431-20. doi:10.1128/mBio.03431-20

20. Deveau H, Garneau JE, Moineau S. CRISPR/Cas System and Its Role in Phage-Bacteria Interactions. Annual Review of Microbiology. 2010;64(1):475–493. doi:10.1146/annurev.micro.112408.134123

21. Donnelly-Wu MK, Jacobs Jr WR, Hatfull GF. Superinfection immunity of mycobacteriophage L5: applications for genetic transformation of mycobacteria. Molecular Microbiology. 1993;7(3):407–417. doi:10.1111/j.1365-2958.1993.tb01132.x

22. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R. Systematic discovery of antiphage defense systems in the microbial pangenome. Science. 2018;359(6379):eaar4120. doi:10.1126/science.aar4120

23. Fineran PC, Blower TR, Foulds IJ, Humphreys DP, Lilley KS, Salmond GPC. The phage abortive infection system, ToxIN, functions as a protein–RNA toxin–antitoxin pair. Proceedings of the National Academy of Sciences. 2009;106(3):894–899. doi:10.1073/pnas.0808832106

24. Gentile GM, Wetzel KS, Dedrick RM, Montgomery MT, Garlena RA, Jacobs-Sera D, Hatfull GF. More Evidence of Collusion: a New Prophage-Mediated Viral Defense System Encoded by Mycobacteriophage Sbash. mBio. 2019;10(2):e00196-19. doi:10.1128/mBio.00196-19

25. Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Charpak-Amikam Y, Afik S, Ofir G, Sorek R. BREX is a novel phage resistance system widespread in microbial genomes. The EMBO Journal. 2015;34(2):169–183. doi:10.15252/embj.201489455

26. Grindley NDF, Whiteson KL, Rice PA. Mechanisms of Site-Specific Recombination. Annual Review of Biochemistry. 2006;75(1):567–605. doi:10.1146/annurev.biochem.73.011303.073908

27. Groth AC, Calos MP. Phage Integrases: Biology and Applications. Journal of Molecular Biology. 2004;335(3):667–678. doi:10.1016/j.jmb.2003.09.082

28. Hatfull GF. Bacteriophage genomics. Current Opinion in Microbiology. 2008;11(5):447–453. (Antimicrobials/Genomics). doi:10.1016/j.mib.2008.09.004

29. HATFULL GF. Molecular Genetics of Mycobacteriophages. Microbiology spectrum. 2014;2(2):1–36.

30. Hatfull GF. MYCOBACTERIOPHAGES: DIVERSITY, DYNAMICS, AND THERAPEUTIC POTENTIAL. :18.

31. Hatfull GF. Mycobacteriophages: From Petri dish to patient. PLOS Pathogens. 2022;18(7):e1010602. doi:10.1371/journal.ppat.1010602

32. Hatfull GF. The secret lives of mycobacteriophages. Advances in Virus Research. 2012;82:179–288. doi:10.1016/B978-0-12-394621-8.00015-7

33. Hatfull GF, Cresawn SG, Hendrix RW. Comparative genomics of the mycobacteriophages: insights into bacteriophage evolution. Research in Microbiology. 2008;159(5):332–339. (Exploring the prokaryotic virosphere). doi:10.1016/j.resmic.2008.04.008

34. Huson DH. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics. 1998;14(1):68–73. doi:10.1093/bioinformatics/14.1.68

35. Hyman P, Abedon ST. Chapter 7 - Bacteriophage Host Range and Bacterial Resistance. In: Advances in Applied Microbiology. Vol. 70. Academic Press; 2010 [accessed 2022 Dec 10]. p. 217–248. (Advances in Applied Microbiology). https://www.sciencedirect.com/science/article/pii/S0065216410700071. doi:10.1016/S0065-2164(10)70007-1

36. IPhane7_EMPic.png (1650×1270). [accessed 2022 Nov 18]. https://phagesdb.org/media/emPics/IPhane7_EMPic.png

37. Jacobs-Sera D, Marinelli LJ, Bowman C, Broussard GW, Guerrero Bustamante C, Boyle MM, Petrova ZO, Dedrick RM, Pope WH, Science Education Alliance Phage Hunters

Advancing Genomics and Evolutionary Science (SEA-PHAGES) program, et al. On the nature of mycobacteriophage diversity and host preference. Virology. 2012;434(2):187–201. (Special issue: viruses of microbes). doi:10.1016/j.virol.2012.09.026

38. de Jonge PA, Nobrega FL, Brouns SJJ, Dutilh BE. Molecular and Evolutionary Determinants of Bacteriophage Host Range. Trends in Microbiology. 2019;27(1):51–63. doi:10.1016/j.tim.2018.08.006

39. Juan AAJ, Tsirigos KD, Kaae SC, Petersen TN, Ole W, Søren B, Gunnar von H, Nielsen H. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature Biotechnology. 2019;37(4):420–423. doi:10.1038/s41587-019-0036-z

40. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583–589. doi:10.1038/s41586-021-03819-2

41. Katsura I, Hendrix RW. Length determination in bacteriophage lambda tails. Cell. 1984;39(3, Part 2):691–698. doi:10.1016/0092-8674(84)90476-8

42. Kim AI, Ghosh P, Aaron MA, Bibb LA, Jain S, Hatfull GF. Mycobacteriophage Bxb1 integrates into the Mycobacterium smegmatis groEL1 gene. Molecular Microbiology. 2003;50(2):463–473. doi:10.1046/j.1365-2958.2003.03723.x

43. Lee MH, Hatfull GF. Mycobacteriophage L5 integrase-mediated site-specific integration in vitro. Journal of Bacteriology. 1993;175(21):6836–6841. doi:10.1128/jb.175.21.6836-6841.1993

44. Lee S, Lewis DEA, Adhya S. The Developmental Switch in Bacteriophage λ: A Critical Role of the Cro Protein. Journal of molecular biology. 2018;430(1):58–68. doi:10.1016/j.jmb.2017.11.005

45. Li W, Kamtekar S, Xiong Y, Sarkis GJ, Grindley NDF, Steitz TA. Structure of a Synaptic γδ Resolvase Tetramer Covalently Linked to Two Cleaved DNAs. Science. 2005;309(5738):1210–1215. doi:10.1126/science.1112064

46. Luong T, Salabarria A-C, Roach DR. Phage Therapy in the Resistance Era: Where Do We Stand and Where Are We Going? Clinical Therapeutics. 2020;42(9):1659–1680. doi:10.1016/j.clinthera.2020.07.014

47. lysogenic.html 19_06LambdaLyticLysoCycle-L.jpg. [accessed 2022 Dec 10]. http://bio3400.nicerweb.com/bio1151/Locked/media/ch19/lysogenic.html

48. Mavrich TN, Hatfull GF. Evolution of Superinfection Immunity in Cluster A Mycobacteriophages. mBio. 2019;10(3):e00971-19. doi:10.1128/mBio.00971-19

49. McGinnis RJ, Brambley CA, Stamey B, Green WC, Gragg KN, Cafferty ER, Terwilliger TC, Hammel M, Hollis TJ, Miller JM, et al. A monomeric mycobacteriophage immunity repressor

utilizes two domains to recognize an asymmetric DNA sequence. Nature Communications. 2022;13(1):4105. doi:10.1038/s41467-022-31678-6

50. Monarch® Plasmid DNA Miniprep Kit Protocol (NEB #T1010) | NEB. [accessed 2022 Oct 10]. https://www.neb.com/protocols/2015/11/20/monarch-plasmid-dna-miniprep-kit-protocol-t1010

51. Montgomery MT, Guerrero Bustamante CA, Dedrick RM, Jacobs-Sera D, Hatfull GF. Yet More Evidence of Collusion: a New Viral Defense System Encoded by Gordonia Phage CarolAnn. mBio. 2019;10(2):e02417-18. doi:10.1128/mBio.02417-18

52. Montgomery MT, Guerrero Bustamante CA, Dedrick RM, Jacobs-Sera D, Hatfull GF. Yet More Evidence of Collusion: a New Viral Defense System Encoded by Gordonia Phage CarolAnn. mBio. 2019;10(2):e02417-18. doi:10.1128/mBio.02417-18

53. NEBcloner. [accessed 2022 Oct 7]. https://nebcloner.neb.com/#!/protocol/re/single/DpnI

54. NEBuilder HiFi DNA Assembly Reaction Protocol | NEB. [accessed 2022 Oct 10]. https://www.neb.com/protocols/2014/11/26/nebuilder-hifi-dna-assembly-reaction-protocol

55. Ofir G, Melamed S, Sberro H, Mukamel Z, Silverman S, Yaakov G, Doron S, Sorek R. DISARM is a widespread bacterial defence system with broad anti-phage activities. Nature Microbiology. 2018;3(1):90–98. doi:10.1038/s41564-017-0051-0

56. Ojha A, Anand M, Bhatt A, Kremer L, Jacobs WR, Hatfull GF. GroEL1: A Dedicated Chaperone Involved in Mycolic Acid Biosynthesis during Biofilm Formation in Mycobacteria. Cell. 2005;123(5):861–873. doi:10.1016/j.cell.2005.09.012

57. Ongenae V, Briegel A, Claessen D. Cell wall deficiency as an escape mechanism from phage infection. Open Biology. 2021;11(9):210199. doi:10.1098/rsob.210199

58. Ophardt CE. Amino Acids. Virtual Chembook, Elmhurst College. 2003 [accessed 2022 Nov 18]. http://chemistry.elmhurst.edu/vchembook/561aminostructure.html

59. Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, Lewis JA, Jacobs-Sera D, Falbo J, Gross J, Pannunzio NR, et al. Origins of Highly Mosaic Mycobacteriophage Genomes. Cell. 2003;113(2):171–182. doi:10.1016/S0092-8674(03)00233-2

60. Petrova ZO, Broussard GW, Hatfull GF. Mycobacteriophage-repressor-mediated immunity as a selectable genetic marker: Adephagia and BPs repressor selection. Microbiology. 2015;161(Pt 8):1539–1551. doi:10.1099/mic.0.000120

61. Phylogenetic analysis pipeline by ETE3. 2022 Nov 18 [accessed 2022 Nov 18]. https://www.genome.jp/tools-bin/ete?id=221119021619a9a45b3a64eb1b129a82b71e6330be5735438d5d

62. Pope WH, Anders KR, Baird M, Bowman CA, Boyle MM, Broussard GW, Chow T, Clase KL, Cooper S, Cornely KA, et al. Cluster M mycobacteriophages Bongo, PegLeg, and Rey with

unusually large repertoires of tRNA isotypes. Journal of Virology. 2014;88(5):2461–2480. doi:10.1128/JVI.03363-13

63. Protocol 2.9: Chemical Transformation of Bacteria. [accessed 2022 Oct 10]. https://seagenes.helpdocsonline.com/2-9-protocol

64. Protocol 3.3: Defense Assay. [accessed 2022 Oct 10]. https://seagenes.helpdocsonline.com/3-1-protocol-2

65. Protocol for DNA Cleanup and Concentration Using the Monarch® PCR & DNA Cleanup Kit (5 μg) (NEB #T1030) | NEB. [accessed 2022 Oct 7]. https://www.neb.com/protocols/2015/11/23/monarch-pcr-and-dna-cleanup-kit-protocol

66. Protocol for Q5® Hot Start High-Fidelity 2X Master Mix | NEB. [accessed 2022 Oct 6]. https://www.neb.com/protocols/2012/08/30/protocol-for-q5-hot-start-high-fidelity-2x-master-mix-m0494

67. Ptashne M. A genetic switch: Gene control and phage. lambda. 1986 Jan 1 [accessed 2023 Jan 24]. https://www.osti.gov/biblio/5413898

68. Recipe Cards. [accessed 2022 Oct 6]. https://seagenesinstructorguide.helpdocsonline.com/12-0-recipes

69. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Research. 2014;42(W1):W320–W324. doi:10.1093/nar/gku316

70. Rutherford K, Van Duyne GD. The ins and outs of serine integrase site-specific recombination. Current Opinion in Structural Biology. 2014;24:125–131. (Folding and binding / Nucleic acids and their protein complexes). doi:10.1016/j.sbi.2014.01.003

71. Rutherford K, Yuan P, Perry K, Sharp R, Van Duyne GD. Attachment site recognition and regulation of directionality by the serine integrases. Nucleic Acids Research. 2013;41(17):8341–8356. doi:10.1093/nar/gkt580

72. Rybniker J, Nowag A, Van Gumpel E, Nissen N, Robinson N, Plum G, Hartmann P. Insights into the function of the WhiB-like protein of mycobacteriophage TM4 – a transcriptional inhibitor of WhiB2. Molecular Microbiology. 2010;77(3):642–657. doi:10.1111/j.1365-2958.2010.07235.x

73. Salmon KA, Freedman O, Ritchings BW, DuBow MS. Characterization of the Lysogenic Repressor (c) Gene of the Pseudomonas aeruginosa Transposable Bacteriophage D3112. Virology. 2000;272(1):85–97. doi:10.1006/viro.2000.0341

74. Sarkis GJ, Hatfull GF. Mycobacteriophages. In: Parish T, Stoker NG, editors. Mycobacteria Protocols. Totowa, NJ: Humana Press; 1998 [accessed 2022 Dec 10]. p. 145–173. (Methods in Molecular Biology™). https://doi.org/10.1385/0-89603-471-2:145. doi:10.1385/0-89603-471-2:145

75. Sassi M, Gouret P, Chabrol O, Pontarotti P, Drancourt M. Mycobacteriophage-drived diversification of Mycobacterium abscessus. Biology Direct. 2014;9(1):19. doi:10.1186/1745-6150-9-19

76. Sharma V, Hardy A, Luthe T, Frunzke J. Phylogenetic Distribution of WhiB- and Lsr2-Type Regulators in Actinobacteriophage Genomes. Microbiology Spectrum. 9(3):e00727-21. doi:10.1128/Spectrum.00727-21

77. Singh S, Ghosh P, Hatfull GF. Attachment Site Selection and Identity in Bxb1 Serine Integrase-Mediated Site-Specific Recombination. PLoS Genetics. 2013;9(5):e1003490. doi:10.1371/journal.pgen.1003490

78. Smith KC, Castro-Nallar E, Fisher JN, Breakwell DP, Grose JH, Burnett SH. Phage cluster relationships identified through single gene analysis. BMC Genomics. 2013;14(1):410. doi:10.1186/1471-2164-14-410

79. Smith MCM. Phage-encoded Serine Integrases and Other Large Serine Recombinases. Microbiology Spectrum. 2015;3(4):3.4.06. doi:10.1128/microbiolspec.MDNA3-0059-2014

80. Srividhya KV, Alaguraj V, Poornima G, Kumar D, Singh GP, Raghavenderan L, Katta AVSKM, Mehta P, Krishnaswamy S. Identification of Prophages in Bacterial Genomes by Dinucleotide Relative Abundance Difference. PLoS ONE. 2007;2(11):e1193. doi:10.1371/journal.pone.0001193

81. Stark WM. Making serine integrases work for us. Current Opinion in Microbiology. 2017;38:130–136. (Mobile genetic elements and HGT in prokaryotes * Microbiota). doi:10.1016/j.mib.2017.04.006

82. The Actinobacteriophage Database | Cluster M Phages. 2022 Dec 10 [accessed 2022 Dec 10]. https://phagesdb.org/clusters/M/

83. The Actinobacteriophage Database | Clusters. 2022 Dec 10 [accessed 2022 Dec 10]. https://phagesdb.org/clusters/

84. The Actinobacteriophage Database | Phage IPhane7. [accessed 2022 Dec 11]. https://phagesdb.org/phages/IPhane7/

85. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research. 1994;22(22):4673–4680.

86. Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. Nucleic Acids Research. 2015;43(W1):W401-407. doi:10.1093/nar/gkv485

87. Uc-Mass A, Loeza EJ, de la Garza M, Guarneros G, Hernández-Sánchez J, Kameyama L. An orthologue of the cor gene is involved in the exclusion of temperate lambdoid phages. Evidence

that Cor inactivates FhuA receptor functions. Virology. 2004;329(2):425–433. doi:10.1016/j.virol.2004.09.005

88. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Research. 2022;50(D1):D439–D444. doi:10.1093/nar/gkab1061

89. Vidhani S, Mehndiratta P. Bacteriophage Therapy : An Alternative to Conventional Antibiotics. 2003;51:4.

90. Villanueva VM, Oldfield LM, Hatfull GF. An Unusual Phage Repressor Encoded by Mycobacteriophage BPs. PLOS ONE. 2015;10(9):e0137187. doi:10.1371/journal.pone.0137187

91. Vuksanovic N, Zhu X, Serrano DA, Siitonen V, Metsä-Ketelä M, Melançon CE, Silvaggi NR. Structural characterization of three noncanonical NTF2-like superfamily proteins: implications for polyketide biosynthesis. Acta Crystallographica Section F: Structural Biology Communications. 2020;76(8):372–383. doi:10.1107/S2053230X20009814

92. Waldor MK, Friedman DI. Phage regulatory circuits and virulence gene expression. Current Opinion in Microbiology. 2005;8(4):459–465. (Host--microbe interactions: fungi / edited by Howard Bussey · Host--microbe interactions: parasites / edited by Artur Scherf · Host--microbe interactions: viruses / edited by Margaret CM Smith). doi:10.1016/j.mib.2005.06.001

93. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. Journal of Molecular Biology. 2018;430(15):2237–2243. (Computation Resources for Molecular Biology). doi:10.1016/j.jmb.2017.12.007

APPENDIX A: ADDITIONAL FIGURES

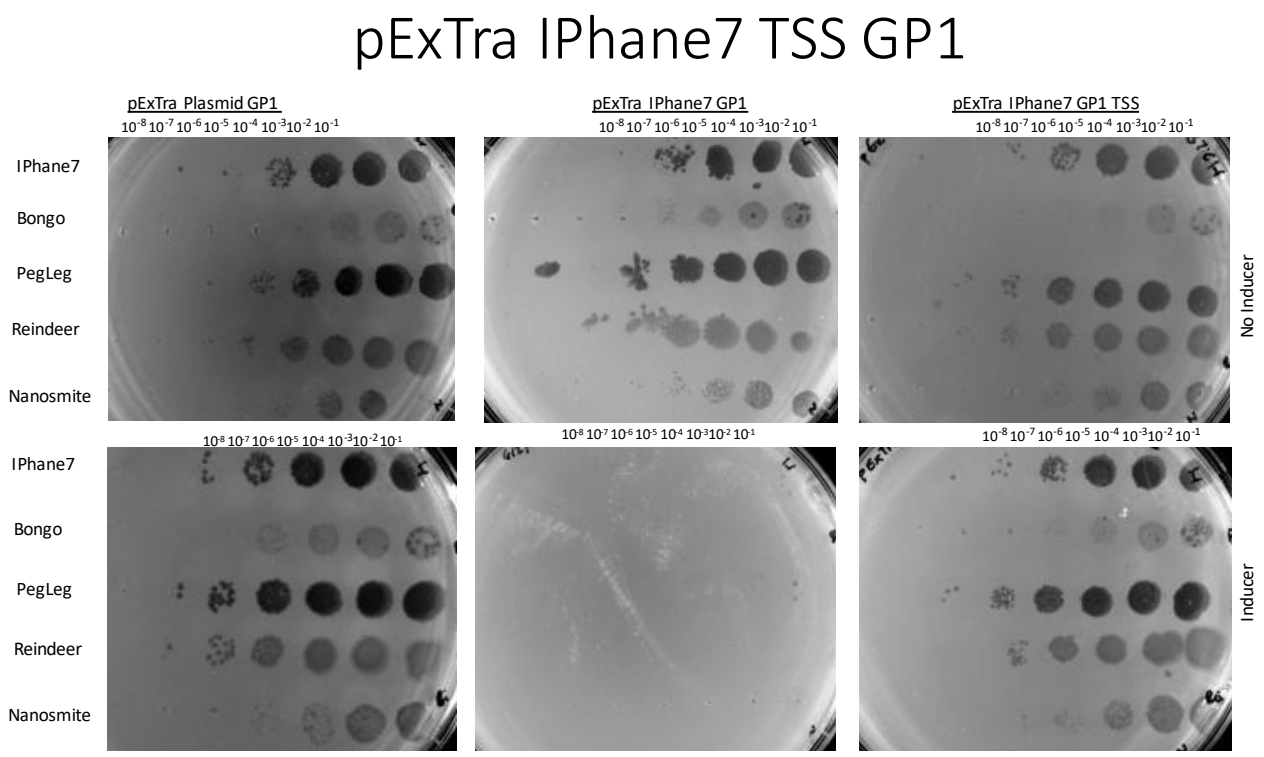



Figure AI. pExTra IPHane7 TSS gp1 Viral Panel challenge results. pExTra vector clone was incubated at 37°C on a shaker for 48 hours in 20% tween and 7H9Neat mixture, and then incubated on a shaker with no 20% tween in a 7H9 mixture that either contained 10ng/mL aTc or no aTc for 48 hours. Upon completion of the 48 hours, the mixtures that contained 10ng/mL aTc were mixed with TA containing 10ng/mL aTc and plated on 7H9+Kan plates and allowed to gel. Similarly, the mixtures that did not contain any inducer were mixed with TA and plated on 7H9+Kan plates. The plates were spotted with serial dilutions of Viral Panel member phages and incubated at 37°C for 72 hours and then imaged.

Table A1. Genbank accession numbers of phages used in the study.

| Bacteriophage: | Genbank Accession Number: |
|---|---|
| IPhane7 (M1) | MH697587.1 |
| Reindeer (M1) | MT658803.1 |
| PegLeg (M1) | KC900379.1 |
| Rey (M2) | JF937105.1 |
| Nanosmite (M3) | MW578836.1 |
| *Mycobacterium smegmatis mc$^2$155* | CP009494.1 |
| Bongo (M1) | JN699628.3 |
| Charlie (N) | JN256079.1 |
| Xeno (N) | KU935728.1 |
| Phayonce (P5) | KR080195.1 |
| Island3 (I1) | HM152765.1 |
| Che9C (I2) | AY129333.1 |

Table A2. Summation of viral challenge results from Figures 15, 30, 32, 33. (A) Summation of IPhane7 viral challenge results evaluated by log inhibitions. (B)Summation of Reindeer viral challenge results evaluated by log inhibitions. (C) Summation of PegLeg viral challenge results evaluated by log inhibitions. (D) Summation of Rey viral challenge results evaluated by log inhibition. (E) Summation of Nanosmite viral challenge results evaluated by log inhibition. **Bongo samples reported in the tables exhibit a different scaling, as the viral stock grew poorly during challenge experiments. Thus, full inhibition is observed at 10^-4 (Lysogen, pMH94 gp1/2, pExTra gp1) and 10^-3 (pMH94 gp1) and used as the baseline for full inhibition in the respective experiments.*

**A**

| IPhane7 (M1) | Viral Challenge: | | | |
|---|---|---|---|---|
| Virus | Lysogen | pMH94 gp1 | pMH94 gp1/2 | pExTra gp1 |
| IPhane7 | 1.0E-06 | 1.0E-01 | 1.0E-01 | 1.0E-06 |
| Bongo* | 1.0E-04 | 1.0E-02 | 1.0E-04 | 1.0E-04 |
| PegLeg | 1.0E-03 | 1.0E-01 | 1.0E-01 | 1.0E-06 |
| Reindeer | 1.0E-06 | 1.0E-03 | 1.0E-06 | 1.0E-06 |
| Rey | 1 | 1 | 1 | 1.0E-06 |
| Nanosmite | 1.0E-06 | 1 | 1.0E-06 | 1.0E-06 |

**B**

| Reindeer (M1) | Viral Challenge: | | | |
|---|---|---|---|---|
| Virus | Lysogen | pMH94 gp1 | pMH94 gp1/2 | pExTra gp1 |
| IPhane7 | 1.0E-06 | 1 | 1.0E-01 | 1.0E-06 |
| Bongo* | 1.0E-04 | 1.0E-01 | 1.0E-04 | 1.0E-04 |
| PegLeg | 1.0E-06 | 1.00 | 1.0E-01 | 1.0E-06 |
| Reindeer | 1.0E-06 | 1.0E-02 | 1.0E-06 | 1.0E-06 |
| Rey | 1.0E-02 | 1 | 1 | 1.0E-06 |
| Nanosmite | 1.0E-06 | 1 | 1.0E-02 | 1.00 |

**C**

| PegLeg (M1) | Viral Challenge: | | | |
|---|---|---|---|---|
| Virus | Lysogen | pMH94 gp1 | pMH94 gp1/2 | pExTra gp1 |
| IPhane7 | 1.0E-06 | n/a | n/a | n/a |
| Bongo* | 1.0E-04 | n/a | n/a | n/a |
| PegLeg | 1.0E-06 | n/a | n/a | n/a |
| Reindeer | 1.0E-06 | n/a | n/a | n/a |
| Rey | 1 | n/a | n/a | n/a |
| Nanosmite | 1.0E-06 | n/a | n/a | n/a |

**D**

| Rey (M2) | Viral Challenge: | | | |
|---|---|---|---|---|
| Virus | Lysogen | pMH94 gp1 | pMH94 gp1/2 | pExTra gp1 |
| IPhane7 | 1.0E-06 | 1 | 1 | 1 |
| Bongo* | 1.0E-04 | 1 | 1 | 1 |
| PegLeg | 1.0E-04 | 1 | 1 | 1 |
| Reindeer | 1.0E-05 | 1 | 1 | 1 |
| Rey | 1.0E-06 | 1 | 1 | n/a |
| Nanosmite | 1.0E-06 | 1 | 1 | 1 |

**E**

| Nanosmite (M3) | Viral Challenge: | | | |
|---|---|---|---|---|
| Virus | Lysogen | pMH94 gp1 | pMH94 gp1/2 | pExTra gp1 |
| IPhane7 | 1.0E-06 | 1 | 1 | 1 |
| Bongo* | 1.0E-04 | 1 | 1 | 1 |
| PegLeg | 1 | 1 | 1 | 1 |
| Reindeer | 1.0E-05 | 1 | 1 | 1 |
| Rey | 1.0E-06 | 1 | 1 | n/a |
| Nanosmite | 1.0E-06 | 1 | 1.0E-03 | 1.0E-06 |

| 1 = viral full infection | 10E-1 to 10E-2 = Mild Inhibition | 10E-3 to 10E-4 = Moderate Inhibition | 1.0E-5 to 1.0E-7 = Full Inhibition |
|---|---|---|---|