# IDENTIFICATION OF A NOVEL TRANSCRIPTIONAL REPRESSOR GENE IN KUMAO, A GENETICALLY UNIQUE BACTERIOPHAGE

A thesis presented to the faculty of the Graduate School of
Western Carolina University in partial fulfillment of the
requirements for the degree of Master of Science in Chemistry

By

Lori Marie Neri

Director: Dr. Maria Gainey
Assistant Professor of Biochemistry
Chemistry and Physics Department

Committee Members: Dr. Carmen Huffman, Chemistry and Physics Department
Dr. Jamie Wallen, Chemistry and Physics Department

March 2020

# ACKNOWLEDGMENTS

I would like to thank my committee members and director for their encouragement and support the last two years. I would also like to thank the Howard Hughes Medical Institute, the Hatful lab at the University of Pittsburgh, Viknesh Sivanathan and Daniel Heller at the Howard Hughes Medical Institute, without whom this thesis would not be possible. Lastly, I would like to thank my parents and sister for their continued support and always believing in me. Thank you for being my biggest supporters and always being there for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# ABSTRACT

IDENTIFICATION OF A NOVEL TRANSCRIPTIONAL REPRESSOR GENE IN KUMAO, A GENETICALLY UNIQUE BACTERIOPHAGE

Lori Marie Neri, M.S.

Western Carolina University (March 2020)

Director: Dr. Maria Gainey

Kumao is a genetically unique bacteriophage that infects *Mycobacterium smegmatis (M. smegmatis)* and was selected for further study because of its noncanonical genome arrangement. As an example, integrase genes are typically located at the center of Mycobacteriophage genomes. However, Kumao and cluster M bacteriophages contain an integrase gene located on the right side of the genome. Kumao and cluster M bacteriophages also contain a leftward transcribed operon containing genes of unknown function directly upstream of their structural genes. Kumao lysogens were successfully generated, proving that Kumao is indeed a temperate bacteriophage, but bioinformatics has failed to reveal an obvious repressor gene candidate. Repressor genes are typically located at a transcriptional directional change and are the first gene of a leftward transcribed operon. Repressor genes are also usually located near the integrase gene and contain an easily predicted helix-turn-helix DNA binding motif.
Current work is focused on identifying the location of Kumao's repressor gene using two different functional screens. The first involves overexpressing individual Kumao genes in *M. smegmatis* and performing a viral challenge experiment. If the gene is the repressor then productive viral replication should be inhibited. This approach was recently used to identify the

repressor gene of cluster M bacteriophages, gene 1. Gene 1 is located at the end of the leftward transcribed operon upstream of the structural genes. Twenty-one Kumao genes were selected for testing, based on their proximity to the integrase gene, nearness to a transcriptional direction change, or potential promoter region. Excitingly, the results reveal that genes 5 and 44 have been found to be promising repressor candidates. The second approach involves CRISPRi of Kumao's operons during lysogeny. If the operon containing the repressor is silenced, Kumao should reenter the lytic replication cycle. CRISPRi vector constructs that will target Kumao's operons and a positive control targeting IPhane7's repressor operon have been successfully generated. The CRISPRi system constructs were validated using genes 1 and 2 from IPhane7. However, further testing is needed to optimize the CRISPRi system.

# CHAPTER ONE: INTRODUCTION

Bacteriophages, or phages, are the most abundant and diverse life forms on earth. Bacteriophages are viruses that infect bacteria and can either undergo a lytic or lysogenic replication cycle. Both cycles begin when a viral particle attaches to receptors located on the bacterial cell surface. The virus then injects its DNA inside the host cell, and it is after this event that the lytic and lysogenic cycles diverge.[1,2] During lytic replication, viral DNA will serve as the template for transcription, and many new viral particles will be produced resulting in cell death.[3] In the lysogenic replication cycle, the injected DNA integrates into the host's genome, (which is called a prophage, and the bacterial cell containing a prophage is known as a lysogen) and remains dormant for long periods of time.[1] While remaining dormant, the host is unaware of the invasion and will undergo cell division like normal, with the prophage transmitted to newly produced cells. This process will continue until the lytic replication cycle is induced.[1] Phages that undergo lysogeny are also referred to as temperate bacteriophages and these phages typically have an integrase and a repressor protein. The integrase is responsible for integrating viral DNA into the hosts DNA, and the repressor is responsible for stopping transcription of viral genes that allows the phage to enter the lytic replication cycle.[4] Due to their coevolution with their hosts, bacteriophages are also great resources for understanding the evolution and genetic diversity of the bacterial cells they infect.[5,6]

Mycobacteriophages are viruses that infect Mycobacterium hosts such as *M. smegmatis* and *M. tuberculosis*. Over 1,877 Mycobacteriophages have been sequenced and archived through the Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Sciences (SEA-PHAGES) program, resulting in the largest collection sequences of

phages that infect a single host currently in exhistence.[4,6,7] Most of the Mycobacteriophage genomes have a canonical structure in which the genome starts with the assembly and structural genes transcribed in the forward (rightward) direction. If it is a temperate phage both the integrase and the repressor are found relatively near each other in the center of the genome, typically near a transcription directional change.[4] However, as was first discovered in Cluster M phages this is not always the case and that noncanonical genome arrangements do exist. Cluster M phage genomes begin with a leftward transcribed operon containing genes of unknown function. In addition, unlike other temperate bacteriophages, the integrase protein is located in the right arm of the genome, and no repressor gene can be called using bioinformatics.[4] IPhane7 is a Cluster M bacteriophage discovered at Western Carolina University that has recently been under study in the Gainey laboratory by Erin Cafferty. Excitingly, Erin was able to recently identify a novel transcriptional repressor in the IPhane7 genome.[8]

It was then sought to determine if other Mycobacteriophages also exhibited this noncanonical arrangement. The genome architecture of every cluster of Mycobacteriophage was examined using the Actinobacteriophage database[9] and Phamerator[10] Bacteriophage Kumao, a singleton, was discovered and annotated at Lehigh University in Pennsylvania.[9] Kumao is considered a "singleton" because it was unable to be classified into a cluster with other Mycobacteriophages because it does not contain greater than fifty-percent nucleotide similarity to any other sequenced bacteriophages. However, upon further inspection of Kumao's genome architecture it was observed that Kumao did exhibit some similarities to Cluster M bacteriophages. As an example, Kumao's genome also begins with a leftward transcribed operon, and an integrase gene is found on the right arm of Kumao's genome.

A visual representation of Kumao's genome is shown below in Figure 1.

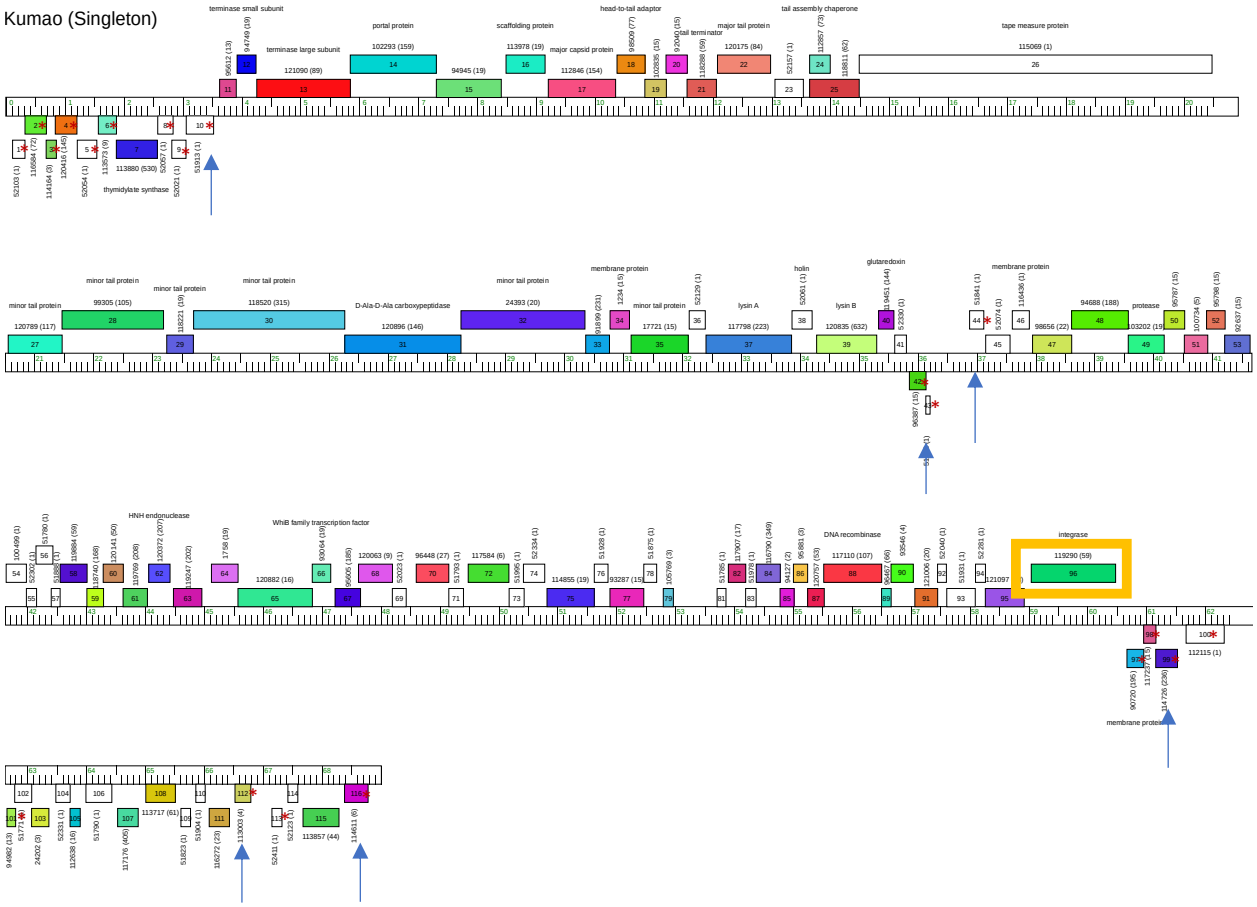


Figure 1. A Phamerator map of Kumao's genome divided into four tiers. Each box represents a gene in Kumao's genome. The ruler indicates how long the genome is, and each number represents 1,000 base pairs. The genes located at the bottom of the ruler is moving left in the reverse direction and the genes located above the ruler is moving right in the forward direction. The different colors of the genes indicate that those genes are found in other bacteriophages and the genes that are white indicate they are unique to that bacteriophage. Stars represent genes that were selected for testing in the overexpression assay, while arrow represent areas of the genome that CRISPRi targets were generated against. The yellow box indicates the integrase protein.

Kumao was selected for further study because temperate phages make up most of the bacteriophage population. However, little is known about how they control and regulate their lytic verses lysogenic replication cycles.[4] Control of gene expression is essential for biological systems to function properly. The more these phages are studied, especially the unique ones, the

more we will learn about the evolution of gene expression control mechanisms.[5]

There is currently a rise of antibiotic resistant bacteria, and new alternatives are needed to treat these pathogens. One of these alternatives is phage therapy, because phages are naturally antibacterial and are active against even multi-drug resistant bacteria. Phage therapy is more specific than existing drugs and is able to limit the death to non-targeted bacteria.[1,11] However, current limitations of phage therapy include that the exact microorganism causing the infection along with a panel of phages that infect it needs to be known for accurate treatment. Also, therapy is currently limited to the use of lytic phages, or temperate bacteriophages that have been genetically altered such that they can no longer enter into the temperate cycle.[11,12] If their repressor systems are not known or fully understood, these viruses cannot be genetically engineered for bacteriophage therapies. The goal of my thesis work is to prove that Kumao is a temperate bacteriophage, identify a novel repressor system in bacteriophage Kumao, and develop a CRIPSRi method as another way to indentify repressors.

# CHAPTER TWO: METHODS

## Viral Stock Growth

Bacteriophage Kumao was discovered by a student at Lehigh University in Easton, PA through the SEA-PHAGES program in 2015.[8] Bacteriophage IPhane7 was discovered at Western Carolina University in 2016 by Dylan Rood from Dr. Charles Marth's compost. Bacteriophage Larva was discovered at the college of William and Mary in Williamsburg, VA in 2010.[13] Bacteriophage Nanosmite was discovered at Del Mar College in Corpus Christi, TX in 2015.

All viruses except IPhane7 were obtained from the Hatfull laboratory SEA-PHAGES archive. IPhane7 was obtained from the Western Carolina University bacteriophage archive. Viruses were plaque picked once for purity. Viral stocks were created by flooding webbed plates with phage buffer followed by 0.22 µM filtration as described by the SEA-PHAGES laboratory protocols.[14] All virus stocks were stored at 4 °C.

## Lysogen Creation and Testing

A Kumao lysogen was created by following the Lysogeny experiment protocol.[15] Briefly, bacteriophage Kumao was serially diluted and spotted onto a top agar lawn containing *M. smegmatis* cells. This plate was inverted and incubated at 37 °C until mesas were observed. Mesas are an overgrowth of bacterial cells. Bacteria in the mesas were then streaked onto fresh plates, and 6-8 colonies were selected for a second round of streaking, followed by patch testing. During a patch test, potential lysogen cells are streaked onto a top agar layer containing M. *smegmatis* and should show evidence of virus "leakage" from these cells. Colonies that had positive patch tests were grown in liquid culture and a Kumao viral challenge experiment was performed to confirm that they were indeed Kumao lysogen cells.[15]

## pSMEG Overexpression Vector

The pSMEG overexpression vector was obtained from Danielle Heller (Howard Hughes Medical Institute). Individual viral genes were amplified by PCR from boiled virus stocks using Q5 Hot Start High Fidelity 2x Master Mix (New England Biolabs) protocols. Forward and reverse primers were designed to amplify the indicated viral genes from start codon to stop codon and are shown in Table 1. If the gene start codon was not ATG, the forward primer was altered such that the start codon in the final construct would be changed to ATG. The pSMEG vector sequence 5'-ATGCGGAGGAATCACTTCCAT-3' was then added to the 5'end of the forward primer and 5'-TGCAGGATCCGACTCGAGTGTCGAC-3' to the 5'end of the reverse primer to enable HiFi assembly of amplified genes into the linearized pSMEG vector.[16] The pSMEG vector was linearized and amplified using primers shown in Table 1. All primers were ordered from Integrated DNA Technologies. PCR products were size-verified by gel electrophoresis and purified following protocols from New England Biolabs Monarch PCR and DNA clean up kit. Individual genes were then ligated into the linearized pSMEG vector following the HiFi assembly master mix protocols from New England Biolabs.[16]

Plasmids were transformed following manufacture protocols into NEB 5-alpha competent *E. coli* cells and plated onto Luria Broth (LB) agar plates plus 50 µg/mL kanamycin. Successful transformants were then identified using colony PCR.[16] Successful colonies were amplified by using 5 mL overnight LB 50 µg/mL kanamycin cultures and plasmids purified following protocols from New England Biolabs Monarch Plasmid Miniprep. These plasmids were quantified using a Thermo Scientific Nanodrop 2000 Spectrophotometer and sequences were verified using Sanger sequencing.[16]

Table 1. pSMEG and Kumao gene primers used to perform the overexpression assays. All primers are listed in the 5' to 3' direction, and the pSMEG vector sequence for each gene is highlighted in yellow.

| Gene | Forward Primer Sequence 5'-3' | Reverse Primer Sequence 5'-3' |
|---|---|---|
| pSMEG | ATGCGGAGGAATCACTTCCAT | TGCAGGATCCGACTCGAGTGTCGAC |
| Colony PCR pSMEG | GTACGACCAGCACGGCATACATC | CTCCTTGATGATCGCCATGTTGTCC |
| 1 | ATGCGGAGGAATCACTTCCATATGGCAGACGAGGTCCAGTTCCC | TGCAGGATCCGACTCGAGTGTCGACTCACGCCGCCTCCCGG |
| 2 | ATGCGGAGGAATCACTTCCATATGACCACACCCGTCACCAAG | TGCAGGATCCGACTCGAGTGTCGACTCAGCGCTCCCAGACGG |
| 3 | ATGCGGAGGAATCACTTCCATATGAGCAGGCACACGCGTGAG | TGCAGGATCCGACTCGAGTGTCGACTCACAGGTGTCCTTCCAGAATGGC |
| 4 | ATGCGGAGGAATCACTTCCATATGCCCATGGCAAGAGTCACTGAGATCAAG | TGCAGGATCCGACTCGAGTGTCGACTCACGCGTGTGCCTGCTCATA C |
| 5 | ATGCGGAGGAATCACTTCCATATGAAGAAAGCGCTCACTGCCACAG | TGCAGGATCCGACTCGAGTGTCGACTCAAGTGATCGACGCTGGAAAAACGGTATC |
| 6 | ATGCGGAGGAATCACTTCCATATGAAACTGTCAGCAATGAGCACCGAG | TGCAGGATCCGACTCGAGTGTCGACTCACTTATCCACAGGCATGTGGATAATCGA |
| 8 | ATGCGGAGGAATCACTTCCAT ATGAGCACGAAGACATCGTTTAGGGATCG | TGCAGGATCCGACTCGAGTGTCGACTCAGAAGTCAGCGAGAATGCGCTTCAGGAC |
| 9 | ATGCGGAGGAATCACTTCCATATGAAGTGGACCAACGAATCCGCCCAG | TGCAGGATCCGACTCGAGTGTCGACTCAAACGATGTCTTCGTGCTCACTGGC |
| 10 | ATGCGGAGGAATCACTTCCATATGGAGCTGACCGAAGTTAG | TGCAGGATCCGACTCGAGTGTCGACTCACTCCTGGGGGAGAACC |
| 42 | ATGCGGAGGAATCACTTCCATATGAAGCAATTTGACACACGGGTCG | TGCAGGATCCGACTCGAGTGTCGACTCAGGAACCGTCCTCTTCCACC |
| 43 | ATGCGGAGGAATCACTTCCATATGCCAAACATGGACTACAACCACGTCAAG | TGCAGGATCCGACTCGAGTGTCGACTCACCCGATCTCCTCTCCGGTC |
| 44 | ATGCGGAGGAATCACTTCCATATGGCCGAACGCACTGCTG | TGCAGGATCCGACTCGAGTGTCGACTCATTGGGCCACGAGGTTGGC |
| 97 | ATGCGGAGGAATCACTTCCATATGACCAACTACCCGTACTACCCGCAG | TGCAGGATCCGACTCGAGTGTCGACTCACCGGTTCTTGACGGCCATCGC |
| 98 | ATGCGGAGGAATCACTTCCATATGAGACGCGAGTGTCCGACG | TGCAGGATCCGACTCGAGTGTCGACTCATCTCTCTCCTTCGTATTCGAATCGTGG |
| 99 | ATGCGGAGGAATCACTTCCATATG CCGACTGATCACCAGATCG | TGCAGGATCCGACTCGAGTGTCGACTCATGCGGCCTCAGCCTC |
| 100 | ATGCGGAGGAATCACTTCCATATGGTCGCCATGGCCATCAG | TGCAGGATCCGACTCGAGTGTCGACTCACACCACACTCCTTTCCGCAATTC |
| 101 | ATGCGGAGGAATCACTTCCATATGGCGGCGATGAATGAGAGGGAGG | TGCAGGATCCGACTCGAGTGTCGACTCAGGCAGGCTTGCGATCAAGC |
| 112 | ATGCGGAGGAATCACTTCCATATGGAAATGAGCAATCTGTGGGTGGAAGCTG | TGCAGGATCCGACTCGAGTGTCGACTCAAGCGGTGAGGGGTCGACGGTT |
| 113 | ATGCGGAGGAATCACTTCCATATGACACTGGCACGCATTGCTTTCG | TGCAGGATCCGACTCGAGTGTCGACTCATCGCGCAGTGCGTTCCT |
| 114 | ATGCGGAGGAATCACTTCCATATGGCACGCACTGAGACTTACGCG | TGCAGGATCCGACTCGAGTGTCGACTCACGAATTCGCCCAGAGAACGC |
| 116 | ATGCGGAGGAATCACTTCCATATGAACATCAATACCGTTGATCAGCATTTCGCTA | TGCAGGATCCGACTCGAGTGTCGACTCATGCCGCTTGCCGTGC |

## CRISPR interference: pLJR965 Cloning

The pLJR965 plasmid was a kind gift from the Fortune laboratory. Forward and reverse primers were designed to target the operon indicated in Table 2 according to Fortune lab protocol version 4.[17] Briefly, a PAM sequence was identified for the non-template strand for each sgRNA sequence to maximize gene knockdown.[17] The sgRNA targeted sequence is ~20 nucleotides in length, with the last nucleotide being either an A or G. The pLJR965 vector sequence 5'-GGGA-3' was then added to the 5' end of the forward primer and 5'-AAAC-3' was added to the 5'end of the reverse primer to enable insertion via gateway assembly into the pLJR965 vector. Forward and reverse primers used are shown in Table 2. Individual colonies were picked and amplified using liquid culture and then plasmids were purified using New England Biolabs Monarch Plasmid Miniprep kit. Plasmid DNA was quantified using a Thermo Scientific Nanodrop 2000 Spectrophotometer and sequences were verified using Sanger sequencing.[17]

Table 2. CRISPRi, Kumao, and Iphane7 gene primers used to perform CRISPRi. All primers are listed in the 5' to 3' direction and highlighted in yellow is the CRISPRi vector sequence added for each gene.

| Gene | PAM Sequence 5'-3' | Forward Sequence 5'-3' | Reverse Sequence 5'-3' |
|---|---|---|---|
| CRISPRi | Not Applicable | GGGA | AAAC |
| 10 | TGAGAAC | GGGAGCTCCAGCGGGTTCACCTGGAAGA | AAACTCTTCCAGGTGAACCCGCTGGAGC |
| 43 | CCAGAAC | GGGAATCTCCTCTCCGGTCAGACGGCCACCGAT | AAACATCGGTGGCCGTCTGACCGGAGAGGAGAT |
| 44 | GTAGAAC | GGGAAGAGCACGGTTGTAACCGTC | AAACGACGGTTACAACCGTGCTCT |
| 99 | GTAGAAG | GGGAACGCGCCGGATCTTCTCGGG | AAACCCCGAGAAGATCCGGCGCGT |
| 112 | CCAGGAC | GGGAGCCAAACGCGTTGTTGATAG | AAACCTATCAACAACGCGTTTGGC |
| 116 | GAAGAAC | GGGAAGCTCGCCGCGTTCCTGGGC | AAACGCCCAGGAACGCGGCGAGCT |
| Iphane7 Gene 1 | GAGGAG | GGGAGTTCTCATGAGCCTTCACAGT | AAACACTGTGAAGGCTCATGAGAAC |
| Iphane7 Gene 2 | CAGGAAT | GGGAGTCTCCCCCGTGCGGGAGTGAATCCG | AAACCGGATTCACTCCCGCACGGGGGAGAC |

## pSMEG Overexpression Assay

Approximately 100 ng of pSMEG DNA alone, or pSMEG containing the indicated gene was electroporated into electrocompetent *M. smegmatis* cells.[16] A volume of 2.5 µL of plasmid DNA was added to 50 µL of cells and incubated on ice for 10 min. The cells were then transferred into a 1mm electroporation cuvette, and a single electric pulse at 1.8 kV, with no time constant, was delivered using a BioRad GenePulser Xcell system.[16] Cells were then transferred from the cuvette to a microcentrifuge tube using 1 mL of 7H9 complete, which consisted of 7H9 neat, AD supplement, and 100 mM concertation of $CaCl_2$. The cells were incubated with shaking at 37 °C at 250 rpm for 2 hr and then 150-250 µL of the cells were plated on to 7H9 agar plates with 5 µg/mL of kanamycin. The plates were incubated for 4-5 days at 37 °C to allow colony growth. After successful colony growth, a 5 mL culture with and without the inducer, 5 ng/mL anhydrous tetracycline (aTC), was grown for two days.[16] All cultures included 7H9 complete media, 5 µg/mL of kanamycin, 12.5 µL 20% tween80, and a single bacterial colony. A bacterial lawn was then made with 500 µL of bacterial culture, mixed with 4.5 mL of 7H9 top agar plus 5 µg/mL kanamycin, and plus or minus 5 ng/mL anhydrous tetracycline. The top agar mixture was plated onto 7H9 plates plus kanamycin (5 µg/mL) plus or minus aTC (5 ng/mL).

About 2.5 µL of ten-fold serial dilutions of the indicated viruses were then plated onto the top agar layers. Plates were then incubated at 37 °C for two days.[16]

## Kumao Gene Toxicity Testing

A single bacterial colony containing pSMEG, or pSMEG plus the indicated Kumao gene, was inoculated into 500 µL of 7H9 neat media and then resuspended. Ten-fold serial dilutions of these cells were performed using 7H9 neat media ($10^{-1}$ to $10^{-5}$ log), and 5 µL was spotted onto 7H9 agar plates plus kanamycin (5 µg/mL) with or without aTC (5 ng/mL) plates. Kanamycin is used as an antibiotic. The plates were then incubated at 37 ºC for 5 days.[16]

## CRISPRi

100 ng of pLJR965 DNA was electroporated into electrocompetent IPhane7 lysogen cells as described above for the pSMEG plasmids.[16] After successful colony growth, a 5 mL culture with 7H9 complete media plus 2.5 µL 20% tween80 and kanamycin (5µg/mL) was incubated for 2 days at 37 °C.[17] A 50-fold dilution of this culture was added to a 125 mL baffled flask containing 25 mL of 7H9 complete media, 2.5 kanamycin (5 µg/mL), and plus or minus aTC (5 ng/mL). Before incubating, a time point of $t_0$ was recorded by removing 1 mL of culture and collecting the supernatant after 1 minute of max speed centrifugation. Subsequent time points were taken once a day for 5 days and stored at 4 °C. Spot titers were performed in triplicate for each time point. The spot test was performed on LB agar plates with a bacteria lawn consisting of *M. smegmatis* and top agar. Approximately, 2 µL of sample was spotted and incubated at 37 °C for 1-2 days.[16]

## Creating Electrocompetent IPhane7 Lysogen Cells

Fresh colonies of IPhane7 lysogen cells were grown from an IPhane7 lysogen glycerol

stock created by former student Erin Cafferty. Approximately, 3 mL liquid cultures (7H9 complete) were grown from a single bacterial colony for 2 days at 37 °C. After 2 days the culture was inoculated into a 250 mL baffled flask containing 50 mL of 7H9 complete media to an optical density of 600 ($OD_{600}$) and incubated with shaking at 37 °C overnight. Once the $OD_{600}$ reached between 0.8-1, cells were pelleted using the Beckman Coulter Allegra X-I5R Centrifuge at 4,500 rpm for 10 min at 4 °C. The cells were washed and resuspended with 10% cold glycerol. This process was repeated four times to ensure all salts were removed. Cells were resuspended in 10% cold glycerol and aliquoted 1 mL into Eppendorf tubes, then flash frozen using dry ice and stored at -80 °C until electroporation.

# CHAPTER THREE: RESULTS

## Lysogen Creation and Testing

To determine if Kumao is a temperate bacteriophage, Kumao lysogens were attempted to be made. When ten-fold serial dilutions of Kumao were spotted onto *M. smegmatis* lawns mesas (islands of bacterial growth in the presence of high virus concentrations) readily formed after several days of incubation. An example of a mesa formed by Kumao is shown below in Figure 2. Cells from this mesa were then subject to several rounds of purification and tested for the presence of virus using a patch test shown below in Figure 3. As shown in Figure 3, candidate cells 1, 5, 8, 10, and 11 show evidence of small virus plaques around the area where the cells were streaked, indicating that these could be lysogen cells. Positive lysogen candidate cells from the patch test were then used to create a bacterial lawn onto which serial dilutions of Kumao, IPhane7, and Larva viruses were spotted. Larva was used as a control because it has been previously studied in the Gainey Laboratory and has a canonical temperate bacteriophage genome architecture. If the cells were indeed a Kumao lysogen, they should be resistant to cell death from bacteriophage Kumao (homoimmunity) but not bacteriophages IPhane7 and Larva. An example of the results from colony 10 are shown in Figure 4. As expected, bacteriophage Kumao was able to readily infect *M. smegmatis* control cells, but not a lawn of the potential Kumao lysogen cells, indicating that these cells are indeed likely to be Kumao lysogens and that Kumao is a temperate bacteriophage.

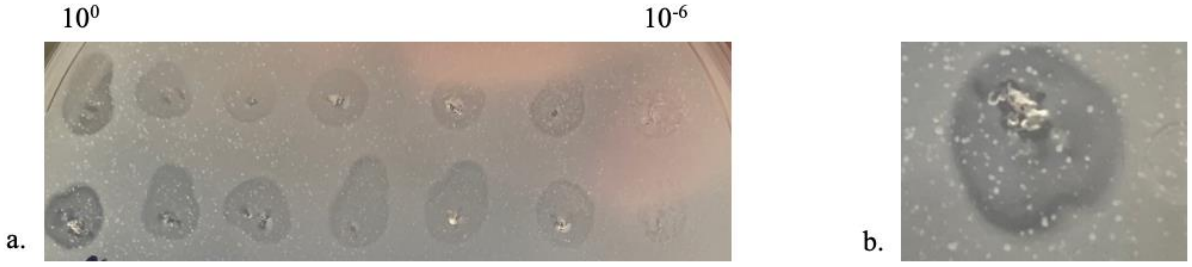


Figure 2a. Ten-fold serial dilution of Kumao. Dilutions were spotted in duplicate onto a top agar layer containing *M. smegmatis*. A close up of the bottom $10_0$ mesa is shown in 2b. The puncture in the agar shows where potential lysogen cells were collected from the mesa for further purification and testing.

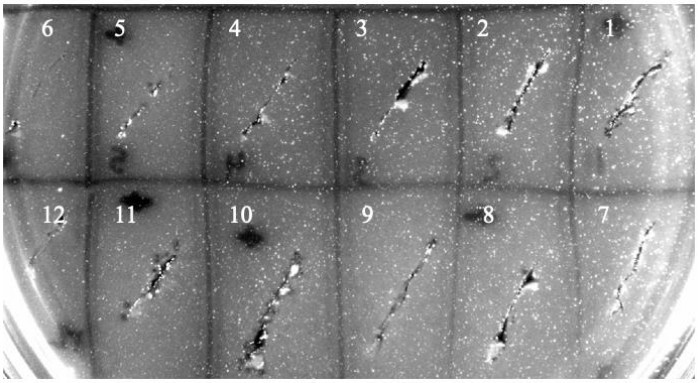


Figure 3. Patch test of twelve different bacterial colonies streaked from the mesa. Bacteria from colonies 1-12 were individually streaked onto a top agar layer containing *M. smegmatis* cells as indicated by the number above the streak. The dark circles around the streak as seen in 1, 5, 8, 10, and 11 indicate cell lysis, where virus is leaking from lysogen cells.

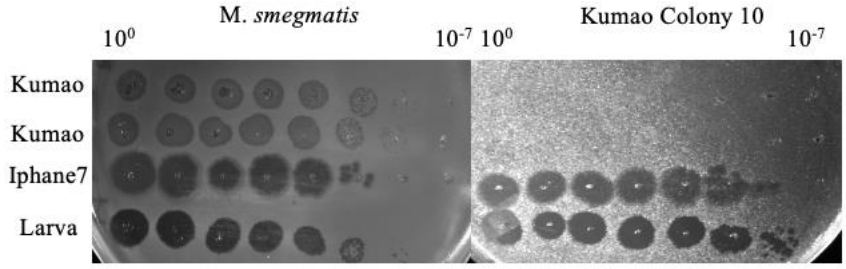


Figure 4. Kumao lysogen challenge experiment. Bacteriophages Kumao, IPhane7, and Larva were serially diluted and spotted onto top agar layers containing control *M. smegmatis* cells or lysogen candidate 10 cells.

## pSMEG Overexpression Assay

After successfully proving that Kumao is indeed a temperate bacteriophage, twenty-one repressor gene candidates were selected for testing. Candidate genes were selected based on

previous knowledge of repressor location, such as proximity to the integrase gene and switches in transcriptional direction. The non-canonical genome location of IPhane7's recently discovered repressor gene was also taken into account.[8] Candidate genes selected for testing are shown in Table 1. Figure 5 shows a representative experiment from this experimental series. Candidate genes were individually cloned into the pSMEG overexpression vector and electroporated into *M. smegmatis* cells. Liquid cultures of cells containing pSMEG (negative control), pSMEG plus a candidate gene, or pSMEG plus the repressor gene from bacteriophage Larva (positive control) were then grown in the presence or absence of inducer. The indicated viruses were then serially diluted and spotted onto plates with top agar layers of these cells. Gene 42 from bacteriophage Larva performed as expected and reduced the growth of bacteriophage Larva by ~3 logs even in the absence of inducer, and 6 logs in the presence of inducer, but had no effect on the growth of bacteriophages Kumao or Iphane7. Excitingly, both gene 5 and 44 from Kumao were able to inhibit the growth of bacteriophage Kumao. A summary of the results from the remaining genes tested can be found in Table 3.

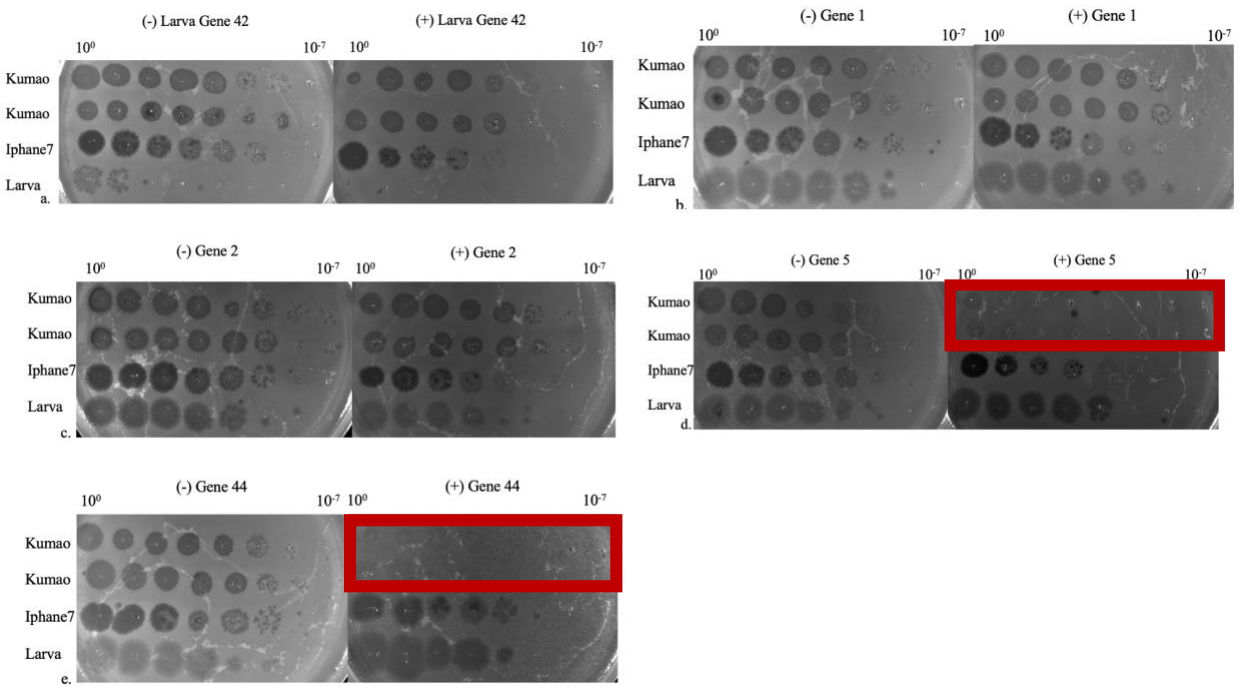


Figure 5a. Bacteriophages Kumao, IPhane7, and Larva were serially diluted and spotted onto top agar layers of *M. smegmatis* cells containing pSMEG plasmids plus the indicated Kumao genes, in the presence of absence of an inducer. Kumao was spotted in duplicate. Bacteriophage Larva's repressor, gene 42 was used as a positive control (a). The (-) indicates without inducer, the (+) indicates plus inducer. The images located on the left side without inducer, all viruses should produce plaques, which is indicated by the dark circles. The images on the right-hand side are with inducer and if the repressor is present then no plaques will be formed because bacteriophages are homoimmune to themselves, meaning they cannot infect themselves. For 5a Larva was a control and you can see that gene 42 is its repressor because Larva produced no plaques with inducer. For Kumao's genes 1 and 2, plaques appeared for both genes indicating that those genes are not the repressor. Gene 5 and 44 Kumao produced very faint plaques, and the red boxes indicate the area of interest for both genes. Since both produced very faint plaques, they are possible candidates for being Kumao's repressor.

Table 3. Kumao repressor gene candidates tested.

| Gene | Transcriptional Direction | Proximity to the Integrase | Repressor | Sequenced | Number of Experiments Performed | Notes |
|---|---|---|---|---|---|---|
| 1 | Reverse | Far Away | No | Yes | 2 | |
| 2 | Reverse | Far Away | No | Yes | 3 | |
| 3 | Reverse | Far Away | Could not test | Yes | 1 | Could not grow cultures with and without inducer |
| 4 | Reverse | Far Away | No | Yes | 2 | |
| 5 | Reverse | Far Away | Maybe | Yes | 3 | |
| 6 | Reverse | Far Away | No | Yes | 2 | |
| 8 | Reverse | Far Away | No | Yes | 1 | |
| 9 | Reverse | Far Away | No | Yes | 1 | |
| 10 | Reverse | Far Away | Could not test | Yes | 1 | Did not grow culture with inducer; Is toxic. |
| 42 | Reverse | Far Away | No | Yes | 2 | |
| 43 | Reverse | Far Away | No | Yes | 1 | |
| 44 | Forward | Far Away | Maybe | Yes | 3 | |
| 97 | Reverse | Very Close | No | Yes | 2 | Could not grow cultures with and without the inducer |
| 98 | Reverse | Very Close | Could not test | Yes | 2 | |
| 99 | Reverse | Very Close | Could not test | Yes | 2 | Could not grow cultures with the inducer |
| 100 | Reverse | Very Close | No | Yes | 1 | |
| 101 | Reverse | Very Close | No | Yes | 1 | |
| 112 | Reverse | Close | No | Yes | 1 | |
| 113 | Reverse | Close | No | Yes | 1 | |
| 116 | Reverse | Close | Could not test | Yes | 2 | Did not grow any colonies when electroporated into Mycobacterium Smegmatis |

## Gene Toxicity

In addition to repressor testing, a toxicity test was performed using the bacterial colonies containing pSMEG plus individual Kumao genes. The bacterial colonies for each are resuspended in media and then serially diluted onto plates with and without the inducer. Figure 6 shows example results for genes 1, 10, 99, and 100. An absence of bacterial growth in the presence of the inducer such as is seen for gene 10 indicates that expression of the gene may be

toxic. Gene 10 also has very minimal bacterial growth without inducer, further indicating its toxicity. A summary of the results from all genes tested can be found in Table 4.

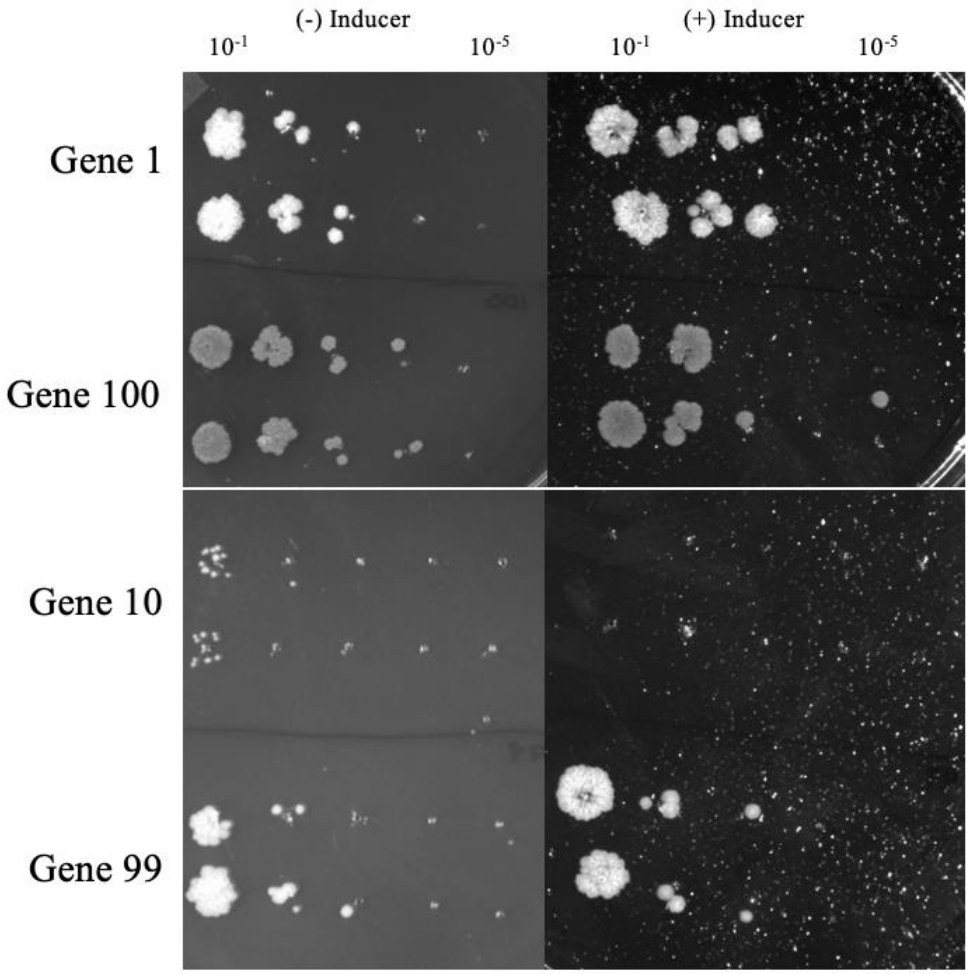


Figure 6. *M. smegmatis* bacterial colonies. Kumao genes 1, 10, 99, and 100 were serially diluted, and then spotted onto plates in the presence (+) and absence (-) of the inducer. If a gene is toxic no bacterial colonies will grow in the presence of inducer, gene 10 is a great example of that.

Table 4. Kumao gene candidates tested for toxicity. The dashes indicate that those genes have not been tested yet.

| Gene | Toxicity | Notes |
|---|---|---|
| 1 | No | |
| 2 | --- | Was not able to test |
| 3 | --- | Was not able to test |
| 4 | --- | Was not able to test |
| 5 | --- | Was not able to test |
| 6 | --- | Was not able to test |
| 8 | No | |
| 9 | No | |
| 10 | Yes | |
| 42 | --- | Was not able to test |
| 43 | --- | Was not able to test |
| 44 | --- | Was not able to test |
| 97 | --- | Was not able to test |
| 98 | Could Not Test | Was not able to grow colonies when electroporated into M. smegmatis |
| 99 | No | |
| 100 | No | |
| 101 | No | |
| 112 | No | |
| 113 | No | |
| 116 | Could Not Test | Was not able to grow colonies when electroporated into M. smegmatis |

## CRISPRi

An alternative approach to determine the location of a repressor gene is to transcriptionally silence candidate operons using CRISPRi in lysogen cells. Transcription of the operon that contains the repressor gene should be active during lysogeny. If the transcription of the repressor gene is silenced, the concentration of repressor protein inside the cells should drop over time to levels that are too low to maintain lysogeny, and the lytic replication cycle should

resume. Kumao gene candidates for transcriptional silencing were selected based on their relation to the start of an operon and on previous knowledge of repressor location.[17] Table 2 shows the genes candidates selected for testing. However, since this system has never been tested, IPhane7 lysogens were used as a positive control, since its repressor location is known. Sequences designed to target IPhane7's gene 1 and 2 were cloned into the pLJR965 vector and electroporated into *M. smegmatis* cells. Liquid cultures of the cells with tween were grown first and then split from this culture into media plus or minus an inducer that will turn on expression of the CRISPRi system. Time points were taken of each sample for five days. Figures 7 and 8 show the results of this experiment. Figure 7 shows the raw spot titer data which was spotted in triplicate for each time point tested. Figure 8 shows the graphical representation from Figure 7, in which the average of each data was graphed. This experiment was performed in three time and data spotted in triplicate. In Figure 7, on the first day, $t_0$, both genes show very minimal spot titer with and without inducer, this is expected as the cultures did not have a chance to incubate yet. After twenty-four hours, there is an increase for both genes in the amount of virus that is present with and without inducer. After twenty-four hours for both genes there is a significant drop in the amount of virus present. Figure 8 shows the average calculated spot titer for Figure 7, and gene 1 there is a slight increase in the amount of virus with inducer, compared to without inducer. For gene 2 there is not much of a difference with and without inducer, as compared to gene 1.

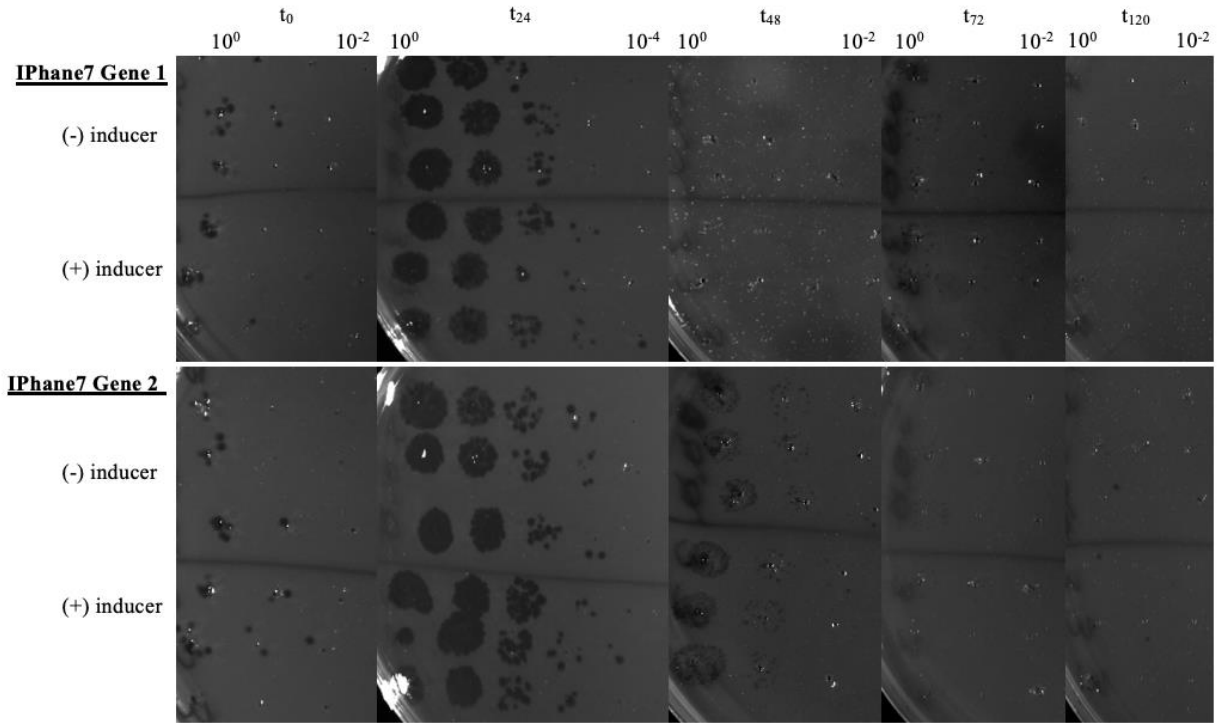


Figure 7. Viral titer of IPhane7 genes 1 and 2. Each timepoint with and without inducer was spotted in triplicate. $t_0$-$t_{120}$ represents the twenty-four hour time points taken over five days. The dark circles represent the viral plaques.

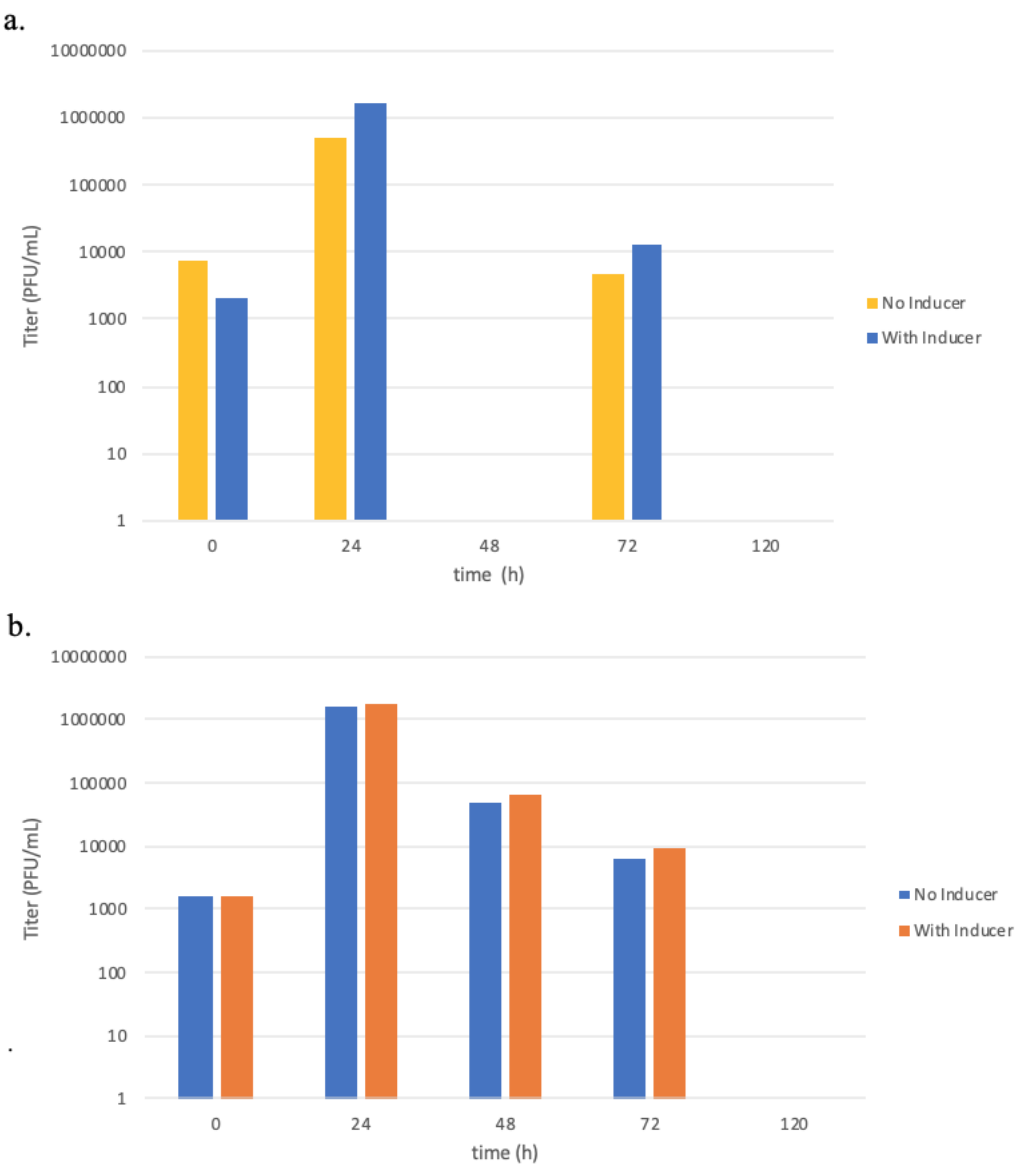


Figure 8a. IPhane7 Viral Titer for Gene 1 with and without inducer. (b) IPhane7 Viral Titer for Gene 2 with and without inducer. The plots represent the average of the viral titer plaques that was performed in triplicate as seen in Figure 7.

# CHAPTER FOUR: DISCUSSION

## Lysogeny Creation and Testing

Kumao was successfully proven that it is a temperate bacteriophage because Kumao lysogens were able to be isolated as seen in Figures 2-4. This result was expected due to the features in Kumao's genome because Kumao has an integrase protein even though bioinformatics failed to identify a repressor protein.[10] Temperate phages need an integrase protein so that the virus's DNA is able to insert its DNA into the hosts DNA. The repressor protein is then usually found relatively near the integrase because it stops transcription from entering the lytic replication cycle. Even though Kumao only had the integrase protein called, this was a good indicator that it was likely to be a temperate bacteriophage.

## pSMEG Overexpression Assay

As shown in Figure 5a, overexpression of Larva's repressor gene completely inhibits Larva viral replication, but not the replication of IPhane7 or Kumao. In fact, even without the presence of inducer a small amount of leaky expression from the pSMEG vector was sufficient to cause an ~3 log inhibition of Larva virus replication.

Kumao's four candidate genes tested, as seen in Figure 5, are genes 1, 2, 5, and 44. Genes 1 and 2 were selected because previous results from our laboratory have[8] revealed that IPhane7 gene 1 is the repressor and gene 2 seems to synergize with gene 1 to completely inhibit virus replication (data not shown). Testing of Kumao's gene 1 and 2 indicated they are not the repressor, as overexpression of these genes did not cause a significant decrease in virus replication of Kumao. In addition, gene 1 and gene 2 showed no significant effect on IPhane7 replication. This is because gene 1 for both Kumao and IPhane7 are from different families.[4,10]

Even though gene 2 for both bacteriophages belong to the same gene family, when their sequences are aligned, they do not align well and are very different from one another. This is why Kumao's gene 2 had no significant effect on IPhane7 replication. Kumao's gene 5 and gene 44, seen in Figures 3d and 3e, showed the most promising results because both genes had almost complete inhibition of Kumao replication. Kumao gene 5 is very promising as the repressor gene candidate compared to gene 44 because the few spots that did appear are very faint, and gene 5 overall caused an ~5 log inhibition. Gene 5 is also considered an orpham because it is not found in any other bacteriophage and is unique to Kumao. This makes it a more promising candidate as the repressor because it is unique to Kumao. This is not normal for repressors that is currently known as repressors are typically able to be identified by bioinformatics and are found in other phages and are not considered orphams.[12] Gene 5 is also found in Kumao's first operon, which is where IPhane7's repressor was also found. The cell lawns for containing gene 44, as compared to gene 5, were lighter, indicating some toxicity from overexpression of gene 5 that could have reduced viral replication because the cells were sick. Also, gene 44 is likely not the repressor because it is in the same operon as the integrase and typically the integrase and the repressor are found going in different transcriptional directions.[12] Gene 5 and gene 44 amino acid sequences were determined through Phyre[18] and I-TASSER[19], which are protein structure prediction software's. Phyre for both genes did not show a helix-turn-helix motif structure, which is characteristic of most. However, I-TASSER did show that gene 5 did have a helix-turn-helix and gene 44 did not. This further suggests gene 5 is the repressor compared to gene 44. More testing will have to be done on both genes to verify gene 5 is indeed Kumao's repressor.

## Gene Toxicity

This experimental series was conducted as a part of a collaboration with Viknesh Sivanathan and Danielle Heller and their new SEA-GENES (Gene-Function Elucidation by a Network of Emerging Scientists) initiative. The goal of this initiative is to determine the function of novel bacteriophage proteins. They are initially focused on discovering novel toxic genes and then determining if their cellular pathways are good therapeutic targets. Gene 10 was the only gene tested so far that was proven to be toxic, which was expected because it did not grow any bacterial cultures with inducer. Gene 99 was predicted to be toxic because it did not grow in liquid with inducer. Although it did produce some colonies, they were small and not as fluffy as compared to the colonies gene 1 and 100 produced. This is usually an indicator that the cells are not well but the gene is not completely toxic. Both genes 10 and 99 are only found in Kumao and haven an unknown function.

## CRISPRi

Due to this method never being done before within a lysogen, it was initially hypothesized that if the repressor is present, the culture with inducer would turn the repressor off and the virus would come out and kill the host, resulting in a clear culture. This was not the case, and the cultures with inducer remained cloudy. This result led to taking time point samples of the cultures of IPhane7's gene 1 and 2 over the course of five days to see if there was an increase in virus activity. As seen in Figure 7, time points were taken over the course of five days for both genes 1 and 2 with and without inducer. The most amount of virus activity was on the second day for both genes 1 and 2, and there was not much of a difference between the two genes and not a significant difference with and without inducer. After the second day there is little to no

virus activity as indicated by the lack of plaques. Gene 1 there is no virus activity at $t_{48}$ and then there is very minimal activity at $t_{72}$. There is a possibility that there was an experimental error, since gene 2 showed virus activity at $t_{48}$. Figure 8 shows the graphical representation of the viral titer on both genes to see if there was a difference with and without inducer since visually there was not much of a difference. There is a slight increase within viral titer with the inducer for gene 1. This increase is the expected outcome because more virus should come out. Although there is not a drastic difference with and without the presence of inducer. There is a possibility with this method that there is not going to be a drastic difference with cultures in the presence and absence of inducer it may only be a subtle difference. For future experiments, it will be important to compare how the cultures look testing an operon in IPhane7 that does not have the repressor to see if there is a difference. Also, future experiments could target Larva's repressor to see if it produces the same results that IPhane7 produced. Although optimization is still required, the foundation has been laid to use CRISPRi to find a repressor within a lysogen.

# CHAPTER FIVE: CONCLUSIONS AND FUTURE WORK

By understanding and further analyzing temperate bacteriophages and their repressor proteins it can help contribute to further understanding genetic evolution, and contribute to phage therapy being more assessable to treat antibiotic resistant bacteria. Phages are naturally antibacterial and are active against even multi-drug resistant bacteria. Phage therapy is more specific than existing drugs and is able to limit the destruction done to non-targeted bacteria.[1,11] If gene 5 is Kumao's repressor, it is novel because it does not fit the standard structure of where repressors are located, and this information will help provide further understanding of temperate phages.[1,5] Control of gene expression is essential for biological systems to function properly. The more these phages are studied, especially the unique ones, the more that can be learned about the evolution of gene expression control mechanisms.[1,5]

Kumao's gene 5 will undergo further testing to verify that it is the repressor. One of the methods is using a single expression assay, with pMH94 as the vector instead of the overexpression assay that was used. The difference between the two methods is the single expression assay a single copy of the gene is expressed, whereas in with the overexpression assay multiple copies of the gene are expressed. pMH94 also uses a natural promoter compared to pSMEG, which uses an inducer to turn the repressor on. However, some genes when overexpressed become toxic which is what is believed gene 44 is doing. Since the single expression assay only gives a single copy of the gene, it can then be determined that gene 44 is not the repressor. Once the repressor is validated, biochemical characterization will be performed on the repressor protein. This includes small-and large-scale protein expression and purification, determining what sequence the repressor binds, and then determining binding stoichiometry

using small-angle x-ray scattering (SAXS), and crystallization.

## CHAPTER SIX: REFERNCES


1) Wittebole, X.; De Roock, S.; Opal, S. M. A Historical Overview of Bacteriophage Therapy as an Alternative to Antibiotics for the Treatment of Bacterial Pathogens. *Virulence* **2014**, *5* (1), 209–218. https://doi.org/10.4161/viru.25991.
2) Hatfull, G. F. Molecular Genetics of Mycobacteriophages. *Microbiol. Spectr.* **2014**, *2* (2). https://doi.org/10.1128/microbiolspec.mgm2-0032-2013.
3) Jain, S.; Hatfull, G. F. Transcriptional Regulation and Immunity in Mycobacteriophage Bxb1. *Mol. Microbiol.* **2000**, 38(5), 971-985. https://doi.org/10.1046/j.1365-2958.2000.02184.x.
4) Pope, W. H.; Anders, K. R.; Baird, M.; Bowman, C. A.; Boyle, M. M.; Broussard, G. W.; Chow, T.; Clase, K. L.; Cooper, S.; Cornely, K. A.; et al. Cluster M Mycobacteriophages Bongo, PegLeg, and Rey with Unusually Large Repertoires of TRNA Isotypes. *J. Virol.* **2014**, *88* (5), 2461–2480. https://doi.org/10.1128/jvi.03363-13.
5) Broussard, G. W.; Oldfield, L. M.; Villanueva, V. M.; Lunt, B. L.; Shine, E. E.; Hatfull, G. F. Immunity Provides Insights into the Evolution of Genetic Switches. *Mol. Cell* **2013**, *49* (2), 237–248. https://doi.org/10.1016/j.molcel.2012.11.012.
6) Hatfull, G. F. Mycobacteriophages: Genes and Genomes. *Annu. Rev. Microbiol.* **2010**, *64* (1), 331-356. https://doi.org/10.1146/annurev.micro.112408.134233.
7) Jacobs-Sera, D.; Marinelli, L. J.; Bowman, C.; Broussard, G. W.; Guerrero Bustamante, C.; Boyle, M. M.; Petrova, Z. O.; Dedrick, R. M.; Pope, W. H.; Modlin, R. L.; et al. On the Nature of Mycobacteriophage Diversity and Host Preference. *Virology*. **2012**, *434* (2), 187-201. https://doi.org/10.1016/j.virol.2012.09.026.

8) Cafferty, E; Gainey, M. In *The Search for a Novel Repressor Gene in Cluster M Bacteriophages*, Western Carolina University Graduate Research Symposium, Cullowhee, NC, April 3-4 2019.
9) The Actinobacteriophage Database. https://phagesdb.org (accessed 2018-12-01).
10) Phamerator. *Genome Maps*. https://phamerator.org/phages (accessed 2018-12-01).
11) Cebriá-Mendoza, M.; Sanjuán, R.; Domingo-Calap, P. Directed Evolution of a Mycobacteriophage. *Antibiotics* **2019**, *8* (2), 46. https://doi.org/10.3390/antibiotics8020046.
12) Hatfull, G. F. *The Secret Lives of Mycobacteriophages*; Elsevier Inc., 2012; Vol. 82. https://doi.or g/10.1016/B978-0-12-394621-8.00015-7.
13) The Actinobacteriophage Database. *Mycobacterium phage Larva. https://phagesdb.org/phages/Larva/ (accessed 2020-01-23).*
14) Phage Discovery Guide. *Chapter 7: Phage Amplification*. https://seaphagesphagediscoveryguide.helpdocsonline.com/7-0-overview (accessed 2020-01-23)
15) Phage Discovery Guide. *Chapter 11: Further Discovery*. https://seaphagesphagediscoveryguide.helpdocsonline.com/11-0-toc (accessed 2020-01-23)
16) SEA-PHAGES. https://seaphages.org (accessed 2019-05-01)
17) Rock, J. M. CRISPRi Gene Silencing in Mycobacteria. 2017, pp 1–8.
18) Phyre2. http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index (accessed 2020-01-23).

19) University of Michigan. *Zhong Lab I-TASSER Protein Structure and Function Predictions.* https://zhanglab.ccmb.med.umich.edu/I-TASSER/ (accessed 2020-01-23).