Spring 5-2016

Characterization of the Mycobacteriophage Ukulele Integration System; Identification of Integration Site Attp and the Role of the Integrase in Lysogeny Regulation

Emily E. Whitaker
University of Maine

CHARACTERIZATION OF THE MYCOBACTERIOPHAGE UKULELE INTEGRATION SYSTEM; IDENTIFICATION OF INTEGRATION SITE ATTP AND THE ROLE OF THE INTEGRASE IN LYSOGENY REGULATION

by

Emily E. Whitaker

Thesis Submitted for Partial Fulfillment
of the Requirements for a Degree with Honors
(Biochemistry)

The Honors College

University of Maine

May 2016

Advisory Committee:

Sally D. Molloy, Assistant Professor of Genomics, Advisor
Keith W. Hutchison, Professor Emeritus of Molecular Biology
John T. Singer, Professor of Microbiology
Dorothy E. Croall, Professor of Biochemistry
David Gross, Part-time Associate Professor in the University of Maine Honors College

ABSTRACT

Mycobacteriophage (phage) are a group of viruses that infect bacteria in the genus *Mycobacterium*. Two phage lifestyles are lytic and temperate. Lytic phage only carry out the lytic life cycle, resulting in host cell lysis. Temperate phage are capable of completing both lytic and lysogenic life cycles. During the lysogenic life cycle, a phage-encoded integrase facilitates integration at sites *attP* in the phage genome and *attB* in the host to form a lysogen. The cluster E mycobacteriophage integration system is poorly understood. Ukulele, a lysogenic cluster E phage, is being used to identify the Cluster E *attP* and characterize lysogeny regulation. A putative *attP* containing sequence was identified in the Ukulele genome by computational analysis. To confirm the presence of *attP*, this sequence will be inserted into a plasmid and transferred into integrase expressing *M. smegmatis* (pST-KT-int). Cells will be screened for plasmid integrated into the genome. To characterize the role of the integrase in lysogeny regulation, we will determine the impact of integrase expression levels on induction event frequency in *M. smegmatis* (pST-KT-int) – Ukulele lysogens.

ACKNOWLEDGEMENTS

I would like to thank the following people without who, this work wouldn't be possible

Sally Molloy and Keith Hutchison; The members of my advisory committee; The University of Maine Honors College and Department of Molecular and Biomedical Sciences; Former members of the Molloy Lab, Gwendolyn Beacham and Katrina Harris; Current Undergraduates working in the Molloy Lab, Campbell Belisle-Haley, Emily Illingworth, Robert Soohey, Erica Sewell Hannah Melotto, Daniel Lesko, Christian Zwirner, Sarai Smith, and Jackson Foley; The SEA-PHAGES program and Dr. Graham Hatfull; The University of Maine HON and BMB 150/155 classes;

The Croall lab for assisting with protein sample preparation and protein separation by SDS-PAGE gel electrophoresis with special thanks to Dr. Dorothy Croall and undergraduates Michael MacLean and Charm Karunasiri; The Maginnis lab for assisting with Western Blot analysis with special thanks to Dr. Melissa Maginnis and graduate student Colleen Mayberry.

Thanks to the following sources of funding

- IDeA Network Of Biomedical Research Excellence (INBRE)/Honors College Comparative Functional Genomics Thesis Fellowship
- INBRE Summer Research Fellowship
- INBRE/Honors College Comparative Functional Genomics Junior Year Research Award
- University of Maine Center for Undergraduate Research

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 Introduction

Mycobacteriophage (phage) are a diverse group of viruses that infect bacteria in the genus *Mycobacterium,* including non-pathogenic *M. smegmatis* and medically relevant *M. tuberculosis*. *M. tuberculosis* is the causative agent of tuberculosis, one of the oldest known human diseases (Smith, I *et al,*. 2003). Although some treatments exist to fight tuberculosis, further investigation is required to combat emerging antibiotic resistant strains. Mycobacteriophage are valuable molecular tools that provide insight into their hosts, including pathogenic *M. tuberculosis*. Phage are the most abundant and diverse biological entity on Earth, and mycobacteriophage appear to be a diverse population as well (Rohwer, 2003) (Hatfull, 2010). Further characterization of mycobacteriophage is required to increase our understanding of the diversity and evolution of phage and their hosts.

In order to examine the diversity and the relationships within this population, mycobacteriophage are divided into clusters. Clusters are groups of phage sharing nucleotide sequence similarity spanning 50% or more of their genomes (Hatfull, 2012). Phage that do not share 50% nucleotide sequence similarity across their genome with any other identified phage are known as singletons (Hatfull, 2012). As of January 31st, 2016, there are 1064 sequenced phage divided into clusters A–Z with 6 singletons. Some clusters are well characterized such as clusters A, O, M K, and G (Hatfull , 2010) (Cresawn *et al*., 2015)(Pope *et al*., 2014)(Pope *et al*., 2011)(Broussard *et al*., 2012). Other clusters, like cluster E, remain poorly understood.

Ukulele is a temperate cluster E phage (Beacham *et al*., 2015). Temperate phage are capable of integrating their genome into that of the host at phage integration site, *attP*

and bacterial site, *attB*, forming a lysogen. Ukulele is capable of forming stable lysogens that demonstrate superinfection immunity to infection by Ukulele and other cluster E phage (Beacham *et al*., 2015). Superinfection immunity is the ability to prevent infection by closely related phage, a common characteristic of lysogens (Hatfull, 2010). Analysis of the Ukulele genome reveals further evidence that Ukulele is a temperate phage.

The Ukulele genome encodes genes required to regulate lysogeny. Ukulele gp49 encodes a tyrosine integrase, the enzyme responsible for facilitating site-specific integration into the host genome (Beacham *et al.,* 2015)(Grindley *et al*., 2006). Although Ukulele encodes an integrase, the cluster E *attP* sequence acted upon by the integrase, has not been identified. Ukulele gp52 gene encodes a predicted winged-helix tertiary structure consistent with DNA binding domains in regulatory proteins such as a repressor or excise (Beacham *et al.,* 2015). Phage immunity repressors are responsible for lysogeny maintenance by suppressing expression of lytic genes and preventing superinfection (Donnelly-Wu *et al*., 1993). An excise gene provides directionality for the integrase to facilitate excision of the phage genome from that of the host (Lewis *et al.,* 2003). Comparison with other phage cluster integration cassettes suggests that gp52 likely encodes a repressor or Cro-like protein (Beacham *et al.,* 2015). Further investigation is required to determine gp52 function. If gp52 does not encode an excise, and other putative excise genes are not identified, this suggests that Ukulele requires an alternative mechanism for facilitating excision from the host genome. One potential alternative mechanism for excision is through altering expression levels of integrase as in cluster G phage (Broussard *et al*., 2012). Cluster G phage do not encode an excise, but controls integration and excision by regulating integrase expression levels (Broussard *et*

al., 2012). Although Ukulele is a temperate phage, further investigation is required to characterize cluster E integration and excision mechanisms and to identify the *attP* site. We are using Ukulele as a model to identify the Cluster E *attP* site and characterize lysogeny regulation.

In this study, we performed computational analyses of the Ukulele genome to predict the location of the Ukulele *attP*. The Ukulele integration cassette was compared to that of well-characterized phage Lambda and L5 (Grindley *et al*., 2006) (Peña *et al*., 1997). Regulatory elements in the integration cassette, such as promoters and repeat motifs were identified to narrow down regions of the integration cassette that likely contained the *attP* site. In order to perform *attP* integration assays and study the effects of integrase expression levels on lysogeny regulation, we constructed a strain of *M. smegmatis* that expresses the Ukulele integrase (*M. smegmatis* (pST-KT-Int)). To confirm the function of the putative *attP* sequence we are cloning the putative *attP*-containing sequence into a non-mycobacterial plasmid and screening for integration into the genome of the *M. smegmatis* strain that expresses the Ukulele integrase. To investigate the potential role of integrase expression levels in the excision process, integrase was induced in *M. smegmatis* (pST-KT-Int)-Ukulele lysogens to observe the effects of integrase expression levels on lysogeny maintenance.

2.0 Literature Review

2.1 Mycobacteriophage and *Mycobacterium* Research

Mycobacteriophage (phage) are a diverse group of viruses that infect bacteria in the genus mycobacterium (Hatfull, 2012). The *Mycobacterium* genus includes *M. tuberculosis*, the causative agent of tuberculosis (TB), and closely related, non-pathogenic, *M. smegmatis*. Phage are used as molecular tools and serve as valuable models for investigating virus/host interactions (Hatfull, 2010). Phage are valuable agents for gaining insight into their host bacterium, including *M. tuberculosis,* the most prevalent bacterial killer among humans (Smith *et al*., 2003).

Evidence of TB was identified in human remains dating back 4,000 years ago, making *M. tuberculosis* one of the oldest known infectious diseases (Smith *et al*., 2003). *M. tuberculosis* is transferred by aerosol particles, initially infecting the respiratory system (Smith *et al*., 2003). TB is primarily a pulmonary disease, but when active the infection can be carried by the lymphatic system and blood to disseminate further than the respiratory system (Smith *et al*., 2003). Antibiotic treatments exist to fight *M. tuberculosis* infection, but the need for more potent therapies still remains (Smith *et al*., 2003). Phage are valuable molecular tools for increasing understanding of virus/host interactions and gaining insight into their pathogenic hosts. Temperate phage integration systems are useful for genetic manipulation of *M. tuberculosis* (Smith *et al*., 2003).

2.2 Phage Clustering

Mycobacteriophage are highly abundant and diverse biological entities (Hatfull, 2012). In order to study relationships between between phage, they are sorted into clusters (Hatfull, 2012). Clusters are groups of phage sharing nucleotide sequence similarity across 50% or more of their genomes (Hatfull, 2012). Singleton phage do not share nucleotide sequence similarity across 50% or more of their genomes with any other phage (Hatfull, 2012). Organizing phage into clusters based on sequence similarity allows for insights into the genetic evolution of phage and characterization of groups of phage with similar genetic features (Hatfull, 2012). As of April 25th, 2016, 1142 phage have been sequenced and organized into clusters A–Z with 6 singletons (phagesDB). Some phage clusters, such as Clusters A and G, are well characterized (Hatfull, 2012) (Broussard *et al*., 2012). Other clusters, such as Cluster E, remain poorly understood (Hatfull, 2012).

2.3 Phage Lifestyles

Two common phage lifestyles are lytic and temperate. Lytic phage are only capable of completing the lytic life cycle. Some lytic phage circularize their genomes immediately after infecting the host genome and begin the process of replicating progeny (28). Replication of progeny and expression of phage encoded lysins cause the host cell to burst. Host cell lysis, the lytic life cycle, can be seen on a lawn of infected bacteria as a clear plaque (Hatfull, 2012). Temperate phage are capable of completing both the lytic and lysogenic life cycle.

During the lysogenic life cycle, the phage genome circularizes after infection and expresses a phage-encoded integrase (Hatfull, 2010). The integrase facilitates integration into the host genome at phage integration site, *attP,* and homologous bacterial site, *attB* (Grindley *et al*., 2006). Recombined *attP* and *attB* form identical sites *attL* and *attR,* which flank the integrated phage genome, or prophage (Grindley *et al*., 2006). Cells carrying prophage are called lysogens. The integrated phage genome is replicated with the host genome when producing daughter cells (28). Phage can excise their genome from the host when signaled by a DNA damage stressor (Dodd *et al*., 2005). Excision from the host genome leads to expression of lytic phage genes, production of progeny, and host cell lysis. The lysogenic life cycle can be seen on a lawn of infected bacteria as a turbid plaque (Hatfull, 2012).

2.4 Lysogeny Regulation

Temperate phage are capable of completing both the lytic and lysogenic life cycles. The decision of which life cycle to complete is dependent on the growth rate of the bacterial host population, and expression of transcription factors in early phage infection (Herskowitz *et al*., 1980)(Maslov *et al*., 2015). Lytic growth is favorable during rapid bacterial growth and low phage multiplicity of infection (Herskowitz *et al*., 1980). During this period, bacteria are actively expressing host factors required for phage replication, and there is an increasing population of bacterial host for phage to infect (Maslov *et al*., 2015). This makes it favorable for phage to propagate infection throughout the population and produce progeny (Maslov *et al*., 2015). Alternatively, during periods of slow or declining bacterial growth, it is favorable for phage to choose

the lysogenic life cycle (Maslov *et al*., 2015). During the lysogenic life cycle, the phage genome can be replicated with the host genome.

2.4.1 Lambda and L5 Lysogeny Regulation

Lysogeny is maintained by a phage encoded immunity repressor protein that prevents expression of lytic genes and prevents infection of closely related invading phage (Hatfull, 2012). The phage genome is replicated with the host genome when the bacterial host produces daughter cells. The advantage of this lifestyle is that the phage genome continues to replicate , and the phage remains dormant until resources required to produce progeny are available.

In *E. coli* phage Lambda, the lifestyle decision is determined by transcriptional factors binding an operator site affecting both lytic and lysogenic gene expression (Herskowitz *et al*., 1980). The transcriptional factors are CI and Cro, which promote lysogenic and lytic lifestyles respectively (Dodd *et al*., 2005). CI binds to O_R operator to shut down lytic gene expression and promote lysogenic gene expression and lysogeny maintenance (Dodd *et al*., 2005). Cro binds the O_R to promote lytic gene expression and prevent CI expression, preventing lysogeny (Dodd *et al*., 2005). Mycobacteriophage L5 utilizes a similar lifestyle regulation system (Nesbit *et al*., 1995). L5 gp71 encodes the L5 repressor protein responsible for regulating lytic and lysogenic gene expression (Brown *et al*., 1997). Directionality of the L5 integrase, which facilitates both integration and excision, is determined by a phage encoded excise protein (Lewis *et al*., 2003). In the absence of the excise protein, the integrase facilitates integration into the host genome

(Lewis *et al*., 2003). In the presence of the excise, the integrase binds to the *attL* and *attR* sites to facilitate excision from the host genome (Lewis *et al*., 2003).

2.4.2 Cluster G Lysogeny Regulation

Cluster G phage utilize an alternative method to regulate lysogeny and switch lifestyles . Upon infection, Cluster G phage begin expressing both an integrase and Cro-like protein (Broussard *et al*., 2012). The integrase is unstable and is degraded quickly (Broussard *et al*., 2012). However, if the bacteria are in a stagnant growth phase, fewer proteases are available to degrade the integrase protein. Under these conditions, increased levels of integrase facilitates integration of the phage genome into the host genome (Broussard *et al*., 2012). The Cluster G *attP* is present within the 3′ end of the phage repressor gene (Broussard *et al*., 2012). Upon integration, the constitutively expressed repressor gene is truncated and activated (Broussard *et al*., 2012). After integration into the host genome, integrase levels greatly decrease and the active repressor maintains lysogeny by competitively binding an operator site to shut down lytic gene expression (Broussard *et al*., 2012). The phage genome is excised from the host genome when integrase expression increases and facilitates excision from the host genome (Broussard *et al*., 2012). Alternatively, if during initial infection of the host, the bacteria are in a rapid growth phase, the protease levels will be high and the integrase quickly degraded (Broussard *et al*., 2012). These conditions allow a cro-like protein to bind the operator site responsible for lytic and lysogenic growth. Binding of the cro-like protein shuts down lysogenic gene expression and promotes lytic growth (Broussard *et al*., 2012). Cluster G

does not encode an excise protein (Broussard *et al*., 2012). Binding to the *att* sites, *attP, B, L,* and *R*, is dependent on the amount of integrase present (Broussard *et al*., 2012).

2.5 *attP* Integration Sites

Phage integrate into the host genome at phage integration site, *attP,* and host integration site, *attB*. Site-specific recombinases, phage-encoded integrases, recognize specific sites in the phage and host genomes and facilitate recombination at these homologous sites (Grindley *et al*., 2006). Some tyrosine integrases use particular motifs, or arm-type binding sequences, to orient their positioning on DNA and facilitate efficient integration (Grindley *et al*., 2006).

Lambda and L5 use arm-type binding motifs to aid in efficient integration (Grindley *et al*., 2006)(Peña *et al*., 1997). Lambda has 5 arm-binding sites flanking either side of the *attP* core (Grindley *et al*., 2006). The two sites located upstream of the *attP* are arranged in an inverted repeat surrounding the integration host factor and the three downstream of the *attP* are all oriented in the same direction (Grindley *et al*., 2006). L5 *attP* is located upstream of the L5 integrase gene. L5 has 7 arm-type binding sites, P1–7. P1–3 are located upstream of the *attP* and P4–7 are downstream (Peña *et al*., 1997). P1 and 2 are oriented in the forward direction, while P3 is in the reverse (Peña *et al*., 1997). Downstream from the *attP*, P4 and 5 are oriented in the forward direction and P6 and 7 in the reverse (Peña *et al*., 1997).

3.0 Materials and Methods

3.1 Bacterial strains and growth conditions.

Mycobacterium smegmatis mc^2155 (ATCC: 700084; NC_008596.1) was used for mycobacteriphage isolation and propagation, lysogen isolation, and integrase expression experiments. *M. smegmatis* was grown at room temperature or at 37° C with shaking at 200 rpm in complete 7H9 broth (Becton Dickinson (BD) Franklin Lakes, NJ), supplemented with 10% albumin dextrose, 1 mM $CaCl_2$, 50 μg mL^{-1} carbenicillin (Sigma, St. Louis, MO), and 10 μg mL^{-1} cyclohexamide (Sigma) or on 7H10 agar plates (BD) supplemented with 10% albumin dextrose, 1 mM $CaCl_2$ (Sigma).

Escherichia coli XL1-Blue (7), obtained from John T. Singer at the University of Maine, was used for construction of recombinant plasmids. *E. coli* was grown at 37° C in Luria broth (L-broth) (1% tryptone, 0.5% yeast extract, 0.5% NaCl, and 0.4% glucose) (BD) and on L-agar (L-broth with 1.6% agar) (BD).

When appropriate, antibiotics were added to the media. Kanamycin (Sigma) was added to media to a final concentration of 20 μg mL^{-1} for *M. smegmatis* and 50 μg mL^{-1} for *E. coli*. Hygromycin (Sigma) was added to media to a final concentration of 50 μg mL^{-1} for *M. smegmatis* and 200 μg mL^{-1} for *E. coli*. Ampicillin (Sigma) was added to media to a final concentration of 50 μg mL^{-1} for *E. coli*.

3.2 Viruses and growth conditions.

Mycobacteriophage Ukulele was isolated by students at the University of Maine in 2011 from a soil sample collected from Old Orchard Beach, ME (3). Phage samples were stored in phage buffer solution (10 mM Tris (pH 7.5), 10 mM $mgSO_4$, 69 mM NaCl, and 1 mM $CaCl_2$).

3.3 Bioinformatic analysis of Ukulele integration cassette.

Ukulele DNA was isolated and prepared by students at the University of Maine and sequenced at the University of Pittsburgh in 2012 using 454 sequencing methods. The Ukulele genome was annotated by Gwendolyn M. Beacham (Beacham *et al*., 2015). Putative promoters in the integration cassette were identified in Geneious (Kearse *et al*., 2012) using an outlined model of *M. smegmatis* promoters (Newton-Foot *et al*., 2013) and programs in DNA Master (Lawrence, 2007) that identify potential −10 and −35 sequences. Putative arm-type binding sites were identified using Multiple EM for Motif Elucidation (Bailey *et al*., 1994). Comparison of putative *attP* containing region was completed using alignment tools in Phamerator (Cresawn *et al.,* 2011) and NCBI BLAST (Blast).

3.4 Plasmids.

Mycobacterial expression plasmid pST-KT (Parikh *et al*., 2013) was used to express the Ukulele integrase gene in *M. smegmatis.* pST-KT contains a multiple cloning site (MCS), kanamycin resistance gene, a modified Tet repressor (TetR) gene, and mycobacterial promoter P_{myc1} with TetR operator site (*tetO*) (Parikh *et al*., 2013).

Expression of a cloned gene is controlled by anhydrotetracycline (ATc) induction of the P_{myc1}*tetO* promoter. Proteins expressed from cloned genes are tagged with hexahistidine and FLAG tags (Parikh *et al*., 2013). pST-KT replicates extrachromosomally in *E. coli* and *M. smegmatis* with a copy number ~23 (Parikh *et al*., 2013).

attP-containing sequences were cloned into plasmid pBR322 (Bolivar *et al.,* 1977). pBR322 contains a MCS, ampicillin resistance gene, and a tetracycline resistance gene (Bolivar *et al.,* 1977). pBR322 replicates extrachromasomally in *E. coli* with a copy number of 15–20 (Bolivar *et al.,* 1977).

3.5 Primer design and DNA sequencing.

Primers were designed based on the target sequence and desired restriction sites using Primer3 (Primer3) (Table 1). All recombinant plasmids were sequenced at the University of Maine DNA sequencing lab.

3.6 PCR amplification.

The Ukulele integrase gene was amplified using flanking primers gp49*Xba*I-R and gp49*Hin*dIII-L (Table 1) in a 25-μL reaction containing 0.5 μM of each primer, 1 ng of Ukulele template DNA, *Taq* polymerase, and PCR master mix according to the manufacturer's recommendations (Promega, Madison, WI). The reaction was heated for 2 cycles at 95°C for 2.5 min, 53° C for 45 s, and 72° C for 1 min followed by 35 cycles of 95° C for 1 min, 63° C for 45 s, and 72° C for 1 min, and a final extension at 72° C for 5 min.

The region of the Ukulele genome containing the putative *attP* and integrase gene was amplified using flanking primers pot*attP*intSpeI-L and pot*attP*intPacI-R in a 25-µL reaction containing 0.5 µM of each primer, 1 ng of template DNA, and Q5 polymerase with PCR master mix as recommended by the manufacturer (New England BioLabs (NEB), Ipswitch, MA). The reaction was heated for 2 cycles at 95° C for 2.5 min, 53° C for 45 s, and 72° C for 1 min, followed by 35 cycles at 95° C for 1 min, 53° C for 1 min, and 72° C for 1 min, and a final extension at 72° C for 1 min.

The region in the Ukulele genome containing the putative *attP* was amplified using flanking primers *attP*intergenic-L and *attP*intergenic-R in 25-µL reactions containing 0.5 µM of each primer, 1 ng of template DNA, Q5 polymerase with PCR master mix as recommended by the manufacturer (NEB). The reaction was heated at 98 ° C for 30 s, followed by 2 cycles at 98° C for 10 s, 67° C for 30 s, and 72° C for 1 min, 35 cycles at 98° C for 10 s, 67° C for 30 s, and 72° C for 30 s, and a final extension at 72° C for 1 min.

The Ukulele potential *attP* containing insert was identified by amplification of pBR-*attP* using flanking primers pbr*attP*-L and Dipbr*attP*-R in 25-µL reaction containing 0.5 µM of each primer, 1 ng, of template DNA, *Taq* polymerase with PCR master mix as recommended by the manufacturer (Promega). The reaction was heated at 95° C for 30 s, followed by 2 cycles at 95° C for 10 s, 57° C for 30 s, and 68° C for 1 min, 35 cycles at 98° C for 10 s, 57° C for 30 s, and 68° C for 30 s, and a final extension at 68° C for 1 min.

PCR products were purified using a QiaQuick PCR Purification kit according to the manufacturer's recommendations (Qiagen).

3.7 Agarose gel electrophoresis.

Gels for electrophoresis were prepared using 1 – 2 % SeaKem LE agarose (Lonza, Rockland, ME) and SeaPlaque (Lonza) agarose dissolved in TAE (40 mM Tris base, 2 mM $Na_2EDTA\ 2H_2O$, 20 mM Glacial Acetic Acid.) buffer. DNA samples were prepared for electrophoresis in a running dye (3 % Ficoll 400, 0.04 % bromophenol blue, 0.04 % xylene cyanol, and 10 mM EDTA). Samples were visualized by staining with 0.5 ng mL^{-1} ethidium bromide and photographed using a ChemImager.

3.8 Restriction endonuclease digests.

Restriction endonuclease digests were completed according to the manufacturer's recommendations (NEB). Reactions contained buffers supplied by the manufacturer (NEB). Reactions were terminated by heat inactivation at 68° or 80° C for 20 min or by performing phenol chloroform extractions

3.9 Ligations.

Ligations were performed using ratios of insert:vector ranging from 1:1 – 21:1. Reactions were performed in 20-µL volumes containing 1 unit of T4 DNA ligase (NEB), T4 ligase buffer (NEB), 100 ng of DNA, and when necessary 0.4 mM dATP and 50 µg mL^{-1} BSA. Reactions were incubated at 13° C overnight and terminated by heating at 65° C for 10 min.

3.10 Plasmid construction.

pST-KT-int was generated by cloning the Ukulele integrase gene into the MCS of inducible mycobacterial expression plasmid pST-KT (Figure 1) (Parikh *et al.*, 2013). Amplification of the integrase gene with primers gp49(XbaI)-L and gp49(HindIII)-R (Table 1) produced a PCR product with *Xba*I and *Hin*dIII restriction sites immediately upstream and downstream, respectively, of the integrase open reading frame. Purified PCR product and pST-KT plasmid were digested with restriction endonucleases *Xba*I and *Hin*dIII (NEB). Digested products were purified by gel electrophoresis and gel extracted using QIAquick Gel Extraction Kit (QIAgen) according to the manufacturer's recommendations. Competent *E. coli* cells were transformed with 10 µl of ligation reaction according to a previously described procedure (Cohen *et al*., 1972). Using a QiaPrep spin miniprep kit (Qiagen), plasmid DNA was isolated from kanamycin resistant *E. coli* XL1-Blue transformants and screened for the presence of the integrase insert by restriction endonuclease assays with *Bam*HI and *Hin*dIII. pST-KT and pST-KT-int were electroporated into electrocompetent *M. smegmatis* mc^2155 as described previously (Kessel *et al*., 2008).

To construct pUV15-*attP*-int, a Ukulele genomic sequence containing the putative *attP* and entire integrase gene (*attP*-int) was PCR amplified and digested with *Spe*I and *Pac*I prior to ligation reactions with *Spe*I and *Pac*I digested mycobacterial vector pUV15-tet-ORm (Figure 2) (Guo *et al*., 2007). Ligation reactions were transformed into One Shot Top10 *E. coli* competent cells according to the manufacturer's recommendation (Life Technologies, Grand Island, NY). Plasmid DNA was isolated from overnight cultures of hygromycin resistant *E. coli* transformants using the QIAprep Spin Miniprep kit

according to the manufacturer's recommendation (Qiagen Inc. USA, Valencia, CA). Transformants were screened for the presence of the *attP*-int insert by restriction endonuclease assays with *Hind*III.

To construct pBR-*attP*, a Ukulele genomic sequence containing the entire putative *attP* containing region within the intergenic region between gp48 and gp49 (*attP*) was PCR amplified and digested with *Bam*HI and *Hin*dIII prior to ligation reactions with *Bam*HI and *Hin*dIII digested *E. coli* vector pBR322 (Figure 3) (Bolivar *et al*., 1977). PCR amplified sequence was purified by phenol chloroform extraction, and ethanol and ammonium acetate precipitation. A 4014 bp digested pBR322 fragment was purified by gel electrophoresis and purified by gel extraction using the Qiagen gel extraction kit (Qiagen, Valentia, CA) according to the manufacturer's instructions . Competent *E. coli* cells were transformed with 10-µL of ligation reaction as previously described (Cohen *et al.*, 1972). Using a QiaPrep spin miniprep kit (Qiagen), plasmid DNA was isolated from ampicillin resistant, tetracycline sensitive *E. coli* XL1-Blue transformants and screened for the presence of the putative *attP* containing sequence by PCR with primers pbr*attP*-L and pbr*attP*-R (Table 1).

3.11 Expressing Ukulele integrase from pST-KT-Int.

M. smegmatis mc^2155 was transformed with pST-KT-Int using a previously described procedure (Kessel *et al*., 2008). *M. smegmatis* containing pST-KT-Int (*M. smegmatis* (pST-KT-Int)) was cultured in complete media with TWEEN 80 (0.0025%) and kanamycin, shaking at 220 rpm for 4 d. The culture was subcultured to an OD_{600} of 0.02 in complete media with kanamycin. Once the culture reached an OD_{600} of 0.60,

cultures were induced with a 0-, 5-, 10-, or 25 ng mL^{-1} anhydrous tetracycline (ATc). ATc induces integrase expression by inhibiting TetR from binding to the *tetO* site in the promoter upstream of the integrase. Induced cultures were incubated at 37° C for 14 h and then room temperature without shaking for 12 h. In volumes of 0.5 ml, aliquots of each culture were infected with 0 and 160 PFU of Ukulele. After 20 min, the cells were plated in 4.5 ml of 7H9 top agar on L-agar plates containing 50 μg mL^{-1} carbenecillin, 10 μg mL^{-1} cyclohexamide, and 20 μg mL^{-1} kanamycin and incubated at 37° C.

3.12 Lysogen Isolation.

Ukulele *M. smegmatis* (psT-KT) and Ukulele *M. smegmatis* (pstkt-int) lysogens were isolated plating serial 10-fold serial dilutions of each *M. smegmatis* strain onto 7H10 Kanamycin plates seeded with 10^9 particle forming units (PFUs) of Ukulele. Resulting colonies were streaked onto new 7H10 Kanamycin plates to remove the colonies from free viral particles. The colonies were tested for superinfection immunity to Ukulele by spotting 10^7, 10^6, 10^5, 10^4, 10^3, 10^2, 10^1, and 0.5 Ukulele PFUs on lawns of potential lysogens. *M. smegmatis* ($mc155^2$) was used as a control. Lysogens that were resistant to lysis by Ukulele lysate were kept for future procedures.

3.13 Inducing integrase expression in Ukulele lysogens.

M. smegmatis (pST-KT-Int)-Ukulele lysogens were cultured in complete media with TWEEN 80 (0.0025%) and kanamycin, shaking at 220 rpm at 25° C for 8 d. The culture was subcultured to an OD_{600} of 0.02 in complete media with kanamycin. Once the culture reached an OD_{600} of 0.60, cultures were induced with 0-, 5-, 10-, or 25 ng mL^{-1}

anhydrous tetracycline (ATc). Induced cultures were incubated at 25° C for 48 h. In volumes of 0.5 ml, aliquots of each culture was plated in 4.5 ml of 7H9 top agar on L-agar plates containing kanamycin and incubated at 25° C.

3.14 Western Blot Analysis.

Expression of Ukulele integrase in *M. smegmatis* (pST-KT-Int) from plasmid pST-KT-Int was confirmed using Western Blot analysis. Cells were prepared for lysis in lysis buffer (Phosphate buffer saline containing 5% glycerol, and 0.1 mm silica beads). Cell lysate was obtained by bead beating 1 mL of cell sample in the Qiagen TissueLyser II (Qiagen) for 3 min. Lysate was clarified by centrifugation at 16,100 x *g* for 10 min. Total protein concentration from each lysate was determined by measuring A_{280} absorbance, assuming Absorbance:protein is 1:1 mg/ml.

20 – 30 µg total protein from each lysate was resolved on a 9% SDS-PAGE by electrophoresis as previously described (Schagger *et al*., 1987). Protein was transferred from the SDS gel to a nitrocellulose membrane with 0.45 µm pores (Bio-Rad Laboratories, Hercules, CA) by electrotransfer. Membranes were treated with blocking buffer (0.1% Tween20, 20 mM Tris, 0.5M NaCl, 5% BSA, pH 7.5) to decrease background signal. The membrane was incubated with monoclonal primary antibody and 1 µg ml^{-1} Mouse Anti-His 6X (ThermoFisher Scientific Waltham, MA) in blocking buffer for 30 min. Primary antibodies were detected by incubaction with Immun-Star Goat Anti-Mouse (GAM)-HRP Conjugate (Bio-Rad Laboratories) 1 hour at .04 µg ml^{-1} in blocking buffer. Antibodies were detected by incubation with HRP substrate according to the manufacturer's instructions (Bio-Rad Laboratories).

4.0 Results

4.1 Computational analysis of Ukulele integration cassette.

Because *attP* sequences are typically located upstream of the integrase gene, a close analysis of the intergenic region between gp48 and and the integrase gene, gp49, was carried out. gp48 and gp49 are divergently transcribed and therefore the region was analyzed for the presence of promoters. A putative leftward and rightward promoter were identified in the intergenic region upstream of gp49 (Figure 3A). To identify potential integrase arm-type binding sites, the integrase cassette was analyzed for repeat motifs (Grindley *et al*., 2006) (Peña *et al.* 1997). There are eight EWMotif1sites within a 625-bp region of the *attP* site (Figure 3B). Based on the location of the putative arm-type binding motifs and promoter sequences, we predict that the *attP* site is located between coordinates 37,101 – 37,170 of the Ukulele genome (Figure 3C).

4.2 Plasmid Constructions.

In order to confirm that the gp48/gp49 intergenic sequence contains the *attP* site, we developed an assay to detect integration activity of this sequence. This required cloning the gp48/gp49 intergenic sequence into an *E. coli* cloning vector and the integrase gene into an inducible mycobacterial expression vector (Figures 1 and 2).

An inducible plasmid containing the Ukulele integrase gene, pST-KT-Int, was constructed by cloning the integrase gene into mycobacterial expression plasmid pST-KT (Parkih *et al*., 2013). The integrase gene was amplified with the addition of restriction tags for *Xba*I and *Hin*dIII. Prior to adding to ligation reactions, pST-KT and the integrase PCR product were digested with *Xba*I and *Hin*dIII yielding expected fragment sizes of

5297-and 891-bp products, respectively (Figure 4). Recombinant plasmid was isolated from kanamycin resistant *E. coli* transformants and analyzed by restriction endonuclease digest analysis with *Bam*HI and *Hin*dIII (Figure 5). Digestion of plasmid from three of the transformants produced the expected fragment sizes of 5296 and 896 bp, indicating the presence of the integrase insert (Figure 2). pST-KT-Int was sequenced to confirm the orientation and sequence of the integrase gene.

We were unable to construct pUV15-*attP*-int. The Ukulele sequence spanning both the putative *attP* and the integrase gene, Ukulele *attP*-Int, was amplified with the addition of restriction tags for *Spe*I and *Pac*I (Figure 6) and digested with restriction endonucleases *Spe*I and *Pac*I. We were unable to clone the *attP*-Int fragment into *Spe*I- and *Pac*I- digested pUV15-tet-ORm. Only non-recombinant pUV15-tet-ORm was isolated from hygromycin resistant *E. coli* transformants. *Hin*dIII digestion of plasmid isolated from transformants produced the fragment sizes of 276, 1199, 3200, and 3351 bp, as predicted for the empty plasmid. None of the digested plasmid samples contained the 4418-bp fragment expected for plasmids carrying the *attP*-Int insert (Figure 7).

We were unable to successfully construct pBR322-*attP*. Ukulele putative *attP* containing sequence was successfully amplified with the addition of restriction tags for *Bam*HI and *Hin*dIII (Figure 8.). Amplification of the intergenic sequence was confirmed by DNA sequencing. pBR322 and *attP* were digested with *Bam*HI and *Hin*dIII to produce 4014- and 308-bp, respectively. We were unable to clone the *attP* fragment into *Bam*HI and *Hin*dIII digested pBR322. All transformants contained recombinant plasmid with inserts other than Ukulele sequence. *Bam*HI endonuclease digestion was used to determine the length of the complete plasmid isolated from ampicillin resistant *E. coli*

transformants. Of the 20 transformants, four of them contained plasmids that were larger in size than the expected 4323 or 4361 bp expected for pBR-*attP* and pBR322, respectively (Figure 9). Recombinant plasmids greater than expected product sizes were not considered for sequencing (Figure 9). Plasmids isolated from ampicillin resistant *E. coli* transformants were analyzed by restriction endonuclease digestions and DNA sequencing. The total length of plasmids isolated from ampicillin resistant *E. coli* transformants was determined by *Bam*HI endonuclease digestion. (Figure 9). *Bam*HI and *Hin*dIII restriction endonuclease digestion was used to determine the length of sequence between *Bam*HI and *Hin*dIII cut sites in plasmid isolated from ampicillin resistant *E. coli*. Only ~4,300 bp plasmids, the expected length for pBR322 and recombinant pBR322-*attP*, that also contained a *Bam*HI and *Hin*dIII endonuclease digestion product 300–400 bp in length were submitted for sequencing. 14 plasmids were sequenced using primer pbr322F to determine the sequence between the plasmid *Bam*HI and *Hin*dIII cut sites. Only non-recombinant pBR322 was isolated from ampicillin resistant *E. coli*.

4.3 Generation of an integrase competent strain of *M. smegmatis* and the impact of expression on Ukulele plaque morphology.

pST-KT-Int was electroporated into *M. smegmatis* to generate an integrase competent strain of *M. smegmatis*. This strain will be used to determine the effects of integrase expression in lysogeny regulation and to perform *attP* integration assays. To induce expression of integrase in *M. smegmatis* cells, cultures were treated with 0-, 5-, 10-, and 25-ng mL^{-1} ATc.

M. smegmatis (pST-KT) and *M. smegmatis* (pST-KT-int) cells induced with ATc were infected with Ukulele lysate to observe plaque morphologies. Ukulele lysate produced turbid plaques on both strains of *M. smegmatis* (data not shown). There was no detectable difference in plaque formation frequency between the strains.

To further characterize the role of integrase expression levels in lysogeny regulation, *M. smegmatis* (pST-KT-Int) and *M. smegmatis* (pST-KT)-Ukulele lysogens were isolated. To confirm the isolation of true lysogens, putative *M. smegmatis* (pST-KT) and (pST-KT-Int)-Ukulele lysogens were tested for superinfection immunity to Ukulele lysate. Varying doses of Ukulele lysate (10–10^7pfu) were applied to lawns of each *M. smegmatis* lysogen. Ukulele infection produced clearings on confluent lawns of the *M. smegmatis* control but not on lawns of *M. smegmatis* (pstKT) #2 and *M. smegmatis* (pstKT-Int) #3-Ukulele lysogens. Superinfection immunity assays confirmed the isolation of true lysogens.

4.4 Western Blot Analysis of Ukulele Integrase expression in *M. smegmatis*

We were unable to detect integrase expression in *M. smegmatis* (pST-KT-Int). The primary antibody detected an off target protein of unknown size in all samples (data not shown).

5.0 Discussion

Temperate mycobacteriophage integrate their genomes into the host genome at integration site *attP*, a crossover exchange facilitated by a phage-encoded integrase. Temperate phage are of interest for their potential use as molecular tools and to better

study lysogenic relationships with their bacterial hosts. Some temperate mycobacteriophage systems such as that of L5 and Cluster G, are better characterized than others (Peña *et al.* 1997)(Broussard *et al.* 2012). By characterizing more integration systems, we enhance our understanding of how to use phage integrases as tools for genetic manipulation. Cluster E is a poorly studied cluster with an uncharacterized integration system and unidentified integration site, *attP*. We are using temperate cluster E mycobacteriophage Ukulele as a model to identify and characterize the cluster E integration site and the role of the integrase in lysogeny regulation. We have identified a putative *attP* containing region and begun to characterize the role of the integrase in lysogeny regulation.

A putative *attP* containing region (37, 101 – 37170) is located upstream of the Ukulele integrase gene. Phage *attP* sites are typically located in close proximity to the integrase gene (Peña et. al. 1997)(Dodd et. al., 2006). Therefore we focused our search within the intergenic region upstream of gp49, the Ukulele integrase gene. To more precisely predict the location of the *attP* site within the intergenic region between gp48 and gp49, we identified SigA-like putative promoters for gp48 and gp49 (Figure 5A) (Newton-Foot *et al*., 2013). Although possible, it is unlikely that the *attP* site is located with the promoter for a gene. Integration at a promoter site would likely disrupt the expression of the promoters target gene. Additionally, identifying arm-type binding sequences in this region may also help us better define the location of *attP.* Tyrosine integrases, like that of Ukulele, often require arm-type binding sites flanking the *attP* core to facilitate efficient integration. We identified eight repeat sequences surrounding

the predicted *attP* integration region that may represent arm-type binding sites for the Ukulele integrase (Figure 5B).

The repeat sequences, EWMotif1, are potential arm-type binding sites for the Ukulele integrase. EWMotif1 is similar to previously identified L5 and Lambda arm-type binding sites Like L5 and Lambda, it is 11-bp long and alternating forward and reverse sequences flanking a putative *attP* core (Figure 5B)(Peña *et al*., 1997) (Grindley *et al*., 2006). Identification of putative arm-type binding sites is further indication that the *attP* site is likely located between coordinates 37,101 – 37,170 (Figure 5C). The region of the Ukulele genome (37,101 – 37,170) predicted to contain the *attP* site is highly conserved in cluster E phage, suggesting that this sequence likely has a critical function in cluster E phage (Table 2).

The Ukulele *attP* sequence is likely conserved amongst other temperate cluster E phage. Because phage integrases are site-specific recombinases, the *attP* site is likely conserved amongst temperate cluster E phage (Grindley *et al*., 2006). The integrase is conserved among Cluster E phage and will only act upon a specific *attP* sequence, therefore the *attP* site must also be conserved for the integrase to be functional. We have determined the most likely region to contain the Ukulele *attP* is the intergenic region upstream of the Ukulele integrase gene, gp49. Previously identified *attP* sites, putative regulatory sequences, and high sequence conservation among Cluster E phage all indicate that the *attP* is likely located upstream of the integrase gene. Given the data from computational analysis, we expect the predicted *attP* containing region to be sufficient for integration into the genome of *M. smegmatis* expressing Ukulele integrase.

To confirm the presence of *attP* upstream of the integrase, we will clone the putative *attP* containing sequence and screen for integration in a strain of *M. smegmatis* expressing Ukulele integrase. If the intergenic region upstream of the integrase contains the *attP*, this sequence should be sufficient for Ukulele integrase to act upon and facilitate integration into the *M. smegmatis* genome. For our first approach, we planned to clone both the *attP* containing sequence and the integrase gene into a single mycobacterial plasmid, pUV15-tet-ORm. If the intergenic region does contain the *attP* site, cloning both the *attP* containing sequence and the integrase gene into a plasmid should allow for integration of pUV15-*attP*-int into the *M. smegmatis* genome. We successfully amplified the Ukulele sequence containing both the intergenic region and integrase gene by PCR; however, we were unable to isolate a recombinant plasmid. It is possible that uncontrolled expression of the Ukulele integrase in *E. coli* could be inhibiting cell growth and preventing isolation of pUV15-*attP*-Int plasmid. Because the *attP*-Int PCR product contains the sequence upstream of gp49, it includes the promoter for the integrase gene (Figure 2C). The putative integrase gene promoter is a SigA-like promoter and is likely constitutively active in mycobacterial cells. While most phage integrases, such as the Lambda integrase, can be expressed in *E. coli,* other phage integrases, such as *intB* a lambda-like integrase, *int*-B13, from an integrative and conjugate element (ICE) in *Pseudomonas knackmussi*, are toxic and impossible to clone without tight gene regulation in *E. coli* (Miyazaki *et al*. 2013). Uncontrolled expression of the Ukulele integrase in *E. coli* may have been toxic, thus preventing the isolation of recombinant transformants. It is also reasonable to speculate that cloning both the *attP* site and integrase gene into the

same plasmid would be difficult because of the reactivity of uncontrolled integrase expression in the presence of the *attP*.

To prevent this reactivity and potential toxic effects of uncontrolled integrase expression, we planned to clone the putative *attP* containing region and integrase gene separately into *E. coli* plasmid pBR322 and mycobacterial vector pST-KT, respectively. pBR322 encodes an ampicillin resistance gene and lacks a mycobacterial origin of replication. If the cloned intergenic sequence contains the *attP* site, pBR322-*attP* in the presence of integrase expressed from pSTKT-int, should integrate into the *M. smegmatis* genome. The intergenic sequence upstream of gp49 was successfully amplified by PCR; however, we were unable to isolate a recombinant plasmid. It is possible that the *attP* sequence is reactive within the *E. coli* transformants, preventing isolation of a recombinant plasmid. Additionally, it is possible that hybridization flanking restriction sites PCR cloned onto the target insert are not sufficient for efficient *Bam*HI and/or *Hin*dIII restriction endonuclease digestion. If either end of the insert is not cleaved by a restriction endonuclease, either *Bam*HI or *Hin*dIII, the insert sequence and pBR322 will not have homologous sticky ends to allow for directionally inserting the *attP* sequence into the plasmid vector. Further investigation of the length and GC content of the insert hybridization sequence may aid in creating the pBR322-*attP* recombinant plasmid.

Once pBR322-*attP* recombinant plasmids are isolated, integration assays will be completed to confirm the presence of *attP* upstream of gp49. The plasmid pBR322-*attP* will be electroporated into competent *M. smegmatis* (pST-KT-Int), which is capable of expressing integrase after treatment with ATc. Because pBR322 cannot be maintained extrachromosomally in mycobacterial cells, only cells with integrated pBR322-*attP* will

be selected on plates containing both ampicillin and kanamycin. Identification of an integrated pBR322-*attP* plasmid will confirm the presence of *attP* within the intergenic region upstream of gp49. If the integration assay reveals that the cloned sequence contains the *attP*, we will further characterize the sequence required for efficient integration. In addition to identifying the Ukulele *attP* site, we plan to determine if integrase expression levels are implicated in lysogeny regulation.

To observe the effects of integrase expression on Ukulele lifestyle regulation, we aimed to alter the expression levels of Ukulele integrase in *M. smegmatis* (pSTKT-Int) cells. Integrase expression was induced with ATc at concentrations of 0, 5, 10, and 25 ng μL^{-1}. ATc induces integrase expression by competitively binding TetR to prevent it from acting on its target site. TetR is constitutively expressed repressor encoded by pSTKT-int, that binds an operator *tetO* within the promoter P_{myc}*tetO* to prevent expression of the downstream gene (Parikh *et al.,* 2013). Because increasing ATc decreases the amount of TetR bound to tetO, we expect integrase expression levels to increase with increased ATc concentrations (Parikh *et al.,* 2013). To determine if integrase expression levels play a role in Ukulele lifestyle regulation, *M. smegmatis* (pSTKT-int) cultures treated with ATc were infected with Ukulele lysate and plated to observe plaque morphologies. If integrase expression levels do play a role in lifestyle regulation, we expect to see a difference in plaque morphology between cultures expressing different amounts of integrase.

Ukulele integrase expression levels do not appear to regulate the directionality of the integrase. If the integrase expression levels are implicated in lysogeny regulation, we might expect to see a difference in Ukulele plaque morphology when altering integrase expression levels in *M. smegmatis*. However, there was no detectable difference in plaque

morphology in the presence of increased integrase expression levels. These results suggest that Ukulele integrase expression levels are not implicated in integrase directionality or lysogeny regulation. One possibility is that Ukulele does encode an unidentified excision protein that controls directionality of the integrase. Another possibility is that the increased integrase expression levels were promoting excision but because the Ukulele control plaques were clear, we could not visually detect increased excision activity. However, it is difficult to interpret these results without knowing if *M. smegmatis* (pST-KT-Int) is expressing Ukulele integrase. We have not yet completed Western Blot analysis to confirm that *M. smegmatis* (pST-KT-Int) is expressing Ukulele integrase, and if the integrase levels are increasing when we induce expression with increased ATc. It is also possible that Ukulele uses an alternative mechanism to control integrase directionality and lysogeny regulation is unique from that of L5 and Cluster G phage (Lewis *et al.,* 2003) (Broussard *et al.* 2012). Further investigation is required to determine if Ukulele integrase expression levels are responsible for controlling integrase directionality in lysogeny regulation.

Altering Ukulele integrase expression levels may be a mechanism used to regulate the integration system by controlling the directionality of the integrase. Since we have been unable to identify the excision gene in the Ukulele genome, it is reasonable to speculate that the integrase expression levels may be involved in regulating the directionality of the phage integrase. However, it is possible that Ukulele encodes an excise protien. Ukulele gp52 has been identified as a potential repressor or excise. Before we can rule out the possibility of an excise assisting in integrase directionality, we need to determine the function of gp52. Additionally, because we have been unable to confirm

that the integrase is being expressed in *M. smegmatis*, we can only speculate that the integrase expression levels may play a role in lysogeny regulation. Further investigation of the Ukulele genome, effects of altering integrase expression levels, and identification of the *attP* site are required to better characterize the Ukulele integration system.

5.1 Future Directions

If integration assays of the putative *attP* containing sequence confirm the presence of *attP* upstream of the integrase, continued efforts are required to determine the sequence required for efficient integration. To further characterize the putative *attP* containing region, we propose to introduce site directed mutagenesis of the intergenic sequence to determine the minimum sequence required for efficient integration. We will complete integration assay on the mutated sites to identify the minimum sequence required for efficient integration and the Ukulele *attP* core. Some integrases, like that of L5, require arm-type binding sites in addition to the *attP* core to facilitate integration (Peña *et al.*, 1997). It is currently unclear if the Ukulele integrase requires arm-type binding sites for efficient integration.

Once the *attP* core has been identified, we plan to determine if arm-type binding sites are essential for Ukulele integrase to facilitate efficient integration. One way to investigate this is to complete an integration assay comparing the efficiency of integration of the *attP* core to that of the Ukulele sequence containing the *attP* core and all potential arm-type binding sites. Equal integration efficiency for both sequences suggests that only the *attP* core is essential for efficient integration. However, greater integration efficiency of the sequence containing both the *attP* core and putative arm-type binding sites than

that containing only the *attP* core suggests that the arm-type binding sites enhance integration efficiency. Identification of the *attP* core will allow us to further characterize the regulation mechanisms of the Ukulele integration system.

If integration assays reveal that the intergenic sequence does not contain the *attP* I propose refocusing the search for the *attP* within predicted repressor genes. We have been unable to identify a Ukulele excise gene. Cluster G phage do not encode an excise which makes is reasonable to speculate that the Ukulele integration system may function similar to that of cluster G, including using a similar *attP* site. Cluster G phage *attP* site is located within the repressor gene (Broussard *et al.,* 2012). Because the cluster G *attP* is located within the repressor gene, integration results in expression of a truncated repressor protein (Broussard *et al.,* 2012). Because the cluster G repressor is only active in its truncated form, cleavage of the *attP* site and integration into the host genome acts as a genetic switch to maintain lysogeny (Broussard *et al*., 2012). Although identification of the Ukulele repressor, the protein required for lysogeny maintenance, is currently in progress, members of the lab have identified putative repressor genes, gp52 and gp53 (Beacham *et al*., 2015). An integration assay similar to that used to test the putative *attP* containing region could be applied to examine the reactivity of putative repressor genes with the Ukulele integrase. If the *attP* site is not located upstream of the integrase, I suggest cloning the putative repressor gene into pBR322 and completing integration assays to screen for integration of the repressor gene into the *M. smegmatis* genome. If the Ukulele integration system functions similar to Cluster G, it is also possible that integrase expression levels are used to control integrase directionality in lysogeny regulation.

We propose to alter integrase expression levels in *M. smegmatis* (pST-KT-Int)-Ukulele lysogens as an alternative method to characterize the role of integrase expression levels in lysogeny regulation. We isolated both *M. smegmatis* (pST-KT-Int) and *M. smegmatis* (pST-KT)-Ukulele lysogens. These lysogens will be induced with varying levels of ATc to induce integrase expression, *M. smegmatis* (pST-KT) and *M. smegmatis* mc^2155 will be used as a non-integrase expressing control strains. If increased Ukulele integrase expression levels promote excision in Ukulele lysogens, as seen in Cluster G, we expect to see an increase in induction events in the *M. smegmatis* (pST-KT-Int) lysogens compared to the control strains. The induction events would be seen as plaques on a lawn of lysogens. If there is an increase in induction events after treatment with ATc, it is likely that integrase expression levels are used to control integrase directionality in Cluster E phage.

We will continue our efforts to confirm the presence of the Ukulele *attP* upstream of the integrase gene. Because lysogeny regulatory mechanisms and the *attP* site are likely conserved amongst cluster E phage, it is likely that the Ukulele lysogeny regulation system can be used as a model for other cluster E phage. Identifying the Ukulele *attP* and determining the role of integrase expression levels in Ukulele lysogeny regulation will greatly enhance our understanding of cluster E integration systems. Well characterized integration systems also have the potential to be used as molecular tools to study their hosts. Characterization of cluster E integration systems can contribute to the greater understanding of mycobacteriophage integration systems and virus-host relationships.

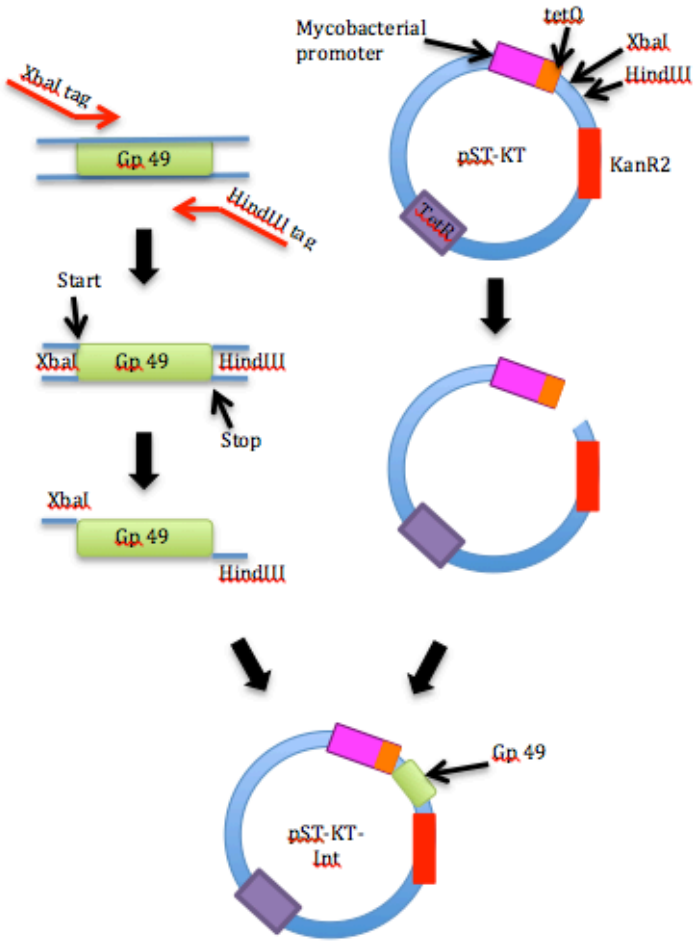

Figure 1.
Experimental design for plasmid pST-KT-Int construction. Ukulele integrase, gp49, was directionally inserted into plasmid pST-KT at complementary sticky ends created by *Xba*I and *Hin*dIII restriction endonuclease digestion.

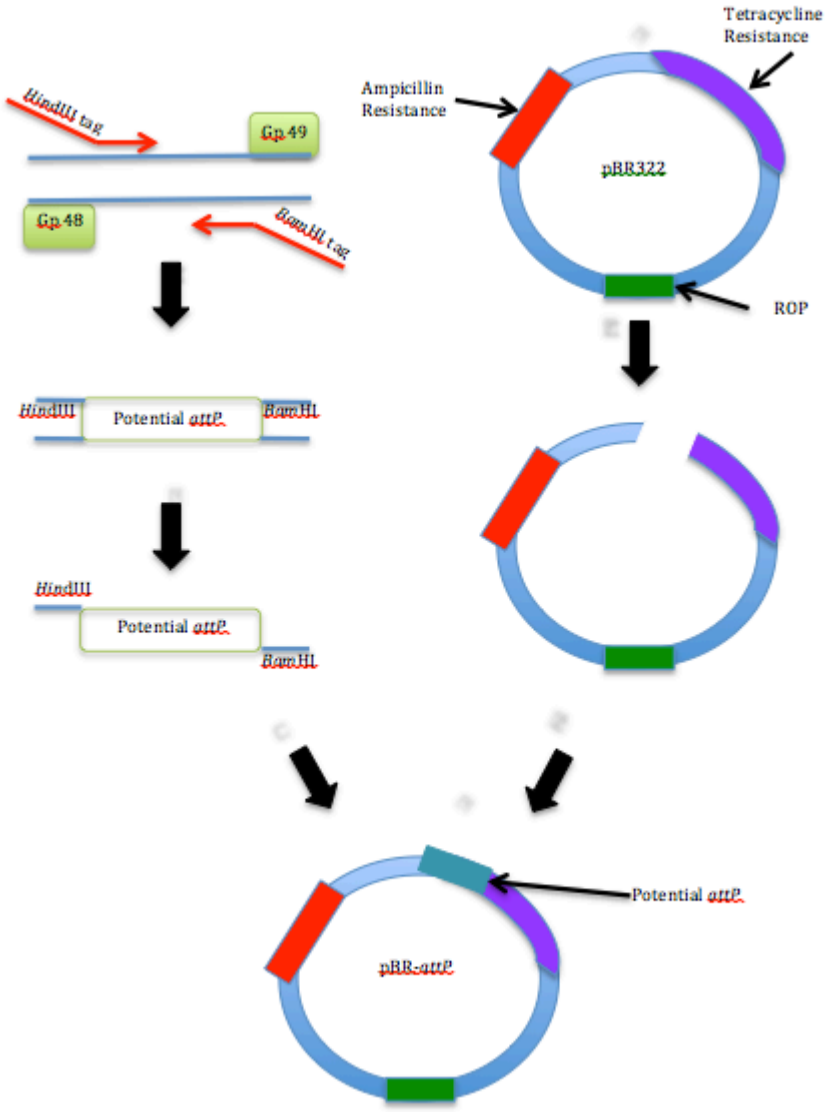

Figure 2. Experimental design for plasmid pbr-*attP* construction. Intergenic sequence upstream of Ukulele gp49 will be directionally inserted into plasmid pBR322 at complementary sticky ends created by *Hin*dIII and *Bam*HI restriction endonuclease digestion.

Figure 3. Identification of promoters and integrase arm-type binding sites in the Ukulele integration cassette. A.) Geneious (Kearse et al., 2012) map of intergenic region between gp48 and gp49 containing putative promoter elements (green arrows). B.) Consensus sequence for EWMotif1, the potential integrase arm-type binding sites. C.) Geneious map of intergenic region between gp48 and gp49, containing putative promoters (green arrows), arm-type binding sites (blue arrows) and putative *attP* containing region (red box, 37,101 – 37,170).

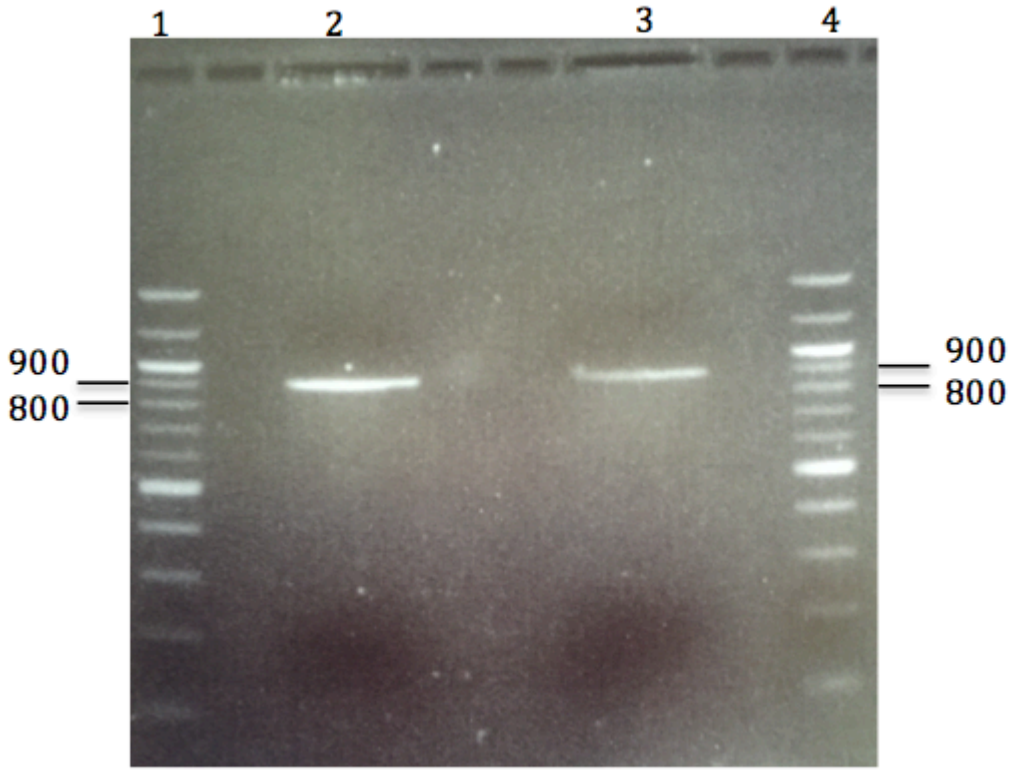

Figure 4. Agarose gel electrophoresis of *Xba*I and *Hin*dIII digested Ukulele integrase gene. Lanes 1 and 4 contain the molecular size markers. Lanes 2 and 3 contain the 891 bp *Xba*I and *Hin*dIII digested integrase PCR product.

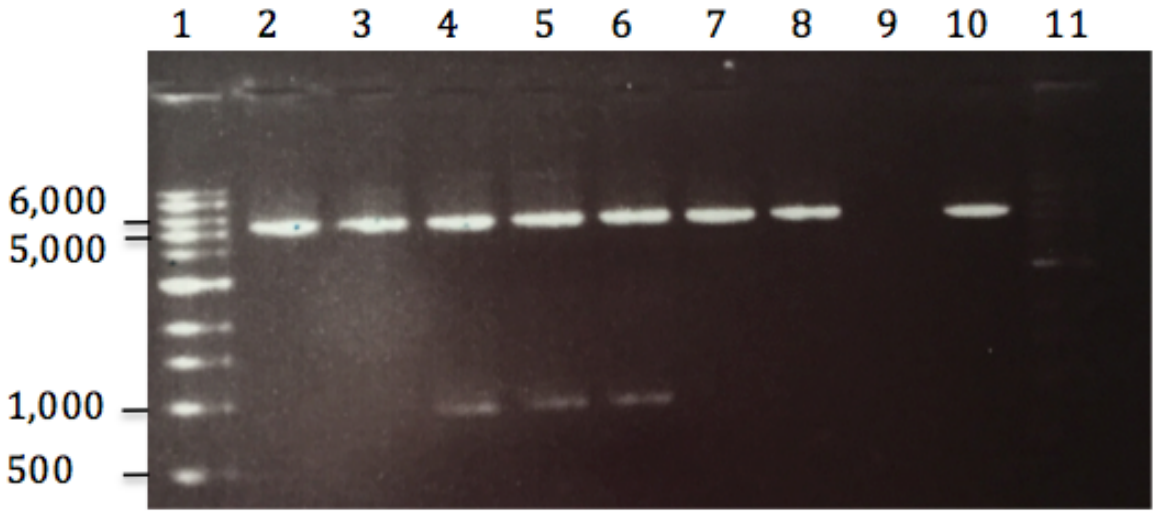

Figure 5. Agarose gel electrophoresis of *Bam*HI and *Hin*dIII restriction endonuclease digest analysis of plasmid isolated from kanamycin resistant *E. coli* transformants. Lanes 1 and 11 contain molecular size markers. Plasmid DNA isolated from transformants (lanes 2 – 8), or control plasmid pST-KT (lane 10) was digested with *Bam*HI and *Hin*dIII.

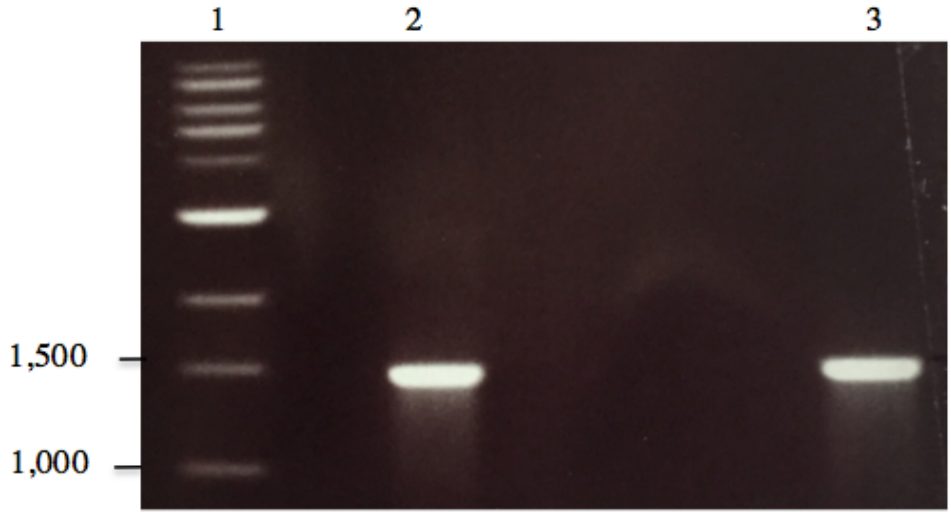

Figure 6. Agarose gelelectrophoresis of a 1460 bp region of the Ukulele genome containing the putative *attP* site and the integrase gene (*attP*-Int) (36,700 – 38,139). Lane 1 contains the molecular size marker. Lanes 2 and 3 contain the PCR product *attP*-Int.

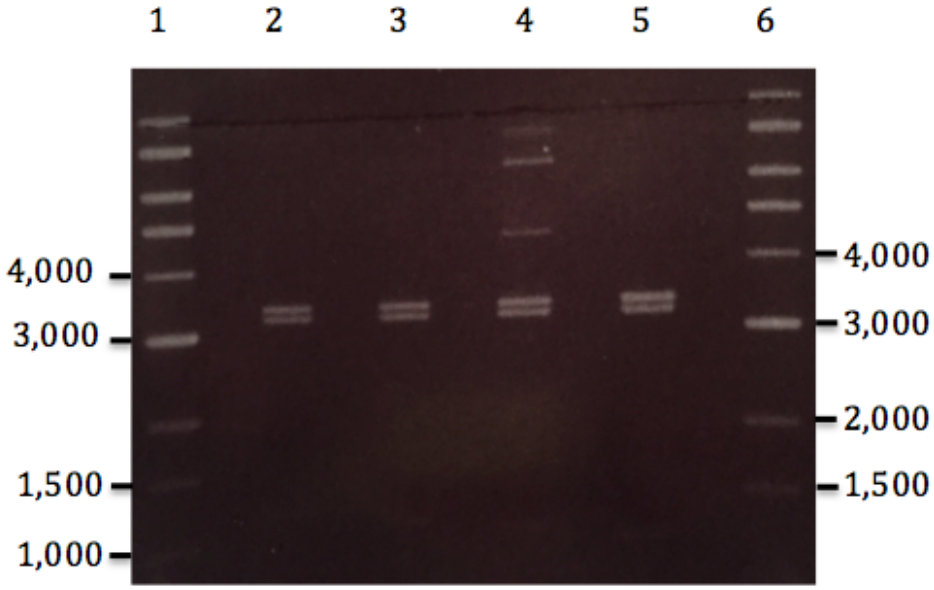

Figure 7. Agarose gel electrophoresis of *Hin*dIII restriction endonuclease digest analysis of plasmid DNA from hygromycin resistant *E. coli* transformants. Lanes 1 and 6 contain molecular size markers. Plasmid DNA isolated from transformants (lanes 2 – 4) or control plasmid pUV15-tet-ORm (lane 5) was digested with *Hin*dIII.

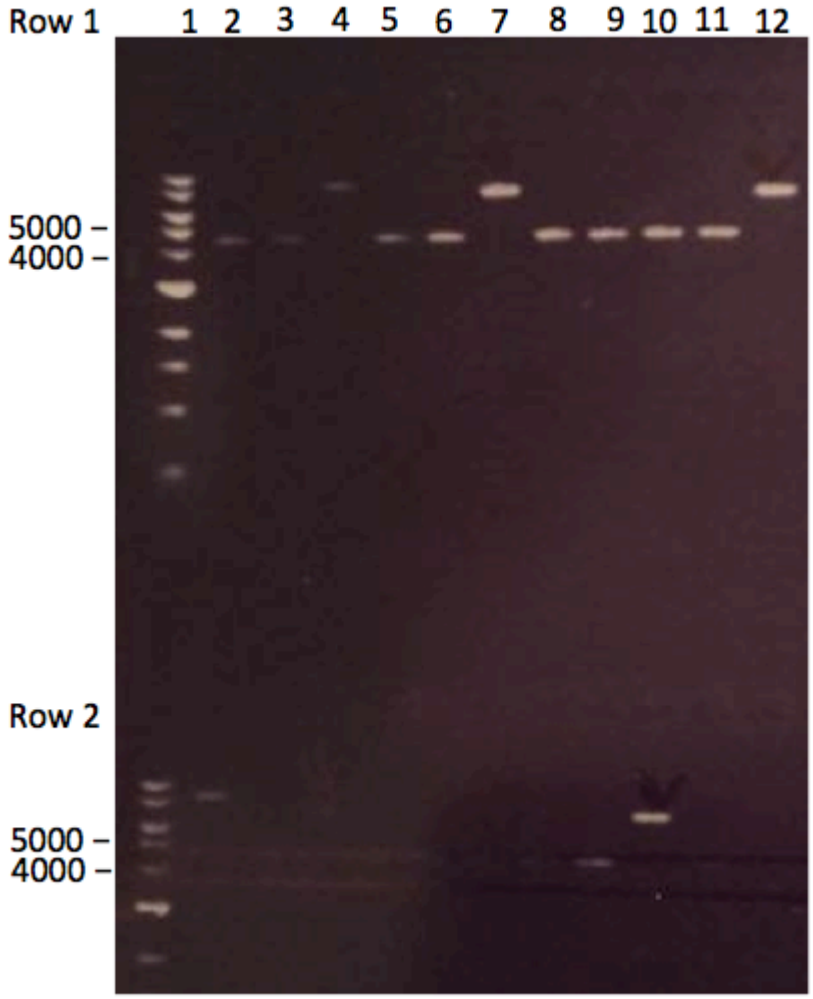

Figure 8. Agarose gel electrophoresis of *Bam*HI restriction endonuclease digest analysis of plasmid isolated from ampicillin resistant *E. coli* transformants. Lanes 1 in rows 1 and 2 contain molecular size markers. Plasmid DNA isolated from transformants (lanes 2 – 12 row 1 and lanes 2 – 10 row 2), or control plasmid pBR322 (lane 11 row 2) was digested with *Bam*HI.

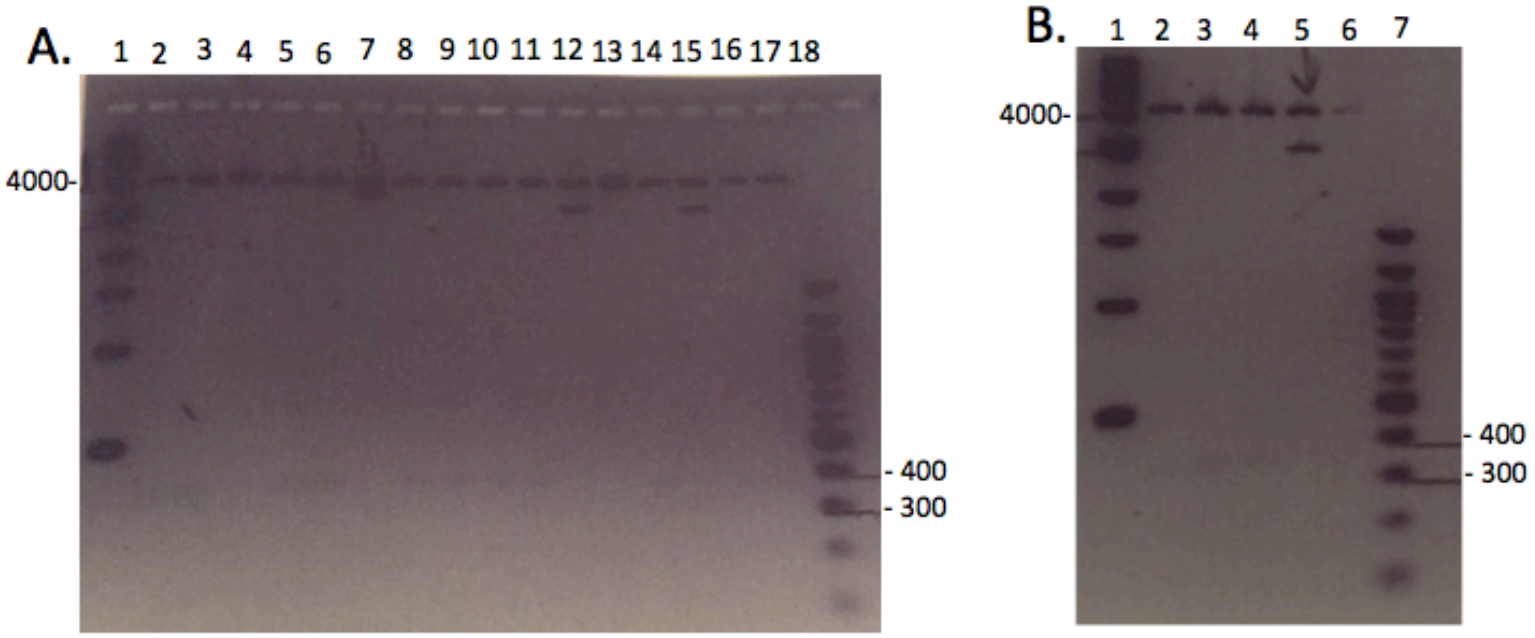

Figure 9. Agarose gel electrophoresis of *Bam*HI and *Hin*dIII restriction endonuclease digest analysis of plasmid isolated from ampicillin resistant *E. coli* transformants. A. Lanes 1 and 18 contain molecular size markers. Plasmid DNA isolated from transformants (lanes 2 – 17), or control plasmid pbR322 (lane 10) was digested with *Bam*HI and *Hin*dIII. B. Lanes 1 and 7 contain molecular size markers. Plasmid DNA isolated from transformants (lanes 2 – 5), or control plasmid pBR322 (lane 6) was digested with *Bam*HI and *Hin*dIII.

Primer Name	Sequence 5' – 3'
Gp49(XbaI)-L	TAATCTAGAGTGACAGACGATTTGCT
Gp49(HindIII)-R	ATAAAGCTTCTACTCCACGCTCAAATTC
potattPintSpeI-L	AATACTAGTACTTCTCAGGAATCGAG
potattPintPacI-R	TTAATTAACTACTCCACGCTCAAA
attPintergenic-L	TATAAGCTTGACGATCCCCCACAGG
attPintergenic-R	CTTCCTAGGAGCAAATCGTCTGTCAC
pbrattp-L	GGACACCGTAAAAGCTGGTC
Diagpbrattp-R	ATCGGTGATGTCGGCGATAT
pbr322F	CCTGACGTCTAAGAAACAT

Table 1. Primers.

Phage name	Max Score	Total Score	Query Cover	E value	% Identity
Lilac	130	130	100%	5e-30	100%
Bruin	130	130	100%	5e-30	100%
HufflyPuff	130	130	100%	5e-30	100%
Nala	130	130	100%	5e-30	100%
PhatBacter	130	130	100%	5e-30	100%
Quink	130	130	100%	5e-30	100%
Goku	130	130	100%	5e-30	100%
Contagion	130	130	100%	5e-30	100%
DrDrey	130	130	100%	5e-30	100%
Phaux	130	130	100%	5e-30	100%
Phrux	130	130	100%	5e-30	100%
Dumbo	130	130	100%	5e-30	100%
Murphy	130	130	100%	5e-30	100%
Kostya	130	130	100%	5e-30	100%
Porky	130	130	100%	5e-30	100%
244	130	130	100%	5e-30	100%
Cjw1	130	130	100%	5e-30	100%

Table 2. NCBI BLAST alignment of Ukulele putative *attP* (37,101-37,170) with other cluster E mycobacteriophage genomes.

REFERENCES

1. The Actinobacteriophage Database | Home. (n.d.). Retrieved April 16, 2016, from http://phagesdb.org/

2. Beacham, G. M. (2015). Complete Annotations of the Cluster E Mycobacteriophage Ukulele Genome and Characterization of Cluster E Lysogeny Regulation. *Honors College*. Paper 226. http://digitalcommons.library.umaine.edu/honors/226

3. BLAST ®. (n.d.). Retrieved April 16, 2016, from https://blast.ncbi.nlm.nih.gov/Blast.cgi

4. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., . . . Falkow, S. (1977, 11). Construction and characterization of new cloning vehicle. II. A multipurpose cloning system. *Gene, 2*(2), 95-113. doi:10.1016/0378-1119(77)90000-2

5. Broussard, G., Oldfield, L., Villanueva, V., Lunt, B., Shine, E., & Hatfull, G. (2013, 01). Integration-Dependent Bacteriophage Immunity Provides Insights into the Evolution of Genetic Switches. *Molecular Cell, 49*(2), 237-248. doi:10.1016/j.molcel.2012.11.012

6. Brown, K. L., Sarkis, G. J., Wadsworth, C., & Hatfull, G. F. (1997, 10). Transcriptional silencing by the mycobacteriophage L5 repressor. *EMBO J The EMBO Journal, 16*(19), 5914-5921. doi:10.1093/emboj/16.19.5914

7. Cohen, S. N., Chang, A. C., & Hsu, L. (1972, 08). Nonchromosomal Antibiotic Resistance in Bacteria: Genetic Transformation of Escherichia coli by R-Factor DNA. *Proceedings of the National Academy of Sciences, 69*(8), 2110-2114. doi:10.1073/pnas.69.8.2110

8. Cresawn, S. G., Pope, W. H., Jacobs-Sera, D., Bowman, C. A., Russell, D. A., Dedrick, R. M., . . . Hatfull, G. F. (2015, 03). Comparative Genomics of Cluster O Mycobacteriophages. *PLOS ONE PLoS ONE, 10*(3). doi:10.1371/journal.pone.0118725

9. Cresawn, S. G., Bogel, M., Day, N., Jacobs-Sera, D., Hendrix, R. W., & Hatfull, G. F. (2011). Phamerator: A bioinformatic tool for comparative bacteriophage genomics. *BMC Bioinformatics, 12*(1), 395. doi:10.1186/1471-2105-12-395

10. Dodd, I. B., Shearwin, K. E., & Egan, J. B. (2005, 04). Revisited gene regulation in bacteriophage λ. *Current Opinion in Genetics & Development, 15*(2), 145-152. doi:10.1016/j.gde.2005.02.001

11. Donnelly-Wu, M. K., Jacobs, W. R., & Hatfull, G. F. (1993, 02). Superinfection immunity of mycobacteriophage L5: Applications for genetic transformation of mycobacteria. *Molecular Microbiology Mol Microbiol, 7*(3), 407-417. doi:10.1111/j.1365-2958.1993.tb01132.x

12. Grindley, N. D., Whiteson, K. L., & Rice, P. A. (2006, 06). Mechanisms of Site-Specific Recombination*. *Annu. Rev. Biochem. Annual Review of Biochemistry, 75*(1), 567-605. doi:10.1146/annurev.biochem.73.011303.073908

13. Guo, X. V., Monteleone, M., Klotzsche, M., Kamionka, A., Hillen, W., Braunstein, M., . . . Schnappinger, D. (2007, 05). Silencing Essential Protein Secretion in Mycobacterium smegmatis by Using Tetracycline Repressors. *Journal of Bacteriology, 189*(13), 4614-4623. doi:10.1128/jb.00216-07

14. Hatfull, G. F. (2010, 10). Mycobacteriophages: Genes and Genomes. *Annual Review of Microbiology Annu. Rev. Microbiol., 64*(1), 331-356. doi:10.1146/annurev.micro.112408.134233

15. Hatfull, G. F. (2012). The Secret Lives of Mycobacteriophages. *Bacteriophages, Part A Advances in Virus Research,* 179-288. doi:10.1016/b978-0-12-394621-8.00015-7

16. Herskowitz, I., & Hagen, D. (1980, 12). The Lysis-Lysogeny Decision of Phage

lambda: Explicit Programming and Responsiveness. *Annu. Rev. Genet. Annual Review of Genetics, 14*(1), 399-445. doi:10.1146/annurev.ge.14.120180.002151

17. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Drummond, A. (2012, 04). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics, 28*(12), 1647-1649. doi:10.1093/bioinformatics/bts199

18. Kessel, J. C., Marinelli, L. J., & Hatfull, G. F. (2008, 11). Recombineering mycobacteria and their phages. *Nature Reviews Microbiology Nat Rev Micro, 6*(11), 851-857. doi:10.1038/nrmicro2014

19. Lewis, J. A., & Hatfull, G. F. (2003, 02). Control of Directionality in L5 Integrase-mediated Site-specific Recombination. *Journal of Molecular Biology, 326*(3), 805-821. doi:10.1016/s0022-2836(02)01475-4

20. Lawrence, J. DNA Master. http://cobamide2.bio.pitt.edu/. 2007.

21. Maslov, S., & Sneppen, K. (2015, 06). Well-temperate phage: Optimal bet-hedging against local environmental collapses. *Sci. Rep. Scientific Reports, 5*, 10523. doi:10.1038/srep10523

22. Miyazaki, R., & Meer, J. R. (2013, 05). A New Large-DNA-Fragment Delivery

System Based on Integrase Activity from an Integrative and Conjugative Element. *Applied and Environmental Microbiology, 79*(14), 4440-4447. doi:10.1128/aem.00711-13

23. Nesbit, C. E., Levin, M. E., Donnelly-Wu, M. K., & Hatfull, G. F. (1995, 09). Transcriptional regulation of repressor synthesis in mycobacteriophage L5. *Molecular Microbiology Mol Microbiol, 17*(6), 1045-1056. doi:10.1111/j.1365-2958.1995.mmi_17061045.x

24. Newton-Foot, M., & Pittius, N. C. (2013, 01). The complex architecture of mycobacterial promoters. *Tuberculosis, 93*(1), 60-74. doi:10.1016/j.tube.2012.08.003

25. Parikh, A., Kumar, D., Chawla, Y., Kurthkoti, K., Khan, S., Varshney, U., & Nandicoori, V. K. (2013, 01). Development of a New Generation of Vectors for Gene Expression, Gene Replacement, and Protein-Protein Interaction Studies in Mycobacteria. *Applied and Environmental Microbiology, 79*(5), 1718-1729. doi:10.1128/aem.03695-12

26. Peña, C. E., Lee, M. H., Pedulla, M. L., & Hatfull, G. F. (1997, 02). Characterization of the mycobacteriophage L5 attachment site, attP 1 1 Edited by M. Gottesman. *Journal of Molecular Biology, 266*(1), 76-92. doi:10.1006/jmbi.1996.0774

27. Pope, W. H., Anders, K. R., Baird, M., Bowman, C. A., Boyle, M. M., Broussard, G. W., . . . Imperiale, M. J. (2013, 12). Cluster M Mycobacteriophages Bongo, PegLeg, and Rey with Unusually Large Repertoires of tRNA Isotypes. *Journal of Virology, 88*(5), 2461-2480. doi:10.1128/jvi.03363-13

28. Pope, W. H., Ferreira, C. M., Jacobs-Sera, D., Benjamin, R. C., Davis, A. J., Dejong, R. J., . . . Hatfull, G. F. (2011, 10). Cluster K Mycobacteriophages: Insights into the Evolutionary Origins of Mycobacteriophage TM4. *PLoS ONE, 6*(10). doi:10.1371/journal.pone.0026750

29. Primer3 Input. (n.d.). Retrieved April 16, 2016, from http://primer3.ut.ee/

30. Rohwer, Forest. "Global Phage Diversity." *Cell* 113.2 (2003): 141. Print.

31. Schägger, Hermann, and Gebhard Von Jagow. "Tricine-sodium Dodecyl Sulfate-polyacrylamide Gel Electrophoresis for the Separation of Proteins in the Range from 1 to 100 KDa." *Analytical Biochemistry*166.2 (1987): 368-79. Print.

32. Smith, I. (2003, 07). Mycobacterium tuberculosis Pathogenesis and Molecular Determinants of Virulence. *Clinical Microbiology Reviews, 16*(3), 463-496. doi:10.1128/cmr.16.3.463-496.2003

BIOGRAPHY OF THE AUTHOR

Emily E. Whitaker is from Westport Island, Maine. Emily graduated from Lincoln Academy in Newcastle, ME in 2012. She is a Mitchell Scholar and recipient of the Susan Elliot Judd Roxby Memorial Scholarship. Emily was a proud member of Operation H.E.A.R.T.S. and All Maine Women. Emily plans to pursue her PhD at the University of Vermont Cellular, Molecular and Biomedical Science program in Burlington, Vermont.