UNIVERSITY OF CENTRAL OKLAHOMA

Edmond, Oklahoma

Jackson College of Graduate Studies

ISOLATION, IDENTIFICATION, AND *IN SILICO* ANALYSIS OF MYCOBACTERIOPHAGE OKCENTRAL2016

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements

For the degree of

MASTER OF SCIENCE IN BIOLOGY

By:

Christopher James Patton

Edmond, Oklahoma

2018

ISOLATION, IDENTIFICATION, AND *IN SILICO* ANALYSIS OF MYCOBACTERIOPHAGE OKCENTRAL2016

A THESIS

APPROVED FOR THE DEPARTMENT OF BIOLOGY

May 2018

By _____

Name **Committee Chairperson**

_____

Name **Committee Member**

_____

Name **Committee Member**

ACKNOWLEDGMENTS

This project would not have been possible without the patience, assistance and expertise of many individuals. I would like to thank Mr. Greg Strout and the Samuel Roberts Noble electron microscopy lab for their help in taking electron microscopy images. I would also like to thank Mr. Daniel Russell for his assistance in the assembly of the viral genome. I would like to thank Ms. Deborah Jacobs-Sera and Dr. Welkin H. Pope for their comments and edits to the genome annotation. I would like to thank Dr. Michelle Haynie and Dr. Allyson Fenwick for their expertise and assistance in constructing dendrograms. I would like to thank Dr. James Creecy for his assistance in sequencing and utilizing bioinformatics. I would like to thank Dr. Robert Brennan for helping me with biofilms and troubleshooting assistance.

I also want to thank all my friends and family who gave me nothing but love and support throughout this arduous process. Lastly, I would like to thank my thesis advisor, Dr. Hari Kotturi for all the time and effort he has poured into me as both as an undergraduate and a graduate student. He has pushed me and truly galvanized me to be a better student, researcher, and a better man.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

Table **Page**

Glossary

Anticodon – the nucleotide triplet that is complementary to the messenger RNA (mRNA) codon and is present on the transfer RNA.

Amplification – Obtaining high titer phage lysate for downstream applications, which includes: seeding the plates, pooling the phage, and precipitating the phage.

Bacteriophage (phage) – a virus that infects bacteria cells.

Biofilms – a community of microbial cells encased in an exopolysaccharide attached to a solid surface.

Capsid – a protein coat composed of capsomeres forming a protein shell that protects the viral genome.

Capsomere – the monomer subunit that is joined with other monomers to form the capsid.

Clustering – a method used in comparative genomics to depict relationships between organisms using nucleotide sequence similarities.

Colony – visible microbial growth on a solid medium that originated from a single parent cell and gave rise to identical progeny.

Colony forming unit (CFU) – the unit used to describe the bacterial concentration within a bacterial stock solution.

Dendrogram – a tree that depicts relationships based on clustering similarities.

Etiological agent – the disease-causing agent of interest.

Enrichment – a method used to promote the growth of a specific microorganism.

Gene – unit of heredity consisting of sequence of nucleotides present on DNA that can be transcribed and translated into a polypeptide.

Glycanase – an enzyme capable of degrading polysaccharides or glycans.

Holin – a membrane protein coded in some phage genomes, produced at the end of the lytic cycle that degrades the cytoplasmic membrane.

Homologous recombination – an exchange of genetic material that requires sequence similarities between the genomic DNA and foreign DNA.

Illegitimate recombination – an exchange of genetic material that requires no homology between the genomic DNA and foreign DNA for DNA recombination to occur. This process is in eukaryotes, prokaryotes, and viruses.

Lymphadenitis – the enlargement of one or more lymph nodes, typically due to an infection.

Lysin A – a protein found in mycobacteriophage lysis cassette produced by bacteriophages. This protein degrades the peptidoglycan within the bacterial cell wall.

Lysin B – a protein that may be found in the mycobacteriophage lysis cassette that cleaves ester bonds releasing free mycolic acid.

Lysis cassette – a sequence of nucleotides comprising of three to four genes found in phage genomes that functions in lysing the host cell.

Lytic phage – a type of virus that kills the host cell at the end of the replication cycle.

McFarland – a standard used to determine the amount of CFUs/mL based on the turbidity of the solution (0.5 McFarland = 1.0×10^8 CFUs/mL).

Microliter (uL) – A unit of volume that represents 10^{-6} liters or 1/1000 milliliter.

Milliliter (mL) – a unit of volume that represents 10^{-3} liters.

Mycobacteriophage – a specific type of bacteriophage that infects bacteria belonging to the genus *Mycobacterium*.

Mycobacterium smegmatis – an acid-fast, non-pathogenic species that has been used as a model organism to study the genus *Mycobacterium*.

Nosocomial infections – infections that were acquired or originated from a hospital.

Open reading frame (ORF) – a sequence of DNA that begins with a start codon and ends with a stop codon that potentially codes for a polypeptide.

Phage therapy – the utilization of bacteriophages to treat bacterial infections.

Phamilies (phams) – organized families of proteins based on high sequence similarities.

Phenotype – the observable traits displayed by an organism.

Plaque – a clearing of bacterial growth formed by infection of one bacteria by a single virus, which then progresses in a circular pattern

Plasmid – extrachromosomal or naked DNA that codes for non-essential genes and is taken up by bacterial cells.

Plaque forming unit (PFU) – the unit to describe the number of infectious viral particles within a stock solution.

Prophage – a phage genome that is integrated into the host's genome.

Purification – a method used to obtain one type of virus.

Quorum sensing molecules – molecules that regulate metabolic activity in bacteria and play a role in biofilm establishment.

Recombinant phage – a phage that has been bioengineered through genetic editing.

Temperate phage – a phage capable of replicating through the lysogenic cycle.

Titer – the concentration of viruses expressed as infectious units per mL.

Transcription – the process by which a DNA template is synthesized into RNA.

Transfer RNA (tRNA) – a type of RNA that possesses an anticodon that is complementary to the codon nucleotide sequence on the mRNA that is used to transport specific amino acids to the ribosome during polypeptide synthesis.

Translation – the process by which a messenger RNA sequence is converted into a polypeptide.

Transposase – an enzyme that binds to a transposon and cuts and pastes the transposon into a different location in the genome.

Transposon – a DNA sequence that can change its position.

Sebaceous gland secretions – a produced substance composed of oils and fats, often referred to as sebum.

Synteny – describes the order of genes present in a genome that are conserved and found among various genomes.

Virion – a complete infectious viral particle.

Virulence factors – molecules produced by a microorganism that can aid in invasion, evasion, or contribute to the overall pathogenicity of the microorganism.

Abbreviations

ADC – albumin dextrose catalase

CFU – colony forming unit

CV – crystal violet solution

gDNA – genomic DNA

MOI – multiplicity of infection

mRNA – messenger RNA

NTM – nontuberculosis mycobacteria

OD – optical density

ORF – open reading frame

PFU – plaque forming unit

PRF – programmed ribosomal frameshift

QS – quorum sensing

RGM – rapid-growing mycobacteria

SD – standard deviation

SE – standard error

SGM – slow-growing mycobacteria

TB – tuberculosis

TEM – transmission electron microscopy

tRNA – transfer RNA

VBR – virus to bacteria ratio

ABSTRACT OF THESIS

University of Central Oklahoma

Edmond, Oklahoma

NAME: Christopher James Patton

TITLE OF THESIS: Isolation, identification, and *in silico* analysis of mycobacteriophage OKCentral2016

DIRECTOR OF THESIS: Hari Shankar R. Kotturi Ph.D.

PAGES: 63

ABSTRACT: Bacteriophages or phages are specific viruses that are capable of infecting bacterial cells without harming eukaryotic cells. Bacteriophages have been utilized to treat and eliminate bacterial infections, which is known as phage therapy. Phages are the most diverse and abundant group of organisms. Mycobacteriophages are a specific type of phage that infect bacteria belonging to the genus *Mycobacterium*. Mycobacteriophages have been studied due to their potential in killing virulent mycobacteria, such as the causative agents of tuberculosis and leprosy. *Mycobacterium smegmatis* (*M. smegmatis*) mc^2155 has been used as the model organism for studying mycobacteria due to its increased growth rate and non-pathogenicity. *M. smegmatis* mc^2155 is commonly used to isolate mycobacteriophages from soil.

Comparative genomic studies of mycobacteriophages have revealed vast diversity among their genomes. The mycobacteriophage OKCentral2016 was isolated from enriched soil obtained at the University of Central Oklahoma. OKCentral2016 produced transparent plaques, and had a morphology consistent with the *Siphoviridae* morphotype. The phage remained stable at temperatures below 55°C and under neutral pH conditions. The replication cycle took

approximately 4 hours to complete under ideal growing conditions. I also examined the effect of OKCentral2016 on biofilms. OKCentral2016 did not infect any pathogenic mycobacteria tested (*M. avium*, *M. abscessus*, *M. intracellulare*, *M. kansasii*, and *M. simiae*). However, it did decrease biofilm formation by the host bacteria.

Genomic analysis showed that OKCentral2016 closely resembles phages belonging to the subcluster A10. The genome of OKCentral2016 has a GC content of 65.1%, which codes for 83 open reading frames (ORFs). The annotated genomic sequence is available on GenBank (MF773750) and has been published (1). OKCentral2016 is notable because it was the first mycobacteriophage to be isolated, sequenced, and annotated from Oklahoma soil.

Chapter 1: Introduction

1.1 Bacteriophages – Viruses have been described as obligate intracellular parasites. Viral particles are composed predominately of nucleic acids and proteins. The viral genome is composed of only one type of nucleic acid, either DNA or RNA, which is encased and protected by the viral capsid. The capsid is made up of various protein monomers known as capsomeres (2, 3), which are linked together by peptide bonding, disulfide bonding, hydrogen bonding, and van der Waals interactions (4-6). The genetic material and protein shell collectively are referred to as the nucleocapsid. Some viruses have their nucleocapsid surrounded by a structure known as an envelope, which is derived from the host's plasma membrane, endoplasmic reticulum, or Golgi membrane. Most enveloped viruses infect animal cells. However, some viruses lack an envelope and are referred to as nonenveloped or naked viruses. A key difference in enveloped and nonenveloped viruses is their escape strategy. Viruses that are enveloped escape the infected cell through a process known as budding, where the virus leaves the cell without killing the infected cell. Nonenveloped viruses escape the host cell by lysing the infected cell (3, 7). This is a common strategy utilized by bacteria-infecting viruses, which are known as bacteriophages. For example, bacteriophage λ (Figure 1) is a naked bacteriophage that lyses the host cell at the end of its replication cycle (8).

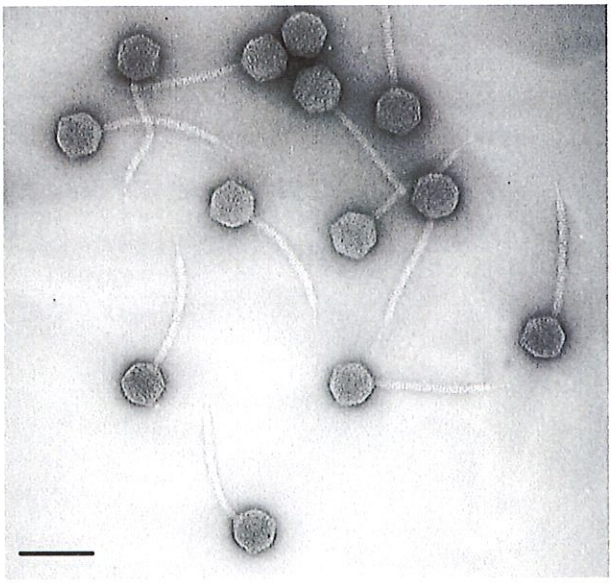

Figure 1. Electron micrograph of the bacteriophage λ.

Source: Todar, Kenneth. *Todar's Online Textbook of Bacteriology.*

Bacteriophage replication may occur in one of two ways and is dependent on gene expression. The first mode of replication is the lytic pathway; the second mode of replication is known as the lysogenic pathway. Bacteriophages that utilize both pathways are known as temperate phages (Figure 2). In both pathways, the virus attaches to the bacterial receptor, and the viral genome is injected into the bacterial cell. In the lytic pathway, the viral genome uses the replicative machinery of the host cell to copy its genome and synthesize its proteins. The proteins are assembled and the genome is packaged inside the viral capsid resulting in an assembled virus. The assembled virus induces bacterial cell lysis allowing viral progeny to escape; once the assembled virus exits the host cell it is referred to as a virion. The virion then infects a new host cell and the replication cycle is repeated.

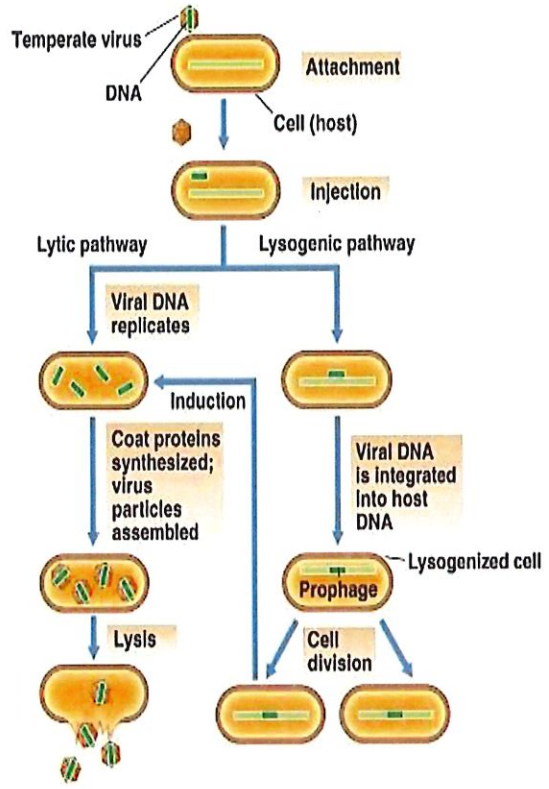

Figure 2. The lifecycle of a temperate bacteriophage undergoing replication in both the lytic and lysogenic pathways.

Source: Maddigan, *et al.* (2009). *Brock Biology of Microorganisms* San Francisco, CA, Pearson Education Inc.

In contrast, during the lysogenic pathway the injected viral DNA is integrated into the bacterial chromosome, or can be present in the bacterial cytoplasm in the form of a plasmid.

Integration of the bacteriophage into the host genome, which is referred to as a prophage, can occur at preferred sites such as transfer RNA (tRNA) sequences (9). Integration at preferred sites is mediated by a phage-encoded integrase, which catalyzes recombination between the *attB* site on the bacterial genome and the *attP* site on the bacteriophage genome (10, 11). It also has been shown that the *attB* site can be near a tRNA sequence on the bacterial genome (12). However, a less common form of integration also can occur by using a phage produced transposase to cause phage integration at random locations in the bacterial chromosome (13). Bacteriophage genomic replication only occurs when the host genome undergoes replication. However, under unfavorable conditions such as pH, temperature, ultraviolet (UV) light, or the presence of antibiotics (14, 15), DNA damage, or other stress factors trigger the SOS DNA repair mechanism in host cells (16). The SOS response also causes prophage induction, resulting in the excision of the integrated phage from the bacterial chromosome (17-19). The bacteriophage will exit and enter the lytic pathway, inducing bacterial cell lysis (3).

1.2 Bacteriophage diversity – Phages are considered the most abundant and diverse group of organisms in the biosphere (20). It is estimated that there are approximately 10^{32} phages on our planet (21), and a predicted number of 10^{8} different phage species (22). In one gram of topsoil, it is estimated that there are approximately 10^{7} to 10^{9} phages present (23, 24). This results in approximately 10^{23} phage infections per second (25). Bacteriophages occupy the vital role of regulating bacterial populations in the environment (26). Approximately 30% to 70% of all aquatic bacterial cells are lysed by bacteriophages every day (27, 28). In aquatic environments, phages are abundant and account for 94% of all nucleic acid containing particles. Despite this, phages only account for 5% of the total biomass (Figure 3) (29). The virus to bacteria ratio (VBR) has been

shown to be 100 to 1,000 fold higher in various soils when compared to aquatic environments (30-32).

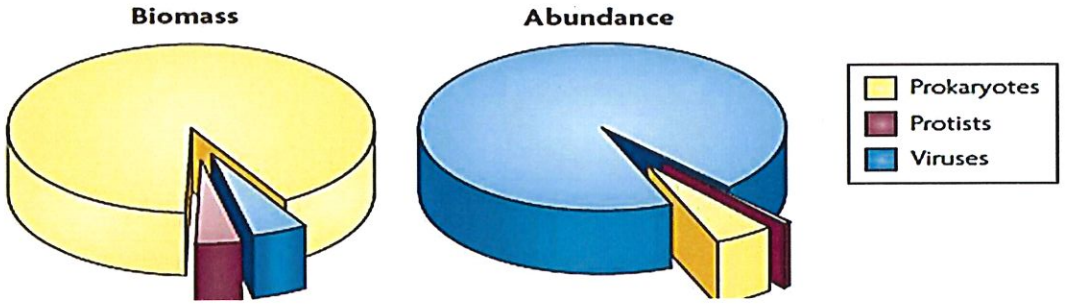

Figure 3. The comparison of biomass between prokaryotes (≥90%) and viruses (5%). However, viruses are abundant and are responsible for the majority of nucleic acid containing particles (94%), in comparison to prokaryotes (10%≤)

Source: Suttle, C. A. (2007). "Marine viruses - major players in the global ecosystem " Nature Reviews Microbiology 5: 801-812.

1.3 *Mycobacterium* – The genus *Mycobacterium* is composed of over 170 species and is the only genus within the Mycobacteriaceae family (33). *Mycobacterium* is composed of acid-fast, obligatory aerobic bacteria, and are Gram-variable due to the inability to be readily stained using Gram staining (34). The Gram-variability is due to the presence of mycolic acid in the cell wall. Mycobacteria are ubiquitous microorganisms and commonly are isolated from household plumbing (35), and various soil types such as those found on ranches, landfills, and boreal coniferous forests (36-39). Two of the more medically important mycobacterial species are *M. tuberculosis* and *M. leprae*, which are the causative agents of tuberculosis and leprosy, respectively. Mycobacterial species that do not cause tuberculosis are termed nontuberculosis mycobacteria (NTM), and are divided into two distinct groups, slow-growing mycobacteria (SGM) and rapid-growing mycobacteria (RGM) (40). Due to slow generation time, some SGM species take up to 60 days for colonies to appear on an agar plate (41).

1.4 *M. smegmatis* mc^{2}155 – *M. smegmatis* is the model organism to study the genus *Mycobacterium*, because it is a RGM and is not pathogenic (42, 43). *M. smegmatis* also has been shown to be part of the normal flora in human sebaceous gland secretions (2). One of the best-studied strains of *M. smegmatis* is the mc^{2}155 strain (44, 45). *M. smegmatis* mc^{2}155 was derived from the parent strain mc^{2}6 (46) and is resistant to carbenicillin.

1.5 Mycobacteriophages – Mycobacteriophages are bacteriophages that are capable of infecting a mycobacterial host (47). Their genomes are more diverse than other bacteriophage groups that infect *Staphylococcus* or *Pseudomonas* (48). According to The Actinobacteriophage Database (phagesdb.org), 9,668 mycobacteriophages have been isolated and 1,566 have been sequenced. These sequenced genomes are highly variable and have a percent GC (GC%) content that ranges from 50% to 70% (49, 50). The GC% content of bacteriophages is similar to their host GC%; this could result in mycobacteriophages having large host ranges. Therefore, mycobacteriophages have been studied for their therapeutic potential against NTM and *M. tuberculosis* (51, 52) infections. For example, mycobacteriophages have been utilized as a delivery system for inserting transposons into *M. tuberculosis* (53).

1.6 Pathogenic mycobacteria and biofilm formation – Tuberculosis (TB) is one of the most notorious diseases caused by mycobacteria, and has killed more than one billion people in the last two centuries (54). According to the World Health Organization, TB continues to be one of the top ten causes of deaths worldwide (55). However, there has been an increase in NTM pulmonary infections. The most common NTM agents that cause pulmonary disease are *M. avium* subspecies *avium*, *M. intracellulare*, *M. kansasii*, *M. simiae*, and *M. abscessus*. These bacterial pathogens can infect various organs. However, the most common are pulmonary disease, lymphadenitis, and soft tissue infections (56). NTM are ubiquitous and have been isolated in

various water sources; *M. avium* was even detected in water on Mir, the Russian Space Station (57). The prevalence of bacteria in potable water is attributed to the formation of biofilms (58).

Biofilm formation is one of the key survival strategies implemented by bacteria. Biofilms are formed when bacteria aggregate together on a solid surface and become encased in an exopolysaccharide matrix, which is secreted by the bacteria. Cell-cell communication occurs through the accumulation of quorum sensing (QS) molecules, which allows for expression of specific genes that can regulate metabolic activity and contribute to biofilm formation (59, 60). Biofilms allow bacteria to overcome the pressures of the natural environment such as pH, UV light, and desiccation (61), which is why biofilms are described as the most successful form of life (62). Approximately 80% of all bacterial infections are associated with bacterial biofilms (63).

The synthesis of mycolic acid has been shown to be closely related to biofilm formation in *M. smegmatis*. The mycolic acid profile was shown to differ in *M. smegmatis* when comparing planktonic cells to biofilm-encased cells (64, 65). Mycolic acid is a substance within the mycobacterial cell wall which causes the bacterium to be acid-fast (2, 66). The presence of mycolic acid is thought to influence bacterial attachment by increasing cell surface hydrophobicity (67), which would affect biofilm formation.

The ability for NTM to produce biofilms has been demonstrated to increase their resistance to disinfectants and antibiotics (68, 69). This has become a problem in hospital settings due to the buildup of biofilms on medical equipment (70). Biofilms that persist through chemical treatment remain on medical devices and serve as bacterial reservoirs (71). These biofilms have contributed to the rise of post-surgical nosocomial infections caused by NTM (72, 73), leading to chronic infections. For example, some chronic pulmonary infections have been linked to *M. abscessus* biofilm formation (74, 75).

1.7 Bacteriophage applications – Before the discovery of the first antibiotic, penicillin in 1929 (76), lytic phages had been used to treat bacterial infections. This was known as phage therapy (77). In 1896, phages were first used to treat diarrheal disease caused by *Vibrio cholerae* (cholera) (78, 79). Eventually, phages were used for treating infections caused by bacteria such as *Yersina pestis* (plague), *Salmonella typhi* (typhoid fever), *Shigella dysenteriae* (dysentery), *Neisseria meningitis* (meningitis) (80), and acute surgical infections (81). However, phage therapy was later abandoned due to the production of antiphage antibodies by the host, which mandated lengthier treatments (82, 83).

The advantages of using phage therapy to treat bacterial infections are that phages are bactericidal agents, while some antibiotics are only bacteriostatic (84). Additionally, bacteriophages are self-replicating and self-limiting. Once administered, if the phage infects the target bacteria, it will replicate. However, if the target bacteria is absent, the phage will not replicate making it a self-limiting treatment. Bacteriophage replication would occur at the infection site, therefore minimal doses would be required to treat the infection (85). Bacteriophages are nontoxic and have been shown to treat biofilms (86, 87). Minimal side effects also have been observed with phage therapy (85).

Some of the disadvantages of phage therapy are that phages generally have a narrow host range, only infecting a few strains, generally requiring a cocktail of phages to be used (80). However, the narrow host range also has been described as beneficial by limiting the phage's infectivity of the host's normal flora (85). In comparison, broad spectrum antibiotics target a large range of bacteria, eliminating host commensal bacteria, which can result in secondary bacterial infections (88). Bacterial resistance to the phage through CRISPR-Cas system or modification of bacterial surface receptors is a drawback to phage therapy. If resistance does occur, new phages

can be readily isolated from the environment, which is why phage discovery has become a heavily studied area (89). However, not all bacterial infections can be treated using phage therapy. Not all phages can be used in phage therapy (90). The phages selected for phage therapy must be evaluated prior to their implementation.

Bacteriophages and their modes of replication have been studied and used in various molecular techniques. For example, the T7 RNA polymerase and its associated promoter were derived from a T7 bacteriophage that infected *E. coli*. The T7 promoter is utilized to increase the transcription rate of cloned genes (91, 92). Bacteriophages also have been used as a delivery vector for recombinant subunit vaccines, by attaching specific proteins onto the viral capsid. The host then begins producing antibodies against the attached protein (93). Phages have been shown to be valuable in detecting specific pathogens. Currently, different assays have been developed for the detection of *M. tuberculosis (94)*, *Yersinia pestis* (95), and *Listeria* (96). There also have been efforts to develop a phage detection assay against the endospores of *Bacillus anthracis*, the causative agent of anthrax. However, the assay had poor efficacy due to the phage binding to endospores of other *Bacillus* species (97).

Recently, developments in biotechnology have been applied to bacteriophages. Recombinant engineering of bacteriophage tail fibers have been used to broaden the host range of a T3 phage that infects *E. coli*. These recombinant phages have been shown to increase infectivity when compared to the original phages (98, 99). Temperate bacteriophages have been used to deliver the CRISPR-Cas system to destroy regions within plasmids that code for antibiotic resistance within bacteria. This resulted in the targeted bacteria reverting back to being sensitive to antibiotic treatment (100), thus demonstrating the efficacy and likely future applications of recombinant bioengineering.

1.8 Research Objectives – The main goal of this project was to isolate, characterize, and sequence a mycobacteriophage from Oklahoma soil. If successful, my virus will be the first mycobacteriophage sequenced from Oklahoma soil. I will explore the therapeutic potential of my virus against common NTM. My research also will yield valuable insights into the diversity and evolutionary relationships between my virus and other mycobacteriophages.

Chapter 2: Materials

Table 1. Bacterial strains used in this study and their source.

Bacterial Strain	Source
Mycobacterium avium DJO-44271	BEI resources
Mycobacterium abscessus MC1518	BEI resources
Mycobacterium intracellulare 1956	BEI resources
Mycobacterium kansaii 824	BEI resources
Mycobacterium simiae MO – 323	BEI resources
Mycobacterium smegmatis mc^2155	Hatfull Lab

Table 2. The composition of different buffers and media used in this study.

Buffer/Media	Composition (1 L)
7H9 Broth	4.7 g 7H9 Broth Base (Middlebrook) 5 mL 40% Glycerol 100 mL ADC 1 mM $CaCl_2$ 900 mL DI H_2O
7H10 Agar Plates	19 g 7H10 Agar Base 12.5 mL 40% Glycerol 12.5 mL 40% Dextrose 990 mL DI H_2O
Albumin Dextrose Catalase (ADC)	20 g Dextrose 145 mM NaCl 50 g Albumin 950 mL ddH_2O
Phage Buffer	10 mM Tris pH 7.5 10 mM $MgSO_4$ 70 mM NaCl 1 mM $CaCl_2$ Dissolved in 980 mL ddH_2O
Soft Agar Overlay	1 L 7H9 Broth 0.4% Agar

Table 3. Restriction enzymes used in this study and their source.

Restriction Enzyme	Restriction Site	Source
BamHI	5'...G▼GATC C...3' 3'...C CTAG▲G...5'	New England Biolabs
ClaI	5'...AT▼CG AT...3' 3'...TA GC▲TA...5'	New England Biolabs
EcoRI	5'...G▼AATT C...3' 3'...C TTAA▲G...5'	New England Biolabs
HaeIII	5'...GG▼CC...3' 3'...CC▲GG...5'	New England Biolabs
HindIII	5'...A▼AGCT T...3' 3'...T TCGA▲A...5'	New England Biolabs

Table 4. List of software used to annotate the viral genome or build trees.

Software	Purpose	Reference
Aragorn	Identify tRNA sequences	(101)
Consed	Genome Assembly	(102)
DNA Master	Genome Annotation	http://cobamide2.bio.pitt.edu/
Glimmer	Predict Gene Coding Potential	https://www.ncbi.nlm.nih.gov/genomes/MICROBES/glimmer_3.cgi
GeneMark	Predict Gene Coding Potential	(103, 104)
MEGA7	Construct Dendrograms	(105)
Newbler	Genome Assembly	https://swes.cals.arizona.edu/maier_lab/kartchner/documentation/index.php/home/docs/newbler
Phamerator	Predict Gene Function	(106)
Phage Enzyme Tools 2.0	Preliminary Cluster Assignment	(107)
Starterator	Evaluate Gene Start Codons	https://seaphages.org/software/
tRNAscan-SE	Identify tRNA sequences	(108)

Table 5. Evolutionary models used for dendrograms.

Cluster/Subcluster	Protein	Evolutionary Model Used	Model Reference
A	DNA polymerase	Whelan and Goldman (+ Gamma)	(109)
A	Lysin A	Whelan and Goldman (+ Gamma and Freq.)	(109)
A	Major capsid protein	Whelan and Goldman (+ Gamma)	(109)
A3, A4, A10	DNA polymerase	Whelan and Goldman (+ Gamma)	(109)
A3, A4, A10	Lysin A	Whelan and Goldman (+ Gamma)	(109)
A3, A4, A10	Major capsid protein	Le Gascuel (+ Gamma)	(110)

Methods

2.1 Culturing bacterial strains – All bacterial strains (Table 1), including the host bacterium (*M. smegmatis* mc^2155), were grown in 7H9 broth supplemented with albumin dextrose catalase (ADC)(10% V/V) and 1 mM $CaCl_2$ unless stated otherwise. Liquid bacterial cultures were incubated in a shaking incubator set at 37°C. On solid media, the bacteria were grown on 7H10 agar supplemented with 0.2% glucose and 1 mM $CaCl_2$. The bacteria grown on 7H10 agar plates were incubated at 37°C. The composition of media and buffers used for experimentation are in Table 2.

2.2 Soil enrichment – Soil was collected from the southwest corner of the University of Central Oklahoma (35.653371 N, 97.474118 W) near the Coyner Health Sciences building. The collected soil was enriched using a double enrichment method (1). Three grams of the collected

soil was enriched with 10 mLs of 7H9 broth and 1 mL of host bacterium. The enriched culture was incubated at 37°C for 24 hours. Following incubation, 10 mLs of phage buffer was added to the mixture; the mixture was vortexed and centrifuged at 1,589 x g for 10 min. Following centrifugation, 9 mLs of supernatant was removed and added to 1 mL of 10X 7H9 broth, which was supplemented with ADC and 1 mM $CaCl_2$. Approximately 1 mL of bacterial culture was added and incubated at 37°C for 24 hours.

2.3 Bacteriophage isolation – Following the soil enrichment step, phage lysate was centrifuged at 12,000 x g for 10 min. The supernatant was filtered using a 0.2 μm filter to remove cell debris, and the phage suspension was serially diluted. Fifty μL of diluted phage was added to 450 μL of host bacterium. The phage and host bacterium were mixed and incubated at room temperature for 10 min. Phage was plated using the agar overlay method in which the incubated phage and bacteria mixture was added to 0.4% soft agar, and then overlayed on top of a 7H10 agar plate (111). Once the soft agar had solidified, the plates were inverted and incubated at 37°C for 24 hours. Following incubation, the plates were examined for plaques, which were clear zones that formed due to the lysis of the host bacterium (112). Plaques were measured in mm following 24 hours of incubation.

2.4 Plaque purification – Once a single isolated plaque was identified, the plaque was removed using a sterile 10 μL micropipette tip and transferred into a host bacterial culture. This was repeated three times to ensure isolation and purification of a single type of mycobacteriophage. During this process, the dilution factor of a webbed plate, or webbed plaque distribution, also was determined. A webbed plate was a plate possessing a concentrated amount of phage resulting in minimal bacterial lawn growth (113).

2.5 Amplification and precipitation of phage – The isolated phage was amplified by seeding 30 plates, using the webbed plate titer. Approximately 7 mLs of phage buffer was added to webbed plates. The plates were incubated with phage buffer in a static environment for 8 hours at room temperature; plates were gently agitated every two hours. The phage lysate was pooled and centrifuged at 5500 x g for 10 min. at 4°C. The supernatant was transferred to a sterile flask. Phage was precipitated by adding 1 M NaCl (114) and polyethylene glycol 8,000 (10% V/V), and incubated overnight at 4°C (115, 116). Following incubation, the solution was centrifuged at 5500 x g for 10 min. at 4°C. The supernatant was discarded and the pellet containing the phage was resuspended in phage buffer. This solution was incubated at 4°C for 24 hours with continuous gentle agitation on an orbital shaker, which allowed for phage diffusion into the buffer. The phage was centrifuged at 5500 x g for 10 min at 4°C. The supernatant, containing the concentrated phage, was removed and was used as the phage stock. The phage stock's titer was determined and the phage stock was used for all experiments.

2.6 Phage titer – Upon evaluation of plaque formation the phage titer was determined to know the concentration of phage in the phage stock. The phage titer was described as the number of plaque forming units (PFUs)/mL. The phage titer was determined using the following equation (117): phage titer $= \frac{\#\ of\ Plaques\ on\ Plate}{Total\ Dilution\ of\ Plate}$.

2.7 Electron microscopy – Electron microscopy was used in the determination of the viral morphotype. High titer phage (10^9 PFUs/mL) lysate was added to a carbon coated electron microscope grid and negatively stained using 1% uranyl acetate (118). Images were obtained using a JEM – 2000FX scanning transmission electron microscope (TEM) (119) at 80,000 X total magnification. Bacteriophage measurements of the head and tail were obtained using the ImageJ software package (120). The diameter of the icosahedral head was obtained by measuring from

one vertex to another vertex on the opposite side. Three different measurements were taken and averaged. The tail was measured from the beginning of the tail, which was adjacent to the icosahedral head, to the bottom of the tail.

2.8 Viral gDNA extraction – The genomic DNA (gDNA) was extracted using a sodium dodecyl sulfate (SDS)/phenol:chloroform: isoamyl alcohol (PCI) (25:24:1 V/V) extraction method following the protocol of Green and Sambrook (122, 123). A phage suspension of approximately 10^9 PFUs/mL was used for gDNA extraction. Phage lysate was treated with 1 M $MgCl_2$, 7 µL of DNase, and 7 µL of RNase, and then incubated for 30 min. at room temperature. The lysate was treated with 0.5 M ethylenediaminetetraacetic acid (EDTA), 10 µL of proteinase K, and 10% SDS, and incubated for 1 hour at 55°C. A 1:1 ratio of PCI was added to the DNA and centrifuged at 17,532 x g for 5 min. at room temperature. The top aqueous layer, containing the DNA, was removed and another 1:1 ratio of PCI was added and centrifuged as previously described. The top aqueous layer was removed and placed in a sterile microcentrifuge tube; 3 M sodium acetate (10% V/V) and ice-cold 95% ethanol (250% V/V) was added to the DNA and the DNA was incubated at -80°C for 10 min. The DNA was centrifuged as previously described at 17,532 x g for 10 min. at room temperature. The DNA was washed with 70% ethanol and centrifuged at 17,532 x g for 5 min. at room temperature. The pelleted DNA was decanted and resuspended in diethyl pyrocarbonate treated nanopure water. The DNA quantity and quality was determined by using UV – spectrophotometry (Nanodrop) (124, 125), and DNA quality was further assessed using gel electrophoresis. A 1% agarose gel containing ethidium bromide at a final concentration of 0.5 µg/mL (126) was used to visualize the DNA under UV light.

2.9 Restriction enzyme digestion – The extracted gDNA was treated with five different restriction enzymes (Table 3). The restriction digest was performed according to the

manufacturer's recommended protocols (New England Biolabs). Preliminary viral cluster assignment was determined by the pattern of DNA fragmentation produced by enzymatic activity (127). DNA fragmentation was observed by using slab-gel electrophoresis as previously described. An undigested DNA sample was used as a control. A DNA ladder was used to determine the approximate genome size. The results for restriction digestion were confirmed using the DNA Master software package (Table 4).

2.10 Determining the thermal stability of phage – Thermostability of the phage particle was determined by incubating the phage at three different temperatures, 50°C, 55°C, and 60°C at three different time points (20, 40, and 60 min.). The incubated phage was diluted and plated using the previously described agar overlay method. Plates were incubated at 37°C for 24 hours. Plaques were counted, and the number of PFUs/mL was determined as discussed previously.

2.11 Determining pH stability of phage – The pH stability of the virion was determined by incubating the phage in phage buffer with adjusted pHs (5, 6, 7, 8, and 9) for 1 hour. Adjustment of the pH was done using a pH meter and the addition of HCl or NaOH to the phage buffer; the phage buffer, with an adjusted pH, was then autoclaved to ensure sterilization. Phage was added to phage buffer at a specific pH and incubated at 50°C for 1 hour. Following incubation, the phage was diluted, plated, and incubated using the agar overlay method as described above.

2.12 One-step growth curve – This method was describe by Catalao et al. (121) at multiplicity of infection (MOI) of 1.0 (one bacterial cell to one bacteriophage). The host bacterium and phage were incubated at 37°C for 50 min. Following incubation, 0.4% H_2SO_4 was added to inactivate the unattached phage particles (121), and the solution was incubated again for 5 min. at room temperature. The H_2SO_4 was neutralized by adding 0.4% NaOH. The bacteriophage suspension was diluted in 7H9 broth, and the suspension was brought to a total volume of 10 mLs.

Every 30 min., for a duration of 4 h, 1 mL of suspension was diluted and plated using the agar overlay method as previously described. Plates were incubated at 37°C for 24 hours. Plaques were counted and the viral titer was determined. This experiment was repeated four times.

2.13 Effect of phage on biofilm formation – The biofilm formation media used in these experiments was composed of working 7H9 broth supplemented with 100 μM $CuSO_4$ (128). A 1.0 MOI was used by initially seeding 50 μL of phage (10^8 PFU/mL) onto a 96-well tissue culture plate. Approximately 100 μL of host bacterium (*M. smegmatis* mc^2 155) at a 0.5 McFarland (1.0 x 10^8 CFU/mL) (129) concentration was added to the 96-well plate, and plates were incubated statically at 37°C for 72 hours.

To investigate the effect of OKCentral2016 on a pre-established biofilm, host bacterium was seeded onto 96-well tissue plates and incubated with biofilm-forming media at 37°C for 72 hours. Planktonic cells were removed and biofilms were washed one time with 200 μL of sterile nanopure water. Fifty μL of phage and 150 μL of biofilm forming media were added to the wells, and incubated statically at 37°C for 72 hours.

Biofilm quantification for both assays was conducted in the same manner. Planktonic cells were removed, and biofilms were washed two times with 200 μL of sterile nanopure water. Approximately 100 μL of a 10% crystal violet solution (CV) was added to the wells and incubated for 10 min. at room temperature. CV was removed and the wells were washed three more times with sterile nanopure water. Biofilms were decolorized by adding 100 μL of 95% ethanol, and plates were incubated for 10 min. at room temperature. The ethanol was removed and added to a sterile 96-well plate. Biofilm growth was analyzed using an ELx808 microplate reader (BioTek) at an OD_{570} nm. Phage buffer treated wells were used as a negative control for both assays. In each experiment, 12 sample replicates were used, and experiments were repeated two additional times.

2.14 Determination of host range – The host range of the virus was tested using pathogenic, BSL-2 mycobacterial species. The different hosts that were tested were: *M. avium*, *M. abscessus*, *M. intracellulare*, *M. kansasii*, and *M. simiae*. The phage was enriched with the host bacterium for 48 hours. Following incubation, the phage and host bacterium were filtered and plated using the agar overlay method as described above. The plates were incubated at 37°C and, due to the diverse growth rates of mycobacterial species, the plates were inspected for plaque formation every day for a duration of 30 days. *M. smegmatis* was used as a positive control and was evaluated using a spot assay (130). In this assay, approximately 10 µL of high titer phage ($\geq 10^9$ PFUs/mL) was added to a mature lawn of host bacterium and incubated for 30 min. at room temperature to allow for phage adsorption. Plates were incubated at 37°C for 24 hours. Phage buffer also was added and served as a negative control. After obtaining results from host range, all bacterial lawns were subjected to acid-fast staining (131) to ensure the bacterial lawn was exclusively composed of an acid-fast bacterium.

2.15 Genomic sequencing and annotation – The bacteriophage genome was sequenced using the Illumina (Solexa) miseq platform with an approximate coverage of 121,165 times the length of the viral genome. Genome assembly was done using consed (102) and Newbler. DNA Master was used to annotate the viral genome. DNA Master uses the software Glimmer and GeneMark (132) to predict the coding potential of open reading frames (ORFs). Starterator (https://seaphages.org/software/) was used to evaluate all potential gene start codons. Gene function was determined using Phamerator (106) and BLASTp on phagedb.org. Aragorn (101) and tRNAscan-SE (108) were used to analyze tRNA sequences within the viral genome. All bioinformatic and dendrogram construction software used are depicted in Table 4.

2.16 Construction of trees – Bacteriophage amino acid sequences, GC%, and protein phams were obtained from phagesdb.org. Amino acid sequences were aligned using MUSCLE (133) and evolutionary models were analyzed using the MEGA software package (105), by which various dendrograms were constructed and analyzed. The evolutionary models used for maximum likelihood analyses are depicted in Table 5.

2.17 Statistical analyses – All experiments were performed using at least three biological replicates, and experiments were repeated at least two additional times to confirm the results. Data points for the one-step viral growth curve and biofilm assays were pooled and graphed with ± standard error (SE). However, for pH stability and thermal stability, data points were graphed with ± standard deviation (SD). For pH stability and biofilm assays, a two sample independent t-test was used to determine statistical significance. The student t-tests were done using the analysis tool pack within the Microsoft Excel software package (2016).

Chapter 3: Results

3.1 Plaque formation – The plaques formed by our virus (Figure 4) were transparent and approximately 4 mm in diameter after 24 h of incubation. The virus was named OKCentral2016 because it was isolated from campus soil at the University of Central Oklahoma in the year 2016.

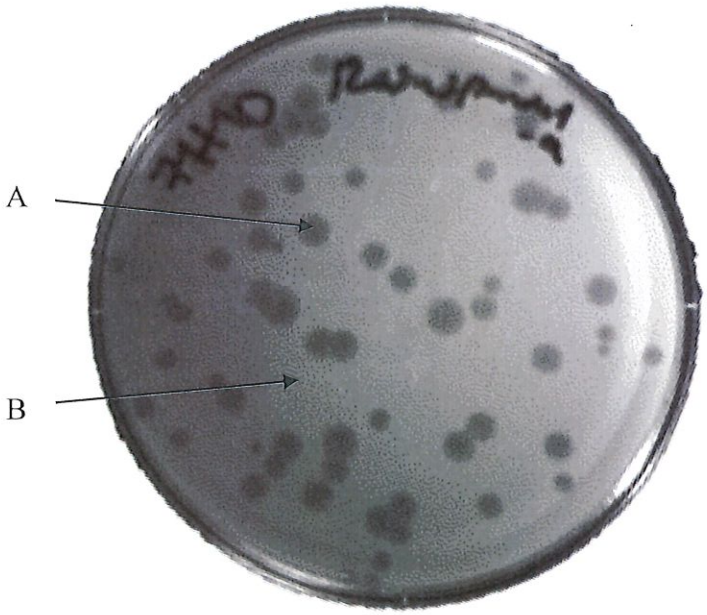

Figure 4. Plaques (A), formed by mycobacteriophage, OKCentral2016, on a bacterial lawn of *M. smegmatis* mc^2155 (B).

3.2 Restriction enzyme sensitivity of OKCentral2016 gDNA – The gDNA of OKCentral2016 contained restriction sites for two enzymes, Bam HI and HaeIII (Figure 5). Digestion by BamHI, resulted in the formation of 14 bands at various molecular sizes between 48,000 to 1,000 bp. The second enzyme that cleaved the gDNA was HaeIII, which resulted in the formation of a DNA smear. However, ClaI, EcoRI, and HindIII did not digest the gDNA,

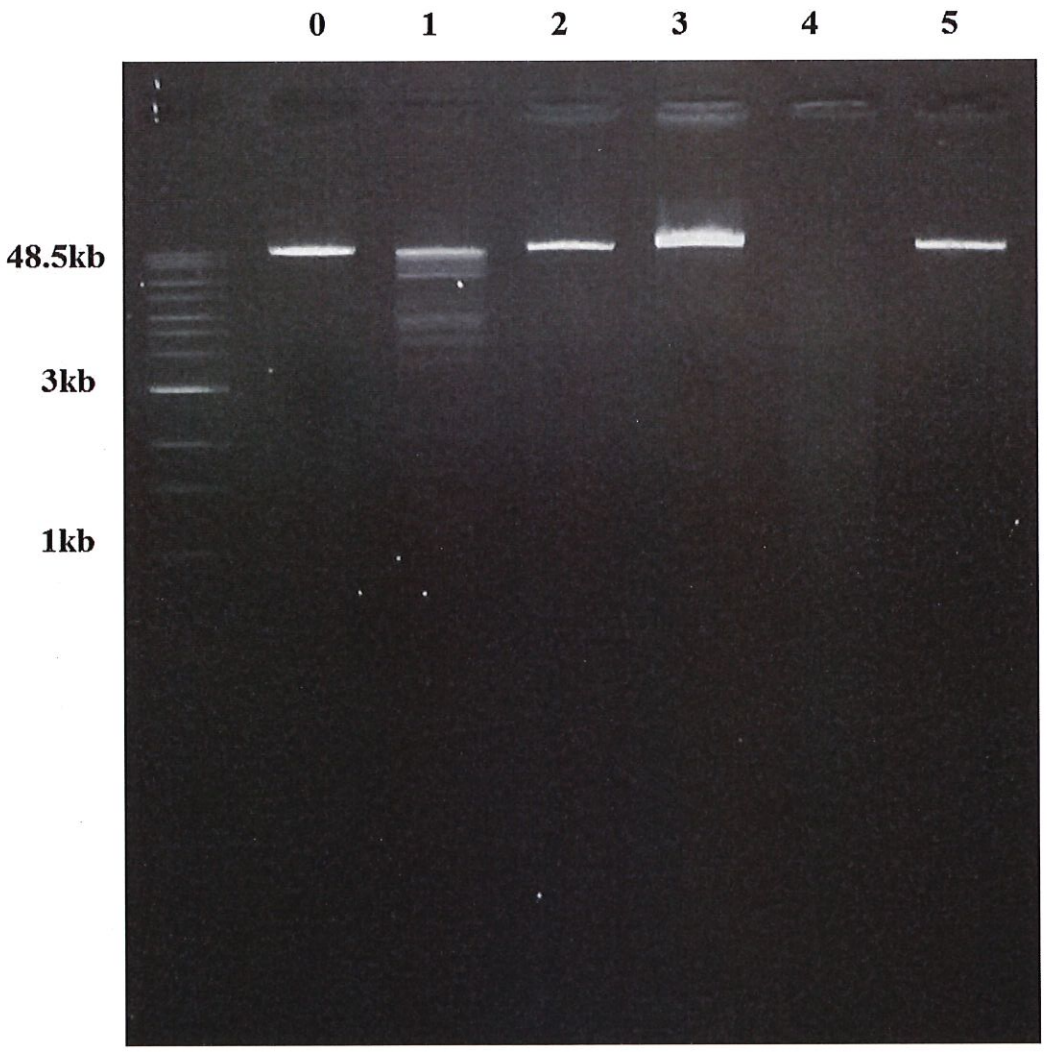

Figure 5. The treatment of OKCentral2016 gDNA with restriction enzymes. Untreated gDNA served as a control (0); gDNA was treated with the following restriction enzymes: BamHI (1), ClaI (2), EcoRI (3), HaeIII (4), and HindIII (5).

indicating absence of restriction sites for these enzymes. These results were compared to the untreated gDNA. Other members of subcluster A10 had similar results. The *in silico* analysis of the restriction digestion showed the same results, where BamHI and HaeIII cleaved the DNA 14 and 542 times, respectively.

3.3 Phage morphology – The morphology of the OKCentral2016 virion was observed using TEM (Figure 6). OKCentral2016 was shown to have a *Siphoviridae* morphotype. A characteristic of the *Siphoviridae* morphotype is an icosahedral head that is estimated to be 60 nm in diameter. The average diameter of our phage's head was 56.152 (±5.187) nm. *Siphoviridae* have a thin long tail, which is flexible and non-contractile (134). The full length of the non-contractile tail was approximately 130.485 nm.

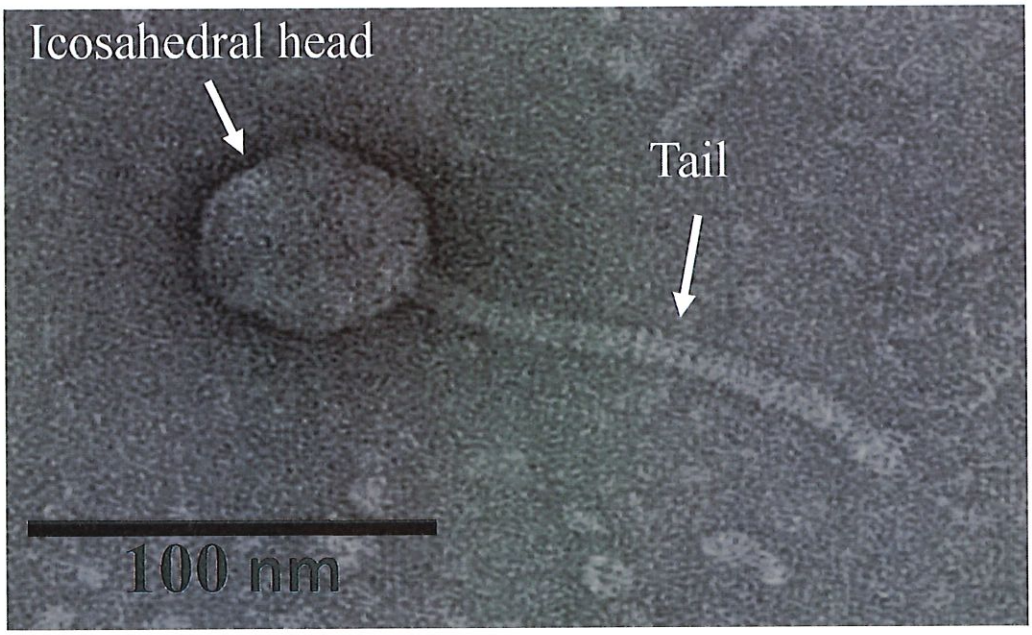

Figure 6. A transmission electron micrograph of OKCentral2016.

3.4 Stability of phage – The thermal stability assay was used to determine if OKCentral2016 virion was stable at various temperatures (Figure 7A). A decrease in viable PFUs was found in all three temperatures (50°C, 55°C, and 60°C). Following 1 hour of incubation at 50°C, there was a minor decrease in viable phage (<1 log). However, following 1 hour of incubation at 55°C, phage viability decreased by 1 log. More than 99% of phages lost their infectious ability after 20 min of incubation at 60°C. Lastly, after 40 min. of incubation at 60°C, no viable PFUs were detected. OKCentral2016 demonstrated a decrease in viable PFUs at all three tested temperatures, which agrees with the current literature on mycobacteriophages (135).

The pH stability of the OKCentral2016 virion was analyzed subjecting the phage to various pHs for one hour (Figure 7B). There was no reduction in infectious PFUs after 1 hour of incubation in a pH 6, 7, or 8. A decrease in infectious PFUs was observed at a more acidic pH of 5 and a basic pH of 9. A student's t test determined a significant decrease in phage viability, where $p<0.05$ was considered significant.

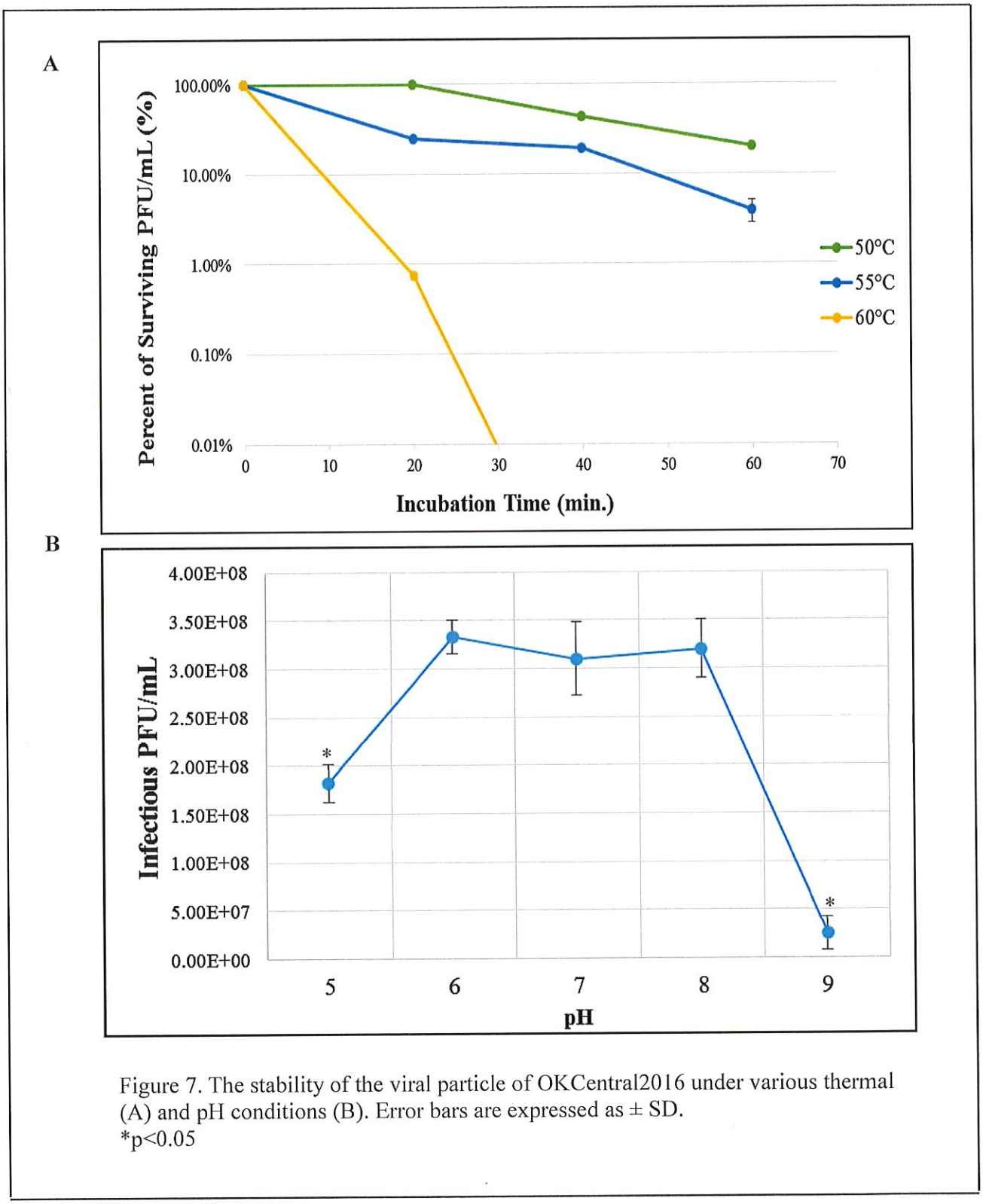

Figure 7. The stability of the viral particle of OKCentral2016 under various thermal (A) and pH conditions (B). Error bars are expressed as ± SD.
*$p<0.05$

3.6 One-step growth curve of OKCentral2016 – Viral replication of OKCentral2016 was characterized by using a one-step growth curve assay (Figure 8). It took approximately 1.5 h for OKCentral2016 to exit the latent period and being producing viral progeny in the rise period. Following the rise period, which lasted approximately 1.5-2 hours, the number of PFUs plateaued, which was indicative that viral replication had stopped in host cells. However, other phages such as SWU1 had a latent period that lasted only 30 min. (136) whereas other phages such as BO1 and BO2a had latent periods that lasted 150 min. and 260 min., respectively (137). This further demonstrates the great diversity among mycobacteriophages.

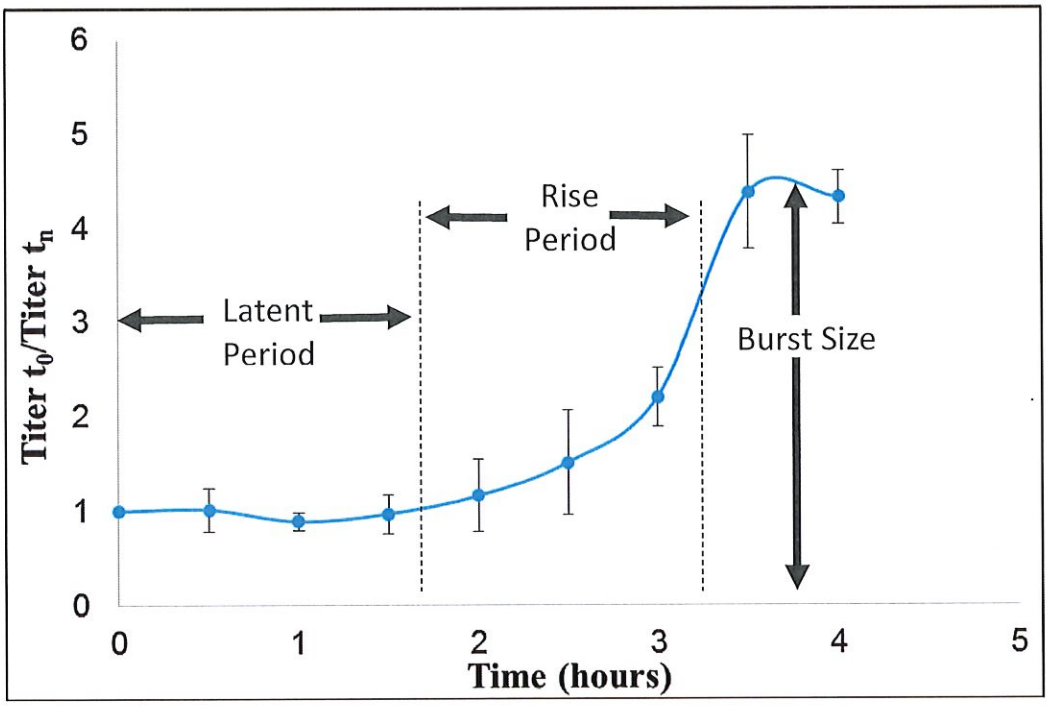

Figure 8. The one-step viral growth curve of OKCentral2016 infecting *M. smegmatis* mc^{2}155. Error bars are expressed as ± SE.

3.7 Phage's effect on biofilm formation – A significant decrease in biofilm formation was observed when planktonic cells were mixed with phage (Figure 9A). However, when phage was added to a pre-formed biofilm, a significant increase in biofilm was observed (Figure 9B). Phage buffer, which lacked phage, was used as the negative control in both assays, where $p<0.05$ was considered statistically significant.

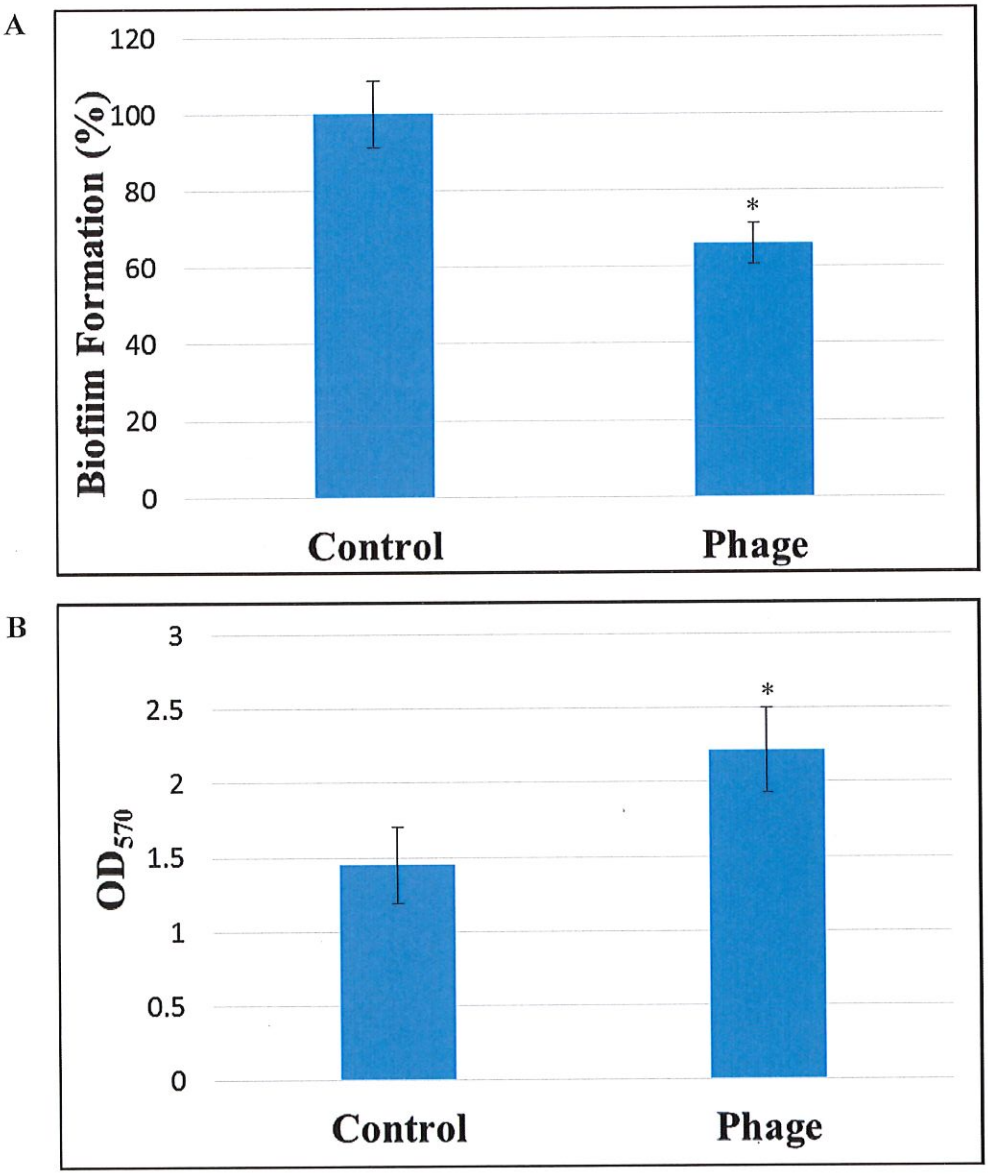

Figure 9. Adding OKCentral2016 to a planktonic culture of *M. smegmatis* mc^2155 significantly decreased biofilm formation (A). However, adding phage to an established biofilm, significantly increases biofilm (B). Error bars are expressed as ± SE.
*$p<0.05$

3.12 OKCentral2016 other mycobacterial species – Five other mycobacterial species (Table 1) were shown to be unaffected when OKCentral2016 was added to the bacterial broth and plated (Figure 10). The phage was enriched in the host culture for 48 hours and then plated. None of the plates had any plaques form on the bacterial lawns. *M. smegmatis* mc^2 155 was used as a positive control using a spot assay.

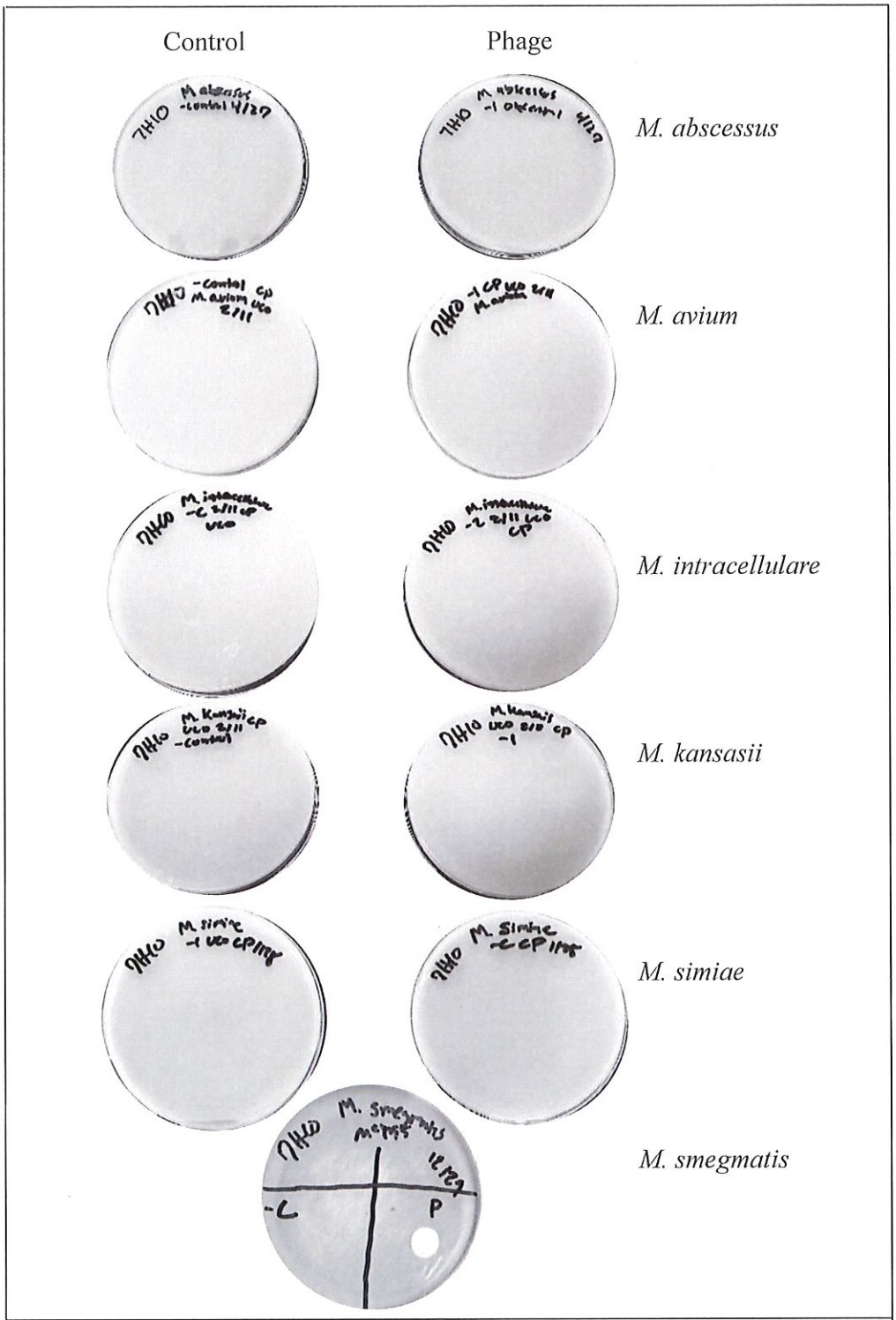

Figure 10. OKCentral2016 was unable to infect five other mycobacterial species. Phage buffer lacking phage was added to the control plates, and served as a negative control. *M. smegmatis* was used as a positive control.

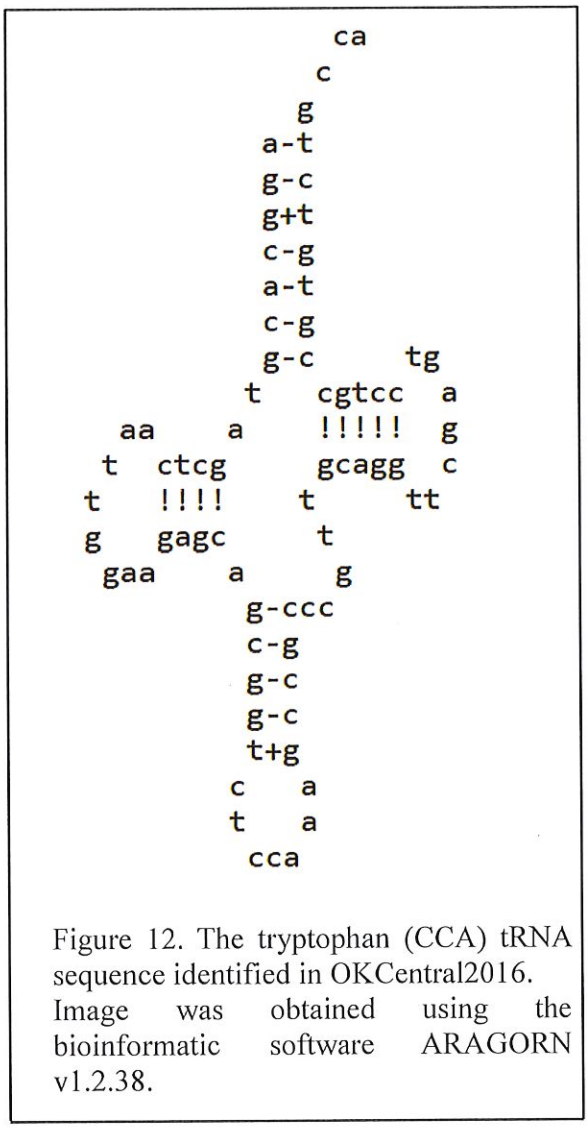

Figure 12. The tryptophan (CCA) tRNA sequence identified in OKCentral2016. Image was obtained using the bioinformatic software ARAGORN v1.2.38.

3.9 Genome analysis of OKCentral2016 – The mycobacteriophage, OKCentral2016, has a double-stranded DNA (dsDNA) genome composed of 50,072 bp containing 83 predicted open reading frames (Figure 11), with 35 known putative protein function (Table A1). Thirty-two of the ORFs are in the forward direction (5'-3') and the remaining 51 ORFs are in the reverse direction (3'-5'). After analysis of all of the ORFs it was determined that 53 ORFs use ATG, 22 use GTG, and 8 use TTG as the start codon. One tRNA sequence was identified and coded for tryptophan ($tRNA^{trp}$). The GC content of this genome was approximately 65.1%. The $tRNA^{trp}$ (Figure 12) sequence was shown to be 76 bases in length and have a GC content of 60.5%. The start site of the $tRNA^{trp}$ was 3483 and the stop site was 3558 in the forward direction. The anticodon sequence in this $tRNA^{trp}$ was found to be CCA. A 10 base 3' cohesive overhang also was observed. The 3' overhang sequence was shown to be 5' CGGCCGGTAA 3.' OKCentral2016 was shown to have strong sequence similarities to subcluster A10, where OKCentral2016 aligned most with Chupacabra and Goose with a minimum identify value of 99%. The average GC content of the subcluster A10 is 64.9% and, on average, members belonging to this subcluster have 85.5 genes and code for one tRNA.

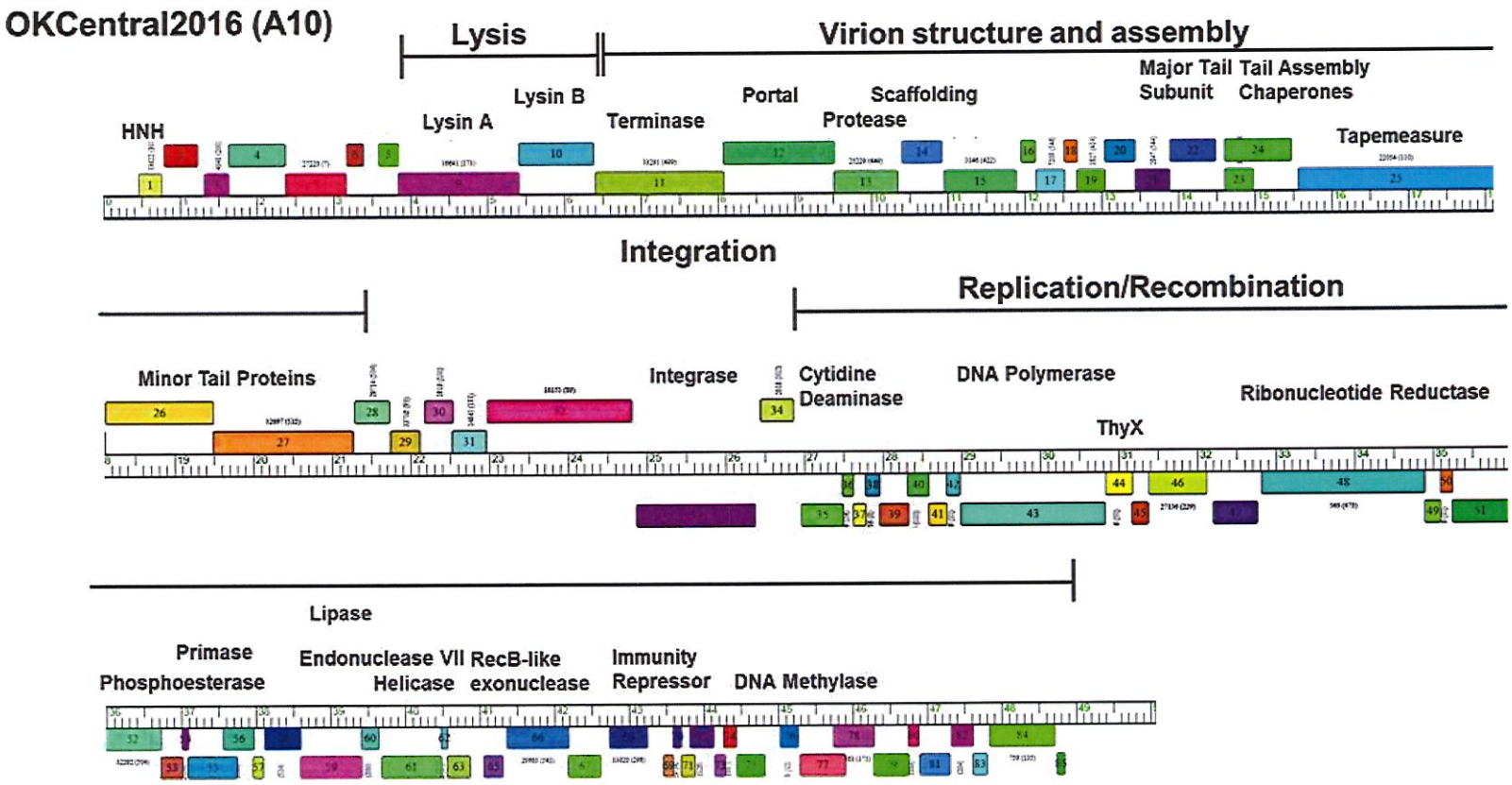

Figure 11. The organizational features of the genome of OKCental2016. The genome is represented as a horizontal bar with markers. The ORFs are depicted as colored boxes. If the ORF is above the horizontal bar the gene is rightward transcribed and if the ORF is below the horizontal bar the gene is leftward transcribed.

Table 6. Mycobacteriophages used for clustering analysis of the A cluster.

Phage	Subcluster
Doom	A1
20ES	A2
Margo	A3
Burger	A4
UnionJack	A5
ToneTone	A6
Sheen	A7
Astro	A8
Catalina	A9
OKCentral2016	A10
Jabith	A11
DarthPhader	A12
Phlei	A13
Luchador	A14
EagleEye	A16
40AC	A17
MyraDee	A18

3.10 Clustering similarities of OKCentral2016 and other A cluster mycobacteriophages – Following genomic sequencing and annotation, three genes were used to evaluate clustering similarities within the A cluster. One individual was selected as a representative from each subcluster within the A cluster (Table 6). A representative from subcluster A15 was not selected due to lack of a member that infected *M. smegmatis* mc^2155. The three genes that were used for comparative analysis of the A cluster were DNA polymerase, lysin A, and the major capsid protein. Trees were condensed showing only bootstrap values that indicate support for that specific branch (≥ 50).

Maximum likelihood analysis of DNA polymerase within the A cluster was determined using the Whelan and Goldman (109) model of evolution. The length of the DNA polymerase sequences ranged from 595 to 692 amino acids. Clustering similarities within the DNA polymerase translated sequences (Figure 13) revealed that the A3 and A4 subclusters were most similar to A10, with bootstrap values of 63 and 67, respectively. Analysis of the lysin A amino acid sequences, which ranged from 226 to 526 amino acids in length, showed A10 was most closely related to A3 (Figure 14). The bootstrap value of the A10 and A3 node was 100, indicating that relationship was strongly supported. The final gene that was evaluated in the A cluster was the major capsid protein. The amino acid sequences ranged from 310 to 401 residues. The dendrogram

evaluating cluster similarities of the major capsid protein (Figure 15) showed A10 was most closely related to A4. All three trees were built using the Whelan and Goldman (109) model of evolution. However, for lysin A analysis the Whelan and Goldman model +Freq. was determined to be the best model for dendrogram construction. A variance in tree topology also was observed between the three proteins. The major capsid protein tree had more unresolved branches when compared to the DNA polymerase tree. Thus, based on maximum likelihood analyses of these three genes, and using only one representative from each subcluster, the subcluster A10 was shown to be more closely related to the A3 and A4 subclusters.

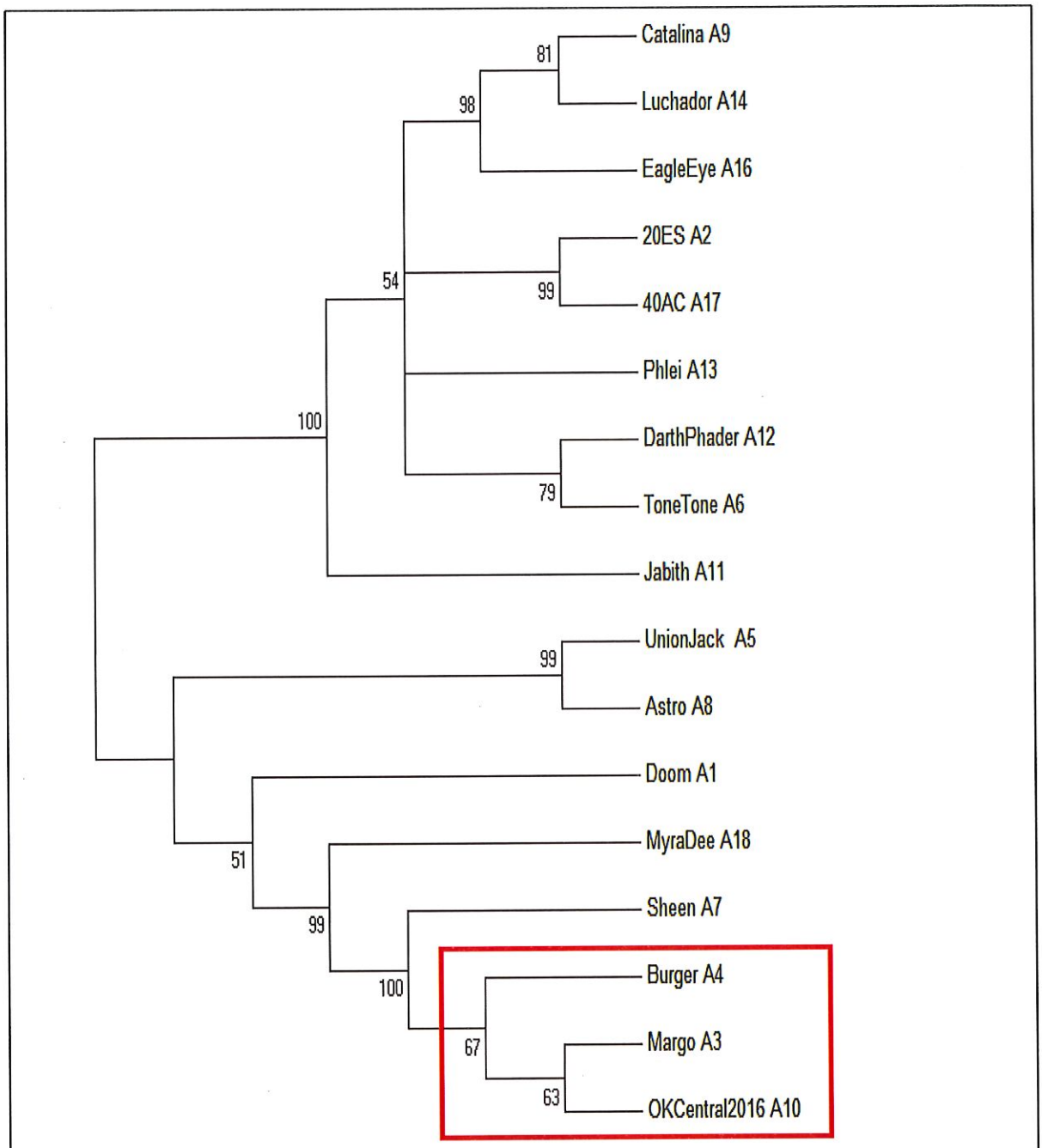

Figure 13. A condensed dendrogram comparing clustering similarities of the DNA polymerase amino acid sequences among the A cluster. Clustering was inferred by using the maximum likelihood method based on the Whelan and Goldman (109) model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 1.3434). The support for this tree was inferred from 1,000 bootstrap replications.

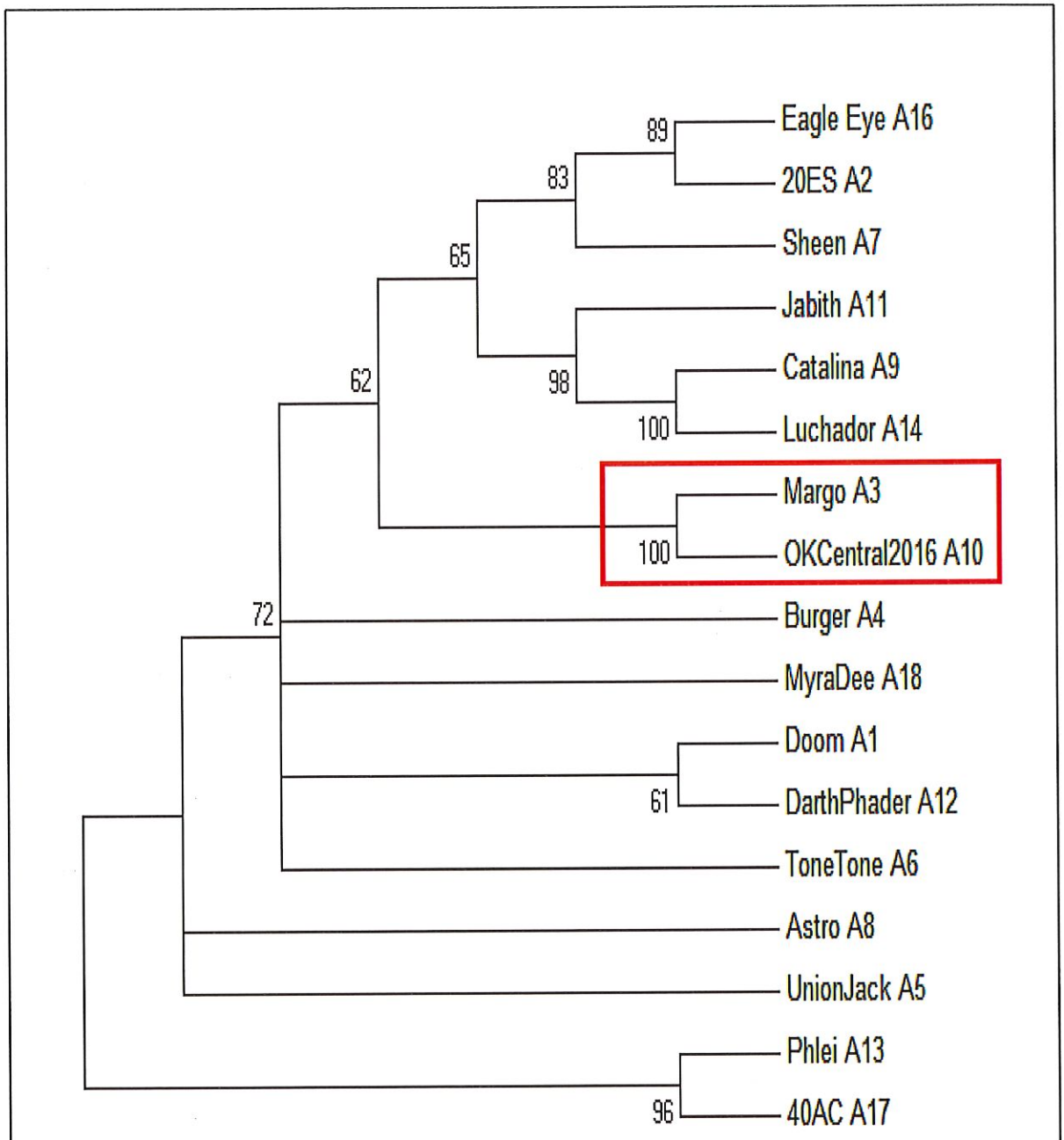

Figure 14. A condensed dendrogram comparing clustering similarities of the lysin A amino acid sequences of the A cluster. Clustering was inferred by using the maximum likelihood method based on the Whelan and Goldman + Freq. model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 1.8754)). Support for this tree was inferred from 1,000 bootstrap replications.

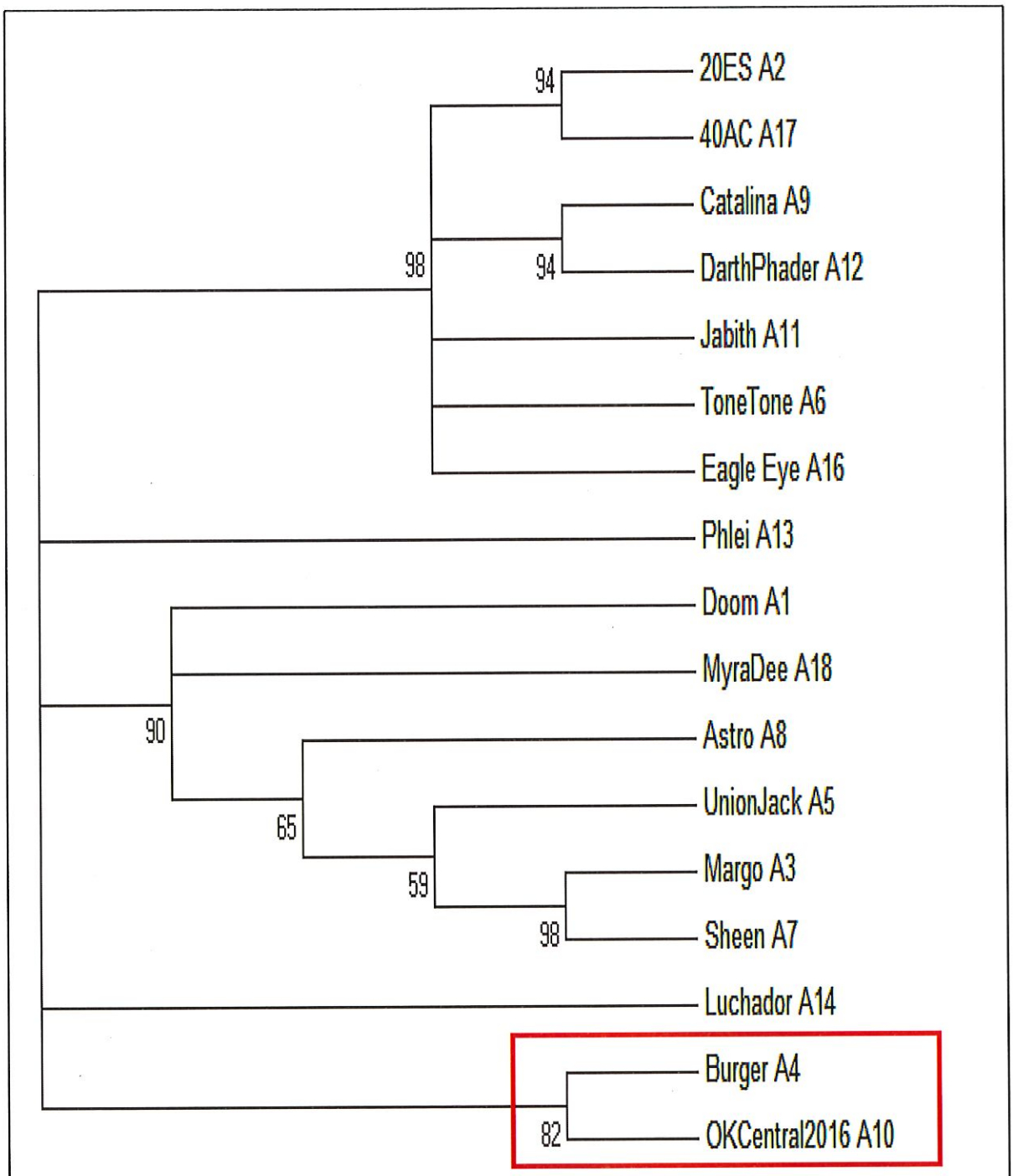

Figure 15. A condensed dendrogram comparing clustering similarities of the major capsid protein amino acid sequences among the A cluster. Clustering was inferred by using maximum likelihood based on the Whelan and Goldman (109) model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 0.7572)). Support for this tree was inferred from 1,000 bootstrap replications.

Table 7. Mycobacteriophages used for comparative analysis of the A3, A4, and A10 subclusters.

Phage	Subcluster
Margo	A3
Microwolf	A3
Phoxy	A3
Rockstar	A3
Taurus	A3
Vix	A3
Bruiser	A4
Burger	A4
Eris	A4
Maverick	A4
Peaches	A4
Goose	A10
OKCentral2016	A10
Rebeuca	A10
Trike	A10
Twister	A10

3.11 Clustering similarities of OKCentral2016 and other A3, A4, and A10 subclusters – The subclusters A3, A4, and A10 were further analyzed for clustering similarities among the DNA polymerase, lysin A, and major capsid protein genes. Depending on the number of applicable representatives within that subcluster, 4 to 5 phages from each subcluster were subjected to further analyses. Mycobacteriophages that were used in these analyses are depicted in Table 7. Clustering analysis of DNA polymerase sequences (Figure 16) showed that A10 clustered more closely with A4 for that specific protein. However, when evaluating the DNA polymerase dendrogram, two representatives from the A3 subcluster (Rockstar and Margo), also showed clustering similarities with A10. The number of amino acids ranged from 604 to 612 for this gene. In contrast, the lysin A (Figure 17) and major capsid protein dendrograms (Figure 18) revealed that A3 and A10 cluster together more than A4 and A10 for specific proteins. The number of amino acids within the lysin A and major capsid protein sequences ranged from 498 to 522 and 310 to 330, respectively. In general, A3 and A4 subclusters had more undetermined relationships in the three proteins evaluated. All dendrograms have been condensed to show only those relationships that are statistically supported.

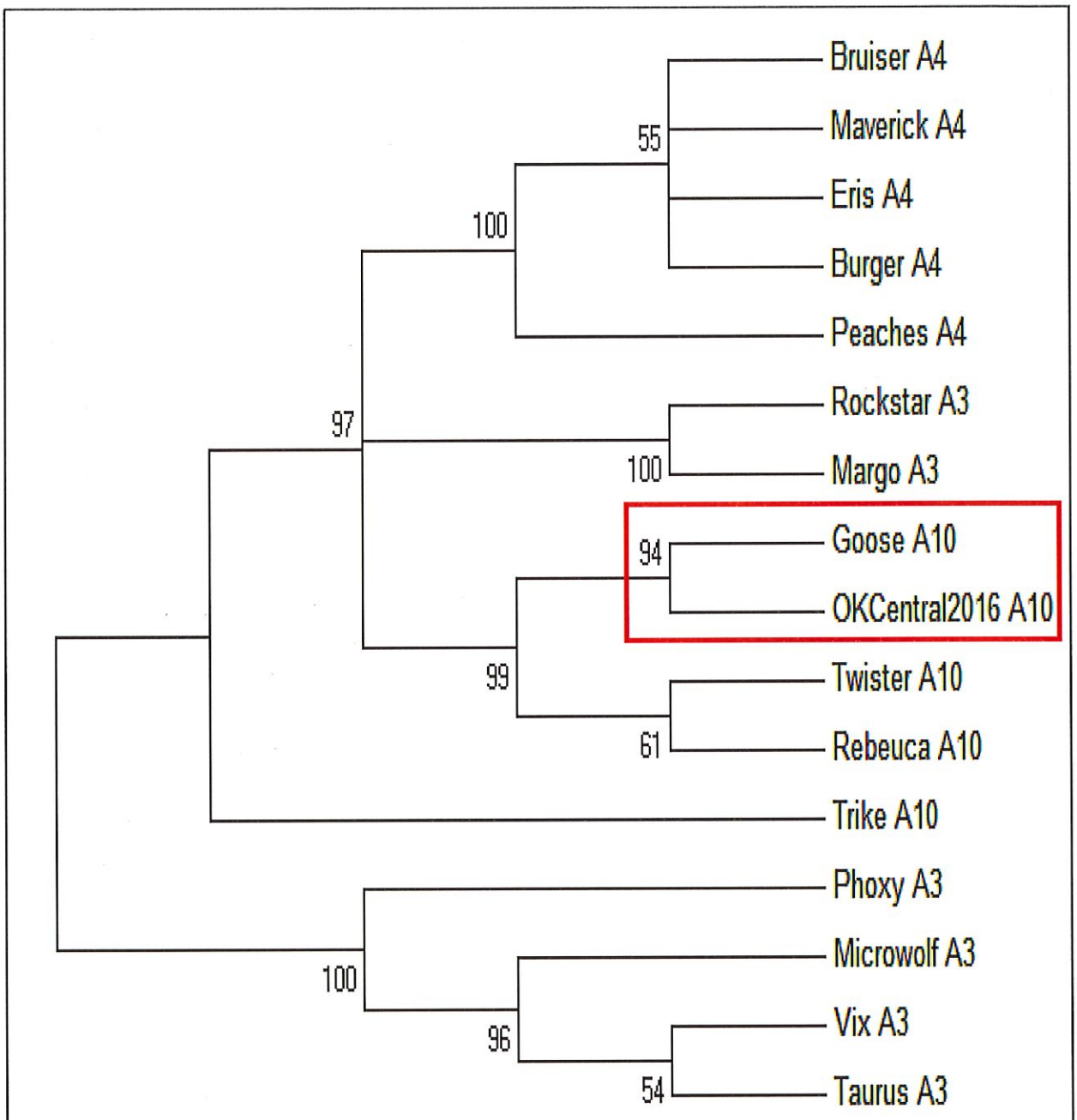

Figure 16. A condensed dendrogram comparing clustering similarities of the DNA polymerase amino acid sequences among the three subclusters A3, A4, and A10. Clustering was inferred using the maximum likelihood method based on the Whelan and Goldman (109) model. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 0.6168)). Support for this tree was inferred from 1,000 bootstrap replications.

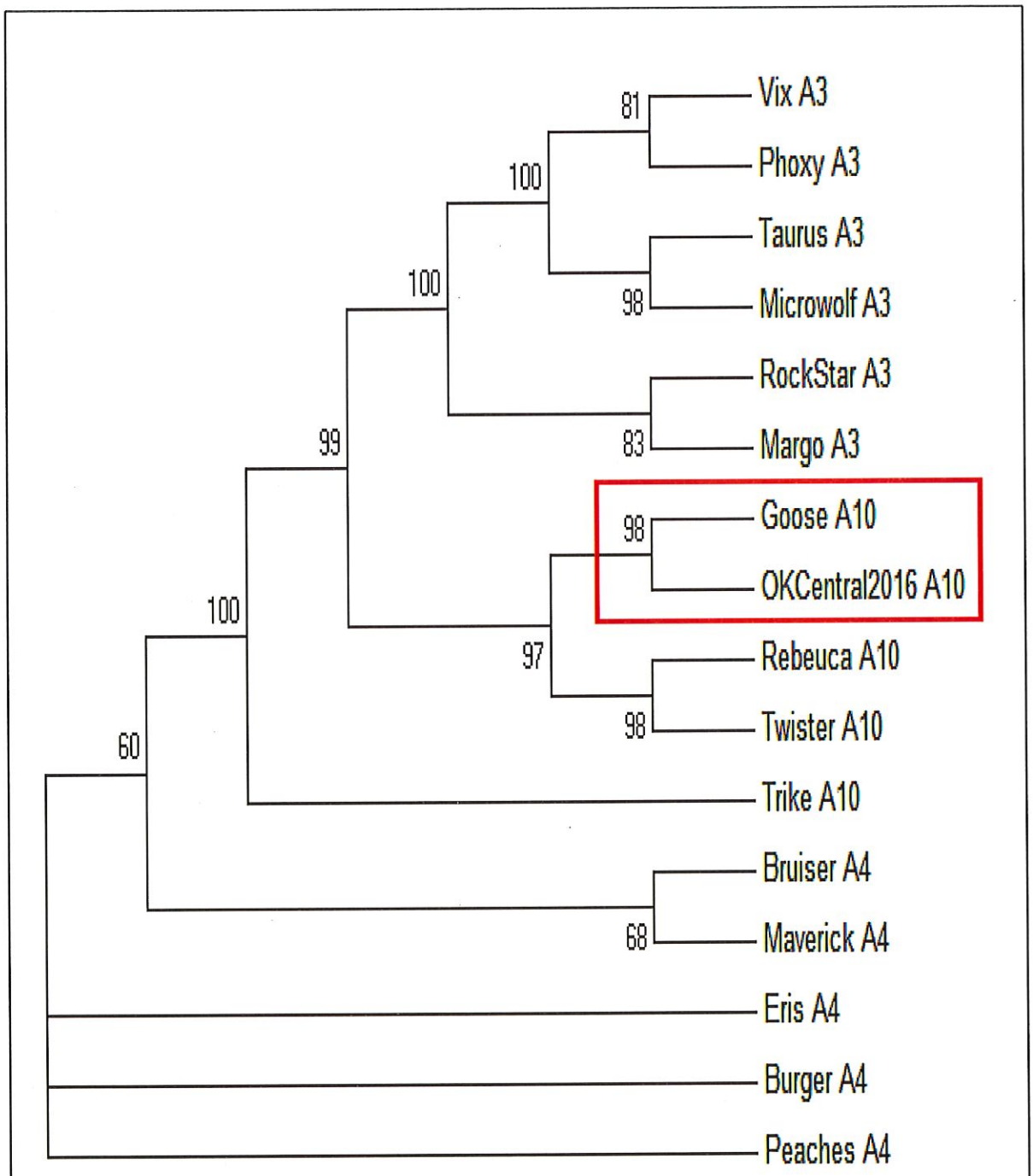

Figure 17. A condensed dendrogram comparing the clustering similarities of the lysin A amino acid sequences among the three subclusters A3, A4, and A10. Clustering was inferred by using the Maximum Likelihood method based on the Whelan and Goldman (109) model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 3.7929)). Support for this tree was inferred from 1,000 bootstrap replications.

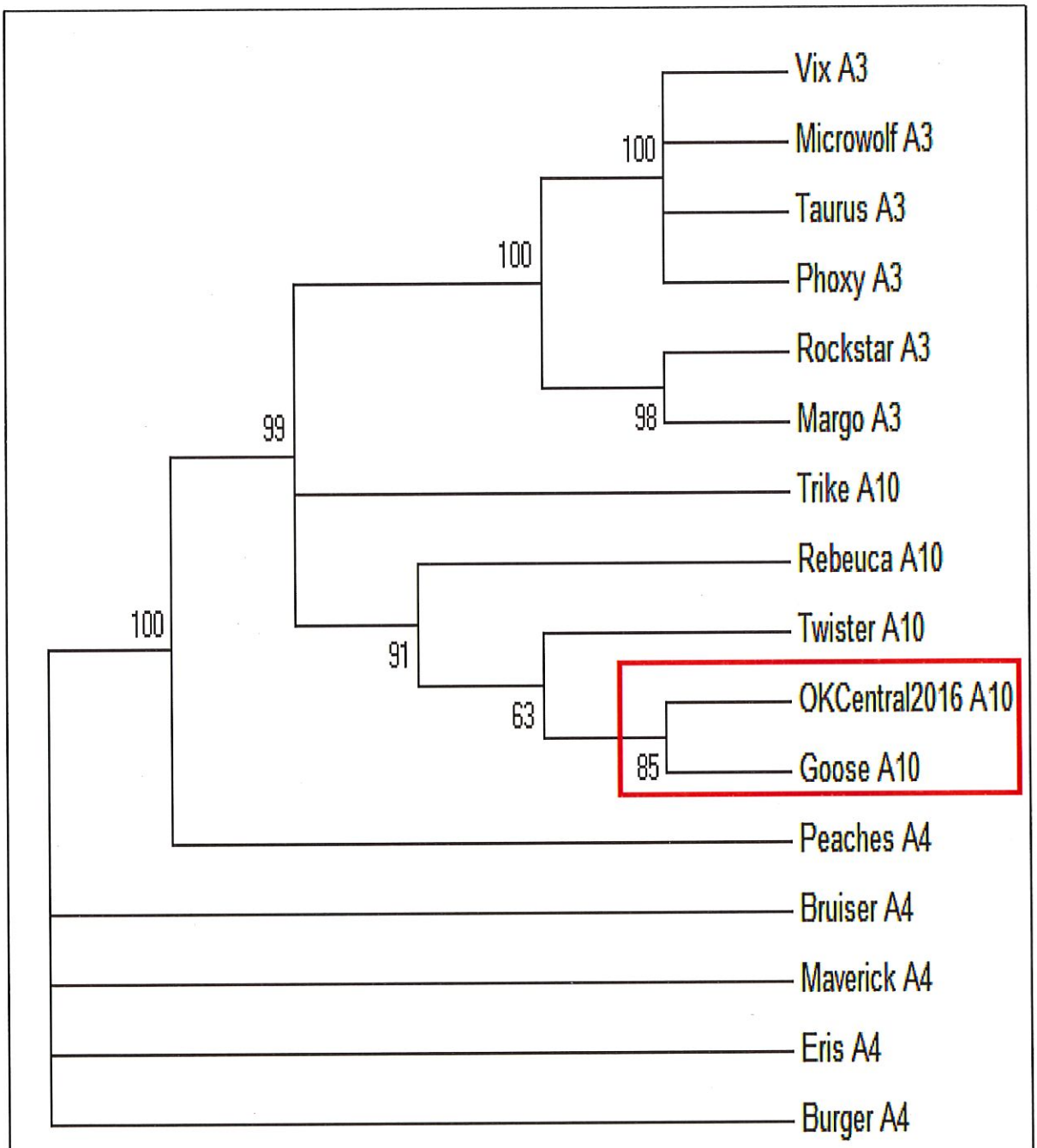

Figure 18. A condensed dendrogram comparing the clustering similarities of the major capsid protein amino acid sequences among the three subclusters A3, A4, and A10. Clustering was inferred by using the Maximum Likelihood method based on the Le Gascuel (110) model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 1.4333)). Support for this tree was inferred from 1,000 bootstrap replications.

Chapter 4: Discussion and Conclusion

4.1 Physical characteristics and stability of OKCentral2016 – The isolated phage, OKCentral2016, had a complex structure and belonged to the viral family *Siphoviridae*, which is the most abundant viral family (138). *Siphoviridae* tail structure is described as conserved and is assembled by stacking major tail proteins around multiple copies of the tapemeasure protein. The distal end of the tail is termed the baseplate. The baseplate initiates viral adsorption and infection (138-140). The presence of divalent cations, such as $CaCl_2$, has been shown to aid in viral adsorption and DNA injection (141, 142).

Kaiser (143) described the turbidity of the plaque corresponding to the phage's ability to lysogenize or integrate into the host genome. Mycobacteriophage OKCentral2016 formed clear transparent plaques. I observed no changes in plaque morphology during our work, indicating viral replication by lysogeny. Mycobacteriophage lysogens have been isolated from turbid plaques Hatfull (144) described all Cluster A phages as being temperate phages, or recent derivatives from temperate phages, which also explains why genes used in lysogenic replication are present in the genome of OKCentral2016. For example, gp33 functions as an integrase, which is responsible for the integration of the prophage into the bacterial genome (145, 146).

Mycobacteriophages have been shown to be thermally inactivated between 54-60°C. (135). The thermal inactivation of protein have been shown to occur by breaking hydrogen bonds, which results in protein folding changes within the secondary structure (147). OKCentral2016 demonstrated a decrease in viable PFUs at both 55 and 60 °C, which agrees with the current literature.

Changes in pH also have been shown to affect viable phage PFUs. The proton concentration can alter protein folding states, which results in different electrostatic interactions between amino acid residues (148). However, hydrogen ion concentration also was shown to

induce phage aggregation of the coliphage, MS2 (149). The viral particle of bacteriophages has been shown to remain stable between the pH range of 6.0-8.0 (135, 150). Sellers and Runnals (135) tested three mycobacteriophages D28, D29, and D32, and showed that survivability for D28 was a pH of 6.0 to 10.0, for D29 it was 7.0 to 10.0, and for D32 it was 7.0 to 9.0. I juxtaposed these results with OKCentral2016's pH stability. When compared with our phage, OKCentral2016 was shown to have a narrower pH stability range, which ranged from 6.0 to 8.0. I suspect that this variation in pH stability can be attributed to the bacteriophage genome. For example, mycobacteriophage D29 is a member of the A2 subcluster, whereas OKCentral2016 belongs to the A10 subcluster. The GC% of D29 and OKCentral2016 were 63.5% and 65.1%, respectively.

The one-step growth curve of OKCentral2016 revealed the time required for the virus to infect a bacterial cell, replicate intracellularly, and release the viral progeny. Following adsorption, the two important time periods in a viral growth curve are the latent period and the rise period. The latent period is the period between initial infection and release of the viral progeny. The rise period is the period between the release of the viral progeny and when the progeny may be detected through plaque formation (151). The rise period of OKCentral2016 started at approximately 1.5 hours and ended at approximately 3 hours.

4.2 Genomic characteristics of OKCentral2016 – Treating OKCentral2016's gDNA with five different endonucleases aided in the preliminary determination of the viral cluster and subcluster of our phage (107, 127, 152), which was cluster A and subcluster A10. This was later confirmed with genomic sequencing. There were two detected programmed ribosomal frameshifts (PRF) found in the genome of OKCentral2016. The first of these was on the tail assembly chaperone protein, and the second can be found on the primase; both PRFs were a -1 frameshift. PRFs occur due to the mRNA having a slippery sequence. A slippery sequence has been identified

to consist of three codons with the nucleotide model sequence of X XXY YYZ. In this model X and Z can be any nucleotide, however Y must be an adenine or uracil (153).

PRFs have been demonstrated as an alternative way viruses can regulate their gene expression (154, 155). PRFs on the tail assembly protein are thought to be conserved among dsDNA bacteriophages (156). The PRFs have been shown to produce a fixed ratio of tail assembly chaperone protein, and the fusion protein of the tail assembly chaperone protein and the adjacent gene downstream. The ratio of the assembly chaperone protein and the fusion protein has been demonstrated to be essential for phage assembly and viability (157, 158).

The mycobacteriophages belonging to the A cluster has been shown to be diverse. The presence of various tRNA sequences bolsters the diversity of the A cluster. The reason why phages contain tRNAs has been viewed as perplexing. However, it has been proposed that tRNA presence in bacteriophage genomes is due to the integration and excision of prophages from the bacterial genome (159, 160). tRNA sequences within the bacterial genome have been shown to be a prophage integration target site (9). However, during excision some bacteriophages will excise incorrectly, resulting in the uptake of host genes during excision (3), which could explain the presence of specific tRNAs within bacteriophage genomes. Mycobacteriophage tRNAs also have been suggested to have been acquired in an ancestral mycobacteriophage genome; where the ancestral genome would have obtained a tRNA cassette (161). Phages also may produce their own tRNA to increase protein synthesis efficiency (161-163).

4.3 Mycobacteriophage gene comparison – Studies conducted on bacteriophage and mycobacteriophage genomics suggest that phages should not be ordered phylogenetically, using a single hierarchal system (164-167). This is due largely to the gain and loss of genes through various mechanisms of genetic exchange, homologous (168) and illegitimate recombination (169), which

exacerbates phage complexity. Another problem with organizing bacteriophages phylogenetically is that phages lack a universal marker as seen in bacteria. This has resulted in the inability to delineate phylogenetic relationships among mycobacteriophages.

The organization of phages into clusters based on genomic similarities has been shown to be the most effective organizational method. Hatfull (167) used various methods to cluster phages such as dotplot analysis, average nucleotide identity, gene content analysis, and pairwise alignment. Phage proteins were grouped into phamilies or phams. These phamilies consisted of similar amino acid sequences that had a minimal sequence similarity E-value of 0.001, or 25% amino acid identity with another protein (170). As per phagesdb.org there are just over 40,500 different phams.

Single gene analysis was used to delineate clustering similarities among subclusters within Cluster A. I used amino acid sequences coding for lysin A, DNA polymerase, and major capsid subunit to carry out these comparisons. Due to the high amount of diversity among viral nucleotide sequences, amino acids sequences are commonly used in single gene analyses.

Some phages, such as mycobacteriophage D29 possess a lytic cassette, which codes for three genes, lysin A, lysin B, and holin. Lysin A is an essential protein that is used to hydrolyze the peptidoglycan present within the bacterial cell wall. The function of lysin B in D29 was shown to act as a mycolylarabinogalactan esterase and cleave the ester bond connecting the outer membrane, which is rich in mycolic acid, to arabinogalactan releasing free mycolic acids (171). The final gene that makes up the lytic cassette is a holin. Holins are small membrane proteins that degrade the cytosolic membrane, allowing the lysin proteins to reach the peptidoglycan within the cell wall, so that bacterial lysis may occur (172). However, not all phage genomes code for a holin,

because cell lysis is achieved in a holin-independent manner (173, 174). OKCentral2016 does not code for a holin protein.

It also has been suggested that the phage DNA polymerase may have originated from the bacterial DNA polymerase (162). For example, some phages code for a DNA polymerase that is similar to DNA polymerase I found in *E. coli*. However, some phage polymerases have been shown to be more closely related to subunits of DNA polymerase III, found in bacteria. (175). Both DNA polymerase I and III are both involved in chromosomal DNA replication (3).

The viral capsid is a durable structure composed of various proteins. Two main proteins make up the capsid, the major capsid subunit and the scaffolding protein. Hundreds of copies of a capsid subunit are produced and arranged to have icosahedral symmetry (176). Dozens of copies of the scaffolding protein are produced to aid in assembly by acting as a chaperone. However, these proteins are not incorporated into the final capsid (177, 178).

Amino acid sequences were used to analyze gene clustering similarities. All dendrograms were constructed without an outgroup. The lytic cassettes among mycobacteriophages are diverse, which is why lysin A was used. Comparing lysin A amino acid sequences, seven different phams were identified (11875, 16641, 17792, 17824, 18207, 22627, and 24417). DNA polymerase is an important enzyme used in DNA replication and is highly conserved (179). Every DNA polymerase amino acid sequence belonged to the same pham (32894), which further demonstrated the conservation of this protein. The final protein that I examined was the major capsid protein. The major capsid protein of every phage used belonged to the same pham (3146), excluding mycobacteriophage Doom (5376), belonging to subcluster A1. All phages used with their proteins and protein phams for the A cluster and A3, A4, and A10 subclusters are depicted in Tables A1 and A2, respectively.

To better understand the clustering similarities of A10 to other subclusters within cluster A, more representatives were used. There was more consistency within the A3, A4, and A10 protein phamilies. For example, the DNA polymerase and major capsid protein sequences used all belonged to the same phams, 32894 and 3146, respectively. In regards to lysin A phams, subclusters A3 and A10 both belonged to the pham 16641, except for Trike (A10). Trike's lysin A belonged to the pham 22627, which was the same pham found in the A4 subcluster.

Initial analysis of the DNA polymerase amino acid sequences showed that A4 and A3 clustered together, as well as A3 and A10 clustered together. Comparison between these three subclusters revealed a split within the A3 subcluster. Phages Rockstar (A3) and Margo (A3) were not grouped with the other A3 subcluster phages. Trike (A10) was shown not to cluster with any other phages' DNA polymerase. The comparison for the A cluster shows that A3 and A10 cluster more tightly together. These results were confirmed when the A3, A4, and A10 were further investigated using more representatives. Relationships among Eris (A4), Burger (A4), and Peaches (A4) were not able to be determined. The final analysis was done on the major capsid protein and showed that A4 and A10 initially clustered together when compared to the A cluster. However, a different relationship was observed when more representatives were utilized for dendrogram construction. A10 was shown to cluster more tightly with A3 than A4. However, phage Peaches (A4) was clustered with the A3 and A10 subclusters and was highly supported. Esposito (180) constructed a dendrogram using the relative synonymous codon usage values. Although OKCentral2016 was isolated after this study, they showed that the A3, A4, and A10 subcluster clustered together (180).

I believe that the large variation among these phage phamilies was due to a less rigorous standard of only 25% similarity in the amino acid sequence. Due to various recombination events

as well as DNA viruses having a mutation rate of approximately 1.1 x 10^{-6} substitutions per nucleotide per strand copying (181), relationships among members of the same pham are unable to be delineated. Our data suggests that there is a need for the integration of more subphamilies to better classify these viral proteins more precisely.

4.4 Host range and phage therapy – Due to the high pathogenicity and transmission of *M. tuberculosis*, mycobacteriophages have been studied as plausible phage therapy agents. It has been suggested that there is a close correlation between the phage cluster and the phages ability to infect other mycobacterial species such as *M. tuberculosis* (175). The A2 and A3 subclusters have been shown to infect mycobacterial hosts other than *M. smegmatis*. For example, phages Bxb1 and U2 (A1 subcluster), L5 and D29 (A2 subcluster), and Bxz2 (A3 subcluster) were able to infect *M. tuberculosis* (47, 182). However, it has been shown that single amino acid substitutions on the putative tail fiber protein allows the mutant to efficiently infect *M. tuberculosis* (175). Phage infectivity of a specific host is postulated to be determined by the specific receptors present on the bacterial cell. However, very few phage receptors have been identified (183).

The GC% has been shown to vary in range in mycobacteria. The lowest GC% content is found in *M. leprae* and the highest GC% content is found in *M. avium*; the GC% of *M. leprae* and *M. avium* are 57.8% and 69.0%, respectively. The GC% of *M. tuberculosis* has been shown to be 65.6% (47). However, the host bacterium used in this study, *M. smegmatis* mc^2155, had a GC% content of 67.4% (45), which is close to the GC% content of OKCentral2016 of 65.1%.

M. smegmatis mc^2 155 is not necessarily the ideal host for isolated mycobacteriophages if phage therapy is the focus (47). If isolation of mycobacteriophages for therapeutic purposes such as phage therapy is the end goal, using specific host trains will prove beneficial.

Bacteriophages have been shown to effectively prevent and degrade various bacterial biofilms, such as those produced by *Campylobacter jejuni*, *E. coli*, and *Proteus mirabilis* (184-186). Phages capable of degrading biofilms possess a glycanase enzyme, which breaks down biofilm polysaccharide. For this to occur, the phage binds to biofilm polysaccharide, which serves as a secondary receptor. The glycanase degrades the biofilm by hydrolyzing β – glycosidic linkages until the phage has reached the primary receptor on bacterial cell exterior (Figure 19). The phage adsorb to the bacterium and will begin replication in the lytic or lysogenic state (187, 188). OKCentral2016 was shown to inhibit the formation of biofilms by planktonic cells, however could not degrade a pre-established biofilm. Therefore, I suspect that this phage does not possess machinery capable of breaking down the biofilm exopolysaccharide. However, it would be fair to speculate the emergence of phage resistant bacteria, or the environmental stress caused by the phage presence induces biofilm proliferation. Bacteriophages possess narrow host ranges. Thus,

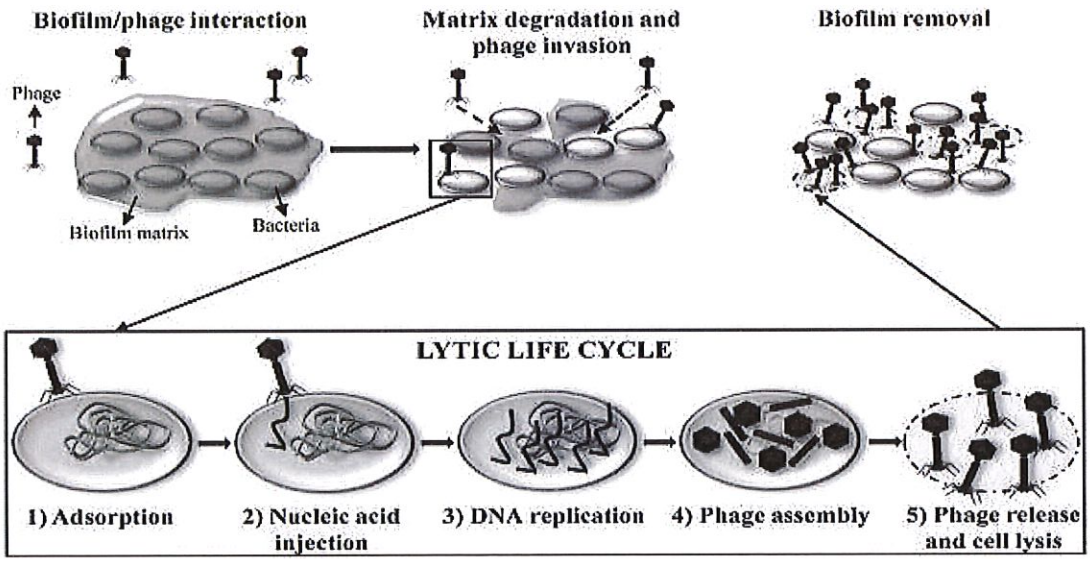

Figure 19. A schematic representation of a bacteriophage degrading the bacterial biofilm and inducing bacterial cell lysis.

Source: Gutiérrez, D., et al. (2016). "Bacteriophage as weapons against bacterial biofilms in the food industry." Frontiers in Microbiology 7.

OKCentral2016 may possess more therapeutic potential demonstrated in this study. Further studies are warranted to better evaluate the application of OKCentral2016 as a potential anti-biofilm agent.

Conclusion - OKCentral2016 was the first mycobacteriophage to be isolated and sequenced from Oklahoma soil. I was able to successfully isolate and characterize the virus. I also was able to examine clustering relationships with other members present in the cluster. This study generated important preliminary data for determining the potential application of this bacteriophage in phage therapy.

Chapter 5: Bibliography

1. **Patton CJ, Kotturi H.** 2018. Genomic sequence of mycobacteriophage OKCentral2016. Genome Announcements **6**.
2. **Brooks G, F., Carroll KC, Butel JS, Morse SA, Mietzner TA.** 2010. *Jawetz, Melnick & Adelberg's Medical Microbiology*. McGraw-Hill, United States of America.
3. **Madigan MT, Martinko JM, Stahl DA, Clark DP.** 2012. Brock Biology of Microorganisms 13th ed. Pearson Education Inc., San Francisco, CA.
4. **Yoshiyuki I, Saori O, Keiko T, Tadahito K.** 2005. Human papillomavirus 16 minor capsid protein L2 help capsomeres assemble indepdently of intercapsomeric disulfide bonding. Virus Genes **31**:321-328.
5. **Reguera J, Carreira A, Riolobos L, Almendral JM, Mateu MG.** 2004. Role of interfacial amino acid residues in assembly, stability, and conformation of a spherical virus capsid. Proceedings of the National Academy of Sciences of the United States of America **101**:2724-2729.
6. **Zlotnick A, Reddy VS, Dasgupta R, Schneemann A, Ray WJJ, Ruckert RR, Johnson JE.** 1994. Capsid assembly in a family of animal viruses primes an autoproteolytic maturation that depends on a single aspartic acid residue. The Journal of Biological Chemistry **269**:13680-13684.
7. **Knipe DM, Howley PM.** 2001. Fundamental Virology, 4th ed. Lippincott Williams & Wilkins, Philadelphia, PA.
8. **Chang CY, Nam K, Young R.** 1995. S gene expression and the timing of lysis by bacteriophage lambda. Journal of Bacteriology **177**:3283-3294.
9. **Reiter W-D, Palm P, Yeats S.** 1989. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. Nucleic Acids Research **17**:1989.
10. **Auvray F, Coddeville M, Ordonez RC, Ritzenthaler P.** 1999. Unusual structure of the attB site of the site-specific recombination system of *Lactobacillus delbrueckii* bacteriophage mv4. Journal of Bacteriology **181**:7385-7389.
11. **Campbell AM.** 1992. Chromosomal insertion sites for phages and plasmids. Journal of Bacteriology **174**:7495-7499.
12. **Dupont L, Bonhoure-Boizet B, Coddeville M, Auvray F, Ritzenthaler P.** 1995. Characterization of genetic elements required for site-specific integration of *Lactobacillus delbrueckii* subsp. *bulgaricus* bacteriophage mv4 and construction of an integration-proficient vector for *Lactobacillus plantarum*. Journal of Bacteriology **177**:586-595.
13. **Saha RP, Lou Z, Meng L, Harshey RM.** 2013. Transposable prophage Mu is organized as a stable chromosomal domain of *E. coli*. PLoS Genetics **9**.
14. **Weinbauer MG, Suttle CA.** 1996. Potential Significance of Lysogeny to Bacteriophage Production and Bacterial Mortality in Coastal Waters of the Gulf of Mexico. Applied and Environmental Microbiology **62**:4374-4380.
15. **Shan J, Korbsrisate S, Withatanung P, Adler NL, Clokie MRJ, Galyov EE.** 2014. Temperature dependent bacteriophages of a tropical bacterial pathogen. Frontiers in Microbiology **5**:599.
16. **Janion C.** 2008. Inducible SOS Response System of DNA Repair and Mutagenesis in Escherichia coli. International Journal of Biological Sciences **4**:338-344.

17. **Maiques E, Ubeda C, Campoy S, Salvador N, Lasa I, Novick RP, Barbe J, Penades JR.** 2006. β-lactam antibiotics induce SOS response and horizontal transfer of virulence factors in *Staphylococcus aureus*. Journal of Bacteriology **188:**2726-2729.
18. **Nanda AM, Heyer A, Krämer C, Grünberger A, Kohlheyer D, Frunzke J.** 2014. Analysis of SOS-Induced Spontaneous Prophage Induction in Corynebacterium glutamicum at the Single-Cell Level. Journal of Bacteriology **196:**180-188.
19. **Lemire S, Figueroa-Bossi N, Bossi L.** 2011. Bacteriophage Crosstalk: Coordination of Prophage Induction by Trans-Acting Antirepressors. PLOS Genetics **7:**e1002149.
20. **Rohwer F, Edwards R.** 2002. The phage proteomic tree: a genome-based taxonomy for phage. Journal of Bacteriology **186:**4529-4535.
21. **Brussow H, Kutter E.** 2005. Bacteriohphages: biology and applications. CRC Press, Boca Raton, FL.
22. **Rohwer F.** 2003. Global phage diversity. Cell **113:**141.
23. **Hewson I, O'Neil JM, Fuhrman JA, Dennison WC.** 2001. Virus-like particle distriubtion and abundance in sediments and overlying waters along eutrophication gradients in two subtropical estuaries. Limnology and Oceanography **47:**1734-1746.
24. **Williamson KE, Wommack KE, Radosevich M.** 2003. Sampling natural viral communites from soil for culture-independent analyses. Applied and Environmental Microbiology **69:**6628-6633.
25. **Guzina J, Djordjevic M.** 2015. Inferring bacteriophage infection strategies from genome sequence: analysis of bacteriophage 7-11 and related phages. BMC Evolutionary Biology **15.**
26. **Atterbury RJ, Dillon E, Swift C, Connerton PL, Frost JA, Dodd CER, Rees CED, Connerton IF.** 2005. Corrleation of *Campylobacter* bacteriophage with reduced prescence of hosts in broiler chicken ceca. Applied and Environmental Microbiology **71:**4885-4887.
27. **Proctor LM, Fuhrman JA.** 1990. Viral mortality of marine bacteria and cyanobacteria. Nature (London) **343:**60-62.
28. **Jiang SC, Paul JH.** 1994. Seasonal and diel abundance of viruses and occurence of lysongey/bacteriocinogeny in the marine environment. Marine Ecology Progress Series **104:**163-172.
29. **Suttle CA.** 2007. Marine viruses - major players in the global ecosystem Nature Reviews Microbiology **5:**801-812.
30. **Srinivasiah S, Bhavsar J, Thapar K, Liles M, Schoenfeld T, Wommack KE.** 2008. Phages across the biosphere: contrasts of viruses in soil and aquatic environments. Research in Microbiology **159:**349-357.
31. **Williamson KE, Radosevich M, Smith DW, Wommack KE.** 2007. Incidence of lysogeny within temperate and exterme soil environments. Environmental Microbiology **9:**2563-2574.
32. **Williamson KE, Radosevich M, Smith DW, Wommack KE.** 2005. Abundance and diversity of viruses in six Deleware soils. Applied and Environmental Microbiology **71:**3119-3125.
33. **Garrity GM, Bell JA, Lilburn TG.** 2004. Bergey's Manual of Systematic Bacteriology Second ed. Springer-Verlag, New York.
34. **Holt JG, Krieg NR, Sneath PHA, Staley JT, Williams ST.** 2000. Bergey's Manual of Determinative Bacteriology, 9th ed. Lippincott Williams and Wilkins, Philadelphia PA.

35. **Mullis SN, Falkinham JOI.** 2013. Adherence and biofilm formation of *Mycobacterium avium, Mycobacterium intracellulare,* and *Mycobacterium abscessus* to houshold plumbing materials Journal of Applied Microbiology **115:**908-914.
36. **Norby B, Fosgate GT, Manning EJB, Collins MT, Roussel AJ.** 2007. Enivronmental mycobactera in soil and water on beef ranches: associated between prescence of cultivable mycobactera and soil and water physiochemical charateristics. Veterinary Microbiology **124:**153-159.
37. **Iivanainen EK, Martikainen PJ, Raisanen ML, Katila M-L.** 1997. Mycobacteria in boreal coniferous forest soils. Microbiology Ecology **23:**325-332.
38. **Neumann M, Schulze-Robbecke R, Hagenau C, Behringer K.** 1997. Comparison of methods for isolation of mycobacteria from water. Applied and Environmental Microbiology **63:**547-552.
39. **Wang Y, Ogawa M, Fukuda K, Miyamoto H, Taniguchi H.** 2006. Isolation and identification from spoils at an illegal dumping site and landfills in Japan. Microbiology and Immunology **50:**513-524.
40. **Kim C-J, Kim N-H, Song K-H, Choe PG, Kim ES, Park SW, Kim H-B, Kim N-J, Kim E-C, Park WB, Oh M-d.** Differentiating rapid- and slow-growing mycobacteria by difference in time to growth detection in liquid media. Diagnostic Microbiology and Infectious Disease **75:**73-76.
41. **Holt JG, Krieg NR, Sneath PHA, Staley JT, Williams ST.** 2000. Bergey's Manual of Determinative Bacteriology vol 9th. Williams & Wilkins, Philadelphia, PA.
42. **Beltan E, Horgen L, Rastogi N.** 2000. Secretion of cytokines by human macrophages upon infection by pathogenic and non-pathogenic mycobacteria. Microbial Pathogenesis **28:**313-318.
43. **Gordon RE, Smith MM.** 1953. Rapidly growing acid fast bacteria. Journal of Bacteriology **66:**41-48.
44. **He Z, Buck JD.** 2010. Cell wall proteome analysis of *Mycobacterium smegmatis* strain MC2 `155. BMC Microbiology **10**.
45. **Mohan A, Padiapu J, Baloni P, Chandra N.** 2015. Complete genome sequences of *Mycobacterium smegmatis* laboratory strain (MC^2 155) and isoniazid-resistant (4XR1/R2) mutant strains. Genome Announcements **3**.
46. **Snapper SB, Melton RE, Mustafa S, Kieser T, Jr. WRJ.** 1990. Isolation and charaterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis* Molecular Microbiology **4:**1911-1919.
47. **Jacobs-Sera D, Marinelli L, Bowman C, Broussard G, Guerrero C, Boyle M, Petrova Z, Dedrick R, Pope W, Science SEAPHAGaE, Modlin RL, Hendrix RW, Hatfull GF.** 2012. On the nature of mycobacteriophage diversity and host preference. Virology **434:**187-201.
48. **Hatfull GF, Cresawn SG, Hendrix RW.** 2008. Comparative genomics of the mycobacteriophages: Insights into bacteriophage evolution. Research in microbiology **159:**332-339.
49. **Pope WH, Bowman CA, Russel DA, Jacobs-Sera D, Asai DJ, Cresawn SG, Jr. WRJ, Hendrix RW, Lawrence JG, Hatfull GF.** 2015. Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. eLife **4**.

50. **Hatfull GF.** 2015. Dark matter of the biosphere: the amazing world of bacteriophage diveristy. Journal of Virology **89:**8107-8110.
51. **Danelishvili L, Young LS, Bermudez LE.** 2006. *In vivo* efficacy of phage therapy for *Mycobacterium avium* infection delievered by a nonvirulent mycobacterium. Microbial Drug Resistance **12:**1-6.
52. **Broxmeyer L, Sosnowska D, Miltner E, Chacon O, Wagner D, McGarvey J, Barletta RG, Bermudez LE.** 2002. Killing of *Mycobacterium avium* and *Mycobacterium tuberculosis* by a nonvirulent mycobacterium: a model for phage therapy of intracellular bacterial pathogens. The Journal of Infectious Diseases **186:**1155-1160.
53. **Bardarov S, Kriakov J, Carriere C, Yu S, Vaamonde C, McAdam RA, Bloom BR, Hatfull GF, Jacobs WRJ.** 1997. Conditionally replication mycobacteriophages: a system for transposon delivery to *Mycobacterium tuberculosis*. PNAS **94:**10961-10966.
54. **Koehler CW.** 2002. Consumption, the great killer. Modern Drug Discovery **5**.
55. **Organization WH.** 2017. Global Tuberculosis Report.
56. **Johnson MM, Odell JA.** 2014. Nontuberculosis Mycobacterial pulmonary infections. Journal of Thoracic Disease **6:**210-220.
57. **Kawamura Y, Li Y, Liu H, Huang X, Li Z, Ezaki T.** 2001. Bacterial population in Russian Space Station "Mir". Microbiology and Immunology **45:**819-828.
58. **van der Wende E, Characklis WG, Smith DB.** 1989. Biofilms and bacterial drinking water quality. Water Research **23:**1313-1322.
59. **Domka J, Lee J, Bansal T, Wood TK.** 2007. Temporal gene-expression in Escherichia coli K-12 biofilms. Environmental Microbiology **9:**332-346.
60. **Deep A, Chaudhary U, Gupta V.** 2011. Quorum sensing and Bacterial Pathogenicity: From Molecules to Disease. Journal of Laboratory Physicians **3:**4-11.
61. **Kiefer B, Dahl JL.** 2015. Distruption of *Mycobacterium smegmatis* biofilms using bacteriophages alone or in combination with mechanical stress. Advances in Microbiology **5:**699-710.
62. **Flemming H-C.** 2002. Biofouling in water systems - cases, causes and countermeasures Applied Microbiology and Biotechnology **59:**629-640.
63. **Anonymous.** 2002. Research on Mmcrobial biofilms. National Institute of Health.
64. **Ojha AK, Baughn AD, Sambandan D, Hsu T, Trivelli X, Guerardel Y, Alahari A, Kremer L, Jacobs WRJ, Hatfull GF.** 2008. Growth of *Mycobacterium tuberculosis* biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. Wiley Molecular Microbiology **69:**164-174.
65. **Ojha A, Anand M, Bhatt A, Kremer L, Jacobs WR, Hatfull GF.** 2005. GroEL1: A Dedicated Chaperone Involved in Mycolic Acid Biosynthesis during Biofilm Formation in Mycobacteria. Cell **123:**861-873.
66. **Stoodley-Hall L, Stoodley P.** 2009. Evolving concepts in biofilm interactions. Cellular Microbiology **11:**1034-1043.
67. **Donlan RM.** 2002. Biofilms: microbial life on surfaces. Emerging Infectious Diseases **8:**881-890.
68. **Carter G, Wu M, Drummond DC, Bermudez LE.** 2003. Characterization of biofilm formation by clinical isolates of *Mycobacterium avium*. Journal of Medical Microbiology **52:**747-752.

69. **Steed KA, Falkinham JOI.** 2006. Effect of growth in biofilms on chlorine susceptibility of *Mycobacterium avium* and *Mycobacterium intracellulare*. Applied and Environmental Microbiology **72:**4007-4011.
70. **Sommerstein RR, Kohler P, Bloemberg G, Kuster S, H. S.** 2016. Transmission of *Mycobacterium chimaera* from heater-cooler units durign cardiac surgery despite an Ultraclean Air Ventilation System. Emerging Infectious Diseases **22:**1008-1013.
71. **Sousa S, Bandeira M, Carvalho PA, Duarte A, Jordao L.** 2015. Nontuberculous mycobacteria pathogenesis and biofilm assembly. International Journal of Mycobacteriology **4:**36-43.
72. **Ghosh R, Das S, Kela H, De A, Haldar J, Maiti PK.** 2017. Biofilm colonization of *Mycobacterium abscessus*: new theat in hospital-acquired surgical site of infection. Indian Journal of Tuberculosis **64:**178-182.
73. **Phillips MS, Reyn CFv.** 2001. Nosocomial infections due to notuberculosis Mycobacteria. Clinical Infectious Disease **33:**1363-1374.
74. **Bjarnsholt T.** 2013. The role of bacterial biofilms in chronic infections. APMIS **121:**1-58.
75. **Qvist T, Eickhardt S, Kragh KN, Andersen CB, Iversen M, Høiby N, Bjarnsholt T.** 2015. Chronic pulmonary disease with *Mycobacterium abscessus* complex is a biofilm infection. European Respiratory Journal **46:**1823.
76. **Flemming A.** 1929. On the antibacterial actin of culture of a *Penicillium* with special reference to their use in the isolation of *B. influnezae*. The British Journal of Experimental Pathology **10:**226-236.
77. **Lin DM, Koskella B, Lin HC.** 2017. Phage therapy: An alternative to antibiotics in the age of multi-drug resistance. World Journal of Gastrointestinal Pharmacology and Therapeutics **8:**162-173.
78. **Hankin HE.** 1896. L'action bactericide des eaux de la Jumna et du Gange sur le vibrion du cholera. Ann Inst Pasteur **10**.
79. **Broeck DV, Horvath C, Wolf MJSD.** 2007. *Vibrio cholerae*: cholera toxin. The International Journal of Biochemistry and Cell Biology **30:**1771-1775.
80. **Ackermann H-W, DuBow M.** 1987. Practical applications of bacteriophges. CRC Press, Florida.
81. **d'Herelle F.** 1931. Bacteriophage as a treatment in acute medical and surgical infections. Bulletin of the New York Academy of Medicine **7:**329-348.
82. **Stent GS.** 1963. *Molecular Biology of Bacterial Viruses*, San Francisco, CA.
83. **Kucharewicz-Krukowska A, Slopek S.** 1987. Immunogenic effect of bacteriophage in patients subjected to phage therapy. Archivum Immunologiae Et Therapiae Experimentalis **35:**553-561.
84. **Stratton CW.** 2003. Dead Bugs Don't Mutate: Susceptibility Issues in the Emergence of Bacterial Resistance. Emerging Infectious Diseases **9:**10-16.
85. **Haq IU, Chaudhry WN, Akhtar MN, Andleeb S, Qadri I.** 2012. Bacteriophages and their implication on future biotechnology: a review. Virology Journal **9**.
86. **Stephen TA, Cameron T-A.** 2010. Phage Therapy Pharmacology. Current Pharmaceutical Biotechnology **11:**28-47.
87. **Harper DR, Parracho HMRT, Walker J, Sharp R, Hughes G, Werthén M, Lehman S, Morales S.** 2014. Bacteriophages and Biofilms. Antibiotics **3:**270-284.

88. **Langdon A, Crook N, Dantas G.** 2016. The effects of antibiotics on the microbiome throughout development and alternative approaches for therapeutic modulation. Genome Medicine **8:**39.
89. **Doss J, Culbertson K, Hahn D, Camacho J, Barekzi N.** 2017. A Review of Phage Therapy against Bacterial Pathogens of Aquatic and Terrestrial Organisms. Viruses **9:**50.
90. **Loc-Carrillo C, Abedon ST.** 2011. Pros and cons of phage therapy. Bacteriophage **1:**111-114.
91. **Studier FW, Moffatt BA.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. Journal of Molecular Biology **189:**113-130.
92. **Tabor S, Richardson CC.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proceedings of the National Academy of Sciences **82:**1074-1078.
93. **Tao P, Mahalingam M, Kirtley ML, Lier CJv, Sha J, Yeager LA, Chopra AK, Rao VB.** 2013. Mutated and bacteriophage T4 nanoparticle arrayed F1-V immunogens from *Yersinia pestis* as next generation plague vaccines. PLOS Pathogens **9**.
94. **McNerney R, Kambashi BS, Kinkese J, Tembwe R, Godfrey-Faussett P.** 2004. Development of a bacteriophage phage replication assay for diagnosis of pulmonary tuberculosis. Journal of Clinical Microbiology **42:**2115-2120.
95. **Schofield DA, Molineux IJ, Westwater C.** 2011. 'Bioluminescent' reporter phage for the detection of category A bacterial pathogens Journal of Visualized Experiments **53**.
96. **Loessner MJ, Rees CED, Stewart GSAB, Scherer S.** 1996. Construction of luciferase reporter bacteriophage A511::*luxAB* for rapid and sensitive dection of viable *Listeria* cells. Applied and Environmental Microbiology **62:**1133-1140.
97. **Brigati J, Williams DD, Sorokulova IB, Nanduri V, Chen I-H, Turnbough CLJ, Petrenko VA.** 2004. Diagnostic probes for *Bacillus anthracis* spores selected from a landscape phage library. Clinical Chemistry **50:**1899-1906.
98. **Lu TKT, Ando H, Lemire S.** 2015. Tuning bacteriophage host range. Google Patents.
99. **Lin T-Y, Lo Y-H, Tseng P-W, Chang S-F, Lin Y-T, Chen T-S.** 2012. A T3 and T7 recombinant phage acquires efficient adsorption and a braoder host range. PLOS ONE **7**.
100. **Yosef I, Manor M, Kiro R, Qimron U.** 2015. Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria. Proceedings of the National Academy of Sciences of the United States of America **112:**7267-7272.
101. **Laslett D, Canback B.** 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Research **32:**11-16.
102. **Gordon D.** 2004. Viewing and editing assembled sequences using consed. Current Protocols in Bioinformatics **2:**1-11.
103. **Besemer J, Lomsadze A, Borodovsky M.** 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Research **29:**2607-2618.
104. **Borodovsky M, McIninch J.** 1993. GeneMark: parallel gene recognition for both DNA strands. Computers and Chemistry **17:**123-133.
105. **Tamura K, Stecher G, Peterson D, Filipski A, Kumar S.** 2013. MEGA6: Molecular evoultionary genetic analysis version 6.0. Molecular Biology and Evolution **30:**2725-2729.

106. **Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF.** 2011. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. BMC Bioinformatics **12**.
107. **Gissendanner CR, Wiedemeier AMD, Wiedemeier PD, Minton RL, Bhuiyan S, Harmson JS, Findley AM.** 2014. A web-based restriction endonuclease tool for mycobacteriophage cluster prediction. Journal of Basic Microbiology **54:**1140-1145.
108. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research **25:**955-964.
109. **Whelan S, Goldman N.** 2001. A general empircal model of protein evolution dervied from multiple protein familues using a maximum-likelihood approach. Molecular Biology and Evolution **18:**691-699.
110. **Le SQ, Gascuel O.** 2008. An improved general amino acid replacement matrix. Molecular Biology and Evolution **25:**1307-1320.
111. **Gratia A.** 1936. Des relations numeriques entre bacteries lysogenes et particules de bacteriophage. Annales de l'Institut Pasteur **57:**652-676.
112. **Delbruck M.** 1940. The growth of the bacteriophage and lysis of the host. The Journal of General Physiology **23:**643-660.
113. **Jensen KC, Hair BB, Wienclaw TM, Murdock MH, Hatch JB, Trent AT, White TD, Haskell KJ, Berges BK.** 2015. Isolation and host range of bacteriophage with lytic activity against methicillin-resistant *Staphylococcus aureus* and potential use as a fomite decontaminant. PLoS One **10**.
114. **Tardieu A, Bonnete F, Finet S, Vivares D.** 2002. Understanding salt or PEG inducted attractive interactions to crystalize biological macromolecules. Acta Crystallographica **58:**1549-1553.
115. **Colombet J, Sime-Ngando T.** 2012. Use of PEG, Polyethylene glycol, to characterize the diversity of environmental viruses. Current Microscopy Contributions to Advances in Science and Technology.
116. **Yamamoto KR, Alberts BM.** 1970. Rapid bacteriophage sedimentation in the prescence of polyethylene glycol and its application to large-scale virus purification. Virology **40:**734-744.
117. **Powell LJ, Brennan R, Kotturi H, Overbo CI, Ahmad R, Jones R.** 2017. Microbiology.
118. **Thiberge S, Nechushtan A, Sprinzak D, Gileadi O, Behar V, Zik O, Chowers Y, Michaeli S, Schlessinger J, Moses E.** 2003. Scanning electron microscopy of cells and tissues under fully hydrated conditions. PNAS **101:**3346-3351.
119. **Ishikawa I, Okunishi E, Sawada H, Okura Y, Yamazaki K, Ishikawa T, Kawazu M, Hori M, Terao M, Kanno M, Tanba S, Kondo Y.** 2009. Development of a 200kV atomic resolution analytical electron microscope. Microscopy and Microanalysis **15:**188-189.
120. **Schnieder CA, Rasband WS, Eliceiri KW.** 2012. NIH Image to ImageJ: 25 years of image analysis. Nature Methods **9:**671-675.
121. **Catalao MJ, Gil F, Moniz-Pereira J, Pimentel M.** 2010. The mycobacteriophage Ms6 encodes a chaperone-like protein involved in the endolysin delievery to the peptidoglycan. Molecular Microbiology **77:**672-686.
122. **Green MR, Sambrook J.** 2012. Molecular Cloning: A Laboratory Manual, 4th ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

123. **Greco M, Sáez CA, Brown MT, Bitonti MB.** 2014. A Simple and Effective Method for High Quality Co-Extraction of Genomic DNA and Total RNA from Low Biomass Ectocarpus siliculosus, the Model Brown Alga. PLoS ONE **9:**e96470.
124. **Thomas NC.** 1991. The Early History of Spectroscopy. Journal of Chemical Education **68:**631-634.
125. **Desjardins P, Conklin D.** 2010. NanoDrop microvolume quantitation of nucleic acids. Journal of Visualized Experiments **45.**
126. **Sharp PA, Sugden B, Sambrook J.** 1973. Detection of two restriction endonuclease activities in *Haemophilus parainfluenzae* using analytical agarose-ethidium bromide electrophoresis. Biochemistry **12:**3055-3063.
127. **Norberg P, Bergstrom T, Lijeqvist J-A.** 2006. Genotyping of clinical herpes simplex virus type 1 isolates by use of restriction enzymes. Journal of Clinical Microbiology **44:**4511-4514.
128. **Nguyen KT, Piastro K, Gray TA, Derbyshire KM.** 2010. Mycobacterial biofilm faciliatate horizontal DNA transfer between strains of *Mycobacterium smegmatis*. Journal of Bacteriology **192:**5134-5142.
129. **McFarland J.** 1907. The Nephelometer: an instrument for estimating the number of bacteria in suspensions used for calculating the opsonic index and for vaccines. The Journal of the American Medical Association **XLIX:**1176-1178.
130. **Mirzaei MK, Nilsson AS.** 2015. Isolation of phages for phage therapy: a comparison of spot tests and efficiency of plating analyses for determination of host range and efficacy. PLoS One **10.**
131. **Kinyoun JJ.** 1914. A note on Uhlenhuths method for sputum examionation for tubercle bacilli. The American Journal of Public Health **5:**867-870.
132. **Besemer J, Borodovsky M.** 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes, and viruses. Nucleic Acids Research **33.**
133. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32:**1792-1797.
134. **King AMQ, Lefkowitz E, Adams MJ, Carstens EB.** 2011. Virus Taxonomy: ninth report of the international committee on taxonomy of viruses.
135. **Sellers MI, Runnals HR.** 1960. Mycobacteriophage I. Phsyiochemical characterization. Journal of Bacteriology **39:**442-447.
136. **Fan X, Yan J, Xie L, Zeng L, Young RF, Xie J.** 2015. Genomic and proteomic features of mycobacteriophage SWU1 isolated from China soil. Gene **561:**45-53.
137. **Kraiss JP, Gelbart SM, Juhasz SE.** 1973. A comparison of three mycobacteriophages. Jounral of General Virology **20:**75-87.
138. **Sciara G, Bebeacua C, Bron P, Tremblay D, Ortiz-Lombardia M, Lichiere J, Heel Mv, Campanacci V, Moineau S, Cambillau C.** 2010. Structure of lactococcal phage p2 baseplate and its mechanism of activation. PNAS **107:**6852-6857.
139. **Yap ML, Klose T, Arisaka F, Speir JA, Veesler D, Fokine A, Rossmann MG.** 2016. Role of bacteriophage T4 baseplate in regulating assembly and infection. PNAS **113:**2654-2659.
140. **Bebeacua C, Lai L, Vegge CS, Brondsted L, Heel Mv, Veesler D, Cambillau C.** 2012. Visualizing a complete *Siphoviridae* member by single-particle electron microscopy: the strcutre of lactococcal phage TP901-1. Journal of Virology **87:**1061-1068.

141. **Karnik SS, Gopinathan KP.** 1980. Possible Involvement of Calcium-Stimulated ATP-Hydrolyzing Activity Associated with Mycobacteriophages I3 in the DNA Injection Process. Journal of Virology **33:**969-975.
142. **Potter NN, Nelson FE.** 1953. Role of calcium and related ions in proliferation of lactic streptococcus bacteriophage. Journal of Bacteriology **66:**508-516.
143. **Kaiser AD.** 1957. Mutations in a temperate bacteriophage affecting its ability to lysogenize *Escherichia coli*. Journal of Virology **3:**42-61.
144. **Hatfull GF.** 2012. The secret life of mycobacteriophages, p 179-234, Advances in Virus Research, vol 82. Elsevier, Burlington.
145. **Kim AI, Ghosh P, Aaron MA, Bibb LA, Jain S, Hatfull GF.** 2003. Mycobacteriophage Bxb1 integrates into the *Mycobacterium smegmatis groEL1* gene. Molecular Microbiology **50:**463-473.
146. **Campbell AM.** 2001. Bacteriophages, p 503-526, Fundamental Virology, Fourth Edition ed. Lippincott Williams & Wilkins, Philadelphia, PA.
147. **Woese C.** 1960. Thermal inactivation of animal viruses. Annals of the New York Academy of Sciences **83:**741-751.
148. **Anderson DE, Becktel WJ, Dahlquist FW.** 1990. pH-induced denaturation of proteins: a single saltbridge contriubtes 3-5 kcal/mol to the free energy of folding of T4 lysozyme. Biochemsitry **29:**2403-2408.
149. **Langlet J, Gaboriaud F, Gantzer C.** 2007. Effects of pH on plaque forming unit counts and aggregation of MS2 bacteriophage. Journal of Applied Microbiology **103:**1632-1638.
150. **Feng YY, Ong SL, Hu JY, Tan XL, Ng WJ.** 2003. Effects of pH and temperature on the survival of coliphage MS2 and Qβ. Journal of Industrial Microbiology and Biotechnology **30:**549-552.
151. **Norkin LC.** 2010. *Virology: Molecular Biology and Parthenogenesis* American Society for Microbiology, Washington D.C.
152. **Clokie MRJ, Millard AD, Letarov AV, Heaphy S.** 2011. Phages in nature. Bacteriophage **1:**31-45.
153. **Moon S, Byun Y, Han K.** 2004. Computational identification of -1 frameshift signals. Springer, Berlin, Heidelberg.
154. **Melian EB, Mendelin SH-, Du F, Owens N, Lauth AM-B, Nagasaki T, Rudd S, Brault AC, Bowen RA, Hall RA, Hurk AFvd, Khromykh AA.** 2014. Programmed robosomal frameshift alters expression of west nile virus genes and facilitates virus replicatio nin birds and mosquitoes. PLOS Pathogens **10**.
155. **Choi J, Xu Z, Ou J-h.** 2003. Triple decoding of hepatitis C virus RNA by programmed translational frameshifting. Molecular and Cellular Biology **23:**1489-1497.
156. **Xu J, Hendrix RW, Duda RL.** 2004. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. Molecular Cell **16:**11-21.
157. **Maxwell KL, Davidson AR.** 2014. A shifty caperone for phage tail assembly. Journal of Molecular Biology **426:**1001-1003.
158. **Levin ME, Hendrix RW, Casjens SR.** 1993. A programmed translational frameshift is required for synthesis of a bacteriophage λ tail assembly protein. Journal of Molecular Biology **234:**124-139.
159. **Bechet-Bailly M, Vergassola M, Rocha E.** 2007. Causes for the intriguing prescence of tRNAs in phages. Genome Research **17:**1486-1495.

160. **Canchaya C, Fournous G, Brussow H.** 2004. The impact of prophages on bacterial chromosomes. Molecular Microbiology **53:**9-18.
161. **Kunisawa T.** 2000. Functional role of mycobacteriophage transfer RNAs. Journal of Theoretical Biology **205:**167-170.
162. **Kunisawa T, Kanaya S, Kutter E.** 1998. Comparison of synonymous codon distribution patterns of bacteriophage and host genomes. DNA Research **5:**319-326.
163. **Hassan S, Mahalingam V, Kumar V.** 2009. Synonymous codon usage analysis of thirty two mycobacteriophage genomes. Advances in Bioinformatics
164. **Lawrence JG, Hatfull GF, Hendrix RW.** 2002. Improglios of viral taxonomy: genetic exchange and failing of phenetic approaches. Journal of Bacteriology **184:**4891-4905.
165. **Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, Lewis JA, Jacobs-Sera D, Falbo J, Gross J, Peannunzio NR, Brucker W, Kumar V, Kandasamy J, Keenan L, Bardarov S, Kriakov J, Lawrence JG, Jacobs WRJ, Hatfull GF.** 2003. Origins of highly mosaic mycobacteriohpage genomes. Cell **113:**171-182.
166. **Carbone A.** 2008. Codon bias is a major factor explaining phage evolutin in translationally biased hosts. Journal of Molecular Evolution **66:**210-223.
167. **Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russel DA, Ko C-C, Weber RJ, Patel MC, Germane KL, Edgar RH, Hoyte NH, Bowman CA, Tantoco AT, Paladin EC, Myers MS, Smith AL, Grace MS, Pham TT, O'Brien MB, Vogelsberger AM, Hryckowian AJ, Wynalek JL, Keller HD-, Bogel MW, Peebles CL, Cresawn SG, Hendrix RW.** 2010. Comparative genomic analysis of sixty mycobacteriophage genomes: genome clustering, gene acquision and gene size. Journal of Molecular Biology **397:**119-143.
168. **Nilsson AS, Ljungquist EH.** 2001. Detection of homologous recombination among bacteriophage P2 relatives. Molecular Phylogenetics and Evolution **21:**259-269.
169. **Hendrix RW, Smith MCM, Burns RN, Ford ME, Hatfull GF.** 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. PNAS **96:**2192-2197.
170. **Hatfull GF, Pedulla ML, Jacobs-Sera D, Cichon PM, Foley A, Ford ME, Gonda RM, Houtz JM, Hryckowian AJ, Kelchner VA, Namburi S, Pajcini KV, Popovich MG, Schleicher DT, Simanek BZ, Smith AL, Zdanowicz GM, Kumar V, Peebles CL, William JRJ, Lawrence JG, Hendrix RW.** 2006. Exploring the mycobacteriophage metaproteome: phage genomics as an eductational platform. PLoS Genetics **2.**
171. **Payne K, Sun Q, Sacchettini J, Hatfull GF.** 2009. Mycobacteriophage Lysin B is a novel mycolylarabinogalactan esterase. Molecular Microbiology **73:**367-381.
172. **Blasi U, Young R.** 1996. Two beginnings for a single purpose: the dual-start holins in the regulation of phage lysis. Molecular Microbiology **21:**675-682.
173. **Xu M, Struck DK, Deaton J, Wang I-N, Young R.** 2004. A single-arrest-release sequence mediates export and control of the phage P1 endolysin. PNAS **101:**6415-6420.
174. **Frias MJ, Cristino JM-, Ramirez M.** 2012. Export of the pneumococcal phage SV1 lysin requires choline-containing teichoic acids and is holin-independent. Molecular Microbiology **87:**430-445.
175. **Hatfull GF.** 2014. Molecular Genetics of Mycobacteriophages. Microbiology Spectrum **2:**1-36.

176. **Caspar DLD, Klug A.** 1962. Physical principle in the construction of regular viruses. Cold Spring Harbor Symposia on Quantitative Biology **27:**1-23.
177. **Chang JR, Spilman MS, Rodenburg CM, Dokland T.** 2009. Functional domains of the bacteriophage P2 scaffolding protein: identification of residues involved in assembly and protease activity. . Virology **384:**144-150.
178. **Duda RL, Martincic K, Hendrix RW.** 1995. Genetic basis of bacteriophage HK97 prohead assembly. Journal of Molecular Biology **247:**636-647.
179. **Aravind L, Walker DR, Koonin EV.** 1999. Conserved domains in DNA repair proteins and evolution of repair systems. Nucleic Acids Research **27:**1223-1242.
180. **Esposito LA, Gupta S, Streiter F, Prasa A, Dennehy JJ.** 2016. Evolutionary interpretations of mycobacteriophage biodiversity and host-range through the analysis of codon usage bias. Microbial Genomics **2:**e000079.
181. **Sanjuan R, Nebot RR, Chirico N, Mansky LM, Belshaw R.** 2010. Viral mutation rates. Journal of Virology **84:**9733-9748.
182. **Rybniker J, Kramme S, Small PL.** 2006. Host range of 14 mycobacteriophages in *Mycobacterium ulcerans* and seven other mycobactera including *Mycobacterium tuberculosis* - an application for identification and susceptibility testing. Journal of Medical Microbiology **55:**37-42.
183. **Hatfull GF.** 2014. Mycobacteriophages: windows into tuberculosis. PLOS Pathogens **10.**
184. **Carson L, Gorman SP, Gilmore BF.** 2010. The use of lytic bacteriophages in the prevention and eradication of biofilms of *Proteus mirabilis* and *Escherichia coli*. FEMS Immunology & Medical Microbiology **59:**447-455.
185. **Chibeu A, Lingohr EJ, Masson L, Manges A, Harel J, Ackermann H-W, Kropinski AM, Boerlin P.** 2012. Bacteriophages with the ability to degrade uropathogenic *Escherichia coli* biofilms. Viruses **4:**471-487.
186. **Siringan P, Connerton PL, Payne RJH, Connerton IF.** 2011. Bacteriophage-mediated dispersal of *Campylobacter jejuni* biofilms. Applied and Environmental Microbiology **77:**3320-3326.
187. **Hughes KA, Sutherland IW, Jones MV.** 1998. Biofilm susceptibility to bacteriophage attack: the role of phage-born polysaccharide depolymerase. Microbiology **144:**3039-3047.
188. **Reiger-Hug D, S. S.** 1981. Comparative study of host capsule depolymerases associated with *Klebsiella* bacteriophages. . Virology **113:**363-378.

Chapter 6: Appendix

Table A1: Identified proteins in OKCentral2016.

Protein	Known Putative Function
gp1	HNH Endonuclease
gp2	Function Unknown
gp3	Function Unknown
gp4	Function Unknown
gp5	Function Unknown
gp6	Function Unknown
gp8	Function Unknown
gp9	Lysin A
gp10	Lysin B
gp11	Terminase
gp12	Portal Protein
gp13	Capsid Maturation Protease
gp14	Scaffolding Protein
gp15	Major Capsid Subunit
gp16	Head-to-Tail Connector Protein
gp17	Function Unknown
gp18	Function Unknown
gp19	Function Unknown
gp20	Function Unknown
gp21	Function Unknown
gp22	Major Tail Subunit
gp23	Tail Assembly Chaperone
gp24	Tail Assembly Chaperone
gp25	Tapemeasure Protein
gp26	Minor Tail Protein
gp27	Minor Tail Protein
gp28	Function Unknown

Protein	Known Putative Function
gp30	Function Unknown
gp31	Function Unknown
gp32	Function Unknown
gp33	Integrase
gp34	Function Unknown
gp35	Cytidine Deaminase
gp36	Function Unknown
gp37	Function Unknown
gp38	Function Unknown
gp39	Function Unknown
gp40	Function Unknown
gp41	Function Unknown
gp42	Function Unknown
gp43	DNA Polymerase
gp44	HTH Binding Domain Protein
gp45	Function Unknown
gp46	Thymidylate Synthase
gp47	Function Unknown
gp48	Ribonucleotide Reductase
gp49	Function Unknown
gp50	Function Unknown
gp51	Function Unknown
gp52	Phosphoesterase
gp53	Function Unknown
gp54	Function Unknown
gp55	DNA Primase
gp56	DNA Primase
gp57	Function Unknown

Protein	Known Putative Function
gp58	Endonuclease VII
gp59	Esterase/Lipase
gp60	Function Unknown
gp61	DnaB - Like Helicase
gp62	Function Unknown
gp63	Function Unknown
gp64	Function Unknown
gp65	Function Unknown
gp66	RecB - Like Exonuclease
gp67	Function Unknown
gp68	Immunity Repressor
gp69	Function Unknown
gp70	Function Unknown
gp71	Function Unknown
gp72	Function Unknown
gp73	Function Unknown
gp74	Function Unknown
gp75	Function Unknown
gp76	Function Unknown
gp77	DNA Methylase
gp78	DNA Methylase
gp79	SprT - Like protease
gp80	Function Unknown
gp81	Function Unknown
gp82	Function Unknown
gp83	Function Unknown
gp84	Function Unknown
gp85	Function Unknown

Table A2: The protein phams of the A cluster amino acid sequences used to construct dendrograms.

Phage	Subcluster	DNA Polymerase Pham	Lysin A Pham	Major Capsid Protein Pham
Doom	A1	32894	22627	5376
20ES	A2	32894	17824	3146
Margo	A3	32894	16641	3146
Burger	A4	32894	22627	3146
UnionJack	A5	32894	24417	3146
ToneTone	A6	32894	11875	3146
Sheen	A7	32894	17792	3146
Astro	A8	32894	24417	3146
Catalina	A9	32894	18207	3146
OKCentral2016	A10	32894	16641	3146
Jabith	A11	32894	22627	3146
DarthPhader	A12	32894	16641	3146
Phlei	A13	32894	18207	3146
Luchador	A14	32894	18207	3146
EagleEye	A16	32894	17824	3146
40AC	A17	32894	22627	3146
MyraDee	A18	32894	22627	3146

Table A3: The protein phams of representatives of the A3, A4, and A10 subcluster

Phage	Subcluster	DNA Polymerase Pham	Lysin A Pham	Major Capsid Protein Pham
Margo	A3	32894	16641	3146
Microwolf	A3	32894	16641	3146
Phoxy	A3	32894	16641	3146
Rockstar	A3	32894	16641	3146
Taurus	A3	32894	16641	3146
Vix	A3	32894	16641	3146
Bruiser	A4	32894	22627	3146
Burger	A4	32894	22627	3146
Eris	A4	32894	22627	3146
Maverick	A4	32894	22627	3146
Peaches	A4	32894	22627	3146
Goose	A10	32894	16641	3146
OKCentral2016	A10	32894	16641	3146
Rebeuca	A10	32894	16641	3146
Trike	A10	32894	22627	3146
Twister	A10	32894	16641	3146