

CHARACTERIZATION OF GORDONIA RUBRIPERTINCTA BACTERIOPHAGE  
WITH MYOVIRIDAE MORPHOLOGY

by

Bonnie Madden  
A Thesis  
Submitted to the  
Graduate Faculty  
of  
George Mason University  
in Partial Fulfillment of  
The Requirements for the Degree  
of  
Master of Science  
Biology

Committee:

\_\_\_\_\_

Dr. Anne Scherer, Thesis Chair

\_\_\_\_\_

Dr. Geraldine Grant, Committee  
Co-Chair

\_\_\_\_\_

Dr. Patrick Gillevet, Committee  
Member

\_\_\_\_\_

Dr. Donald Seto, Committee  
Member

\_\_\_\_\_

Dr. Iosif Vaisman, Director,  
School of Systems Biology

\_\_\_\_\_

Dr. Cody W. Edwards, Senior Associate  
Dean for Faculty and Academic Affairs,  
College of Science

\_\_\_\_\_

Dr. Fernando R. Miralles-Wilhelm,  
Dean, College of Science

Date: \_\_\_\_\_

Spring Semester 2023  
George Mason University  
Fairfax, VA

Characterization of *Gordonia Rubripertincta* Bacteriophage with *Myoviridae*  
Morphology

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of  
Science at George Mason University

by

Bonnie Madden  
Bachelor of Science  
George Mason University, 2015  
Associate of Science  
College of DuPage, 2004

Director: Anne Scherer, Chair  
Department of Biology

Spring Semester 2023  
George Mason University  
Fairfax, VA

Copyright 2023 Bonnie Madden  
All Rights Reserved

## **DEDICATION**

I dedicate this work to my family for all the sacrifices they've made to support me on my journey. John, I would have never begun this journey if not for you. Johnny and Josh, you are never too old to grow, learn, and explore new paths. To my mother, whom I wish could be here to see this. My friends, for being there for me and believing in me. You've helped me to be patient and trust in the process. To all the women in The Elsinore Book Club who've witnessed my journey and cheered for me on since I was an undergraduate. Each one of you is an inspiration.

## ACKNOWLEDGEMENTS

I wish to thank the members of my committee for their guidance, support, and most valuable input into this thesis: Dr. Anne Scherer, Dr. Geraldine Grant, Dr. Patrick Gillevet, and Dr. Donald Seto. Dr. Scherer, this work would not exist if it were not for you! I would also like to thank Dr. Deborah Polayes, who opened so many doors for me. If not for her infinite wisdom, my work would not have been possible. Joanne Andersen-Zimmerman, who has been there for me every step of the way. All the amazing students at George Mason University, each and every one of you are inspirations and superstars. Dr. Larry Rockwood, thank you for bringing me into the Biology Department. I still miss listening to the music from your office. Finally, I would like to thank all the faculty and staff in the Biology Department for their help in making this happen.

## TABLE OF CONTENTS

	Page
List of Tables .....	vii
List of Figures .....	viii
List of Abbreviations and/or Symbols .....	ix
Abstract .....	xi
Chapter One   INTRODUCTION .....	1
1.1   <i>Background</i> .....	1
1.1.1   <i>Overview</i> .....	1
1.1.1.1   <i>History</i> .....	2
1.1.1.2   <i>Evolution</i> .....	3
1.1.1.3   <i>Applications</i> .....	6
1.1.2   <i>Bacteriophages</i> .....	9
1.1.2.1   <i>Genomes</i> .....	9
1.1.2.2   <i>Structures</i> .....	11
1.1.2.3   <i>Lifecycles</i> .....	13
1.1.3   <i>SEA-PHAGES</i> .....	15
1.1.3.1   <i>Program</i> .....	15
1.1.3.2   <i>Workflow</i> .....	16
1.1.4   <i>Optimizing workflow, introducing a new host, and a new challenge.</i> .....	17
1.1.5   <i>Reducing phylogenetic gaps with a new protocol</i> .....	19
Chapter Two   Materials and Methods.....	21
2.1   <i>Overview of collecting, isolating, purifying, and amplifying phages</i> .....	21
2.2   <i>Combined and enriched and direct isolation methods</i> .....	22
2.3   <i>DNA extraction methods using a Promega DNA Wizard Clean-Up System</i> .....	23
2.4   <i>Other methods of extraction</i> .....	24
2.5   <i>Modified phenol-chloroform protocol</i> .....	24
2.5.1   <i>Day One</i> .....	24
2.5.2   <i>Day Two</i> .....	25
2.6   <i>Quantification and Characterization</i> .....	26
2.6.1   <i>Quantification</i> .....	26

2.6.2   <i>Restriction Digest</i> .....	26
2.6.3   <i>Gel Electrophoresis and Imaging</i> .....	26
2.7   <i>Sequencing and Assembly</i> .....	27
2.8   <i>TEM Imaging</i> .....	28
2.9   <i>Phamerator analysis</i> .....	28
2.10   <i>Gene content similarity</i> .....	29
2.11   <i>Host Range</i> .....	29
Chapter Three   <i>Results</i> .....	31
3.1   <i>Overview of Methods used to extract DNA from phage capsids.</i> .....	31
3.2   <i>DNA extracted using a Promega DNA Wizard Clean-Up System per SEA-PHAGES instruction.</i> .....	32
3.3   <i>Detection for the presence of Gordonia phage and identification of morphology</i> .....	35
3.4   <i>Other extraction methods</i> .....	36
3.5   <i>Modified phenol-chloroform extraction</i> .....	38
3.6   <i>Phamerator analysis DQ cluster phages</i> .....	40
3.6.1   <i>Synteny among DQ cluster phages</i> .....	40
3.6.2   <i>Areas of Interest within the genome identified in Phamerator.</i> .....	42
3.7   <i>Gene content similarity</i> .....	44
3.8   <i>Host Range - Testing infectivity against other hosts</i> .....	50
Chapter Four   <i>Discussion</i> .....	57
4.8   <i>Future Directions</i> .....	60
Appendix   1 .....	62
References .....	63

## LIST OF TABLES

Table	Page
Table 1   List of current phage trials in the United States.....	7
Table 2   List of Phages used for investigations.....	21
Table 3   List of phages showing titers, concentrations, and relationship to each other...	33
Table 4   Percentage of GCS for DQ cluster phages.....	50
Table 5   Comparison table for percentage of GCS for DQ cluster phages.....	50
Table 6   Levels of infectivity against different <i>Gordonia</i> hosts.....	52

## LIST OF FIGURES

Figure	Page
Figure 1   <i>Gordonia rubripertincta</i> bacteriophages, isolated at George Mason .....	2
Figure 2   Representation of the diversity of bacteriophages. Phages may have .....	6
Figure 3   Figure 3 shows the process of creating a phage display library .....	8
Figure 4   Structural and bonding differences between.....	10
Figure 5   Classic morphologies of tailed phages .....	12
Figure 6   Three life cycles of bacteriophage: lysogenic, lytic, and chronic.....	14
Figure 7   Big Flow chart downloaded from the SEA-PHAGES.....	18
Figure 8   Flow chart showing different methods investigated.....	32
Figure 9   Gels of DNA extraction using Promega DNA Wizard Clean Up (Promega)...	34
Figure 10   Transmission Electron Microscopy (TEM) images.....	36
Figure 11   Electrophoresis gel of DNA extracted from ChisanaKitsune.....	38
Figure 12   Image of an electrophoresis gel of DNA .....	39
Figure 13   Map showing synteny among DQ cluster phages.....	41
Figure 14   Map showing a close-up of an area of interest .....	43
Figure 15   Map showing GC/AT content for DQ cluster phages. ....	45
Figure 16   Phylogenetic tree of <i>Gordonia rubripertincta</i> .....	48
Figure 17   Phylogenetic tree of nine DQ cluster phages.....	49
Figure 18   Petri dishes of ChisanaKitsune showing plaque formations.....	53
Figure 19   Petri dishes of Hanem showing plaque formations .....	54
Figure 20   Petri dishes of Kabocha showing plaque formations.....	55
Figure 21   Petri dishes of Schomber showing plaque formations.....	56
Figure 22   Diagram of modified phenol-chloroform protocol.....	62

## LIST OF ABBREVIATIONS AND SYMBOLS

Absorbance .....	A
Approximately .....	~
Basic Local Alignment Tool Nucleotide .....	BLASTN
Basic Local Alignment Tool.....	BLAST
Celsius.....	C
Chloride.....	Cl
Clustered Regularly Interspaced Short Palindromic Repeats .....	CRISPR
CRISPR-associated enzyme.....	Cas
Cytosine .....	C
Degree .....	°
Deoxyribonucleic acid .....	DNA
Double-Stranded Deoxyribonucleic acid.....	dsDNA
Double-Stranded Ribonucleic Acid .....	dsRNA
<i>Escherichia coli</i> .....	<i>E. coli</i>
Ethylenediaminetetraacetic acid .....	EDTA
Gene Content Similarity .....	GCS
Gene Product.....	gp
Genomic Sciences Laboratory .....	GSL
George Mason University.....	GMU
<i>Gordonia</i> .....	<i>G.</i>
Greater than or equal to.....	≥
Greater Than .....	>
Guanidine.....	G
Guanine.....	G
Horizontal Gene Transfer .....	HGT
Howard Hughes Medical Institute .....	HHMI
Kilobase .....	Kb
Magnesium.....	Mg
Microliter .....	μL
Milligram .....	mg
Milliliter .....	mL
Millimeter .....	mm
Molar.....	M
<i>Mycobacterium smegmatis</i> .....	<i>M. smegmatis</i>
Nanogram.....	ng
National Center for Biotechnology Information.....	NCBI
North Carolina State University.....	NCSU
Open Reading Frames.....	ORFs
Peptone Yeast Calcium.....	PYCa
Percent.....	%

Phage Evidence Collection and Annotation Network .....	PECAAN
Phage Hunters Integrating Research and Education .....	PHIRE
Phenol-Chloroform-Isoamyl Alcohol .....	PCI
Plaque Forming Unit.....	PFU
Polyethersulfone .....	PES
Polyethylene Glycol.....	PEG
Relative centrifugal force.....	rcf
Ribonucleic Acid .....	RNA
Rotations per minute .....	rpm
Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science .....	SEA-PHAGES
Science, Technology, Engineering, Math .....	STEM
Single-Stranded Deoxyribonucleic acid .....	ssDNA
Single-Stranded Ribonucleic Acid.....	ssRNA
Sodium Dodecyl Sulfate .....	SDS
Thousand.....	K
Thymine .....	T
Times.....	X or x
Transmission Electron Microscope.....	TEM
Tris/Acetic Acid/EDTA .....	TAE
Ultraviolet .....	UV
United States .....	US
Units.....	U
University of Maryland Baltimore Campus.....	UMBC

## ABSTRACT

### CHARACTERIZATION OF GORDONIA RUBRIPERTINCTA BACTERIOPHAGE WITH MYOVIRIDAE MORPHOLOGY

Bonnie Madden, M.S.

George Mason University, 2026

Thesis Director: Dr. Anne Scherer

Bacteriophages are significant drivers of bacterial evolution. As bacteria mutate to avoid infection, phages evolve and develop ways to break down and avoid bacterial defenses. This race for survival has resulted in a broad spectrum of diversity within the phage population. The *Actinobacteriophages* are a group of double-stranded DNA (dsDNA) phages belonging to the order of *Caudovirales*. Based on their tail structures, these phages are further classified across three families: *Siphoviridae*, *Myoviridae*, or *Podoviridae*. Currently, 343 phages using *Gordonia rubripertincta* NRRL B-16540 (*Gordonia*) as a host are registered in The Actinobacteriophage Database at Phages DB.org. Interestingly, only nine of these phages are known to have a *Myoviridae* morphology. However, the other 334 *Gordonia* phages have *Siphoviridae* morphologies. Five of the nine *Myoviridae* phages were isolated, characterized, and registered by George Mason University students. These five phages presented difficulties

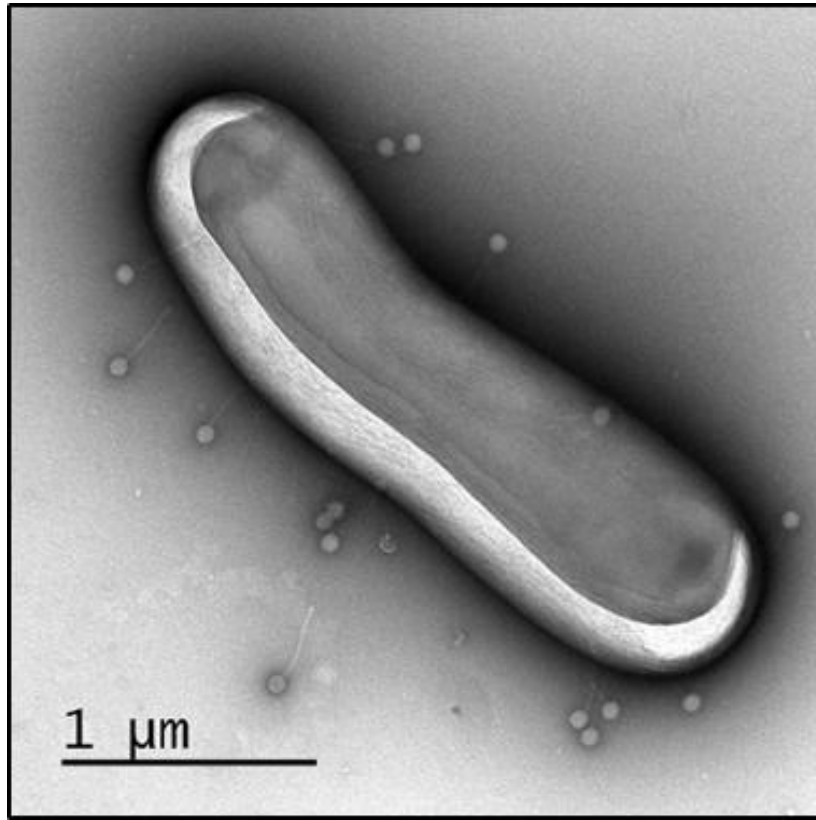
while using the standard protocol during the DNA extraction process and required a modified protocol to obtain the required starting material for downstream applications such as genome sequencing. Within these *Myoviridae* *Gordonia* phages, unique capsid proteins have been identified as being likely responsible for the difficulties experienced using the standard DNA extraction protocol. This study focuses on five novel *Myoviridae* *Gordonia* phage genomes in cluster DQ and the phylogenetic and morphological relationships between these cluster DQ phages and other *Gordonia* phages. In addition, the protocol for the modified DNA extraction is discussed. The information presented expands our knowledge of phages and this will further increase their value in both basic and applied research. Finally, due to their lytic nature, adding these phages will be powerful tools for future therapies, such as treating antibiotic-resistant infections.

## CHAPTER ONE | INTRODUCTION

### 1.1 | *BACKGROUND*

#### 1.1.1 | *Overview*

Since their discovery by Frederick W. Twort in 1915, bacteriophages (phages) have been instrumental in many advances in molecular biology and genetics [1, 2, 6, 7, 10, 12, 14, 15]. Phages are viruses that target and infect bacteria by inserting their genomes into host cells and taking over their host's metabolic processes. Modern gene-editing technologies such as the CRISPR-Cas system, phage display, and novel therapies are possible due to a phage's ability to infect bacteria and integrate their genome into their host's DNA [1, 7, 10, 15]. With an estimated  $10^{31}$  existing phage particles, they are Earth's most abundant and ancient organisms [1–2, 3, 8 – 11, 13]. Despite their abundance, there are significant gaps in our knowledge about phages, their biomes, genetics, and evolution. As their use in basic and applied research has expanded, it has become vital to bridge these gaps in knowledge [1, 6, 8 – 10, 12, 14].



**Figure 1 | *Gordonia rubripertincta* bacteriophages, isolated at George Mason University, shown target its host. Images captured using a FEI Morgagni 268 100 kV TEM equipped with a Gatan Orius CCD camera by Dr. Tagide de Carvalho at the Keith R. Porter Imaging Facility at UMBC.**

#### **1.1.1.1 | *History***

Fredrick Twort and Felix d'Herelle are credited independently with the discovery of phages in 1915 and 1917, respectively [1, 2, 6, 7, 10, 12, 14 - 16]. However, Ernest Hanbury Hankin was the first to establish the effects of bacteriophages on *Vibrio cholera* and published his findings in 1896 [2]. D'Herelle established phage therapy as a mode of treatment when he successfully used it to treat a 12-year-old boy suffering from dysentery in 1919 [2]. Also, D'Herelle noted that the infected bacteria apparently mutated to defend against phages and therefore proposed using a cocktail of different phages to

treat disease and guard against resistance [2]. Even though phage therapy continued in Russia and Poland to the present day, it did not flourish in the rest of the world. With the discovery of penicillin in 1927 and the era of antibiotics, phage therapeutics was pushed to the background [2, 6]. However, research with phages continued in other research areas, such as genetics, and gave rise to the discipline of molecular biology [16].

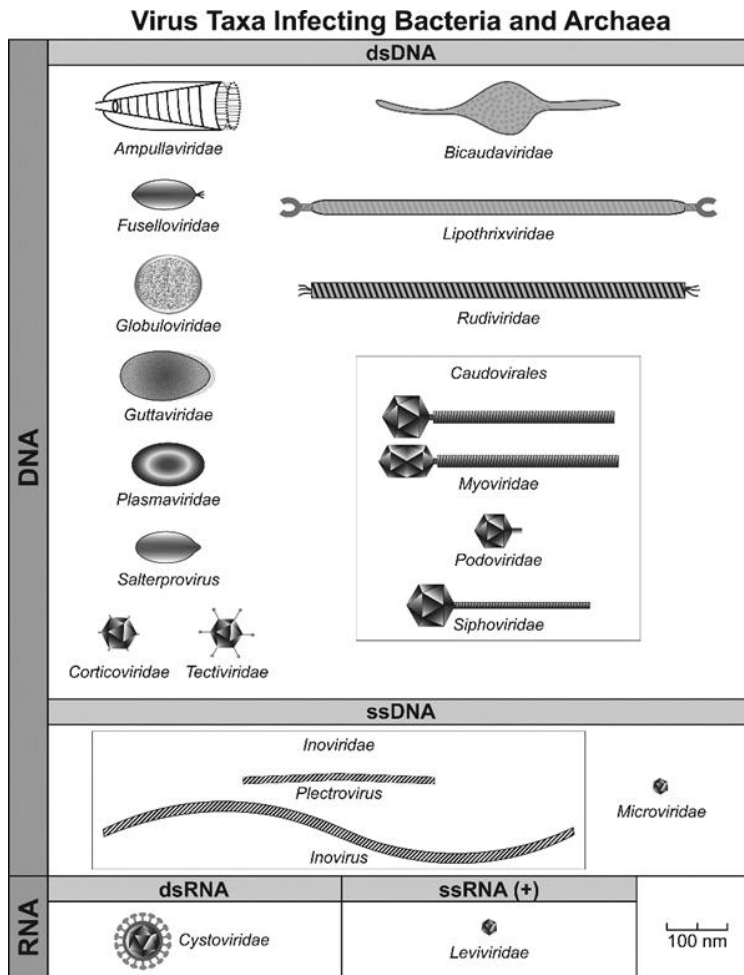
Bacteriophages were no longer used as therapeutics for the rest of the 20th century [2, 16]. However, they were used as models and tools for research. In 1940, the first phage images of structures were observed using electron microscopy, revealing their morphologies [16]. This was state-of-the-art for the identification and characterization of phages. In 1952, T2 phages were essential in proving that genetic material consisted of DNA in Hershey and Chase's pivotal publication [16]. Zinder and Lederberg, in 1952, used phages to reveal that *Salmonella* bacteria could exchange genes using bacteriophages and coined the term "genetic transduction" [16]. These vital works laid the groundwork and framework for modern molecular biology. These early steps have led to the development of powerful techniques and tools, such as gene sequencing, the capabilities to map entire genomes and gene-editing using the CRISPR-Cas 9 system [16]. Along with this progress came new insights into bacteriophages, their diversity, and evolution.

#### **1.1.1.2 | Evolution**

Bacteriophages are incredibly diverse, with a genome comprised of either RNA or DNA, which can be either single-stranded or double-stranded, and a very wide range in genomic sizes [3, 7, 17]. As such, they are initially classified based on these genomic

distinctions and then according to their morphology [18]. As shown in **Figure 2**, phages may be filamentous, spiked, enveloped, or have flagella [5]. Some dsDNA phages may have two tails, like members of the *Bicaudaviridae*, or a single tail, like members in the *Caudovirales* [5]. The *Caudovirales* phages contain even further evidence of morphological diversity, as these tailed phages are further classified by their tail lengths into *Siphoviridae*, *Myoviridae*, and *Podoviridae*. All of the *Caudovirales* carry their dsDNA in icosahedral capsids. The *Siphoviridae* may have either one of two differently shaped capsids, a shorter or an oblong (prolate) capsid [5, 8, 9, 12, 17]. Further classification, involving the establishment of evolutionary histories and phylogenies, have proven to be complicated. Genomic sequencing has revealed that phages with similar morphologies and even shared hosts may be distantly related phylogenetically. These complications arise from the phage's genetic mosaicism, which results from the ability to transfer genes as units through horizontal gene transfer (HGT) [18, 19, 21]. Also, it has been proposed that the degree to which HGT plays a role in phage diversity depends on their lifestyles and levels of access to shared gene pools, leading to two different proposed evolutionary tracts for phages. These evolutionary tracts are based on the rate of HGT. Some have higher rates of HGT, and others have lower rates of HGT. The rates of HGT are influenced by factors such as their host and their mechanism for replication. These factors, combined with their biomes and genomic composition, result in difficulties in establishing phylogenetic relationships and evolutionary histories for bacteriophages [21].

Due to the complex nature of their evolution and phylogenetic relationships, phages may be classified by their hosts and further grouped into clusters based on the percentage of gene content similarity (GCS) [4, 19, 21]. Clustering can make performing bioinformatic tasks like annotating easier when comparing genomes for regions of similarity or synteny. The *Caudovirales* Order contains a group of phages that infect *Actinobacteria*, which are designated as Actinobacteriophages [4]. Even though it appears that many phages share a large amount of sequence similarity with other phages and also fit neatly into a cluster, there are many phages that bridge the gap between clusters and display mosaicism. As more and more novel phages are discovered and sequenced, the data points to a continuum of diversity rather than existing in discrete phylogenetic clusters [21].



**Figure 2 | Representation of the diversity of bacteriophages. Phages may have single stranded or double stranded DNA or RNA and different morphologies. Not all phages are represented in this figure. This image is from the International Committee on Taxonomy of Viruses website: <https://talk.ictvonline.org>.**

### 1.1.1.3 | Applications

Phages are significant drivers of bacterial evolution as they drive bacteria to mutate to avoid infections and survive. Global and human microbiomes are affected in this race for survival between bacteria and phages [1 - 3]. However, for humans, phages can also play a positive role in the defense against multi-drug resistance bacteria. For example, information on how phages evolve to evade host defenses and survive can be determined

and implemented in vaccine development. The structural proteins in phage capsids, spikes, and tail fibers, combined with their ability to hydrolyze capsids and penetrate biofilms, are all practical tools with the potential to overcome mechanisms involved with multi-drug resistant infections like *Pseudomonas aeruginosa* (*Pseudomonas*), *Staphylococcus aureus* (Staph), and *Escherichia coli* (*E. coli*). Several clinical trials are currently underway in the US as shown in **Table 1** [2, 3, 6, 15, 19, 20, 24].

**Table 1 | List of clinical trials using phage in the United States.** Phage trials are noted along with medical conditions being treated, and intervention protocol. This table is adapted from the U.S. National Library of Medicine’s website ClinicalTrials.gov at: Search of phage | Active, not recruiting Studies – List Results – ClinicalTrials.gov.

Title	Conditions	Interventions
Cystic Fibrosis bacteriophage Study Yale (CYPHY)	Cystic Fibrosis	<ul style="list-style-type: none"> <li>• Drug: Standard Dose</li> <li>• Other: Placebo</li> </ul>
Bacteriophage therapy in Tonsillitis	Acute Tonsillitis	<ul style="list-style-type: none"> <li>• Drug: Nebulizer inhalation irrigation of the mucous membranes of the tonsils with a bacteriophage</li> </ul>
Safety and Efficacy of the Bacteriophage Preparation, ShigActive™, in a Human Experimental Model of Shigellosis	Shigellosis	<ul style="list-style-type: none"> <li>• Biological: bacteriophage</li> <li>• Other: Placebo</li> </ul>
Phage 3 Determination of Phage/Probiotic Synergistic Effects of Gastrointestinal Health	Gastrointestinal Dysfunction	<ul style="list-style-type: none"> <li>• Dietary Supplement: PreforPro+B. subtilis DE111 probiotics</li> <li>• Dietary Supplement: Placebo</li> <li>• Dietary supplement: <i>B. subtilis</i> DE 111</li> </ul>
Ph 1/2 Study Evaluating Safety and Tolerability of Inhaled AP-P/a02 in Subjects with Chronic <i>Pseudomonas Aeruginosa</i> Lung Infections and Cystic Fibrosis	Cystic Fibrosis <i>Pseudomonas Aeruginosa</i> Lung Infection Lung Infections Pseudomonas	<ul style="list-style-type: none"> <li>• Biological: AP-PA02</li> <li>• Other: Placebo</li> </ul>
A Study Investigating the Safety, Recovery, and Pharmacodynamics of Multiple Oral Administrations of SNIPR001 in Healthy Subjects	<i>E. coli</i> infections Bloodstream Infection	<ul style="list-style-type: none"> <li>• Drug: SNIPR001</li> <li>• Drug: Placebo</li> </ul>

Phage display as presented in **Figure 3**, is another biomedical application of phage technology [25]. In this technique, a foreign gene or genes are incorporated into phage DNA and subsequently expressed as epitopes on its capsid. T7 is one example of a phage

that has been used for phage display. A common insertion site for the T7 phage is gene product (gp) 10. This site encodes for a capsid protein, and the inserted foreign genes are therefore expressed as epitopes on the surface of the capsid. The expression of these heterologous proteins on the surface of the capsid is being investigated for developing vaccines and in targeting cancerous cells [1 - 3, 6, 13 - 15].

With today's emerging infectious diseases, continued studies of these diverse organisms are essential to advancing our knowledge of how phages evade host defenses and survive while maximizing their potential in therapeutic applications.

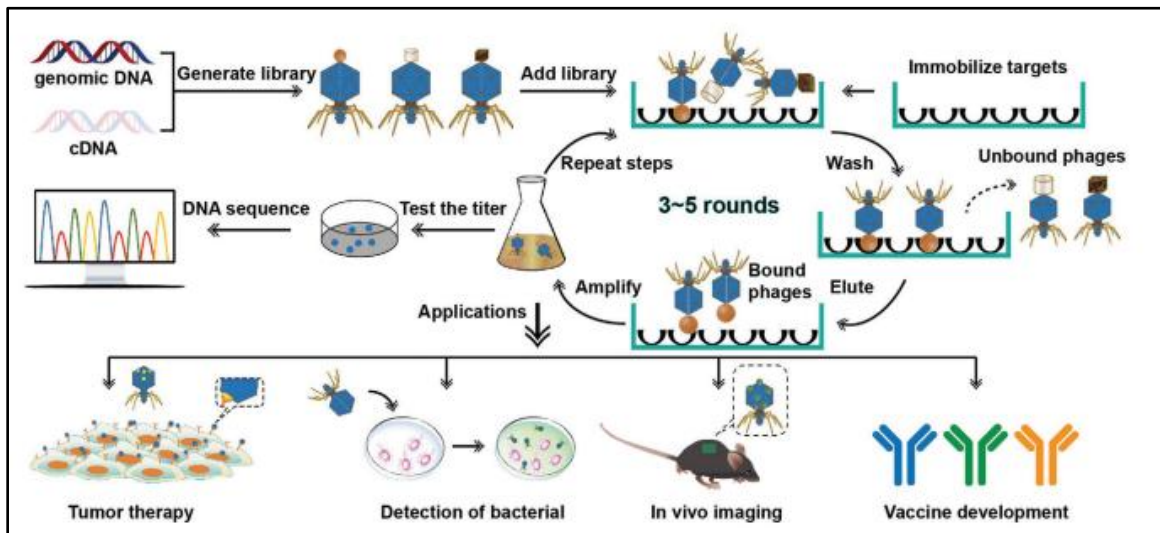


Figure 3 | The process of creating a phage display library and its applications. In this cartoon, a T7 phage is used to model the steps used to create, amplify, different applications of phage display Shown on the bottom, from left to right: phages are being used to target cancer cells, phages are used for detecting the presence of bacteria, reporter molecules are expressed on phages for in vivo imaging, and phages may express antibodies on their capsids for developing vaccines. Image from Yue, H., Li, Y., Yang, M., Mao, C., T7 Phage as an Emerging Nanobiomaterial with Genetically Tunable Target Specificity. *Adv. Sci.* 2022, 9, 2103645. <https://doi.org/10.1002/advs.202103645>.

## 1.1.2 | *Bacteriophages*

### 1.1.2.1 | *Genomes*

Bacteriophages are viruses that reproduce in bacteria and archaea by inserting their genomes into host cells and taking over metabolic processes. Although **Figure 2** highlights the diversity of phages, it does not reveal the true complexity of these viruses [5]. Beyond varied morphology, phages also have a diverse range of hosts, growth requirements, life cycles, and genomes. Their genomes may consist of either single-stranded or double-stranded RNA or DNA within their capsids. The *Caudovirales* bacteriophages have dsDNA genomes surrounded by an icosahedral-shaped protein capsid [1, 2, 6 – 12, 14, 17, 19 – 21]. These genomes may be circularly permuted or linear with 3' sticky overhangs. The size of their genomes can range from 2,435 bp to greater than 540 Kb, and they have a positive correlation between the capsid size and the genome size [9, 21, 26]. Recent studies have revealed that some phages possess a fifth nucleotide, Z, or 2-amino adenine, a purine that is similar to adenine but containing an additional amine group. This amide group adds to the hydrogen bonding capabilities of Z, resulting in a more stable overall genomic structure as shown in **Figure 4**. Having a fifth nucleotide provides thermostability and protection from host defenses [26, 27].

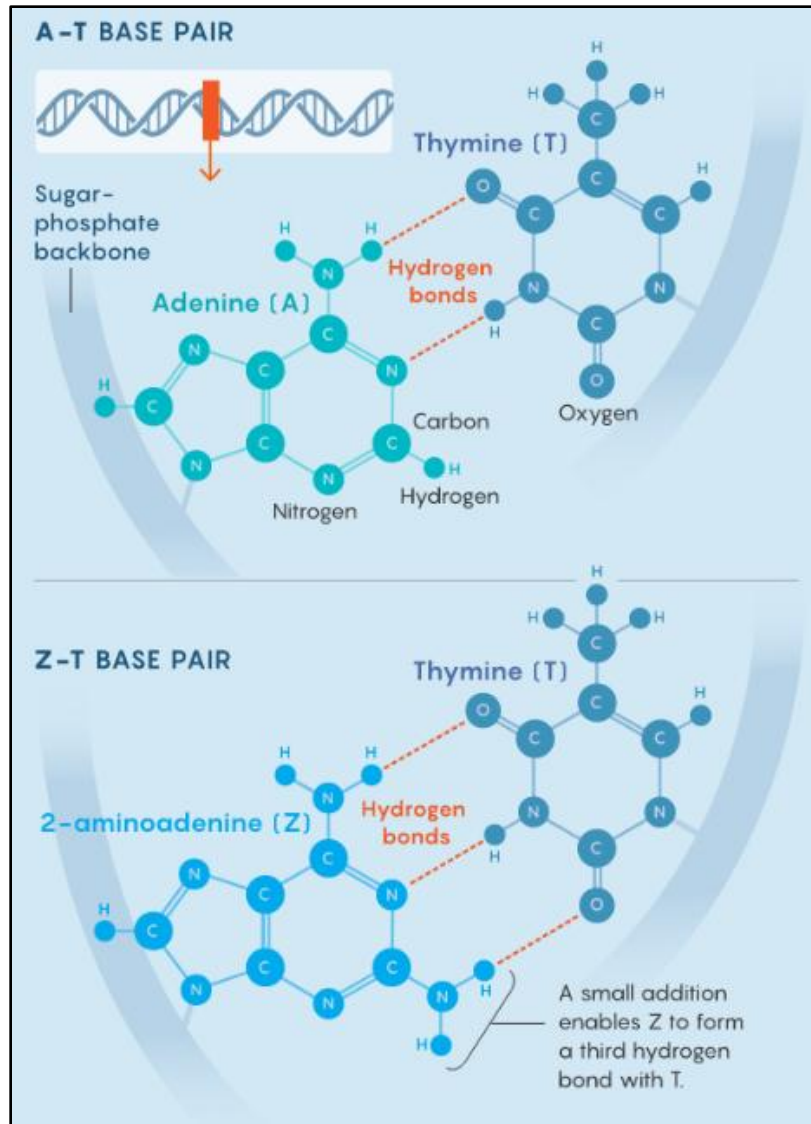


Figure 4 | Structural and bonding differences between the Watson-Crick standard A-T base pairing and the phage-specific Z-T base pairing. Image from Cepelwicz J. News: DNA has four bases. some viruses swap... (wired) - behind the headlines - NLM [Internet]. National Center for Biotechnology Information. U.S. National Library of Medicine; 2021 [cited 2021Nov21]. Available from: <https://www.ncbi.nlm.nih.gov/search/research-news/14163/>

Similarly, the *Gordonia rubripertincta* NRRL B-16540 phages range from 45,245 base pairs (bp) to 91,206 bp and contains 51.1% - 70.6% GC content [4]. As is the case with Z-containing phages, the increased GC content of DNA leads to increased genomic

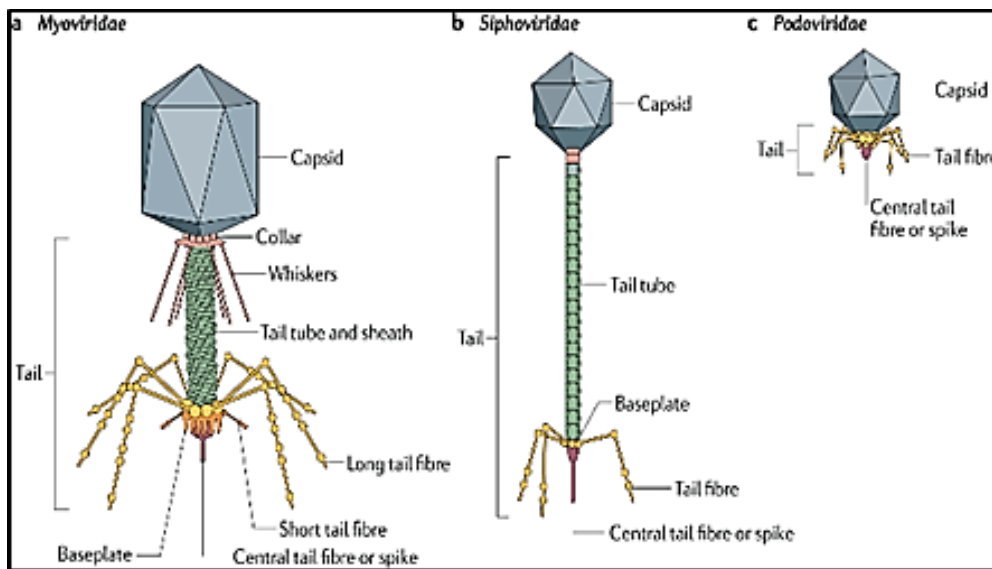
structural stability in harsh environments [4, 26, 27]. Because of the broad range of genomic diversity in *Gordonia* phages at the nucleotide level, phages are clustered in groups based on a 35% gene content similarity for convenience by the Actinobacteriophage Database at PhagesDB.org, where information on phages that infect *Actinobacteria* is curated [4, 19]. Even with a 35% gene content similarity, phages in the same cluster tend to have similar morphologies.

#### 1.1.2.2 | **Structures**

The morphologies of phages are very diverse, and the only commonality linking the organisms as a group is their ability to infect bacteria and that they are viruses, as illustrated in **Figure 2** [5]. The *Caudovirales* represent the largest group of isolated phages comprising over 95% of the phages identified [9, 28]. Their name derives from Latin and means "tailed virus". The *Caudovirales* are the most recognized phage due to their resemblance to a lunar lander. These phages hold great clinical significance because they infect human pathogens such as *Mycobacterium tuberculosis* and *Gordonia bronchialis*, which belong to the *Actinobacteria* phylum [13, 19, 29 - 31]. These distinct phages are further subdivided into three families based on their morphologies: *Siphoviridae*, *Myoviridae*, and *Podoviridae*.

These three families are classified based on their tail structures. *Siphoviridae* phages have long, flexible, non-contractile tails, whereas *Myoviridae* have medium-length, rigid, contractile tails. Finally, the *Podoviridae* phages have short, non-contractile tails, as shown in **Figure 5**. In all three families, the ends of the tails have tail fibers with receptor proteins that are used for binding to their host and are important for host

specificity. Binding to a host triggers conformational changes within the phage, allowing it to penetrate the host's cell wall with a central spike, similar to a syringe. The phage's DNA then travels down the tail tube from the capsid, where it was stored. The tail tube is attached to a vertex which serves as an attachment point for the tail at the end of the capsid [9, 12].



**Figure 5 | Classic morphologies of tailed phages.** Tail length is the basis of classification and identification. A. T4 is representative of Myoviridae B.  $\lambda$  is representative of Siphoviridae C. T7 is representative of Podoviridae. Image from Nobrega FL, Costa AR, Kluskens LD, Azeredo J. Revisiting phage therapy: New applications for old resources. Trends in Microbiology. 2015;23(4):185–91

The *Caudovirales* capsids are icosahedral-shaped and consist of protein sub-units called protomers arranged into groups of three into a triangle-shaped unit called a capsomere. A capsid's minimum sub-unit is one, and each capsid consists of a minimum of 20 capsomeres or 60 protomers. A capsid with only one subunit is given a triangulation number of  $T = 1$ . Triangulation numbers help identify the capsid's composition, shape, and size. Capsids are the limiting factor in genome size, and genome size is regulated through the process of genomic economy. Some phages with higher

triangulation numbers may also have accessory proteins that act as staples stabilizing the capsid from internal pressures created by the DNA packaged within them [32].

#### F1.1.2.3 | Lifecycles

Despite their diversity, all phages propagate by infecting a prokaryotic host and taking over the host metabolic processes for replication. After infecting the host cell, the phage enters one of three life cycles: lytic, lysogenic, or chronic - a less common life cycle as shown in **Figure 6** [10, 12]. As shown in **Figure 5**, T4 is a classic example of a lytic phage. It infects *Escherichia coli* (*E. coli*). During the lytic cycle, T4 attaches to a cell with its tail fibers and undergoes conformational changes. These changes allow the phage to penetrate the cell with a central spike and inject its DNA into its host. The host's metabolic processes are taken over to make and assemble new phage particles inside the cell, after which the cell bursts and releases the new phages to start the cycle with a new host. Lambda is a classic example of a lysogenic phage that also infects *E. coli* bacteria. However, with lambda, the DNA incorporates itself into the host's DNA, where it lies dormant as a prophage. Copies of the prophage are replicated, along with the host's DNA, during the host's growth cycles. Lysogenic phages can become lytic when their hosts experience stressors such as low nutrient levels, pH changes, or cell colony density. Temperate phage is another term used for lysogenic phages. Finally, the chronic life cycle of phage Ff here is worth noting because it also infects *E. coli* bacteria. Ff is a filamentous phage with single-stranded DNA that does not lyse its host. Instead, Ff competes for resources, slowing its host's growth rate. After assembly in the host's cell, the new phages are "extruded" [10, 12, 33].

Phages also show diversity in their host ranges; even phages within the same cluster may exhibit different host ranges. Some phages may have a narrow host range, infecting only their isolation host. In contrast, other phages may demonstrate a broader host range and are able to infect different hosts [10, 12, 30]. Understanding the lifecycles of phages allows us to continue investigating the diversity of their structures and genomes and fill in phylogenetic gaps as they become evident. Therefore, isolating and characterizing phages by extracting, sequencing, and analyzing phage genomes is paramount to filling these gaps [12, 15, 22].

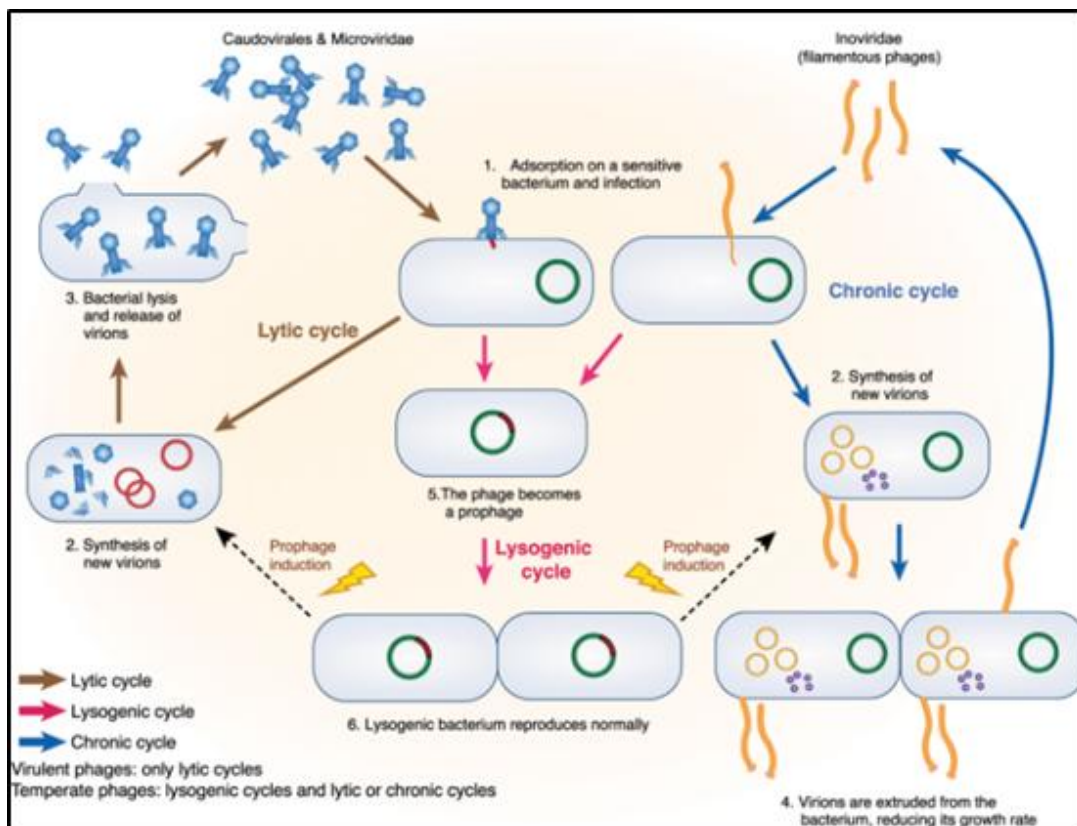


Figure 6 | Three life cycles of bacteriophage: lysogenic, lytic, and chronic. The lytic cycle is represented by brown arrows, the lysogenic cycle in red arrows, and the chronic cycle in blue arrows. Image from 6. Campbell A. The Future of Bacteriophage Biology. Nature Reviews Genetics. 2003;4(6):471–7

### **1.1.3 | *SEA-PHAGES***

#### **1.1.3.1 | *Program***

The Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program is a global Consortium that isolates, characterizes, and archives bacteriophage information in an open-source website, creating opportunities for open collaboration and undergraduate engagement in phage education and research [30]. The program's roots started in 2001 as the Phage Hunters Integrating Research and Education (PHIRE) program, created by Graham Hatfull at the University of Pittsburgh in conjunction with the Howard Hughes Medical Institute (HHMI). In 2008, HHMI, the Hatfull lab, and James Madison University collaborated on a program centered on undergraduate students participating in the research process regardless of their skill levels. This program, the National Genomics Research Initiative, ultimately became SEA-PHAGES [4, 30, 34].

The ethos behind SEA-PHAGES was to test if a national consortium of education institutions could successfully collaborate in meaningful research while recruiting, training, and retaining undergraduate students in STEM fields. The year-long program starts with the scientific discovery process by gathering soil and water samples to isolate a novel virus. Students learn the vital skills needed for success in all investigations, internships, careers, and graduate education by characterizing their own phage during the course. Ownership of their research is a powerful tool in student retention, as sequenced genomes are published in genome announcements and NCBI GenBank [30]. In 2016,

George Mason University joined SEA-PHAGES and has since enrolled nearly 250 students. The course is taught over two semesters, with Phage Discovery in the Fall semester and Phage Genomics in the Spring. Dr. Anne Scherer is the lead faculty member, and I have been assisting Dr. Scherer with all aspects of the lab course.

All isolated phages from participating institutions worldwide are archived at the University of Pittsburgh for future research including some instances where student phage lysates were used as treatments against human pathogenic organisms such as *Mycobacterium tuberculosis* and opportunistic lung infections in cystic fibrosis patients [30]. Interestingly, in the book, *The Perfect Predator* by Stephanie Strathdie, phages from the SEA-PHAGES program were used for life-saving treatments for her husband when conventional medicine and antibiotics failed [35].

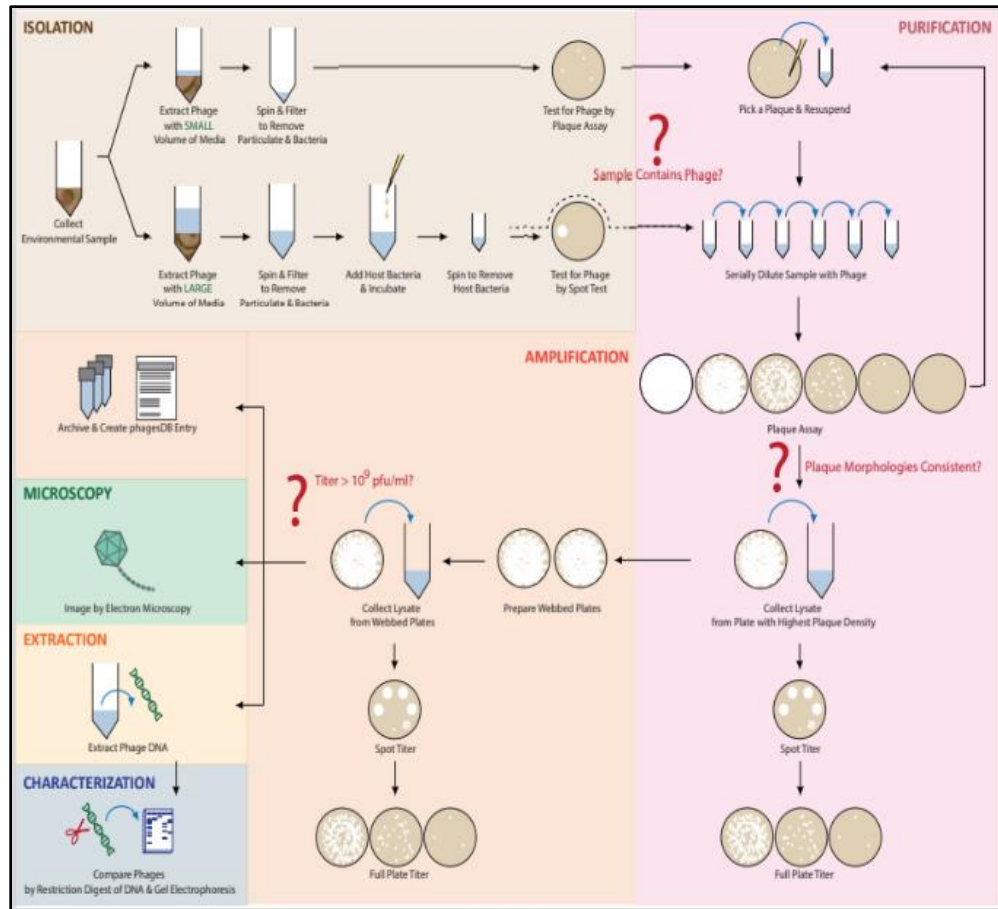
#### **1.1.3.2 | *Workflow***

The Actinobacteriophage Database at PhagesDB.org is currently an archive of over 22,592 phages. These phages infect 11 different Actinobacterial hosts and one non-Actinobacterial host (*Bacillus*) for the SEA-PHAGES program. However, only 4,423 of these registered phages have been sequenced [4]. The reason for this low number is cost restraints. Although each participating institution may register and archive an unlimited number of novel phages, only two genomes are sequenced free of charge per year at the University of Pittsburgh. The downstream applications of genome sequencing and annotation require that the DNA sample submitted meets specific quality control measures. Data for concentration levels, volume, purity, and images are submitted with

the samples to ensure high-quality sequencing and reduce misinterpretation of results [30].

#### **1.1.4 | *Optimizing workflow, introducing a new host, and a new challenge.***

The general standardized operating protocols for isolating and characterizing novel phages are outlined by SEA-PHAGES as noted in **Figure 7** [30]. The isolation process is selective since only the phages that can infect the host used for the procedure are isolated. All other phages in the environmental samples infecting other hosts are lost. Some institutions work with one host for logistical reasons such as space, coordinating incubation times, media preparation, and available resources. The hosts are chosen from a list of tested and reviewed *Actinobacteria*, and new hosts are added to the list after they have gone through the validation process outlined by the program [30]. To expand student success rate for isolation, increase our unique isolation score, and reduce wasted resources, we tested our samples with another host (*Gordonia rubripertincta* NRRL B-16540) in addition to *M. smegmatis*.



**Figure 7 | Big Flowchart** downloaded from the SEA-PHAGES website, depicts the steps use from sample collection to characterization as outlined by the SEA-PHAGES program. Available from: [https://seaphages.org/media/docs/Big\\_Flowchart\\_Outlined\\_2020.pdf](https://seaphages.org/media/docs/Big_Flowchart_Outlined_2020.pdf)

After successfully isolating, purifying, and amplifying phages, students extract genomic DNA by denaturing the capsid proteins. The DNA released from the capsid is collected for quantification and characterization. SEA-PHAGES requires a minimum concentration of 40 ng/μl but prefers 100 - 300 ng/μl of DNA with optical densities ratios of 1.80 for the A260/A280 ratio and a range of 2.0 - 2.2 for the A260/A230 ratio [30]. If enough DNA has been collected that meets SEA-PHAGES requirements for sequencing, then a restriction digest is performed on the sample. A copy of the restriction digest

image accompanies the DNA sample when sent for sequencing. The image is proof of purity, quality, and a genomic fingerprint for the DNA submitted [30]. To help meet these requirements and standardize protocols, SEA-PHAGES specifies using Promega's Wizard® DNA Clean-Up System for extraction which uses guanidinium thiocyanate to denature proteins [37].

### **1.1.5 | *Reducing phylogenetic gaps with a new protocol***

During the Fall of 2019, the Phage Discovery Class at George Mason University experienced difficulties extracting DNA from *Gordonia* phages with *Myoviridae* morphology. The extractions yielded low DNA concentrations despite working with phage lysate titer levels of  $5.0 \times 10^9$  plaque-forming units (PFU) and above.

DNA extractions were performed when minimum phage concentration levels of  $5.0 \times 10^9$  were obtained. DNA at concentrations of 40 ng/ $\mu$ L and above with an A260/280 near 1.80 were restriction enzyme digested and resolved by gel electrophoresis. However, when the gels were imaged, the presence of DNA was not visible even after many attempts.

The inability to extract and sequence DNA from a select group of phages presents several issues:

1. An informational bias is created, and knowledge gaps remain as the phages are stored without further analysis. Without sequencing and analysis, the phages remain phylogenetically unclassified.

2. The opportunity to analyze and identify unknown gene functions and maximize their potential in antibiotic-resistant therapies and other fields is lost.

Further analysis through TEM revealed that only *Gordonia* phages with *Myoviridae* morphology proved problematic for DNA isolation. Therefore, it was essential to establish a new protocol for extracting DNA from these phages, identify any possible suspect capsid or accessory protein structures responsible for this issue, and investigate its potential evolutionary role. This information is vital for developing phage-enhanced vaccines, vectors, and phage display.

This project focused on establishing a reproducible and efficient protocol for extracting and purifying phage DNA from these temperamental phages. Investigating the potential genomic players in this extraction resistance and hypothesize the role of this extraction resistance in the evolution of these phages. In addition, advancing these protocols will be essential for the success of this student-driven, research-based undergraduate class.

## CHAPTER TWO | MATERIALS AND METHODS

For a flow chart of the steps used in the modified phenol-chloroform method used to extract DNA from phages ChisanaKitsune, Hanem, Kabocha, Pakusa, and Schomber, refer to **Appendix 1**.

### 2.1 | OVERVIEW OF COLLECTING, ISOLATING, PURIFYING, AND AMPLIFYING PHAGES

All DNA extractions were performed on phages extracted from soil samples gathered from northern Virginia by students at George Mason University (Fairfax, VA). Information regarding discovery and sequencing may be found on the Actinobacteriophage Database at PhagesDB.org website. All media and solutions used were prepared as per the instructions in the SEA-PHAGES' Phage Discovery Instructor's Guide [30]. Unless otherwise detailed, the SEA-PHAGES protocols found online in the Phage Discovery Guide were followed for phage collection, isolation, purification, and amplification [30]. **Table 2** below exhibits each phage used for this study with the corresponding associated information.

**Table 2 | List of Phages used for investigations.** Phages are divided according to host. Names of founders, year isolated, cluster, morphologies, and other relevant information are presented.

Phage	Host	Finder	Year
ChisanaKitsune	<i>G. rubripertincta</i> NRRL B-16540	Bonnie Madden	2019
Hanem	<i>G. rubripertincta</i> NRRL B-16540	Sophia Vining	2019
Kabocha	<i>G. rubripertincta</i> NRRL B-16540	Naomi Jones	2019
Orla	<i>G. rubripertincta</i> NRRL B-16540	Justin Taylor Browne	2019
Pakusa	<i>G. rubripertincta</i> NRRL B-16540	Kashmala Mahmood	2022
Schombar	<i>G. rubripertincta</i> NRRL B-16540	Thomas Hutchinson	2019
Stark520	<i>M. smegmatis</i> MC <sup>2</sup> 155	Johnny Bui	2019
Usagi	<i>M. smegmatis</i> MC <sup>2</sup> 155	Bonnie Madden	2019

## 2.2 | COMBINED AND ENRICHED AND DIRECT ISOLATION METHODS

Soil samples were collected and added to a sterile 50 mL conical tube up to the 15 mL level. A peptone, yeast, dextrose, and calcium chloride (PYCa) enrichment broth was added to the tube up to the 45 mL level. Samples were then incubated for ~ one hour in a shaking incubator at 37°C, and 180 rpm. After incubation, the samples were centrifuged at 2,000 relative centrifugal force (RCF) for 10 minutes. The supernatant was vacuum filter size-separated, using a 0.22 µm polyethersulfone (PES) membrane filter, into a sterile, 50 mL conical tube.

A volume of 250 µL of *Gordonia* was aseptically added to a sterile 16 x 100 mm borosilicate test tube and inoculated with 500 µL of lysate. The same procedure was repeated with *M. smegmatis*, and both tubes sat undisturbed for 10 minutes for adsorption. After 10 minutes, 5 mL of sterile top agar was aseptically added to the tube containing lysate/*Gordonia* mix and poured onto a PYCa agar plate (100 x 15 mm). This procedure was repeated with the lysate/*smegmatis* mix. The plates were then incubated at 30°C for *Gordonia* and 37°C for *smegmatis* for two days.

Next, half of the lysate from the remaining sample in part A was transferred to a sterile 50 mL conical tube. 500 µL of *G. rubripertincta* to a sample tube labeled *Gordonia*. The tube with the remaining lysate was labeled and inoculated with 500 µL of *M. smegmatis*. Both tubes were incubated in a shaking incubator for two days at 180 rpm with *Gordonia* at 30°C and *smegmatis* at 37°C.

After incubating for two days, 1 mL of *G. rubripertincta* lysate was filtered into a sterile 1.5 ml microcentrifuge tube using a 0.22 µm PES filter and a 5 mL syringe. The same method was repeated with *M. smegmatis*. Next, 250 µL of *G. rubripertincta* was aliquoted to a sterile 16 x 100 mm borosilicate test tube. Then 5 mL of PYCa top agar was mixed with the *G. rubripertincta* and poured onto a PYCa agar plate. The same procedure was repeated for smegmatis. After the top agar solidified, two spots of 10 µL of the appropriate lysate were spotted on the plate. A third spot of 10 µL sterile phage buffer was used as a negative control. The plates were not moved for at least 15 minutes until the spots were absorbed into the agar.

### **2.3 | DNA EXTRACTION METHODS USING A PROMEGA DNA WIZARD CLEAN-UP SYSTEM**

1. Successful DNA extraction was achieved using a Promega DNA Wizard Clean-Up System (Catalog number A7280) per the SEA-PHAGES protocols for Orla, Stark520, and Usagi [30].

2. A concentrated lysate was created from ChisanaKitsune using the following polyethylene glycol 8000 (PEG) precipitation protocol. The protocol was performed in duplicate to keep the centrifuges balanced. 20 mL of phage lysate was added to 5 mL of 20% (w/v) of PEG in a sterile 50 mL conical tube and mixed. The solution was chilled on ice for one hour. After one hour, the tube was centrifuged for 20 minutes at 15,000 relative centrifugal force (g) on a Thermo Scientific Sorvall Legend X1. The supernatant was removed, and the pellet was resuspended with 2 mL of phage buffer. The resuspended solution was transferred to a 2 mL microcentrifuge tube and centrifuged for 10 minutes at 14,000 relative centrifugal force (rcf) on an Eppendorf 5420. The

supernatant was pipetted into a new, sterile microcentrifuge tube. After following this protocol, DNA extraction using the Promega system was performed on the concentrated lysate per the manufacturer's instructions.

3. SEA-PHAGES's protocol 9.1 of the Phage Discovery guide was used on lysate from ChisanaKitsune [30]. The optional step of adding 0.5  $\mu$ L of proteinase K and 50  $\mu$ L of 10% sodium dodecyl sulfate (SDS), written in the manual, was included before using Promega's system according to the directions.

#### **2.4 | OTHER METHODS OF EXTRACTION**

Sigma-Aldrich's GenElute Bacterial Genomic DNA Kit (Sigma) (catalog number NA2110-1KT) and QUIAGEN QIAprep Spin Miniprep Kit (QUIAGEN) (catalog number 27104) were used according to the manufacturer's direction. Also, a modified protocol using heat to denature capsid proteins was attempted by adding 100  $\mu$ L of lysates from ChisanaKitsune, Hanem, Kabocha, and Schomber separately to polymerase chain reaction (pcr) tubes. Using a BioRad T100 thermal cycler, the lysates were heated to 95° C for 10 minutes [38].

#### **2.5 | MODIFIED PHENOL-CHLOROFORM PROTOCOL**

##### **2.5.1 | Day One**

The following protocol is a modified protocol [39]. 2 mL of high titer phage lysate was added to a sterile, 5 mL Eppendorf tube. 25  $\mu$ L of 1 M Magnesium Chloride<sub>2</sub> (MgCl<sub>2</sub>) was added to the lysate and the solution was gently mixed by inverting the tube. Next, 2  $\mu$ L DNase I (2000 U/mL) and 20  $\mu$ L RNase A (10 mg/mL) was carefully added to the lysate-MgCl<sub>2</sub> mixture and mixed by gently inverting ~ 3 times. The lysate was

incubated in a 37°C water bath for 30 minutes. After incubation, the 200 µL of 0.5 M ethylenediaminetetraacetic acid (EDTA) was added to the tube and incubated for 10 minutes in a 75°C water bath. After incubation, the lysate was placed on ice to cool off before the addition of 5 µL of Proteinase K at a concentration of 20 mg/mL and 200 µL of 10% sodium dodecyl sulfate (SDS) solution. The solution was gently mixed by inverting ~ 3 times and incubated overnight at 27°C overnight.

### **2.5.2 | *Day Two***

The tube was gently mixed by inverting it ~ 3 times. 500 µL of the mixture was aliquoted into four sterile 1.5 mL microcentrifuge tubes and the remaining mixture added into a fifth tube. Next, 500 µL of phenol-chloroform-isopropyl alcohol (PCI) (25:24:1) was added to each tube (~ 300 µL was added to the fifth tube containing the remaining volume). The tubes were inverted several times to mix well and centrifuged at room temperature for 5 minutes at 13K rotations per minute (rpm) on an Eppendorf 5420. The top aqueous layer above the white interphase was pipetted into five new sterile 1.5 mL microcentrifuge tubes.

After pipetting into new tubes, 1 mL of ice-cold 100% ethanol and 50 µL of 3 M sodium acetate was added to each tube. The samples were placed on ice for ~ 10 minutes. After 10 minutes, the tubes were gently mixed by inverting, and the DNA formed a white, “dog snot” like precipitate. The tubes were then centrifuged at room temperature for 10 minutes at 13K rpm on an Eppendorf 5420. After centrifugation, the tubes were decanted, and the remaining droplets carefully removed by pipetting. The pellet was then washed with 500 µL of ice-cold 70% ethanol (EtOH) for a total of 3 washes. After

washing, the tubes were centrifuged at room temperature for 10 minutes at 13K rpm on an Eppendorf 5420. The tubes were decanted, and any remaining droplets carefully removed by a pipettor. The pellet was air dried for ~ 10 minutes at room temperature. 50  $\mu$ L sterile RNase, DNase free water was added to dissolve the pellets and the tubes were placed in a 37°C water bath for 10 minutes to ensure complete solvation.

## **2.6 | *QUANTIFICATION AND CHARACTERIZATION***

### **2.6.1 | *Quantification***

DNA concentrations and purity levels were measured on a Thermo Scientific NanoDrop One Microvolume UV-Vis Spectrophotometer (Nanodrop) using 1  $\mu$ L volumes and following the manufacturer's directions for measuring double-stranded DNA. A260/280 were noted. Invitrogen Ultrapure DNase and RNase-free distilled water (ultrapure) (catalog #10977023) was used as a blank.

### **2.6.2 | *Restriction Digest***

Restriction enzyme digests were performed per SEA-PHAGES's Phage Discovery Guide protocol 10.1 [30]. Restriction enzymes Bam HI-HF (catalog #R3136S), ClaI (Catalog #R0197S), EcoRI-HF (Catalog #R3101S), HaeIII (Catalog #R0108S), HindIII-HF (Catalog #R3104S), SalI-HF (Catalog #R3138S), and CutSmart reaction buffer (catalog number B7204 - discontinued) was obtained from New England Biolab (NEB). Ultrapure water was used for setting up the digests.

### **2.6.3 | *Gel Electrophoresis and Imaging***

The products of restriction enzyme digests were imaged per the SEA-PHAGES Phage Discovery Guide protocols 10.2 and 10.3 [30]. NEB's 1 kb DNA ladder (catalog

#N3232S), Gel Loading Dye, Purple (6X) (catalog #B7024S), and Invitrogen's 1 kb Plus DNA Ladder (catalog #10787018) were used to load samples with and as references. A 0.8% agarose gel was made using Bio-Rad's Certified Molecular Biology Agarose (catalog #1613102EDU) and 50x TAE electrophoresis buffer (catalog number 1660742) diluted to 1x. Invitrogen's SYBR Safe DNA Gel Stain (catalog #S33102) was added at 3  $\mu$ L/50 mL of agarose gel. Gels were electrophoresed using an Edvotek M12 apparatus (catalog #502/504) and Edvotek QuadraSource Power Supply (Catalog #5010) at 100V for one hour.

Gels were imaged using a BioRad Gel Doc EZ Gel Documentation system (discontinued) and Image Lab program. Image settings were set to use the UV sample tray with SYBR safe stains and the detection of faint bands.

## **2.7 | *SEQUENCING AND ASSEMBLY***

ChisanaKitsune, Hanem, Kabocha, and Schomber were sequenced and assembled by the staff of Dr. Baltzeger in the Genomic Sciences Laboratory (GSL) at North Carolina State University (NCSU) using Illumina sequencing technology. Orla and Pakusa were sequenced by the Pittsburgh Bacteriophage Institute at the University of Pittsburgh. Stark520 and Usagi have not been sequenced to date. Further sequencing information and links to fasta files may be found in The Actinobacteriophage Database at PhagesDB.org.

## **2.8 | TEM IMAGING**

All imaged phages were stained with a 1% uranyl acetate solution per SEA-PHAGES's Phage Discovery Guide, protocol 8.1a, except for centrifuging and resuspending the phage prior to staining [30]. ChisanaKitsune, Kabocha, and Schomber were imaged with an FEI Morgagni 268 100kV transmission electron microscope (TEM) equipped with a Gatan Orius CCD camera at the Keith R. Porter Imaging Facility, University of Maryland Baltimore Campus (UMBC,) by Dr. Tagide de Carvalho. Hanem was imaged at George Mason University by Bonnie Madden with a JEOL JEM-1400 Flash.

## **2.9 | PHAMERATOR ANALYSIS**

Genomic comparisons for synteny, identification of areas of interest, and conserved domains were performed in Phamerator Actino\_Draft (version 502). Annotations performed on Schomber have yet to be updated, and Pakusa is currently being annotated by George Mason University's Spring 2023 Phage Genomics class. The program creates pairwise alignments on genomes registered on the Actinobacteriophage Database website [4]. Annotations displayed on the genome maps in Phamerator are uploaded by the SEA-PHAGES team and based on data analysis performed by Institutions belonging to the program using the following: SEA-PHAGES website, DNAMaster, Phamerator, Phage Evidence Collection and Annotation Network (PECAAN), National Center for Biotechnology Information's (NCBI) Basic Local Alignment Search Tool (BLAST), Starterater, HHpred, SOSUI, GeneMark, and Glimmer. Colors are used to represent sequence similarity/homology. The color range

displayed starts with violet representing matches with E values of 0 and ends at red, with the cutoff being an E value of  $10^{-4}$  based on NCBI's BLAST searches. White areas do not align or have sequence similarity with other phage sequences [40 - 47].

### **2.10 | *GENE CONTENT SIMILARITY***

Gene content similarity analysis was performed using a resource on The Actinobacteriophage Database website (PhagesDB.org) [4]. Pairwise alignments were made for all phages in the DQ cluster, and the percentage of shared genes was calculated [4, 19]. Pope et al. (2017 paper) uses 35% shared gene percentage as the cutoff for phages being assigned to the same cluster. Confirmation of gene content similarity was performed with a multi-sequence alignment using Clustal Omega in Geneious Prime [49].

### **2.11 | *HOST RANGE***

All *Gordonia* hosts were grown using peptone, yeast, and calcium chloride<sub>2</sub> (CaCl<sub>2</sub>) (PYCa) media per SEA-PHAGES's Phage Discovery Guide instructions in a 30° C shaking incubator at ~ 190 - 200 rpms in baffled flasks [30]. *Gordonia rubripertincta* Grub 38 (Grub) and *Gordonia terrae* CAG3 (*terrae*) were incubated for 48 hours, while *Gordonia lacunae* NRRL B-24531 (*lacunae*) and *Gordonia westfalica* NRRL B-24152 (*westfalica*) incubated for ~ 96 hours. Lawns of the bacterial host were created by mixing 250 µL of the host with 8 mL of PYCa top agar before pouring it on a PYCa plate. Lysates of ChisanaKitsune, Hanem, Kabocha, and Schomber were serially diluted in a 1:10 ratio out to a  $10^{-8}$  dilution using phage buffer. After the top agar solidified, 3 µL of each dilution was spotted onto a host plate. The plates remained stationary until the spots were absorbed into the top agar at room temperature. After

absorption, the plates were incubated using the same incubation temperatures and times the same as when growing the hosts in liquid cultures. After the appropriate time, the plates were removed from incubation, and the results were recorded.

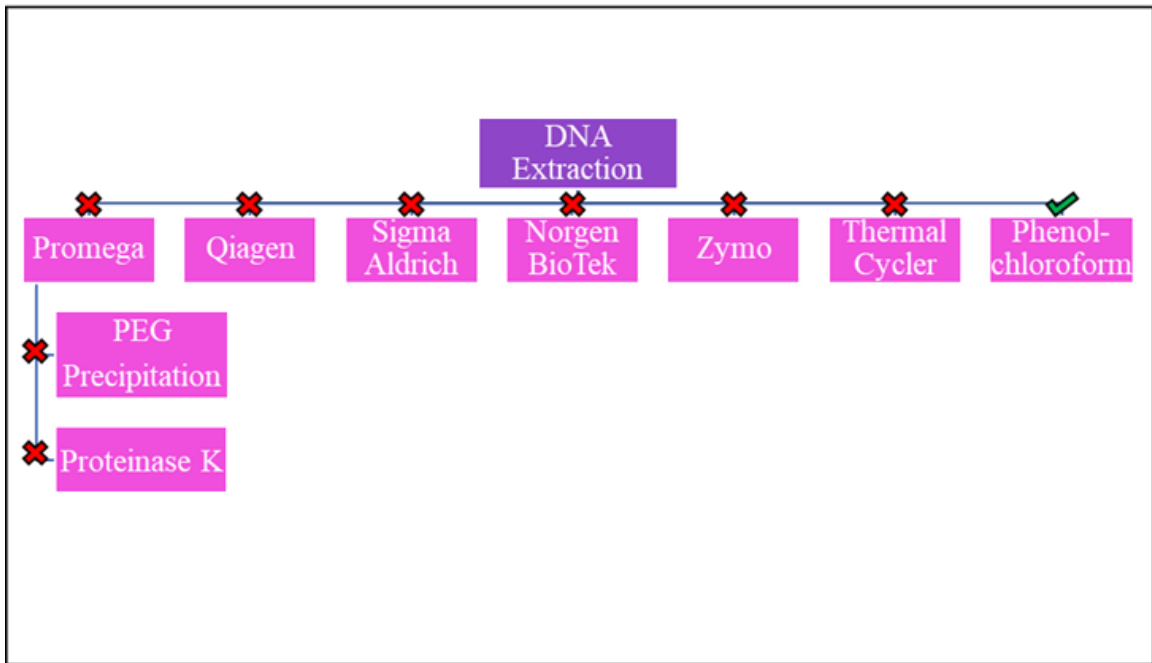
A DNA extraction was performed using Promega on a lysate of ChisanaKitsune grown on Grub. The product of the extraction was measured on a Nanodrop, and a restriction digest followed by gel electrophoresis was performed per protocols 9.1, 10.1, 10.2, and 10.3 of SEA-PHAGES's Phage Discovery Manual [30].

## CHAPTER THREE | RESULTS

### 3.1 | *OVERVIEW OF METHODS USED TO EXTRACT DNA FROM PHAGE CAPSIDS.*

Initially, all phage DNA was extracted using a Promega DNA Wizard Clean Up kit. This kit uses guanidinium thiocyanate, which denatures proteins and releases the DNA from the phage capsid. The DNA is then free to bind to silica beads and transferred to an affinity column. The beads then go through a series of ethanol washes, and the DNA is eluted from the column using a low-salt solution (i.e., water).

Using Promega's DNA Wizard Clean Up kit (Promega) and following SEA-PHAGES protocol 9.1 for DNA extraction as outlined in the Phage Discovery Guide, did not yield the results necessary for characterization and sequencing (Figures 8 and 9) [30]. DNA extraction was also unsuccessful using QIAGEN's QIAprep Spin Miniprep kit (QIAGEN) and Sigma-Aldrich's GenElute kit (Sigma) (**Figure 8**). A decision was made not to use Norgen Biotek's Phage DNA Isolation kit (Norgen Biotek) and Zymo Research's Genomic DNA Clean & Concentrator-25 kit (Zymo) as they are both guanidinium-based like Promega, QIAGEN, and Sigma-Aldrich. Processing the lysate in a BioRad Thermal Cycler at 95°C for 10 minutes to denature the capsid proteins with heat did not yield results (data not shown). Concentrating the phage particles with polyethylene glycol 8000 (PEG) before using the Promega kit and pretreating phage lysate before using the Promega kit was also unsuccessful (data not shown). However, DNA was successfully extracted using a modified phenol-chloroform protocol.



**Figure 8 | Flow chart showing different methods investigated for phage DNA extraction.** Several methods were used to extract DNA from *Gordonia* phage with *Myoviridae* morphology. The red x's indicates the methods that did not work, and the green check mark indicates the protocol that worked.

### **3.2 | DNA EXTRACTED USING A PROMEGA DNA WIZARD CLEAN-UP SYSTEM PER SEA-PHAGES INSTRUCTION.**

In their Phage Discovery Guide, SEA-PHAGES supplies protocols for extracting DNA from isolated phages [30]. DNA was extracted from high-titer lysates, and the calculated titers were close in range with a  $10^9$  order of magnitude. When using the Promega kit, The *M. smegmatis Myoviridae* phage (Usagi) yielded high DNA concentrations compared to *G. rubripertincta Myoviridae* phages (ChisanaKitsune, Schomber and Hanem), which had undetectable levels of DNA even though they were isolated from lysates with similar titers (at least  $1.0 \times 10^9$ ) (Figures 8 and 9). Also, necessary DNA concentration levels were obtained without difficulties for

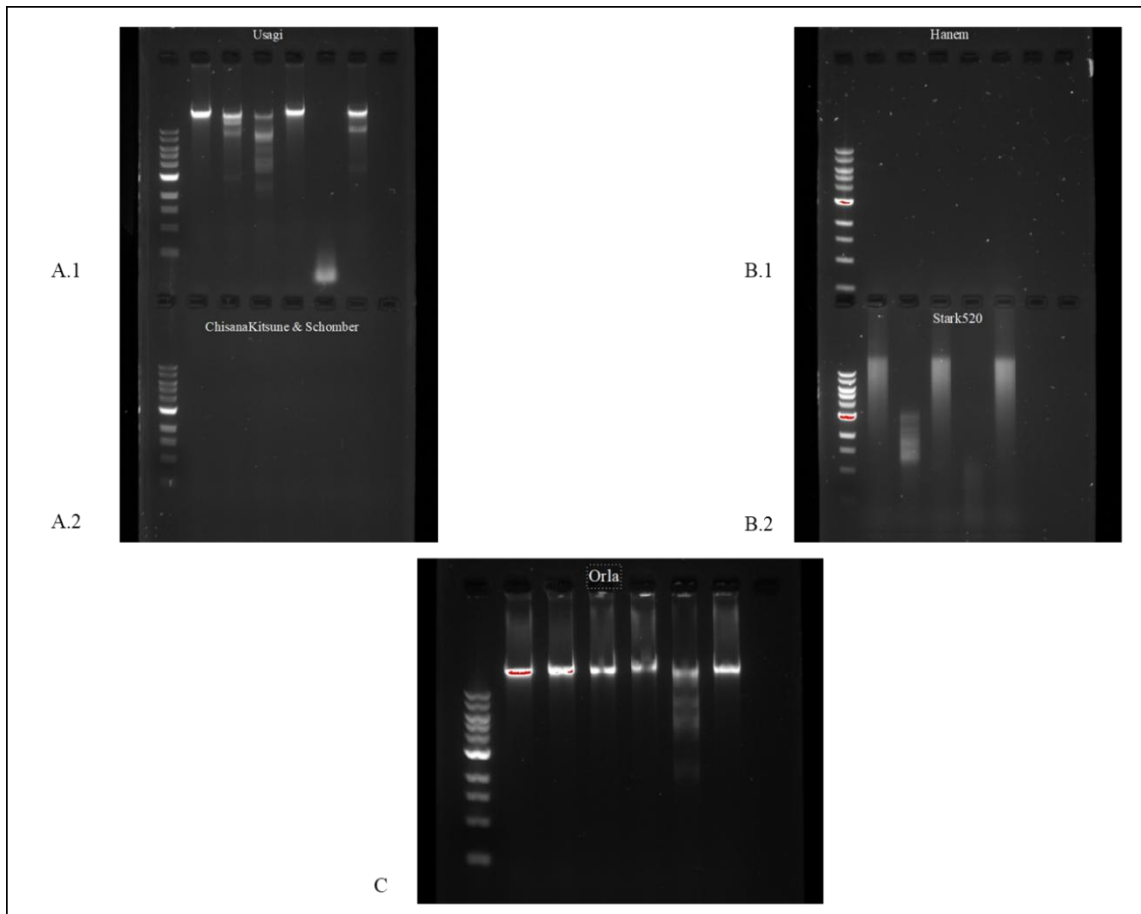
the *Siphoviridae* phages regardless of host. Only DNA extracted from the *Gordonia* phages with *Myoviridae* morphologies was not visible on gels, even when concentration levels of > 50.0 ng/μL were detected on the Nanodrop (**Table 3**).

**Table 3 | List of phages showing titers, concentrations, and relationship to each other.** Cells are color-coded to illustrate their relationship with each other. Orange was used for *Gordonia* and green for *smegmatis*, pink for *Myoviridae* morphologies, and purple for *Siphoviridae* morphologies. DNA concentration levels were measured on a Thermo Fisher NanoDrop One after using a Promega DNA Wizard Clean Up kit for extraction.

Phage	Host	Morphology	Titer (PFU)	Concentration (ng/μL)
ChisanaKitsune	<i>Gordonia rubripertincta</i>	<i>Myoviridae</i>	7.6 x 10 <sup>9</sup>	64.9
Hanem	<i>Gordonia rubripertincta</i>	<i>Myoviridae</i>	6.2 x 10 <sup>9</sup>	51.5
Orla	<i>Gordonia rubripertincta</i>	<i>Siphoviridae</i>	9.0 x 10 <sup>9</sup>	230.6
Schomber	<i>Gordonia rubripertincta</i>	<i>Myoviridae</i>	6.7 x 10 <sup>9</sup>	50
Stark520	<i>Mycobacterium smegmatis</i>	<i>Siphoviridae</i>	7.5 x 10 <sup>9</sup>	72.1
Usagi	<i>Mycobacterium smegmatis</i>	<i>Myoviridae</i>	3.4 x 10 <sup>9</sup>	172.6

Downstream applications such as restriction enzyme digests and gel electrophoresis were conducted on the phage DNA as part of the characterization process for phages submitted for sequencing to the University of Pittsburgh. As shown in **Figure 2**, DNA from the Promega extractions was not visible on the gels for ChisanaKitsune,

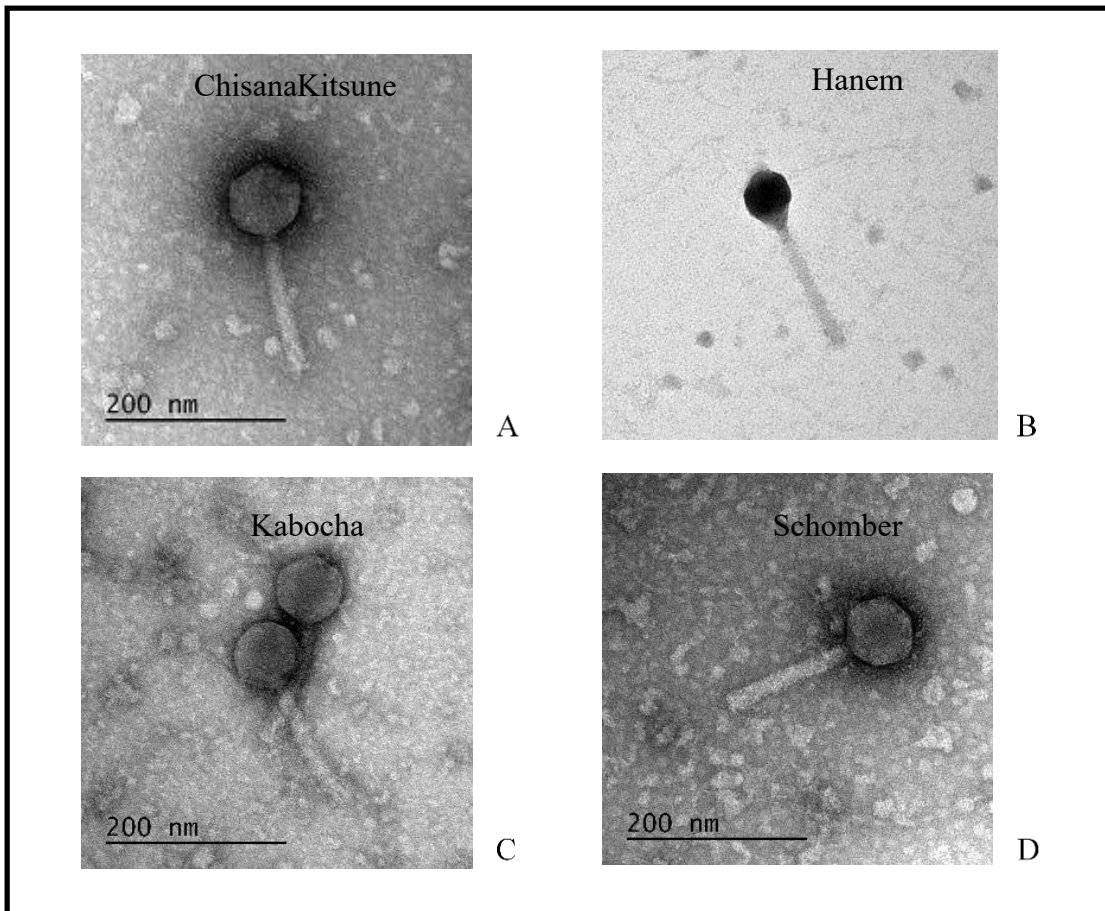
Hanem, and Schomber (Kabocha not shown), which are *Myoviridae* *Gordonia* phages. Bands of uncut DNA extracted from Stark520 and Usagi, both *M. smegmatis* phages, are visible on the gels above the 10.0 Kb markers. Orla also had visible bands above 10.0 Kb and is a *Siphoviridae* phage, the same as Stark520. This data aligns with the difficulties experienced during DNA extraction.



**Figure 9 | Gels of DNA extraction using Promega DNA Wizard Clean Up (Promega).** A.1, and B.2 DNA from a *Mycobacterium smegmatis* MC<sup>2</sup>155 phages. A.2, B.1, and C are images of DNA from *Gordonia rubripertincta* NRRL B-16540 (*Gordonia*) phages. A.1 and B.2 are both positive controls.

### **3.3 | DETECTION FOR THE PRESENCE OF GORDONIA PHAGE AND IDENTIFICATION OF MORPHOLOGY**

EM is a powerful tool used in the isolation and characterization of bacteriophages. Visualizing samples helped to confirm the presence of the isolated *Gordonia* phages described in this study and to determine their morphologies (*Siphoviridae* vs. *Myoviridae*). In **Figure 3**, each sample was inspected for presence, uniformity, and phage morphology. All four samples displayed only one type of morphology, *Myoviridae*. With the presence of phage confirmed, it was decided to use other DNA extraction methods on the *Gordonia* phages with *Myoviridae* morphologies.

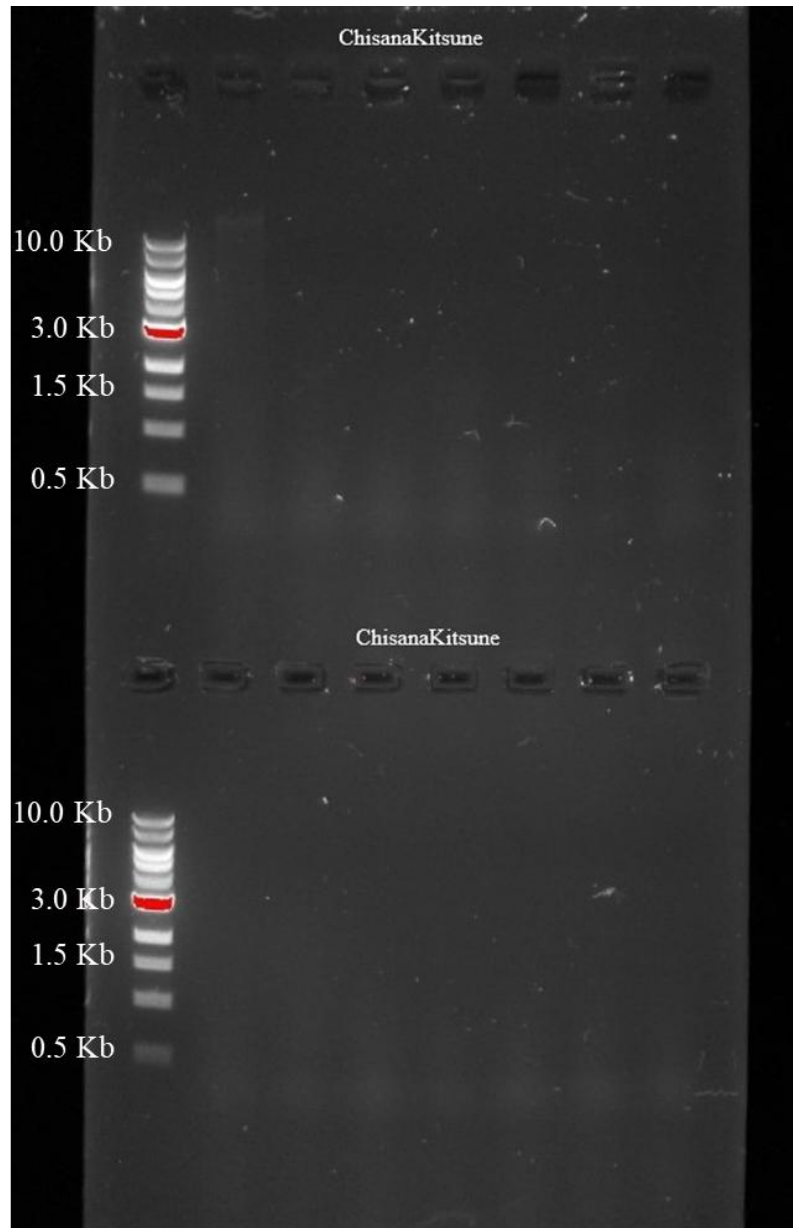


**Figure 10 | Transmission Electron Microscopy (TEM) images of Gordonia phages with Myoviridae morphologies. Image A) ChisanaKitsune. Image B) Hanem. Image C) Kabocha. Image D) Schomber. Images A, C, and D were captured using a FEI Morgagni 268 100 kV TEM equipped with a Gatan Orius CCD camera by Dr. Tagide de Carvalho at the Keith R. Porter Imaging Facility at UMBC. Image B was imaged by Bonnie Madden using a JEOL JEM-1400 Flash at George Mason University.**

### **3.4 | OTHER EXTRACTION METHODS**

Since the presence of phage was confirmed, a PEG precipitation was performed to concentrate the number of phages in the lysate. PEG precipitation of phages can increase DNA yields by increasing the number of phages in a small lysate volume. After concentrating the phages, the Promega extraction kit was used for DNA extraction per SEA-PHAGES protocol 9.1. Evidence of DNA was not observed in either the top or

bottom half of the gel (Figure 11). Other methods attempted were using proteinase K before using the Promega kit to ensure the deactivation of DNase I in case extracted DNA was being degraded by the enzyme. A QUIAGEN QIAprep Spin Miniprep Kit (QUIAGEN), and a Sigma-Aldrich GenElute Kit (Sigma) were also used per the manufacturer's directions (data not shown). Another method used to isolate DNA involved denaturing the capsid proteins by exposing the lysate to extreme heat (95°C for 10 minutes in a BioRad thermal cycler) at the beginning of the DNA isolation protocol. This approach produced no visible DNA compared to other samples treated similarly (data not shown). Gel electrophoresis was not performed on these samples since the proteinase K/Promega, BioRad, QUIAGEN, and Sigma protocols resulted in Nanodrop concentration readings of less than 20 ng/μL. Upon further investigation, it was noted that the commercial kits used were guanidinium-based, and another approach would be needed.

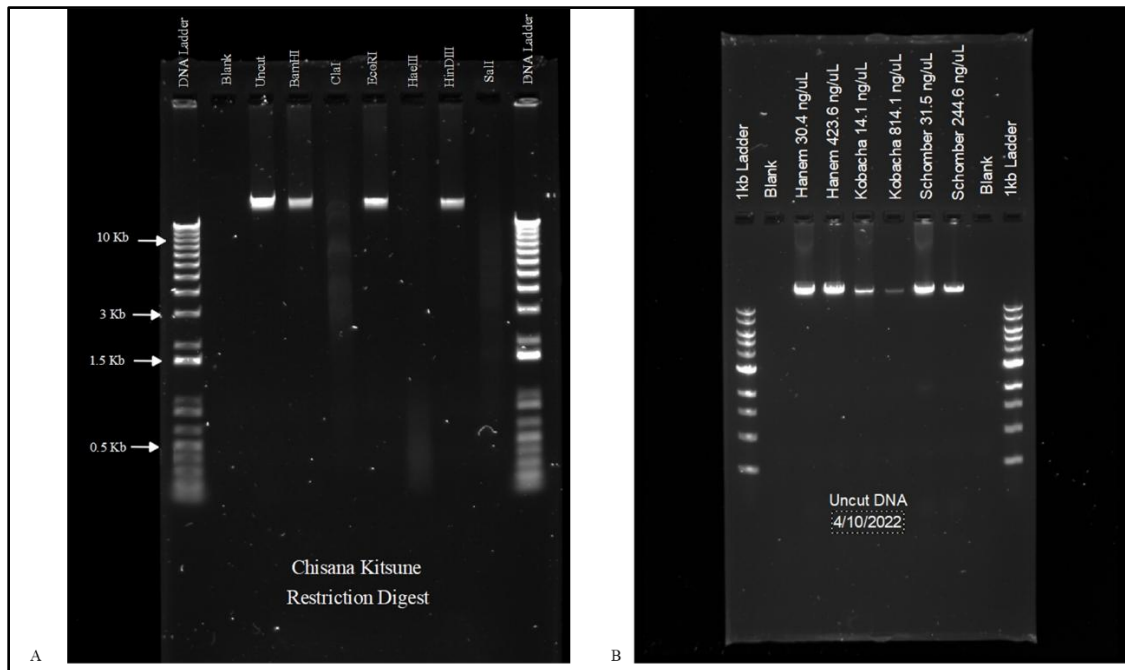


**Figure 11 | Electrophoresis gel of DNA extracted from ChisanaKitsune using a modified Promega protocol.** Lysate was concentrated using PEG 8000 prior to using the Promega kit to extract DNA. Red bands represent saturated pixels.

### 3.5 | *MODIFIED PHENOL-CHLOROFORM EXTRACTION*

Since other methods did not successfully extract the necessary DNA from the *Gordonia* phages with *Myoviridae* morphologies, a modified phenol-chloroform-

based approach was adopted. Below in Figure 12, uncut DNA extracted from ChisanaKitsune, Hanem, Kabocha, and Schomber may be observed. The DNA was extracted from these phages using the modified phenol-chloroform extraction approach followed by ethanol precipitation. For ChisanaKitsune (Figure 12 A), uncut DNA visible above the 10 Kb band in the ladder. The activity observed from ClaI, HaeIII, and Sall are consistent for these enzymes. In Figure 12 B, clear bands of uncut DNA from Hanem, Kabocha, and Schomber are visible above the 10 Kb band regardless of the concentration of DNA extracted. Both gels are confirmation of successful DNA extractions using the modified protocol.

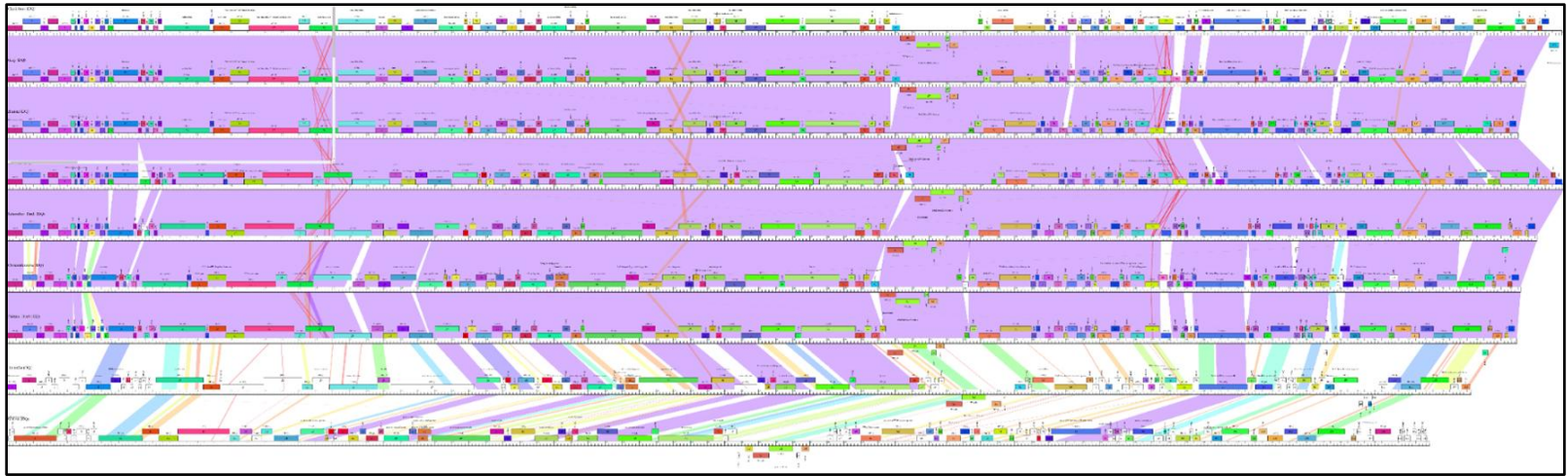


**Figure 12 | Image of an electrophoresis gel of DNA extracted from ChisanaKitsune on a 0.8% agarose gel. Restriction enzymes used are lane four, BamHI; lane five, ClaI; lane six, EcoRI; lane seven, HaeIII; Lane eight, HindIII; and Lane nine, Sall. Lanes one and ten are DNA ladders. Lane two is blank.**

### 3.6 | *PHAMERATOR ANALYSIS DQ CLUSTER PHAGES*

#### 3.6.1 | *Synteny among DQ cluster phages*

Most of the phages in cluster DQ (Chidiebere, ChisanaKitsune, Gray, Hanem, Kabocha, Pakusa, Schomber, DalanDe, and GMA6) show strong synteny (areas of homologous gene order between two genomes) and sequence similarity illustrated using the web-based application called Phamerator (**Figure 9**) [44]. Phamerator uses BLAST to make pairwise alignments at the nucleotide level and then assigns colors to represent amount of sequence similarity between the neighboring phages. Levels calculated with e-values of 0 are shown in violet. As the levels of similarity between genomes decrease, the color changes. Red is used for the least similarity, with BLASTN e-values cutting off at  $10^{-4}$ . White depicts areas with no similarities. High levels of synteny are observed between Chidiebere, ChisanaKitsune, Gray, Hanem, Kabocha, Pakusa, and Schomber, as shown by the violet colors between each pairwise alignment. The other two DQ phages, DalanDe and GMA6, display lower levels of synteny when compared to each other or the other seven phages, as shown by large areas of white [44, 47].



**Figure 13 | Map showing synteny among DQ cluster phages generated by Phamerator.** Phage genomes shown in order from top to bottom are Chidiebere, Gray, Hanem, Kabocha, Pakusa, Schomber, ChisanaKitsune, Pakusa, DalanDe, and GMA6. Levels calculated with e-values of 0 are shown in violet. As the levels of similarity between genomes decrease, the color changes. Red is used for the least similarity, with BLASTN e-values cutting off at  $10^{-4}$ . White depicts areas with no similarities.

### 3.6.2 | *Areas of Interest within the genome identified in Phamerator.*

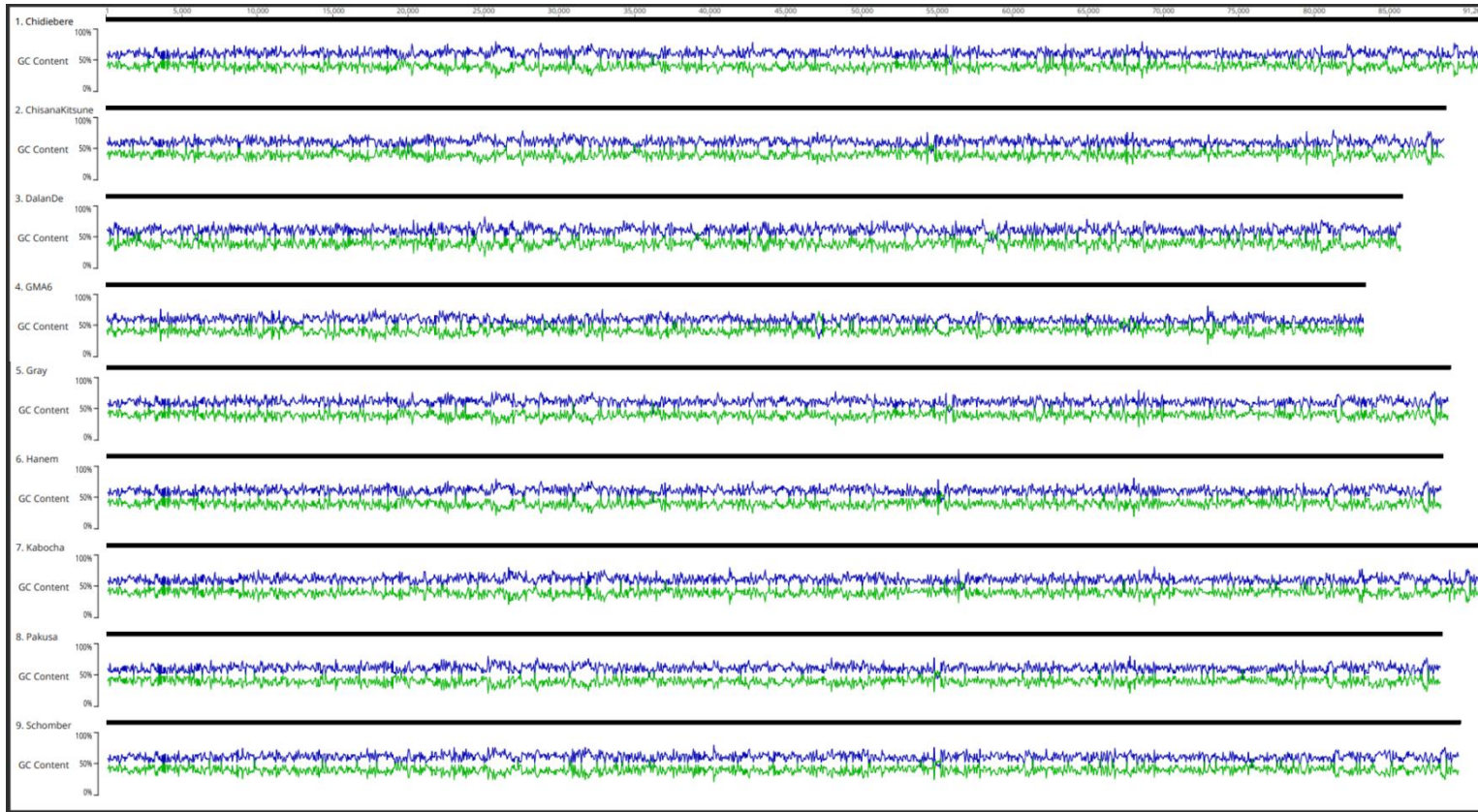
Open reading frames (ORFs) encoding capsid and tail proteins were identified within the mid-section of the genomes for phages (Figure 14 and 15). The red box outlines the region where genes encoding capsid proteins are located between genes encoding tail proteins. Phamerator organizes genes into "phamilies" or "phams" based on a content similarity of  $\geq 32.5\%$  and assigns these groups a number. If a gene is unique and not grouped, it is called an "orpham" [44]. The Pham numbers in this region coding for capsid proteins or proteins with no known function (NKF) shared by all nine DQ phages and unique to cluster DQ are 8559, 8161, 8314, and 8981. The Pham numbers in this region coding for capsid proteins or proteins with NKF not shared by all nine DQ phages or having one or two phages from other clusters are 11011, 11811, 62800, 9688, 9226, 10501, 7619, and 66295 [44]. Pham number 10501, which is shared by the seven phages that shows strong synteny and sequence similarity, contains a conserved domain. This conserved domain codes for PHA03269, envelope glycoprotein C which has been found to aid in adsorption and protection from antibodies in other viruses, such as HIV. Its role in these phages has yet to be identified and was not found in GMA6 and DalanDe (phages which were isolated and sequenced by other institutions) [4, 44].



**Figure 14 | Map showing a close-up of an area of interest among DQ cluster phages generated by Phamerator.** Map showing a close-up of an area of interest among DQ cluster phages generated by Phamerator. Enclosed in the red box are genes encoding capsid proteins, tail proteins, and NKF. The genes encoding for capsid proteins are in the middle of the genes encoding for the tail proteins. Many of these genes are unique to the DQ cluster phages.

### 3.7 GENE CONTENT SIMILARITY

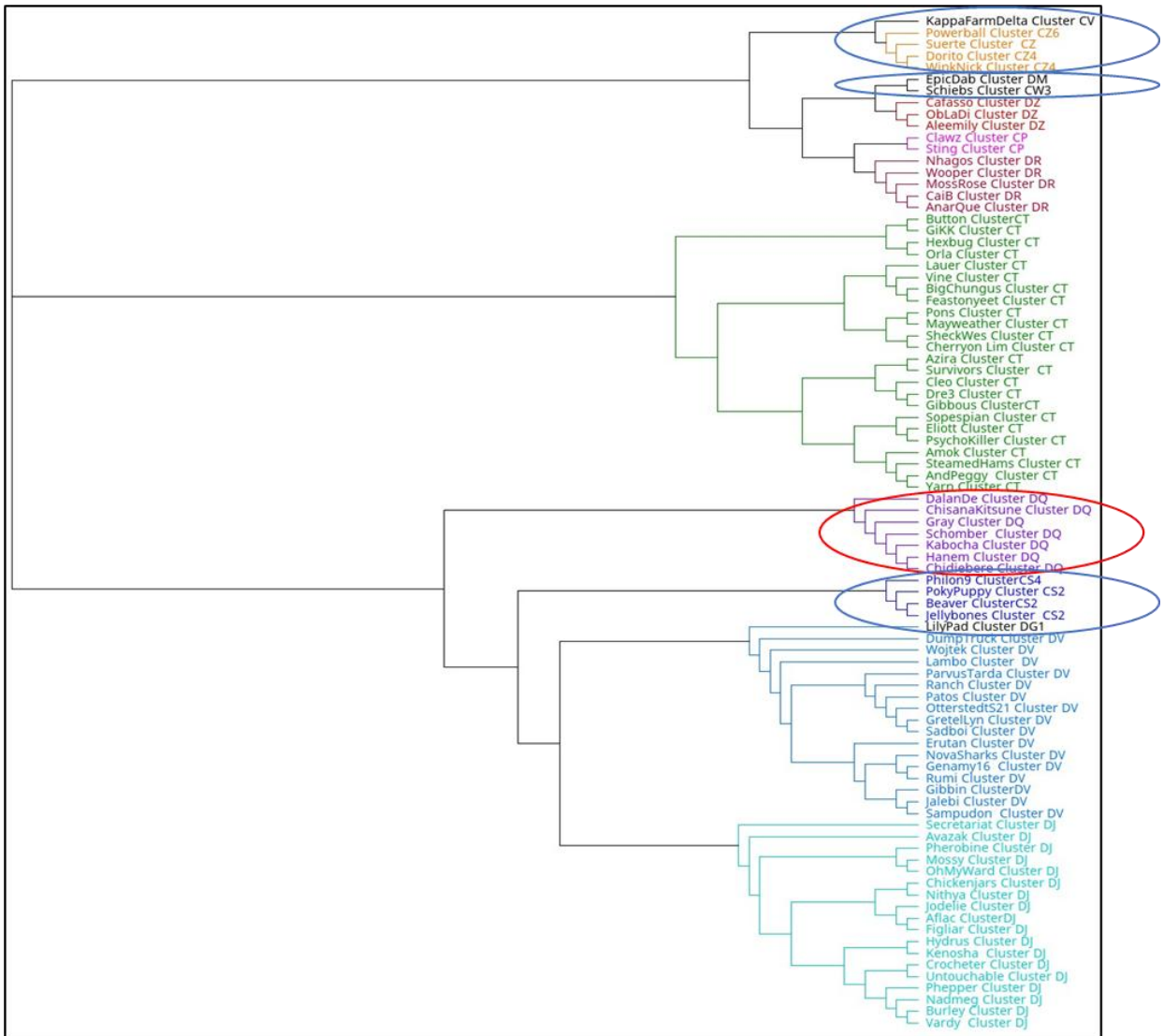
The percentage of gene content similarity reveals that seven of the DQ phages are highly similar (**Table 4**). The two with lower levels of similarity were isolated and sequenced by other institutions, and GMA6 was isolated using a different host. These numbers were calculated by an open-source tool provided by the Actinobacteriophage Database at PhagesDB.org [4]. Across the table, a high percentage of gene content similarity (GCS) trend may be observed among Chidiebere, ChisanaKitsune, Gray, Hanem, Kabocha, Pakusa, and Schomber of  $> 87.5\%$ . GMA6 displayed a lower percentage of GCS, from 52.2% to 53.2%, and DalanDe had even lower percentages of GCS ranging from 45.6% to 47.1%. These figures align with the synteny seen in a multi-sequence alignment that was performed on the same cluster of phages using Clustal Omega in Geneious Prime (data not shown) [45]. The alignment was used to make the same comparisons but at the nucleotide level. Similar trends were observed among the phages, and the resulting table may be found in **Table 5**. Except for GMA6, the percentage of GC content was  $\geq 60.0\%$ . GMA6 had a lower percentage of 58.2%, while ChisanaKitsune was 60.0%; Chidiebere, DalanDe, Kabocha, and Schomber were 60.2%; and Gray, Hanem, and Pakusa were at 60.3% (**Figure 15**).



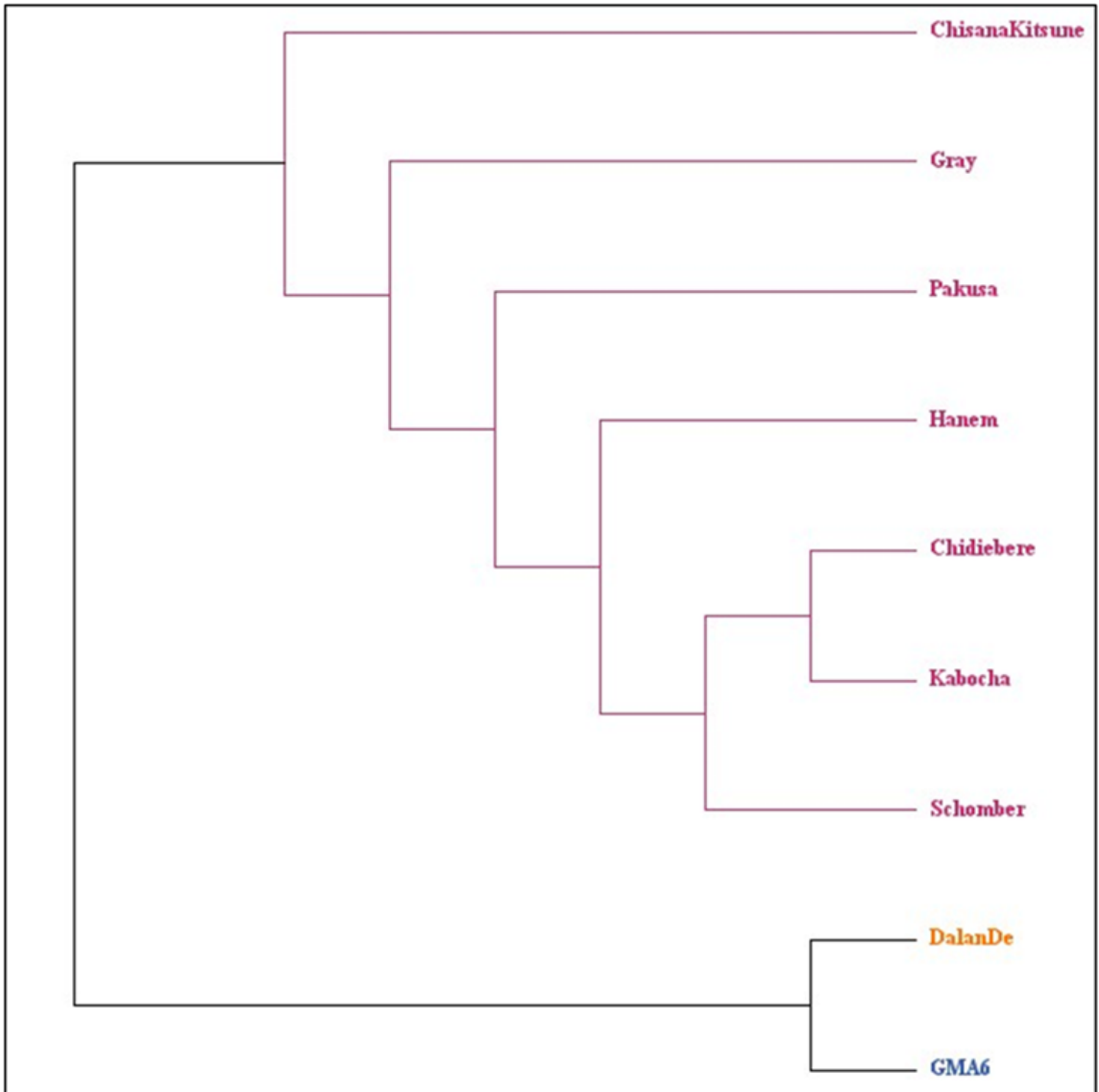
**Figure 15 | Map showing GC/AT content for DQ cluster phages, as generated by Geneious Prime, a bioinformatic software program designed for sequence analysis. The blue line shows GC content above 50% of the genome. The green line shows AT content below 50% for the genomes.**

Also investigated through Geneious Prime, a bioinformatic software program designed for sequence analysis, was the overall clustering of *Gordonia Rubripertincta* NRRL B-16540 phages [48]. A phylogenetic tree was built based on a multi-sequence alignment of 87 phages (**Figure 16**). Four single-

clustered phages are included in the data and appear closely related to other clusters. However, the cluster DQ branch in the phylogenetic tree reveals that they are a distinct group. With DQ phages creating a distinct cluster, a separate tree was created to include all DQ cluster phages regardless of isolating host (GMA6 was isolated using *Gordonia malaquae* BEN700) (**Figure 17**). DalanDe and GMA6 form a separate branch from the other DQ phages, reinforcing the GCS data and suggesting that more genomes are needed from this cluster to bridge this gap.



**Figure 16 | Phylogenetic tree of *Gordonia rubripertincta* NRRL B-16540 bacteriophages created in Geneious Prime at the nucleotide level using progressive pairwise alignments. Each cluster is given a different color to observe the different branches. Singletons are black. The blue circles highlight the four single clustered phages and where they are closely related to other clusters. The red circle highlights the DQ cluster in a distinct cluster.**



**Figure 17 | Phylogenetic tree of nine DQ cluster phages created in Geneious Prime at the nucleotide level using progressive pairwise alignments.** DalanDe and GMA6 appear to be related more to each other as they are on the same branch with each other. However, they are shown to belong to a different branch than the other seven phages.

**Table 4 | Percentage of GCS for DQ cluster phages showing the levels of gene content similarity among DQ cluster phages.** Percentages are from The Actinobacteriophage Database at PhagesDB.org. (\*) = Phages isolated at George Mason University.

Phage	Chidiebere	ChisanaKitsune*	DalanDe	GMA6	Gray	Hanem*	Kabocha*	Pakusa*	Schomber*
Chidiebere		90.2	46.6	52.6	93.3	94.9	96.9	91.3	92.9
ChisanaKitsune			47.1	52.4	92.0	92.1	91.8	87.5	87.6
DalanDe				45.6	46.5	47.1	46.6	47.1	45.9
GMA6					52.8	53.2	52.6	52.6	52.2
Gray						96.0	94.9	93.9	91.5
Hanem							96.5	94.0	93.2
Kabocha								92.1	93.7
Schomber									95.1

**Table 5 | Comparison table for percentage of GCS for DQ cluster phages at the nucleotide level for the genomes.** The table was created from data generated in pairwise alignments made in Geneious Prime. The same DQ cluster phages were used as in Table 4.

	GMA6	DalanDe	ChisanaKitsune	Gray	Hanem	Pakusa	Schomber	Chidiebere	Kabocha
GMA6		36.56	42.37	42.36	42.35	42.24	42.1	42.51	42.18
DalanDe	36.56		43.56	43.65	43.74	43.65	43.25	43.1	42.74
ChisanaKitsune	42.37	43.56		88.43	87.77	88.37	87.26	86.84	86.67
Gray	42.36	43.65	88.43		97.41	97.16	95.28	95.5	95.41
Hanem	42.35	43.74	87.77	97.41		98.16	97.24	96.47	96.51
Pakusa	42.24	43.65	88.37	97.16	98.16		95.96	95.13	95.01
Schomber	42.10	43.25	87.26	95.28	97.24	95.96		97.71	97.71
Chidiebere	42.51	43.1	86.84	95.5	96.47	95.13	97.71		98.12
Kabocha	42.18	42.74	86.67	95.41	96.51	95.01	97.71	98.12	

### 3.8 | *HOST RANGE - TESTING INFECTIVITY AGAINST OTHER HOSTS*

Since GMA6 was isolated from a different host, and the DQ phages have a unique morphology, the host range was tested. ChisanaKitsune, Hanem, Kabocha, and Schomber were tested against four different *Gordonia* hosts, *Gordonia rubripertincta* Grub 38 (Grub), *Gordonia lacunae* (*G. lacunae*), *Gordonia terrae* (*G. terrae*), and *Gordonia westfalica* (*G. westfalica*) (Figures 18 – 21). As seen in Table 6, each phage could infect

each host except for Kabocha, which did not infect *G. westfalica*. However, the levels of infectivity ranged from highly infective to mildly infective. Lysate for each phage was created using *Gordonia rubripertincta* NRRL B-16540 as a host. Titters for ChisanaKitsune were calculated at  $6.9 \times 10^9$ , and positive results were observed at  $10^{-8}$  dilution when tested against Grub. *G. lacunae* had positive results that were observable at the  $10^{-3}$  dilution level. Finally, ChisanaKitsune displayed mild infectivity levels for *G. terrae* and *G. westfalica*, with faint spots observed at  $10^{-0}$

Hanem had titers calculated at  $1.95 \times 10^{11}$ , and positive results were observed at  $10^{-8}$  dilution when tested against Grub. Hanem was less infective against *G. lacunae* and *G. terrae*, with positive results observed at the  $10^{-1}$  dilution levels. For *G. westfalica*, Hanem displayed more infectivity than *G. lacunae* and *G. terrae*, with activity observed at  $10^{-2}$ .

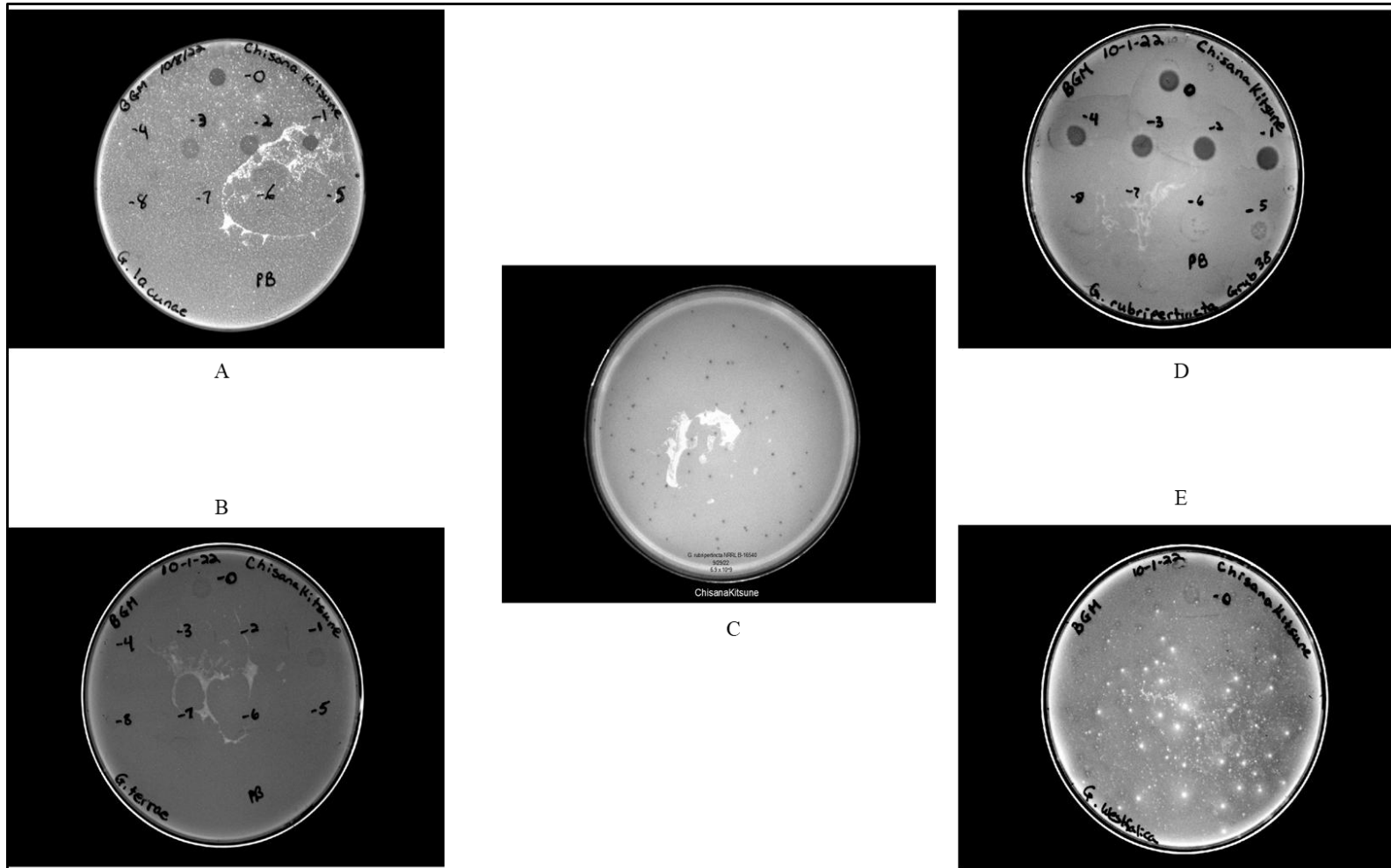
With Kabocha, titers were calculated at  $1.08 \times 10^{10}$ . Again, positive results were observed at the  $10^{-8}$  dilution when tested against Grub. However, when tested against *G. lacunae* and *G. terrae*, very light activity was observed at  $10^{-0}$ , and no activity was observed for *G. westfalica*. Titters for Schomber were calculated at  $7.4 \times 10^9$ , and positive results were observed at  $10^{-7}$  dilution when tested against Grub. *G. lacunae* had positive results that were observable at the  $10^{-2}$  dilution level. Finally, Schomber displayed mild infectivity levels for *G. terrae* and *G. westfalica*, with faint spots observed at  $10^{-0}$ .

Since *G. rubripertincta* Grub 38 is closely related to *G. rubripertincta* NRRL B-16540, the degree to which ChisanaKitsune, Hanem, Kabocha, and Schomber were able to infect *G. rubripertincta* Grub 38 was expected. However, their ability to infect *G.*

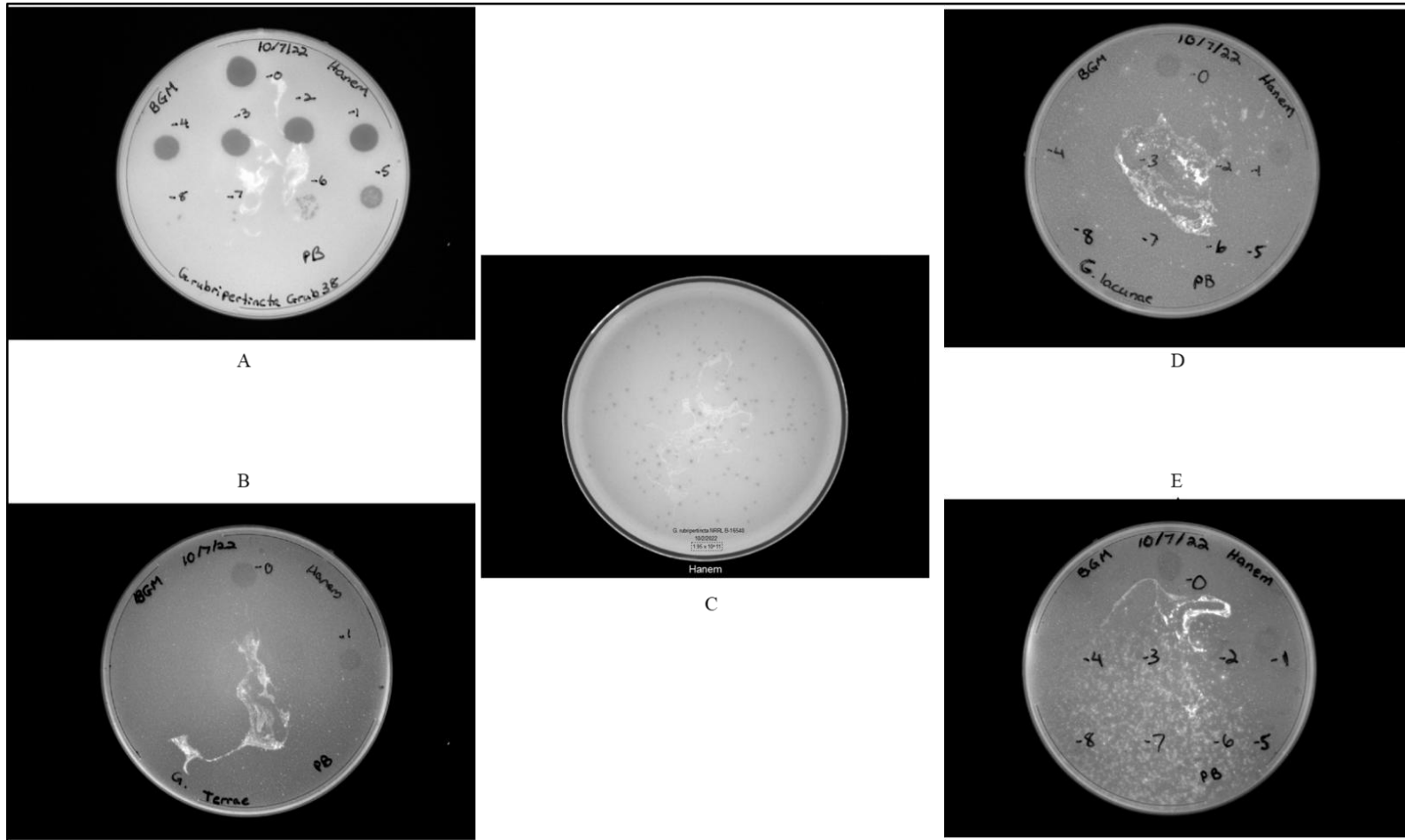
lacunae, *G. terrae*, and *G. westfalica* was more limited than expected since these hosts were within the same species.

**Table 6 | Levels of infectivity against different *Gordonia* hosts.** (+) (+) (+) = infectivity observed at dilution levels 10<sup>-6</sup> – 10<sup>-8</sup>. (+) (+) = infectivity observed at dilution levels 10<sup>-3</sup> – 10<sup>-5</sup>. (+) = infectivity observed at dilution levels 10<sup>-0</sup> – 10<sup>-2</sup>. (-) = no activity observed at any dilution level.

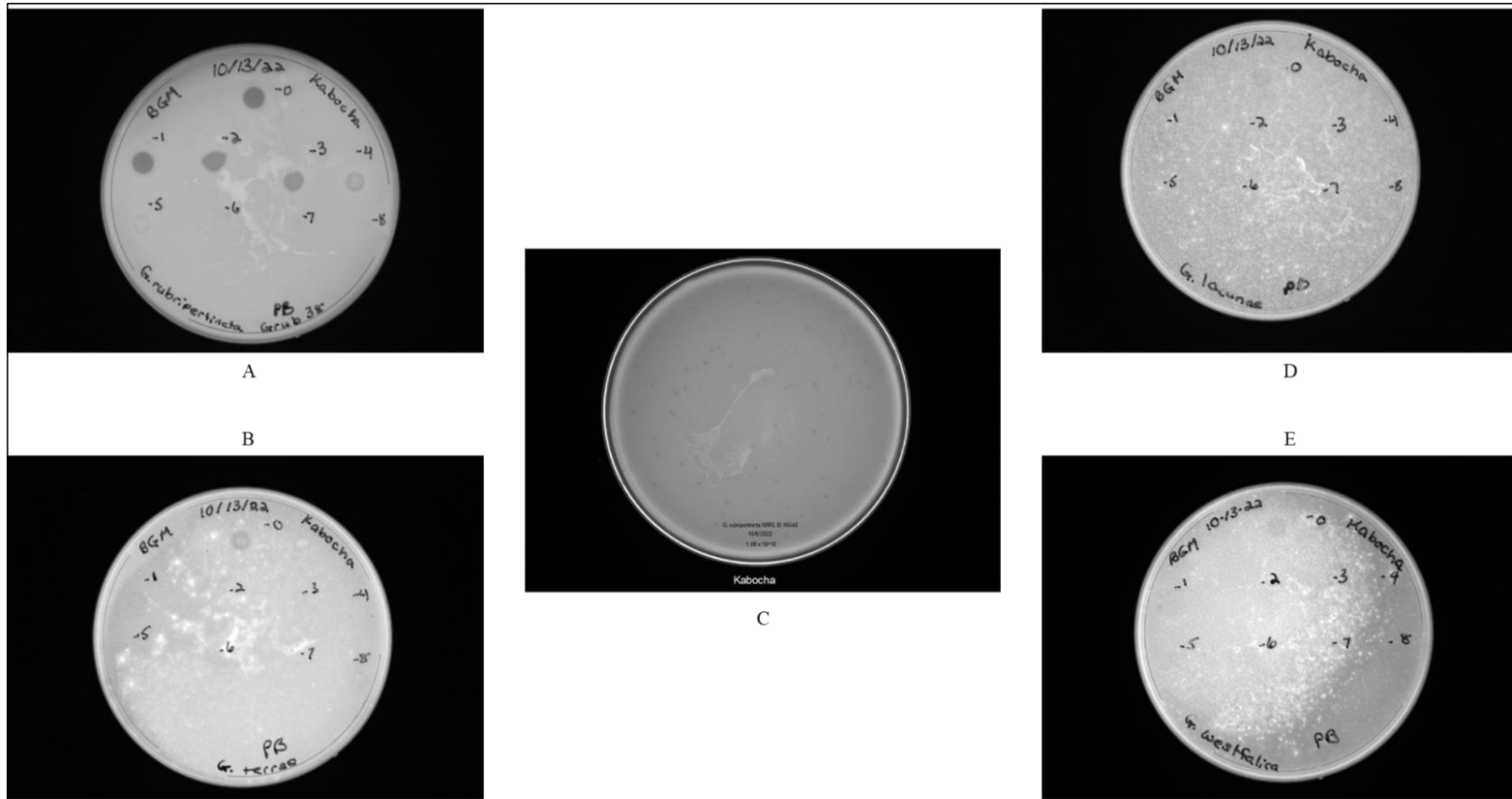
Phage	<i>G. rubripertincta</i> Grub 38	<i>G. lacunae</i>	<i>G. terrae</i>	<i>G. westfalica</i>
ChisanaKitsune	(+) (+) (+)	(+) (+)	(+)	(+)
Hanem	(+) (+) (+)	(+)	(+)	(+)
Kabocha	(+) (+) (+)	(+)	(+)	(-)
Schomber	(+) (+) (+)	(+)	(+)	(+)



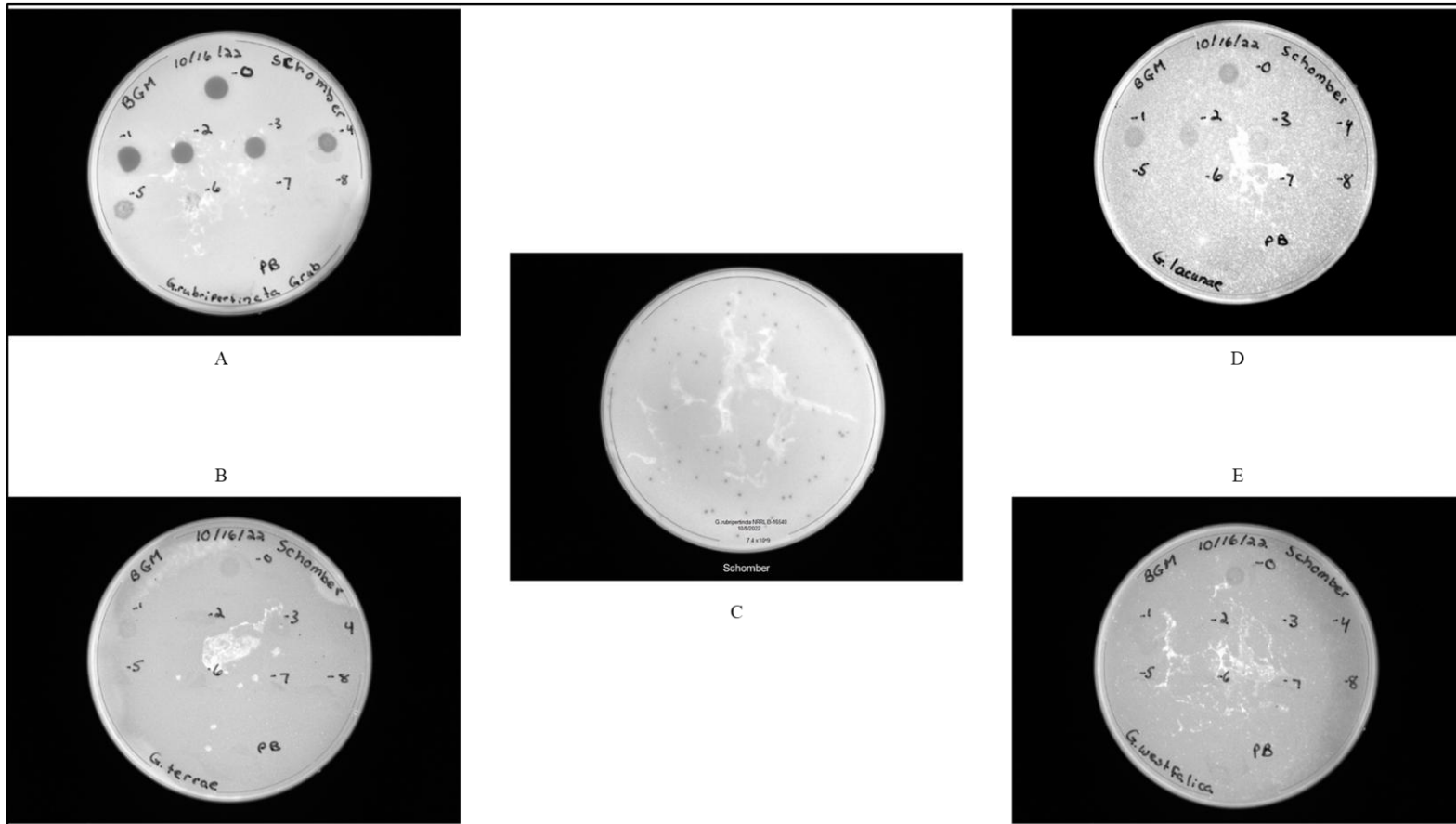
**Figure 18 | Petri dishes of ChisanaKitsune showing plaque formations at different dilution levels with different hosts. A) *G. lacunae* exhibited plaque formation out to  $10^{-4}$  dilution levels. B) *G. terrae* exhibited plaque formation out to  $10^{-1}$  dilution levels. C) A full plate titer of *G. rubripertincta* NRRL B-16540 (isolation host) at  $6.9 \times 10^9$  showing plaque morphology. D) *G. rubripertincta* Grub 38 exhibited plaque formation out to  $10^{-8}$  dilution levels. E) *G. westfalica* exhibited plaque formation at undiluted levels.**



**Figure 19 | Petri dishes of Hanem showing plaque formations at different dilution levels with different hosts.** A) *G. rubripertincta* Grub 38 exhibited plaque formation out to  $10^{-8}$  dilution levels. B) *G. terrae* exhibited plaque formation out to  $10^{-1}$  dilution levels. C) A full plate titer of *G. rubripertincta* NRRL B-16540 (isolation host) at  $1.95 \times 10^{11}$  showing plaque morphology. D) *G. lacunae* exhibited plaque formation out to  $10^{-8}$  dilution levels. E) *G. westfalica* exhibited plaque formation at undiluted levels.



**Figure 20 | Petri dishes of Kabocha showing plaque formations at different dilution levels with different hosts. A) *G. rubripertincta* Grub 38 exhibited plaque formation out to  $10^{-8}$  dilution levels. B) *G. terrae* exhibited plaque formation out to  $10^{-1}$  dilution levels. C) A full plate titer of *G. rubripertincta* NRRL B-16540 (isolation host) at  $1.08 \times 10^{10}$  showing plaque morphology. D) *G. lacunae* exhibited plaque formation at undiluted levels. E) *G. westfalica* did not exhibit plaque formation.**



**Figure 21 | Petri dishes of Schomber showing plaque formations at different dilution levels with different hosts. A) *G. rubripertincta* Grub 38 exhibited plaque formation out to 10<sup>-7</sup> dilution levels. B) *G. terrae* exhibited plaque formation out to 10<sup>-1</sup> dilution levels. C) A full plate titer of *G. rubripertincta* NRRL B-16540 (isolation host) at 1.08 x 10<sup>10</sup> showing plaque morphology. D) *G. lacunae* exhibited plaque formation at out to 10<sup>-2</sup> dilution levels. E) *G. westfalica* exhibited plaque formation at the undiluted level.**

## CHAPTER FOUR | DISCUSSION

SEA-PHAGES is an international consortium designed to provide undergraduate students with meaningful course-related research experience so they may learn the fundamentals and gain valuable skills for future success in the lab. Led by Graham Hatfull and his lab at the University of Pittsburgh in conjunction with HHMI, this project aims to curate bacteriophages that infect *Actinobacteria*. These phages are isolated from environmental samples gathered worldwide, characterized, and then sent to be archived in a library at the University of Pittsburgh for future research and therapies. [30] George Mason University (GMU) is a member of SEA-PHAGES, and over 200 students have successfully gone through the Phage Discovery and Phage Genomic courses.

In 2019, students in the Phage Discovery course at GMU isolated 12 novel bacteriophages from environmental samples. Repeated attempts to extract DNA from four of the 12 isolated phages were unsuccessful, and a pattern emerged upon TEM imaging several of the lysates. Regardless of the morphology, all the *Mycobacterium smegmatis* phages had DNA successfully extracted from their capsids. However, only the *Siphoviridae* *Gordonia* phages had successful DNA extractions. The problematic *Gordonia* phages all had *Myoviridae* morphologies. We hypothesized that the *Gordonia* phages with *Myoviridae* morphologies contained a difference in their capsid, which was responsible for the difficulties experienced during DNA extraction. This suggests that if DNA extractions consistently presented difficulties with these phages, a bias may exist in the sampling data. With a 35 % cutoff in GCS for clustering

and the mosaic nature of their genomes, phages within the same cluster may appear unrelated.

After recognizing the pattern for *Gordonia Myoviridae* phages, several approaches were tried for extracting their DNA. Different DNA extraction kits and heat denaturation of the protein capsids using a thermal cycler were met without success. Performing a PEG precipitation to concentrate the phages before using a DNA extraction kit also proved unsuccessful. It was determined that phage concentration levels did not cause the difficulties experienced. Interestingly, when DNA concentration levels of 40 ng/ $\mu$ L or more were achieved, the results were not visible when imaged on agarose gels. Further investigation revealed that DNA extraction kits use guanidinium hydrochloride or guanidinium thiocyanate to denature proteins. Therefore, another method was required to extract DNA from these problematic phages.

Successful results were achieved using a modified phenol-chloroform protocol. Sequencing of the *Gordonia Myoviridae* phages revealed they belonged to the same cluster of phages, DQ. Information from The Actinobacteriophage Database at PhagesDB.org and Phamerator showed that only three other DQ phages were registered: Chidiebere, Gray, and GMA6. However, viewing the phages on Phamerator revealed that GMA6 was distantly related to the other DQ phages with low GCS. The low number of DQ phages registered online reinforced the question of a sampling bias and raised new questions. Does GMA6 belong to the DQ cluster, or should it be a sub cluster? The DQ cluster is a distinct branch when creating a phylogenetic tree of the *Gordonia rubripertincta* NRRL B-16540 phages. However, GMA6 was isolated with a different

host. Creating a tree of just the DQ phages revealed it to be on an alternate branch from the other phages.

In 2015, Pope, Bowman, et al. proposed that phage diversity is more of a continuum rather than exhibiting discrete groups [13]. Establishing a continuum between GMA6 and the other DQ phages is difficult, at best, with the limited number of phages registered. Recently, two more DQ phages have since been registered; DalanDe and Pakusa. Pakusa shares a high GCS with six DQ phages, and DalanDe has a low GCS with the same six phages, which can be seen in Phamerator. The phylogenetic tree of DQ phages also revealed DalanDe to be more closely related to GMA6. Since GMA6 was isolated with a different host, and DQ phages have different tails, the host range was investigated.

Different hosts were chosen to test the host range based on the clusters infecting *Gordonia rubripertincta* NRRL B-16540. If a cluster was found to infect another host, ChisanaKitsune, Hanem, Kabocha, and Schomber were tested against them. It is noted here that the isolating host for GMA6 was not available. *G. rubripertincta* Grub 38, *G. lacunae*, *G. terrae*, and *G. westfalica* were used in testing the host range. All four DQ phages were able to infect *G. rubripertincta* Grub 38. With *G. lacunae*, *G. terrae*, and *G. westfalica*, lower levels of infectivity were observed. The ability to infect differing hosts plays a vital role in phage applications.

In 2016, Ross, Ward, and Hyman stated that inconsistencies exist in what is defined as a narrow or broad host range [50]. Host range may refer to strains or species. They also propose that the ability of phages to infect multiple hosts may be common.

However, defining the host range of a phage may be more complex due to several factors. First, testing host range occurs in the lab and does not account for all the available hosts in the environment that the phage may encounter. Second, hosts evolve defense mechanisms against phages, and phages evolve to survive so that the host range may be fluid [50]. Finally, Ross, Ward, and Hyman (2016) highlight the importance of using a mixture of phages, with a broad host range, in a cocktail for phage therapy to prevent phage resistance from developing among bacteria and compare it to using a broad-spectrum antibiotic [50].

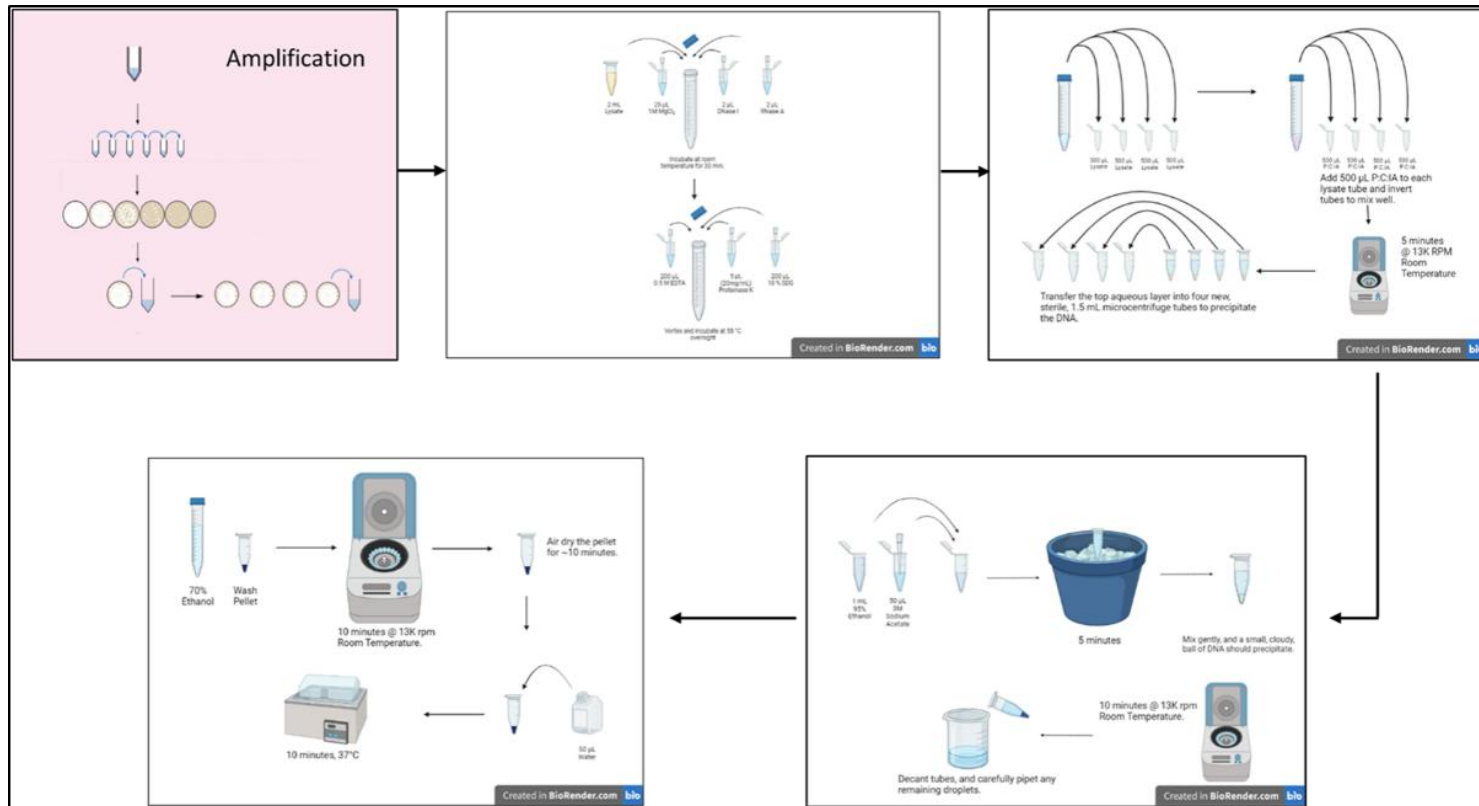
The DQ phages can make vital contributions to several research and therapeutic applications. Capsids resistant to denaturing would provide an excellent base for phage display used in diagnostics, treatments, and vaccines. A unique tail and the ability to infect different hosts make DQ phages a valuable tool in the fight against antibiotic-resistant organisms. With the aid of the CRISPR-Cas system and the mosaic nature of their genomes, phages may be assembled using different tail fibers and expressing different epitopes. However, further investigations are needed for these unique phages.

#### **4.8 | *FUTURE DIRECTIONS***

Several areas remain to be investigated regarding the DQ cluster of phages. Further in-depth studies of the phage capsid proteins are needed to identify the source of their stability, isolate it, and provide a tool to standardize and extend the shelf-life of phage therapeutics. Possible areas to investigate are whether the capsid proteins are methylated or if the sequence for the conserved domain envelope glycoprotein C provides stability and prevents the guanidinium from denaturing the proteins. Also, several genes

that code for proteins without known functions may be accessory or staple proteins. Studies into the tail proteins to understand why DQ cluster phages have a *Myoviridae* morphology, their adsorption process, and how it affects their host range will aid in the fight against antibiotic-resistant organisms. Finally, the ability to extract DNA using the modified phenol-chloroform protocol will provide a powerful tool to decrease the selective characterization process during phage discovery and bridge phylogenetic gaps seen in the continuum of diversity in the DQ cluster of phages.

## APPENDIX | 1



**Figure 22 | Diagram of modified phenol-chloroform protocol.** The work flow presented starts with the amplification process to make a large lysate as outlined in the Phage Discovery manual and through the process of a modified phenol-chloroform protocol for DNA extraction from *G. rubripertincta* NRRL B-16540 phages with Myoviridae morphologies. The diagram was created with BioRender.

## REFERENCES

1. Keen EC. A century of phage research: Bacteriophages and the shaping of Modern Biology. *BioEssays*. 2014;37(1):6–9.
2. Wittebole X, De Roock S, Opal SM. A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens. *Virulence*. 2013;5(1):226–35.
3. Yu P, Mathieu J, Li M, Dai Z, Alvarez PJ. Isolation of polyvalent bacteriophages by sequential multiple-host approaches. *Applied and Environmental Microbiology*. 2016Feb;82(3):808–15.
4. The Actinobacteriophage Database: Home [Internet]. The Actinobacteriophage Database | Home. [cited 2021Nov26]. Available from: <https://phagesdb.org/>
5. Virus taxa infecting vertebrates [Internet]. ICTV. International Committee on Taxonomy of Viruses; 2013 [cited 2021Nov21]. Available from: [https://talk.ictvonline.org/ictv-reports/ictv\\_9th\\_report/introduction/w/9th\\_report\\_intro/88/virus-taxa-infecting-vertebrates](https://talk.ictvonline.org/ictv-reports/ictv_9th_report/introduction/w/9th_report_intro/88/virus-taxa-infecting-vertebrates)
6. Campbell A. The Future of Bacteriophage Biology. *Nature Reviews Genetics*. 2003;4(6):471–7.

7. Citorik RJ, Mimee M, Lu TK. Bacteriophage-based synthetic biology for the Study of Infectious Diseases. *Current Opinion in Microbiology*. 2014Jul3; 19:59–69
8. Dion MB, Oechslin F, Moineau S. Phage diversity, genomics and Phylogeny. *Nature Reviews Microbiology*. 2020;18(3):125–38.
9. Fokine A, Rossmann MG. Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage*. 2014;4(2).
10. Hatfull GF. Bacteriophage genomics [Internet]. *Current opinion in microbiology*. U.S. National Library of Medicine; 2008 [cited 2017Nov21]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2706577/>
11. Newcomer RL, Schrad JR, Gilcrease EB, Casjens SR, Feig M, Teschke CM, et al. The phage L capsid decoration protein has a novel ob-fold and an unusual capsid binding strategy. *eLife*. 2019;8.
12. Nobrega FL, Costa AR, Kluskens LD, Azeredo J. Revisiting phage therapy: New applications for old resources. *Trends in Microbiology*. 2015;23(4):185–91.
13. Pope WH, Mavrich TN, Garlena RA, Guerrero-Bustamante CA, Jacobs-Sera D, Montgomery MT, et al. Bacteriophages of *Gordonia* spp.. display a spectrum of diversity and genetic relationships. *mBio*. 2017;8(4).
14. Sausset R, Petit MA, Gaboriau-Routhiau V, De Paepe M. New insights into intestinal phages. *Mucosal Immunology*. 2020;13(2):205–15.
15. Gibb B, Hyman P, Schneider C. The many applications of engineered bacteriophages—an overview. *Pharmaceuticals*. 2021;14(7):634.

16. Salmond, G. P., Fineran, P. C. (2015). A century of the phage: Past, present and future. *Nature Reviews Microbiology*, 13(12), 777–786.  
<https://doi.org/10.1038/nrmicro3564>
17. Sanz-Gaitero, Marta, Mateo Seoane-Blanco, and Mark J van Raaij. “Structure and Function of Bacteriophages.” *Bacteriophages*. Cham: Springer International Publishing, 2021. 19–91. Web.
18. Brussow, H., Hendrix, R. (2002). *Phage Genomics: Small Is Beautiful*. Cell Press, 108, 13–16.
19. Mavrich TN, Hatfull GF. Bacteriophage evolution differs by host, lifestyle and genome. *Nature Microbiology*. 2017;2(9).
20. Nobrega, F.L et al. “Targeting Mechanisms of Tailed Bacteriophages.” *Nature reviews. Microbiology* 16.12 (2018): 760–773. Web.
21. Pope, W. H., Bowman, C. A., Russell, D. A., Jacobs-Sera, D., Asai, D. J., Cresawn, S. G., Jacobs, W. R., Hendrix, R. W., Lawrence, J. G., & Hatfull, G. F. (2015). Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. *ELife*, 4. <https://doi.org/10.7554/elife.06416>
22. Deng, X., Wang, L., You, X., Dai, P., & Zeng, Y. (2017). Advances in the T7 phage display system (review). *Molecular Medicine Reports*, 714–720.  
<https://doi.org/10.3892/mmr.2017.7994>
23. Gamkrelidze, M., Dąbrowska, K. (2014). T4 bacteriophage as a phage display platform. *Archives of Microbiology*, 196(7), 473–479.  
<https://doi.org/10.1007/s00203-014-0989-8>

24. U.S. National Library of Medicine. Search of: Phage: Active, not recruiting studies - list results. Home - ClinicalTrials.gov. (n.d.). Retrieved April 12, 2023, from <https://clinicaltrials.gov/ct2/results?term=phage&Search>
25. Yue, H., Li, Y., Yang, M., Mao, C. (2021). T7 phage as an emerging nanobiomaterial with genetically tunable target specificity. *Advanced Science*, 9(4), 2103645. <https://doi.org/10.1002/advs.202103645>
26. Cepelewicz J. News: DNA has four bases. some viruses swap... (wired) - behind the headlines - NLM [Internet]. National Center for Biotechnology Information. U.S. National Library of Medicine; 2021 [cited 2021Nov21]. Available from: <https://www.ncbi.nlm.nih.gov/search/research-news/14163/>
27. Pezo V, Jaziri F, Bourguignon P-Y, Louis D, Jacobs-Sera D, Rozenski J, et al. Noncanonical DNA polymerization by aminoadenine-based siphoviruses. *Science*. 2021;372(6541):520–4.
28. Yang, M., Liang, Y., Huang, S., Zhang, J., Wang, J., Chen, H., Ye, Y., Gao, X., Wu, Q., & Tan, Z. (2020). Isolation and characterization of the novel phages vb\_vps\_ba3 and vB\_VpS\_CA8 for lysing vibrio parahaemolyticus. *Frontiers in Microbiology*, 11. <https://doi.org/10.3389/fmicb.2020.00259>
29. Jacobs-Sera D, Marinelli LJ, Bowman C, Broussard GW, Guerrero Bustamante C, Boyle MM, et al. On the nature of Mycobacteriophage Diversity and host preference. *Virology*. 2012Dec20;434(2):187–201.
30. SEA-Phages: Home [Internet]. SEA-PHAGES Home. [cited 2021Nov26]. Available from: <https://seaphages.org/>

31. Welkin PH, Aleks BK, Fisher DJ, Nicholas OH, Savage KA, German BA, et al. Genome Sequence of Gordonia Bacteriophage Lucky 10. *American Society for Microbiology*. 2016;4(3):1–2.
32. Acheson, N. (2011). *Fundamentals of Molecular Virology*. WILEY-BLACKWELL; 2011
33. Youle, M., & Pantéa Leah. (2017). Thinking like a phage: The genius of the viruses that infect bacteria and Archaea. Wholon.
34. Phire. Hatfull Lab. (n.d.). Retrieved April 13, 2023, from <http://www.hatfull.org/courses>
35. Strathdee, S. A., Patterson, T. L., Barker, T. (2020). The perfect predator: A scientist's race to save her husband from a deadly superbug. Hachette Books.
36. Big Flowchart outlined - sea-phages [Internet]. SEA-PHAGES. HHMI; 2013 [cited 2021Nov22]. Available from: [https://seaphages.org/media/docs/Big\\_Flowchart\\_Outlined\\_2020.pdf](https://seaphages.org/media/docs/Big_Flowchart_Outlined_2020.pdf)
37. SDS Wizard® DNA Clean-up System. Promega Corporation. (n.d.). Retrieved April 16, 2023, from <https://www.promega.com/resources/msds/msdss/a7000/a7280/>
38. Smith, K. C., Castro-Nallar, E., Fisher, J. N. B., Breakwell, D. P., Grose, J. H., & Burnett, S. H. (2013). Phage cluster relationships identified through single gene analysis. *BMC Genomics*, 14(1). <https://doi.org/10.1186/1471-2164-14-410>
39. Phage DNA extraction by PhenolChloroform protocol - Texas A&M university. (2018, September 21). Retrieved September 2019, from

- <https://cpt.tamu.edu/wordpress/wp-content/uploads/2018/09/Phage-DNA-Extraction-by-PhenolChloroform-Protocol.pdf>
40. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. *J Mol Biol.* 2018 Jul 20. S0022-2836(17)30587-9.
  41. DNA master 5.22.1 [Internet]. DNA Master 5.22.1 DNA sequence editor and analysis package - DNA Sequence Analysis - Biology Software Net. [cited 2021Nov26]. Available from: <http://en.bio-soft.net/dna/dnamaster.html>
  42. Glimmer. (n.d.). Retrieved April 13, 2023, from <http://ccb.jhu.edu/software/glimmer/index.shtml>
  43. John Besemer, Alexandre Lomsadze and Mark Borodovsky, GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research* (2001) 29, pp 2607-2618
  44. Pecaan. [cited 2021Nov26]. Available from: <https://discover.kbrinsgd.org/login>
  45. Phamerator. [cited 2021Nov26]. Available from: <https://phamerator.org/>
  46. Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN, Alva V. *Curr Protoc Bioinformatics.* 2020 Dec;72(1):e108. doi: 10.1002/cpbi.108.
  47. Sosui: Submit protein sequences. SOSUI: submit protein sequences. (n.d.). Retrieved April 13, 2023, from <https://harrier.nagahama-i-bio.ac.jp/sosui/mobile/>

48. U.S. National Library of Medicine. (n.d.). Blast: Basic local alignment search tool. National Center for Biotechnology Information. Retrieved April 13, 2023, from <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
49. Geneious Prime 2023.x.x <https://www.geneious.com>
50. Ross, A., Ward, S., Hyman, P. (2016). More is better: Selecting for Broad Host Range bacteriophages. *Frontiers in Microbiology*, 7. <https://doi.org/10.3389/fmicb.2016.01352>

## **BIOGRAPHY**

Bonnie Madden received her Associates in Science at the College of DuPage in Illinois. She went on to earn her Bachelor of Science in Biology with a concentration in Microbiology in 2015 from George Mason University as a First-Generation student and graduating Cum Laude. She earned her Master of Science in Biology with a concentration in Microbiology and Infectious Disease from George Mason University in 2023, where she will continue her work as the Microbial and Phage Lab Manager.