The Isolation and Characterization of Novel Bacteriophages

Tobias A. Konieczka ’22 and Michael J. Wolyniak

Department of Biology, Hampden-Sydney College, Hampden-Sydney, VA 23943

Abstract

Bacteriophages are the most ubiquitous organisms found within the biosphere, composed of an estimated 1031 particles. The evolutionary “arms-race” between bacteriophages and bacteria have set the stage for life as we know it today. Because of the sheer volume of phage particles, it is practically impossible to identify every unique phage, however the combined power of laboratory techniques and computer prediction software has made the identification of a few of these phages possible. Under the guidance of the SEA-PHAGES isolation guide, solid and liquid samples from three locations across the Hampden-Sydney campus were tested for phage presence. The solid sample obtained from Lake Chalgrove initially showed promise of phage presence, but was ultimately unproven. The full annotation of mycobacterium phage LESTYG was reformatted into a DNAMasterfile from a PECAAN file and sent to SEA-PHAGES for publication. Using PECAAN, peer review of a previous annotation of mycobacterium phage Samcro found that the genome was submitted in reverse order, disproving its existence. The genome was then imported into PECAAN in the correct order and annotated under the name Dawnguard. The annotation of Dawnguard still needs to undergo reformatting and publication.

Background Information

Bacteriophages, more simply referred to as phages, are nature’s perfect weapon. In the evolutionary history of the world, they—along with bacteria—have dominated the battlefield and have helped pave the way for human evolution. Phages are the most ubiquitous organism on the planet earth, composed of roughly 10^31 particles. To put this into perspective, there are an estimated one trillion phage particles for every grain of sand found on Earth (Keen, 2015). Numbers aside, Bacteriophages operate with acute precision during the infection of host organisms. As the prefix “bacterio-” suggests, phages exclusively target bacteria. All bacteria can fall prey to phages, however not all phages can infect all bacteria. Due to variations in genetic makeup, all species of bacteriophage target unique species of bacteria. A phage that naturally infects mycobacterium hosts will not infect staphylococcus hosts, and vice versa.

First contacts.

Although phages preceded the existence of Homo sapien by billions of years, they were not formally discovered until the early 20th century, during the development of the smallpox vaccine. The synthesis of the vaccine was dependent upon the successful culture of vaccinia virus, which was unfortunately prone to Staphylococcus contamination. In 1915, English Physician Frederick Twort discovered the presence of “plaques” on the agar plates, which he determined to be zones of dead bacteria. In his paper entitled An Investigation on the Nature of Ultra-Microscopic Viruses, Twort posits the culprit of this phenomenon as “either a stage in its life-history which will not grow on ordinary media but stimulates fresh cultures of the micrococcus to pass into the same stage, or an enzyme secreted by the micrococcus which leads to its own destruction and the production of more enzyme” (Twort). The arrival of World War I halted Twort’s efforts to pursue the matter any further, due to the reallocation of funding towards the war effort.

With the war came the outbreak of Shigella dysenteriae, the bacterium responsible for dysentery infection in troops. In 1915, French-Canadian microbiologist Felix d’Herelle observed similar zones of clearing in cultures containing the bacteria. d’Herelle published his findings to the French Academy of Science two years later. Unlike Twort, d’Herelle concluded that the plaques were caused by the presence of an virus-like organism that he called “bacteriophages.” He theorized that phages could be used to treat dysentery infections, although this theory was left untested until 1919, when a young boy suffering from dysentery was cured of the infection after drinking a phage solution under d’Herelle’s guidance (University of California San Diego).

Phage Biology.

While characterized as organisms in primary literature, bacteriophages are non-living and function as bacteria-specific viruses. Because of this, there is no risk of bacteriophage infection in human cells, hence their use in treating bacterial infections. **Figure 1** gives an overview of phage infection. The process of phage infection can be broken down into two distinct cycles. In the Lytic Cycle, the Phage attaches to the host and inserts its genetic material through the use of

lysozymes. The genetic material flows to the nucleus, where it integrates into the host's genome. From here, copies of the phage- “progeny”- are rapidly produced by the host. Phage progeny then breaks through the

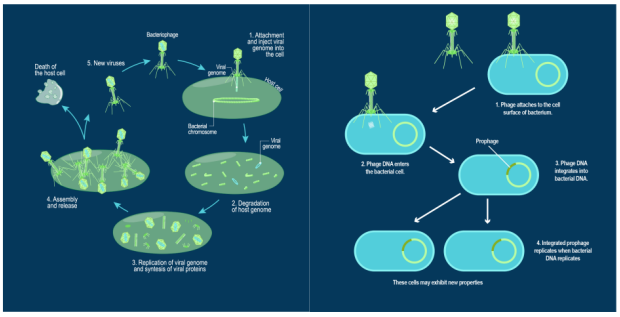

Figure 1: A visual overview of the Lytic (left) and Lysogenic (right) cycles of bacteriophage infection.

(Courtesy TechnologyNetworks)

host cell, resulting in the host’s death and the spread of virulent phage particles. The Lysogenic cycle begins in very much the same way as in the Lytic cycle, where the phage attaches to the host and transmits its genetic material. From here, the genetic material interfaces with the host’s genome and remains dormant. Environmental effects, including host damage and exposure to UV radiation, result in the production of progeny, transitioning into the Lytic Cycle. The CRISPR system of gene editing was developed after the discovery of bacterial immune systems to fight off phage infection.

Sea Phages.

The Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science program was founded at the University of Pittsburgh in 2008. 199 institutions based in the United States are allied with the program. Hampden-Sydney College is one of 13 schools allied within the Commonwealth of Virginia. The function of SEA-PHAGES is to offer collegiate students the experience of individual scientific discovery and the advancement of the identification of phages across the country.

Materials and Methods

All phage isolation techniques were taken from the SEA-PHAGES Discovery Guide, which is completely free to the public. The use of an instructor version allowed access to the toolbox, a section devoted to recipes for various laboratory techniques. The Discovery Guide also provided the proper procedures for carrying out all steps in the phage isolation protocol.

Bacterial Hosts.

Mycobacterium neoaurum and *Rhodococcus erythropolis* were selected based on multiple mechanistic similarities to *Mycobacterium tuberculosis*, namely the use of mycolic acid as a protective matrix. *M. tuberculosis* infection causes a multitude of symptoms within an infected patient, including coughing, chest pain, fatigue, fever, chills, and death. According to an article entitled “The U.S. Government and Global Tuberculosis Efforts" published by the *Kaiser Family Foundation*, approximately ⅓ of the world’s population is infected with *M. tuberculosis*, annually. Of these infections, approximately 9 million result in illness, while 1.8 million deaths occur (KFF, 2021). As treatments have been developed, known strains of *M. tuberculosis* have continued to evolve into Multidrug-Resistant (MDR) strains that can only be treated with top-of-the-line medications, as well as Extensively-Drug-Resistant (XDR) strains that are practically unkillable. Due to this, the use of phage therapy as a tuberculosis treatment has garnered increased attention, in the same vein as d'Herelle’s dysentery treatment.

Freeze dried samples of the two bacterial hosts were purchased from the American Type Culture Collection (ATCC). These samples were immediately stored at -80℃ in small tubes. Two initial overnight cultures were created by taking 100 uL of each freeze dried host and combining this with 5 mL of Luria Broth and stored in a 4℃ walk in fridge. To distinguish each host throughout the isolation process, yellow tape was used to identify *M. neoaurum* and blue tape was used to identify *R. erythropolis*. Subsequent overnight cultures were created through the addition of 0.5 mL of host culture supplemented into falcon tubes with 4.5 mL of either LB, Peptone Yeast Calcium Agar (PYCa) media, or Middlebrook 7H9 broth (MB). The use of the different media in each culture proved arbitrary and were used based on availability. Plated cultures of each host bacterial were grown in different temperatures. *R. erythropolis* cultures were grown in a 29 ℃ oven for 3 days, on average. This was based on a SEA-PHAGES fact sheet for the optimal growing conditions of the host. Conversely, *M. neoaurum* cultures were grown in a 37 ℃ oven for 2 days on average. These conditions were based on the protocols for the growth of *M. smegmatis*, due the genetic similarities of the two hosts.

Sample Collection.

The protocols used for the collection of environmental samples were taken from “Chapter 5.1: Collecting Environmental Samples” in the SEA-PHAGES Discovery Guide. Solid and liquid samples were obtained from 3 different locations across the campus of Hampden-Sydney College. The samples were taken from Mayes Lake, Lake Chalgrove, and Tadpole Hole (Table 1). Solid samples from each

location were collected near the banks and stored in 50mL conical tubes. Liquid samples were collected from water within arm's length from the banks and stored in 50mL conical tubes.

Location	Sample Name	Sample Type	Description (if applicable)	Coordinates
Lake Chalgrove	Chal	Solid		37.242, -78.463
		Liquid		37.243, -78.463
Mayes Lake	KFH	Solid		37.243, -78.454
		Liquid		37.243, -78.454
Tadpole Hole	Tad	Solid		37.245, -78.452
		Liquid		37.245, -78.452

Table 1: Information on environmental samples used during phage isolation.

Isolation Overview.

The protocols used for the isolation of phages from the previously described environmental samples were taken from "Chapter 5: Isolation" of the SEA-PHAGES Discovery Guide. The isolation protocols were done in two sections. The procedures for direct isolations were pulled from Protocols 5.2 and 5.3. The procedures for enriched isolations were pulled from Protocols 5.5 and 5.6. PYCa plates were used during the Plaque Assay and Spot Test procedures. Two batches of molten top agar were also synthesized (MB 7H9 top agar and PYCa top agar) and stored in a 55 ℃ bath. All Erlenmeyer flasks and glass bottles were autoclaved prior to use and sealed with aluminum foil. All samples were stored in a 4 ℃ walk-in fridge.

Direct Isolation.

The purpose of direct isolation is to provide a "snapshot" of all phages potentially contained within an environmental sample at the time of collection. The expectation of a direct isolation is a collection of diverse phages with the ability to infect multiple hosts. The average turn-around time of a direct isolation is 2 days. 15 mL conical tubes were filled to approximately ½ to ⅓ full with environmental samples, submerged beneath 2-3 mL of the appropriate media (MB 7H9 for Mycobacterium and PYCa for Rhodococcus), and then inverted. The tubes were then incubated via shaker for 2 hours at 250 rpm. After removal from the shaker, the tubes were centrifuged at 2,000 x g for at least ten minutes, allowing the formation of a pellet. The supernatant of each tube was then filtered into labeled microcentrifuge tubes via a 0.2 um syringe filter. The use of filtration in this step removed any bacteria from the sample.

After the direct isolation concluded, a Plaque Assay was performed for all samples, based on Protocol 5.3. 250 uL of each host culture were injected into microcentrifuge tubes and labeled accordingly. 500 uL of phage samples were then injected into the corresponding tubes containing host bacteria. A negative control tube was created by adding 10 uL of phage buffer to the bacterial cultures. 3 mL of host preferred molten top agar were transferred into the microcentrifuge, aspirated back into the pipette, and plated. This procedure was completed for each sample. Each plate was left to sit for a variable amount of time- between 20 and 30 minutes- to allow for top agar solidification. The plates were inverted and incubated at the appropriate temperatures for at least 48 hours. After incubation, the plates were observed for the presence of
plaques.

Enriched Advanced Lab.

The purpose of enriched isolation is to amplify the concentration of all phage particles in a given environmental sample by introducing them to a nutrient rich environment, thereby promoting phage replication. The average turn-around time of an enriched isolation is one week. 50 mL conical tubes were filled to the 15 mL mark with environmental samples and labelled appropriately. The appropriate liquid media was then transferred into the 50 mL conical tube to the 35 mL mark and shaken at 250 rpm for 2 hours. After shaking, the tubes were centrifuged at 2,000 x g for 10 minutes. The supernatant of each tube was then filtered through a 0.2 um syringe filter into a properly labelled Erlenmeyer Flask. 0.5 mL of the corresponding bacterial host were transferred into the flask and shaken at ~220 rpm for 5 days. For liquid environmental samples, the same procedure was used, albeit with 10X concentration liquid media.

After incubation, 1.4 mL of each sample were pipetted into properly labelled microcentrifuge tubes. This process was repeated once for each sample. The tubes were then centrifuged at 14,000 rpm for 1 minute. The supernatants of each tube were transferred into new appropriately labelled microcentrifuge tubes (Supernatants that were not clear were filtered with a 0.2 um syringe filter before addition to the new tubes). The flasks containing the culture were stored in the walk-in fridge.

After the Enriched Isolation procedure was complete, a Spot Test was performed with each sample. The procedure for the Spot Test was obtained from Protocol 5.6. 250 uL of the appropriate host bacteria were transferred into a 15 mL conical tube. A grid pattern was drawn on the bottom of each plate to establish zones. As in the Plaque Assay protocol, 3 mL of the appropriate molten top agar were pipetted into the tube, aspirated, and plated. Top Agar was left to sit for 20 minutes to solidify. After solidification, 10 uL of phage sample was pipetted into 8 of the 9 grid squares (All spots came from the same sample for that plate). 10 uL of phage buffer was added to the final grid

square to represent the negative control. The plates were given variable timebetween 30 minutes and 3 hours- to absorb into the appropriate square. After this, the plates were inverted and incubated at the proper temperature for at least 24 hours. After incubation, the plates were observed for the presence of plaques.

Annotation Software.

DNAMaster and PECAAN served as the primary annotation software during the bioinformatics component of the research. Annotation is the process of assigning functions to gene candidates found within the phages genome, based on statistical models. The Genemark and Glimmer algorithms were utilized by both annotation softwares to call the locations of genes. Genemark predictions are based on the location of start and stop codons within the genome, as well as comparison to similar organisms. Glimmer predictions are based on coding-potential and open reading frames (orfs). This data is then compiled into a numerical value that represents the start site of best prediction. The data produced by the two algorithms is compiled into a list of gene candidates containing information about start and stop sites, gaps, Z-Scores, Final Scores, and whether the gene contains the longest open reading frame. To settle disagreements between the two algorithms, the information presented in the list of gene candidates is observed. The gap, or distance between the stop site of one gene and the start site of another is very useful for determining the proper gene candidates. A slightly negative value indicates a small overlap, which is common. Large gaps between genes, while not unheard of, discredit the candidate. The Z-Score is interpreted by finding the number closest to 2.000. The Final Score is interpreted by finding the highest negative number. The Longest Open Reading Frame is assigned to one candidate and implies that all bases are used in that gene.

Both annotation softwares also obtain information from the PhagesDB Basic Local Alignment and Search Tool (BLAST), HHPRED, and NCBI BLAST tools. PhagesDB BLAST contains the genetic information of published phages and includes information about observed functions, sequence lengths, clusters, phams, and Clusters. Phams are generated through the use of the bioinformatic tool Phamerator, a program that "assorts protein-coding genes into families of the related sequences using pairwise comparisons to generate a database of gene relationships" (Cresawn et.al.). The inclusion of pham information in PECAAN allows for the creation of Pham Maps. Pham Maps serve as a visual representation of the alignment of genes in comparison between two phages, as seen in **Figure 2**.

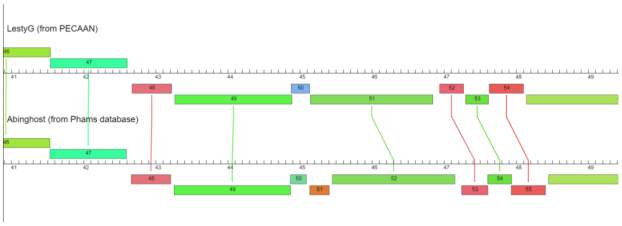

Figure 2: Pham map generated comparing phages LestyG and Abinghost

Similarly, clusters identify phages based on the species of bacteria phages infect and the nucleotide sequence of those phages. Good evidence to add to an annotation will take the form of similar or identical protein numbers, function, and sequence length. Additionally, hits through PhagesDB BLAST list scores and e-values. A high score and an e-value closest to zero indicate a good match.

HHPRED is a 3D modeling prediction software that takes the primary structure of the investigated gene and predicts its tertiary structure. This information is then compared to proteins with similar tertiary structures as a means to draw similarities and propose a function. Hits from HHPRED also include information about a function's probability, % coverage, and e-value. Due to the nature of the information and the variance in statistical importance, HHPRED hits were seldom added as evidence.

NCBI BLAST is used to find regions of similarity between a nucleotide sequence and published sequences in other databases. Hits from NCBI BLAST were frequently added as evidence, due to the inclusion of information regarding % identity, % alignment, % coverage, gaps, and e-values.

Discussion

Findings.

Of the plates documented, only one plate showed promise of phage activity. The solid sample from Lake Chalgrove that underwent enriched culture contained what appeared to be a halo, as seen in **Figure 3**. The presence of this could've been an indication of phage activity, in the form of a light plaque. Repeated experiments did not produce any similar results, though.

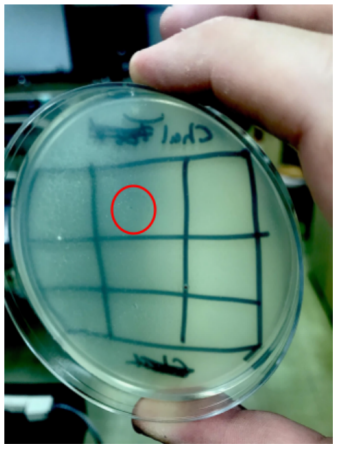

Figure 3: Halo (Highlighted in red) found on an enriched isolation plate of the chalgrove solid sample. *Courtesy Dr. Michael Wolyniak*

Contamination.

In the final weeks of the research, rampant contamination plagued the experiment. To alleviate this, new batches of PYCa plates, molten top agar of MB 7H9 and PYC, and Phage Buffer were created. A sample of molten MB 7H9 top agar was tested for contamination. 3 mL of top agar was plated in a fresh PYCa plate and incubated for 2 days at 37 ℃ and 29 ℃. Cultures of bacteria were found to have grown in these plates, indicating the top agar was a major source of contamination.

Phage LestyG (Subcluster B3).

Mycobacterium phage LestyG (Figure 4) was discovered by Caleb Manu, a Hampden-Sydney student, in the summer of 2020. Specifically, it was found in the Aztec grass behind the Rhetoric Studio. The host bacterium used during the isolation techniques was Mycobacterium Smegmatis mc2155. The annotation of the 68,198 base pairs was conducted through the use of the DNAMaster and PECAAN softwares, by students taking the Genomics and Bioinformatics course in the Fall of 2020. 100 genes were included in the final draft of the phage. The majority of these genes have an unknown function (NKF), but a few functions were identified, including: reductase, nuclease, terminase, resolvase, capsid proteins, tail proteins and adaptors, helicase, polymerase, and recombinase. After annotation of LestyG concluded, the completed file was imported into DNAMaster and reformatted for publication to SEA-PHAGES on 6/11/2021, pending further review.

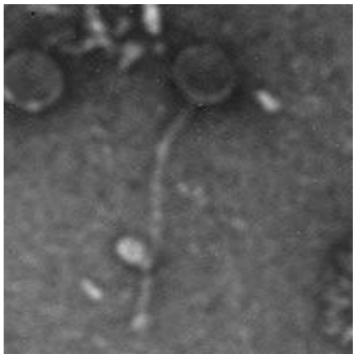

Figure 4: *Mycobacterium* phage LestyG . *Courtesy University of Mary Washington*

Phage Samcro (Subcluster F1).

Mycobacterium phage Samcro was discovered in 2019 at James Madison University and belongs to the F1 cluster. Dr. Louise Temple of JMU sent the sequenced genome to Hampden-Sydney College for annotation soon after. It was fully annotated by Aidan Sloan in the summer of 2020 using DNAMaster. The host bacterium used during the isolation techniques was Mycobacterium smegmatis mc2155. The genome of phage Samcro was found to be approximately 56,573 base pairs long with 62.8% being CG. In-house peer review of this phage began after the unsuccessful attempts to isolate phage from the collected environmental samples. From the start of the review a few problems became apparent. While the genome was annotated in rising numerical order, the protein numbers included in NCBI Blast counted in decreasing numerical order. For example, gene 1 of Samcro corresponded to gene 97 of another phage contained within the same subcluster and pham. The pham maps generated in PECAAN between Samcro and another host also presented an issue, in that no sites lined up. On the annotation of the 83rd gene, it was finally determined that the sequence of the phage had been uploaded in reverse order, hence the unalignment and discrepancy in protein numbers. The nucleotide sequence was reversed and re uploaded into PECAAN under the working name “Tobias.”

Phage Dawnguard (Subcluster F1).

Dawnguard is the finished name of the corrected genome formerly known as “Samcro.” The annotation of this phage was completed entirely in PECAAN. Annotation information from Samcro was used to help gather evidence to speed up the process. The only major change from the two versions was the removal of gene 45 from inclusion in the genome. The files will need to be imported into DNAMaster and formatted for publication to SEA-PHAGES. An entry for Dawnguard was accepted and published in PhagesDB.

Acknowledgements

Dr. Michael Wolyniak for providing limitless information, help/clarification, and access to the necessary materials/software. Mrs. Beverly Hines for providing chemical compounds not found on the bio floor. Nate Jones and Taisho Miller for quick traversal around campus during environmental sample collection. The Office of Undergraduate Research for providing the funding for this research.

REFERENCES

Cresawn, S. G., Bogel, M., Day, N., Jacobs-Sera, D., Hendrix, R. W., & Hatfull, G. F. (2011). Phamerator: A bioinformatic tool for COMPARATIVE BACTERIOPHAGE GENOMICS. BMC Bioinformatics, 12(1). https://doi.org/10.1186/1471-2105-12-395

Manu, C. (2020). Mycobacterium phage LestyG. https://phagesdb.org/phages/LestyG/. Published: Mar 02, 2021. (2021, August 9). The U.S. government and GLOBAL TUBERCULOSIS EFFORTS. KFF. https://www.kff.org/global-health-policy/fact-sheet/the-u-s-government-and-global-tuberculosis-efforts/.

SEA-PHAGES. (n.d.). SEA-PHAGES Discovery Guide. Home. https://seaphagesphagediscoveryguide.helpdocsonline.com/home.

Steward, K. (2018, August 28). Lytic vs Lysogenic – understanding BACTERIOPHAGE life cycles. Technology Networks. https://www.technologynetworks.com/immunology/articles/lytic-vs-lysogenic-understanding-bacteriophage-life-cycles-308094.

Thomas, G. H. (2014, May 29). Frederick William Twort: Not just bacteriophage. Microbiology Society. https://microbiologysociety.org/publication/past-issues/world-war-i/article/frederick-william-twort-not-just-bacteriophage.html.

University of California San Diego. (n.d.). Phage 101. UC Health - UC San Diego. https://health.ucsd.edu/news/topics/phage-therapy/pages/phage-101.aspx.