Proceedings of the National Conference
On Undergraduate Research (NCUR) 2020
Montana State University, Bozeman MT
March 26-28, 2020

Genomic Characterization of the Newly Isolated Bacteriophage Vanisius

Gregory Markov, Tessa Cote, and Jenna R. St. Pierre
STEM Division
Columbia State Community College
1665 Hampshire Pike
Columbia, TN 38401

Faculty Advisor: Dr. Elvira Eivazova

Abstract

This project highlights the genome annotation of the new bacteriophage Vanisius. This bacteriophage was isolated from soil using the bacterial host *Microbacterium foliorum*. Purified phage DNA samples were sequenced at the Pittsburgh Bacteriophage Institute and annotated at CSCC. The bacteriophage Vanisius belongs to the small EE cluster exhibiting the lytic life cycle and is a member of the Siphoviridae phage family. Vanisius' genome was analyzed based on its sequence's start and stop codons, cross-references in Phamerator, and between NCBI and HHPred toolkits. Vanisius was shown to have a tightly packed genome, with the brevity of its sequence corresponding to its small size, at 152 nm. Among EE cluster members, the average genome size was 17,407 bp, with Vanisius' DNA sequence being 17453 bp long with 25 total genes. The annotation of Vanisius showed no orphams, with all genes matched in the phage genome database. Our future goal is to further evaluate phage genome communities isolated from various habitats. This study contributes to the understanding of genomic diversity of bacteriophages, essential for developing phage taxonomy.

Keywords: bacteriophage, genomic annotation, comparative analysis

1. Introduction

Viruses that infect bacteria, known as bacteriophages, are the most abundant organisms on the planet. The phage population is immense, at 10^{31} particles, with only around 3,000 phages having been characterized genomically. [1] Phages have been of interest to scientists because they can effectively destroy bacteria, and therefore, be used for phage therapy. [2] Cataloguing novel phages grows the national database, while also allowing for research into the similarities and differences of bacteriophages within recognized groups or clusters that share genetic similarities. Comparative studies give insight into structural genes, as well as genes governing phage mechanisms. [1] [3] The purpose of this research was to successfully isolate a bacteriophage from soil samples and submit it to the National Registry to be named, catalogued and analyzed. This isolation yielded the siphoviridae, or long non-contractile tail, phage Vanisius, whose purified DNA sample was sequenced at the Pittsburgh Bacteriophage Institute. The host used for the bacteriophage was *Microbacterium foliorum*, a soil-dwelling bacterium. This paper will compare the genomic characteristics of the newly annotated Vanisius sequence with three other bacteriophages of the EE cluster in the national database.

2. Methodology

Soil samples were collected from damp and wet areas around the Columbia State Community College campus. The bacteriophage Vanisius was isolated from 35.918 ˚N, 86.8˚W, near an HVAC unit. Alongside the collected soil samples, the bacteria were pelleted in microcentrifuge tubes using the enriched isolation method, which involved the

amplification of the number of phages present before identification on solid media. Supernatant was separated. A spot test was then needed to determine if bacteriophages were present. The chosen host bacterium, *Microbacterium foliorum*, was added to the top agar and the mixture was poured onto PYCa agar plates. PYCa agar was made from 15.0g agar at 1% concentration, 1g yeast extract at 1g/L, 15g peptone at 15g/L, 4.5mL 1 M $CaCl_2$ at 4.5mM, 2.5mL 40% dextrose at 0.1% concentration, 990mL ddH_2O and 1.0mL CHX stock at 10 µg/mL. After the plates solidified, 10 µl of each phage sample was spotted onto designated sections of the plate. The plates were incubated at 28˚C for one week. Positive spot test samples were used to create a plaque assay. Using stock (filtered enriched isolation), serial dilutions were made up to 10^{-5} and then the host bacteria were inoculated with the phage sample. The plates were incubated at 28˚C for 24-48 hours to allow for phage infection and bacterial growth. An isolated plaque was then picked from the serial dilution plaque assays.

Using the picked plaques, liquid phage samples of decreasing concentrations were prepared using ten-fold serial dilutions with phage buffer. Phage buffer was prepared from 10mL 1 M Tris stock (pH 7.5) at 10mM, 10mL 1 M $MgSO_4$ stock at 10mM, 4g NaCl at 68mM, 1L ddH_2O and 10mL 100mM $CaCl_2$ stock at 1mM. Plaque assays for purification were created in order to ensure that there was a clonal population of phages. The plaque morphologies for Vanisius were consistent; they were medium-sized and clear, indicative of a lytic phage (Figure 1). The webbed plates, plates with a very high density of plaques made with a lysate of known titer, were flooded with phage buffer, and the plate lysates were collected and filtered through a 0.22 µm filter. A spot titer, a titer used to determine the concentration of phage particles using a spot test, and a full plate titer, a titer used to determine the concentration of phage particles using plaque assay, were made for each sample. It was confirmed that there was a ten-fold decrease in the number of plaques for each plate.

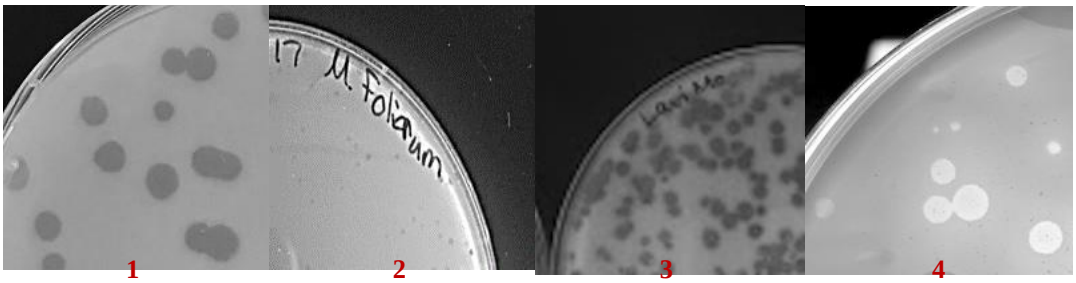

Figure 1. Plaque pictures of phages Vanisius (1), with comparison to Scamander (2), Lavimo (3), and Bonaevitae (4). All four phages belong to the EE cluster.

Phage particles were bracketed by making one plate with fewer phage particles and one plate with more phage particles (three plates total per each sample) to ensure a webbed plate. Duplicates were created for each sample to obtain a greater volume of lysate. The HHMI plaque assay was used, but only three dilutions were needed to make the webbed plates. For samples with titers less than $5x\ 10^9$, the amplification steps were repeated. Adequate phage samples were mounted onto a grid, stained and observed under a transmission electron microscope (Figure 2).

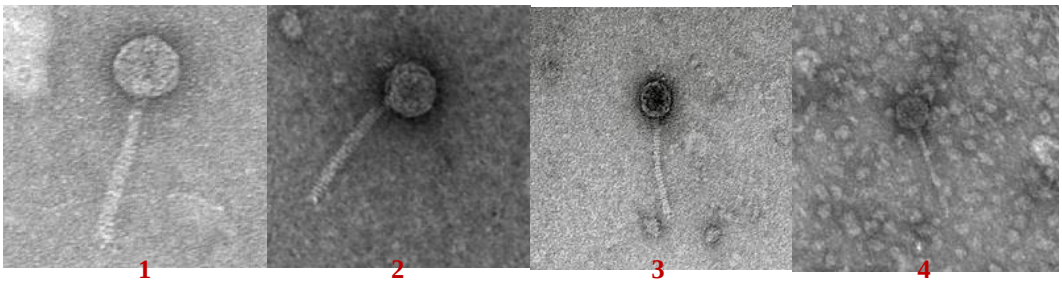

Figure 2. TEM (Transmission Electron Microscope) micrographs of phages Vanisius (1), with comparison to Scamander (2), Lavimo (3) and Bonaevitae (4). Phages 1-4 display the non-contractile tail of the siphoviridae family, a shared characteristic of all EE cluster members.

In order to extract phage DNA, the protein capsid was denatured and DNA was precipitated and purified, as described in the HHMI guide [4] The Vanisius DNA sample concentration (352 ng/ul) was measured with a Nanodrop Spectrophotometer at A260/280:1.95. The purified DNA was sent to be sequenced at the Pittsburgh Bacteriophage Institute.

The nucleotide sequence file received from the PBI was imported into PECAAN, which is an umbrella program created for genomic annotation and includes Glimmer/Starterator, GeneMark, a PhagesDB bank, HHPred, NCBI, Aragorn and Phamerator plugins. The first step in annotation was to determine the starts and stops of all potential genes in the sequence. GeneMark was used to create a map; this map showed the coding potential of each potential sequence, and along with Starterator and the already existing Phagesdb database, the starts and stops for each coding sequence was chosen. It was ensured that each potential sequence began with ATG, TTG or GTG codons, that frameshifts were noted and that each genome had only one portal protein. With this complete, functions could be assigned. Phamerator was used to find preliminary matches with other closely matching sequences, allowing the prediction of potential genes as part of a visible, mapped context. Finally, HHPred and NCBI were used to definitively predict the functions of these DNA sequences, using the best matching hits that had a high probability and e-value, as listed evidence, for eventual submission. The data was exported from PECAAN into DNA Master, another annotation program, to proofread the two sequences and then generate files for official submission. Phamerator was used to analytically compare phages within the EE cluster.

3. Data

Four bacteriophages, Vanisius, BonaeVitae, LaviMo and Scamander, all from the relatively new EE cluster and available in the national database, were selected for comparative analysis. Reasons for their selection include: Phagesdb only contains the plaque images for thirteen bacteriophages from the EE cluster, work was done on Vanisius, and the bacteriophages BonaeVitae, LaviMo and Scamander frequently appeared in the annotation process. At the time of the annotation in 2019, BonaeVitae, LaviMo, Scamander, Rhysand, and Burtonthepup had been submitted and were approved, while other bacteriophage sequences were considered drafts. The sequences for bacteriophages Rhysand and Burtonthepup did not have their corresponding plaque photographs available.

Phamerator is a key tool used in the analysis and comparison of bacteriophages. Phamerator displays a set of genes along a marked sequence, with each gene represented by a colored box. The gene boxes are numbered from 1 to the last gene of the sequence. These boxes can also differ in length, which is based upon the gene's selected starts and stops, imported from GeneMark. The program can also display any number of sequences. When more than one sequence is selected, phamerator generates colored boxes that are either the same color for matching genes or different colors for genes that diverge in function. Phamerator also displays lines drawn from a gene on one sequence to a gene on another sequence. This way, even if the gene number is different, gene functions can still be matched from one sequence to another, e.g. from a green box representing gene 4 to a green box representing gene 7.

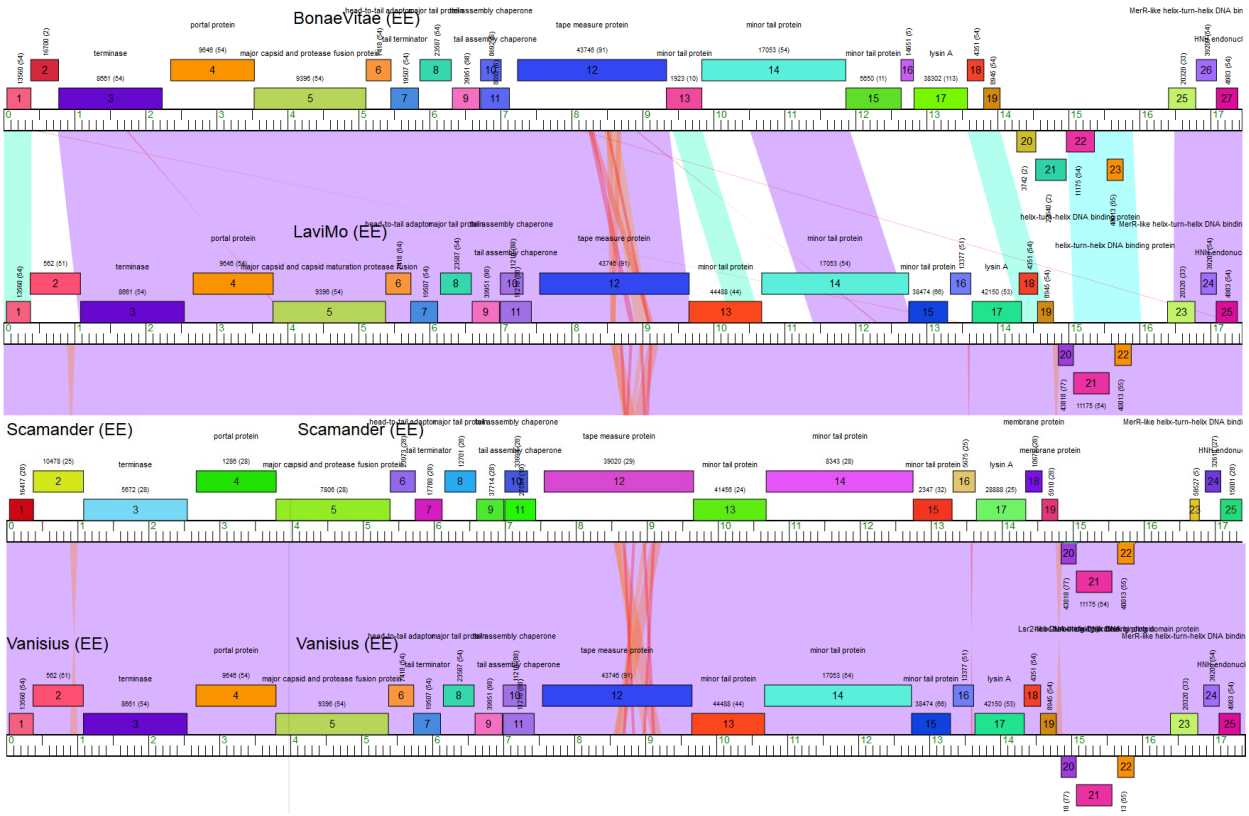

Figure 3. Phamerator map comparing four bacteriophage sequences in the EE cluster: BonaVitae, LaviMo, Scamander and Vanisius.

In small genomic sequences, such as those of bacteriophages of the EE cluster, genetic conservation is readily apparent. Comparing genes 1-22, it can be seen that many of the genes are identical in function, with some points of difference.

4. Discussion

Cluster EE bacteriophages infect bacteria of the genus Microbacterium and are lytic bacteriophages. [1] They are characteristically short, with an average length of 17,360 base pairs, 27 genes and a GC% of 68.6 in a cluster consisting of 54 members (Table 1). Prior published research with bacteriophages has shown that, “only a limited number of structural solutions are used in order to produce a functional virion.” [3] This is the concept of genetic conservation, and a way of organizing similar bacteriophages within a cluster of other similar bacteriophages. The genomes of Cluster EE are relatively small, therefore displaying a high degree of genetic conservation to produce a functional phage, in comparison with much larger and longer genomes, such as those of the A cluster, which consist of an average of 100 genes.

Table 1. Comparison of the cluster EE phage genome sizes in base pairs (bp)

	Phage Genome Size (bp)
Vanisius	17,453
Scamander	17,452
Lavimo	17,453
Bonaevitae	17,451

Genetic conservation is most common towards the beginning of the sequence (Figure 3). This is the place in the sequence with the most structural genes, without which the phage would be nonfunctional. An example of a conserved structural gene is the presence of tail assembly chaperones across each sequence in the EE cluster, which are present in the same location, always following the tail terminator gene but always coming before the gene coding for the tape measure protein (Figure 3). Cluster EE members are all phages with tails, the siphoviridae family, and in comparing BonaeVitae, LaviMo, Scamander and Vanisius, each of these four bacteriophage sequences have genes number 9-11 coding for tail assembly chaperones. In terms of conservation, apart from this presence of tail assembly chaperones, the presence of a scheduled frameshift on exactly gene 10 across each of the four bacteriophages in question can be noted as well.

In spite of this genetic conservation, however, there are still 54 individually unique members of the EE cluster of bacteriophages, and therefore differences among them at distinguishable points in their genomic sequence. Vanisius, newly discovered in 2018, is itself a novel bacteriophage. As seen from Figure 3, Vanisius shares 23 identical genes with Scamander, the closest genetic member, yet with a difference of two genes towards the end of the sequence. The study of the dissimilarity between Vanisius and Scamander, in particular the difference on gene 23, is important. Gene 23 does not fall within the beginning of the sequence and the other structural genes, such as those determining phage tail characteristics. Both the genomic sequences of Vanisius and Scamander contain gene 22 which codes for a helix-turn-helix binding domain and contain gene 24, which codes for a MerR-like helix-turn-helix binding domain; these are followed by an HNH endonuclease coding gene on gene 25 in each phage genome. This is the region of the phage genome corresponding to DNA identification for virus assembly, and therefore a region corresponding to the phage's unique mechanism for infection. DNA-binding proteins, specifically the helix-turn-helix binding domain, have been shown to be "motifs made up of a first alpha-helix, a connecting turn and a second helix, which specifically interacts with the DNA and is known as the recognition helix." [7] Within the sequence, protein families associated with helix-turn-helixes are areas of the greatest genomic difference and logically, the greatest evolutionary divergence. In considering Vanisius and Scamander, it is seen that both genes 23 have experienced different mutations rendering the function unknown. The unknown gene's function cannot be determined through attempted matching against other bacteriophages within the current database. This is in contrast to those genes involved with structure rather than mechanism, which can all be largely be identified and can be found in dozens of other bacteriophages.

In spite of genetic conservation, new phages can and do arise through the evolution of those genes in the sequence responsible for mechanisms much more frequently than those responsible for structure [5]. This is simply due to the fact that structural genes are required for the most basic functionalities, and a reason why structural, and not mechanism-related, genes are used for phage classification – terminase subunits and portal proteins, as an example, are common proteins looked for in this type of analysis. [5] Phamerator analysis in the EE cluster shows genetic conservation of so-called structural genes, while those genes governing specific mechanisms of how the phage behaves, such as DNA-binding genes, show a high degree of evolutionary or mutation-related change.

5. Conclusion

Future research goals include the further exploration of phage genome communities isolated from the immediate environment. After all, evaluating and understanding both phage biology and taxonomy is essential to the development of phage therapy approaches and biotechnology application. The creation of models which can recognize helix-turn-helix motifs that have evolved independently in unrelated proteins, or the use of hidden Markov models as a profiling method for pattern detection are potential avenues for the identification of unknown genes that have arisen in divergent genomic sequences and await identification as part of ongoing research. [7]

6. Acknowledgements

The authors wish to express their appreciation to the TBR Course-based Research Experiences for Student Success Grant and the CSCC Foundation Grant, as well as the HHMI SEA-PHAGES Program and Dr. Elvira R. Eivazova, CSCC, as the project's mentor.

7. References

1. Jacobs-Sera, D., Abad, L. A., Alvey, R. M., Anders. K. R., Aull, H. G., Bhalla, S. S., et al. (2020). Genomic diversity of bacteriophages infecting *Microbacterium spp*. *PLoS ONE*, *15*(6): e0234636. https://doi.org/10.1371/journal.pone.0234636

2. Altamirano, F. L. G., & Barr, J. J. (2019). Phage therapy in the postantibiotic era. *Clinical Microbiology Reviews*, *32*(2): e00066-18. https://dx.doi.org/10.1128%2FCMR.00066-18

3. Lopes, A., Tavares, P., Petit, M., Guerois, R., & Zinn-Justin, S. (2014). Automated classification of tailed bacteriophages according to their neck organization. *BMC Genomics*, *15*(1): 1027. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4362835/

4. The Howard Hughes Medical Institute phage discovery and annotation custom guide (HHMI Press, 2019).

5. Carrias, A., Welch, T. J., Waldbieser, G. C., Mead, D. A., Terhune, J. S., & Liles, M. R. (2011). Comparative genomic analysis of bacteriophages specific to the channel catfish pathogen Edwardsiella ictaluri. *Virology Journal*, *8*(6). https://dx.doi.org/10.1186%2F1743-422X-8-6

6. The Actinobacteriophage Database, https://phagesdb.org/

7. Pellegrini-Callace, M., & Thornton, J. (2005). Detecting DNA-binding helix-turn-helix structural motifs using sequence and structure information. *Nucleic Acids Research 33*(7): 2129-2140. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1079965/