

Characterization of Intergenic Recurring Elements in Actinobacteriophage Cluster BI1

Sonya R. Layton and Lee E. Hughes

Department of Biological Sciences, University of North Texas

Email: SonyaRamzanali@my.unt.edu

Abstract

Bacteriophages have been used in many important discoveries, such as DNA as the hereditary material of cells, and tools, such as CRISPR/Cas systems. Actinobacteriophages are a diverse group of phages clustered by gene content similarity, with phages that share at least 35% of gene content similarity are clustered together. Cluster BI contains *Streptomyces* phages and is divided into four subclusters. Subcluster BI1 contains 19 phages, including OlympicHelado, Rima, DrGrey, Bing, and Spectropatronm, which are the focus of this study. Members of this subcluster contain small intergenic recurring elements within a small region on the right half of their genomes. These small repeats were visualized in a Phamerator map, then verified using a motif finder and searching the genomes for the predicted motifs. Motifs were found using MEME suite. DNA master was then used to search for these motifs within each phage genome. The motifs were found almost exclusively within the intergenic regions on the right side of the phage genome and each motif was in multiple intergenic regions within each phage genome. A phagesdb blast shows that the motifs are found throughout other clusters of Actinobacteriophages as well. We have identified and characterized these repeated regions bioinformatically, which will inform future experimental studies toward understanding their functional role in the phage genome.

Introduction

Bacteriophages have long been used in research because of their small size and relative simplicity. One notable experiment involving bacteriophage was the Hershey and Chase experiments in 1952. They used a T2 phage to determine that DNA, as opposed to protein, was the hereditary material of the cell (1). CRISPR/cas systems are being used in genomic editing. This system is naturally found within bacteria as an adaptive immunity to bacteriophages. CRISPR/cas play an important role in the coevolution of bacteria and bacteriophage (2). Actinobacteriophages are phages that infect Actinobacteria, a phylum of Gram-positive bacteria that includes *Streptomyces spp.* The Actinobacteriophage database, available at phagesdb.org, groups bacteriophages by gene content similarity. Those with at least 35 percent gene content similarity are placed into a cluster (3). A cluster can then be broken down into subclusters. The subcluster that is the focus of the research presented here is Subcluster BI1. The 19 members of subcluster BI1 have a genome size ranging from 54664 bp to 56472 bp. The GC content ranges from 59.1% to 59.6% (4). Typically, phage genomes are compact with few intergenic regions. The left side of the genome contains the structural genes of the phage and the right side has the

functional genes of the phage. The Phamerator map of the B11 bacteriophages show an interesting pattern outlined in the results section (Figure 2).

Methods

Motif Hunting:

MEME suite was used on OlympicHelado, Bing, Spectropatronm, Rima, and DrGrey. The parameters used were as follows:

The image shows a screenshot of the MEME Data Submission Form. The form is titled "Data Submission Form" and includes the following sections and options:

- Select the motif discovery mode:** Classic mode (selected), Discriminative mode, Differential Enrichment mode.
- Select the sequence alphabet:** DNA, RNA or Protein (selected), Custom.
- Input the primary sequences:** Enter sequences in which you want to find motifs. Includes an "Upload sequences" button and a "Choose file" button.
- Select the site distribution:** How do you expect motif sites to be distributed in sequences? Any number of repetitions (selected).
- Select the number of motifs:** How many motifs should MEME find? 10 (selected).
- Input job details:** (Optional) Enter your email address. (Optional) Enter a job description.
- Advanced options:** (Reset)
 - What should be used as the background model? (Default model of sequences)
 - How wide can motifs be? Minimum width: 6, Maximum width: 50
 - How many sites must each motif have? Minimum sites: 1, Maximum sites: 10
 - Can motif sites be on both strands? (DNA/RNA only) search gives strand only
 - Should MEME restrict the search to palindromes? (DNA only) look for palindromes only
 - Should MEME shuffle the sequences? Shuffle the sequences

At the bottom, there is a note: "Note: if the combined form inputs exceed 10MB the job will be rejected." and two buttons: "Start Search" and "Clear Input".

Locating Motifs in the Genome:

OlympicHealado was used as the test for the procedure and all the motifs were searched in DNA Master. A 10 base pair segment was chosen to better identify psudeomatches as well. The locations of the various motifs were mapped on the Phamerator map using arrows.

BLAST searches:

Each motif was searched using NCBI BLAST as well as a local PhagesDB BLAST.

Promoter Search:

A promoter search was conducted within DNA Master.

Results

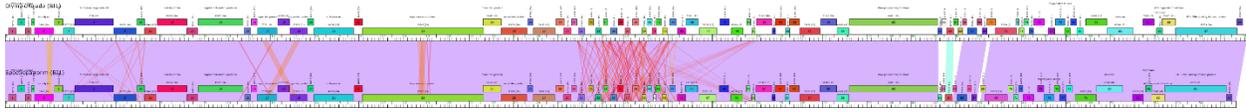


Figure 1: Phamerator map showing a comparison of the genomes of OlympicHelado and Spectropatrom. The purple shading shows the most nucleotide similarity and white is the least similar. The genes and gene phamilies are represented by the number and color in the boxes, respectively.

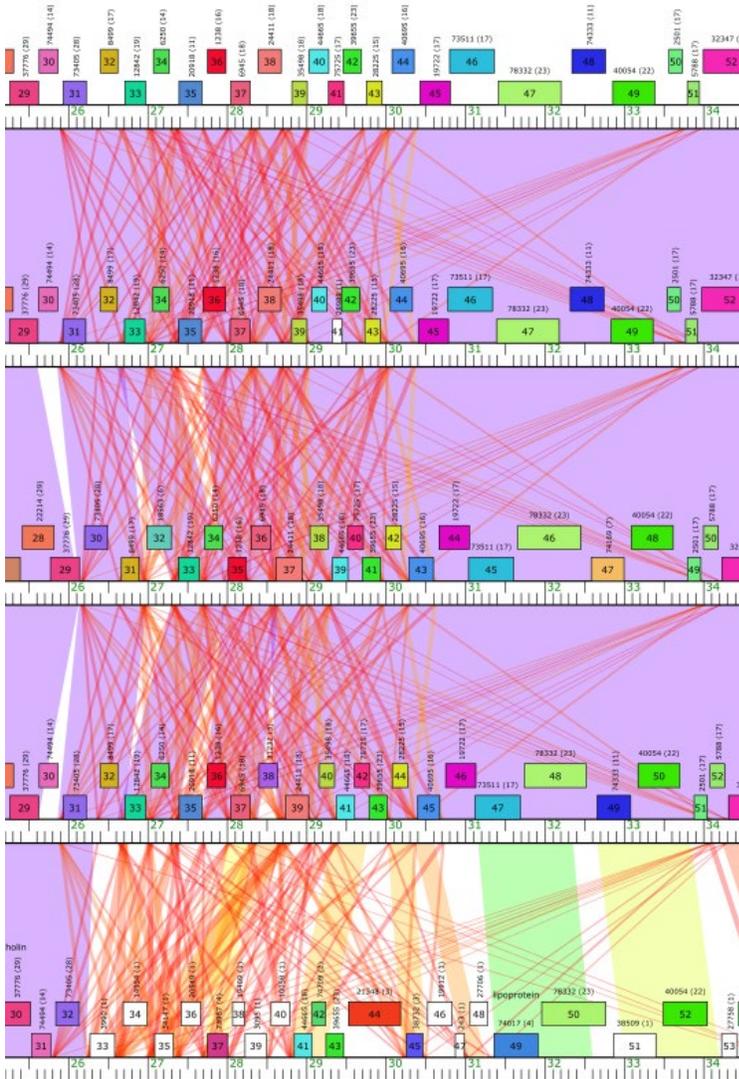


Figure 2: A zoomed in area of a Phamerator map showing the area of interest in OlympicHelado, Spectropatrom, DrGrey, Rima, and Bing. The red indicates regions where similar nucleotides are found.

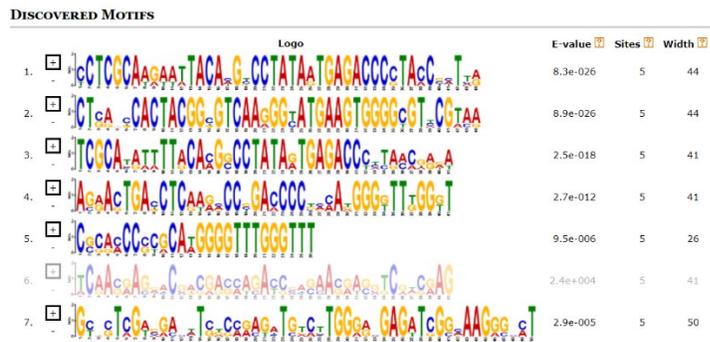


Figure 3: MEME output of motif search for DrGrey. The ones with significant e-values are highlighted. Letter size indicates conservation of bases at different occurrences within the genome.



Figure 4: MEME output of motif search for Rima. Letter size indicates conservation of bases at different occurrences within the genome.



Figure 5: MEME output of motif search for OlympicHelado. Letter size indicates conservation of bases at different occurrences within the genome.

DISCOVERED MOTIFS

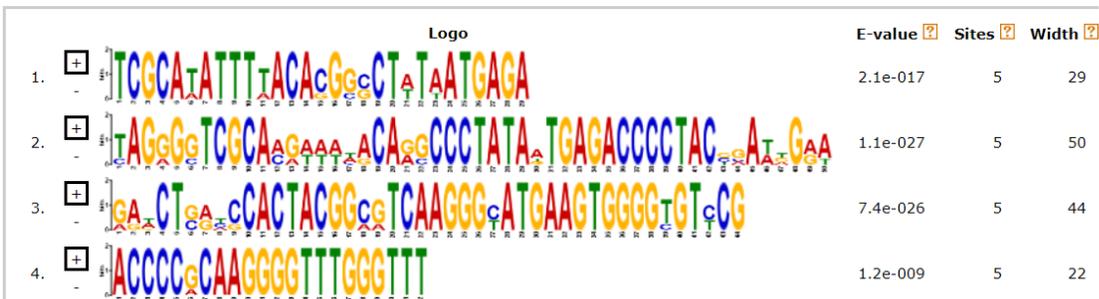


Figure 6: MEME output of motif search for Bing. Letter size indicates conservation of bases at different occurrences within the genome.

DISCOVERED MOTIFS

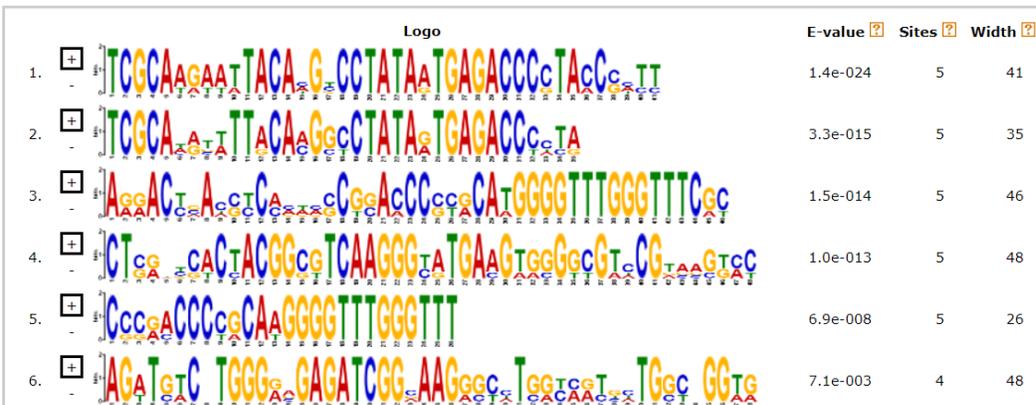


Figure 7: MEME output of motif search for Spectropatromm. Letter size indicates conservation of bases at different occurrences within the genome.

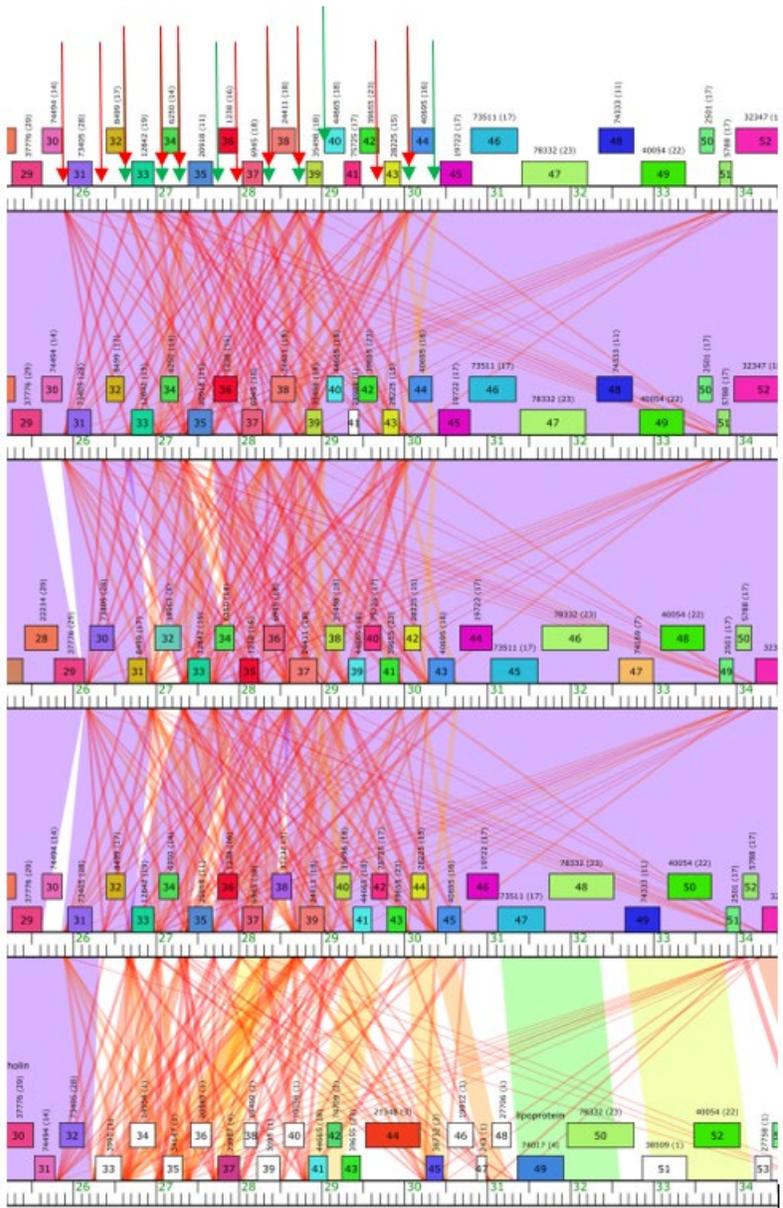


Figure 8: Zoomed in Phamerator map with the region of interest with arrows indicating the location of motifs found in OlympicHelado using MEME suite. Red arrows are motif 1 and 2, Green arrows are motifs 3 and 5. All are intergenic with the exception of the motif on gene 40.

Streptomyces phage Spectropatronm complete sequence, 55707 bp in...	72	1e-12
Streptomyces phage Soshi complete sequence, 56311 bp including 9...	72	1e-12
Streptomyces phage Rima complete sequence, 56168 bp including 9-...	72	1e-12
Streptomyces phage Popy complete sequence, 56470 bp including 9-...	72	1e-12
Streptomyces phage OlympicHelado complete sequence, 56189 bp inc...	72	1e-12
Streptomyces phage Namu complete sequence, 56168 bp including 9-...	72	1e-12
Streptomyces phage Meibysrarus complete sequence, 55097 bp inclu...	72	1e-12
Streptomyces phage Jaylociraptor complete sequence, 55699 bp inc...	72	1e-12
Streptomyces phage IceWarrior complete sequence, 55532 bp includ...	72	1e-12
Streptomyces phage Hoshi complete sequence, 55728 bp including 9...	72	1e-12
Streptomyces phage FrodoSwaggins complete sequence, 56472 bp inc...	72	1e-12
Streptomyces phage FidgetOrca complete sequence, 56115 bp includ...	72	1e-12
Streptomyces phage Esketit complete sequence, 54664 bp including...	72	1e-12
Streptomyces phage DrGrey complete sequence, 56076 bp including ...	72	1e-12
Streptomyces phage CherryBlossom complete sequence, 55563 bp inc...	72	1e-12
Streptomyces phage GirlPower complete sequence, 55521 bp includi...	70	4e-12
Streptomyces phage Madamato complete sequence, 55349 bp includin...	66	7e-11
Gordonia phage Schwartz33 complete sequence, 59457 bp including ...	64	3e-10
Gordonia phage Untouchable complete sequence, 61623 bp including...	60	4e-09
Gordonia phage Kenosha complete sequence, 60899 bp including 9-b...	58	2e-08
Gordonia phage Crocheter complete sequence, 60531 bp including 9...	58	2e-08
Gordonia phage Tanis complete sequence, 59727 bp including 9-bas...	56	6e-08
Rhodococcus phage Poco6 complete genome, 78064 bp including 9-ba...	56	6e-08
Gordonia phage Odesza complete sequence, 59746 bp including 9-ba...	56	6e-08
Gordonia phage Kerry complete sequence, 59608 bp including 9-bas...	56	6e-08
Gordonia phage Gravy complete sequence, 59545 bp including 9-bas...	56	6e-08
Gordonia phage Duffington complete sequence, 61343 bp including ...	56	6e-08
Gordonia phage Chikenjars complete sequence, 61544 bp including ...	56	6e-08
Gordonia phage Avazak complete sequence, 60278 bp including 9-ba...	56	6e-08
Gordonia phage AlainaMarie complete sequence, 61315 bp including...	56	6e-08
Gordonia phage Rickmore complete sequence, 60218 bp including 9-...	54	3e-07
Gordonia phage OhMyWard complete sequence, 60978 bp including 9-...	50	4e-06
Streptomyces phage Bing complete sequence, 56341 bp including 9-...	48	2e-05
Arthrobacter phage Sporto complete sequence, 54780 bp including ...	46	6e-05
Streptomyces phage SendItCS complete sequence, 55993 bp includin...	46	6e-05
Streptomyces phage Rainydai complete sequence, 57623 bp includin...	46	6e-05
Rhodococcus phage Pepy6 complete genome, 76797 bp including 9-ba...	46	6e-05
Arthrobacter phage Loretta complete sequence, 56953 bp including...	46	6e-05
Arthrobacter phage Ingrid complete sequence, 56953 bp including ...	46	6e-05

Figure 9: A screenshot of the PhagesDb BLAST search of one of the motifs found in the OlympicHelado genome. Some of the hits are with other BI1 phages not represented in this study but other significant hits include phages infecting *Gordonia spp*, *Rhodococcus spp*, etc.

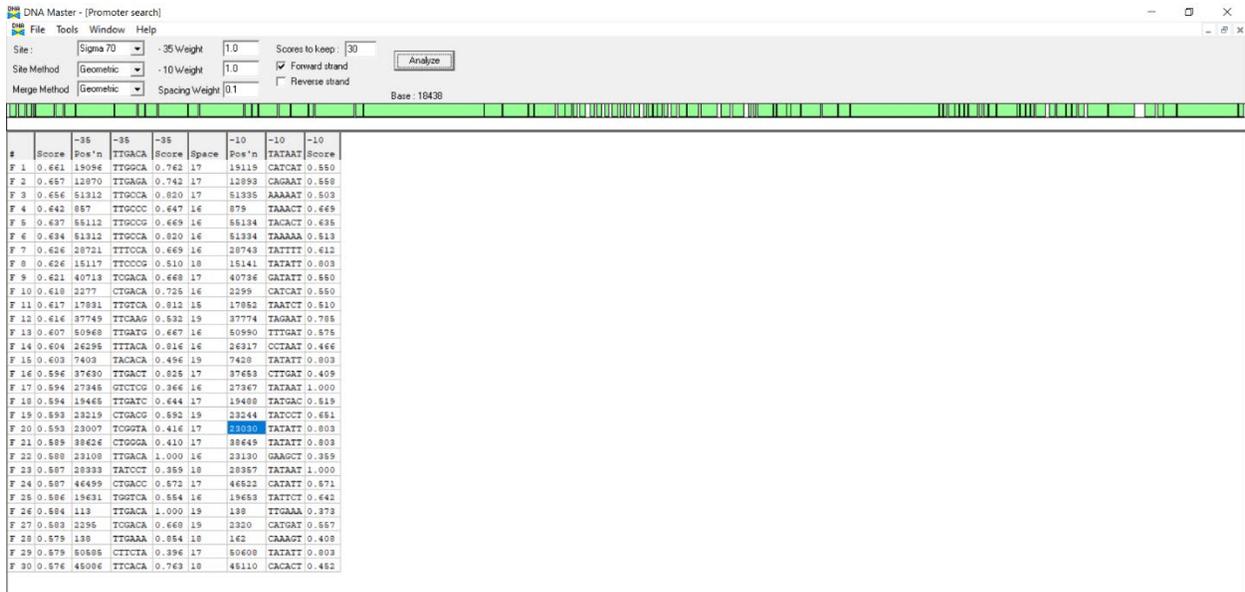


Figure 10: Output of the promoter search of OlympicHelado on DNA Master. The location of promoters are given and the scores are best when closest to 1.

Observations/Conclusions:

Motifs were found in all 5 of the bacteriophages that were searched. These motifs are very similar in all of the phages. Some of the motifs are very similar to other motifs found within the same genome.

The motifs were found to be mainly in the region of interest. Multiple copies of the motifs were found in the intergenic spaces between the genes of this region.

An NCBI BLAST shows no significant hits for any known function so a potential function could not be determined.

The PhagesDb BLAST search shows that these motifs are found within phages belonging to other clusters and within phages that infect other genera of bacteria.

The search for promoters shows multiple potential promoters in the region of interest leading to the hypothesis that these motifs are involved in DNA binding.

Future Work:

The potential promoters will be mapped out to determine exactly where they may be within this region of interest.

A gel shift mobility assay will be performed to check for binding properties.

References

1. National Academy of Sciences. 2001. *Biographical Memoirs: Volume 80*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/10269>.
2. Szczepankowska A. 2012. Role of CRISPR/cas system in the development of bacteriophage resistance. *Adv Virus Res* 82:289–338. doi:10.1016/B978-0-12-394621-8.00011-X
3. Maverich, TN and Hatfull, GF. (2017). Bacteriophage evolution differs by host, lifestyle, and genome. *Nature Microbiology*, 2: e17112.
4. Pittsburgh Bacteriophage Institute. The Actinobacteriophage Database. University of Pittsburgh, Web. 06 Nov. 2019.
5. <http://phamerator.org>