Characterization of Intergenic Recurring Elements in Actinobacteriophage Cluster BI1 Sonya R. Layton and Lee E. Hughes **University of North Texas, Denton, Texas**

Introduction

Bacteriophages are interesting organisms to study because they are small and relatively simple and have led to important discoveries such as DNA as the hereditary material of cells and CRISPR/Cas systems. Actinobacteriophages are grouped into clusters and subclusters, with subcluster BI1 being the major focus of this research. Bacteriophages typically have small genomes and with few intergenic regions. The BI1 phages have recurring sequences in the intergenic space on the right half of the genome. This half of the genome usually contains the functional genes, as opposed to structural genes, of the bacteriophage. This has led to the hypothesis that these intergenic sequences have a role in DNA binding.

Methods

- Each phage genome was searched for potential motifs using MEME suite.
- The motifs were searched for within the genome using \bullet DNA Master,
- The motifs were located and mapped on the genome to \bullet determine if they were in the interesting region on the Phamerator map.
- A BLAST search in PhagesDb was performed to check for the motifs in other phages.
- An NCBI BLAST was performed to look for potential functions.
- Potential promoters were searched for in the genomes using DNA Master.

Observations

- Potential motifs were found in each of the five phages • searched.
- These motifs were found to be recurring and intergenic on the half of the genome containing functional genes.
- The motifs were found within the region of interest (based on the Phamerator map below).
- The PhagesDb BLAST search showed that these motifs are found in many other phages belonging to other clusters.
- The BLAST shows no significant hits for any known functions.
- A preliminary search for promoters show potential promoters within these intergenic regions.

OlympicHelado (BI1) (99) 95 (1) 51 (1) 51 (1	terminase large subunit 9738 (90) 8	32175 (18)	45938 (90)	capsid maturation protease 45152 (90) 14	(68) 69 16	n ^{45325 (90} major tail protein 9
∑ [±] / ₂ <	3 4	19655 (18) 43761 (67) 9 11 5 6	13 7		§ 75791 (86) 15 17 11 12	73373 (61) 19 13 14
Spectropatronm (BI1)	terminase large subunit 9738 (90)	32175 (18)	portal protein 45938 (90) 🛱	capsid maturation protease 45152 (90)	(%) 9 9 9 9 9 7 9 7 8 7 8 7 7 7 7 8 7 7 7 7	n ^{45325 (99} major tail protein
9 2 9 4 3570 (26) 6 1 3 5 7	8	10655 (18) 10 43761 (67) 9 11	12 1659E	14	5 16 75791 (86) 15 17	18 73373 (61) 19

Figure 9: Phamerator map showing a comparison of the genomes of OlympicHelado and Spectropatronm

chitosanase

There are intergenic recurring sequences within the genomes of BI1 phages that we hypothesize have a role in DNA binding.

minor tail protei

minor tail protein



75347 (29)



46041 (90)

Logo	E-value 🙎	Sites ?	Width 🙎
_* G _* CCTATAATGAGACCCcTAcC _{7*} T _{**}	8.3e-026	5	44
TCAAgGG _I ATGAAGTGGGGGGGT _# CG _{IAA}	8.9e-026	5	44
_B CCTATA _R TGAGACC _{C#} taec _{re#} a	2.5e-018	5	41
CC ₌ GAcCCC ₌₌₌ cA ₌ GGGeTT _z GGeT	2.7e-012	5	41
GIIIGGGIII	9.5e-006	5	26
ççagAçç _{ə \$} şAAççageTÇ _{ə x} çaAG	2.4e+004	5	41
ç _{şş} əç _ə Tq _i ç _ı TÇÇqq, ÇAÇA TçÇq _ə AAÇqq, ş	2.9e-005	5	50
: MEME output of	mot	tif	

Logo	E-value ?	Sites ?	Width ?
G ₌ CCTATAATGAGACCCcTAcCc ₌ tt	1.5e-024	5	41
JecCTATAgTGAGACCcsta	1.7e-015	5	35
_z sCqsAcCCsqsCAtGGGGTTTGGGTTTC <u>s</u> c	1.6e-014	5	46
;TCAAGGG _{IA} TGAaGiggGgCGigCG	6.4e-014	5	41
GGIIIGGGIII	7.2e-008	5	26

Logo	E-value ?	Sites ?	Width ?	
A _* G _* CCTATAATGAGACCCcTAcCc _* tt	1.4e-024	5	41	
AaggçCTATARTGAGACCç _{ri} a	3.3e-015	5	35	
scGeeAcCCseeCAtGGGGTTTGGGTTTCsc	1.5e-014	5	46	
GcqTCAAGGG _I aTGAaGiggGqqGi _s CG _{igg} g _{is} ç	1.0e-013	5	48	
GGGGIIIGGGIII	6.9e-008	5	26	
çAGATCGG _a AAGqqç _a Tgqzqq _{aa} Tq _{aa} TGqq_GGtq	7.1e-003	4	48	
: MEME output of me	otif			
or Spectropotropm				

Results

	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 Namo 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	<u>_72</u>	1e-12
	Streptomyces phage FidgetOrca complete sequence, 56115 bp includ	<u>_72</u>	1e-12
	Streptomyces phage Esketit complete sequence, 54664 bp including	_72	1e-12
	Streptomyces phage DrGrey complete sequence, 56076 bp including	<u>_72</u>	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	<u>_58</u>	2e-08
	Gordonia phage lanis complete sequence, 59727 bp including 9-bas		66-08
	Condenie phage Odecze complete genome, 78064 bp including 9-ba		60-00
	Gondonia phage Volesza complete sequence, 59746 bp including 9-ba	56	60-00
	Gordonia phage Gravy complete sequence, 59505 bp including 9-bas	56	60-08
	Gordonia phage Duffington complete sequence, 55545 bp including 5 bas	56	60-08
	Gordonia phage Chikanians complete sequence, 61544 bp including	56	60-08
	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 OhMvWard 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 2: A screenshot of the		
	PhagesDb BLAST search of on	e o	f
	the motifs found in the		
?	OlympicHelado genome		
	DISCOVERED MOTIFS		

Logo	E-value ?	Sites ?	Width
\blacksquare $(]$	6.9e-026	5	42
\blacksquare $ICGCA_{abit}ITACA_{s}G_{s}CCTATA_{s}TGAGACCC_{s}TA_{s}$	3.6e-021	5	37
\blacksquare \square	1.3e-013	5	41
	1.3e-012	5	28
	1.0e-010	5	50
Figure 1: MEME output of m	otif		

Figure 4. MEME output of moun search for Rima

S	COVERED MOTIFS			
	Logo	E-value ?	Sites ?	Width 🙎
	TCGCATATTTTACAcGggCT+TAATGAGA	2.1e-017	5	29
	$\blacksquare_{JAGgGgTCGCAegeee_{z}ACAggCCCTATA_{z}TGAGACCCCTAC_{z}A_{Iz}Gee$	1.1e-027	5	50
	🗄 🖞 🚱 🗸 CT Se 🚛 CCACTACGG C STCAAGGG FATGAAGTGGGG TGT FCG	7.4e-026	5	44
		1.2e-009	5	22

Figure 6: MEME output of motif search for Bing



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.

