Spring 2019

Bioinformatic Analysis of The Flathead Lake Monster Bacteriophage

Jake Plagenz
Carroll College

SIGNATURE PAGE

This thesis for honors recognition has been approved for the

Department of Life and Environmental Sciences.

Daniel Gretch — 4/29/19

Director — Date

Dr. Daniel Gretch

Print Name

4/29/19

Reader — Date

Stefanie Otto-Hitt, PhD

Print Name

04/29/19

Reader — Date

EDWARD GLOWIENKA

Print Name

Bioinformatic Analysis of The Flathead Lake Monster Bacteriophage

Jake Plagenz

Honors Thesis

Carroll College

Department of Life & Environmental Sciences

Table of Contents

Abstract

The Flathead Lake Monster (FLM) bacteriophage was noted to have an abnormally-long tail upon its discovery. Once its genome was sequenced, this research sought out to identify the 117 FLM gene products using the BLASTp sequence alignment algorithm. This resulted in the discovery of five genes that are considered to be novel to the FLM. A specific gene within the FLM genome called the tape measure gene (TMG) was further analyzed once it was identified based on homology with other phages. Previous literature has suggested that a longer TMG can manifest itself as a longer bacteriophage tail length. This observation led to the hypothesis that a long tail length should be encoded by a correspondingly-long TMG within the FLM. The bioinformatic investigation involved comparing the FLM tail length and FLM tape measure gene length to other phages. The results found that the FLM does not have an abnormally long TMG when compared to how long its tail is, indicating that the FLM is an anomaly when compared to other phages. Future examination of phage mosaicism may yield more information as to why the FLM tail length is abnormally long.

Introduction

Bacteriophages, commonly known as phages, are a type of virus that infect bacteria (Veiga-Crespo et al., 2007). This feature of phages gave rise to its name of "bacteriophage," which literally means "bacteria eater" (Lin et al., 2017). Phages are the most abundant genetic entities on Earth (Brussow & Hendrix, 2002). In fact, the population of phages in the biosphere is estimated to be over 10^{31} phages, outnumbering their bacterial counterparts by roughly tenfold (Brussow & Hendrix, 2002). A phage's structure includes a capsid, which contains the phage's DNA, and a tail that allows penetration of bacterial cell membranes, using its fibers to facilitate binding to bacteria (Mayer, 2016).

Phages were first characterized in 1915 by Frederick William Twort (Veiga-Crespo et al., 2007). Twort's discovery gave way to a new paradigm in the field of medicine, allowing for the use of phages in therapeutic settings (Veiga-Crespo et al., 2007). For example, phage therapy has been an alternative to the use of antibiotics in treating bacterial infections (Lin et al., 2017). Phage therapy works by using phages to infect and kill the bacteria that is causing bacterial infections, similar to how antibiotics target bacteria cells (Lin et al., 2017). Although antibiotic therapy was initially favored over bacteriophage therapy, antibiotic resistance has elicited an interest in re-evaluating the effectiveness of phage therapy (Lin et al., 2017). Antibiotic resistance is the ability of bacteria to combat the effects of antibiotic, often resulting from the inappropriate and excessive prescribing of antibiotics by physicians (Ventola, 2015). Having knowledge of

the threat of antibiotic resistance drives the significance of studying practical uses of bacteriophages.

The Flathead Lake Monster bacteriophage was isolated in 2015 by Ian Lorang, an alumnus of Carroll College. Lorang isolated the phage in Bigfork, MT near Flathead Lake and named the phage the Flathead Lake Monster (Lorang, 2016; Figure 1). He predicted that the phage belonged to the F1 sub-cluster (Phage Data Bank, 2018).

Figure 1: Electron micrograph of the Flathead Lake Monster (Lorang, 2016).

This research studied the bioinformatics of the Flathead Lake Monster, which entailed examining each of the 117 genes in the phage's genome using a process called functional gene annotation. The corresponding products of these genes were analyzed using the Standard Protein BLAST (BLASTp) sequence

alignment algorithm. It was discovered that the Flathead Lake Monster's genome contains novel genes as well as genes that are conserved in other genetic entities. One gene of interest, the tape measure gene, appears to be conserved among a multitude of phages.

Many of the gene products discovered in this process are thought to play a role in the assembly of the Flathead Lake Monster's tail. Lorang noted that the phage he isolated contains an abnormally large tail length of 375 nm (Lorang, 2016). To investigate this further, the relationship between the phage's tail length and the size of the tape measure gene was analyzed. Studies performed by Katsura and Hendrix (1984) resulted in the observation of significantly shorter tail length in phages with base pair deletions in the tape measure gene. Knowing this, it was hypothesized that an abnormally long tail length would correlate with a correspondingly long tape measure gene within the FLM. Testing this hypothesis involved cross-examination of the bioinformatics of other phages related to the Flathead Lake Monster, including correlation analysis between the number of base pairs found in the tape measure gene and the phage tail length.

Methods

Functional Gene Annotation

Amino acid sequences for each of the 117 gene products encoded by the Flathead Lake Monster's genome were accessed from the Actinobacteriophage Database (Phage Data Base). The sequences were analyzed to produce a list of potential gene products using the BLASTp sequence alignment algorithm from the National Center for Biotechnology Information (NCBI) website. BLASTp matched Flathead Lake Monster sequences with specific gene products from other genetic entities. This analysis indicated the most likely function of each Flathead Lake Monster gene product. The "percent identity" parameter was used to quantify how significant each match was. The gene product with the highest percent identity for each Flathead Lake Monster sequence was named as the best match for that gene product. After identification, the "Accession Number" was logged for each match, which contained information about the candidate gene product and the genetic entity that shares the gene with the Flathead Lake Monster. Any gene that was not matched was considered novel to the Flathead Lake Monster and could not be identified.

Qualitative Analysis of Tape Measure Gene

Once the Flathead Lake Monster's tape measure gene was identified based on homology with other tape measure genes, its number of base pairs was recorded. The Flathead Lake Monster's tail length was also recorded as described by Lorang (Lorang, 2016). Phages reported by Pedulla et. Al (2003) were used to create a table of tape measure gene lengths and phage tail lengths. These values were then plotted on a scatter plot showing tail length vs. tape measure gene length. A best fit trend line was placed on this plot, allowing for qualitative comparison of the Flathead Lake Monster's position to other phages.

Quantitative Analysis of Tape Measure Gene

The trend line from the qualitative analysis was used to create a model function in a separate data set. Then, another data set of tail length vs. tape measure gene length was created using information from all of the phages except the Flathead Lake Monster. Residual values for these other phages were generated by subtracting the y-value of the phage's position from the y-value of the model function, which produces a measure of how far each phage is from the model function. The absolute value was taken for any negative residuals, allowing for the inclusion of positive residuals only.

The residuals were averaged to produce a single average residual value. The Flathead Lake Monster's tail length and tape measure gene length were then plotted on this graph, and the above procedure was repeated to find a second residual value for the Flathead Lake Monster. Comparison of the Flathead Lake Monster's residual to the average value of the other phages' residuals used a t-test, with a significance level of $\alpha = 0.05$.

Results

Functional Gene Annotation

Each of the 117 gene products encoded by the Flathead Lake Monster's genome were analyzed to generate a list of potential gene product functions (Table 1)

Table 1: List of potential gene product functions for each of the 117 gene products encoded by the Flathead Lake Monster's genome. Included in this table is the gene number, the number of amino acids of the gene product, accession number of the homologous organism, and the percent identity parameter. Gene products with no identifiable function are listed in boldface text.

Gene #	Number of Amino Acids	Potential Gene Product Function	Accession # of Best Match	% Identity
1	110	HNH endonuclease	YP_009198105.1	109/110(99%)
2	84	Small terminase subunit	YP_009198106.1	84/84(100%)
3	495	Large terminase subunit	YP_009198107.1	495/495(100%)
4	473	Portal protein	YP_009198108.1	473/473(100%)
5	241	Capsid maturation protease	YP_009198109.1	241/241(100%)
6	178	Scaffolding protein	YP_009124558.1	177/178(99%)
7	37	ADP-ribosylglycohydrolase family protein	WP_033204792.1	14/34(41%)
8	302	Major capsid protein	AVP42311.1	300/302(99%)
9	123	Head-to-tail connector complex protein	AUX81934.1	121/123(98%)
10	109	Head-to-tail connector complex protein	AUX81935.1	106/109(97%)
11	108	Head-to-tail connector complex protein	YP_009124562.1	108/108(100%)
12	134	Head-to-tail connector complex protein	AUX81937.1	128/134(96%)
13	269	Major tail subunit	YP_009016901.1	268/269(99%)
14	183	Tail assembly chaperone	YP_001469246.1	183/183(100%)
15	144	Tail assembly chaperone	AWN04937.1	144/144(100%)

Gene #	Number of Amino Acids	Potential Gene Product Function	Accession # of Best Match	% Identity
16	1176	Tape measure protein	AWH13716.1	1175/1176(99%)
17	569	Minor tail protein	YP_009199699.1	569/569(100%)
18	37	**Unknown**		
19	569	Minor tail protein	YP_008409554.1	569/569(100%)
20	276	Minor tail protein	YP_009214377.1	275/276(99%)
21	845	Minor tail protein	AVP41674.1	838/845(99%)
22	628	D-Ala-D-Ala carboxypeptidase	YP_009016098.1	622/628(99%)
23	388	Minor tail protein	YP_009125015.1	386/388(99%)
24	113	Minor tail protein	YP_655017.1	113/113(100%)
25	84	gp42	NP_818343.1	84/84(100%)
26	212	gp23	YP_655019.1	210/212(99%)
27	326	gp44	NP_818345.1	315/326(97%)
28	43	gp45	NP_818346.1	42/43(98%)
29	94	gp26	YP_655022.1	93/94(99%)
30	149	O-methyltransferase	WP_043125434.1	31/92(34%)
31	45	gp28	YP_655024.1	37/44(84%)
32	77	gp1	AAG48317.1	75/77(97%)
33	381	lysin A	YP_004123853.1	378/381(99%)
34	39	**Unknown**		
35	338	lysin B	YP_009125312.1	318/338(94%)
36	78	hypothetical protein WIVsmall_48	YP_008059949.1	60/71(85%)
37	77	holin	AOQ28489.1	76/77(99%)
38	124	gp32	YP_003495173.1	124/124(100%)
39	77	gp33	YP_003495174.1	77/77(100%)
40	258	DnaQ	YP_008408994.1	255/258(99%)
41	71	gp35	YP_003495176.1	71/71(100%)
42	170	Immunity repressor protein	ASZ72922.1	170/170(100%)
43	35	FAD-containing monooxygenase EthA	OUR97592.1	16/35(46%)
44	73	integrase	WP_087674176.1	17/50(34%)
45	48	hypothetical protein SEA_XERXES_34	AMS01980.1	48/48(100%)
46	65	HicB family antitoxin	YP_009302344.1	52/66(79%)
47	67	hypothetical protein SEA_PIPSQUEAKS_32	AMS02185.1	67/67(100%)
48	81	exodeoxyribonuclease V subunit gamma	WP_082674444.1	24/60(40%)
49	57	hypothetical protein PBI_PINTO_87	YP_009043871.1	22/31(71%)
50	372	integrase	YP_009013242.1	372/372(100%)
51	156	hypothetical protein MUTAFORMA13_46	AEJ93181.1	150/156(96%)
52	222	immunity repressor protein	ASZ72927.1	222/222(100%)
53	98	HTH DNA binding domain protein	AVJ50686.1	97/98(99%)
54	103	HTH binding domain protein	AVP41703.1	45/68(66%)
55	86	helix-turn-helix DNA binding domain protein	AWN04870.1	85/86(99%)
56	68	HNH endonuclease	YP_008409010.1	17/17(100%)
57	63	DNA replication protein	YP_009199735.1	57/63(90%)

Gene #	Number of Amino Acids	Potential Gene Product Function	Accession # of Best Match	% Identity
58	50	NAD(P)/FAD-dependent oxidoreductase	WP_077624790.1	16/36(44%)
59	161	NKF protein	YP_009199736.1	124/198(63%)
60	238	RecB-like protein	YP_009210425.1	195/262(74%)
61	93	hypothetical protein SEA_EMMA_56	ASZ72935.1	93/93(100%)
62	62	hypothetical protein POPTART_50	YP_009214410.1	62/62(100%)
63	111	hypothetical protein PBI_DLANE_52	AEK08596.1	111/111(100%)
64	133	WhiB family transcription factor	AWH14149.1	132/133(99%)
65	166	HTH domain protein	YP_008409124.1	163/164(99%)
66	121	TC3 transposase	P_008409125.1	121/121(100%)
67	51	Zn-dependent exopeptidase M28	WP_025484100.1	15/31(48%)
68	76	hypothetical protein PBI_TWEETY_61	YP_001469294.1	76/76(100%)
69	48	hypothetical protein PBI_TWEETY_62	YP_001469295.1	48/48(100%)
70	97	hypothetical protein PBI_BIPOLAR_62	YP_009200689.1	88/89(99%)
71	192	interferon-induced protein 44-like isoform X2	XP_014060891.1	30/123(24%)
72	165	**Unknown**		
73	49	HTH DNA binding protein	ARM70659.1	36/46(78%)
74	475	DNA methylase	AEK07834.1	459/475(97%)
75	46	hypothetical protein HAMULUS_66	YP_008409130.1	42/46(91%)
76	113	hypothetical protein SEA_MELISSAUREN88_68	AWH14114.1	110/113(97%)
77	288	site-specific DNA-methyltransferase	WP_108676300.1	188/226(83%)
78	44	hypothetical protein PBI_WEE_79	YP_004123901.1	44/44(100%)
79	283	hypothetical protein PBI_SISI_71	YP_008051197.1	283/283(100%)
80	87	hypothetical protein SEA_KIMBERLIUM_75	YP_009198179.1	81/87(93%)
81	40	NAD-dependent dehydratase	OGD20686.1	16/25(64%)
82	88	DNA binding protein	YP_009199731.1	21/53(40%)
83	91	beta-enolase	NP_001133193.1	14/36(39%)
84	64	hypothetical protein SEA_MELISSAUREN88_71	AWH14117.1	63/64(98%)
85	126	hypothetical protein PBI_WEE_84	YP_004123906.1	122/125(98%)
86	125	HNH endonuclease	YP_008531118.1	115/124(93%)
87	70	HNH endonuclease	YP_008531061.1	69/70(99%)
88	94	gp77 [Mycobacterium virus Llij]	YP_655073.1	94/94(100%)
89	67	**Unknown**		
90	62	gp78 [Mycobacterium virus Llij]	YP_655074.1	62/62(100%)
91	110	gp89 [Mycobacterium virus Che8]	NP_817427.1	110/110(100%)
92	57	hypothetical protein PBI_TWEETY_85 [Mycobacterium phage Tweety]	YP_001469318.1	57/57(100%)
93	123	hypothetical protein PBI_INVENTUM_80 [Mycobacterium phage Inventum]	YP_009125361.1	115/123(93%)
94	71	gp74 [Mycobacterium virus Pacc40]	YP_002241658.1	69/71(97%)
95	62	gp88 [Mycobacterium virus Ramsey]	YP_002241875.1	62/62(100%)
96	41	Peptidase T [uncultured Eubacterium sp.]	SCJ02761.1	13/20(65%)
97	63	hypothetical protein JABBAWOKKIE_93 [Mycobacterium phage Jabbawokkie]	YP_008410764.1	63/63(100%)
98	59	maltose phosphorylase [Bacteroides uniformis]	WP_057256365.1	19/50(38%)
99	51	hypothetical protein SEA_BYOUGENKIN_87 [Mycobacterium phage Byougenkin]	AWN05010.1	49/51(96%)
100	40	Recombination protein RecR [uncultured Ruminococcus sp.]	SCJ13520.1	14/31(45%)
101	53	hypothetical protein SEA_EMMA_95 [Mycobacterium phage Emma]	ASZ72970.1	49/50(98%)
102	87	cation acetate symporter [Pseudoalteromonas sp. T1lg24]	WP_105170464.1	28/77(36%)

Gene #	Number of Amino Acids	Potential Gene Product Function	Accession # of Best Match	% Identity
103	50	**Unknown**		
104	25	IS66 family transposase, partial [Moraxella osloensis]	WP_095356986.1	12/18(67%)
105	175	GIY-YIG endonuclease [Mycobacterium phage Job42]	YP_008126683.1	175/175(100%)
106	223	gp90 [Mycobacterium virus Pacc40]	YP_002241674.1	223/223(100%)
107	55	PREDICTED: E3 ubiquitin-protein ligase UPL4 isoform X1 [Eucalyptus grandis]	XP_010046832.1	15/32(47%)
108	65	lysophospholipase [Micrococcales bacterium]	PID55370.1	15/32(47%)
109	137	PREDICTED: telomere-associated protein RIF1-like isoform X12 [Salmo salar]	XP_014020835.1	25/58(43%)
110	80	MFS transporter [Mycobacteroides abscessus]	WP_079669700.1	41/79(52%)
111	53	hypothetical protein PBI_TWEETY_101 [Mycobacterium phage Tweety]	YP_001469334.1	53/53(100%)
112	478	glycosyltransferase [Mycobacterium phage Phatniss]	YP_009202615.1	465/478(97%)
113	70	DNA-directed RNA polymerase II subunit RPB2 [Valsa mali var. pyri]	KUI56648.1	16/41(39%)
114	157	serine/threonine kinase [Mycobacterium phage SimranZ1]	AQT25914.1	155/157(99%)
115	50	glycosyltransferase [Mycobacterium phage Ovechkin]	YP_009211270.1	23/36(64%)
116	69	hypothetical protein PBI_TWEETY_107 [Mycobacterium phage Tweety]	YP_001469340.1	69/69(100%)
117	204	glycosyltransferase [Mycobacterium phage Bobi]	YP_008409064.1	194/204(95%)

The tape measure gene was found to be the 16th gene in the Flathead Lake Monster's genome (Table 1). In addition, there were five genes with unknown function, which are considered to be novel to the Flathead Lake Monster.

Qualitative Analysis of the Tape Measure Gene

Qualitative comparison of the tape measure gene size to phage tail length was performed for the 14 phages reported by Pedulla et al. (2003), with the addition of the Flathead Lake Monster (Table 2). A graph of tail length vs. tape measure gene size was generated (Figure 2).

Table 2: Table of values for tape measure gene size (in base pairs) and tail length (in nm). All values for non-Flathead Lake Monster phages were reported by Pedulla et al. (2003).

Phage Name	Tape Measure Gene Size (bp)	Tail Length (nm)
HK022	2448	122
λ	2643	135
L5	2550	140
Bxz2	3027	138
Che9c	3750	152
HK97	3249	158
Che9d	3612	176
TM4	3804	190
Che8	3621	184
Omega	4815	188
Corndog	4299	232
Cjw1	4764	243
Rosebush	5613	226
Barnyard	6153	270
Flathead Lake Monster	3528	375

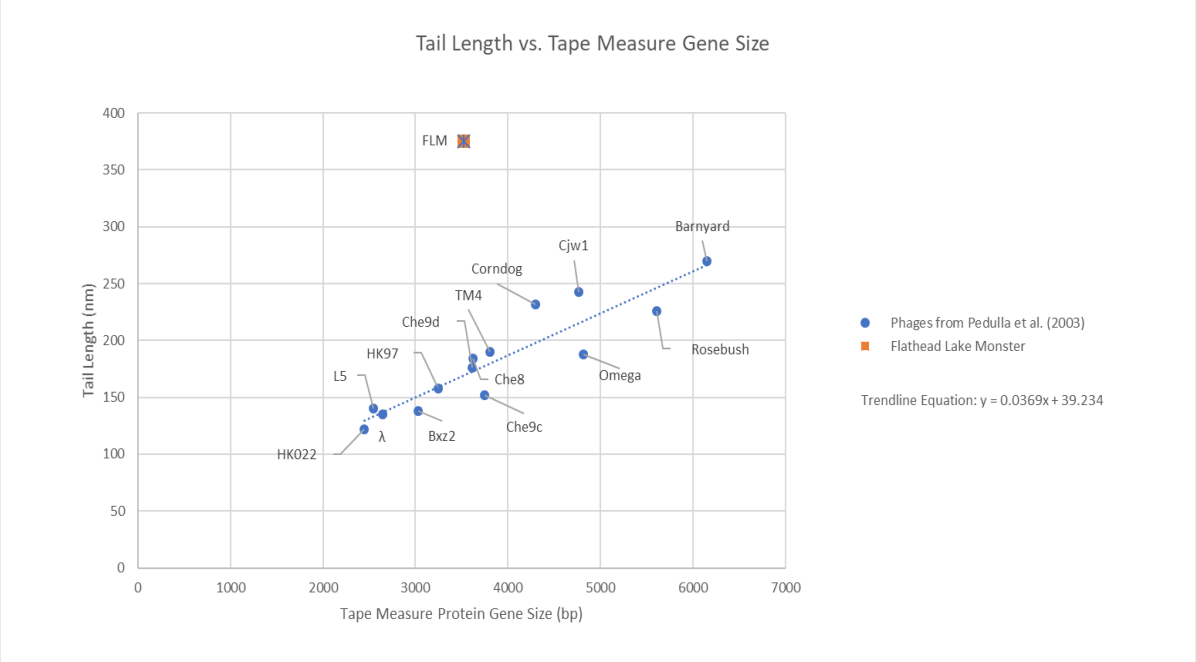

Figure 2: Graph of phage tail length vs. tape measure gene size. The trendline equation of the phages reported by Pedulla et al. (2003) was found to be $y = 0.0369x + 39.234$. The Flathead Lake Monster was included in this graph for qualitative comparison.

Quantitative Analysis of the Tape Measure Gene

The trendline from the qualitative analysis was used to create a model function that predicted the tail length for a given tape measure gene size. The values generated from the model function for each phage were subtracted from its actual tail length value to produce a residual. The average value for the non-Flathead Lake Monster phages was subsequently compared to the Flathead Lake Monster residual using a t-test.

Table 3: Table of values showing model function values based off of the equation $y = 0.0369x + 39.234$, the residual values, and the average residual values for the non-Flathead Lake Monster phages. It should be noted that the model function would predict that the Flathead Lake Monster would have a tail length of approximately 170 nm based upon how long its tape measure gene size is.

Phage Name	**Tape Measure Gene Size (bp)**	**Tail Length (nm)**	**Model Function Value for Tail Length (nm) {y=0.0369x+39.234}**	**Residual**	**Average Residual (Non-FLM Phages)**
HK022	2448	122	129.5652	7.5652	**13.98427143**
λ	2643	135	136.7607	1.7607	
L5	2550	140	133.329	6.671	
Bxz2	3027	138	150.9303	12.9303	
Che9c	3750	152	177.609	25.609	
HK97	3249	158	159.1221	1.1221	
Che9d	3612	176	172.5168	3.4832	
TM4	3804	190	179.6016	10.3984	
Che8	3621	184	172.8489	11.1511	
Omega	4815	188	216.9075	28.9075	
Corndog	4299	232	197.8671	34.1329	
Cjw1	4764	243	215.0256	27.9744	
Rosebush	5613	226	246.3537	20.3537	
Barnyard	6153	270	266.2797	3.7203	
Flathead Lake Monster	3528	375	169.4172	**205.5828**	

The t-test comparing the Flathead Lake Monster residual value of 205.5828 to the average residual value of 13.98427 for Non-Flathead Lake Monster phages yielded a p-value of $p < 0.0000001$, with $\alpha = 0.05$. From this, it can be concluded that the Flathead Lake Monster residual value is significantly different from the average residual value of the Non-Flathead Lake Monster phages.

Discussion

Functional Gene Annotation

Lorang performed a restriction enzyme analysis on the Flathead Lake Monster that predicted its assignment to the F1 sub-cluster (Lorang, 2016). His prediction was confirmed via genomic sequencing and alignment analysis (Phage Data Base, 2018). There are 152 phages in the F1 sub-cluster (Phage Data Base, 2018). Characteristics of the F1 sub-cluster include an average genome size of 57,439 base pairs, an average GC content of 61.5%, and an average gene number of 103.7 genes (Phage Data Base, 2018). For comparison, the Flathead Lake Monster has 57,663 base pairs in its genome, a GC content of 61.2%, and a total of 117 genes (Phage Data Base, 2018).

25 hypothetical proteins were identified among the 117 gene products (Table 1). It was found that 80% of those 25 hypothetical proteins showed highest alignment

scores with other F1 phages (Phage Data Base, 2018). This finding is consistent with the assignment of the Flathead Lake Monster to the F1 sub-cluster.

In addition to these 25 hypothetical proteins, there were five gene products that failed to align with any other known sequences. These represent genes that are novel to the Flathead Lake Monster (Table 1). This means that roughly ¼ (30/117) of the Flathead Lake Monster's genome is yet to have a function assigned to it.

Quantitative Analysis of Tape Measure Gene

The results of the quantitative analysis of the tape measure gene fail to confirm the initial hypothesis stating that a longer Flathead Lake Monster tail length will manifest itself as a longer Flathead Lake Monster tape measure gene. The statistically-significant difference between the Flathead Lake Monster's residual value and the average residual value for the Non-Flathead Lake Monster phages suggests that the Flathead Lake Monster is somewhat of an anomaly here. It is evident from inspection that the Flathead Lake Monster possesses a much longer tail length than what the model function predicted.

These findings suggest that the tape measure gene may not be the sole predictor of tail length in every phage. The Flathead Lake Monster has many gene products that appear to have a role in tail development and assembly, such as the gene products for genes 9 through 24 (Table 1). These gene products have names such as "major tail subunit" and "tail assembly chaperone" (Table 1). It could be speculated that a combination of the tape measure gene (gene 16) with these other genes predicts tail

length, rather than solely the tape measure gene, an idea that lends itself to further investigation.

Potential Exploration of Phage Mosaicism

It has been suggested that the concept of phage mosaicism contributes to the genetic diversity of bacteriophages (Hatfull, 2008). Phage mosaicism is the result of horizontal genetic exchange, which leads to a genome that appears to contain multiple individual modules that have been obtained from other phages (Hatfull, 2008). Mosaic analysis can be done by creating mosaic graphs that illustrate genomic overlap between multiple phages (Belcaid et al., 2010). An example of a mosaicism analysis map is shown in Figure 3 (Proux et al., 2002).

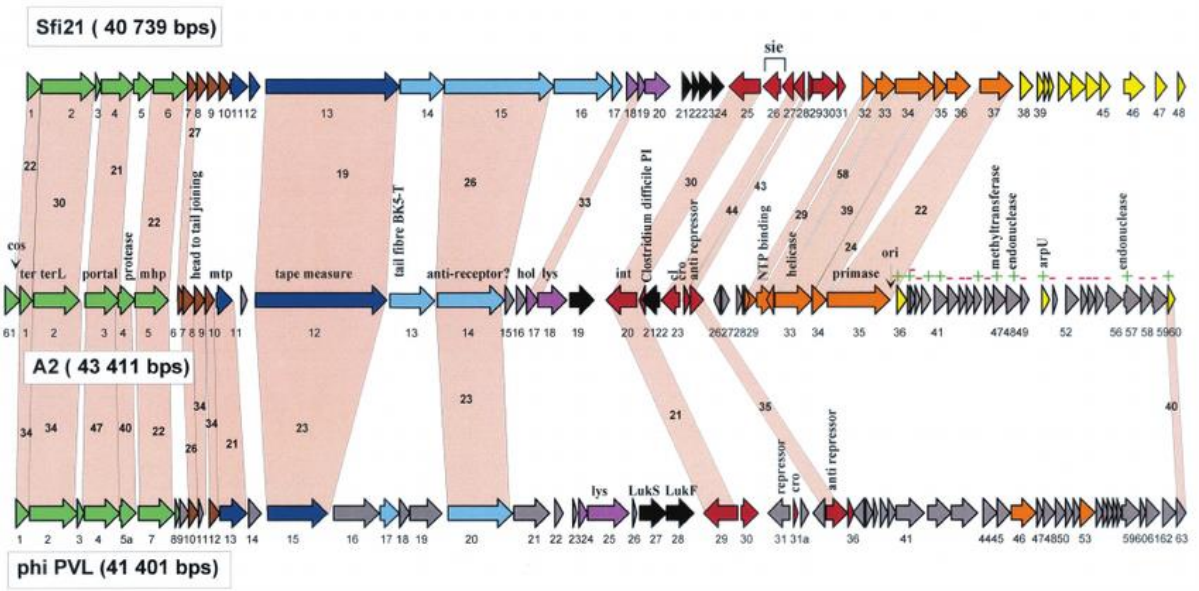

Figure 3: An example of a phage genome mosaicism map created by Proux et al. (2003). This demonstrates how genetic exchange events induced conservation of particular genes across phages.

As noted above, the Flathead Lake Monster has many analyzable gene products that might play a role in tail length. Mosaic analysis of these gene products can be investigated in efforts to quantify their role in other phages' tail lengths. Further exploration of mosaicism may yield an explanation of why the Flathead Lake Monster has such an anomalous tail length when compared to other phages of similar tape measure gene length.

Acknowledgements

I would like to express my sincere gratitude to Dr. Dan Gretch for his guidance and patience during this research. He equipped me with the tools necessary to perform this project independently, while providing mentorship whenever he saw fit. Thank you for allowing an inquisitive student like myself to expand my intellectual horizons. I couldn't have asked for a better advisor.

This research would not have been possible without the generous endowed scholarship provided by Thomas and Carolyn Paul. My thesis readers, Dr. Stefanie Otto-Hitt and Dr. Edward Glowienka, deserve special recognition for providing extensive edits and constructive feedback. I would also like to thank Ian Lorang for providing the foundation of this research through his work of characterizing the Flathead Lake Monster.

Lastly, I would like to express my indebtedness to my wonderful family: to my mother, Vicki, my father, Chad, my stepfather, Matt, and my siblings, Emily, Lauren, and Abram. Any of my accomplishments would not have been possible without your support. I love you all so much.

References

Belcaid M, Bergeron A, Poisson G. 2010. Mosaic graphs and comparative genomics in phage communities. Journal of Computational Biology. 17(9):1315-1326.

Brussow H, Hendrix R. 2002. Phage genomics: Small is beautiful. Cell. 108:13-16.

Hatfull G. 2008. Bacteriophage genomics. Curr Opin Microbiology. 11(5):447-453.

Katsura I, Hendrix R. 1984. Length determination in bacteriophage lambda tails. Cell. 39(3):691-698.

Lin D, Koskella B, Lin H. 2017. Phage therapy: An alternative to antibiotics in the age of multi-drug resistance. World Journal of Gastrointestinal Pharmacology and Therapeutics. 8(3):162-173.

Lorang I. 2016. Isolation and characterization of a bacteriophage The Flathead Lake Monster [thesis]. [Helena (MT)]: Carroll College.

Mayer G. 2016. Virology and bacteriophages. University of South Carolina School of Medicine. [updated November 23, 2016; accessed September 17, 2018]. http://www.microbiologybook.org/mayer/phage.htm

Pedulla M, Ford M, Houtz J, Karthikeyan T, Wadsworth C, Lewis J, Jacobs-Sera D, Falbo J, Gross J, Pannunzio N. 2003. Origins of highly mosaic mycobacteriophage genomes. Cell. 113(2):171-182.

Phage Data Base. 2018. The Actinobacteriophage Database. *Mycobacterium* phage Flathead. https://phagesdb.org/phages/Flathead

Proux C, Van Sinderen D, Suarez J, Garcia P, Ladero V, Fitzgerald G, Desiere F, Brussow H. The dilemma of phage taxonomy illustrated by comparative genomics of Sfi21-like *Siphoviridae* in lactic acid bacteria. Journal of Bacteriology. 184(21):6026-6036.

Rowley J. 2018. Modeling pseudo first-order chemical kinetics. Lecture presented at Carroll College Physical Chemistry Class.

Veiga-Crespo P, Barros-Velazquez J, Villa T. 2007. What can bacteriophages do for us? Communicating Current Research and Educational Topics and Trends in Applied Microbiology. 885-893.

Ventola C. 2015. The Antibiotic Resistance Crisis: Part 1: Causes and Threats. Pharmacy and Therapeutics. 40(4):277–283.