Fall 12-11-2016

Genomic Description and Annotation of a New Colivirus; KB1

Maggie Caswell
cmaggie@bgsu.edu

Genomic Description and Annotation of a New Colivirus; KB1

Maggie Caswell

December 2016

Advisors:

Raymond A. Larsen

Timothy Messer-Kruse

ABSTRACT

KB1 is a bacteriophage with relatively few established characteristics. However, it is thought to be similar to the novel environmental isolate bacteriophage, JK5. The aim of this study was to describe and annotate the genome of this new colivirus, relative to the genome of JK5. This comparison confirms the hypothesis that these are similar, but distinct viruses, and provides a foundation for the further understanding of both of these novel environmental isolates.

INTRODUCTION

Bacteriophages are bacterial viruses with the capability to infect bacterial cells (Sulakvelidze et al. 2001; Alavidze et al. 2001; Morris et al. 2001). Also known as phages, they can disrupt bacterial metabolism, causing cell lysis. Bacteriophages can be split into several classifications on the basis of nucleic acids and morphology, such as tailed, cubic, pleomorphic phage and helical (Ackerman, 2006). Furthermore, phages are can be split up by the chemical characteristics of their genomes. This encompasses phages with single and double stranded DNA as well as single and double stranded RNA. The subjects of the present study, JK5 and KB1, are tailed double stranded DNA viruses.

Since JK5 is similar to KB1, a similar study was reviewed to analyze the genomic components of JK5. This study analyzed the similarity between JK5 and both RTP and T1 (Kotlarsic et al. 2014; Kotlarsic et al. in prep). To infect *Escherichia coli* cells, JK5 makes use of the TonB system couple with the siderophore transporter FhuA. Additionally, it

was found that JK5 contains many different open reading frames on its genome ends. Many of these open reading frames are less than 100 codons long. JK5 carries a tRNA gene on the left end of its genome. The gene encodes a tRNA with an anticodon that specifies arginine. However, the role of the tRNA genes in bacteriophage still remains unclear (Wietzorrek et al. 2006). Furthermore, it was proposed that many of the open reading frames of JK5 with unknown functions take part in the infection and replication in the host.

METHOD

As stated before, KB1 is a relatively new virus with similarity to a more known virus, JK5. In this study, the relationship between bacteriophages KB1 and JK5 was analyzed by determining the genomic sequence of KB1 and comparing it to that of JK5. The genome of KB1 was sequenced by next generation ion torrent sequencing, with subsequent assembly using algorithms provided by the CLC genomic workbench as previously described for JK5 (Kotlarsic et. al 2014). KB1 spans a distance of roughly 44,000 base pairs. Similarly, the genome of JK5 also spans a distance of 44,000 base pairs with many similarities to KB1. Further analysis of the KB1 genome using a CLC DNA analysis program generated a map of potential protein coding sequences (termed “open reading frames: ”ORFs”). The ORFs of KB1 were analyzed relative to those of JK5, with similarities and differences noted.

RESULTS

The results are displayed as a genomic map in Figure 1 and the genes of KB1 are annotated to Table 1.

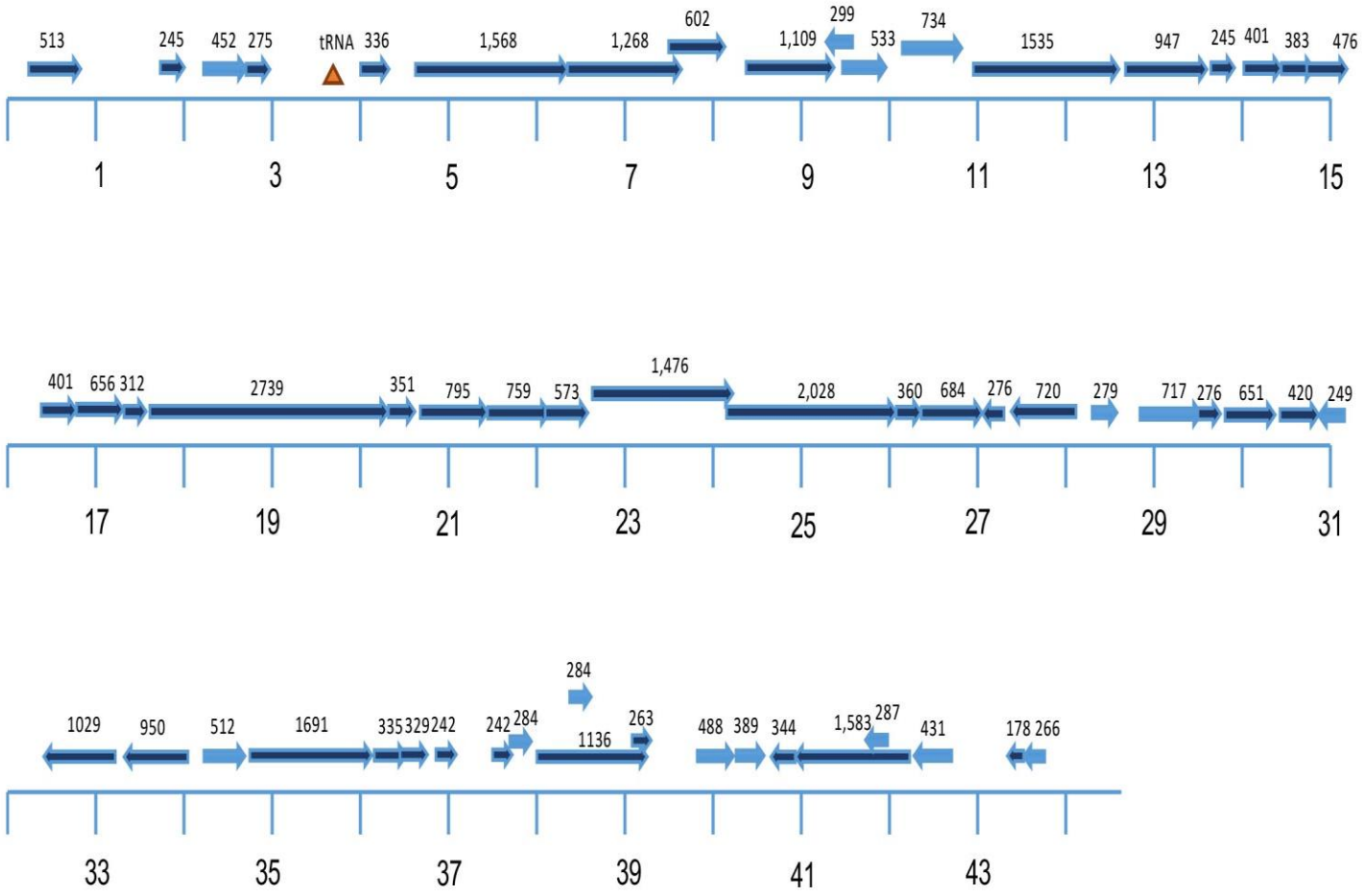

Figure 1. Gene map of the KB1 genome. The genes are indicated by arrows. Dark blue arrows indicate genes shared with JK5 while light blue arrows indicate genes that are not shared with JK5. The orange triangle identifies a tRNA gene.

Table 1. Putative genes of JK5.

	Range	bp	Strand	JK5 homolog	Name
1	489-1001	513	+	JK5-01	Hypothetical protein ACG-M12_0001 [Enterobacteria phage vB_EcoS_ACG-M12]
2	1967-2212	245	+	JK5-08	Hypothetical protein rtp7 [Enterobacteria phage RTP]
3	2425-2877	452	+		Hypothetical protein ACG-M12_0009
4	2919-3194	275	+	JK5-12	Hypothetical protein rtp13 [Enterobacteria phage RTP]
	3819-3893	74	+		tRNA
5	4207-4543	336	+	JK5-16	Putative terminase small subunit [Enterobacteria phage vB_EcoS_ACG-M12]
6	4832-6400	1568	+	JK5-17	Terminase large subunit Xanthomonas phage vB_XveM_RJBBI]
7	6451-7719	1268	+	JK5-18	Portal protein
8	7676-8278	602	+	JK5-19	Putative phage head morphogenesis protein [Escherichia phage RES-2009a]
9	8452-9561	1109	+	JK5-20	Prohead (putative) protease [Enterobacteria phage vB_EcoS_ACG-M12]
10	9366-9665	299	-		Hypothetical protein JK_27 [Enterobacteria phage JK06]
11	9572-10105	533	+		Hypothetical protein ACG-M12_0019 [Enterobacteria phage vB_ACG-M12]
12	10203-10937	734	+		DNA methylase N-4IN-6 domain containing protein partial [Pseudomonas amygdali]
13	11042-12577	1535	+	JK5-22	Putative tail fiber protein [Escherichia phage er/1c]
14	12656-13603	947	+	JK5-23	Major capsid protein [Escherichia phage bV_EcoS_AHS24]
15	13696-13941	245	+	JK5-24	Hypothetical protein [Escherichia phage e4/1c]
16	13984-14385	401	+	JK5-25	Halo29 [Escherichia phage RES-2009a]

17	14370-14753	383	+	JK5-26	Halo30 [Escherichia phage RES-2009a]
18	14707-15183	476	+	JK5-27	Conserved phage-related protein [Enterobacteria phage RTP]
19	15183-15584	401	+	JK5-28	Hypothetical protein ACG-M12_0025
20	15602-16258	656	+	JK5-29	Putative major tail protein [Enterobacteria phage cB_EcoS_ACG-M12]
21	16477-16789	312	+	JK5-30	ACG-M12-0029 Enterobacteria phage_EcoS_ACG_M12
22	16824-19563	2739	+	JK5-32	Putative tape measure protein Enterobacteria phate RPT
23	19593-19944	351	+	JK5-33	Putative minor tail protein Enterobacteria phage RTP Enterobacteria phage vB_EcoS_ACG-M12
24	19970-20765	795	+	JK5-34	Putative minor tail pprotein (Enterobacteria phage VB_EcoS_ACG-M12) and RTP
25	27045-21504	759	+	JK5-34	Putative minor tail protein [Enterobacteria phage vB_EcoS_ACG-M12]
26	21484-22057	573	+	JK5-36	Putative tail assembly protein [Enterobacteria phage RTP]
27a	22098-23574	1476	+	JK36 5' end	Putative tail fiber protein [Enterobacteria phage RTP]
27b	23561-25589	2028	+	JK36-3'end	Putative tail fiber protein [Enterobacteria phage RTP]
28	25589-25949	360	+	JK5-37	Hypothetical protein [Enterobacter asburiae] Hypothetical protein [Enterobacter sp. MGH8]
29	25948-26632	684	+	JK5-38	Hypothetical protein mEp 043_021 [Enterobacteria phage mEp043 c-1]
30	26656-26932	276	-	JK5-39	T1-30p cor gene
31	27059-27779	720	-	JK-40	C-5 cytosine specific methylase
32	28034-28313	279	+		Hypothetical protein ACG-M12_0040A [Enterobacteria phage vB_EcoS_ACG-M12]

33	28419-29208	717	+		Putative exodeoxyribonucleaseIII [Enterobacteria phage vB_EcoS_Acg-M12]
34	29180-29456	276	+	JK45-42	Punative exooeoxyribonuclease VIII (RecE) [Enterobacteria phage RTP]
35	29529-30180	651	+	JK5-44	Puntative recombination protein [Enterobacteria phage RTP]
36	30221-30641	420	+	JK5-45	Single-stranded DNA binding protein [Escherichia phage bV_EcoS_AHP24]
37	30954-31983	1,029	-	JK5-46	Tall fiber protein [Enterobacteria phage vB_EcoS_ACG-M12]
38	31982-32932	950	-	JK5-47	[Escheria phage e4/1c] [Escheria phage RTP]
39	32961-33473	512	+		Putative transcriptional regulator [Enterobacteria phage vB_EcoS_ACG-M12]
40	33568-35259	1,691	+	JK5-49	Hypothetical protein yejH [Enterobacteria phage JK06]
41	35226-35561	335	+	JK5-49	Putative ATP-dependent helicase [Enterobacteria phage VB_EcoS_ACG-M12]
42	35530-35859	329	+	JK5-50	Hypothetical protein rtp54 [Enterobacteria phage RTP]
43	36009-36251	242	+	JK5-56	Hypothetical protein JK_68 [Enterobacteria phage JK06]
44	36748-36990	242	+	JK5-57	Hypothetical protein ACG_M12_0058 [Enterobacteria phage vB_EcoS_ACG-M12]
45	36963-37247	284	+		Hypothetical protein Hp 60 [Enterobacteria phage RTP]
46	37330-38466	1,136	+	JK5-58	Hypothetical protein [Enterobacteria phage F20]
47	38511-38774	263	+	JK5-05	Hypothetical protein rtp5
48	39360-39848	488	+		Putative endolysin [Enterobateria phage RTP]
49	39824-40213	389	+		Hypothetical protein ACG-M12_0067 [Enterobacteria phage vB_EcoS_ACG-M12]
50	40228-40572	344	-	JK5-63	Hypothetical protein ACG-M12_0068

51	40569-42152	1583	-	JK5-65	Hypothetical protein rtp69 [Enterobacteria phage RTP]
52	42227-42658	431	-		Hypothetical protein rtp69 [Enterobacteria phage RTP]
53	43210-43388	178	-	JK5-70	Hypothetical protein
54	43384-43650	266	-		Hypothetical protein AHs4_49 [Escherichia phage bv_EcoS_AHS24]

Genome components of KB1: Looking at the genome of KB1, it is shown that this genome contains a wide array of open reading frames, many of which are homologous to genes in JK5. In particular, structural genes are very well conserved between the two viruses. For example, the JK5 homolog of KB1-9 is JK-20 as seen in Table 1. KB1-9 is responsible for helping cut structural proteins. Similarly, KB1-13 is a structural gene with a JK5 homolog, JK5-22. KB1-13 is a tail fiber gene. While this gene is shared with JK5, most other viruses don't have this gene. In viruses, the tail fiber is highly variable. Only a few viruses have it, including JK5 and KB1.

While there are many similarities between the genomes of JK5 and KB1, there are also differences. These differences will help us determine the unique characteristics of KB1 and similar viruses. While many structural genes are conserved between the two genomes, there are insertions and deletions that create variability. An example of this is the gene JK5-21, which is not a homolog to a gene in KB1. This gene, a minor head protein, is not highly conserved and is highly variable for those that have this gene.

Instead, in KB1, three other genes are present in—KB1-10, 11, and 12. While we are not entirely sure of the function of genes 10 and 11, KB1-12 appears to be DNA methylase.

DISCUSSION

Moving forward, a couple things will need addressed and resolved. First of all, we are unsure of the exact start site of KB1. Figures 2 and 3 show the suspected first and last genes of KB-1 with the homology to other similar viruses, including JK5.

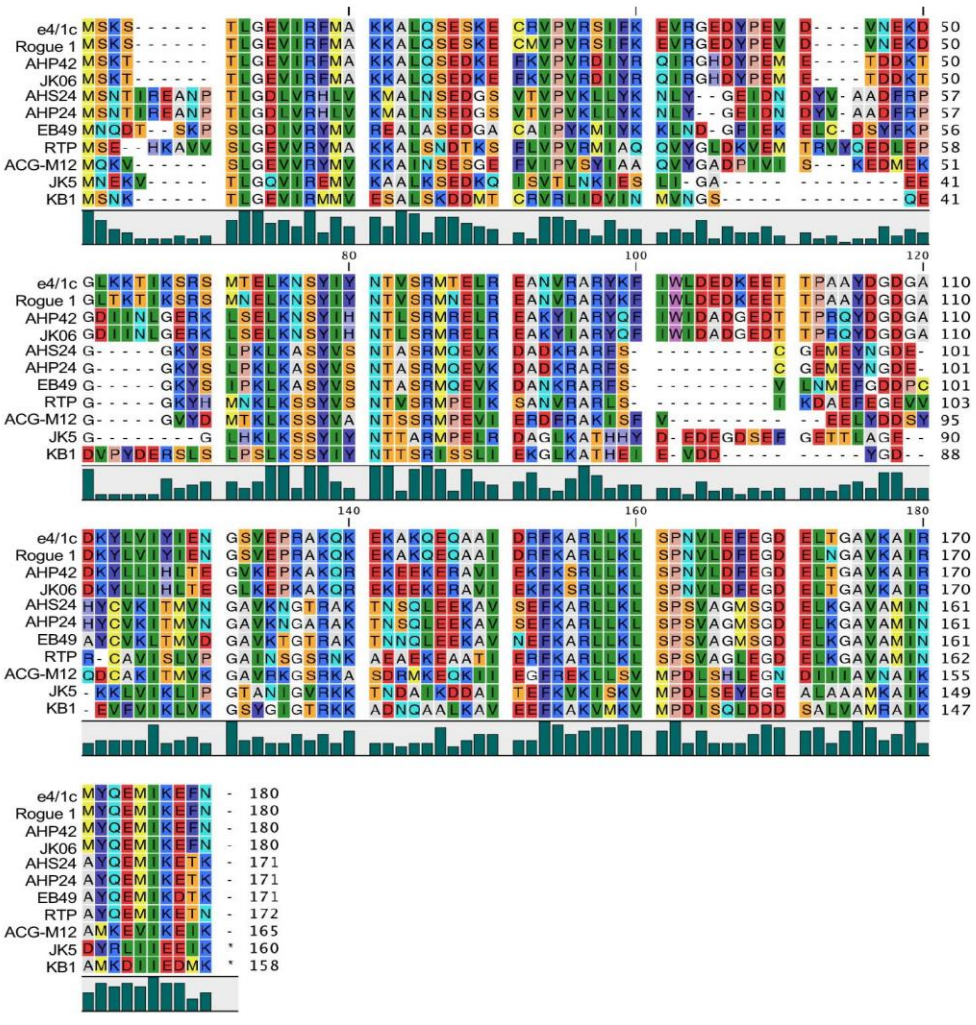

Figure 2. Our proposed first gene of KB1 with homology to the first gene of other similar viruses, including JK5.

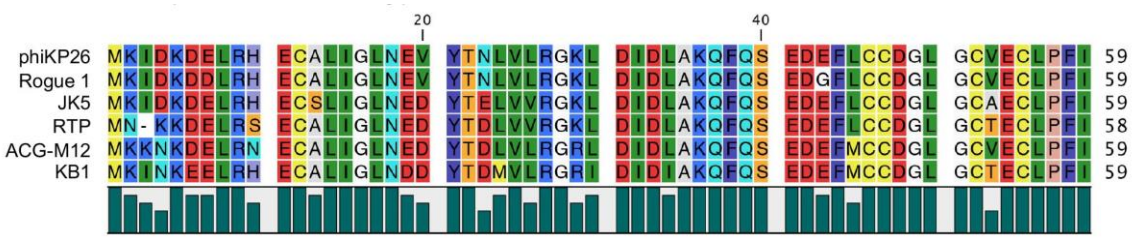

Figure 3. Hypothetical last gene of KB1 shown with its homology to the last gene of other viruses, including JK5.

Fortunately, we do know the true start and stop site of JK5. We also know that the official start site of KB1 occurs between the homolog of JK5-1 and JK5-70. We have used our knowledge of the start site of JK5 to determine the possible start site of KB1 relative to JK5. Moving forward, however, this issue will need solidified and the true ends of KB1 will need to be identified. The plan for addressing this will be to use restriction mapping. The restriction site within that specific region will be mapped and a range of sites will be analyzed. Restriction analysis will be completed, in which we can expect to see both large and small bands. The small bands will be indicative of sites near the ends of the virus. This will aid us in localizing the ends of the virus down to a few bases. We can also be on the lookout as more viruses are discovered that might resemble KB1 and therefore might give us further insight into the true start and stop site of KB1.

The second issue that needs addressed is the nature of the tail spike protein of KB1. JK5-36 is a tail fiber protein. The tail spike protein is homologous to a gene, J, in bacteriophage λ, which is responsible for initiating the process of tail assembly (Kotlarsic et al. 2014). The tail spike protein, usually the largest of the virus, is what the virus

utilizes to attach itself to a host, penetrate membranes, and transfer its DNA into phage-infected cells (Pukatzki et al. 2007; Ma et al. 2007; Revel et al. 2007; Sturtevant et al. 2007; Mekalanos et al. 2007). The tail spike protein also has great homology. In the case of KB1, the homolog of JK5-36 (tail spike protein) is divided into two different proteins, KB1-27a and KB1-27b, due to a premature stop codon. We do not believe the tail spike actually exists as two genes. Therefore, a few possibilities of the reasoning behind this split gene exist. First of all, there is a possibility that this gene should be full length, (not divided in two) and if it is not, this means the C-terminus must be sufficient to carry out the functions of the tail fiber protein. A second possibility that exists is that this is potentially a sequencing error. This has a high probability of being accurate. A third possibility is that this is the real sequence of the tail spike protein in KB1 and that it does in fact exist as two genes, KB1-27a and KB1-27b and functions by ribosomal slippage. In viruses, there are certain sequences that are set up for and rely on ribosomal slipping. However, KB1 does not have sequences like this, making this possibility highly unlikely. Moving forward, PCR will be done to amplify the specified region. This region will be re-sequenced to see if there exists a frameshift mutation or if this was a sequencing error.

REFERENCES

1. **Ackerman, H-W.** 2006. Classification of Bacteriophages "The Bacteriophages" R. Calendar (ed). Oxford University Press (New York) pp. 8-16.
2. **Kotlarsic, J.** 2014. The Coliphage JK5 and the Nature of TonB-dependence. Masters Thesis, Department of Biological Sciences, Bowling Green State Univeristy.
3. **Pukatski, S., A. T. Ma, A. T. Revel, D. Sturtevant, and J. J. Mekalanos.** 2007**.** Type VI Secretion System Translocates a Phage Tail Spike-like Protein into Target Cells Where it Cross-links Actin. *Preceedings of the National Academy of Sciences* 104:15508-5513.
4. **Sulakvelidze, Alexander, Zemphira Alavidze, and J. Glenn Morris**. 2001. Bacteriophage Therapy. *Antimicrobial Agents and Chemotherapy* 45:649–659.
5. **Wietzorrek, A, Schwarz H, Herrmann C, Braun V.** 2006. The genome of the novel phage RTP, with a rosette-like tail tip, is homologous to the genome of phage T1. *Journal of Bacteriology* **188:**1419-1436.