Annotation of Three Actinobacteriophages: TukTuk, Shamu, and Megatron06

Lauren N. Fisk, Kylie A. Foss, Khloe Hester, Yuki Kobayashi, Madelin M. Kvatek, Elsie L. Larson, Joshua T. Lewis, Madilyn L. Manary, Laura F. Nelsen, Thomas E. Ostrem, Cayli H. Penner, Sydney M. Rehmke, Taylor M. Rentschler, Adriana Sandbulte, Dane D. Schoenborn, Jonas D. Tiensvold, Shayler O. Van Gelder, Kallie L. Van Kley, Megan VanGarven, Brianna A. Winford, Byron Noordewier, and Sara S. Tolsma

Department of Biology, Northwestern College, Orange City, IA 51041

ABSTRACT

We annotated three newly discovered bacteriophages. TukTuk and Shamu were isolated on the host *Microbacterium folorium* and Megatron06 on *Mycobacterium smegmatis*. Based on based on gene-content similarity (GCS) of 35% or higher to sequenced bacteriophages present in the Actinobacteriophage database, phagesDB, TukTuk was assigned to cluster EB, Shamu to cluster EA, and Megatron06 to Cluster H1. Here we report a summary of our annotation findings along with one in-depth analysis of an aspect of our annotation for each phage.

General Information

Table 1. Summary Information for TukTuk, Shamu, and Megatron06

Phage	Host	Isolation Source	Genome size (bp)	% GC	Cluster	# Protein coding genes	# tRNA genes	GenBank Accession Number
TukTuk	*M. foliorum*	Carnes, IA	41700	66.7	EB	71	1	PP358746
Shamu	*M. foliorum*	Orange City, IA	40239	62.0	EA2	63	1	PP405134
Megatron06	*M. smegmatis*	Orange City, IA	69095	57.5	H1	TBD	0	TBD

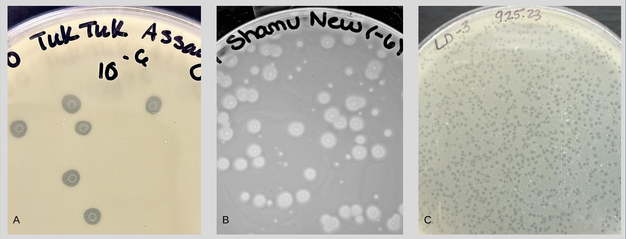


Figure 1. Plaque assays of A: Phage TukTuk B: Phage Shamu and C: Phage Megatron06 after 24 hours at 30ºC (A and B) or 37ºC (C).

Materials and Methods.
Viruses were purified (Discovery Guide Section 6) and amplified (Discovery Guide Section 7) prior to DNA isolation and characterization (Discover Guide Sections 8 and 10). All phages were sequenced at the University of Pittsburgh Bacteriophage Institute with Illumina Sequencing (http://phagesdb.org/phages/) and all genomes were assembled at the University of Pittsburgh (Newbler and Consed).

The sequences were auto-annotated using DNA Master software. Start sites, reading frames, coding potential, missing or mis-annotated genes, and gene functions were determined using Starterator, Phamerator (www.phamerator.org), NCBI BLAST (https://www.ncbi.nlm.nih.gov), GeneMark, Glimmer, Phagesdb (http://phagesdb.org/), and HHPred (https://toolkit.tuebingen.mpg.de/hhpred) directly and as collected in PECAAN (Phage Evidence Collection And Annotation Network) Specific guidelines are outlined in the SEA-PHAGES Bioinformatics Guide (https://seaphagesbioinformatics.helpdocsonline.com/home). We used Aragorn and tRNAscan software (http://mbio-serv2.mbioekol.lu.se/ARAGORN/) to search for tRNA genes.

References

1) Abdelrahman, et al. Antibiotics 2021, 10, 124. https://doi.org/10.3390/antibiotics10020124.
2) Elsa del Val, William Nasser, Hafid Abaibou, Sylvie Reverchon. RecA and DNA recombination: a review of molecular mechanisms. Biochemical Society Transactions, 2019, 1f10.1042/BST20190558ff.ffhal-02320683.
3) Fischetti VA. Bacteriophage lysins as effective antibacterials. Curr Opin Microbiol. 2008 Oct;11(5):393-400. doi: 10.1016/j.mib.2008.09.012. Epub 2008 Oct 14. PMID: 18824123; PMCID: PMC2597892.
4) Jumper, J. et al. "Highly accurate protein structure prediction with AlphaFold." Nature, 596, pages 583–589 (2021). DOI: 10.1038/s41586-021-03819-2.
5) Krishnan A, Burroughs AM, Iyer LM, Aravind L. Comprehensive classification of ABC ATPases and their functional radiation in nucleoprotein dynamics and biological conflict systems. Nucleic Acids Res. 2020 Oct 9;48(18):10045-10075. doi: 10.1093/nar/gkaa726. PMID: 32894288; PMCID: PMC7544218.
6) Pollenz RS, Bland J, Pope WH. Bioinformatic characterization of endolysins and holin-like membrane proteins in the lysis cassette of phages that infect Gordonia rubripertincta. PLoS One. 2022 Nov 17;17(11):e0276603. doi: 10.1371/journal.pone.0276603. PMID: 36395171; PMCID: PMC9671378.
7) Rahman MU, Wang W, Sun Q, Shah JA, Li C, Sun Y, Li Y, Zhang B, Chen W, Wang S. Endolysin, a Promising Solution against Antimicrobial Resistance. Antibiotics (Basel). 2021 Oct 20;10(11):1277. doi: 10.3390/antibiotics10111277. PMID: 34827215; PMCID: PMC8614784.
8) Schmelcher M, Donovan DM, Loessner MJ. Bacteriophage endolysins as novel antimicrobials. Future Microbiol. 2012 Oct;7(10):1147-71. doi: 10.2217/fmb.12.97. PMID: 23030422; PMCID: PMC3563964.
9) Tuteja N, Tuteja R. Prokaryotic and eukaryotic DNA helicases. Essential molecular motor proteins for cellular machinery. Eur J Biochem. 2004 May;271(10):1835-48. doi: 10.1111/j.1432-1033.2004.04093.x. PMID: 15128294; PMCID: PMC7164108.
10) Wang IN, Smith DL, Young R. Holins: the protein clocks of bacteriophage infections. Annu Rev Microbiol. 2000;54:799-825. doi: 10.1146/annurev.micro.54.1.799. PMID: 11018145.
11) X Zhou, W Zheng, Y Li, R Pearce, C Zhang, EW Bell, G Zhang, Y Zhang. I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction. Nature Protocols, 17: 2326-2353 (2022).

We are grateful for the support provided for this project from the Howard Hughes Medical Institute, SEA-PHAGES Program, Northwestern College, the NWC Biology Department. Many thanks to the members of the Hatfull laboratory at the University of Pittsburgh, especially Dan Russell for sequencing our phage genomes, and to Chad Miller at Northwestern College for the many ways he supports our research.

TukTuk (Cluster EB)

Research Question/Hypothesis: Cluster EB phage genomes have endolysins that are members of one of three phamilies:86883, 98619, or 143866. We hypothesize that phages with similar endolysin genes will share more genetic similarity outside of the endolysin gene compared to other EB cluster phages that carry a different endolysin gene.

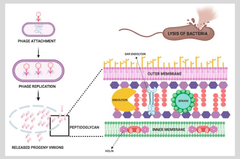


Figure 2. Endolysins are peptidoglycan hydrolase enzymes that function to digest cell wall components so progeny phages are released. Abdelrahman, et al. Antibiotics 2021, 10, 124. https://doi.org/10.3390/antibiotics10020124.

Figure 3. Endolysins are located in a similar position in genome (synteny), a few genes downstream of the tapemeasure gene and among genes that encode structural proteins. In Cluster EB phages, we find three different endolysins as indicated by the three different colors and different Pham numbers.

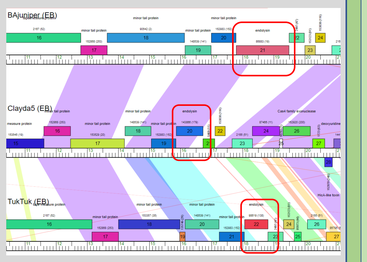
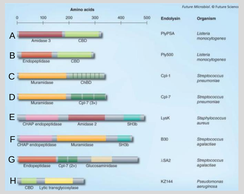


Figure 4. Most endolysins are modular in structure with 1-2 enzymatically active domains + linker + C-terminal cell wall binding domain. Schmelcher M, et al 2012.



Is the type of endolysin present in the genome an indication of general genetic similarity/dissimilarity?

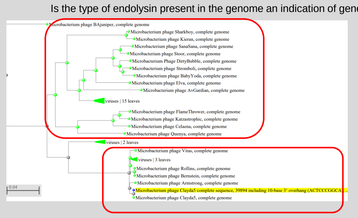


Figure 5. Phages with the same endolysin are more related to each other than they are to other EB phages as indicated in this cladogram so it seems that the type of endolysin present is an indication of general genetic similarity/dissimilarity.

Are the structures of the three endolysins found in Cluster EB phages similar?

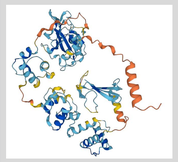


Figure 6. Alphafold predicted structure of the endolysin in BAjuniper (Pham 86883) suggests four possible domains, one with peptidase activity (N terminus), two with peptidoglycan binding activity, and one with amidase activity (C terminus). It seems most similar to structures E and G from Figure 4 (above). It is the most dissimilar of the three endolysins found in Cluster EB phages.

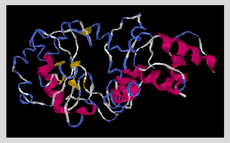


Figure 7. iTASSER predicted structure of the endolysin in Clayda5 (Pham 143886) sugests more similarity to the endolysin of TukTuk than BAjuniper.

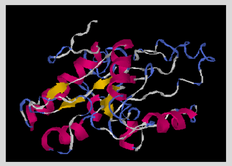


Figure 8. iTASSER predicted structure of the endolysin in TukTuk (Pham 98619) suggests more structural similarity to the endolysin in Clayda5 than that of BAjuniper.

Conclusion: There are three types of endolysins in Cluster EB phages. Each has a unique structure and the genetic similarity of the phages within the cluster can be predicted by examining the type of endolysin the phage uses.

Shamu (Cluster EA2)

Research Question/Hypothesis: Cluster EA2 phages have genes in the same Pham (84815) that are called DNA recombinase, DNA primase/helicase, RecA-like DNA recombinase, and helicase. We investigated which of these putative functions is the best call for genes in Pham 84815.

We began our investigating by asking about each of the proposed functions individually.

DNA recombinase: A fairly genetic term to describe enzymes that homologous recombination and maintenance of genome integrity.
RecA-like DNA recombinase: A more precise term that describes an enzyme that is similar to the RecA DNA recombinase from *E. coli*. This well-studied enzyme binds ssDNA present at a DNA break and searches for a homologous dsDNA to use as a template for to repair that break. There are a number of different, but related RecA proteins that form a superfamily (Figure 8).
DNA primase/helicase: Phage T7 is an example of a DNA primase/helicase (Gp4). It is ATP-dependent and essential for viral DNA replication and recombination. In T7, during replication, it unwinds the DNA double helix and synthesizes short RNA primers on the lagging strand (from which DNA polymerase extends).
Helicase: Helicases are ubiquitous molecular motor proteins use the energy of ATP hydrolysis to break the hydrogen bonds holding a DNA double helix together, thus catalyzing the unwinding of DNA. They play important roles in nearly all aspects of nucleic acid metabolism, including replication, repair, recombination, and transcription.

Conclusion: These enzymes are highly related with overlapping functions.

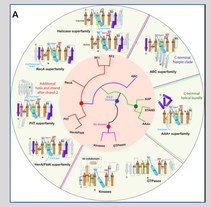


Figure 9. Recombinases and helicases are members of a large protein superfamily. Krishnan, et al. 2020.

Requirements to call a gene product a RecA-like DNA recombinase:

1. Alignment to an N-terminal domain that precedes a core ATPase. The core ATPase should have approximately 30 amino acids forming an alpha helix
2. A core ATPase domain with Walker A and Walker B motifs and a hydrolytic E and ahydrolytic motif
3. A C-terminal domain that includes a DE-rich Mg++ binding tail

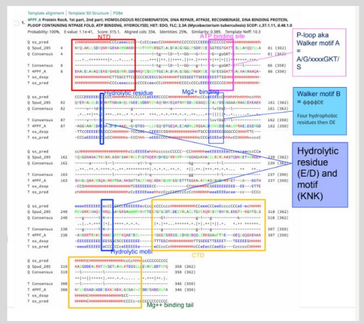
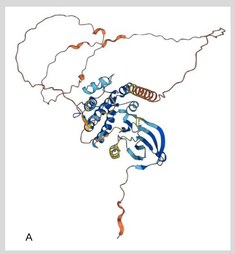


Figure 10. Spud_205 is an example of a gene that includes all these elements and is called a RecA-like DNA recombinase. Sally Malloy, SMART Team Presentation 2023.

Shamu_32 does not include all these elements, so we ruled out calling it a RecA-like DNA recombinase.
DNA recombinase is not on the approved functions list so we did not consider this a good call for Shamu_32.
HHPred hits included RecA-like proteins, ATPases, helicases, primases, and kinases. Our HHPred data indicated quite clearly that Shamu_32 had helicase activity in the C-terminal half of the protein.
HHPred was less clear about the function of the N-terminal half of Shamu_32. Some hits suggested primase activity, others recombinase activity, and others a more generic ATPase activity. We were less confident in the function of the N-terminus function so we are leaning toward a more conservative call of helicase.



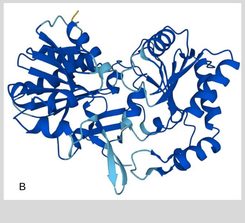


Figure 11. Alphafold predicted structure of Shamu_32 (Panel A) shares structural elements found in DEAD/DEAH box helicase (Panel B).

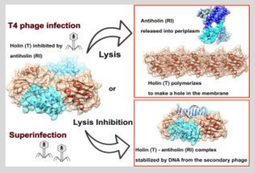
Conclusion:

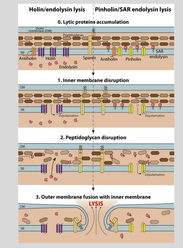
We called Shamu_32 a helicase. We are confident that it does not have the required elements to be called a RecA but it is possible it is a primase/helicase..

Megatron06 (Cluster H1)

Research Question/Hypothesis: Do H1 phages have more than one holin?

Figure 12. Holins are transmembrane proteins. They control the timing of bacterial lysis, contribute to destabilization of the membrane proton motive force, and generate large "pores" in the cell membrane to allow the exit of phage-encoded endolysin so these enzymes can access the peptidoglycan components of the cell wall. Wang, et al. 2000.





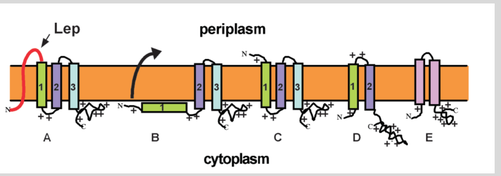


Figure 13. Proposed membrane topologies of holins. Holins and endolysins exhibit a wide range of sequence and structural diversity. Holins have 1-4 helical, hydrophobic transmembrane domains. Wang et al 2000.

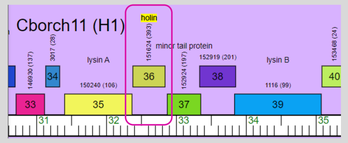


Figure 14. To call a gene a holin, the gene product must have at least two transmembrane domains, the gene must be adjacent to the endolysins, show sequence conservation to other holins when BLASTed, and there should be no additional transmembrane domains in the same area of the genome. Only one gene should be called a holin and if there are multiple genes that fit the required criteria, then all should be called transmembrane proteins.

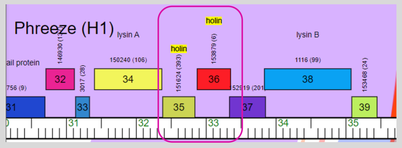


Figure 15. Some H1 phages called two holins (Beckerton, Oaker, Phreeze).

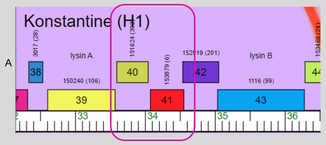


Figure 16. Some H1 phages, Predator and Konstantine, did not call a holin (A) even though there are several genes syntenic to the lysins with 1-4 transmembrane domains (B).

Conclusion:

We need to do more work to fully understand holins, their structures, functions, and abundance in phage genomes. For now, we are calling membrane proteins in Megatron06's genome.