Genetic Annotation of Bacteriophages MScarn, Knocker, and Neos5

David M. Adair*, Lindsey L. Craig*, Keziah J. Knudson*, Kaylee A. Maasdam*, Kaarina M. Marttila*, Katie M. Meyer*, Alexa Olguin*, Mitchell Rentschler*, Emily Schmidt*, April S. Van Tol*, Samantha Blum*, Byron Noordewier, and Sara S. Tolsma

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

*Authors contributed equally to this work

Abstract

We annotated the genomes of three recently discovered bacteriophages to learn more about their genetic composition. MScarn is a lytic bacteriophage that infects Gordonia terrae 3612. It was discovered and purified from soil collected in Iroquois, SD. MScarn is a cluster CT phage, one of only 37 discovered to date. Its genome is 45,677 base pairs long and has 10-nucleotide 3' sticky overhanging ends. Its GC content is 60.3% which is typical of CT cluster members. Knocker is a cluster B9 phage that was isolated on the host Mycobacterium smegmatis mc²155 from soil collected in Watertown, SD. Its circularly permuted genome contains 71,459 base pairs, and it has a high GC content of 69.7%. Similar to the other three members of the B9 cluster, it exhibits a lytic life cycle. Neos5, a lytic bacteriophage, was also isolated on Mycobacterium smegmatis mc²155 from soil collected in Baltimore, MD. It is a cluster B3 phage with a circularly permuted genome of 68,886 base-pairs and a 67.5% GC content, synonymous to the other 37 members of the cluster. All three phages were discovered, purified, and annotated by Northwestern College students.

Introduction

Bacteriophages (phages) are viruses that infect bacteria. They are the most abundant infectious particles on Earth yet are largely unstudied and uncharacterized (Sulakvelidze, 2011).

As part of the Howard Hughes Medical Institute's SEA PHAGES (Science Education Alliance – Phage Hunters Advancing Genomic and Evolutionary Science) Program, Northwestern College students have been phage hunting for more than five years and with an estimated 10^{31} phages on the planet, they can continue to hunt for many years.

This work reports the discovery, characterization, sequencing and annotation of three novel phages: Mscarn, Knocker, and Neos5.

Materials and Methods

MScarn was discovered by Gracelyn Fast in 2019 from an enriched sample. It infects the host *Gordonia terrae* 3612. Knocker was discovered by Cole Kruse and Joseph Kelly in 2018 from soil collected in Watertown, SD. It infects the host *Mycobacterium smegmatis* mc2 155. Neos5 also infects *Mycobacterium smegmatis* mc2 155. It was isolated from soil collected in Baltimore, MD in 2018. (Discovery Guide Protocol 5.2) or by enrichment isolation (Discovery Guide Protocol 5.5).

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). Neos5 and Knocker were sequenced at Northwestern College. MScarn was sequenced at the University of Pittsburgh. All three used Illumina Sequencing (http://phagesdb.org/phages/) and all three 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

Grose, J. H., & Casjens, S. R. (2014). Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family Enterobacteriaceae. Virology, 468-470, 421-443.
Hatfull, G. F. (2015). Dark Matter of the Biosphere: the amazing world of bacteriophage diversity. J Virol 89:8107-8110.
Mavrich TN and G Hatfull, (2017) Bacteriophage evolution differs by host, lifestyle and genome. Nat. Microbiol.
PubMed - NCBI. (n.d.). Retrieved March 23, 2019, from https://www.ncbi.nlm.nih.gov/pubmed/28692019
Pope, WH, et al (2017) Bacteriophages of Gordonia spp. Display a Spectrum of Diversity and Genetic Relationships. mBio. DOI: 10.1128/mBio.01069-17.
Pope, et al., (2015). Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity eLife 4:e06416. DOI: 10.7554/eLife.06416
Sulakvelidze A. (2011). Bacteriophage: A new journal for the most ubiquitous organisms on Earth. Bacteriophage, 1(1), 1-2.
WH Pope, MT Montgomery, JA Bonilla, R Dejong, RA Garlena, ...(2017) Complete genome sequences of 38 Gordonia sp. Bacteriophages Genome announcements 5 (1)
RW Hendrix, GF Hatfull, ME Ford, MCM Smith, RN Burns (2002) Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage Horizontal gene transfer, 133-VI

Results: MScarn

MScarn is a Cluster CT phage. It infects the host *Gordonia terrae* 3612 and was discovered from soil collected in Iroquois, SD. Its 45,677 bp linear genome has 3' sticky, 10 bp overhanging ends (CGGTAGGCTT) and a GC content of 60.3%. There are 37 members in the CT cluster. It is a lytic phage as seen by its clear plaque morphology and lack of an identifiable integrase gene in its annotated genome. Using bioinformatics, we were able to assign gene functions to 35 of MScarn's 72 genes.

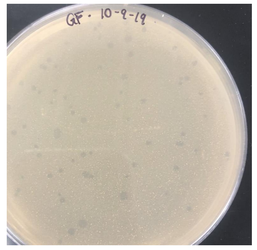

Figure 1. Plaque picture of MScarn shows small, clear plaques.

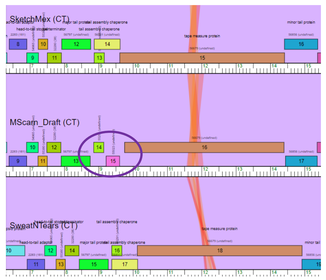

Figure 2. MScarn is most similar to SketchMex, SweatNTears, and Buttermilkdreams, all of which are also Cluster CT phages. The tail assembly chaperone gene contains a slippery sequence similar to that found in other CT cluster phages and therefore likely uses a programmed translational frameshift to regulate the production of proteins from this gene.

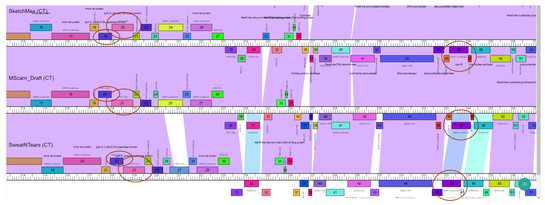

Figure 3. MScarn's lysin genes show an unusual synteny as seen in other CT cluster phages. The Lysin A domains are encoded by two genes: L-Ala-D-Glu peptidase domain and glycosyl hydrolase domain. Lysin B is about 30 genes away from the lysin A genes in the right arm of the phage rather than next to or very near the lysin A genes. We were not able to identify a holin gene but found several membrane proteins in the large region between Lysin A and B.

Results: Neos5

Neos5 is a Cluster B3 phage, which is a subcluster containing 37 members to date. Neos5 was discovered in 2018 by Byron Noordewier from soil collected in Baltimore, MD. It infects *Mycobacterium smegmatis* mc^2 155. Its genome was sequenced at Northwestern College and found to be circularly permuted and contains 68,886 bp. Its GC content is 67.5%. Its plaque morphology suggests that it is a lytic phage like other members of this subcluster. Neos5's genome contains 102 genes, of which we were able to assign functions to 37.

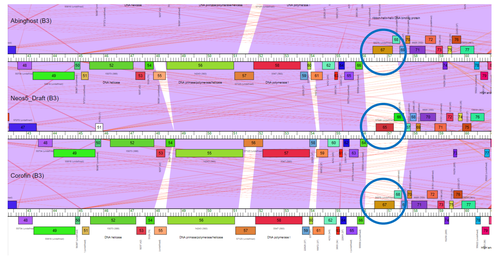

Figure 4. Neos5 is most similar to Abinghost, Chandler, and Corofin which are also Cluster B3 phages—with minor differences except for the genes circled in this phamerator map which showed quite a divergence. We are curious about the gene in this region and how this might make Neos5 different from its close relatives.

Discussion

We discovered and annotated three novel Actinobacteriophages to contribute to the growing understanding of phage biology and evolution (Mavrich and Hatfull, 2017). Most of the genomes conformed to expected synteny and genomic organization with the possible exception of Knocker and its potential holin gene. We are working to confirm this gene function call and polish all three annotations through the rest of the semester. The annotated genomes of all three phages will be submitted to GenBank for publication.

Support for this project was provided by the Howard Hughes Medical Institute, the Hatfull Lab at the University of Pittsburgh, SEA-PHAGES, and Northwestern College.

Results: Knocker

Knocker was discovered by Cole Kruse and Joseph Kelly in 2018 from soil collected in Watertown, SD. It infects the host *Mycobacterium smegmatis* mc2 155. Its genome was sequenced at Northwestern College and found to consist of 71,459 bp with circularly permuted ends. It was classified as a Cluster B9 phage—one of only four B9 phages identified to date. Like other B9 phages, it has a high GC content of nearly 70%. Knocker appears to be lytic by plaque morphology. Knocker's genome encodes 97 protein-coding genes of which we were able to call functions for 39.

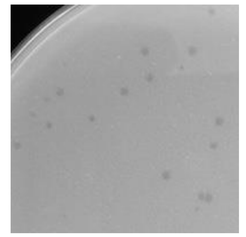

Figure 5. Plaque picture of Knocker shows small, clear plaques.

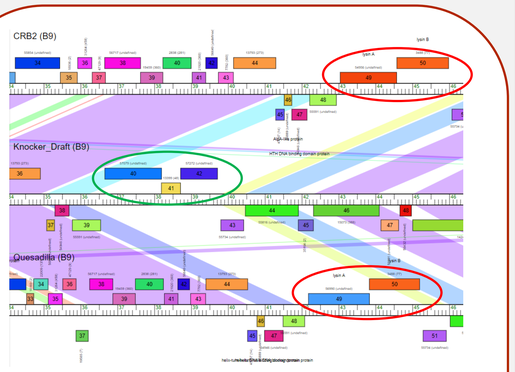

Figure 6. Cluster BCluster B phages tend not to have an identifiable holin gene. Knocker seems different than the other two B9 phages. We identified lysin A and lysin B genes (green oval) but there is a gene between them that looks a lot like a holin. It BLASTs to holins from Cluster B10, B2, and W phages and has two transmembrane domains when analyzed using TMHMM and SOSUI prediction software.

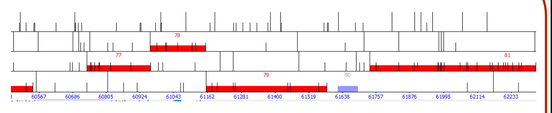

Figure 7. Knocker has one tRNA gene in the right arm of its genome. The tRNA recognizes the codon CAA specifying the amino acid glutamine even though it has a codon bias for using CAG to specify glutamine.