# Genomic Annotation of Bacteriophages Clayda5, GShelby23, Santhid, and Wrigley

**Abigail M. Bastian*, Marcus J. Blankespoor*, Gideon K. Fynaardt*, Sadie A. Gilmeister*, Emily E. Hurley*, Madison Jones*, Erika J. McKenney*, Alexa N. Olguin*, Micah N. Rens*, Garrett Snyder*, Anneka E. Sterk*, Sophie M. Swart*, Chelsea L. Thurm*, Alaena Trevino*, Ashley N. Van Egdom*, Morgan C. Veach*, Lauren R. Pavich, Byron Noordewier, and Sara S. Tolsma**

*Authors contributed equally to this work

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

## Abstract

We annotated the genomes of four recently discovered Actinobacteriophages. **Clayda5** and **GShelby23** were isolated on *Microbacterium foliorum* NRRL B-24224. Clayda5 is a lytic, cluster EB phage, one of only 47 discovered to date. It has 10 base pair 3' sticky overhanging ends and a GC content is 67.2%. It has 70 protein-coding genes and two tRNA genes in its 39,894 bp genome. Clayda5 was purified from soil collected in Hull, IA. GShelby23 was isolated from soil collected in Storm Lake, IA. It is a cluster EM phage, one of only six discovered to date. Its genome is circularly permuted and 53,603 bp long. Its GC content is 64.8%. **Santhid** and **Wrigley** are phages that infect *Gordonia terrae* 3612. Santhid is a cluster DY phage, one of only five discovered to date. It was isolated from soil collected in Orange City, IA. Its genome is 39,295 bp long and includes 60 protein-coding genes. Its GC content is 67.7% and has 10 base pair 3' sticky overhanging ends. Wrigley was isolated using an enrichment protocol from soil collected in Johnston, IA. It is a cluster CY phage, one of only 17 discovered to date. It is a temperate phage whose genome is 51,878 bp long and includes 81 protein-coding genes. It has 10 base pair 3' sticky overhanging ends and a GC content of 66.3%..

## Introduction

The most abundant infectious particles on Earth are bacteriophages (Sulakvelidze, 2011). Bacteriophages or, simply, phages are viruses that infect bacteria. They are categorized by the host they infect. Actinobacteriophages are phages that infect bacteria in the phylum Actinobacteria. They have double-stranded, DNA genomes packaged in an icosahedral protein capsid. They are tailed, thus classified as caudovirales and can be divided into three families: Myoviridae (long contractile tails), Siphoviridae (long noncontractile tails), and Podoviridae (short noncontractile tails).

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 since 2016. This work reports the discovery, characterization, sequencing and annotation of four novel actinobacteriophages: Clayda5, GShelby23, Wrigley, and Santhid. Clayda5 and GShelby23 infect *Microbacterium foliorum* while Wrigley and Santhid infect *Gordonia terrae*. All four appear to be Siphoviridae. Clayda5 is a lytic phage, Wrigley is temperate, but the life cycles of GShelby23 and Santhid are still under investigation.

## Discussion

We discovered and annotated four 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 the right arm of Wrigley but since cluster CY phages tend to display right arm divergence this is not surprising. We were able to annotate programmed translational frameshifts for Clayda5, Santhid, and Wrigley.

Our annotations for Clayda5, Santhid, and Wrigley have already been submitted for publication to GenBank and we will submit GShelby23 before the end of the semester. We plan to present our research at two upcoming meetings: the Iowa Academy of Science Annual Meeting and the annual SEA-PHAGES Symposium. We are also writing a Microbiology Resource Announcement to publish the results of our annotation work. We will submit that paper early in the summer of 2022.

## Results: Clayda5

**Clayda5** is a Cluster EB phage. It infects the host *Microbacterium foliorum* and was discovered from soil collected in Hull, IA. Its 39,894 bp linear genome has 3' sticky ends (ACTCCCGGCA) and a GC content of 67.2%. There are 47 members in the EB cluster. It is a lytic phage as seen by its clear plaque morphology and lack of an identifiable integrase gene in its annotated genome. We found that its genome includes 70 protein-coding genes and two tRNA genes. Using bioinformatics, we were able to assign gene functions to 33 of the 70 protein-coding genes.

*Figure 1. Plaque picture of Clayda5 shows small to medium-sized, clear plaques, typical of a lytic phage.*


*Figure 2. The tail assembly chaperone gene of Clayda5 uses a programmed translational frameshift to regulate gene expression.*

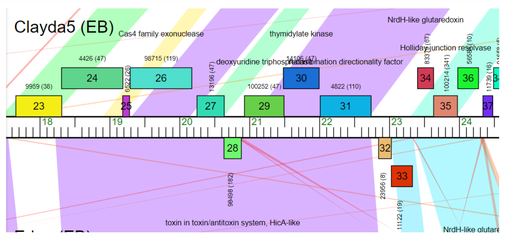


*Figure 3. Clayda5's genome includes elements of a toxin-antitoxin system (HicA-like).*

## Results: GShelby23

**GShelby23** is a Cluster EM phage, which contains six members to date. Gshelby23 was discovered in soil collected in Storm Lake, IA. It infects *Microbacterium foliorum*. Its genome was sequenced and found to be 53.603 bp in length with circularly permuted ends. Its GC content is 64.8%. Its plaque morphology suggests that it might have a temperate life cycle but we need to do more work to confirm this. We are in the process of annotating Gshelby23's genome.

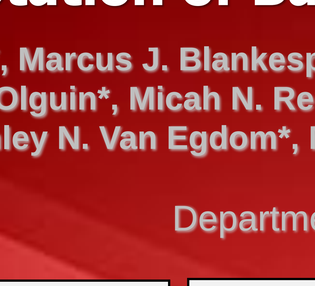

*Figure 4. Plaque picture of GShelby23 shows small to medium, clear plaques with a slight bullseye in the center of the plaques on a lawn of M foliorum.*

## Results: Publications

**Table 1. Published annotations in GenBank**

| Phage | GenBank Accession Number |
|---|---|
| Clayda5 | OM818329 |
| GShelby23 | TBD |
| Santhid | OM818327 |
| Wrigley | TBD |

## Results: Santhid

**Santhid** is a Cluster DY phage, contains five members to date. It was discovered in 2021 from soil collected in Orange City, IA. It infects *Gordonia terrae*. Its genome was sequenced and found to be 39295 bp long with 3' sticky overhanging ends (TCCGGAGGTA). Its GC content is 67.7%. Santhid's genome contains 60 genes, all but three of which are forward genes. We used bioinformatics to assign functions to 36 of Santhid's genes.

*Figure 5. Plaque picture of Santhid shows small, slightly cloudy plaques.*

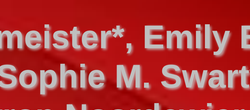


*Figure 6. The tail assembly chaperone gene of Santhid uses a programmed translational frameshift to regulate gene expression, annotated here in Rejya (also DY) but the same in Santhid.*

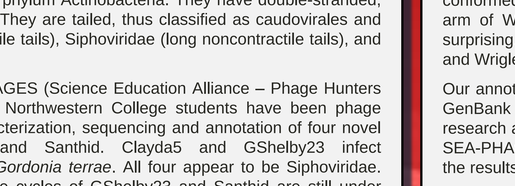

*Figure 7. Santhid's genome includes a tyrosine integrase (gene 30) suggesting that Santhid may be a temperate phage.*

## Results: Wrigley

Wrigley was discovered in 2021 in soil collected in Johnston, IA. It infects the host *Gordonia terrae*. Its genome was sequenced and found to consist of 51878 bp with 3' sticky ends (CGTATGGCAT). It was classified as a Cluster CY phage—one of only 17 CY phages identified to date. Wrigley appears to be temperate by plaque morphology. Wrigley's genome includes 81 protein-coding genes, of which we were able to assign functions to 36 using bioinformatics.

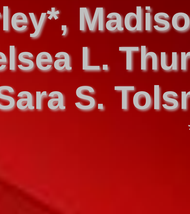

*Figure 8. Plaque picture of Wrigley shows medium-sized plaques with a bullseye appearance typical of temperate phages.*

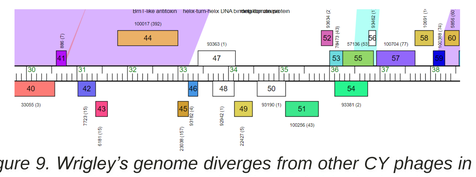

*Figure 9. Wrigley's genome diverges from other CY phages in its right arm and includes several orphams (white in the phamerator map). Many of the genes in this divergent right arm include genes involved in toxin/antitoxin systems and transcriptional regulators*

## Future Directions

- **Complete annotation of GShelby23**
- **Submit MRA**
- **Visualize phages by transmission electron microscopy**

## Materials and Methods

Clayda5 was discovered Ashley Van Egdom in 2021. It infects the host *Microbacterium foliorum* NRRL B-24224 . GShelby23 was discovered by Sadie Gilmeister in 2021 on *Microbacterium foliorum* NRRL B-24224. Santhid was discovered b yKrista Riensche, Ashley VanEgdom, and Gideon Fynaardt from soil collected in Orange City, IA. It infects the host *Gordonia terrae 3612*. Wrigley also infects *Gordonia terrae 3612*. It was isolated from soil collected in Johnston, IA in 2021. All phages were isolated by direct isolation (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). 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**

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

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.