Complete Genome Sequences of Nine Cluster A Mycobacteriophages: AgentM, Ajay, Aragog, Archetta, ForGetIt, Koko, Ph8s, Phlorence, and Wilkins

Rachel M. Lipman*, Amanda C. Freise*, Andrew Kapinos, Canela Torres, Krisanavane Reddi, and Jordan Moberg Parker[#]

Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, Los Angeles, California, USA

*Indicates equal contributions
[#]Address Correspondence to Jordan Moberg Parker, jmobergparker@ucla.edu

Abstract

Background: Bacteriophages are ubiquitous, highly diverse, and relatively understudied. Growing interest in phage therapy has underscored the importance of isolating and characterizing novel bacteriophages. The Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program aims to address this need by involving undergraduates around the world in authentic research.

Materials and Methods: Nine novel mycobacteriophages - AgentM, Ajay, Aragog, Archetta, ForGetIt, Koko, Ph8s, Phlorence, and Wilkins - were isolated from soil samples in southern California using host *Mycobacterium smegmatis* mc^{2}155. Each purified phage was characterized using transmission electron microscopy and genome sequencing and annotation.

Results: All nine bacteriophages were placed into mycobacteriophage Cluster A based on nucleotide similarity with other phages. The average genome length of all nine

phages was 50,706 bp. On average, each phage had 87 total coding genes and GC content between 60-63%, consistent with that of other Cluster A phages. Transmission electron microscopy of phage particles revealed they had icosahedral heads and long, flexible tails, consistent with that of the *Siphoviridae* family. Plaque morphology and genome analysis confirms the nine novel phages are temperate as expected of Cluster A phages.

Conclusions: AgentM, Ajay, Aragog, Archetta, ForGetIt, Koko, Ph8s, Phlorence, and Wilkins are all mycobacteriophages that belong to the *Siphoviridae* family. Comparative genomic analyses revealed genetic mosaicism and diversity among these Cluster A phages. The discovery of these novel phages expands on the existing library of mycobacteriophage genomes.

Introduction

The Howard Hughes Medical Institute Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) initiative is an international research program that harnesses the power of undergraduate researchers to isolate, sequence, and characterize actinobacteriophages across the world for potential real-world applications [1]. To date, students in the program have isolated over 18,600 actinobacteriophages and sequenced over 3,500 (https://PhagesDB.org). These phages have already begun to have translational impact; several novel bacteriophages previously discovered by students involved in the SEA-PHAGES program were recently used to treat a disseminated drug-resistant *Mycobacterium abscessus* infection in a cystic fibrosis patient [2].

Here we report the isolation and characterization of nine novel Cluster A mycobacteriophages – AgentM, Ajay, Aragog, Archetta, ForGetIt, Koko, Ph8s, Phlorence, and Wilkins – by students at the University of California, Los Angeles in collaboration with SEA-PHAGES. These bacteriophages were isolated using bacterial host *Mycobacterium smegmatis* mc^{2}155, a nonpathogenic and easily experimentally manipulated species in the *Mycobacterium* genus.

Materials and methods

Phage collection, isolation, and purification

All soil samples except that which produced phage Ajay were collected from the greater Los Angeles, CA area. Phage Ajay was isolated from a sample obtained 40 miles further east in Eastvale, CA. Phage extraction occurred through either direct or enriched isolation, as detailed in the SEA-PHAGES Discovery Guide (https://seaphagesphagediscoveryguide.helpdocsonline.com/).

Phage Ajay was isolated via direct isolation by combining soil and 10X 7H9 Middlebrook enrichment broth (supplemented with albumin, dextrose, carbenicillin [50 µg/ml] and cycloheximide [10 µg/ml]) in a 15 mL conical tube and shaking at 250 rpm for 1-2 hours. After incubation, the mixture was filter sterilized. 500 µL of the mixture was added to 250 µL of host bacterium *M. smegmatis* mc^{2}155 and incubated for 5-10 minutes. Lastly, 3 mL of molten top agar was added to the inoculated host tube, mixed, and then poured onto an agar plate. The plates were incubated at 30°C and observed for plaques 24-48 hours later. A turbid plaque was picked and purified via two

consecutive plaque assays to produce a single phage population. Purified lysates were amplified via additional plaque assays to produce a high-titer phage lysate.

The other eight phages described were discovered through enriched isolation. Soil samples were similarly incubated with enrichment broth and a 48-hour culture of *M. smegmatis* and shaken for 2 days. After incubation, the sample mixture was centrifuged at 2,000 x g for 10 minutes. Supernatant was removed and filter sterilized. 10 µL of each supernatant sample was spotted on to a lawn of *M. smegmatis*. Plates were incubated at 30 or 35°C and plaques were picked and purified as above.

Transmission electron microscopy

Phage lysates were pipetted onto a carbon-coated Formvar grid. Grids were then stained with 1% uranyl acetate (UA), dried, and viewed using a FEI T12 electron microscope (Thermo-Fisher, USA). Each phage was observed at multiple magnifications.

DNA extraction and genome sequencing

Viral DNA was extracted and purified using the Promega Wizard DNA Clean-Up Kit (Product #A7280). Whole genome sequencing was conducted at the North Carolina State Genomic Sciences Laboratory, the Pittsburgh Bacteriophage Institute, or the UCLA Genotyping and Sequencing Core, using either Illumina sequencing or 454 GS FLX pyrosequencing. The sequencing libraries were created using the NEBNext® Ultra™ II DNA Library Prep Kit (New England Biolabs, MA, USA). Coverage values are reported in Table 1. Genome assembly was performed as described previously [3].

Genome Annotation

All phage genomes underwent auto and manual annotation. DNA Master (http://cobamide2.bio.pitt.edu/computer.htm) was used for auto-annotation. Glimmer [4] and GeneMark [5] were used to predict coding potential and identify open reading frames. ARAGORN [6] and tRNAScan-SE [7] were used to predict and trim tRNAs for each phage. Manual annotation was performed using PECAAN (http://pecaan.kbrinsgd.org), Phamerator [8] and Starterator [9]. NCBI BLAST [10], HHPred [11], and the Conserved Domain Database [12] were used to predict gene functions. TMHMM [13] was used to predict transmembrane domains.

Results

TEM imaging of each phage revealed that they all belonged to the *Siphoviridae* family, based on their long, flexible tails (Figure 1).

The nine novel mycobacteriophages were placed into Cluster A based on their nucleotide similarity with other phages. Ajay and Wilkins were placed into subcluster A1, Ph8s into A2, Agent M, Aragog, Archetta, ForGetIt, and Phlorence into A5, and Koko into A6. All nine phages share a 3' sticky end overhang of 10 base pairs, and phages in the same subcluster have the same overhang sequence (Table 1). The average genome size was 50,706 bp. The nine phages had a median of 84 total protein coding genes (range: 76-102), 0-3 tRNAs, and an average GC content of 61.7% (range: 60.6-63.9%) (Table 2). Each of these values is consistent with overall Cluster A genomes (https://phagesdb.org/clusters/A/). In all nine phages, structural and assembly genes

were conserved on the left arm of the genome [14] (Figure 2). The right arms of the genomes contained a variety of genes including those with DNA replication, processing, and binding functions, membrane proteins, DNA-binding proteins, and numerous genes with no currently known function. The A1 and A5 phages each contained lysin A, lysin B, and holin; the A2 and A6 phages did not have identifiable lysin B genes. Each genome contained an integrase and an immunity repressor, supporting the designation of these phages as temperate.

Discussion

Cluster A, which at the time of publication comprises 662 members, makes up the largest cluster of actinobacteriophages. It has substantial intracluster variation, including unusual and diverse immunity systems [14], and is relatively genetically isolated from other clusters [15,16]. In the present study, nine novel Cluster A mycobacteriophages AgentM, Ajay, Aragog, Archetta, ForGetIt, Koko, Ph8s, Phlorence, and Wilkins were discovered, sequenced, and characterized.

TEM confirmed that these phages, similar to other Cluster A phages, were members of the Siphoviridae family. Genome sequencing and annotation revealed that phages within the same subcluster retained a high degree of genetic similarity, and that all nine phages shared at least some genes with each other. Phages Ph8s and Koko, from subclusters A2 and A6, appear to be more similar to each other than they are to any of the phages from clusters A1 or A5. The isolation and characterization of these phages contributes additional knowledge to the expanding database of actinobacteriophages, and further supports current knowledge of Cluster A phages.

Acknowledgments

Author contributions

Author disclosure statement

The authors declare that there is no conflict of interest regarding the publication of this article.

Figures and Tables

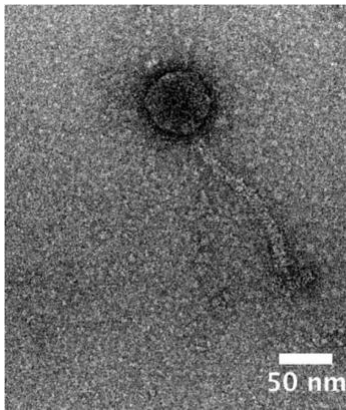
Figure 1. Isolated Cluster A mycobacteriophages have *Siphoviridae* morphology.
Representative image of Cluster A phage Wilkins at 26,000X magnification, showing its long flexible tail and icosahedral head. All phages described had similar features.
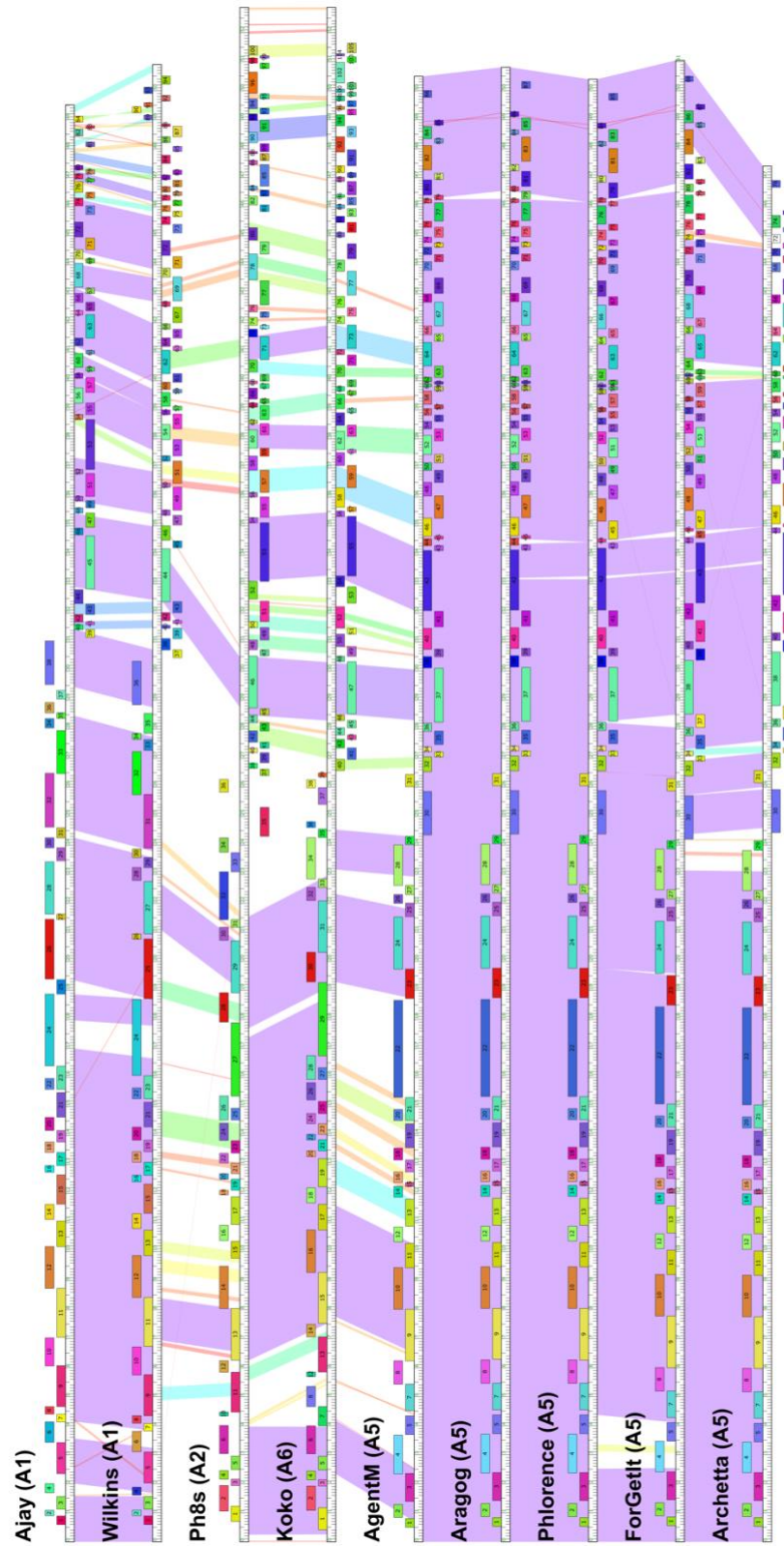
Figure 2. Phamerator map of nine novel Cluster A mycobacteriophages.
Subclusters are arranged in order of approximate genomic similarity.

Table 1. Novel Cluster A mycobacteriophage sequencing information

Phage Name	Cluster	GenBank Accession no.	Sequencing Coverage (x)	3' Sticky Overhang Sequence
Ajay	A1	MN062710	4,218[a]	CGGATGGTAA
Wilkins	A1	MG099951	32[c]	CGGATGGTAA
Ph8s	A2	MG099947	14,824[b]	CGGTCGGTTA
AgentM	A5	MG099934	62[c]	CGGGAGGTAA
Aragog	A5	MG099937	76[c]	CGGGAGGTAA
Archetta	A5	MG099938	103[c]	CGGGAGGTAA
ForGetIt	A5	MG099944	144[c]	CGGGAGGTAA
Phlorence	A5	MG099949	92[c]	CGGGAGGTAA
Koko	A6	MG099945	11,617[b]	CGGTCGGTTA

[a] Sequencing performed at the NC State Genomic Sciences Laboratory (Illumina)
[b] Sequencing performed at the Pittsburgh Bacteriophage Institute (Illumina)
[c] Sequencing performed at the UCLA Genotyping and Sequencing Core (454 GS FLX pyrosequencing)

Table 2. Genome characteristics of phages

Phage Name	Cluster	Genome Length (bp)	GC Content (%)	No. of Protein Coding Genes	No. of tRNAs
Ajay	A1	49,516	63.6	84	0
Wilkins	A1	50,907	63.9	93	0
Ph8s	A2	52,874	62.6	98	3
AgentM	A5	50,503	60.9	82	2
Aragog	A5	50,812	60.8	82	1
Archetta	A5	47,409	60.6	76	0
ForGetIt	A5	51,050	60.6	85	1
Phlorence	A5	50,403	60.9	82	1
Koko	A6	52,879	61.3	102	3

Data availability. The genome sequences of Mycobacterium phages Ajay, AgentM, Aragog, Archetta, ForGetIt, Koko, Ph8s, Phlorence, and Wilkins have been deposited in GenBank under the accession numbers MN062710, MG099934, MG099937, MG099938, MG099944, MG099945, MG099947, MG099949, and MG099951 (NCBI GenBank).

References

1. Hanauer DI, Graham MJ, Sea-Phages, Betancur L, Bobrownicki A, Cresawn SG, et al. An inclusive Research Education Community (iREC): Impact of the SEA-PHAGES program on research outcomes and student learning. PNAS. 2017;114: 13531–13536. doi:10.1073/pnas.1718188115

2. Dedrick RM, Guerrero-Bustamante CA, Garlena RA, Russell DA, Ford K, Harris K, et al. Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant Mycobacterium abscessus. Nat Med. 2019;25: 730–733. doi:10.1038/s41591-019-0437-z

3. Russell DA. Sequencing, Assembling, and Finishing Complete Bacteriophage Genomes. In: Clokie MRJ, Kropinski AM, Lavigne R, editors. Bacteriophages: Methods and Protocols, Volume 3. New York, NY: Springer New York; 2018. pp. 109–125. doi:10.1007/978-1-4939-7343-9_9

4. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. Improved microbial gene identification with GLIMMER. Nucl Acids Res. 1999;27: 4636–4641. doi:10.1093/nar/27.23.4636

5. Besemer J, Borodovsky M. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res. 2005;33: W451-454. doi:10.1093/nar/gki487

6. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic acids research. 2004;32: 11–16.

7. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 2016;44: W54–W57. doi:10.1093/nar/gkw413

8. Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. BMC Bioinformatics. 2011;12: 395. doi:10.1186/1471-2105-12-395

9. Pope WH, Jacobs-Sera D, Russell DA, Cresawn SG, Hatfull GF. Starterator. SEA-PHAGES Bioinformatics Guide. 2017. Available: https://seaphagesbioinformatics.helpdocsonline.com/article-23

10. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology. 1990;215: 403–410. doi:10.1016/S0022-2836(05)80360-2

11. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Research. 2005;33: W244–W248. doi:10.1093/nar/gki408

12. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. Nucleic Acids Res. 2015;43: D222–D226. doi:10.1093/nar/gku1221

13. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305: 567–580. doi:10.1006/jmbi.2000.4315

14. Mavrich TN, Hatfull GF. Evolution of Superinfection Immunity in Cluster A Mycobacteriophages. mBio. 2019;10: e00971-19. doi:10.1128/mBio.00971-19

15. Pope WH, Bowman CA, Russell DA, Jacobs-Sera D, Asai DJ, Cresawn SG, et al. Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. Kolter R, editor. eLife. 2015;4: e06416. doi:10.7554/eLife.06416

16. Hatfull GF. Actinobacteriophages: Genomics, Dynamics, and Applications. Annual Review of Virology. 2020;7: 37–61. doi:10.1146/annurev-virology-122019-070009