Complete Genome Sequences of Eleven Cluster DJ *Gordonia rubripertincta* Bacteriophages

Ping An[1], Meredith Barbieri[1], Dawn C. Bisi[1], Kristen Butela[1], Aimee H. Danley[1], Christina Grogan[1], Ben Hilldorfer[1], Bhaswati Manish[2], Kevin K. McDade[1], Eliza J. Neumann[1], Aparna Nigam[1], Rachelle Patterson[1], Ananya Saini[1], Samantha Walker[1], Otto G. Williams[1], Marcie H. Warner[1§]

[1]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, United States

[2]Department of Math and Natural Sciences, Metropolitan Community College, Omaha, Nebraska, United States

[§]To whom correspondence should be addressed: mwarner1@pitt.edu

Abstract

Eleven novel cluster DJ *Gordonia rubripertincta* phages were isolated from soil collected in Pittsburgh, PA and Omaha, NE. The genomes of these phages have an average length of 60,185 bp and contain an average of 91 predicted genes. Genes with putative roles in structure, assembly, lysis, and nucleic acid processing were identified. Cluster DJ phages have an unusually low average G+C content (51.6%) compared to *G. rubripertincta* (~67%). These phages are also unusual in the organization of genes involved in lysis, the inclusion of two related major tail protein genes, and the absence of a candidate tail assembly chaperone gene.
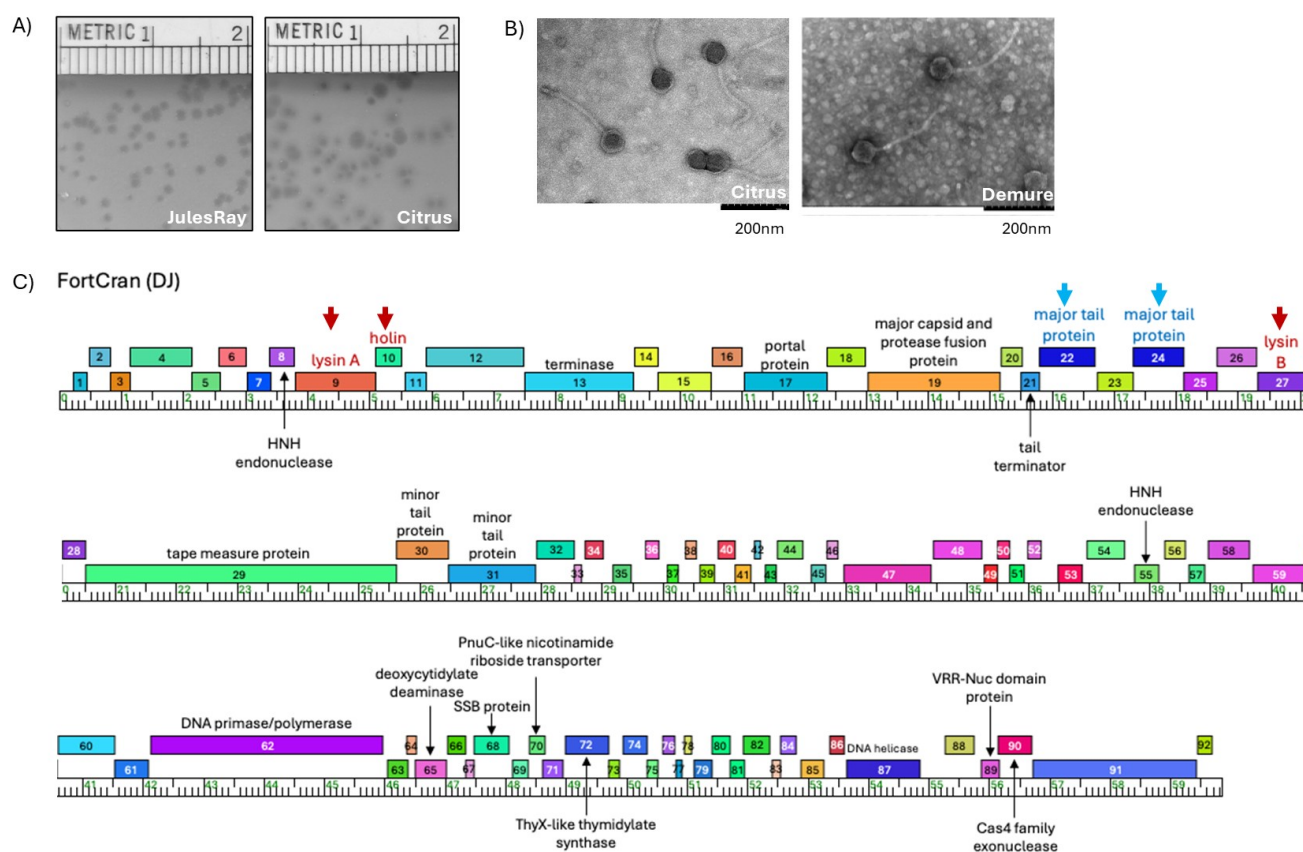
Figure 1. The morphological and genomic characteristics of 11 novel cluster DJ phages:

(A) DJ phages in this group tend to form clear plaques represented by the plaque morphologies of phages JulesRay (Left) and Citrus (Right). A culture of host cell *Gordonia rubripertincta* was inoculated with samples of JulesRay and Citrus separately. The phage and cell mixture were then combined with 0.4% PYCa top agar and overlaid on top of PYCa agar. Plates were incubated at 30°C for 3 days. Sizes of the plaques range from 0.5 to 2mm in diameter. Numbers indicated by the rulers are in cm. (B) Morphology of viral particles was visualized using a HITACHI 7800 Transmission Electron Microscope following negative staining using 1% wt/vol uranyl acetate. Two representative pictures from phages Citrus (Left) and Demure (Right) are shown. A scale bar of 200nm length was included at the bottom right of each picture. (C) Several distinct genome features of DJ cluster phages were annotated using the Phamerator map of phage FortCran as a representative. Colored boxes represent protein coding genes in the genome. Numbers inside the boxes indicate gene numbers. Putative functions are listed above or below the corresponding gene boxes. The measuring tape indicates

nucleotide coordinates and genes transcribed in the forward direction are positioned above the measuring tape. Genes involved in lysis were indicated using dark red arrows and text labels and the two major tail protein genes were indicated with blue arrows and labels. Additionally, note the lack of tail assembly chaperone gene and tRNA genes in the genome.

Description

Due to advances in genome sequencing technology, the number of bacteriophage genomes that have been completely sequenced has increased rapidly in recent years (Hatfull, 2020, 2021; Heller et al., 2024; Pope et al., 2020). Despite this expeditious pace of identification and sequencing, novel phages are still frequently predicted to contain genes with no recognizable homology to those in previously identified phages; this suggests that phages still represent a substantial source of unexplored biological functions. We have focused our efforts on the identification of phages that infect the bacterial host *Gordonia rubripertincta*. This bacterium has potential applications in removal of pollutants from ecosystems (Frantsuzova et al., 2023); an understanding of the abundance and diversity of phages capable of infecting this organism may assist with assessing its viability as a remediation agent in natural environments. Furthermore, the genus *Gordonia* contains several opportunistic human pathogens (Barthel et al., 2023; Lesens et al., 2000; Vergov et al., 2023; Zheng et al., 2024) for which phage therapy may be a suitable treatment strategy.

This report focuses on the discovery and characterization of eleven novel cluster DJ *Gordonia rubripertincta* phages: Arataki, BetterYeti, Captrips, Citrus, Dancer, Demure, FortCran, Gusicorn, JulesRay, NoPickles, and ShaggyRogers. These phages were isolated from soil samples collected in Pittsburgh, Pennsylvania and Omaha, Nebraska during 2023 and 2024 (Table 1). Soil samples were first resuspended in phage buffer (10mM Tris-HCl, pH7.5, 68mM NaCl, 10mM $MgSO_4$, 1mM $CaCl_2$, 10% Glycerol) and agitated at 250 rpm at 30°C for 15 minutes to facilitate separation of phage particles from soil particulates. Samples were then centrifuged, and the supernatant was passed through a 0.22 μm filter to remove residual soil particulates and bacterial cells. For phages identified via direct isolation (Arataki, BetterYeti, Demure, NoPickles, and ShaggyRogers), the resulting filtrate was mixed with PYCa top agar (1.5% Peptone, 0.1% Yeast extract, 4.5mM $CaCl_2$, 0.1% Dextrose, 0.4% agar) containing *G. rubripertincta* NRRL B-16540 and overlaid on a solid PYCa agar (1.5% Peptone, 0.1% Yeast extract, 4.5mM $CaCl_2$, 0.1% Dextrose, 1.5% agar) plate. The resulting plates were incubated at 30°C for 3 days to allow for plaque formation. For phages identified utilizing an enrichment method (Captrips, Citrus, Dancer, FortCran, Gusicorn, and JulesRay), *G. rubripertincta* NRRL B-16540 and liquid PYCa were added to the soil filtrates, and the mixture was incubated at 30°C for 3 days with shaking to allow amplification of any *G. rubripertincta* phage in the sample. The cultured mixture was refiltered, and the supernatant was plated on a lawn of *G. rubripertincta* host cells embedded in PYCa top agar overlaid on a solid PYCa plate. Resulting plates were incubated at 30°C for 3 days to assay for host cell lysis due to phage activity. High titer phage lysates were prepared following two to three rounds of plaque purification. Phages in this group produced clear plaques, indictive of a lytic infectious cycle, that are frequently surrounded by haloes (Figure 1A). Transmission electron micrographs were obtained using a HITACHI 7800 Transmission Electron Microscope following staining with 1% uranyl acetate for 8 of the phages described in this report. All electron micrographs for this set of phages exhibit an icosahedral capsid ranging from 60 to 70nm in diameter and a flexible tail ranging from 235 to 260nm in length consistent with the morphology of siphoviruses (n=1 for each of the imaged phages, Figure 1B).

To obtain viral genomic information, viral DNA was isolated from high-titer phage lysates using the Promega Wizard DNA Cleanup Kit. Individual libraries were constructed using the NEBNext Ultra II FS Kit and an Illumina MiSeq sequencer (v3 reagents) was used for sequencing. The approximate shotgun coverages of single-end 100 or 150 base reads varied between 955 to 15,696-fold (Table 1). For each viral genome, raw reads were assembled using Newbler v2.9 with default parameters (Russell, 2017). Genomes were manually checked for completeness using Consed v29 (Gordon & Green, 2013). The resulting genomes had an average length of 60,185 bp ranging from 59,584 to 61,136 bp, with an average G+C content of 51.6% (range 51.3% to 52%) (Table 1). All 11 phage genomes terminate in 9 base (CGCCGCTCT) 3' single stranded overhangs.

Initial auto-annotation was performed using GLIMMER v3.02 (Delcher et al., 1999) and GeneMark v2.5 (Besemer & Borodovsky, 2005) and manually refined using Phamerator v580 (Cresawn et al., 2011), DNA Master v5.23.6 (cobamide2.bio.pitt.edu), Starterator v587 (github.com/SEA-PHAGES/starterator), and PECAAN (discover.kbrinsgd.org). Functional assignments were evaluated using BLASTp v 2.2.26 (using NCBI nonredundant database and the database at phagesdb.org) (Altschul et al., 1990) and HHpred (using PDB_mmCIF70, PfamA, and NCBI Conserved Domain databases) (Soding et al., 2005). The presence of putative tRNAs was evaluated using Aragorn v1.2.38 (Laslett & Canback, 2004) and tRNAscan-SE v2.0 (Lowe & Chan, 2016). DeepTMHMM v1.0.42 (Hallgren et al., 2022) was utilized to evaluate the presence of predicted transmembrane domains in putative gene products. Default parameters were utilized for all programs other than DNA Master which was set according to instructions in the SEA-PHAGES bioinformatics guide (Pope, Jacobs-Sera et al., 2017).

All predicted genes in the eleven novel cluster DJ phages described in this report are transcribed unidirectionally. The finalized annotations of these genomes include between 90 and 93 protein coding genes (Table 1). Additionally, between

28 to 33 functional assignments were made in each genome (Table 1). Proteins with predicted roles in structure and assembly are interspersed throughout the 5' end of the genome. The 3' half of the genome contains many predicted genes that can be assigned putative functions in nucleic acid modification, processing, and metabolism. None of the genomes described in this report contain any predicted tRNA genes or genes associated with lysogeny.

Among actinobacteriophages, cluster DJ phages have previously been reported to have an unusual arrangement of lysis factors. The predicted lysin A and holin genes in these phages are unusually skewed toward the 5' end of the genome at a position upstream of the predicted structural protein genes (Pollenz et al., 2022). Furthermore, while the genes encoding lysin A, holin, and lysin B are often found as a cassette of adjacent genes in actinobacteriophages (Pollenz et al., 2022), the gene that encodes the predicted lysin B in cluster DJ phages is found approximately 15 kb and 16 intervening genes downstream of the genes predicted to encode lysin A and holin (Figure 1C). All eleven phages reported here display this unusual arrangement of lysis factors.

Another atypical feature of cluster DJ phage genomes is the identification of more than one putative major tail protein gene (Figure 1C) (Benson et al., 2025). All 56 currently sequenced cluster DJ phages contain two major tail protein genes (FortCran genes 22 and 24) separated by a single gene for which a function has not been identified (FortCran gene 23). To explore the similarities and differences between the FortCran major tail protein genes, we conducted several comparative analyses. PhaMMseqs-grouping (Gauthier et al., 2022) of proteins with at least 35% amino acid identity and 80% sequence coverage, assigns both FortCran major tail proteins to the same "phamily" suggesting they likely share an evolutionary origin. We more closely examined the sequence variation between these genes using BLASTp; this analysis revealed that the FortCran major tail proteins share 43% amino acid identity (60% similarity) and that gene *22* encodes an additional 51 C-terminal amino acids. Further analysis demonstrates that the G+C content of gene 24 (49%) closely reflects the average G+C content of the eleven cluster DJ phages reported here (average 51.6%) while gene 22's G+C content (63%) more closely aligns with the G+C content of *G. rubripertincta* (~67%) (Frantsuzova et al., 2025; Parveen et al., 2021). The sequence and G+C content variation between these genes suggests gene 22 was likely acquired as the result of horizontal transfer. It will be interesting to further explore the evolutionary histories of these genes and determine whether they perform overlapping or divergent functions.

Cluster DJ phages are also unusual in that, despite 56 manually annotated cluster members, the putative gene that encodes the tail assembly chaperones has not been identified (Figure 1C). Necessary for proper tail assembly, the tail assembly chaperone gene in actinobacteriophages is typically located immediately, or within a few genes, upstream of the gene encoding the tape measure protein (Christie et al., 2002; Levin et al., 1993; Xu et al., 2004). Interestingly, the only predicted gene in FortCran that aligns to a putative tail assembly chaperone gene is the single gene between the two major tail protein genes (gene 23 of FortCran). However, the homology and alignment are too incomplete for a functional designation and the gene lacks the characteristic slippery sequence that distinguishes this family of genes (Baranov et al., 2006; Christie et al., 2002; Levin et al., 1993; Xu et al., 2004). All cluster DJ members contain a predicted gene in this same gene family in the same position. Additional analyses will be required to determine if this protein has tail assembly chaperone activity, the tail assembly chaperones are located in an atypical genomic location, or if tail assembly occurs by an unexplored mechanism in the cluster DJ phages (Vladimirov et al., 2022).

Phage	Sample location (city, state, GPS)	Approx. read coverage (fold, x)	# of single end reads (size in bp)	Genome length (bp)	%GC content	# of ORFs (# with predicted function)	capsid diameter (nm)	tail length (nm)
Arataki	Pittsburgh, PA 40.446795 N, 79.964165 W	15,696	10,605,554 (100)	60,488	51.60%	91 (31)	66	259
BetterYeti	Pittsburgh, PA 40.446884 N, 79.952998 W	3,659	2,690,994 (100)	60,438	51.90%	90 (30)	62	246
Captrips	Omaha, NE 41.171194 N, 96.032694 W	955	401,741 (150)	59,629	51.60%	91 (30)	ND	ND

Citrus	Pittsburgh, PA 40.442000 N, 79.956000 W	3,150	1,960,987 (100)	61,136	51.30%	90 (32)	68	240
Dancer	Pittsburgh, PA 40.260000 N, 79.570000 W	1,759	1,326,262 (100)	59,584	51.50%	92 (30)	ND	ND
Demure	Pittsburgh, PA 40.264600 N, 79.571100 W	4,305	2,724,930 (100)	60,622	51.60%	90 (28)	67	235
FortCran	Omaha, NE 41.223607 N, 95.990680 W	1,794	753,910 (150)	59,848	51.50%	92 (29)	70	260
Gusicorn	Pittsburgh, PA 40.440000 N, 79.960000 W	6,583	3,978,858 (100)	60,438	52.00%	92 (32)	60	240
JulesRay	Pittsburgh, PA 40.445983 N, 79.953041 W	3,844	3593453 (100)	59,943	51.60%	92 (32)	58	260
NoPickles	Pittsburgh, PA 40.443611 N, 79.955278 W	4,426	2,766,385 (100)	59,701	51.30%	93 (33)	ND	ND
ShaggyRogers	Pittsburgh, PA 40.443611 N, 79.955278 W	4,393	2,834,257 (100)	60,204	52.00%	91 (31)	66	257
Average (Range)				60,185 (59584-61136)	51.6 (51.3-52.0)	~91 (90-93) & ~31 (28-33)	(58-70)	(235-260)

Nucleotide sequence accession numbers

Arataki is available at GenBank with Accession No. PV915802 and Sequence Read Archive (SRA) No. SRX29486052. BetterYeti is available at GenBank with Accession No. PV915818 and Sequence Read Archive (SRA) No. SRX29486053. Captrips is available at GenBank with Accession No. PP978771 and Sequence Read Archive (SRA) No. SRX24892111. Citrus is available at GenBank with Accession No. PV915797 and Sequence Read Archive (SRA) No. SRX29486054. Dancer is available at GenBank with Accession No. PV915891 and Sequence Read Archive (SRA) No. SRX29486055. Demure is available at GenBank with Sequence Read Archive (SRA) No. SRX29486056 (GenBank accession No. to be assigned). FortCran is available at GenBank with Accession No. PP978793 and Sequence Read Archive (SRA) No. SRX24892115. Gusicorn is available at GenBank with Accession No. PV915881 and Sequence Read Archive (SRA) No. SRX29486057. JulesRay is available at GenBank with Accession No. PV915826 and Sequence Read Archive (SRA) No. SRX29486058. NoPickles is available at GenBank with Accession No. PV915879 and Sequence Read Archive (SRA) No. SRX29486052. ShaggyRogers is available at GenBank with Accession No. PV915809 and Sequence Read Archive (SRA) No. SRX29486060.

References

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

Baranov PV, Fayet O, Hendrix RW, Atkins JF. 2006. Recoding in bacteriophages and bacterial IS elements. Trends in Genetics 22: 174-181. DOI: 10.1016/j.tig.2006.01.005

Barthel A, Ursenbach A, Kaeuffer C, Koebel C, Gravet A, De Briel D, et al., Gerber. 2023. Characteristics and Treatment of *Gordonia* spp. Bacteremia, France. Emerging Infectious Diseases 29. DOI: 10.3201/eid2905.221901

Benson E, Blount M, Chauhan S, Ehrhart J, Foster A, Ingber A, et al., Monti D. 2025. Complete Genome Sequence of the Cluster DJ Actinobacteriophage, Petito, isolated on the host *Gordonia rubripertinca*. microPublication Biology. 10.17912/micropub.biology.001456. DOI: 10.17912/micropub.biology.001456

Besemer J, Borodovsky M. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Research 33: W451-W454. DOI: 10.1093/nar/gki487

Christie GE, Temple LM, Bartlett BA, Goodwin TS. 2002. Programmed Translational Frameshift in the Bacteriophage P2 *FETUD* Tail Gene Operon. Journal of Bacteriology 184: 6522-6531. DOI: 10.1128/JB.184.23.6522-6531.2002

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

Delcher A. 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Research 27: 4636-4641. DOI: 10.1093/nar/27.23.4636

Frantsuzova E, Bogun A, Solomentsev V, Vetrova A, Streletskii R, Solyanikova I, Delegan Y. 2023. Whole Genome Analysis and Assessment of the Metabolic Potential of Gordonia rubripertincta Strain 112, a Degrader of Aromatic and Aliphatic Compounds. Biology 12: 721. DOI: 10.3390/biology12050721

Frantsuzova E, Kocharovskaya Y, Vetrova A, Delegan Y, Bogun A. 2025. Complete genome sequence of *Gordonia rubripertincta* 112, a promising degrader of aromatic and aliphatic compounds. Microbiology Resource Announcements 14: 10.1128/mra.00114-25. DOI: 10.1128/mra.00114-25

Gauthier CH, Cresawn SG, Hatfull GF. 2022. PhaMMseqs: a new pipeline for constructing phage gene phamilies using MMseqs2. G3 Genes|Genomes|Genetics 12: 10.1093/g3journal/jkac233. DOI: 10.1093/g3journal/jkac233

Gordon D, Green P. 2013. *Consed:* a graphical editor for next-generation sequencing. Bioinformatics 29: 2936-2937. DOI: 10.1093/bioinformatics/btt515

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

Hatfull GF. 2021. Wildy Prize Lecture, 2020–2021: Who wouldn't want to discover a new virus?. Microbiology 167: 10.1099/mic.0.001094. DOI: 10.1099/mic.0.001094

Heller DM, Sivanathan V, Asai DJ, Hatfull GF. 2024. SEA-PHAGES and SEA-GENES: Advancing Virology and Science Education. Annual Review of Virology 11: 1-20. DOI: 10.1146/annurev-virology-113023-110757

Laslett D. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Research 32: 11-16. DOI: 10.1093/nar/gkh152

Lesens O, Hansmann Y, Riegel P, Heller Rm, Benaissa-Djellouli M, Martinot M, Petit Hln, Christmann D. 2000. Bacteremia and Endocarditis Caused by a *Gordonia* Species in a Patient with a Central Venous Catheter. Emerging Infectious Diseases 6: 382-385. DOI: doi.org/10.3201/eid0604.000410

Levin ME, Hendrix RW, Casjens SR. 1993. A Programmed Translational Frameshift is Required for the Synthesis of a Bacteriophage λ Tail Assembly Protein. Journal of Molecular Biology 234: 124-139. DOI: 10.1006/jmbi.1993.1568

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

Parveen S, Akhtar N, Akram J. 2021. Genomic analysis provides insights into the Gordonia sp. W3S5 taxonomy and sulfur metabolism-related genes. 3 Biotech 11: 10.1007/s13205-021-02850-4. DOI: 10.1007/s13205-021-02850-4

Pollenz RS, Bland J, Pope WH. 2022. Bioinformatic characterization of endolysins and holin-like membrane proteins in the lysis cassette of phages that infect Gordonia rubripertincta. PLOS ONE 17: e0276603. DOI: 10.1371/journal.pone.0276603

Pope WH, Butela KA, Garlena RA, Jacobs-Sera D, Russell DA, Warner MH, University of Pittsburgh SEA-PHAGES, Hatfull GF. 2020. Genome Sequences of 20 Bacteriophages Isolated on Gordonia terrae. Microbiology Resource Announcements 9: 10.1128/mra.01489-19. DOI: 10.1128/MRA.01489-19

Pope WH, Jacobs-Sera DJ, Russell DA, Cresawn SG, Hatfull GF. 2017. HHMI SEA-PHAGES Bioinformatics Guide. https://seaphagesbioinformatics.helpdocsonline.com/home

Russell DA. 2017. Sequencing, Assembling, and Finishing Complete Bacteriophage Genomes. Methods in Molecular Biology,Bacteriophages : 109-125. DOI: 10.1007/978-1-4939-7343-9_9

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

Vergov B, Angelova A, Baldzhieva A, Kalchev Y, Tsochev G, Murdjeva M. 2023. Gordonia sputi as an Arising Causative Agent of Bacteremia in Immunocompromised Comorbid Dialysis Patients—A Case Report. Healthcare 11: 2059. DOI: 10.3390/healthcare11142059

Vladimirov M, Gautam V, Davidson AR. 2022. Identification of the tail assembly chaperone genes of T4-Like phages suggests a mechanism other than translational frameshifting for biogenesis of their encoded proteins. Virology 566: 9-15. DOI: 10.1016/j.virol.2021.11.003

Xu J, Hendrix RW, Duda RL. 2004. Conserved Translational Frameshift in dsDNA Bacteriophage Tail Assembly Genes. Molecular Cell 16: 11-21. DOI: 10.1016/j.molcel.2004.09.006

Zheng W, Liu J, Bai H, Xu X, Wu L, Qin X. 2024. A Patient of Spontaneous Bacterial Peritonitis in Hepatitis C Cirrhosis Caused by Gordonia terrae: A Case Report. Infection and Drug Resistance Volume 17: 1803-1810. DOI: 10.2147/IDR.S459821

Conflicts of Interest: The authors declare that there are no conflicts of interest present.

Author Contributions: Ping An: investigation, data curation, visualization, writing - original draft, writing - review editing, conceptualization, formal analysis, project administration, resources, supervision. Meredith Barbieri: investigation, data curation, writing - review editing, formal analysis. Dawn C. Bisi: investigation, data curation, writing - review editing, formal analysis. Kristen Butela: investigation, data curation, writing - review editing, formal analysis. Aimee H. Danley: investigation, data curation, writing - review editing, formal analysis. Christina Grogan: investigation, data curation, writing - review editing, formal analysis. Ben Hilldorfer: investigation, data curation, writing - review editing, formal analysis. Bhaswati Manish: investigation, data curation, writing - review editing, formal analysis. Kevin K. McDade: investigation, data curation, writing - review editing, formal analysis. Eliza J. Neumann: investigation, data curation, writing - original draft, writing - review editing. Aparna Nigam: investigation, data curation, writing - review editing, formal analysis. Rachelle Patterson: investigation, data curation, writing - review editing, formal analysis. Ananya Saini: investigation, data curation, writing - original draft, writing - review editing. Samantha Walker: investigation, data curation, writing - original draft, writing - review editing. Otto G. Williams: investigation, data curation, writing - original draft, writing - review editing. Marcie H. Warner: conceptualization, investigation, data curation, writing - original draft, writing - review editing, formal analysis, project administration, resources, supervision, validation.

Reviewed By: Anonymous, Randall DeJong