

Creative Commons Attribution 4.0 International (CC BY 4.0)

<https://creativecommons.org/licenses/by/4.0/>

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

**Please provide feedback**

Please support the ScholarWorks@UMBC repository by emailing [scholarworks-group@umbc.edu](mailto:scholarworks-group@umbc.edu) and telling us what having access to this work means to you and why it's important to you. Thank you.



# Complete Genome Sequences of *Streptomyces* Bacteriophages Annihilus, TonyStarch, Thiqqums, CricKo, ClubPenguin, RosaAsantewaa, and PherryCruz

Yvette Genie Park,<sup>a</sup> Gillian Faith McCarthy,<sup>a</sup> Hadeeqa Mustafa,<sup>a</sup> Gemma M. Feild,<sup>a</sup> Snigdha Puram,<sup>a</sup> Hager Aly Younes,<sup>a</sup> Danyah Imam,<sup>a</sup> 2021 UMBC Phage Hunters, STEM BUILD at UMBC Cohort 5,  Ivan Erill,<sup>a</sup>  Steven M. Caruso<sup>a</sup>

<sup>a</sup>Department of Biological Sciences, University of Maryland Baltimore County, Baltimore, Maryland, USA

**ABSTRACT** Seven siphoviruses were isolated from soil using *Streptomyces* hosts. Their genome sequences ranged from 42,730 to 57,624 bp long and had a GC content of approximately 60%. Based on their gene content similarity to actinobacteriophages, all seven phages were assigned to cluster BI. For several of these phages, multiple ribosomal frameshifts were identified.

*Streptomyces* species are well known for the production of antibiotics and other bioactive compounds. Here, we report on seven bacteriophages isolated from soil samples on two members of the genus, *Streptomyces scabiei* RL-34 (ATCC 49173), a plant pathogen that causes potato scab disease (1), and *Streptomyces mirabilis* NRRL-2400 (ARS), a species able to grow in soils containing heavy metals (2), using standard methods (3) (Table 1). Briefly, soil samples were washed in phage buffer (10 mM Tris [pH 7.5], 10 mM MgSO<sub>4</sub>, 1 mM CaCl<sub>2</sub>, 68.5 mM NaCl), and the wash was collected by centrifugation and filtration (0.22- $\mu$ m filter). The filtrate was then plated in tryptic soy soft agar (BD), with either *S. scabiei* or *S. mirabilis* overlaid on nutrient agar (BD Difco) supplemented with 10 mM MgCl<sub>2</sub>, 8 mM Ca(NO<sub>3</sub>)<sub>2</sub>, and 0.5% glucose, and incubated at 30°C for 1 to 2 days to yield bacteriophages Annihilus, TonyStarch, Thiqqums, CricKo, ClubPenguin, and RosaAsantewaa. For one soil sample, the filtrate was first inoculated with *S. scabiei* RL-34 and incubated with shaking for 24 h at 30°C; then, the culture was filtered and plated in soft agar with *S. scabiei* yielding phage PherryCruz. All phages were purified with a minimum of three rounds of plating. Negative stain transmission electron microscopy revealed all seven bacteriophages to be siphoviruses (4) (Fig. 1A). The particle capsid and tail measurements are provided in Table 1.

Genomic DNA of all seven bacteriophages was isolated from crude lysate and purified using a Promega Wizard DNA cleanup system, prepared for sequencing using the NEB Ultra II library kit, and sequenced at the Pittsburgh Bacteriophage Institute using the Illumina MiSeq platform (v3 reagents), producing over 100,000 150-base single-end reads for each phage (Table 1). The raw reads were assembled using Newbler v2.9. Quality control was performed using Consed v29 (5). The genome ends were identified by comparison to similar phages with known ends and confirmed by read start buildups. Based on the gene content similarity, all seven phages were assigned using PhagesDB to actinobacteriophage cluster BI (6–8). The sequencing data, genome characteristics, and cluster assignments are provided in Table 1.

Genome annotation was completed using DNA Master v5.23.6 (9) embedded with Glimmer v3.02b (10), GeneMark v4.28 (11), Phamerator v.Actino\_draft 463 (12), NCBI blastp v2.13.0 (13), and HHpred v57c87 (14). The phages were found to have from 55 (RosaAsantewaa) to 94 (CricKo, Thiqqums) protein coding genes, of which an average 32% were assigned functions. No tRNA coding genes were identified using tRNAscan-SE v2.0 (15) or Aragorn v1.2.41 (16).

**Editor** John J. Dennehy, Queens College CUNY

**Copyright** © 2022 Park et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Steven M. Caruso, [scaruso@umbc.edu](mailto:scaruso@umbc.edu).

The authors declare no conflict of interest.

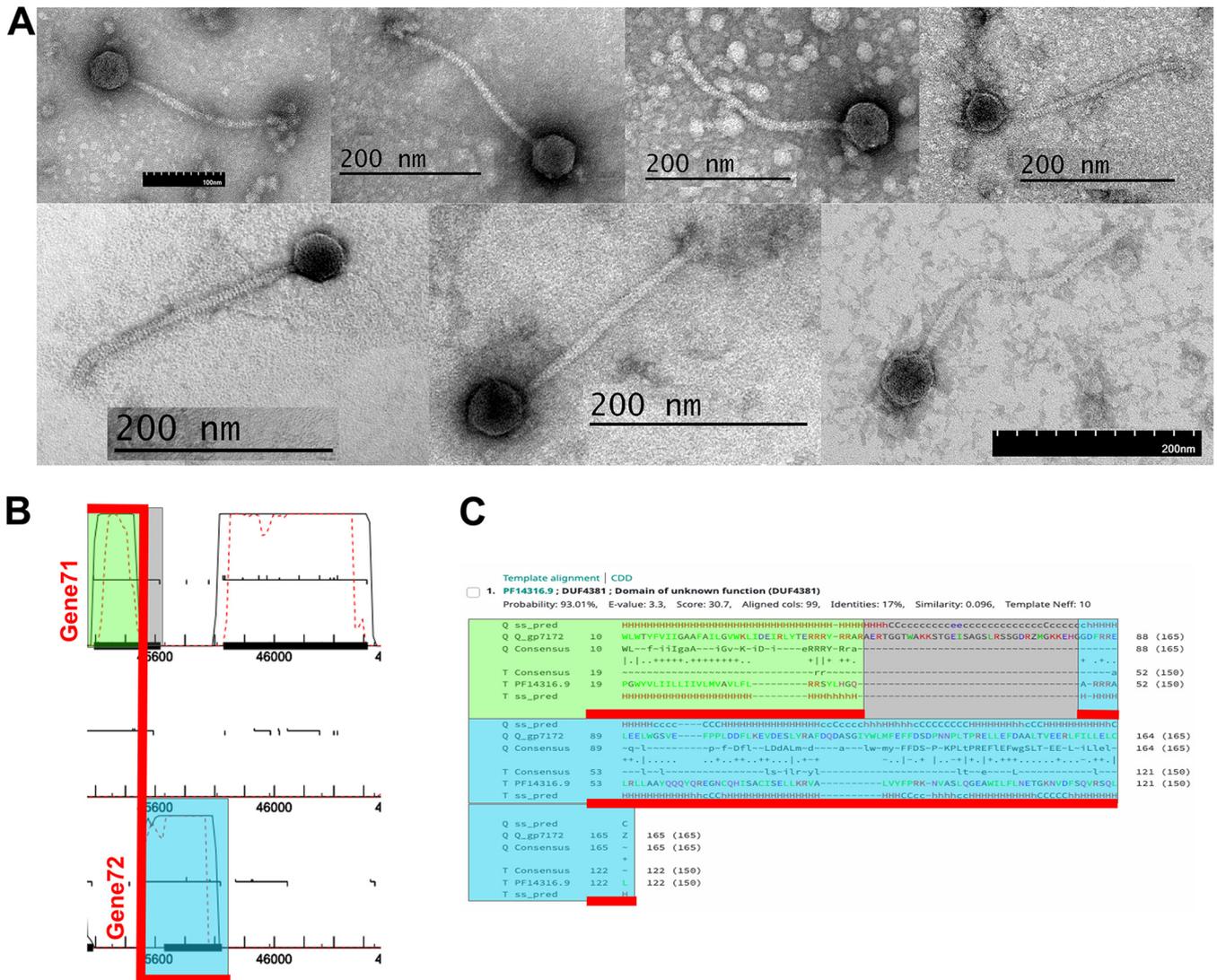
**Received** 1 September 2022

**Accepted** 5 October 2022

**Published** 26 October 2022

**TABLE 1** Summary properties of the analyzed BI *Streptomyces* bacteriophages

Phage name	Collection location	Host	Particle head diam $\pm$ SD, tail length $\pm$ SD (nm); no. of particles	Total no. of 150-bp reads	Cluster	Genome length (bp)	GC content (%)	3' single-stranded end sequence	GenBank accession no.	SRA accession no.
TonyStarch	Ellicott City, MD (39,262 N, 76.818 W)	<i>S. mirabilis</i>	55 $\pm$ 1, 285 $\pm$ 8; 3	972,262	B11	55,469	59.6	5'-CGCCCCGCCT-3'	ON108646	SRX14485098
Annihilus	Glen Burnie, MD (39.133021 N, 76.628615 W)	<i>S. scabiei</i>	52 $\pm$ 2, 238 $\pm$ 24; 4	1,895,132	B12	43,562	61.2	5'-CGCCCGCCCT-3'	ON081336	SRX14443513
PherryCruz	Halethorpe, MD (39.257417 N, 76.704278 W)	<i>S. scabiei</i>	52 $\pm$ 2, 228 $\pm$ 8; 8	1,005,742	B12	43,736	61.0	5'-CGCCCGCCCT-3'	MK686070	SRX14814648
RosaAsantewaa	Accra, Ghana (5.6052560 N, 0.1733080 W)	<i>S. mirabilis</i>	47 $\pm$ 4, 234 $\pm$ 15; 5	832,774	B12	42,730	58.8	5'-CGCCCGCCCT-3'	MK686072	SRX14814650
CricKo	Baltimore, MD (39.25366 N, 76.71331 W)	<i>S. scabiei</i>	62 $\pm$ 3, 271 $\pm$ 6; 4	1,086,915	B14	57,623	58.1	5'-CGCCCGCCCT-3'	MT310854	SRX14814644
Thiqqums	Catonsville, MD (39.2585 N, 76.7131 W)	<i>S. scabiei</i>	50, 200; 1	932,673	B14	57,624	58.1	5'-CGCCCGCCCT-3'	MT657340	SRX14814651
ClubPenguin	Millersville, MD (39.096152 N, 76.573399 W)	<i>S. mirabilis</i>	61 $\pm$ 2, 264 $\pm$ 16; 5	913,162	B17	56,205	59.0	5'-CGCCCGCCCT-3'	MT310852	SRX14814643



**FIG 1** Virion imaging and putative frameshift identification in cluster B1 phages. (A) Representative transmission electron microscopy images of the phages described in this paper. (Left to right, top row) Annihilus, ClubPenguin, CricKo, and PherryCruz; (bottom) RosaAsantewaa, Thiqqums, and TonyStarch. (B) Annotated GeneMark coding potential for TonyStarch genes 71 and 72. Green and blue highlights indicate the coding potential for genes 71 and 72 involved in a hypothetical ribosomal frameshift. Gray highlight indicates the coding potential of gene 71 not included in that ribosomal frameshift. (C) HHpred was used to identify a conserved domain (PF14316) using a literal concatenation of gene product 71 and 72 hits as the query, but the coding potential of gene 71 not included in the ribosomal frameshift was not required for the alignment. Using the predicted frameshift sequence as the query resulted in an improved alignment score (35) and probability (95.85%). Color coding as in panel B.

In the B11 and B12 phages examined here, the heuristic GeneMarkS algorithm (11) was used to predict multiple programmed ribosomal frameshifts. In TonyStarch (B11), most putative frameshifts were in genes located near predicted endolysins and nucleases, while in the cluster B12 phages (Annihilus, PherryCruz, RosaAsantewaa), they were in genes near predicted holins. Some of these putative frameshifts may be functional, since the predicted products for frameshifts involving TonyStarch genes 4 to 5 and 71 to 72 improved the HHpred alignment to known protein domains compared to the literal concatenation of the respective gene products (Fig. 1B and C).

**Data availability.** The GenBank accession numbers for the genome sequences reported here and the SRA accession numbers for the raw sequence reads are available in Table 1.

**ACKNOWLEDGMENTS**

We thank the UMBC Department of Biological Sciences, the SEA-PHAGES program, Viknesh Sivanathan, Daniel Russell, Ralph Murphy, and Tagide deCarvalho for their support.

Members of the 2021–2022 UMBC Phage Hunters class are listed at <https://phages.umbc.edu/home/class-lists/2021-22/>. Members of STEM BUILD at UMBC Cohort 5 are listed at <https://phages.umbc.edu/home/class-lists/stem-build/cohort-5-summer-2020/>.

This research was supported by the Science Education Alliance, Howard Hughes Medical Institute, Chevy Chase, MD, and the National Institute of General Medical Sciences of the National Institutes of Health under award numbers TL4GM118989, UL1GM118988, and RL5GM118987. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This work was also supported by the UMBC Department of Biological Sciences and the Howard Hughes Medical Institute SEA-PHAGES program.

## REFERENCES

- Hiltunen LH, Ojanperä T, Korttemaa H, Richter E, Lehtonen MJ, Valkonen JPT. 2009. Interactions and biocontrol of pathogenic *Streptomyces* strains co-occurring in potato scab lesions. *J Appl Microbiol* 106:199–212. <https://doi.org/10.1111/j.1365-2672.2008.03992.x>.
- Schütze E, Klose M, Merten D, Nietzsche S, Senftleben D, Roth M, Kothe E. 2014. Growth of streptomycetes in soil and their impact on bioremediation. *J Hazard Mater* 267:128–135. <https://doi.org/10.1016/j.jhazmat.2013.12.055>.
- Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA. 2000. *Practical Streptomyces genetics*, 2nd ed. John Innes Foundation, Colney, UK.
- Ackermann HW. 1998. Tailed bacteriophages: the order Caudovirales. *Adv Virus Res* 51:135–201. [https://doi.org/10.1016/s0065-3527\(08\)60785-x](https://doi.org/10.1016/s0065-3527(08)60785-x).
- Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res* 8:195–202. <https://doi.org/10.1101/gr.8.3.195>.
- Russell DA, Hatfull GF. 2017. PhagesDB: the actinobacteriophage database. *Bioinformatics* 33:784–786. <https://doi.org/10.1093/bioinformatics/btw711>.
- Pope WH, Mavrich TN, Garlena RA, Guerrero-Bustamante CA, Jacobs-Sera D, Montgomery MT, Russell DA, Warner MH, Hatfull GF, Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES). 2017. Bacteriophages of *Gordonia* spp. display a spectrum of diversity and genetic relationships. *mBio* 8:e01069-17. <https://doi.org/10.1128/mBio.01069-17>.
- Hatfull GF. 2020. Actinobacteriophages: genomics, dynamics, and applications. *Annu Rev Virol* 7:37–61. <https://doi.org/10.1146/annurev-virology-122019-070009>.
- Pope WH, Jacobs-Sera D. 2018. Annotation of bacteriophage genome sequences using DNA Master: an overview. *Methods Mol Biol* 1681: 217–229. [https://doi.org/10.1007/978-1-4939-7343-9\\_16](https://doi.org/10.1007/978-1-4939-7343-9_16).
- Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679. <https://doi.org/10.1093/bioinformatics/btm009>.
- Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* 29: 2607–2618. <https://doi.org/10.1093/nar/29.12.2607>.
- Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF. 2011. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. *BMC Bioinformatics* 12:395. <https://doi.org/10.1186/1471-2105-12-395>.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. <https://doi.org/10.1093/nar/25.17.3389>.
- Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2018. A completely reimplemented MPI Bioinformatics Toolkit with a new HHpred server at its core. *J Mol Biol* 430: 2237–2243. <https://doi.org/10.1016/j.jmb.2017.12.007>.
- Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol Biol* 1962:1–14. [https://doi.org/10.1007/978-1-4939-9173-0\\_1](https://doi.org/10.1007/978-1-4939-9173-0_1).
- Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32:11–16. <https://doi.org/10.1093/nar/gkh152>.