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 and telling us what having access to this work means to you and why it's important to you. Thank you.

Genome Sequences of *Streptomyces* Bacteriophages Fabian, FlowerPower, Geostin, RetrieverFever, and Vorvolakos

Faith Davis,[a] Baily Kakacek,[a] Delaram Doorandish,[a] Tamar Goldwasser,[a] Neyiasuo Agboha,[a] 2021 UMBC Phage Hunters,[a] STEM BUILD at UMBC Cohort 6,[b] Steven M. Caruso,[a] Ivan Erill[a]

aDepartment of Biological Sciences, University of Maryland, Baltimore County, Baltimore, Maryland, USA
bCollege of Natural and Mathematical Sciences, University of Maryland, Baltimore County, Baltimore, Maryland, USA

ABSTRACT This paper reports the genome sequences of five bacteriophages that were isolated using *Streptomyces scabiei*. Phages Fabian, FlowerPower, Geostin, RetrieverFever, and Vorvolakos were assigned to actinobacteriophage cluster BF based on shared gene content, with each phage containing between 16 and 21 tRNA genes.

Five bacteriophages were isolated using the phytopathogenic bacterium *Streptomyces scabiei* RL-34 and annotated as part of the Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program (1).

Phages were isolated from soil samples collected across Maryland, USA (Table 1), using standard methods (2). Briefly, soil samples were suspended in phage buffer (10 mM Tris [pH 7.5], 10 mM $MgSO_4$, 1 mM $CaCl_2$, 68.5 mM NaCl). Suspensions were centrifuged to pellet soil, and the supernatant was filtered (0.22-μm filter). The filtrate was then plated in tryptic soy soft agar (BD) with *S. scabiei* RL-34 on nutrient agar plates (BD Difco) supplemented with 10 mM $MgCl_2$, 8 mM $Ca(NO_3)_2$, and 0.5% glucose. Incubation at 30°C for 1 to 2 days yielded bacteriophages Fabian, FlowerPower, Geostin, RetrieverFever, and Vorvolakos. A minimum of three rounds of plaque purification were performed for each phage. After 24 to 48 h at 30°C on *S. scabiei*, all phages except Vorvolakos formed round 1.5- to 5-mm-diameter plaques with turbid halos; Vorvolakos formed clear plaques. Negative-staining transmission electron microscopy revealed these phages to be podoviruses, with capsid widths ranging from 50 to 61 nm (Fig. 1A).

Phage DNA was isolated using the Promega Wizard DNA purification system on a freshly prepared lysate as described previously (3). Sequencing of Fabian, Geostin, RetrieverFever, and Vorvolakos was completed by the Pittsburgh Bacteriophage Institute; sequencing of FlowerPower was completed by the North Carolina State University Genomic Sciences Laboratory. All sequencing was performed with the Illumina MiSeq (v3 reagents) sequencing platform using the NEBNext Ultra II library preparation kit and 150-base single-end reads. Raw sequencing reads were assembled using Newbler v2.9 or the CLC Genomics Workbench next-generation sequencing (NGS) *de novo* assembler v6 with default settings. Genome completeness and termini were determined using Consed v29 (4, 5). All phages were found to have linear chromosomes with 253-bp direct terminal repeats and an average GC content of 60.07% (standard deviation, ±0.44%) (Table 1). Based on gene content similarity, the Actinobacteriophage Database (phagesDB) assigned these phages to the BF cluster (6–8).

Genome annotation was completed using DNA Master v5.23.6 (9) embedded with Glimmer v3.02b (10) and GeneMark v4.28 (11), with manual refinement using parameters such as proximity and directionality with respect to protein-encoding sequences, the presence of putative ribosome binding sites, and sequence similarity to previously annotated genes (12). Protein-encoding genes were functionally annotated using BLASTp (13), Phamerator v3.0 (14), and HHPred v57c87 (15, 16). The numbers of genes encoding proteins with known functions ranged from 13 to 20 (Table 1). The numbers of tRNA genes found using ARAGORN

TABLE 1 Properties of five cluster BF *Streptomyces* phage genomes

| Phage	Sample collection location coordinates	No. of genes encoding proteins with assigned functions	No. of tRNA genes	Genome length (bp)	No. of reads	Sequencing coverage (×)	GenBank accession no.	SRA accession no.
Fabian	39.25281N, 76.851488W	18	18	46,505	649,913	2,003	MN284894	SRX14814645
FlowerPower	39.256894N, 76.714128W	13	21	46,133	1,596,259	5,190	MH155868	SRX14814646
Geostin	39.140777N, 76.528469W	16	16	46,175	537,680	1,661	MN119376	SRX14814647
RetrieverFever	39.23N, 77.24W	18	21	46,175	898,381	2,778	MZ622160	SRX14814649
Vorvolakos	39.58972N, 76.28048W	20	21	46,175	4,523,515	13,943	ON108647	SRX14485104
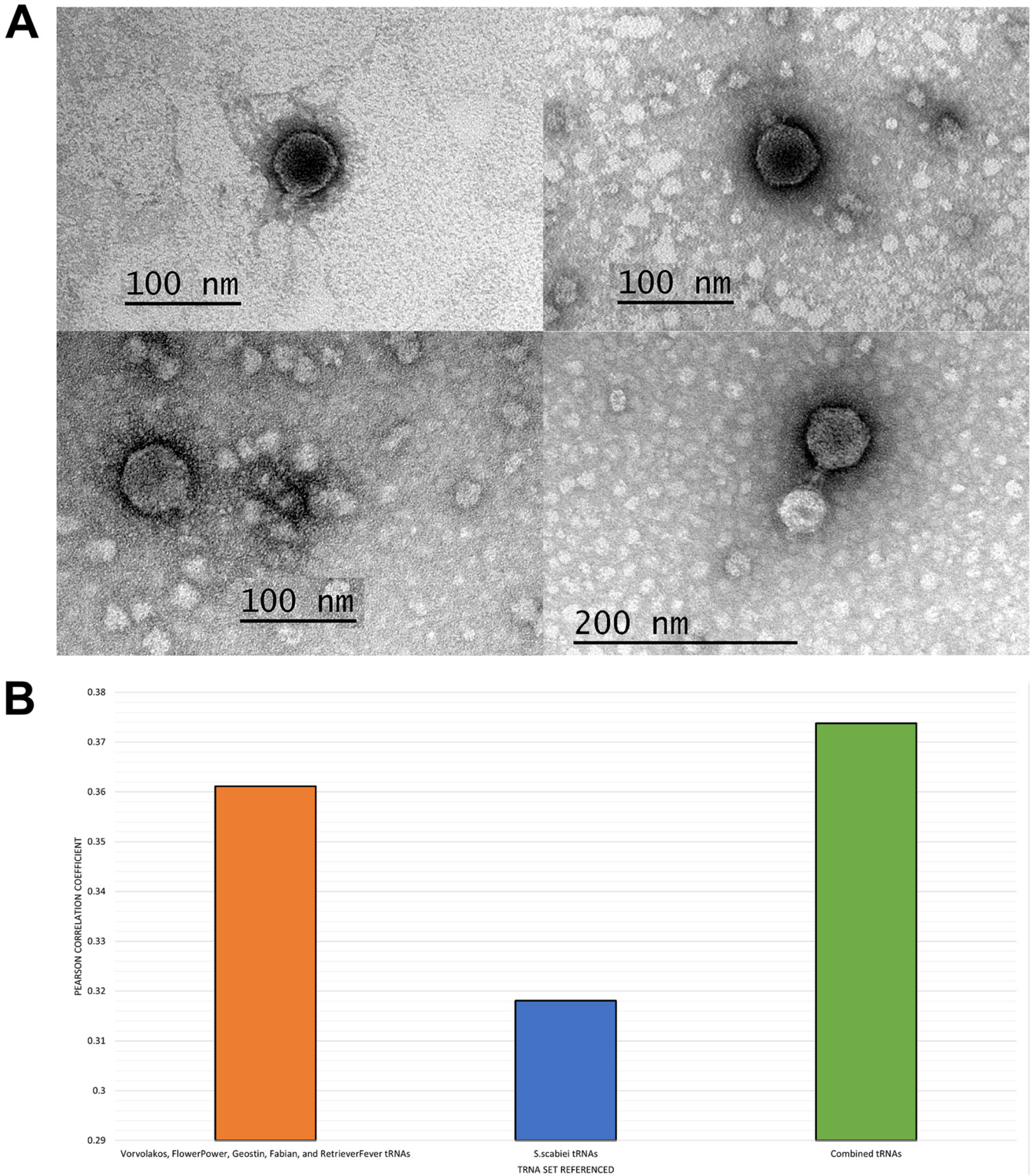
FIG 1 Characterization of phages and tRNA genes. (A) Representative transmission electron microscopy images of the phages described in this paper. (Top left) Fabian. (Top right) FlowerPower. (Bottom left) Geostin. (Bottom right) RetrieverFever. (B) Pearson correlation coefficients for correlations between codons recognized by the products of tRNA genes included in *Streptomyces* BF cluster phage genomes, the *S. scabiei* genome, or the combined phage and *S. scabiei* genomes and the codons present in the gene coding for the MCP in each phage. Codon usage correlates better with the combined tRNA pool.

v1.2.41 ranged from 16 to 21 (17). All cluster BF phages contain 15 to 20 tRNAs in a single genome segment. The numbers of codons in their major capsid protein (MCP) gene sequences, as well as the numbers of host and phage tRNAs recognizing each codon, were tabulated, and their correlation was analyzed. Our data support the observation that phage tRNAs match the codon usage for the MCPs better than do host tRNAs (Fig. 1B) (18).

Data availability. GenBank nucleotide record and Sequence Read Archive (SRA) accession numbers for all genomes reported in this work are provided in Table 1.

REFERENCES

1. Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, DeJong RJ, Dennehy JJ, Denver DR, Dunbar D, Elgin SCR, Findley AM, Gissendanner CR, Golebiewska UP, Guild N, Hartzog GA, Grillo WH, Hollowell GP, Hughes LE, Johnson A, King RA, Lewis LO, Li W, Rosenzweig F, Rubin MR, Saha MS, Sandoz J, Shaffer CD, Taylor B, Temple L, Vazquez E, Ware VC, Barker LP, Bradley KW, Jacobs-Sera D, Pope WH, Russell DA, Cresawn SG, Lopatto D, Bailey CP, Hatfull GF. 2014. A broadly implementable research course in phage discovery and genomics for first-year undergraduate students. mBio 5:e01051-13. https://doi.org/10.1128/mBio.01051-13.
2. Sarkis GJ, Hatfull GF. 1998. Mycobacteriophages. Methods Mol Biol 101:145–173. https://doi.org/10.1385/0-89603-471-2:145.
3. Caruso SM, deCarvalho TN, Huynh A, Morcos G, Kuo N, Parsa S, Erill I. 2019. A novel genus of actinobacterial *Tectiviridae*. Viruses 11:1134. https://doi.org/10.3390/v11121134.
4. Miller JR, Koren S, Sutton G. 2010. Assembly algorithms for next-generation sequencing data. Genomics 95:315–327. https://doi.org/10.1016/j.ygeno.2010.03.001.
5. 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.
6. Hatfull GF. 2020. Actinobacteriophages: genomics, dynamics, and applications. Annu Rev Virol 7:37–61. https://doi.org/10.1146/annurev-virology-122019-070009.
7. Russell DA, Hatfull GF. 2017. PhagesDB: the Actinobacteriophage Database. Bioinformatics 33:784–786. https://doi.org/10.1093/bioinformatics/btw711.
8. Pope WH, Mavrich TN, Garlena RA, Guerrero-Bustamante CA, Jacobs-Sera D, Montgomery MT, Russell DA, Warner MH, Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES), Hatfull GF. 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.
9. 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.
10. 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.
11. 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.
12. Pope WH, Jacobs-Sera D, Russell DA, Cresawn SG, Hatfull GF. 2017. SEA-PHAGES phage discovery guide. Howard Hughes Medical Institute, Chevy Chase, MD. https://seaphagesphagediscoveryguide.helpdocsonline.com/home.
13. 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.
14. 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.
15. 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.
16. Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33:W244–W248. https://doi.org/10.1093/nar/gki408.
17. 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.
18. Bailly-Bechet M, Vergassola M, Rocha E. 2007. Causes for the intriguing presence of tRNAs in phages. Genome Res 17:1486–1495. https://doi.org/10.1101/gr.6649807.