



Complete Genome Sequences of Five Phietaviruses Infecting *Staphylococcus aureus*

 Taylor P. Andrews,^a  J. Steen Hoyer,^b  Nicole L. Fahrenfeld,^c  Jeffrey M. Boyd,^d  Siobain Duffy^b

^aMicrobial Biology Graduate Program, Rutgers University, New Brunswick, New Jersey, USA

^bDepartment of Ecology, Evolution, and Natural Resources, School of Environmental and Biological Sciences, Rutgers University, New Brunswick, New Jersey, USA

^cDepartment of Civil and Environmental Engineering, School of Engineering, Rutgers University, Piscataway, New Jersey, USA

^dDepartment of Biochemistry and Microbiology, School of Environmental and Biological Sciences, Rutgers University, New Brunswick, New Jersey, USA

ABSTRACT The annotated whole-genome sequences of five cultured phietaviruses infecting *Staphylococcus aureus* are presented. They are closely related to prophages that were previously sequenced as part of *S. aureus* genomes.

Staphylococcus aureus is a human commensal bacterium that has the potential to cause life-threatening infection (1). Its interactions with bacteriophages are an increasingly studied part of microbiome studies (2). We present the annotated genomes of five plaque-purified *S. aureus* temperate phages in the genus *Phietavirus* (3). Four aliquots of municipal wastewater influent from a mid-Atlantic, U.S. treatment plant were collected in March 2021. To enrich for *S. aureus* phages, 5 mL of each sample were cocultured with *S. aureus* RN4220 (4) in tryptic soy broth (TSB) containing 10 mM CaCl₂ (5). Phages were isolated using centrifugation and 0.22- μ m filtration before being plated with *S. aureus* RN4220 using the pour-plate technique. Plaques underwent three rounds of subculturing through single plaques to yield purified phage stocks (6). The DNA genomes of five selected phages were extracted using QIAamp MinElute virus spin kits.

Paired-end (2 \times 150-bp) sequencing using the Illumina DNA library preparation kit was performed on the NextSeq 2000 system at the Microbial Genome Sequencing Center (MiGS), which provided quality-controlled and trimmed reads. These reads were analyzed using the CPT Galaxy Phage Genome Assembler v2021.01 Workflow (<https://cpt.tamu.edu/galaxy-pub>) (7) with SPAdes v3.12.0 (<http://cab.spbu.ru/software/spades>) (8), which produced linear contigs with small overlaps at the end, suggesting that the genomes were circular. The overlaps were manually cut. Taxonomic assignment of the five double-stranded DNA (dsDNA) phage genomes was performed with GRAVITY v1.1.0 (<http://gravity.cvr.gla.ac.uk>) (9), which showed that they were phietaviruses (symmetrical Theil's *U* value [reference prediction] of 0.863) related to SAP26 (GenBank accession number [GU477322](https://www.ncbi.nlm.nih.gov/nuccore/GU477322) [arbitrarily linearized]). The genomes were reoriented to reflect the termini of *Staphylococcus* prophages from a closely related genus (e.g., GenBank accession number [DQ530359](https://www.ncbi.nlm.nih.gov/nuccore/DQ530359)). Genome annotation was performed as described previously (10, 11); open reading frames (ORFs) were annotated using Prokka v1.14.6 (parameters Genus: *Phietavirus*, Kingdom: Viruses) in Galaxy (12) and further annotated for functionality with the PHROGs v4 database (<https://phrogs.lmge.uca.fr>) (13) and Phyre2 v2.0 (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>) (14), and non-protein-coding features, including tRNAs (tRNAscan-SE v2.0) (<http://trna.ucsc.edu/tRNAscan-SE>) (15), terminators (ARNold v1.0) (<http://rssf.i2bc.paris-saclay.fr/toolbox/arnold>) (16), noncoding RNAs (Rfam v14.8) (<https://rfam.xfam.org/search#tabview=tab1>) (17), and promoters (Genome2D Prokaryote Promoter Prediction) (http://genome2d.molgenrug.nl/g2d_pepper_promoters.php) (18), were identified. Sequence

Editor John J. Dennehy, Queens College CUNY

Copyright © 2022 Andrews 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 Jeffrey M. Boyd, jeffboyd@sebs.rutgers.edu, or Siobain Duffy, duffy@sebs.rutgers.edu.

The authors declare no conflict of interest.

Received 17 August 2022

Accepted 15 September 2022

Published 29 September 2022

TABLE 1 Summary of SAP phage genomic characteristics

Phage	Genome length (bp)	No. of predicted ORFs	No. of putative promoters	No. of putative rho-independent terminators	Avg sequencing coverage (×)	GC content (%)	No. of reads	GenBank accession no.	SRA accession no.
SAP1	43,962	68	10	22	9,518	34.3	2,896,630	ON911714	SRX16769400
SAP2	43,863	69	6	23	9,069	34.0	2,736,310	ON911715	SRX16769401
SAP3	43,586	66	11	18	11,412	34.6	3,405,310	ON911716	SRX16769402
SAP8	42,981	63	8	20	11,997	34.1	3,539,164	ON911717	SRX16769403
SAP13	43,478	67	10	25	11,145	34.6	3,316,128	ON911718	SRX16769404

coverage was calculated using Map with BWA-MEM v0.7.17.2 (19) and SAMtools Depth v1.13 in Galaxy (20). Default parameters were used except where otherwise noted.

The five SAP genomes are ~43 kb (Table 1), and portions of the genomes are very similar to one another (the most divergent pair, SAP1 and SAP8, are $\geq 94\%$ identical by BLAST over 60% of the genome). There was significant synteny between the 63 to 69 ORFs of the genomes (Fig. 1). The closest BLAST hits to these phage genomes in the NCBI nonredundant database are all prophages within *S. aureus* genomes (e.g., SAP3 is 100% identical, with 100% query coverage, by BLAST to GenBank accession number [CP051919](#)).

Data availability. Genomes are available in GenBank (see Table 1 for the accession numbers). Illumina data are available in the NCBI SRA (BioProject accession number [PRJNA857681](#)) (Table 1). The phages are available by request from the corresponding authors.

ACKNOWLEDGMENTS

This work was supported by NIAID grant 1R01AI139100-01, NSF grant 1750624, and USDA MRF project NE-1028 to J.M.B., NSF grant 1453241 to S.D., and a Rutgers Center for COVID-19 Response and Pandemic Preparedness grant to N.L.F. T.P.A. was supported by NIGMS grant 1T32GM139804-01.

We thank the utility partner (which wished to remain anonymous) for providing the wastewater influent. We thank the Center for Phage Technology for their open resources and Evelien Adriaenssens for her input on analysis.

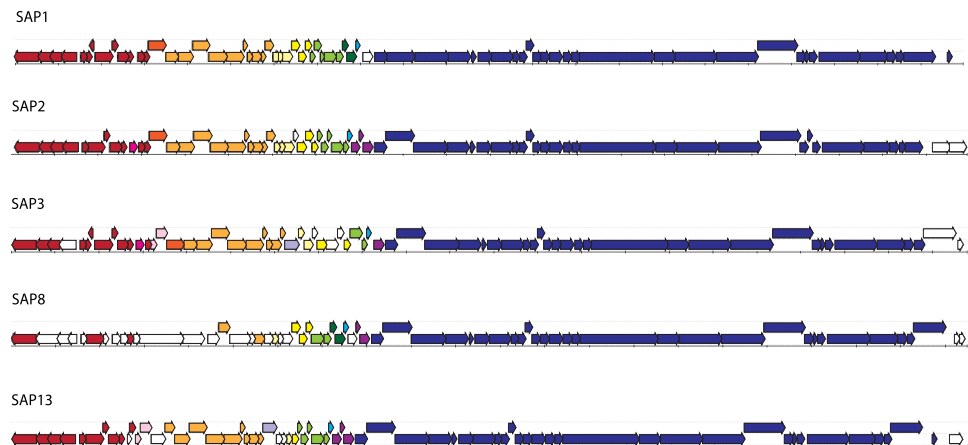


FIG 1 Genomic maps of the five phage genomes. Colors indicate blocks of homology, and ORFs without homology with other SAP genomes are depicted in white. All genomes have integrase genes at the 5' end, indicating that they are likely capable of lysogeny. They share a large, syntenous block of genes toward the 3' end, containing structural and hypothetical proteins.

REFERENCES

1. Tong SYC, Davis JS, Eichenberger E, Holland TL, Fowler VG. 2015. *Staphylococcus aureus* infections: epidemiology, pathophysiology, clinical manifestations, and management. *Clin Microbiol Rev* 28:603–661. <https://doi.org/10.1128/CMR.00134-14>.
2. Federici S, Nobs SP, Elinav E. 2021. Phages and their potential to modulate the microbiome and immunity. *Cell Mol Immunol* 18:889–904. <https://doi.org/10.1038/s41423-020-00532-4>.
3. Gutiérrez D, Adriaenssens EM, Martínez B, Rodríguez A, Lavigne R, Kropinski AM, García P. 2014. Three proposed new bacteriophage genera of staphylococcal phages: “3alikevirus,” “77likevirus” and “Phietalikevirus”. *Arch Virol* 159:389–398. <https://doi.org/10.1007/s00705-013-1833-1>.
4. Kreiswirth BN, Löfdahl S, Betley MJ, O’Reilly M, Schlievert PM, Bergdoll MS, Novick RP. 1983. The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. *Nature* 305:709–712. <https://doi.org/10.1038/305709a0>.
5. Price EE, Rudra P, Norambuena J, Román-Rodríguez F, Boyd JM, Semrau JD. 2021. Tools, strains, and strategies to effectively conduct anaerobic and aerobic transcriptional reporter screens and assays in *Staphylococcus aureus*. *Appl Environ Microbiol* 87:e01108-21. <https://doi.org/10.1128/AEM.01108-21>.
6. Hyman P. 2019. Phages for phage therapy: isolation, characterization, and host range breadth. *Pharmaceuticals* 12:35. <https://doi.org/10.3390/ph12010035>.
7. Ramsey J, Rasche H, Maughmer C, Criscione A, Mijalis E, Liu M, Hu JC, Young R, Gill JJ. 2020. Galaxy and Apollo as a biologist-friendly interface for high-quality cooperative phage genome annotation. *PLoS Comput Biol* 16:e1008214. <https://doi.org/10.1371/journal.pcbi.1008214>.
8. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
9. Aiewsakun P, Simmonds P. 2018. The genomic underpinnings of eukaryotic virus taxonomy: creating a sequence-based framework for family-level virus classification. *Microbiome* 6:38. <https://doi.org/10.1186/s40168-018-0422-7>.
10. Turner D, Adriaenssens EM, Tolstoy I, Kropinski AM. 2021. Phage annotation guide: guidelines for assembly and high-quality annotation. *Phage* (New Rochelle) 2:170–182. <https://doi.org/10.1089/phage.2021.0013>.
11. Shen A, Millard A. 2021. Phage genome annotation: where to begin and end. *Phage* 2:183–193. <https://doi.org/10.1089/phage.2021.0015>.
12. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
13. Terzian P, Olo Ndela E, Galiez C, Lossouarn J, Pérez Bucio RE, Mom R, Toussaint A, Petit M-A, Enault F. 2021. PHROG: families of prokaryotic virus proteins clustered using remote homology. *NAR Genom Bioinform* 3: lqab067. <https://doi.org/10.1093/nargab/lqab067>.
14. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10: 845–858. <https://doi.org/10.1038/nprot.2015.053>.
15. 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.
16. Naville M, Ghuillot-Gaudeffroy A, Marchais A, Gautheret D. 2011. ARNold: a web tool for the prediction of Rho-independent transcription terminators. *RNA Biol* 8:11–13. <https://doi.org/10.4161/rna.8.1.13346>.
17. Kalvari I, Nawrocki EP, Ontiveros-Palacios N, Argasinska J, Lamkiewicz K, Marz M, Griffiths-Jones S, Toffano-Nioche C, Gautheret D, Weinberg Z, Rivas E, Eddy SR, Finn RD, Bateman A, Petrov AI. 2021. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res* 49:D192–D200. <https://doi.org/10.1093/nar/gkaa1047>.
18. Baerends RJ, Smits W, De Jong A, Hamoen LW, Kok J, Kuipers OP. 2004. Genome2D: a visualization tool for the rapid analysis of bacterial transcriptome data. *Genome Biol* 5:R37. <https://doi.org/10.1186/gb-2004-5-5-r37>.
19. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 1303.3997v2. <https://doi.org/10.48550/arXiv.1303.3997>.
20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.