AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of a Phage Infecting *Sphingomonadaceae*

Shang Shen,[a,b] Tetsunobu Anazawa,[a] Tomonari Matsuda,[a] Yoshihisa Shimizu[a]

[a]Department of Engineering, Kyoto University, Otsu, Shiga, Japan
[b]Lake Biwa Branch Office, National Institute for Environmental Studies, Otsu, Shiga, Japan

Shang Shen and Tetsunobu Anazawa contributed equally to this work. Author order was determined in order of increasing seniority.

ABSTRACT We isolated a phage infecting a member of the *Sphingomonadaceae* family from a freshwater lake. The phage has a DNA genome of 41,771 bp, with a GC content of 61.7%. The genome harbors 50 predicted protein-coding genes and an auxiliary metabolic gene, which encodes a protein belonging to the radical *S*-adenosylmethionine superfamily.

Members of the family *Sphingomonadaceae* play important roles in aquatic environments (1–4). However, little is known about the phages infecting *Sphingomonadaceae* and how they control the abundance of this family.

The phage, infecting a member of the *Sphingomonadaceae* family, was isolated from a water sample collected from the surface of a freshwater lake (Lake Biwa, Japan; 35°23′21.0″N, 136°07′51.0″E) and from a strain of *Sphingomonadaceae* family (BSN-002 [deposited in GenBank under accession number CP091804]) (5). The collected lake water was filtered through 0.2-μm-pore-size polycarbonate filters. The filtrates and the preincubated host strain were incubated using the double-agar method at 25°C for 2 weeks. Virions contained in the plaques were purified twice using the same method. Phage DNA was extracted using the DNeasy blood and tissue kit (Qiagen, Hilden, Germany). The library was prepared using a Nextera XT DNA sample preparation kit (Illumina, San Diego, CA) according to the manufacturer's protocol. The sample was then sequenced using a MiSeq sequencing system with a V3 (2× 300-bp reads) reagent kit (Illumina). Bioinformatic analysis was performed as described below, and default parameters were used for all software unless otherwise specified. Raw reads (524,664 paired-end reads) with low-quality regions were removed using Trimmomatic with default settings (v.0.39) (6), and finally, 515,345 paired-end reads were obtained. The phage genome was assembled using SPAdes with the –careful option (v.3.13.1) (7) and identified as a complete (i.e., circular) genome using ccfind (v.1.1) (8). Open reading frames (ORFs) were predicted using Prodigal with the -p meta option (v.2.6.3) (9) and annotated by eggNOG-Mapper v.1.0.3 (10) using the -m diamond option. ORFs were also annotated using HMMER v.3.1b2 software (11) against the Prokaryotic Virus Orthologous Groups databases (12) and using the highly sensitive HMM-HMM search (8) against Pfam v.31.0 databases using HHsearch (13) and JackHMMER (14).

The genome of the phage (VSN-002) infecting a member of the *Sphingomonadaceae* family was found to comprise 41,771 bp, with a GC content of 61.7%. Of the 75 predicted ORFs, 41 were leftward oriented, 65 started with ATG, 8 started with GTG, and 2 started with TTG (Table 1). A total of 50 proteins were encoded in the phage genome, varying in length from 138 to 3,186 bp. Of the 50 proteins, 19 were classified as hypothetical or uncharacterized proteins, and 31 were assigned putative functions. The phage genome was found to encode proteins related to phage replication (e.g., terminase, DNA helicase, capsid, and tail). An auxiliary metabolic gene (AMG), which may be expressed during infection to increase viral production by promoting host metabolism, was also found in the phage genome. The phage genome was found to have a radical *S*-adenosylmethionine (SAM) superfamily that

Editor Kenneth M. Stedman, Portland State University

Address correspondence to Shang Shen, shen.shang@nies.go.jp.

The authors declare no conflict of interest.

Received 11 April 2022
Accepted 19 May 2022
Published 2 June 2022

TABLE 1 Summary of predicted open reading frames and their annotation

Parameter	Value
Length of genome (bp)	41,771
GC content (%)	61.7
No. of predicted ORFs	
Total	75
Leftward	41
Rightward	34
Start codon with ATG	65
Start codon with GTG	8
Start codon with TTG	2
No. of ORFs annotated	
Total	50
Hypothetical or uncharacterized proteins	19
Putative functions	31

may be involved in critical roles in numerous biosynthetic pathways in bacteria (e.g., synthesis of protein cofactors) (15). BLASTn analysis showed that our *Sphingomonadaceae* phage was not homologous (more than 95% similarity) to previously reported bacteriophages, and the host of this phage was also not assigned to any existing species in the Genome Taxonomy Databases (5).

Data availability. The complete bacteriophage genome has been deposited in GenBank (accession number MZ127829). Raw reads from whole-genome sequencing data were deposited in the DDBJ (accession number DRR361504).

ACKNOWLEDGMENTS

We thank Y. Goda and T. Akatsuka (Center for Ecological Research, Kyoto University) for their assistance with field sampling. Computation time was provided by the SuperComputer System, Institute for Chemical Research, Kyoto University.

This research was supported by JSPS KAKENHI (grant numbers JP19J14985 and JP20H04323) and the Kurita Water and Environment Foundation (19B045).

REFERENCES

1. Baik KS, Choe HN, Park SC, Hwang YM, Kim EM, Park C, Seong CN. 2013. Sphingopyxis rigui sp. nov. and Sphingopyxis wooponensis sp. nov., isolated from wetland freshwater, and emended description of the genus Sphingopyxis. Int J Syst Evol Microbiol 63:1297–1303. https://doi.org/10.1099/ijs.0.044057-0.
2. Gomez-Alvarez V, Pfaller S, Revetta RP. 2016. Draft genome sequence of two Sphingopyxis sp. strains, dominant members of the bacterial community associated with a drinking water distribution system simulator. Genome Announc 4:e00183-16. https://doi.org/10.1128/genomeA.00183-16.
3. Stolz A. 2009. Molecular characteristics of xenobiotic-degrading sphingomonads. Appl Microbiol Biotechnol 81:793–811. https://doi.org/10.1007/s00253-008-1752-3.
4. Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. 2011. A guide to the natural history of freshwater lake bacteria. Microbiol Mol Biol Rev 75:14–49. https://doi.org/10.1128/MMBR.00028-10.
5. Shen S, Anazawa T, Matsuda T, Shimizu Y. 2022. Draft genome sequences of *Sphingomonadaceae* strains isolated from a freshwater lake. Microbiol Resour Announc 11:e00070-22. https://doi.org/10.1128/mra.00070-22.
6. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
7. 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.
8. Nishimura Y, Watai H, Honda T, Mihara T, Omae K, Roux S, Blanc-Mathieu R, Yamamoto K, Hingamp P, Sako Y, Sullivan MB, Goto S, Ogata H, Yoshida T. 2017. Environmental viral genomes shed new light on virus-host interactions in the ocean. mSphere 2:e00359-16. https://doi.org/10.1128/mSphere.00359-16.
9. Hyatt D, Chen G, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.
10. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ, von Mering C, Bork P. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res 44:D286–D293. https://doi.org/10.1093/nar/gkv1248.
11. Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol 7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195.
12. Grazziotin AL, Koonin EV, Kristensen DM. 2017. Prokaryotic Virus Orthologous Groups (pVOGs): a resource for comparative genomics and protein family annotation. Nucleic Acids Res 45:D491–D498. https://doi.org/10.1093/nar/gkw975.
13. Söding J. 2005. Protein homology detection by HMM-HMM comparison. Bioinformatics 21:951–960. https://doi.org/10.1093/bioinformatics/bti125.
14. Johnson LS, Eddy SR, Portugaly E. 2010. Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics 11:431. https://doi.org/10.1186/1471-2105-11-431.
15. Broderick JB, Duffus BR, Duschene KS, Shepard EM. 2014. Radical S-adenosylmethionine enzymes. Chem Rev 114:4229–4317. https://doi.org/10.1021/cr4004709.