



Complete Genome Sequence of *Emiliana huxleyi* Virus Strain M1, Isolated from an Induced *E. huxleyi* Bloom in Bergen, Norway

 Amir Fromm,^a  Daniella Schatz,^a  Shifra Ben-Dor,^b  Ester Feldmesser,^b  Assaf Vardi^a

^aDepartment of Plant and Environmental Sciences, Weizmann Institute of Science, Rehovot, Israel

^bBioinformatics Unit, Life Sciences Core Facilities, Weizmann Institute of Science, Rehovot, Israel

ABSTRACT *Emiliana huxleyi* virus strain M1 (EhVM1), a large double-stranded DNA virus from the family *Phycodnaviridae*, was isolated from an *Emiliana huxleyi* bloom during a mesocosm experiment in Raunefjorden, Bergen, Norway. Here, we report its complete genome, composed of one full contig.

Emiliana huxleyi is a unicellular alga that forms massive blooms that cover vast oceanic ranges. *E. huxleyi* blooms are routinely infected by the *Emiliana huxleyi* virus (EhV), leading to their demise (1). EhV is a large, double-stranded DNA virus from the family *Phycodnaviridae* (2). Here, we report the complete genome sequence of EhV strain M1 (EhVM1), which was isolated from an *E. huxleyi* bloom during a mesocosm experiment in Bergen, Norway (3, 4). To isolate EhV strains from the natural environment, water from the induced blooms in mesocosm bags (3) was filtered through a GF/C filter and stored at 4°C.

This water was used to inoculate *E. huxleyi* cells and conduct plaque assays for viral isolation, according to the methods described previously for EhV86 (5). Visible plaques were excised on day 4 postinfection and placed in a fresh *E. huxleyi* CCMP374 culture. Once the culture cleared, the lysate was used for two consecutive plaque assay rounds. Cultures (100 mL) of *E. huxleyi* CCMP374 were grown to exponential phase and then infected with EhVM1. Once the culture cleared, the lysate was filtered to eliminate cell debris, and the viruses were concentrated using a 100-kDa Amicon filter. DNA was extracted from the virions by a conventional phenol-chloroform method (6). The DNA concentration and quality were measured using Qubit and NanoDrop analyses. Library preparation was performed according to the Pacific Biosciences (PacBio) microbial multiplexing protocol for one Sequel single-molecule real-time (SMRT) Cell (7). Polymerase reads were demultiplexed to subreads and assigned to EhVM1 using SMRT Link analysis (Table 1). Highly accurate circular consensus sequence (CCS) reads were generated from subreads using SMRT Link with default parameters; all tools were run with default parameters unless otherwise specified.

Three draft EhVM1 assemblies were constructed by (i) SMRT Link microbial assembly using an expected genome size parameter of 1 Mbp, (ii) Canu assembly (8) with both CCS reads and subreads as the input, and (iii) SPAdes assembly (9) with CCS reads as the input. A final version of the genome was assembled from the three draft assemblies using GFinisher (10). Circularization of the genome was confirmed with CCS reads that spanned both ends. The GC content of the genome was calculated via an in-house script. The genome sequence of EhVM1 consists of one circular contig of 411,976 bp, longer than the EhV86 genome (407 kbp [5]), with an average GC content of 40.32%. Coding DNA sequences (CDSs) longer than 100 amino acids were predicted using GeneMarkS (11) with the virus sequence type parameter. Additional CDSs were predicted using Prodigal (12). tRNAs were predicted with tRNAscan-SE v2.0 (13) and were

Editor Jelle Matthijnssens, KU Leuven

Copyright © 2022 Fromm 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 Assaf Vardi, assaf.vardi@weizmann.ac.il.

The authors declare no conflict of interest.

Received 16 February 2022

Accepted 4 April 2022

Published 19 April 2022

TABLE 1 Sequencing data

Parameter	Finding for EhVM1
No. of polymerase reads	6,520
No. of subreads	5.6E+4
No. of subread bases	2.9E+8
No. of CCS reads	2,995

analyzed by the RNACentral (14) web server for verification. We predict that the EhVM1 genome contains 489 CDSs, more than EhV86 (472 CDSs [5]), and 6 tRNA genes.

Data availability. The complete genome sequence of EhVM1 has been deposited in GenBank under the accession number [OM339720](https://www.ncbi.nlm.nih.gov/nuclseq/OM339720). PacBio sequence reads have been deposited in the NCBI Sequence Read Archive (SRA) under the BioSample accession number [SAMN24818830](https://www.ncbi.nlm.nih.gov/sra/SAMN24818830). The complete record is available in the NCBI BioProject database under the accession number [PRJNA796183](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA796183).

ACKNOWLEDGMENTS

This research was supported by the Simons Foundation (grant 735079 [Untangling the infection outcome of host-virus dynamics in algal blooms in the ocean], awarded to A.V.). The mesocosm experiment VIMS-Ehux was supported by EU Horizon2020-INFRAIA project AQUACOSM (grant no. 731065).

REFERENCES

1. Laber CP, Hunter JE, Carvalho F, Collins JR, Hunter EJ, Schieler BM, Boss E, More K, Frada M, Thamatrakoln K, Brown CM, Haramaty L, Ossolinski J, Fredricks H, Nissimov JI, Vandzura R, Sheyn U, Lehahn Y, Chant RJ, Martins AM, Coolen MJL, Vardi A, Ditullio GR, van Mooy BAS, Bidle KD. 2018. *Coccolithovirus* facilitation of carbon export in the North Atlantic. *Nat Microbiol* 3:537–547. <https://doi.org/10.1038/s41564-018-0128-4>.
2. Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB. 2018. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res* 46:D708–D717. <https://doi.org/10.1093/nar/gkx932>.
3. Vincent F, Sheyn U, Porat Z, Schatz D, Vardi A. 2021. Visualizing active viral infection reveals diverse cell fates in synchronized algal bloom demise. *Proc Natl Acad Sci U S A* 118:e2021586118. <https://doi.org/10.1073/pnas.2021586118>.
4. Vincent F, Gralka M, Schleyer G, Schatz D, Cabrera-Brufau M, Kuhlisch C, Sichert A, Vidal-Melgosa S, Mayers K, Barak-Gavish N, Flores JM, Masdeu-Navarro M, Egge JK, Larsen A, Heheman J-H, Marrasé C, Simó R, Cordero OX, Vardi A. 2021. Viral infection switches the balance between bacterial and eukaryotic recyclers of organic matter during algal blooms. *bioRxiv* 2021.10.25.465659. <https://doi.org/10.1101/2021.10.25.465659>.
5. Wilson WH, Schroeder DC, Allen MJ, Holden MTG, Parkhill J, Barrell BG, Churcher C, Hamlin N, Mungall K, Norbertczak H, Quail MA, Price C, Rabinowitsch E, Walker D, Craigon M, Roy D, Ghazal P. 2005. Complete genome sequence and lytic phase transcription profile of a *Coccolithovirus*. *Science* 309:1090–1092. <https://doi.org/10.1126/science.1113109>.
6. Sambrook J, Fritsch E, Maniatis T. 1989. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
7. Pacific Biosciences. 2020. Procedure & checklist: preparing multiplexed microbial libraries using SMRTbell® Express template prep kit 2.0. Pacific Biosciences, Menlo Park, CA. <https://www.pacb.com/wp-content/uploads/Procedure-Checklist-%E2%80%93-Preparing-Multiplexed-Microbial-Libraries-Using-SMRTbell-Express-Template-Prep-Kit-2.0.pdf>.
8. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>.
9. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski 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>.
10. Guizelini D, Raittz RT, Cruz LM, Souza EM, Steffens MBR, Pedrosa FO. 2016. GFinisher: a new strategy to refine and finish bacterial genome assemblies. *Sci Rep* 6:34963. <https://doi.org/10.1038/srep34963>.
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. Hyatt D, Chen GL, 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>.
13. Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* 44:W54–W57. <https://doi.org/10.1093/nar/gkw413>.
14. RNACentral Consortium. 2017. RNACentral: a comprehensive database of non-coding RNA sequences. *Nucleic Acids Res* 45:D128–D134. <https://doi.org/10.1093/nar/gkw1008>.