

Draft Genome Sequence of Four Coccolithoviruses: *Emiliana huxleyi* Virus EhV-88, EhV-201, EhV-207, and EhV-208

Jozef I. Nissimov,^a Charlotte A. Worthy,^{a,b} Paul Rooks,^a Johnathan A. Napier,^b Susan A. Kimmance,^a Matthew R. Henn,^c Hiroyuki Ogata,^d and Michael J. Allen^a

Plymouth Marine Laboratory, Plymouth, United Kingdom^a; Department of Biological Chemistry, Rothamsted Research, Harpenden, Hertfordshire, United Kingdom^b; The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA^c; and Structural and Genomic Information Laboratory, Mediterranean Institute of Microbiology (IFR-88), Aix-Marseille University, Marseille, France^d

The Coccolithoviridae are a group of viruses which infect the marine coccolithophorid microalga *Emiliana huxleyi*. The *Emiliana huxleyi* viruses (known as EhVs) described herein have 160- to 180-nm diameter icosahedral structures, have genomes of approximately 400 kbp, and consist of more than 450 predicted coding sequences (CDSs). Here, we describe the genomic features of four newly sequenced coccolithoviruses (EhV-88, EhV-201, EhV-207, and EhV-208) together with their draft genome sequences and their annotations, highlighting the homology and heterogeneity of these genomes to the EhV-86 model reference genome.

The coccolithoviruses infect *Emiliana huxleyi*, a globally distributed bloom-forming marine microalga. The abundance of *Emiliana huxleyi* viruses (EhVs) in natural seawater can typically reach 10^7 ml⁻¹ in bloom conditions and 10^8 to 10^9 ml⁻¹ under laboratory culture (6). Following the full sequencing of the model virus EhV-86 (isolated in 1999 from the English Channel) and the partial sequencing of EhV-163 (isolated from a Norwegian fjord the following year), we have recently undertaken to sequence the remaining English Channel isolates currently contained within the Plymouth Virus Collection (PVC) (1). The draft genomes of EhV-84, EhV-203, and EhV-202 have been described previously (4, 5). Here, we present the draft genomes of the remaining four coccolithoviruses in the PVC: EhV-88, EhV-201, EhV-207, and EhV-208.

EhV-88 was isolated from the English Channel (50°15'N/04°13'W) from a depth of 5 m in 1999, while EhV-201, EhV-207, and EhV-208 were isolated from the English Channel (49°56'N/04°19'W, 50°15'N/04°13'W, and 50°15'N/04°13'W, respectively) from a depth of 2 to 15 m in 2001 (1, 8). Their icosahedral virion structure and morphology are similar to those of other coccolithoviruses and of phycodnaviruses in general (9). Phylogenetic analysis of available major capsid protein (MCP) and DNA *pol* gene sequences indicates that the closest relatives to EhV-88 are EhV-84 and EhV-86 (4, 7), while EhV-201, EhV-207, and EhV-208 are closely related to EhV-203 (1, 8).

Genome sequencing, finishing, and annotation were performed by the Broad Institute. The genomes were sequenced using the 454 FLX pyrosequencing technology platform (Roche/454, Branford, CT). Library construction and sequencing were performed as previously described (2). General protocols for library construction can be found at <http://www.broadinstitute.org/annotation/viral/Phage/Protocols.html>. *De novo* genome assembly of resulting reads was performed using the Newbler v2.3 assembly software package (2). A total of 74,782, 78,268, 33,894, and 85,422 reads were produced and assembled into 8, 7, 16, and 17 contigs, comprising 396,598 bp, 406,701 bp, 420,391 bp, and 409,403 bp, for EhV-88, EhV-201, EhV-207, and EhV-208, respectively. Genes were identified using the Broad Institute's Automated Phage Annotation Protocol (2). Additional gene prediction

analysis and functional annotation was performed within the Integrated Microbial Genomes-Expert Review (IMG-ER) platform (3).

General features of the genomes include nucleotide compositions of 40.18%, 40.46%, 40.49%, and 40.42% G + C and 475, 451, 473, and 455 predicted protein coding sequences (CDSs) for EhV-88, EhV-201, EhV-207, and EhV-208, respectively. EhV-201, EhV-207, and EhV-208 have six tRNAs (Arg, Asn, Gln, Glu, Leu, and Lys), while EhV-88 has five (Arg, Asn, Gln, Ile, and Lys). EhV-84 is most similar to the model virus EhV-86, encoding 231 CDSs with identical homologues in the EhV-86 genome. In contrast, EhV-201, EhV-207, and EhV-208 have just 26, 29, and 25 CDSs sharing 100% identity with their EhV-86 homologues, respectively. The majority of CDSs not shared with EhV-86 encode hypothetical proteins of unknown function. Those of predicted function include those encoding glycosyltransferase (EhV-201, EhV-207, and EhV-208), methyltransferase (EhV-88, EhV-207, EhV-208), and RNase (EhV-88 and EhV-201). The genomes of these viruses will provide new insights into coccolithovirus evolution and their coevolution and interaction with *Emiliana huxleyi*.

Nucleotide sequence accession numbers. The nucleotide sequence accession numbers for the draft genomes sequences have been deposited in GenBank under accession numbers [JF974310](https://www.ncbi.nlm.nih.gov/nuccore/JF974310), [JF974311](https://www.ncbi.nlm.nih.gov/nuccore/JF974311), [JF974317](https://www.ncbi.nlm.nih.gov/nuccore/JF974317), and [JF974318](https://www.ncbi.nlm.nih.gov/nuccore/JF974318).

ACKNOWLEDGMENTS

This research was funded in part by the Gordon and Betty Moore Foundation through a grant to the Broad Institute (M.R.H.) and through the NERC Oceans 2025 program (M.J.A.). J.I.N. is supported by an NERC studentship, and C.A.W. is supported by a BBSRC Industrial CASE studentship sponsored by PML Applications. H.O. is supported by IGS/CNRS and ANR (grants ANR-09-PCS-GENM-218, ANR-08-BDVA-003).

Received 8 December 2011 Accepted 9 December 2011

Address correspondence to Michael J. Allen, mija@pml.ac.uk.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.07046-11

Samples G3265, G3266, G3249, and G3250 were sequenced, assembled, and annotated at the Broad Institute.

We thank Konstantinos Mavromatis from JGI who assisted with information regarding the IMG/ER platform and the Broad Institute's Genome Sequencing Platform, Finishing Team, and Annotation Team for their efforts to generate the genomic data. Jean Devonshire and the Centre for Bioimaging at Rothamsted provided technical support for transmission electron microscopy.

REFERENCES

1. Allen MJ, Martinez-Martinez J, Schroeder DC, Somerfield PJ, Wilson WH. 2007. Use of microarrays to assess viral diversity: from genotype to phenotype. *Environ. Microbiol.* 9:971–982.
2. Henn MR, et al. 2010. Analysis of high-throughput sequencing and annotation strategies for phage genomes. *PLOS One* 5:e9083.
3. Markowitz VM, et al. 2009. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 25:2271–2278.
4. Nissimov JI, et al. 2011. Draft genome sequence of the coccolithovirus EhV-84. *Stand. Genomic Sci.* 5:1–11.
5. Nissimov JI, et al. 2011. Draft genome sequence of the coccolithovirus *Emiliania huxleyi* virus 203. *J. Virol.* 85:13468–13469.
6. Schroeder DC, Oke J, Hall M, Malin G, Wilson WH. 2003. Virus succession observed during an *Emiliania huxleyi* bloom. *Appl. Environ. Microbiol.* 69:2484–2490.
7. Wilson WH, et al. 2005. Complete genome sequence and lytic phase transcription profile of a coccolithovirus. *Science* 309:1090–1092.
8. Wilson WH, et al. 2002. Isolation of viruses responsible for the demise of an *Emiliania huxleyi* bloom in the English Channel. *J. Mar. Biol. Assoc. U.K.* 82:369–377.
9. Wilson WH, Van Etten JL, Allen MJ. 2009. The Phycodnaviridae: the story of how tiny giants rule the world. *Curr. Top. Microbiol. Immunol.* 328:1–42.