Journal of Virology

Draft Genome Sequence of the Coccolithovirus Emiliania huxleyi Virus 202

Jozef I. Nissimov, Charlotte A. Worthy, Paul Rooks, Johnathan A. Napier, Susan A. Kimmance, Matthew R. Henn, Hiroyuki Ogata and Michael J. Allen

J. Virol. **2012, 86(4):2380. DOI: 10.1128/JVI.06863-11.**

JVI Journals.ASM.org

GENOME ANNOUNCEMENT

Draft Genome Sequence of the Coccolithovirus *Emiliania huxleyi* Virus 202

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, Herts, United Kingdom[b]; The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA[c]; and Structural and Genomics Information Laboratory, CNRS-UPR2589, Mediterranean Institute of Microbiology (IFR-88), Aix-Marseille University, Marseille, France[d]

***Emiliania huxleyi* virus 202 (EhV-202) is a member of the *Coccolithoviridae*, a group of viruses that infect the marine coccolithophorid *Emiliania huxleyi*. EhV-202 has a 160- to 180-nm-diameter icosahedral structure and a genome of approximately 407 kbp, consisting of 485 coding sequences (CDSs). Here we describe the genomic features of EhV-202, together with a draft genome sequence and its annotation, highlighting the homology and heterogeneity of this genome in comparison with the EhV-86 reference genome.**

Emiliania huxleyi is a globally distributed marine microalgae which forms blooms that can cover up to 100,000 km^2 (10). Coccolithoviruses have been shown to be a major cause of bloom termination, and their role in global biogeochemical cycling is gaining increasing attention (3). Coccolithovirus abundances can reach 10^7 ml^{-1} in natural seawater under bloom conditions (8). *Emiliania huxleyi* virus 86 (EhV-86), the model virus, was isolated in 1999 from the English Channel (50°13.79′N/04°9.59′W) and was sequenced in its entirety in 2005 (9, 10). EhV-163, isolated from a Norwegian fjord, was partially sequenced in 2006 (2), and more recently, permanent draft genomes for EhV-84 and EhV-203 (both isolated in the English Channel) were obtained (6, 7). *Emiliania huxleyi* virus 202 (EhV-202) was isolated from the English Channel (50°00.36′N/04°18.87′W) from a depth of 15 m on 27 July 2001 (1, 10). Like other coccolithoviruses and phycodnaviruses in general, EhV-202 has an icosahedral virion structure (11). Phylogenetic analysis of available major capsid protein (MCP) gene sequences indicates that the closest relatives of EhV-202 are EhV-201, EhV-203, EhV-207, and EhV-208 (1, 10).

EhV-202 genome sequencing, finishing, and annotation were performed by the Broad Institute. The genome was sequenced using the 454 FLX pyrosequencing technology platform (Roche/454, Branford, CT). Library construction and sequencing were performed as previously described (4). General protocols for library construction can be found at www.broadinstitute.org/annotation/viral/Phage/Protocols.html. Genome assembly of resulting reads was performed using the Newbler v2.3 software package (4). A total of 31,250 reads were produced and assembled into 12 contigs comprising 407,516 bp, with a maximum contig length of 137,441 bp, an average contig length of 33,868 bp, and an average sequencing coverage of approximately 32.5 times (± 5.8). Genes were identified using the Broad Institute's automated phage annotation protocol (4). Additional gene prediction analysis and functional annotation were performed within the Integrated Microbial Genomes-Expert Review (IMG-ER) platform (5).

General features of the EhV-202 genome sequence include a nucleotide composition of 40.30% G+C, a total of 485 predicted protein coding sequences (CDSs), and three tRNA genes (Arg, Asn, and Gln). Of the 485 CDSs annotated by IMG-ER, 93 have been annotated with functional product predictions. Four hundred one CDSs have homologues (>26% identity) within the EhV-86 genome, yet only one shares 100% identity. EhV-202 is the most genetically distinct coccolithovirus strain sequenced to date. EhV-202 contains 83 unique CDSs; while the majority are of unknown function, two CDSs have homology to glycosyltransferases, one to a zinc finger protein, one to ADP ribose pyrophosphatase, one to lipopolysaccharide (LPS) biosynthesis protein, and one to polyubiquitin. Among the EhV-202 CDSs displaying the lowest similarities to their EhV-86 homologues is a putative phosphoglycerate mutase (62.30%), an ATPase (60.86%), and a CDS containing a PDZ domain (42.62%). The CDSs displaying the highest similarities to their EhV-86 counterparts include MCP (99.80%), sterol desaturase (99.51%), and DNA polymerase (92.13%).

Nucleotide sequence accession number. The draft genome sequence has been deposited in GenBank under accession number HQ634145.

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 a NERC studentship; C.A.W. is supported by a BBSRC Industrial CASE studentship sponsored by PML Applications. H.O. is supported by the IGS/CNRS and ANR (grants ANR-09-PCS-GENM-218 and ANR-08-BDVA-003).

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. Sample G4014 was sequenced, assembled, and annotated at the Broad Institute.

Received 21 November 2011 **Accepted** 28 November 2011

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

doi:10.1128/JVI.06863-11

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. **Allen MJ, Schroeder DC, Donkin A, Crawfurd KJ, Wilson WH.** 2006. Genome comparison of two coccolithoviruses. Virol. J. **3**:15.
3. **Coyle KO, Pinchuk AI.** 2002. Climate-related differences in zooplankton density and growth on the inner shelf of the southeastern Bering Sea. Prog. Oceanogr. **55**:177–194.
4. **Henn MR, et al.** 2010. Analysis of high-throughput sequencing and annotation strategies for phage genomes. PloS One **5**:e9083.
5. **Markowitz VM, et al.** 2009. IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics **25**:2271–2278.
6. **Nissimov JI, et al.** 2011. Draft genome sequence of the coccolithovirus EhV-84. Stand. Genomic Sci. **5**:1.
7. **Nissimov JI, et al.** Draft genome sequence of the coccolithovirus *Emiliania huxleyi* virus 203. J. Virol. **85**:13468–13469.
8. **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.
9. **Wilson WH, et al.** 2005. Complete genome sequence and lytic phase transcription profile of a *Coccolithovirus*. Science **309**:1090–1092.
10. **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. UK **82**:369–377.
11. **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.