genomeA
Journals.ASM.org

The 474-Kilobase-Pair Complete Genome Sequence of CeV-01B, a Virus Infecting *Haptolina* (*Chrysochromulina*) *ericina* (Prymnesiophyceae)

Lucie Gallot-Lavallée,[a] António Pagarete,[b] Matthieu Legendre,[a] Sebastien Santini,[a] Ruth-Anne Sandaa,[b] Heinz Himmelbauer,[c] Hiroyuki Ogata,[d] Gunnar Bratbak,[b] Jean-Michel Claverie[a,e]

Structural and Genomic Information Laboratory (IGS), Aix-Marseille Université, CNRS UMR 7256 (IMM FR 3479), Marseille, France[a]; Department of Biology, University of Bergen, Bergen, Norway[b]; Institute of Biotechnology, University of Natural Resources and Life Sciences (BOKU), Vienna, Austria[c]; Institute for Chemical Research, Kyoto University, Uji, Japan[d]; Assistance Publique des Hôpitaux de Marseille (APHM), Marseille, France[e]

We report the complete genome sequence of CeV-01B, a large double-stranded DNA virus infecting the unicellular marine phytoplankton *Haptolina* (formerly *Chrysochromulina*) *ericina*. CeV-01B and its closest relative *Phaeocystis globosa* virus define an emerging subclade of the Megaviridae family with smaller genomes and particles than the originally described giant Mimiviridae infecting *Acanthamoeba*.

Received 11 October 2015 **Accepted** 16 October 2015 **Published** 3 December 2015

Citation Gallot-Lavallée L, Pagarete A, Legendre M, Santini S, Sandaa R-A, Himmelbauer H, Ogata H, Bratbak G, Claverie J-M. 2015. The 474-kilobase-pair complete genome sequence of CeV-01B, a virus infecting *Haptolina* (*Chrysochromulina*) *ericina* (Prymnesiophyceae). Genome Announc 3(6):e01413-15. doi:10.1128/genomeA.01413-15.

Copyright © 2015 Gallot-Lavallée et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported license.

Address correspondence to Jean-Michel Claverie, Jean-Michel.Claverie@univ-amu.fr.

The haptophyte *Haptolina* (formerly *Chrysochromulina*) *ericina* is a phytoplankton species with a worldwide distribution. It most commonly occurs in low numbers but has occasionally been observed to form blooms (1). Viruses are abundant in aquatic ecosystems and are increasingly recognized to play a significant role in the regulation of plankton populations, such as in the prevention or termination of blooms (2). However, only a few different algal host-virus systems have been put in culture and studied in details. Comprehensive genome analyses have been performed for six DNA viruses infecting *Chlorella* spp. (3, 4), six infecting Mamiellales green algae (5–8), two infecting marine brown algae (Phaeophyceae) (9, 10), and one each infecting *Emiliania huxleyi* (Coccolithophyceae) (11), *Phaeocystis globosa* (Prymnesiophyceae) (12), and *Aureococcus anophagefferens* (Pelagophyceae) (13).

Haptolina ericina virus CeV-01B was isolated from Norwegian coastal waters in 1998 (1). The virus replicates in the host cytoplasm with a lytic cycle lasting 14 to 19 h resulting in thousands of icosahedral particles 160 nm in diameter. Its genome size was previously estimated around 510 kbp by pulsed-field gel electrophoresis (1). DNA from purified CeV-01B particles was sequenced in 2013 on an Illumina HiSeq 2000 Platform. 3,626,569 × 2 pair-ended 100-nt high-quality reads (approximately 1,400-fold coverage of the CeV-01B genome) were generated and assembled using SOAPdenovo (14) with a stringent k-mer size ($k = 97$). Two scaffolds, with sizes of 67 kb and 410 kb were initially obtained. The scaffolder SSPACE (15), Gapfiller (16), and Bowtie (17) were used to fill up the gaps and to correct remaining sequencing errors.

The final 473,558-bp genome sequence exhibited a high A+T content of 75%. It was predicted to encode 512 open reading frames (ORFs), using GeneMark (18), and 12 tRNAs, using tRNAscan-SE (19). They span over 91% of the genome. Among the 512 predicted ORFs, 274 (53.5%) exhibited a significant homolog in NCBI's nonredundant protein sequence database (BlastP, E value $< 10^{-5}$), of which 163 (59.5%) were most similar to their PgV homolog and an additional 40 had their closest homologue in other Megaviridae. These best matches in PgV include usual phylogenetic markers such as the DNA polymerase B (CeV_365, 45% identical to PgV's PGCG_248), the major capsid protein (CeV_191, 73% identical to PgV's PGCG_157), as well as two enzymes characteristic of Megaviridae: the mismatch DNA repair enzyme MutS7 (20) (CeV_281 47% identical to PgV's PGCG_223) and an asparagine synthetase (21) (CeV_376, 49% identical to PgV's PGCG_327). Moreover, CeV and PgV share gene fusion between their DNA polymerase X and NAD-dependent DNA ligase (CeV_489, 49% identical to PGCG_401) not found in other Megaviridae. Finally, the newly described PgV-MIGE mobile element of which 12 copies were found in the PgV genome (13) was also found in 6 copies in the CeV-01B genome.

The proposed Megaviridae family, initially composed of *Acanthamoeba*-infecting giant viruses (now constituting the proposed Mimiviridae subfamily) (21), progressively expanded to encompass smaller members infecting other unicellular protists (12, 13, 21). Its genome sequence clearly classifies CeV-01B in a subclade of "small" Megaviridae (12) with PgV as the closest, nevertheless distant relative.

Nucleotide sequence accession number. The completely annotated genomic sequence of *Haptolina ericina* virus CeV-01B is available in Genbank under accession number KT820662.

ACKNOWLEDGMENTS

L.G.-L. is a recipient of a Ph.D. award from Aix-Marseille University. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

We thank Guillaume Blanc for advice on sequence annotation. The data were produced at the Genomics Unit of the Center for Genomic

Regulation (CRG), Barcelona, Spain. We thank Anna Ferrer Salvador for assistance and Manuela Hummel for an initial bioinformatic analysis of the data.

REFERENCES

1. **Sandaa R, Heldal M, Castberg T, Thyrhaug R, Bratbak G**. 2001. Isolation and characterization of two viruses with large genome size infecting *Chrysochromulina ericina* (*Prymnesiophyceae*) and *Pyramimonas orientalis* (*Prasinophyceae*).Virology **290:**272–280. http://dx.doi.org/10.1006/viro.2001.1161.
2. **Wommack KE, Colwell RR**. 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol Mol Biol Rev **64:**69–114. http://dx.doi.org/10.1128/MMBR.64.1.69-114.2000.
3. **Van Etten JL, Dunigan DD**. 2012. Chloroviruses: not your everyday plant virus. Trends Plant Sci **17:**1–8. http://dx.doi.org/10.1016/j.tplants.2011.10.005.
4. **Jeanniard A, Dunigan DD, Gurnon JR, Agarkova IV, Kang M, Vitek J, Duncan G, McClung O, Larsen M, Claverie JM, Van Etten JL, Blanc G**. 2013. Towards defining the chloroviruses: a genomic journey through a genus of large DNA viruses. BMC Genomics **14:**158. http://dx.doi.org/10.1186/1471-2164-14-158.
5. **Moreau H, Piganeau G, Desdevises Y, Cooke R, Derelle E, Grimsley N**. 2010. Marine prasinovirus genomes show low evolutionary divergence and acquisition of protein metabolism genes by horizontal gene transfer. J Virol **84:**12555–12563. http://dx.doi.org/10.1128/JVI.01123-10.
6. **Weynberg KD, Allen MJ, Ashelford K, Scanlan DJ, Wilson WH**. 2009. From small hosts come big viruses: the complete genome of a second *Ostreococcus tauri* virus, OtV-1. Environ Microbiol **11:**2821–2839. http://dx.doi.org/10.1111/j.1462-2920.2009.01991.x.
7. **Weynberg KD, Allen MJ, Gilg IC, Scanlan DJ, Wilson WH**. 2011. Genome sequence of *Ostreococcus tauri* virus OtV-2 throws light on the role of picoeukaryote niche separation in the ocean. J Virol **85:**4520–4529. http://dx.doi.org/10.1128/JVI.02131-10.
8. **Derelle E, Ferraz C, Escande M, Eychenié S, Cooke R, Piganeau G, Desdevises Y, Bellec L, Moreau H, Grimsley N**. 2008. Life-cycle and genome of OtV5, a large DNA virus of the pelagic marine unicellular green alga *Ostreococcus tauri*. PLoS One **3:**e2250. http://dx.doi.org/10.1371/journal.pone.0002250.
9. **Delaroque N, Müller DG, Bothe G, Pohl T, Knippers R, Boland W**. 2001. The complete DNA sequence of the *Ectocarpus siliculosus* virus EsV-1 genome. Virology **287:**112–132. http://dx.doi.org/10.1006/viro.2001.1028.
10. **Schroeder DC, Park Y, Yoon HM, Lee YS, Kang SW, Meints RH, Ivey RG, Choi TJ**. 2009. Genomic analysis of the smallest Giant virus—Feldmannia sp. Virus 158. Virology **384:**223–232. http://dx.doi.org/10.1016/j.virol.2008.10.040.
11. **Wilson WH, Schroeder DC, Allen MJ, Holden MT, Parkhill J, Barrell BG, Churcher C, Hamlin N, Mungall K, Norbertczak H, Quail MA, Price C, Rabbinowitsch 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. http://dx.doi.org/10.1126/science.1113109.
12. **Santini S, Jeudy S, Bartoli J, Poirot O, Lescot M, Abergel C, Barbe V, Wommack KE, Noordeloos AA, Brussaard CP, Claverie JM**. 2013. Genome of Phaeocystis globosa virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes. Proc Natl Acad Sci USA **110:**10800–10885. http://dx.doi.org/10.1073/pnas.1303251110.
13. **Moniruzzaman M, LeCleir GR, Brown CM, Gobler CJ, Bidle KD, Wilson WH, Wilhelm SW**. 2014. Genome of brown tide virus (AaV), the little giant of the Megaviridae, elucidates NCLDV genome expansion and host-virus coevolution. Virology **466–467:**60–70. http://dx.doi.org/10.1016/j.virol.2014.06.031.
14. **Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Li S, Yang H, Wang J, Wang J**. 2010. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res **20:**265–272. http://dx.doi.org/10.1101/gr.097261.109.
15. **Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W**. 2011. Scaffolding pre- assembled contigs using SSPACE. Bioinformatics **27:**578–579. http://dx.doi.org/10.1093/bioinformatics/btq683.
16. **Nadalin F, Vezzi F, Policriti A**. 2012. GapFiller: a de novo assembly approach to fill the gap within paired reads. BMC Bioinformatics **13:**S8. http://dx.doi.org/10.1186/1471-2105-13-S14-S8.
17. **Langmead B, Salzberg SL**. 2012. Fast gapped-read alignment with bowtie 2. Nat Methods **9:**357–359. http://dx.doi.org/10.1038/nmeth.1923.
18. **Borodovsky M, Lomsadze A**. 2014. Gene identification in prokaryotic genomes, phages, metagenomes, and EST sequences with GeneMarkS suite. Curr Protoc Microbiol **32:**1–17. http://dx.doi.org/10.1002/9780471729259.mc01e07s32.
19. **Lowe TM, Eddy SR**. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res **25:**955–964. http://dx.doi.org/10.1093/nar/25.5.0955.
20. **Ogata H, Ray J, Toyoda K, Sandaa R, Nagasaki K, Bratbak G, Claverie JM**. 2011. Two new subfamilies of DNA mismatch repair proteins (MutS) specifically abundant in the marine environment. ISME J **5:**1143–1151. http://dx.doi.org/10.1038/ismej.2010.210.
21. **Mozar M, Claverie JM**. 2014. Expanding the Mimiviridae family using asparagine synthase as a sequence bait. Virology **466–467:**112–122. http://dx.doi.org/10.1016/j.virol.2014.05.013.