Science
AAAS

Viral Glycosphingolipids Induce Lytic Infection and Cell Death in Marine Phytoplankton

Assaf Vardi, *et al.*

Science **326**, 861 (2009);
DOI: 10.1126/science.1177322

related prasinovirus was confirmed by PCR amplification and sequencing of a conserved region encoding its major capsid protein (fig. S4). However, the <78% amino acid identity in the major capsid protein and the fact that some regions of the genome were not identified in spite of the high sequence coverage in the virome indicate that this Antarctic prasinovirus is distinct from OtV5. The change in the viral community composition before and after ice melt, observed with five different search strategies, was consistent with the quantification of virus particles (fig. S2) and probably reflects changes in physical factors and host community (*6*).

Restricted metabolic activity under the ice promotes lysogeny in large Antarctic phages (*15*) but may allow the lytic replication of ssDNA viruses with small genomes, which are more abundant in the spring sample. The high prevalence of genes involved in photosynthesis and other cellular metabolic processes in phages and viromes has led others to propose that these genes might facilitate the expansion of their hosts into new ecological niches (*25*–*27*). Similarly, the increase we observed in the Antarctic summer virome of genes involved in carbohydrate and amino acid metabolism, respiration, and stress responses suggests that they may help the infected host to better survive the changing environmental conditions after the ice cover melts (fig. S6). Host expansion will also favor virus populations; for example, the phycodnavirus expansion we have observed in the summer sample is probably a consequence of a Prasinophyceae green alga bloom. Scales of the prasinophyte *Pyramimonas geledicola*, the dominant phytoflagellate in some Antarctic lakes, have been observed as associated with large viral particles (*28*) in Antarctica. The progressive reduction of Lake Limnopolar ice cover during December provided better radiation transmission to the water column triggering the growth of phytoplankton just under the ice sheet. The algal bloom was detected in December both by a clear chlorophyll *a* peak and a reduction in light adsorption below 1 to 2 m (fig. S7), suggesting that it may represent the early expansion of the algal host of the phycodnavirus that dominated the lake in January (summer sample).

The viral assemblage of Lake Limnopolar reported here is unexpected by comparison with other aquatic environments. First, in contrast with other known aquatic viromes, this Antarctic lake viral community is not dominated by bacteriophages infecting prokaryotes (*12*, *21*) but by viruses known to infect eukaryotes, including ssDNA viruses related to animal and plant viruses, and dsDNA viruses infecting algae. Second, undescribed ssDNA viruses possibly belonging to distinct viral families were found. Third, unprecedented taxonomic diversity and high genetic richness in this Antarctic lake illustrates that high virus diversity may be found in systems where biological diversity in other taxa is low (*2*, *3*, *19*). Finally, we found that melting of the ice cover in summer leads to a change in this Antarctic viral community from small ssDNA to large dsDNA viruses. Altogether, the Lake Limnopolar virome sheds light into the largely unknown community of viruses populating Antarctic freshwater ecosystems.

References and Notes

1. L. S. Peck *et al.*, *Polar Biol.* **28**, 351 (2005).
2. P. Convey, M. I. Stevens, *Science* **317**, 1877 (2007).
3. P. Convey *et al.*, *Biol. Rev. Cambridge Philos. Soc.* **83**, 103 (2008).
4. J. Laybourn-Parry, D. A. Pearce, *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **362**, 2273 (2007).
5. D. A. Pearce, W. H. Wilson, *Antarct. Sci.* **15**, 319 (2003).
6. D. A. Cowan, L. A. Tow, *Annu. Rev. Microbiol.* **58**, 649 (2004).
7. J. Laybourn-Parry, *Science* **324**, 1521 (2009).
8. C. A. Suttle, *Nat. Rev. Microbiol.* **5**, 801 (2007).
9. J. A. Fuhrman, *Nature* **399**, 541 (1999).
10. N. H. Mann, A. Cook, A. Millard, S. Bailey, M. Clokie, *Nature* **424**, 741 (2003).
11. S. J. Williamson *et al.*, *PLoS One* **3**, e1456 (2008).
12. F. E. Angly *et al.*, *PLoS Biol.* **4**, e368 (2006).
13. E. A. Dinsdale *et al.*, *Nature* **452**, 629 (2008).
14. M. Toro *et al.*, *Polar Biol.* **30**, 635 (2007).
15. C. Sawstrom, J. Lisle, A. M. Anesio, J. C. Priscu, J. Laybourn-Parry, *Extremophiles* **12**, 167 (2008).
16. R. L. Kepner, R. A. Wharton, C. A. Suttle, *Limnol. Oceanogr.* **43**, 1754 (1998).
17. W. H. Wilson, D. Lane, D. A. Pearce, J. C. Ellis-Evans, *Polar Biol.* **23**, 657 (2000).
18. Materials and methods are available as supporting material on *Science* Online.
19. S. L. Chown, P. Convey, *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **362**, 2307 (2007).
20. J. A. Fuhrman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7774 (2008).
21. C. Desnues *et al.*, *Nature* **452**, 340 (2008).
22. J. Filee, F. Tetart, C. A. Suttle, H. M. Krisch, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12471 (2005).
23. K. H. Kim *et al.*, *Appl. Environ. Microbiol.* **74**, 5975 (2008).
24. E. Derelle *et al.*, *PLoS One* **3**, e2250 (2008).
25. M. B. Sullivan, M. L. Coleman, P. Weigele, F. Rohwer, S. W. Chisholm, *PLoS Biol.* **3**, e144 (2005).
26. M. Breitbart *et al.*, *Environ. Microbiol.* **11**, 16 (2009).
27. F. Rohwer, R. V. Thurber, *Nature* **459**, 207 (2009).
28. J. Laybourn-Parry, J. S. H. R. Sommaruga, *Freshwater Biol.* **46**, 1279 (2001).

29. We thank the Maritime Technology Unit (CSIC) and Las Palmas crew (Spanish Navy) for the logistic help and support that made this expedition possible. We also thank M. Toro, A. Camacho, and other members of the Limnopolar Project for help and discussions; A. Alejo for helpful comments and reviewing the manuscript; and M. Pignatelli for technical support. This work was funded by grants from the Spanish Ministry of Science and Innovation (CGL2005-06549-C02-01/ANT, CGL2007-29843-E/ANT, CTM2008-05134-E/ANT, and BFU2008-04501-E). Spring and summer viral metagenomes from the Antarctic Lake Limnopolar have been submitted to GenBank and assigned the genome project accession number 34669. The metagenomes are also publicly accessible from the ftp server of the SEED public database (ftp://ftp.theseed.org/metagenomes) under the project accession numbers 4441778.3 and 4441558.3 for the spring and summer viromes, respectively. PCR-obtained sequences encoding the gp23 protein of Antarctic T4-phages and the MCP protein of Antarctic phycodnavirus have been also deposited at GenBank as "environmental sequences" and are listed under accession numbers FJ791185-247 and FJ791175-84, respectively.

Supporting Online Material

www.sciencemag.org/cgi/content/full/326/5954/858/DC1
Materials and Methods
Figs. S1 to S7
Tables S1 to S9
References

6 April 2009; accepted 3 September 2009
10.1126/science.1179287

Viral Glycosphingolipids Induce Lytic Infection and Cell Death in Marine Phytoplankton

Assaf Vardi,[1]* Benjamin A. S. Van Mooy,[2] Helen F. Fredricks,[2] Kimberly J. Popendorf,[2] Justin E. Ossolinski,[2] Liti Haramaty,[1] Kay D. Bidle[1]†

Marine viruses that infect phytoplankton are recognized as a major ecological and evolutionary driving force, shaping community structure and nutrient cycling in the marine environment. Little is known about the signal transduction pathways mediating viral infection. We show that viral glycosphingolipids regulate infection of *Emiliania huxleyi*, a cosmopolitan coccolithophore that plays a major role in the global carbon cycle. These sphingolipids derive from an unprecedented cluster of biosynthetic genes in *Coccolithovirus* genomes, are synthesized de novo during lytic infection, and are enriched in virion membranes. Purified glycosphingolipids induced biochemical hallmarks of programmed cell death in an uninfected host. These lipids were detected in coccolithophore populations in the North Atlantic, which highlights their potential as biomarkers for viral infection in the oceans.

Marine phytoplankton are the basis of marine food webs and are responsible for nearly half the global carbon-based net primary production (*1*). The coccolithophorid *Emiliania huxleyi* (Prymnesiophyceae, Haptophyte) is a cosmopolitan unicellular photoautotroph whose intricate calcite skeletons account for about a third of the total marine $CaCO_3$ production. *E. huxleyi* forms massive annual blooms in the North Atlantic that have been shown to be infected and terminated by lytic, giant double-stranded DNA containing coccolithoviruses (*2*, *3*), a subset of the larger *Phycodnaviridae* group that infects microalgae (*4*). As the most abundant biological entities in aquatic environments, viruses turn over more than a quarter of the photosynthetically fixed carbon, thereby fueling microbial food webs and short-circuiting carbon export to higher trophic levels and the deep sea (*5*, *6*). In addition, marine viruses stimulate the lateral trans-

fer of genes from one host cell to another, which contributes to the diversification and adaptation of plankton in the oceans (*7*).

Very little is known about the molecular mechanisms and signal transduction pathways mediating phytoplankton cell death by marine viruses. However, it was recently shown that lytic viral infection of *E. huxleyi* induces hallmarks of autocatalytic, programmed cell death (PCD), including metacaspase expression and caspase activity, which were required for successful viral replication (*8*). The recent availability of genomic resources for an *E. huxleyi* host (*9*) and *E. huxleyi* lytic virus strain 86 (EhV86) (*10*) provides an unprecedented opportunity to explore cellular pathways triggered during execution of viral infection and to gain insights into the origin of PCD in these unicellular photoautotrophs. The genome sequence of EhV86, the type strain for the *Coccolithoviruses*, revealed an unexpected cluster of putative sphingolipid biosynthetic genes (*10*), a pathway that has not been described in a viral genome. Recent phylogenetic evidence for the transfer of seven genes in the sphingolipid biosynthesis pathway between *E. huxleyi* and EhV86 suggests a critical role in host-virus interactions (*11*). De novo sphingolipid biosynthesis is initiated by serine palmitoyltransferase (SPT) (*12*) and leads to ceramide production, a potent inducer of PCD in animals and plants (*13*, *14*). The EhV86-encoded SPT gene is expressed during infection (*10*, *15*) and encodes an active SPT with a unique biochemical preference for myristoyl–coenzyme A (myristoyl-CoA) as a substrate, when expressed heterologously in yeast (*16*). Nonetheless, it is not known whether these lipids play any functional role during infection of *E. huxleyi*.

We examined the polar membrane composition of uninfected and EhV86-infected sensitive (*Ehux374*) and resistant (*Ehux373*) *E. huxleyi* strains during the course of lytic infection. Using high-performance liquid chromatography with electrospray ionization mass spectrometry (HPLC/ESI-MS) (*17*, *18*), we compared the lipid composition of uninfected and infected host cells. We detected glycosphingolipids (GSLs) in uninfected host cells that appeared to be composed of predominantly hydroxyl-sphingoid bases derived from palmitoyl-CoA (Fig. 1A). These host sphingoid bases are consistent with the expected products of the host SPT, which utilizes palmitoyl-CoA, and are common in plants (*19*). However, the lipids from EhV86-infected *Ehux374* had unique GSLs yielding fragmentation ions that were indicative of multiply hydroxylated sphingoid bases derived from myristoyl-CoA (Fig. 1B and fig. S1). These sphingoid bases are the expected products of the viral SPT, based on its preference for myristoyl-CoA (*16*). The virus-induced myristoyl-base GSLs were absent in uninfected cells and were unique to lytic viral infection (Fig. 1B). Both resistant and susceptible hosts produced significant concentrations of host palmitoyl GSLs, which were structurally distinct from the viral myristoyl GSLs (Fig. 1, A and D).

To ascertain the origin of the myristoyl GSLs, viruses were purified using a cesium chloride ($CsCl_2$) density gradient and ultracentrifugation. Identical myristoyl GSLs were dominant components of the lipids extracted from purified EhV86 viruses (Fig. 1C); host palmitoyl GSLs were absent. These observations indicate that the viral GSLs are a component of the membranes packaged with this virus. Although GSLs are distributed in some prokaryotes, they have not been found in phytoplankton or microalgal viral membranes (*20*). Proteomic analysis of the EhV86 virion determined that 23 out of 28 proteins are predicted to be membrane proteins (*21*), which corroborates our observations of GSLs and membrane structures. Sphingolipids, such as GSLs, are common constituents of membrane lipids, which may form lipid rafts in eukaryotes. For example, lipid rafts may be involved in the entry and budding of HIV and hepatitis C, yet their role is not well understood (*20*, *22*). Transmission electron microscopy (TEM) micrographs of infected *Ehux374* cells revealed viral particles within intracytoplasmic vacuoles, consistent with this type of strategy (Fig. 1B, inset).

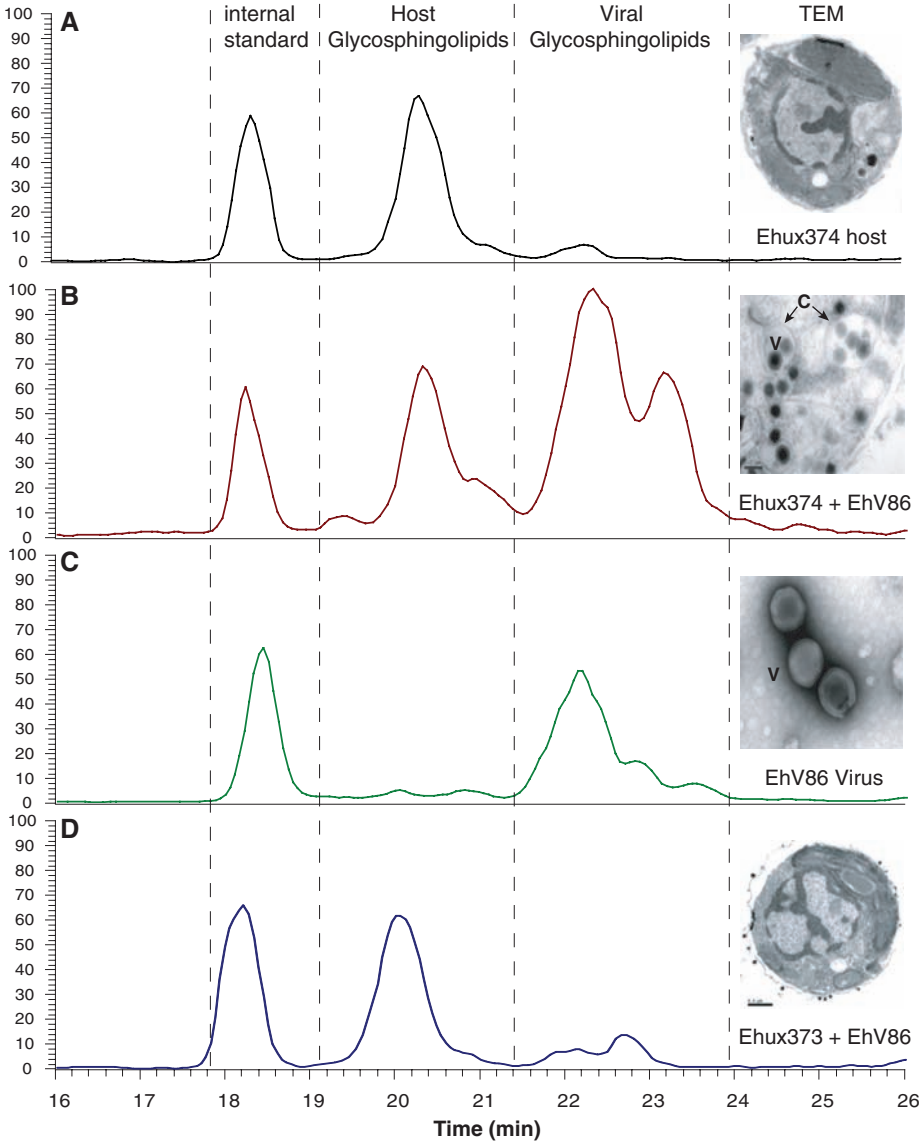

Fig. 1. Production of viral GSLs in EhV86-infected *E. huxleyi* cells and in purified EhV86 virions. Summed ion HPLC/MS chromatograms showing relative abundances (normalized to an internal standard) of GSLs extracted from (**A**) susceptible *Ehux374*, (**B**) *Ehux374* infected with EhV86 52 hours post infection, (**C**) purified EhV86 on a $CsCl_2$ gradient, and (**D**) resistant *Ehux373* infected with EhV86 52 hours post infection. (Insets) TEM micrographs of respective treatments. Arrows in (B) depict intracytoplasmic vacuoles (C) that contain viral particles (V).

[1]Environmental Biophysics and Molecular Ecology Group, Institute of Marine and Coastal Sciences, Rutgers University, 71 Dudley Road, New Brunswick, NJ 08901, USA. [2]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA.

*Present address: Department of Plant Sciences, Weizmann Institute of Science, Rehovot 76100, Israel.

†To whom correspondence should be addressed: bidle@marine.rutgers.edu

The accumulation of viral GSLs in *Ehux374* during EhV86 infection was accompanied by a reduction in cell abundance, a severely compromised photochemical quantum yield of photosystem II (PSII) (declining to 0.22 after 48 hours), and an ~30-fold induction in caspase specific activity (Fig. 2). Induction of this biochemical PCD marker occurred concomitantly with de novo synthesis of viral GSLs and viral replication (reaching 3×10^8 virus/ml), leading to the demise of *Ehux374* at the onset of the lytic phase 25 hours post infection. Although elevated viral GSLs were detected at 3.5 hours in EhV86-infected cultures, presumably because of the presence of free viruses, de novo production of viral GSLs began after the first 12 hours (Fig. 2D, inset), trailing gene expression dynamics of the virally encoded SPT at 2 hours post infection (*15*). In late infection (>50 hours), caspase specific activity and GSL production exceeded levels seen in uninfected cells or in resistant *Ehux373* cells by a factor of >100 (Fig. 2, C and D). Linear regression demonstrated a high correlation ($R^2 = 0.815$) between viral GSL production and caspase specific activity over the course of lytic viral infection (fig. S2). In contrast, the EhV86-resistant strain *Ehux373* exhibited slightly better growth than control, uninfected *Ehux374* cells (Fig. 2A). Only trace levels of viral GSLs were detected in infected *Ehux373*, probably originating from the viral inoculum (Figs. 1D and 2D).

We monitored the ability of viral GSLs to modulate host physiology after purification from EhV86-infected *Ehux374* cells by preparative HPLC and addition to uninfected *Ehxu374* at various concentrations (Fig. 3). The viral GSLs suppressed cell growth compared with control cells treated with dimethyl sulfoxide (DMSO, a solvent) and control cells treated with phosphatidylglycerol (PG), which had a similar HPLC retention time (Fig. 3A). The cells treated with viral GSLs exhibited a dose-dependent induction of cell death above a threshold concentration (>0.06 μg/ml). Induction of cell death compromised photosynthetic efficiency of PSII (Fig. 3A) and resulted in elevated in vivo caspase activity for about 20 to 25% of *Ehux374* cells treated with 0.3 and 1.5 μg/ml of viral GSLs after 48 hours (Fig. 3B). The induction of in vivo caspase activity was assessed by cell staining with the fluorescently labeled caspase probe, VAD-FMK conjugated with fluorescein isothiocyanate (FITC), and flow cytometric analysis (Fig. 3B) (*18*). Likewise, 18 and 56.2% of the cells at these two concentrations were positively stained with SYTOX, a DNA binding fluorescent indicator of compromised cell membrane integrity (Fig. 3C); this observation is consistent with previous findings of late stages of PCD in phytoplankton (*23*). Only 5.6 to 10.3% of cells were positively stained with VAD-FMK-FITC in the control treatments and in cells that were treated with a sublethal GSL concentration (e.g., 0.06 μg/ml), which indicated minimal induction of PCD. Likewise, a visual comparison of control (DMSO or PG) and GSL treatments revealed massive cell lysis only at GSL concentrations >0.06 μg/ml (Fig. 3C).

Given their potent ability to trigger *E. huxleyi*'s PCD response in a dose-dependent manner and their presence in purified EhV virions, we propose that viral GSLs may be part of a timing mechanism for viral release. In such a mechanism, host lysis is dependent on the accumulation of viral myristoyl GSLs to a critical effective concentration, above which host PCD is induced. According to our measurements, an EhV86 virion contains ~0.1 to 0.3 fg of myristoyl GSLs. At a typical burst size of ~800 to 1000 viruses per cell (*8*), an effective intracellular concentration of about 80 to 300 fg per cell is reached at the peak of the lytic phase; this concentration is consistent with the 100 to 200 fg per cell observed at the onset of lytic infection and PCD activation (>50 hours post infection) (Fig. 2, C and D, and fig. S3). Given that ceramide enrichment in cell membranes can serve to modulate entry and release of viruses (*20*, *22*), GSL enrichment in intact EhV86 virions and the profound accumulation of viral GSL during lytic phase may suggest a similar mechanism for the timed release of EhV86 virions.

We hypothesize that such bioactive molecules have the potential also to elicit cell death in surrounding, uninfected cells under natural bloom densities and, hence, may serve as a bloom termination signal. It has been suggested that induction of PCD in *E. huxleyi* can act as a "viral exclusion" strategy (*8*, *24*) to limit production of viruses during infection and, ultimately, their propagation through clonal populations. We calculated the "sphere of influence" distance around a given lysed cell that would contain an adequate viral GSL concentration to induce PCD in surrounding cells. Using our measurements of viral GSL concentrations (~200 fg per cell) and a threshold concentration of 0.2 μg/ml to induce PCD and cell lysis (fig. S4) and assuming the GSL is isotropically dispersed, we estimated that a sphere with a diameter of a few hundred micrometers would be sufficient to elicit cell death. Given that *E. huxleyi* blooms can reach about 100,000 cells/ml (*25*) and that cells clearly respond to external GSL application (Fig. 3), it is plausible that GSLs act as intercellular signals. Similar findings were recently reported for diatom-derived oxylipins found to act as infochemicals either to potentiate PCD or to induce resistance in sublethal doses (*26*, *27*).

Because of their potent bioactivity and unique chemical signature, we also propose virus-derived

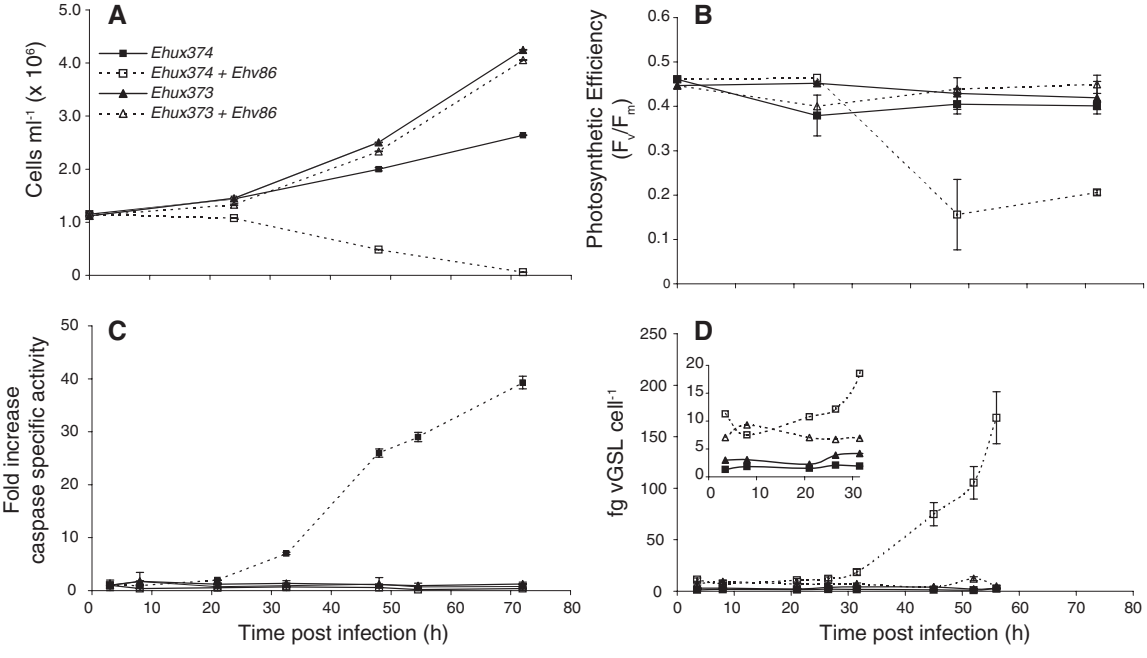

Fig. 2. Onset of the lytic phase during EhV86 infection is mediated by induction of caspase activity and viral GSL production. Viral infection dynamics of susceptible *Ehux374* or resistant *Ehux373* strains as monitored by the following parameters: (**A**) host abundance; (**B**) photochemical quantum yield (F_v/F_m) of PS II , a proxy for photosynthetic health; (**C**) caspase specific activity via normalized cleavage of IETD-AFC in cell extracts; and (**D**) de novo synthesis of viral, myristoyl GSLs. Data in (A) to (C) is derived from four to six biological replicates for infected cultures. Analytical error for GSL measurement is from two independent experiments, each analyzed in triplicate.

GSLs can act as a novel biomarker for viral infection of natural phytoplankton assemblages. We collected natural plankton for GSL analyses during a cruise in the North Atlantic along 65°W between 43°N and 33°N during April of 2008. We used concentrations of 19′-hexanoyloxyfucoxanthin (19′-hex), a pigment marker for prymnesiophytes, including coccolithophores, as a proxy for host abundances (Fig. 4). Palmitoyl GSLs, similar to those seen in uninfected *Ehux374* cultures, were observed throughout the North Atlantic transect at stations characterized by high 19′-hex pigment concentrations, which we postulate are indicative of healthy coccolithophore populations. In contrast, at a local minimum in 19′-hex pigment concentrations, we observed a GSL molecule with an HPLC retention time and mass spectrum that was consistent with the myristoyl GSL observed in EhV86-infected cultures of *Ehux374* (compare Fig. 1, Fig. 4, and fig. S3). Together with our laboratory-based findings showing viral GSL production during viral lysis and its incorporation into virions (Fig. 1, 2), the observation of a putative viral GSL at the 19′-hex minimum is suggestive of virus-induced demise of a natural *E. huxleyi* population at this location. Detected expression of some host and viral sphingolipid biosynthetic genes in natural *E. huxleyi* populations corroborate our findings (*28*). On the basis of the diagnostic preference of the viral SPT for myristoyl-CoA (*16*), these viral GSLs may serve as a proxy for viral infection of coccolithophore populations in the sea. Currently, biomarkers to quantify active viral infection in the oceans are lacking, which hinders our understanding of the role and activity of viruses and virus-mediated

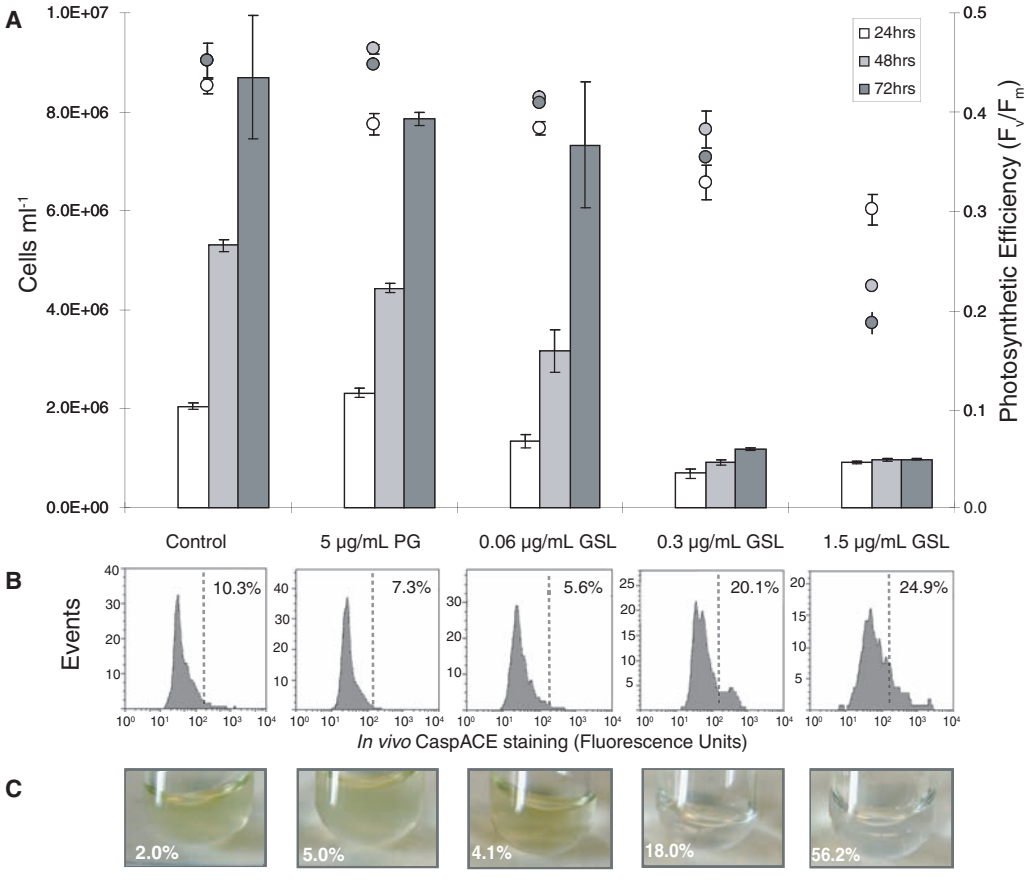

Fig. 3. Application of purified viral GSL to uninfected *E. huxleyi* cells mimics infection by inducing PCD. Dose-dependent induction of cell death in uninfected *Ehux374* cells over 72 hours by application of purified GSLs. (**A**) Cell abundance (bars) and photochemical quantum yield of PS II (circles). Error bars, standard deviation of four biological replicates. (**B**) In vivo caspase activity (measured by flow cytometry). Cytograph plots represent the fluorescence distribution of 5000 cells 48 hours post treatment and after staining with VAD-FMK-FITC (CaspACE). The percentage of positively stained cells is given. The dashed line represents the threshold fluorescence above which cells are positively stained (as determined by unstained cells for each treatment). (**C**) Viral GSL-treated cultures exhibited massive cell lysis after 72 hours. The percentage of SYTOX-positive cells, which serves as a proxy for dying cells, is given. Control treatments consisted of DMSO (solvent) or phosphatidylglycerol (PG).

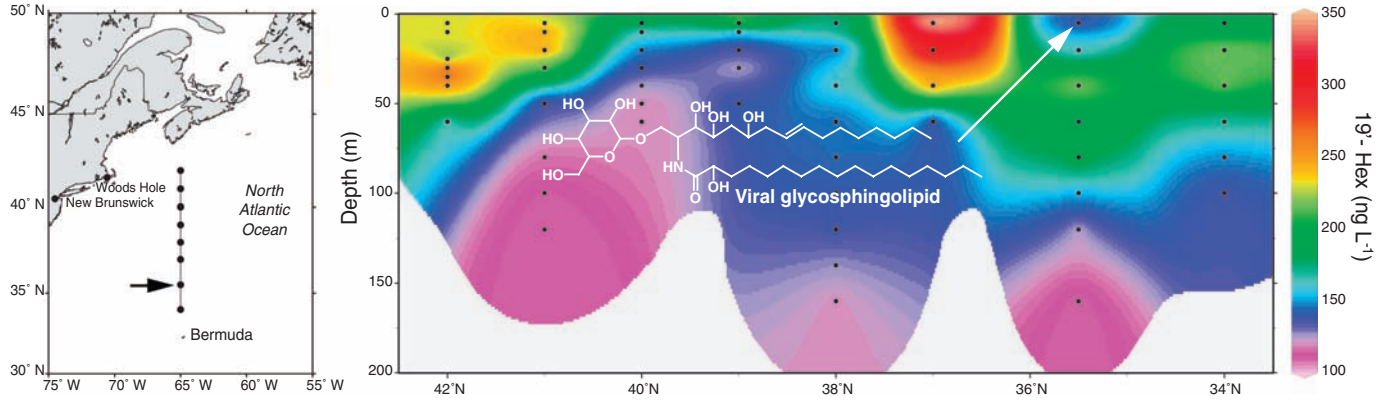

Fig. 4. Viral GSLs represent a potential biomarker for viral infection of natural coccolithophore populations. Vertical distribution of 19′-hex along a meridianal transect at 65°W in the North Atlantic (43°N to 33°N). A multiply hydroxylated myristoyl GSL, which we posit is of viral origin, was only detected in a distinct zone of minimum 19′-hex in pigment values (35°N 65°W, 5-m depth; designated by arrows). This observation is consistent with population demise by viral infection. (**Inset**) Tentative chemical structure of the observed myristoyl GSL.

processes in the oceans. We posit that future studies will elucidate the relative importance of biotic (viral infection as in this study) and abiotic [phosphorus limitation (*17*)] stress conditions, in shaping community structure of marine microbes, through the detection of different classes of stress-specific lipids.

Although the origin of PCD in unicellular organisms is still unclear, its functional conservation among phylogenetically diverse phytoplankton lineages suggests key evolutionary and ecological drivers in aquatic environments (*29*). The retention and expression of a nearly complete, virus-based sphingolipid biosynthetic pathway, along with its requirement for viral replication and regulation of the host cell fate, now underscores the pivotal role of a chemical-based, coevolutionary "arms race" in mediating host-virus interactions.

References and Notes

1. M. J. Behrenfeld *et al.*, *Science* **291**, 2594 (2001).
2. G. Bratbak, J. K. Egge, M. Heldal, *Mar. Ecol. Prog. Ser.* **93**, 39 (1993).
3. D. C. Schroeder, J. Oke, G. Malin, W. H. Wilson, *Arch. Virol.* **147**, 1685 (2002).
4. J. L. Van Etten, M. V. Graves, D. G. Müller, W. Boland, N. Delaroque, *Arch. Virol.* **147**, 1479 (2002).
5. J. A. Fuhrman, *Nature* **399**, 541 (1999).
6. C. A. Suttle, *Nat. Rev. Microbiol.* **5**, 801 (2007).
7. D. Lindell, J. D. Jaffe, Z. I. Johnson, G. M. Church, S. W. Chisholm, *Nature* **438**, 86 (2005).
8. K. D. Bidle, L. Haramaty, J. B. E. Ramos, P. Falkowski, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6049 (2007).
9. *E. huxleyi* CCMP1516 Main Genome Assembly, http://genome.jgi-psf.org/Emihu1/Emihu1.home.html.
10. W. H. Wilson *et al.*, *Science* **309**, 1090 (2005).
11. A. Monier *et al.*, *Genome Biol.* 10.1101/gr.091686.109 (2009).
12. K. Hanada, *Biochim. Biophys. Acta* **1632**, 16 (2003).
13. Y. A. Hannun, L. M. Obeid, *Trends Biochem. Sci.* **20**, 73 (1995).
14. H. Liang *et al.*, *Genes Dev.* **17**, 2636 (2003).
15. M. J. Allen *et al.*, *J. Virol.* **80**, 7699 (2006).
16. G. Han *et al.*, *J. Biol. Chem.* **281**, 39935 (2006).
17. B. A. S. Van Mooy *et al.*, *Nature* **458**, 69 (2009).
18. Materials and methods are available as supporting material on *Science* Online.
19. D. Lynch, T. Dunn, *New Phytol.* **161**, 677 (2004).
20. B. Brügger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2641 (2006).
21. M. J. Allen, J. A. Howard, K. S. Lilley, W. H. Wilson, *Proteome Sci.* **6**, 11 (2008).
22. H. Sakamoto *et al.*, *Nat. Chem. Biol.* **1**, 333 (2005).
23. A. Vardi *et al.*, *Curr. Biol.* **9**, 1061 (1999).
24. M. Frada, I. Probert, M. J. Allen, W. H. Wilson, C. de Vargas, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 15944 (2008).
25. M. Turkoglu, *J. Mar. Biol. Assoc. U. K.* **88**, 433 (2008).
26. A. Vardi *et al.*, *Curr. Biol.* **18**, 895 (2008).
27. A. Vardi *et al.*, *PLoS Biol.* **4**, e60 (2006).
28. A. Pagarete, M. J. Allen, W. H. Wilson, S. A. Kimmance, C. de Vargas, *Environ. Microbiol.* 10.1111/j.1462 (2009).
29. K. D. Bidle, P. G. Falkowski, *Nat. Rev. Microbiol.* **2**, 643 (2004).

30. The authors acknowledge M. Soule and E. Kujawinski and the funding sources of the Woods Hole Oceanographic Institution Fourier-Transform Mass Spectrometry Facility (NSF's Major Research Instrumentation Program OCE-0619608 and the Gordon and Betty Moore Foundation) for assistance with MS data acquisition. We thank D. Repeta for assistance and advice on pigment analysis, V. Starovoytov for TEM analysis, and C. Brown and M. Frada for helpful discussions and valuable feedback. This work was supported by NSF grant IOS-0717494 to K.D.B. and A.V., as well as NSF grant OCE-0646944 to B.V.M.

Supporting Online Material

www.sciencemag.org/cgi/content/full/326/5954/861/DC1
Materials and Methods
SOM Text
Figs. S1 to S4
References

5 June 2009; accepted 2 September 2009
10.1126/science.1177322

Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse

C. M. Wade,[1,2,3]* E. Giulotto,[4] S. Sigurdsson,[1,5] M. Zoli,[6] S. Gnerre,[1] F. Imsland,[5] T. L. Lear,[7] D. L. Adelson,[8] E. Bailey,[7] R. R. Bellone,[9] H. Blöcker,[10] O. Distl,[11] R. C. Edgar,[12] M. Garber,[1] T. Leeb,[11,13] E. Mauceli,[1] J. N. MacLeod,[7] M. C. T. Penedo,[14] J. M. Raison,[8] T. Sharpe,[1] J. Vogel,[15] L. Andersson,[5] D. F. Antczak,[16] T. Biagi,[1] M. M. Binns,[17] B. P. Chowdhary,[8] S. J. Coleman,[7] G. Della Valle,[6] S. Fryc,[1] G. Guérin,[19] T. Hasegawa,[20] E. W. Hill,[21] J. Jurka,[22] A. Kiialainen,[23] G. Lindgren,[24] J. Liu,[25] E. Magnani,[4] J. R. Mickelson,[26] J. Murray,[27] S. G. Nergadze,[4] R. Onofrio,[1] S. Pedroni,[14] M. F. Piras,[4] T. Raudsepp,[8] M. Rocchi,[28] K. H. Røed,[9] O. A. Ryder,[30] S. Searle,[15] L. Skow,[18] J. E. Swinburne,[31] A. C. Syvänen,[23] T. Tozaki,[32] S. J. Valberg,[26] M. Vaudin,[31] J. R. White,[1] M. C. Zody,[1,5] Broad Institute Genome Sequencing Platform,[1] Broad Institute Whole Genome Assembly Team,[1] E. S. Lander,[1,33] K. Lindblad-Toh[1,5]*

We report a high-quality draft sequence of the genome of the horse (*Equus caballus*). The genome is relatively repetitive but has little segmental duplication. Chromosomes appear to have undergone few historical rearrangements: 53% of equine chromosomes show conserved synteny to a single human chromosome. Equine chromosome 11 is shown to have an evolutionary new centromere devoid of centromeric satellite DNA, suggesting that centromeric function may arise before satellite repeat accumulation. Linkage disequilibrium, showing the influences of early domestication of large herds of female horses, is intermediate in length between dog and human, and there is long-range haplotype sharing among breeds.

As one of the earliest domesticated species, the horse, *Equus caballus*, has played an important role in human exploration of novel territories. Belonging to the order perissodactyla (i.e., odd-toed animals with hooves), the genus *Equus* radiated into 8 or 9 species around three million years ago (*1*). Members of the family equidae exhibit diverged karyotypes (*2*) and variable centromeric positioning (*1*). With over 90 hereditary conditions, which may serve as models for human disorders (*3*, *4*) (such as infertility, inflammatory diseases, and muscle disorders), the horse has much to offer as a model species.

DNA from a single mare of the Thoroughbred breed was sequenced to 6.8× coverage [supporting online material (SOM) text], resulting in a high-quality draft assembly (designated EquCab2.0) with a 112-kb N50 contig size (SOM) and a 46-Mb N50 scaffold size (tables S1 and S2), and >95% of the sequence anchored to the 64 (2N) equine chromosomes. The 2.5- to 2.7-Gb genome size is somewhat larger than the dog genome (2.5 Gb) and smaller than the human and bovine genomes (2.9 Gb) (*5*–*7*). Segmental duplications (*8*) make up <1% of the equine genome, and most are intrachromosomal duplications such as are seen in many other mammalian genomes (SOM). Repetitive sequences, many equine-specific, make up 46% of the genome assembly (SOM). The predominant repeat classes include long interspersed nuclear elements, dominated by L1 and L2 types (tables S3 and S4) (19% of bases), and short interspersed nuclear elements, including the recent ERE1 and ERE2 and the ancestral main immunogenic regions (7% of bases). Comparison of horse and human chromosomes reveals strong conserved synteny between these species (fig. S1). Indeed, 17 horse chromosomes (53%) comprise material from a single human chromosome (in the dog, it is 29%).

One unexpected feature of the horse genome landscape was the identification of an evolutionary new centromere (ENC) on chromosome 11 (ECA11), captured in an immature state. Several ENCs have been generated in the genus *Equus* by centromere repositioning (a shift of centromeric position without chromosome rearrangement) (*1*). Mammalian centromeres are typically complex structures characterized by the presence of satellite tandem repeats. ENCs are believed to form initially by unknown mechanisms in repeat-free regions and then progressively acquire extended arrays of satellite tandem repeats that may contribute to functional stability (*9*). The centromere of ECA11 resides in a large region of conserved synteny in many mammals, where the horse is the only species with a centromere present, strongly suggesting that this centromere is evolutionarily new. The ECA11 centromere is the only horse centromere lacking any hybridization signal in fluorescence in situ hybridization