Biochemical diversity of glycosphingolipid biosynthesis as a driver of *Coccolithovirus* competitive ecology

Jozef I. Nissimov [1,2] David Talmy [3] Liti Haramaty,[1]
Helen F. Fredricks [4] Ehud Zelzion,[5] Ben Knowles,[1]
A. Murat Eren [6,7] Rebecca Vandzura,[1]
Christien P. Laber,[1] Brittany M. Schieler,[1]
Christopher T. Johns,[1] Kuldeep D. More,[8]
Marco J. L. Coolen [8] Michael J. Follows,[9]
Debashish Bhattacharya [5]
Benjamin A. S. Van Mooy [4] and Kay D. Bidle [1*]

[1]*Department of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ 08901, USA.*

[2]*Scottish Association for Marine Science, Scottish Marine Institute, Oban, Argyll, PA37 1QA, Scotland, UK.*

[3]*Department of Microbiology, University of Tennessee-Knoxville, Knoxville, TN 37996, USA.*

[4]*Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA.*

[5]*Department of Biochemistry and Microbiology, Rutgers University, New Brunswick, NJ, 08901, USA.*

[6]*Marine Biological Laboratory, Josephine Bay Paul Center, Woods Hole, Massachusetts 02543, USA.*

[7]*Department of Medicine, The University of Chicago, Chicago, IL 60637, USA.*

[8]*WA-Organic and Isotope Geochemistry Centre, School of Earth and Planetary Sciences, Curtin University, Bentley, WA 6102, Australia.*

[9]*Department of Earth, Atmosphere and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.*

Summary

Coccolithoviruses **(EhVs) are large, double-stranded DNA-containing viruses that infect the single-celled, marine coccolithophore *Emiliania huxleyi*. Given the cosmopolitan nature and global importance of *E. huxleyi* as a bloom-forming, calcifying, photoautotroph, *E. huxleyi*–EhV interactions play a key role in oceanic carbon biogeochemistry. Virally-encoded glycosphingolipids (vGSLs) are virulence factors that are produced by the activity of virus-encoded serine palmitoyltransferase (SPT). Here, we characterize the dynamics, diversity and catalytic production of vGSLs in an array of EhV strains in relation to their SPT sequence composition and explore the hypothesis that they are a determinant of infectivity and host demise. vGSL production and diversity was positively correlated with increased virulence, virus replication rate and lytic infection dynamics in laboratory experiments, but they do not explain the success of less-virulent EhVs in natural EhV communities. The majority of EhV-derived SPT amplicon sequences associated with infected cells in the North Atlantic derived from slower infecting, less virulent EhVs. Our lab-, field- and mathematical model-based data and simulations support ecological scenarios whereby slow-infecting, less-virulent EhVs successfully compete in North Atlantic populations of *E. huxleyi*, through either the preferential removal of fast-infecting, virulent EhVs during active infection or by having access to a broader host range.**

Introduction

Emiliania huxleyi–Coccolithovirus (EhV) interactions are dependent upon a near-complete, *de novo* sphingolipid biosynthesis pathway encoded in EhV genomes (Wilson *et al.*, 2005; Monier *et al.*, 2009), which leads to a retrenchment of the host lipid biosynthetic machinery (Monier *et al.*, 2009; Vardi *et al.*, 2009; Michaelson *et al.*, 2010; Rosenwasser *et al.*, 2014; Ziv *et al.*, 2016) and the production of virus-derived glycosphingolipids (vGSLs). These lipids are critical to and diagnostic of successful infection (Vardi *et al.*, 2009; Ziv *et al.*, 2016) and are associated with the production of reactive oxygen species, autophagy, programmed cell death (PCD) and lysis of *E. huxleyi* cells (Evans *et al.*, 2006; Bidle *et al.*, 2007; Schatz *et al.*, 2014; Sheyn *et al.*, 2016). vGSLs are incorporated into the EhV virion membrane envelope, where they comprise >80% of the total lipid content (Vardi *et al.*, 2009; Fulton *et al.*, 2014). Exogenous addition of vGSLs mimics infection by inducing PCD and *E. huxleyi* cell lysis in a dose-dependent manner (Vardi *et al.*, 2009). However, the impact of these lipids on the competitive ecology of EhVs is largely unknown. EhV infections of natural *E. huxleyi*

populations in the North Atlantic (Sorensen *et al.*, 2009; Martínez Martínez *et al.*, 2012) and in induced-bloom meso-cosms in Norwegian fjords (Martínez Martínez *et al.*, 2007; Pagarete *et al.*, 2014) suggest that some EhVs preferentially propagate. Likewise, the dominance of EhV207 during competitive coinfection of *E. huxleyi* with EhV86 was thought to be due to a more efficient replication, release and lysis of its *E. huxleyi* host (Nissimov *et al.*, 2016). However, the functional importance of vGSLs as virulence factors determining EhV infection remains unknown.

Serine palmitoyltransferase (SPT) is the first and rate-limiting enzyme in the biochemical pathway for the production of ceramide (Han *et al.*, 2006; Monier *et al.*, 2009; Bidle and Vardi 2011), a precursor for the synthesis of glycosphingolipids (GSLs). GSLs are important for cell function (Hakomori 2008) and SPT genes are found in both *E. huxleyi* and EhV genomes (Wilson *et al.*, 2005; Monier *et al.*, 2009). The EhV-encoded SPT is expressed as early as 2 h post-infection (hpi) (Allen *et al.*, 2006) and has a unique substrate preference for myristoyl-CoA and pentadecylic-acid-CoA over palmitoyl-CoA (Han *et al.*, 2006; Ziv *et al.*, 2016), which allows for a diagnostic distinction between EhV and host-derived GSL production. Due to the importance of these vGSLs as critical virulence factors (Vardi *et al.*, 2009; Bidle and Vardi 2011; Fulton *et al.*, 2014; Rose *et al.*, 2014; Hunter *et al.*, 2015), we characterized the levels, rate and structural diversity of vGSL production for several different EhV genotypes during infection, as key determinants of infectivity and the rate of host demise. We link the aforementioned features with measurements of biochemical catalytic differences of SPT during infection and explore the functional diversity of this protein in natural populations of the North Atlantic as a proxy for the EhV genotypes.

Results and discussion

EhV-derived SPT

The EhV-encoded SPT protein consists of a large (LCB2) and a small (LCB1) subunit, with a connecting linker region (Han *et al.*, 2006; Michaelson *et al.*, 2010; Nissimov *et al.*, 2013) (Fig. 1A). Neighbor joining phylogenetic analysis of amino acid alignments places the full-length protein, the conserved LCB2 and LCB1 subunits, and the more variable linker region from cultured EhV strains in distinct sub-clades (Fig. 1B). All sequenced EhV genomes to date encode genes for sphingolipid biosynthesis with 11 of the 13 containing the linker sequence (Nissimov *et al.*, 2017). Oligonucleotide primers targeting this region successfully amplify natural EhV populations in the North Atlantic and South Atlantic oceans (Nissimov *et al.*, 2013) and do not amplify host-derived SPTs (Fig. S1). The more variable sequence composition of the

linker region may serve to impact SPT catalytic activity by influencing tertiary folding (Nissimov *et al.*, 2013). Taken together, this suggests that the linker is a common and widespread feature of EhV-encoded SPT proteins and a useful genetic biomarker for studies of functional diversity and competitive ecology. There are instances, however (e.g. EhV18 and EhV145), in which the two LCB subunits are encoded by different loci on the genome (Nissimov *et al.*, 2014, 2017) and, hence, do not contain a linker. These cases are less common and, in our study, they represent a comparative 'control' for the impact of subunit structural organization on associated SPT activity and vGSL production dynamics in the absence of a linker.

Infection-dynamics, GSL production, and GSL species diversity

We characterized the dynamics of infection and vGSL production, along with the relative contribution of myristoyl-CoA and palmitoyl-CoA-based activity for a highly sensitive *E. huxleyi* host strain CCMP374 (Fulton *et al.*, 2014) challenged with representatives from each EhV sub-clade. EhV207 was consistently the most virulent, fast-infecting virus strain tested, with a lytic cycle of 8–24 hpi, whereas EhV18 and EhV99B1 displayed similar, less-aggressive, slow-infecting infection dynamics with a lytic cycle of ~48–72 hpi (Fig. 1C and D). EhV207-infected cultures produced significantly more virus progeny than EhV18- and EhV99B1-infected cultures between 24 and 72 hpi. Free virus abundances ultimately converged and were similar across strains at the end of the experiment at 120 hpi. Infection dynamics for EhV86 from a prior, independent study (Fulton *et al.*, 2014; Fig. S2) was also slower than EhV207, with host lysis and virus production occurring at ~24–48 hpi. The more rapid, virulent infection dynamics of EhV207 supports previous findings in which it outcompeted EhV86 in simultaneous laboratory infection experiments (Nissimov *et al.*, 2016) and are consistent with early (2–4 hpi) upregulation of SPT gene transcripts in *E. huxleyi* cells infected with EhV201 (Ziv *et al.*, 2016), the closest phylogenetic relative to EhV207 (Fig. 1B) (Nissimov *et al.*, 2017).

The temporal dynamics (Fig. 2A) and structural composition patterns of vGSL production (Fig. 2B) differed significantly among EhV strains. The earlier and significantly higher level of vGSL production for EhV207-infected cells 8–72 hpi (Fig. 2A) was accompanied by a more diverse pool of vGSL species, based both on their mass-to-charge ratio (*m/z*) and on component long chain bases (Fig. 2B). Fifteen different vGSL species were detected in EhV207-infected cells, in contrast to 14 in EhV86-infected cells and 11 in the EhV18- and EhV99B1-infected cells, with most EhV207-derived vGSL species being synthesized relatively early in the infection cycle (<24 hpi). Several vGSL
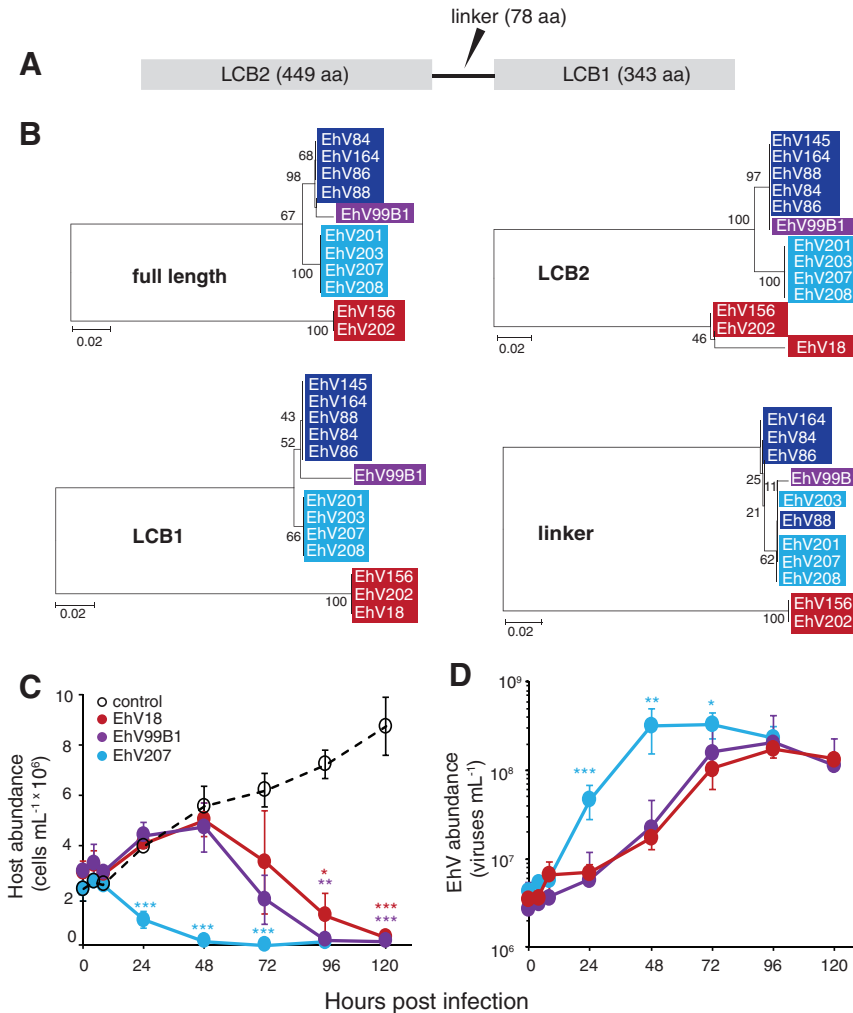
Fig. 1. EhV-encoded SPT and its different components cluster into distinct groups and are associated with strain specific infection dynamics.

A. Schematic representation of the three components that make the virus SPT protein.

B. Neighbor Joining phylogenetic analyses of the full-length EhV-encoded SPT protein (870 aa), the conserved, large (LCB2, 449 aa) and small (LCB1, 343 aa) subunits, cluster EhV genotypes into distinct groups (highlighted by dark-blue, purple, light-blue, and red shading). Note that EhV18 and EhV145 are missing from the full-length and linker region trees, since the two SPT subunits in these strains are encoded by separate genes, each residing at different locations in their respective genomes (there is no linker connecting the two subunits in these strains). Trees were generated with a total of 1000 permutations using the bootstrap method in MEGA6.
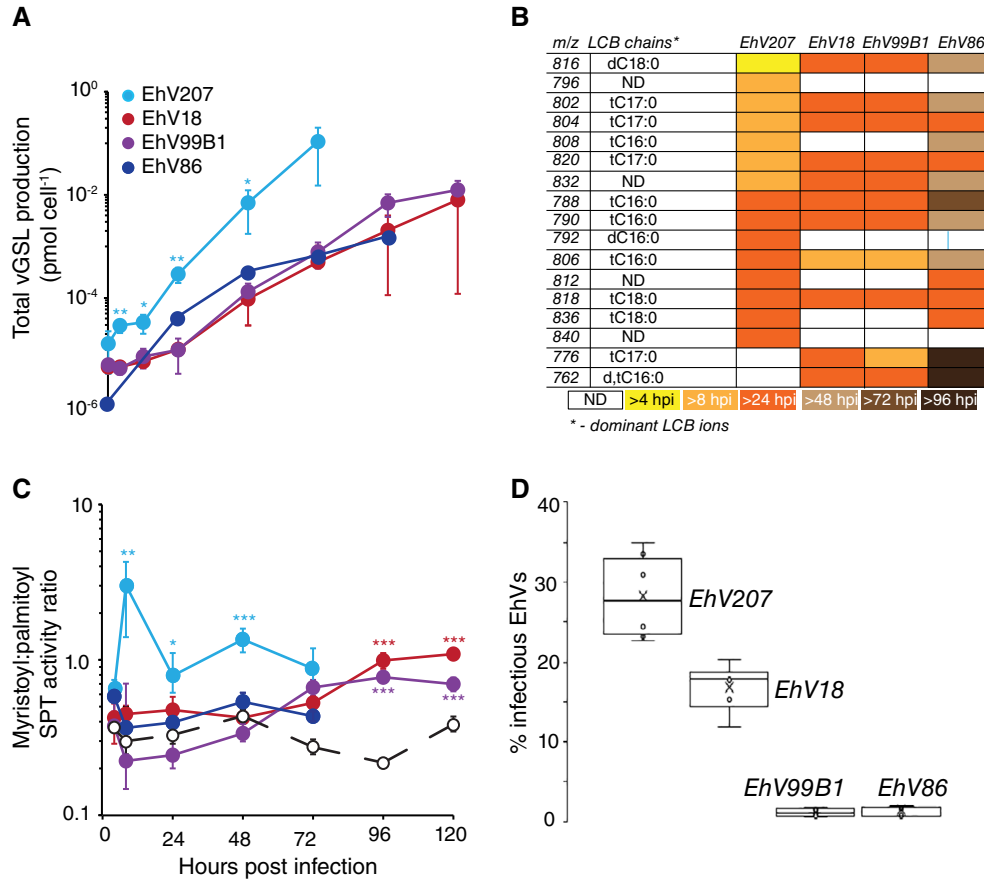
C and **D.** EhV strains (EhV207, EhV99B1, EhV18) representing the aforementioned SPT clusters exerted different infection dynamics with a susceptible host strain *E. huxleyi* CCMP374 (for EhV86 see Fig. S2). Average *E. huxleyi* (C) and EhV (D) abundances ($\pm$SD; $n = 3$) were measured via flow cytometry (FC) for up to 120 hpi. EhV207-, EhV99B1- and EhV18-infected *E. huxleyi* cultures are shown in solid light-blue, purple and red lines respectively. Control cultures ($n = 3$) contained heat denatured EhV particles and are shown in dashed lines and open circles. EhV207 infections had significantly lower host abundance ($p < 0.05$, $n = 3$, One-way ANOVA between infected and uninfected host pairs) and significantly higher EhV production ($p < 0.05$, $n = 3$; one-way ANOVA between infected and uninfected host pairs) between 24 and 72 hours post infection (hpi) compared with all other EhV strains and the uninfected controls. EhV18 and EhV99B1 had significantly lower host populations from uninfected controls at 96 and 120 hpi (***$p < 0.001$; **$p < 0.005$; *$p < 0.05$; $n = 3$ for each treatment; one-way ANOVA between infected and uninfected host pairs). EhV particles were not detected in uninfected controls.

species (*m/z* 792, 796 and 840) were only produced in EhV207-infected cells. Approximately one-half of the total vGSL species in the EhV207-infected cells were detected within 8 hpi, whereas only a fifth of the total vGSL species were detected for the other EhVs during the same time period. MS2 mass spectrometry analysis revealed a mixture of C16, C17 and C18 long-chain base structures, derived from pentadecylic acid and myristic acid (Fig. 2B).

Importantly, C17 long-chain bases are enriched in infected *E. huxleyi* cells and are essential for the assembly of virus particles (Ziv *et al.*, 2016).

SPT activity

EhV207-infected cells had significantly higher myristoyl: palmitoyl (myr:palm) SPT activity ratios (Fig. 2C) and

A.

B.

m/z	LCB chains*	EhV207	EhV18	EhV99B1	EhV86
816	dC18:0				
796	ND				
802	tC17:0				
804	tC17:0				
808	tC16:0				
820	tC17:0				
832	ND				
788	tC16:0				
790	tC16:0				
792	dC16:0				
806	tC16:0				
812	ND				
818	tC18:0				
836	tC18:0				
840	ND				
776	tC17:0				
762	d,tC16:0				

Legend: ND | >4 hpi | >8 hpi | >24 hpi | >48 hpi | >72 hpi | >96 hpi

* - dominant LCB ions

C.

D.

Fig. 2. EhVs possess fundamental differences in GSL production, SPT activity and infectivity.
A. Dynamics of vGSL production during infection of *E. huxleyi* CCMP374 with EhV207 (light-blue), EhV18 (dark-red), EhV99B1 (purple) or EhV86 (dark-blue). Data for EhV86 derive from a previous study (Fulton *et al.*, 2014) (Fig. S2) and are superimposed with the other EhV strains analyzed in our study. EhV207 infections had significantly higher total vGSL production from 4 to 48 hpi compared with the other EhV strains and controls (***$p < 0.001$; **$p < 0.005$; *$p < 0.05$; $n = 3$ for each treatment; one-way ANOVA between infected host pairs).
B. A heat-map of the specific vGSL species scaled to the time at which they were first detected in each infection treatment (ND = not determined; dC = dihydroxy; tC = trihydroxy). MS2 analysis identified the fatty acid chain base length of most vGSL species except *(m/z)* 796, 832, 812 and 840.
C. EhV207 infections had significantly higher myristoyl- to palmitoyl-based SPT activity ratios within the first 8–48 hpi compared with infections with the other EhV strains and controls ($p < 0.05$, $n = 3$ for each treatment; one-way ANOVA between infected and uninfected host pairs). EhV18 and EhV99B1 were significantly different to each other and the uninfected controls (dotted black line) at 96 and 120 hpi (***$p < 0.001$; **$p < 0.005$; *$p < 0.05$, $n = 3$ for each treatment; one-way ANOVA between infected and uninfected host pairs). Note that SPT activity measurements started at 4 hpi, due to sampling constraints.
D. EhV strains differed in the percentage of infectious virus particles at the end of the experiments (ANOVA for pairwise strain comparisons: EhV99B1-EhV18: $p < 0.001$; EhV99B1-EhV207: $p < 0.001$; EhV18-EhV207: $p < 0.05$; EhV86-EhV18: $p < 0.001$; EhV86-EhV207: $p < 0.001$; EhV86-EhV99B1: $p > 0.05$; $n = 3$).

C16:C18 vGSL species ratios (Fig. S3) than those infected with the other EhVs and uninfected controls. By quantifying the respective incorporation of ¹⁴C-labelled L-serine into myristoyl- or palmitoyl-CoA, these measurements are a proxy for the catalytic activity of the EhV-encoded SPT enzyme and, hence, its ability to produce shorter chain vGSLs compared with longer chain GSLs derived from the host's SPT enzyme. It has been previously established that the EhV-encoded SPT has a preferential affinity for myristoyl-CoA over palmitoyl-CoA (Han *et al.*, 2006), the latter representing the classic SPT substrate (Merrill 2002; Hanada 2003). The myristoyl:palmitoyl SPT activity ratio for EhV207-infected cells was ~10-fold higher at 8 hpi than

that observed for EhV18-, EhV99B1- and EhV86-infected cells (Fig. 2C; ratios of 3.02 versus 0.44, 0.3, and 0.36, respectively) and remained significantly elevated to 48 hpi. This manifested in significantly higher C16:C18 vGSL relative abundance ratios for EhV207-infected cells (Fig. S3). Although the myristoyl:palmitoyl SPT activity ratio dropped to ~1.0 at 24 hpi, it remained higher than that observed for the other EhV strains, except for late phase infection (>72 hpi), at which point the myristoyl:palmitoyl SPT activity ratios for EhV18-, EhV99B1- and EhV86-infected cells approached ~1.0. These were still higher than, and significantly different from, uninfected controls, which ranged from 0.2 to 0.4 throughout the experiments (Fig. 2C),

indicating the preferential use of the myristoyl-CoA in all infected treatments.

Virulence of EhVs

Together with previous findings that GSLs comprise >80% of the EhV membrane envelope (Fulton *et al.*, 2014) and that EhVs may adsorb to and exit from host cells via GSL-enriched lipid rafts (Rose *et al.*, 2014), we sought to determine if the production and composition of vGSL species coincides with increased virulence (defined in our study as faster infection-dynamics, virus replication and infectivity rate, and quicker production and accumulation of vGSLs). Plaque assays conducted on fresh virus lysates after the aforementioned experiments revealed that nearly 30% of the EhV207 progeny were infectious, compared with only ~17%, ~1% and ~1.2% of EhV18, EhV99B1 and EhV86 progeny, respectively (Fig. 2D), despite the total number of viruses produced being similar. The inherently higher infectivity of EhV207 virions appeared to explain their more virulent, faster-infecting phenotype. Subsequent infections with EhV207 and EhV18 at the same infectious MOI (5:1) also showed a significantly more rapid host demise for the former (Fig. S4). This enhanced infectivity of EhV207 was not accompanied by higher adsorption coefficients than its slow-infecting EhV counterparts (Fig. S5). Adsorption rates averaged across the 'slow' viruses (EhV86, EhV99B1 and EhV18) were ~35% higher than 'fast' EhV207, arguing that enhanced virulence was not driven by a more efficient, interactive contact with host cells. Given the documented association of EhVs with lipid rafts (Rose *et al.*, 2014), which are themselves enriched in GSLs, and the high vGSLs composition of the EhV membrane envelope (Vardi *et al.*, 2009; Fulton *et al.*, 2014), our data are consistent with enhanced vGSL production as an important determining factor of virulence.

We argue that the differences in SPT biochemical activities exert an important control over EhV virulence and infection dynamics (Fig. S6). The EhV207-encoded SPT is associated with earlier and higher myristoyl: palmitoyl SPT activity ratios, rapid vGSL production, a greater C16:C18 vGSL abundance ratio (and vGSL species diversity), faster EhV production, faster host demise and higher virulence (Fig. 1, Fig. 2, Fig. S2, Fig. S3). We acknowledge that a variety of subcellular processes, including ROS, PCD, autophagy (Evans *et al.*, 2006; Schatz *et al.*, 2014; Bidle 2015) and unknown host defence responses, likely also play important roles in successful infection, which are not accounted for in our study. Nonetheless, the fact that vGSLs are essential for EhV infection (Ziv *et al.*, 2016) and can themselves elicit the aforementioned cell processes when added exogenously (Vardi *et al.*, 2009) suggests that biochemical differences on vGSL production at the level of SPT are critical.

Functional diversity of EhVs in the North Atlantic Ocean

We extended our analyses to natural EhV communities associated with (i.e. replicating within) natural bloom populations of *E. huxleyi* cells in the North Atlantic (Lehahn *et al.*, 2014; Laber *et al.*, 2018; Sheyn *et al.*, 2016) (Fig. 3A-B). Given the high representation of the EhV SPT linker region in 11 out of 13 EhV genomes (Nissimov *et al.*, 2017) and its specific amplification from EhVs (Fig. S1), including populations in the Atlantic Ocean (Nissimov *et al.*, 2013), along with the importance of the virus SPT in the synthesis of GSLs, we targeted this region as a functional fingerprint and gene proxy of EhV functional diversity and an indicator of the competitive success of EhV genotypes in nature. Illumina sequencing of EhV-encoded SPT linker amplicons from infected *E. huxleyi* cells collected between 5 and 56 m (Table S1) within the mixed layer and euphotic zone reflected the relative inventories of infecting EhVs at each location (see Methods; Fig. 3C-D; Table S1). Of the 8.64 million merged and translated reads, 77% (6.3 million) mapped to one of three EhV reference sequences: EhV86, EhV207 and EhV99B1 at the 98% amino acid identity threshold (Fig. 3C-D). Slow-infecting EhV86-like and EhV99B1-like genotypes, respectively, encompassed 42% (most frequent) and 1.4% (fifth most frequent) of the >3 million merged reads that comprise the top 45 most frequent SPT linker clusters (Fig. 3D), whereas the fast-infecting EhV207-like genotypes comprised 18% (third most frequent).

When these reads were binned into their respective sampling locations and were depth integrated, the EhV86-like linker sequences represented the majority of reads, followed by EhV207- and EhV99B1-like sequences (Fig. 3E, Fig. S7). EhV86 sequences were significantly more prevalent (higher percent of reads) than EhV207 and EhV99B1 sequences in upper, euphotic zone waters ($p = 5 \times 10^{-5}$ and 0, respectively); EhV207 also significantly exceeded EhV99B1 ($p = 8 \times 10^{-5}$) (Fig. S7B). EhV86 composed >50% of reads in almost all casts (15 out of 17 CTD casts; 88%), a dominance that was also confirmed by Kruskal–Wallis analyses [$\chi^2 = 46.82$, degrees of freedom = 2, $p = 6.8 \times 10^{-11}$; see Experimental Procedures and Supporting Information].

These results demonstrate that more efficient SPT activity and vGSL production, as observed in our lab-based experiments for highly virulent EhV207, do not predict numerical dominance in natural populations. Given their overrepresentation in an enormously diverse SPT linker pool (696,191 nonredundant sequences clustered at 100% identity and 513,993 singletons), slow- and fast-infecting EhV genotypes with specific SPT linker compositions clearly coexist. We confirmed that the differential virus composition across stations and depths was not due to primer bias in polymerase chain reaction (PCR) amplification (Fig. S7C). Two-way ANOVA of
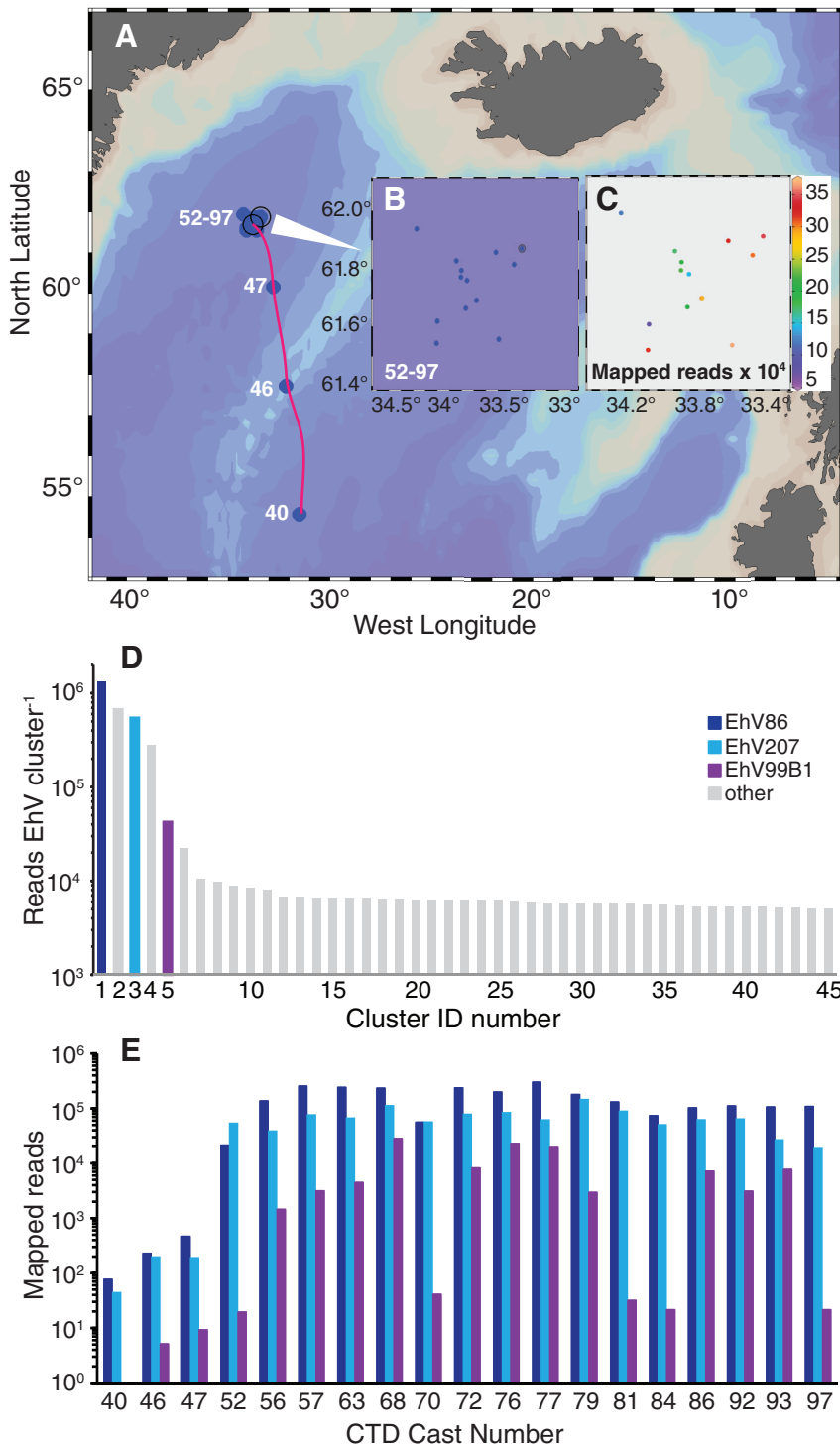
Fig. 3. Relative representation of EhV-encoded SPT linker sequences for cell-associated, replicating viruses during an *E. huxleyi* bloom in the North Atlantic.

A. Partial cruise track (blue dots; red line) of the 2012 *NA-VICE* expedition between the Azores and Iceland (Laber *et al.*, 2018; Sheyn *et al.*, 2018), showing CTD casts (blue dots) that were taken at locations with actively infected *E. huxleyi* populations (Laber *et al.*, 2018).

B. CTD casts 52–97 (and their exact locations) corresponded to a large *E. huxleyi* bloom (~4000 cells mL⁻¹) within an anticyclonic eddy, which was repeatedly sampled in Lagrangian mode for ~9 days (Laber *et al.*, 2018).

C. The distribution and total number of EhV-derived SPT linker amplicon sequences associated with *E. huxleyi* cells (collected onto 0.8-μm pore-size filters; see Methods) at CTD casts 52–97 and mapped onto EhV reference strains (see Supporting Information for details).

D. Frequency distribution of the 45 most abundant SPT linker clusters integrated across all *NA-VICE* CTD casts, for the 8.65 million merged reads. Slow-infecting EhV86 and EhV99B1-like genotypes, respectively, represented 42% (most frequent) and 1.4% (fifth most frequent) of the data (combined total 43.4%), whereas fast-infecting EhV207-like genotypes comprised 18% (third most frequent). Other EhV cluster representatives (grey), including two of the top five clusters, have unknown infectivity dynamics or no close representatives in culture.

E. Amplicon sequence reads of the SPT linker binned into their respective sampling locations and depth integrated down to ~50 m (see Table S1). The EhV86-like linker protein sequences represented the majority of reads, followed by EhV207- and EhV99B1-like sequences. Given their over-representation in an enormously diverse SPT linker pool (696,191 unique sequences at 100% identity clustering and 513,993 singletons), these data suggest that slow- and fast-infecting strains with specific SPT linker compositions clearly coexist in a natural, ecologically relevant setting.

amplification efficiencies of the primer set on the SPT genes of EhV207, EhV86 and EhV99B1 revealed no significant primer bias ($p < 0.05$, $r^2 = 0.92$, d.f. = 21). Hence, the observed distributions of EhV genotypes as inferred by SPT linker sequences reflected their *in situ* distributions.

Ecological scenarios of EhV infection

We explored three nutrient-phytoplankton-virus ecosystem models representing contrasting ecological scenarios that might promote the coexistence of slow- and fast- infecting EhVs in the North Atlantic (Fig. 4, Fig. S8, Fig. S9). These modelled scenarios are driven by empirical laboratory- and

field-based observations (this study; Nissimov *et al.*, 2016; Laber *et al.*, 2018), with their output aimed at testing the hypothesis that competitive coexistence between slow- and fast- infecting EhVs is stable and is ecologically and biogeochemically relevant. We first allowed for an '*all-else-even*' competition, in which modelled fast-infecting, highly virulent (EhV207-like) genotypes compete with slow-infecting, less virulent (EhV86-, EhV99B1-, and EhV18-like) ones in steady-state conditions. Simulations in this scenario (Fig. 4) stochastically allocate key traits (i.e. virus replication rate and burst size) using plausible ranges informed by model fits to laboratory experiments (Table S2, Fig. S8, Fig. S9). Fast-infecting EhVs consistently dominate under stable growth conditions in this model due to faster replication rates and higher burst sizes (Fig. 4A and D; Fig. S8), but it does not support our data from the North Atlantic.

We also explored a scenario for '*differential removal*' of virulent EhVs from the euphotic zone (Fig. 4B and E), due to preferential entrainment of viruses into aggregated, sinking particles. The rationale for this approach was several fold. First, EhV infection induces the production of sticky transparent exopolymeric particles (TEP), enhancing aggregate formation of infected cells and facilitating downward flux in natural populations in the North Atlantic (Vardi *et al.*, 2012; Laber *et al.*, 2018; Nissimov *et al.*, 2018); this phenomenon is accentuated in fast-infecting EhV207 and occurs early in infection, absent of lysis (Nissimov *et al.*, 2018). Second, we observed an enrichment of EhV207-like SPT linker amplicons in sinking biomass collected at 150 m at 5 of seven CTD cast stations (Fig. S10A-B), which differed from the relative representation of SPT linker amplicons observed in overlying mixed layer and euphotic zone (Fig. S7). Statistical analysis confirmed the shift from dominance by EhV86 in shallower, euphotic zone water towards EhV207, deeper in the water column across the stations sampled in the North Atlantic. A higher prevalence of EhV207 was observed at depth (150 m) than surface waters; no significant changes in the proportion of other lineages were observed (Fig. S10C; see Experimental Procedures and Supporting Information). Further, the EhV mapped reads in CTD casts 52–97 (Fig. 3E) were associated with high particulate organic carbon and particulate inorganic carbon sinking fluxes (Collins *et al.*, 2015), which may have been facilitated by virus-induced, TEP production and aggregate formation (Vardi *et al.*, 2012; Laber *et al.*, 2018). The model illustrates how higher population densities of slow, less virulent EhVs can be maintained and their relative success enhanced, when fast-infecting viruses are mechanistically removed at a higher rate through entrainment into particles (Fig. 4B and E). Taken together with field data showing entrainment of fast viruses at depth (Fig. S10), the '*differential removal*' model supports enhanced removal of virulent

strains from the upper ocean and implicates them in the enhanced production of TEP, particle aggregation and sinking flux (Laber *et al.*, 2018; Nissimov *et al.*, 2018).

Finally, we explored a '*differential infectivity*' model to address whether the success of less-virulent EhV genotypes could also be explained by their inherent ability to infect a wider range of host population genotypes (Fig. 4C and F). The approach here was empirically informed by Illumina sequencing of *E. huxleyi* cytochrome oxidase I (COI) in the North Atlantic (Fig. S11), which revealed the presence of nine host genotypes with broad representation to *E. huxleyi* strains in culture. While one genotype (Eh1) accounted for >95% of the COI copies, it is possible that resistant strains were present within this microdiversity, which could impact relative EhV infectivity. Indeed, several of the North Atlantic genotypes, including Eh1, cluster with COI sequences from *E. huxleyi* strains with known resistance to EhVs (Fig. S12). Also, experimental data on the host range of EhV99B1 and EhV207 showed that, out of an array of 16 different *E. huxleyi* strains, they successfully infected 12 and 8, respectively (Table S3). EhV86 was also recently shown to have a broader host range than EhV207 (33 strains vs 27 strains, respectively for a slow:fast infectivity ratio of 1.2; Ruiz *et al.*, 2017), further supporting this idea. Our host-range experiments yielded a slow:fast infectivity ratio of ~1.5 (75%:50%), which was subsequently used as an empirically derived parameter to inform the model. In the cases where both EhV99B1 and EhV207 were capable of infecting the same host strain, the latter had a more rapid lytic dynamic, demonstrating that this fast-infecting phenotype was generally conserved across host strains (Table S3). The inclusion of host populations with different susceptibilities to EhVs in our '*differential infectivity*' simulations yielded coexistence across a range of slow: fast infectivity ratios (Fig. 4F) consistent with our data from the North Atlantic.

Our lab and model-based data/simulations argue for a competitive advantage of high vGSL producing, virulent, fast-infecting EhVs at high host densities (~10^5–10^6 cells mL^{-1}) associated with lab-based culture experiments (Fig. 1, Fig. 2). Conversely, our field- and model-based data/simulations support the competitive survival of slow EhVs in natural populations, where host densities are significantly lower (10^1–10^3 cells mL^{-1}) (Laber *et al.*, 2018) and fast-infecting, 'aggressive' strains appear to be preferentially removed or are unable to infect a subpopulation of susceptible host genotypes (Fig. 4, Fig. 5). We argue that it is likely a combination of these two factors that contribute to the competitive success of slow-infecting EhVs in natural populations. While we acknowledge that multifaceted processes likely drive virus ecology, our empirical lab- and field-based observations here and elsewhere
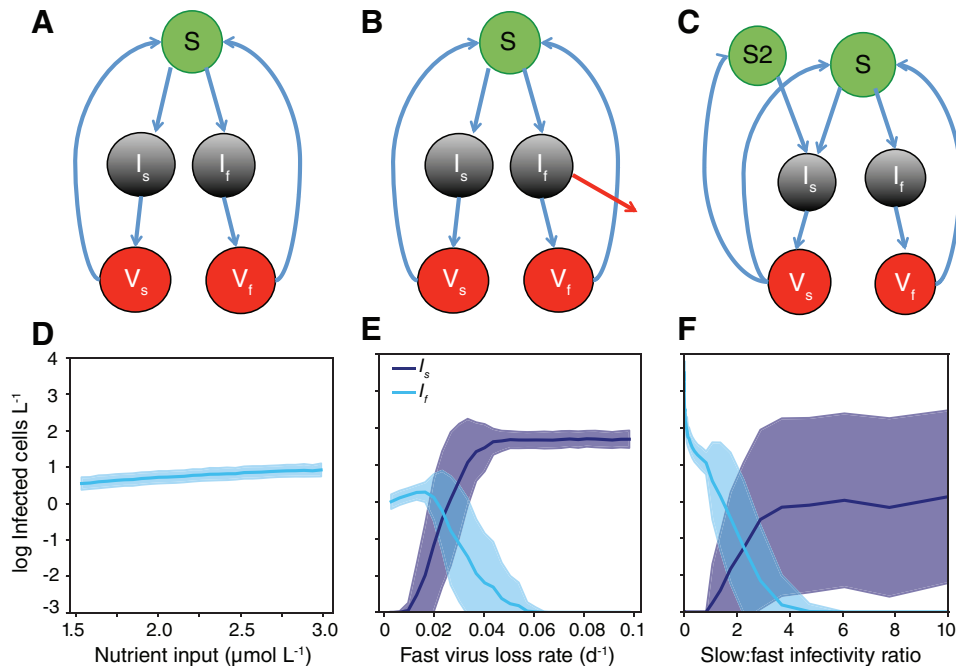
Fig. 4. Mathematical models depicting scenarios where EhV strains with different virulence can coexist. Contrasting hypothetical interactions (**A–C**) and population dynamics (**D–F**) of modelled 'slow'-infecting (V_s; EhV86-, EhV99B1- and EhV18-like) and 'fast'-infecting (V_f; EhV207-like) EhVs. Each model has at least six state variables, but for comparison with observations (see Fig. 3) and for clarity, we only show infected host populations. Nutrients were supplied as if in a chemostat and the models were integrated to steady state with zero rate of change for all state variables. Simulations were repeated 10,000 times, drawing key parameters informed by model fitting to laboratory experimental data (see Fig. 1C and D; Fig. S8). Mean values (solid lines) and one standard deviation either side of the mean (shaded regions) are plotted. I_s and I_f represent host cells infected with 'slow' and 'fast' EhVs respectively. Scenario #1 (A and D) assumes equal losses of the two virus types, which results in the more aggressive EhV genotypes (V_f) being dominant when one host subpopulation is present (S). Scenario #2 (B and E) allows for enhanced removal of V_f-infected host cells (I_f) (red arrow), due to potential entrainment in sinking particles (when only S is present). Modelled EhV genotypes coexist when the virus loss rate of fast EhV genotypes (0.01–0.03 d^{-1}) offsets their replication advantage. Scenario #3 (C and F) allows for an additional host subpopulation (S2) for which slow-infecting EhV genotypes (e.g. V_s) have a larger host range (as per empirical observations in our study; Table S3). The proportion of the host subpopulation that is susceptible to slow and fast viruses were drawn randomly from uniform distributions within the range (0–1), and results are plotted as a function of the slow:fast infectivity ratio. Slowly infecting EhVs can coexist with fast-infecting ones when the slow:fast infectivity ratio is 2–3 (i.e. slow EhVs can infect at least ~2–3 times the subpopulation than the fast EhVs).

(Vardi *et al.*, 2009, 2012; Fulton *et al.*, 2014; Hunter *et al.*, 2015; Laber *et al.*, 2018; Nissimov *et al.*, 2018; Sheyn *et al.*, 2018) are consistent with SPT functional diversity and GSL production being a driver that helps to explain the competitive ecology of EhV strains infecting *E. huxleyi* populations in the North Atlantic. There may indeed be other mechanistic drivers that help to shape EhV competitive ecology in the North Atlantic, but they have yet to be demonstrated.

Quantitatively, our mathematical models suggest that a fast virus removal rate between 0.01 and 0.03 d^{-1} combined with a slow:fast infectivity ratio of >2–3 represent relevant ecological boundary conditions that can explain our field observations and provide a framework for future work. Given the aforementioned model simulations assume steady-state conditions and do not account for ephemeral dynamics that characterize North Atlantic blooms, we embedded the traits and trade-offs of EhV coexistence in simulated North Atlantic coccolithophore blooms within the Darwin ecosystem model (Fig. 5). The '*differential removal*' and the '*differential infectivity*' scenarios were both

consistent with the competitive coexistence of slow EhVs observed in the North Atlantic (Fig. 3); the former scenario resulted in a clear subsurface maximum of fast EhVs consistent with SPT sequencing observations of enrichments at depth (Fig. S10).

The considerable EhV functional diversity, as evidenced by >500K unique SPT linker sequences (singletons) and 45 SPT genotypes with >5K copies (Fig. 3D), argues for numerous EhV 'winners' in natural populations. Such an immense gene pool is likely to provide EhVs with a sufficient adaptive potential against evolving host defences, with 'winners' and 'losers' rising and falling in response. More broadly, our data underline an immense genetic variation of EhVs in nature and highlight an extremely vigorous evolutionary tempo of the host-virus 'arms-race' at the level of SPT and GSL production. It is intriguing to speculate that competitive survival through SPT diversity (and associated functional outcomes outlined in this work) may serve to drive *E. huxleyi*-EhV temporal and spatial dynamics in natural populations, whereby vastly different relative frequencies in a diverse EhV gene pool are driven by the
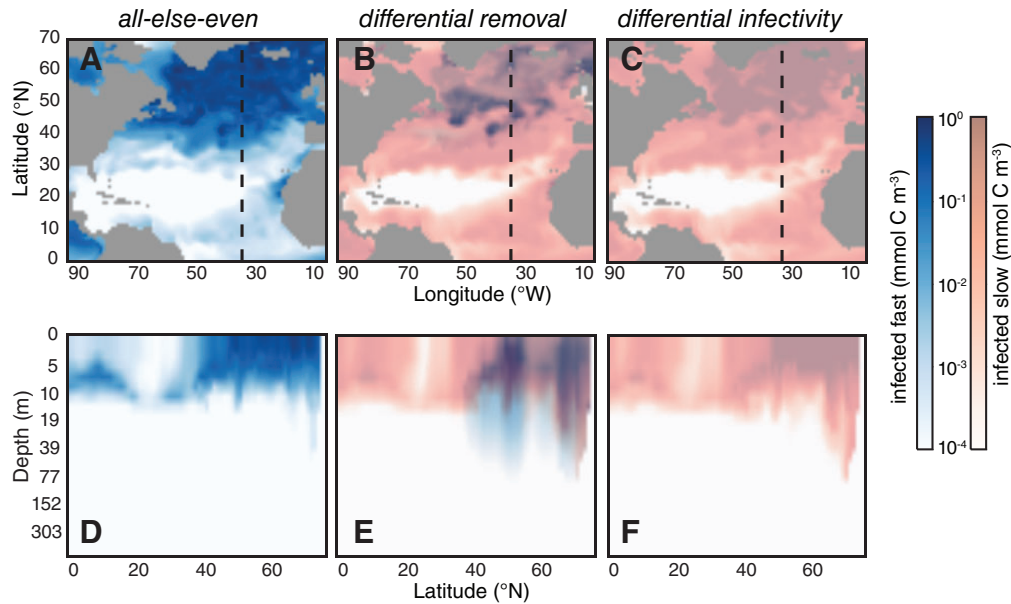
Fig. 5. Darwin ecosystem model simulates the competitive ecology of slow EhV viruses in the North Atlantic. Shown are the simulated biomass of *E. huxleyi* infected with fast (blue) and slow (red) EhVs during the month of July in the North Atlantic for the three modelled ecological scenarios (described in Fig. 4) embedded in the Darwin ecosystem model.

A–C. July surface averages for simulations of '*all-else-even*' competition, '*differential removal*', and '*differential infectivity*' are shown (see Methods, Supporting Information and Fig. 4, Fig. S8, and Fig. S9 for details).

D–F. Depth profile of the fast- and slow-infected populations along the 30° W longitude transect (marked by the vertical dashed lines in panels A–C), which roughly corresponds to the *NA-VICE* transect (Fig. 3A). Each panel in D–F corresponds to the same simulation and ecological scenario directly above (A–C). The '*differential removal*' (B and E) and the '*differential infectivity*' scenarios (C and F) are both consistent with the numerical dominance of slow EhVs observed in the North Atlantic (Fig. 3). '*Differential removal*' of fast EhVs results in a subsurface maximum in panel E that is consistent with SPT sequencing observations of enrichments at depth (Fig. S10).

aforementioned ecological scenarios and environmental conditions.

Experimental procedures

Laboratory experimental set-up and sampling

A 24L culture of *E. huxleyi* CCMP374 was grown in f/2-Si at 18 °C, a light:dark cycle of 14:10 h and a light intensity of ~180 μmol photons m^{-2} s^{-1}. An exponentially growing culture was split into replicates (×12), sterile glass culture flasks (2 L), which served as the experimental vessels in this study. Three were inoculated with the EhV strain EhV207 (English Channel origin, isolated in 2001; Nissimov *et al.*, 2012), three with EhV99B1 (Norwegian fjord origin, isolated in 1999; Pagarete *et al.*, 2013) and three with EhV18 (English Channel origin, isolated in 2008; Nissimov *et al.*, 2014), at virus: host ratios of 5:1 (as determined by SYBR green-staining and flow cytometry (FC); see below). Additional control cultures contained a heat-denatured and UV-irradiated mixture of the three EhV strains (all added in equal proportions). Samples were collected for FC (1 mL fixed in 0.5% glut), biomass (upon 1.2 μm pore-size membrane filters) and lipid analysis (0.2 μm pore-size Durapore filters) at 0–120 hours post-virus

infection (hpi). At the end of the experiments, subsamples of each EhV lysate (0.45 μm pore-size filtered) were retained to measure virus infectivity via plaque assays (details below).

Cell and virus abundance measurements

Subsamples from the aforementioned cultures were analyzed for host and virus abundances by FC according to Brussaard *et al.* (Brussaard *et al.*, 2000; Brussaard 2004) using a BD Accuri C6 flow cytometer equipped with a 488-nm laser and a standard filter setup (i.e. 533/30 in FL1, 585/40 in FL2, 670 LP in FL3 and 675/25 in FL4). Fluorescent beads (3–10 μm) were used for calibration prior to sample analysis. Counts of *E. huxleyi* cells and EhVs were conducted at a flow rate of 66 and 13 μL min^{-1}, respectively. *E. huxleyi* cells and EhVs were identified on the basis of their side scatter (SSC) versus chlorophyll fluorescence (692 nm) and SSC versus SYBR green fluorescence (520 nm) respectively.

In vitro *SPT activity assays*

One-hundred fifty millilitres from the above cultures were collected at 4, 8, 24, 48, 72, 96 and 120 hpi onto

47mm diameter, 1.2μm pore-size Isopore Millipore membrane filters for downstream protein extractions and *in vitro* SPT activity assays. Collected biomass was scrapped off from filters and resuspended in 1 mL lysis buffer (a stock mixture of 1-mL HEPES-NaOH [0.5 M, pH 8], 20 μL of EDTA [0.5 M, pH 8] and 100 μL of SML [10% (w/v)], brought to a final volume of 10 mL with dH_2O) to create a cell concentrate. Concentrates were sonicated on ice in 15 s intervals for a total of 2 min (15-s sonicaton followed by 15s rest on ice) and centrifuged ($2500 \times g$; 2 min; 4°C). Supernatants were transferred to sterile tubes and protein concentrations were determined using the Thermo Scientific PierceBCA Protein Assay Kit (Fischer *et al.*, 1999; Adilakshmi and Laine 2002; Prozialeck *et al.*, 2002; Roberts 2002) and Molecular Devices Spectra Max M3 microplate reader. Prior to the SPT activity assays, two 20X [^{14}C] assay mixes were made based on a modified protocol from Rutii *et al.* (Rütti *et al.*, 2009). One contained a mixture of L-serine (200 mM), pyridoxal 5-phosphate (5 mM), palmitoyl-CoA (5 mM), L-[U-^{14}C] serine (50 mCi mL^{-1}) and dH_2O, and the second contained myristoyl-CoA (5 mM) instead of palmitoyl-CoA. The SPT activity assay mixtures were prepared on ice to which 190 μL of total cell lysate (~2 mg mL^{-1}, diluted with lysis buffer if needed) and the corresponding 20X [^{14}C] assay mix were added. A myriocin SPT-inhibitor solution (1 mg mL^{-1}) was added to negative controls. SPT reactions were initiated by transferring tubes to a 37°C water bath in which they incubated for 120 min. Mixtures were cleaned by the addition of methanol/KOH: $CHCl_3$ (4:1), $CHCl_3$ (chloroform), alkaline water and 2 N NH_4OH (ammonium hydroxide), centrifuged ($12,000 \times g$; 1 min; RT) and the upper aqueous phase removed by gentle aspiration. The lower organic phase was washed twice with alkaline water after which it was transferred into polyethylene scintillation vials. The $CHCl_3$ was let to evaporate under a stream of nitrogen for up to 10 min. The amount of incorporated ^{14}C-labelled serine was determined by liquid scintillation counting for 10 min (Fisher Scientific ScintiVerse BD cocktail) using the ^{14}C detection window (Beckman LS 6000IC).

Cell-associated GSL extraction and analyses

Twenty millilitres of total biomass samples were collected at the aforementioned sampling times from the above cultures onto 25 mm diameter, 0.2 μm pore-size Durapore membrane filters under gentle vacuum and used for cell-associated lipids. Polar lipids were extracted from Durapore membrane filters as described elsewhere (Popendorf *et al.*, 2013). Lipid extracts from Durapore membrane filters were analyzed using normal-phase liquid chromatography followed by electrospray ion-trap mass spectrometry (Van Mooy and Fredricks 2010). Individual GSLs were quantified against a calibration curve of an authentic standard (Matreya LLC, glucocerebroside, #1522) and normalized to an internal standard (1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine-N-[2,4-dinitrophenyl], Avanti Polar Lipids #810508). vGSLs are defined as GSLs appearing during infection as compared with GSLs that were present in uninfected cultures (so-called host-GSLs or hGSLs), and which also showed a dominant (positive ion) neutral loss of 162 (glyco-headgroup) as compared with 180 for other GSLs (glycol-headgroup plus water loss) (Vardi *et al.*, 2009, 2012). In addition, long-chain content of vGSLs was determined by examination of MS2 spectra using high resolution/accurate mass data obtained from a Thermo QExactive instrument. The same chromatography and ionization were employed while preselected vGSL molecular ions were isolated and fragmented to yield long-chain base fragments in positive ion mode. Accurate mass data allowed unambiguous determination of trihydroxy-C17:0 long-chain base [304.2846 m/z, 1.4 δppm; Ziv *et al.*, 2016 (Ziv *et al.*, 2016)], trihydroxy-C16:0 long-chain base (290.2690, 1.4 δppm) and trihydroxy-C18:0 long-chain base (318.3003, 2.0 δppm. The tetrahydroxy-C16:1 long-chain base (304.2482 m/z) described by Vardi *et al.* (2009) was not observed, which supports the assertion by Ziv *et al.* (2016) that the trihydroxy-C17:0 vGSL was previously misidentified as a tetrahydroxy-C16:1 vGSL by Vardi *et al.* (2009) who lacked high resolution/accurate mass MS2 spectra.

Plaque assays and infectivity

Plaque assays were performed on triplicate lysates for all EhV strains at the end of the experiment as previously described (Schroeder *et al.*, 2002) in order to elucidate the percentage of infectious viruses produced following infection by the different viruses. Prior to assays, the new virus lysates were passed through sterile 0.45μm pore-size syringe filters in order to separate EhVs from uninfected cells and cell debris. Plaque formation was monitored daily for one week visually and imaged via chlorophyll fluorescence using a ChemiDoc Bio-Rad imager.

Host infectivity range

The virulence of fast-infecting (e.g. EhV207) and slow-infecting (e.g. EhV99B1) EhV genotypes was tested against 15 additional host strains and reported in Table S3. Briefly, exponentially growing cultures of *E. huxleyi* strains CCMP373, CCMP374, CCMP379, CCMP1516, CCMP2090, DHB606, DHB607, DHB611, DHB621, DHB623, DHB624, DHB629, DHB631, DHB641, DHB655 and DHB659 were infected by the two viruses and monitored visually, as well as spectrophotometrically for cell abundance decrease for up to one and a half weeks after virus addition. The lytic

period for each combination of virus and host was determined as a 10% difference in the cell abundances between the controls and the infected treatments. Data for infectivity of CCMP374 are taken from Fig. 1.

Virus adsorption

Adsorption assays to *E. huxleyi* CCMP374 host cells were performed empirically based on methods previously described (Brown and Bidle 2014). Briefly, exponentially growing *E. huxleyi* cultures (n = 3) were infected with EhV207 and EhV86 at an MOI of 5:1 and were sampled up to 2 h post-virus addition for free virus abundances, as described above for virus abundance measurements. The adsorption coefficients (Cd) were determined by plotting the natural logarithm of the fraction of free (i.e. unadsorbed) viruses against elapsed time. The Cd (mL min^{-1}) was calculated as: $Cd = a/N$, where N is the cell concentration and a is the slope of the regression line between the natural logarithm of the remaining fraction of free viruses plotted over time (Cottrell and Suttle 1995). The analysis included a correction factor for potential virus adsorption to culture flasks in cell-free controls (Bratbak *et al.*, 1993; Schroeder *et al.*, 2002).

Statistical analyses

Goodness of fit was ensured using the plot (model) function in R with log transformation of data being used where appropriate. Significant differences were identified in two-way ANOVA outputs using the TukeyHSD (model) *post hoc* test using the R statistical package and are indicated with asterisks where relevant in the appropriate figures.

Field-based sampling during the NA-VICE

Biomass samples of natural *E. huxleyi* populations (and associated, infecting EhVs) were collected at 17 stations (CTD casts; Fig. 3 and Table S1) along a transect between the Azores and Iceland during the *North Atlantic Virus Infection of Coccolithophore Expedition* (*NA-VICE*; https://www.bco-dmo.org/project/2136; Laber *et al.* 2018). Water was collected from various depths from 5 to 150 m, with most depths residing in the upper mixed layer and euphotic zone (see Table S1) and filtered as described below.

Sample collection and DNA extraction from natural samples

Water was prefiltered through a 200 μm mesh to eliminate large organisms. The prefiltered water (3–5 L) was then filtered through 0.8 μm pore-sized polycarbonate filters to collect predominantly protists (including *E. huxleyi* and their cell-associated viruses). The filters were immediately submerged in 10 mL extraction buffer (100 mM Tris HCl pH 8, 250 mM EDTA pH 8, 100 mM NaCl, and 1 vol% SDS) and homogenized at maximum speed on a Vortex Genie (MO BIO, Carlsbad, CA) for 10 min in the presence of 2 mL molecular-grade zirconium beads (equal amount of 100 and 400 μm diameter beads) (OPS diagnostics, Lebanon, NJ). The homogenized filters were subject to three freeze–thaw cycles via submerging in liquid nitrogen, thawed in a water bath set to 50°C, homogenized for 5 min at maximum speed using the Vortex Genie after each freeze–thaw cycle and stored at −80°C until further DNA extraction and purification steps.

Back in the laboratory, the homogenized filters in extraction buffer were thawed, and 5mL buffer was incubated for 1 h at 50°C in the presence of 100 μg Proteinase K, followed by standard phenol/chloroform/isoamyl alcohol 25:24:1 v/v extraction. Nucleic acids were precipitated with two volumes of 100% ethanol in the presence of 0.2 M NaCl. Following centrifugation for 20 min at 10,000 RCF, salts were removed by washing the DNA pellet with 20 mL of 70% ethanol. Upon centrifugation for 10 min at 10,000 RCF and removal of traces of ethanol, the DNA pellets were dissolved in 1X TE buffer (pH 8). Traces of impurities that could inhibit enzymatic amplification reactions (PCR) were removed using the PowerClean DNA Clean-Up Kit (MO BIO).

PCR amplification of EhV-derived SPT linker and host-derived COI sequences from NA-VICE samples

PCR amplification of EhV-derived SPT linker sequence used the following primers sets: vSPT-F: 5′-GARCATA CACGIRAGCAACTT and vSPT-R: 5′-CGGTCCACAT GTACCACA-3′, deriving a 303-bp amplicon product. PCR reactions using these primers was first performed on virus lysates of EhV207, EhV99B1 and EhV86 (~10^8 viruses mL^{-1}) and on DNA (~100 ng) extracted from *E. huxleyi* strains CCMP1516 and DHB611. Positive amplification controls for *E. huxleyi* strains used previously published primers (Coolen 2011) for the COI gene (EhuxGoce-COIF: 5′-TCGGAATTGTTTCTCACA-3′ and EhuxGoce-COIR: 5′-AATCCTACCGCAAAAAGC-3′) that produce a 276-bp amplicon fragment. Each PCR reaction (25 μL) consisted of 1X Green GoTaq Flexi Buffer, 200 μM dNTPs, 3 mM MgCl$_2$, 1 U GoTaq Flexi DNA polymerase, 0.4 μM of each forward and reverse primer and 2 μL template (i.e. either virus lysate or DNA extract). Reactions with vSPT primers were conducted at initial 95°C (for 5 min), followed by 30 cycles at 95°C (for 30 s), 53°C (for 60 s) and 74°C (for 90 s), and a final step at 74°C (for 5 min). PCR reactions of the host COI were conducted at initial 94°C (for 3 min), followed by 30 cycles at 94°C (for 45 s), 55°C (for 30 s) and 72°C (for 30 s), and a final step at 72°C (for 5 min). Both vSPT and COI primer sets were used to amplify the respective gene fragments in environmental DNA derived

from biomass collected in the North Atlantic as part of the NA-VICE cruise.

Illumina MiSeq sequencing and bioinformatic analysis of environmental EhV-derived SPT amplicons

PCR amplification and library preparation was conducted in the following manner. First, semi-quantitative PCR reactions were performed in order to characterize PCR kinetics for each sample and identify a threshold cycle (C_t) for which amplicons would most accurately reflect the relative template abundance in environmental nucleic acid samples. The SPT linker primers (vSPT-F: 5' TCGTCG GCAGCGTCAGATGTGTATAAGAGACAGGARCATAC-ACGIRAGCAACTT-3'; vSPT-R:5'GTCTCGTGGGCTCG GAGATGTGTATAAGAGACAGCGGTCCACATGTACC-ACA-3') were anchored within the flanking large (LCB2) and small (LCB1) SPT subunits, spanned the linker region and included overhangs (underlined above) for sequencing and library preparation, based on manufacturer protocols (Caporaso *et al.*, 2012). These reactions used reagents from the GoTaq Promega kit and consisted of 1 μL template of total extracted nucleic acid, 1X PCR reaction buffer, 2.5 mM MgCl$_2$, 0.04 U μL^{-1} Taq DNA polymerase, 10 pmol of each primer (IDT, US), 0.2 mM dNTPs, 5% dimethylsulphoxide (DMSO), 0.5 μg μL^{-1} BSA, 2X SYBR Green mix and dH$_2$O to a final reaction volume of 25 μL, in a Mx3000P thermocycler (Stratagene). Thermocycler conditions included an initial cycle at 95°C (for 5 min), followed by 40 cycles at 95°C (for 30 s), 53°C (for 60 s) and 74°C (for 90 s), finishing with a dissociation curve. Dissociation curves and visual inspection on agarose gels confirmed desired amplicon lengths. Once the threshold cycle was determined for each sample, subsequent PCR reactions were performed (as described above) without SYBR green and DMSO for each sample at the empirically determined cycle number. Amplicon products were purified using the Agencourt AMPure XP beads DNA purification kit (Beckman Coulter) and used for downstream library preparation of *NA-VICE* virus SPT linker amplicons. Amplicons were used to prepare dual-index libraries using the Nextera XT Index Kit, according to manufacturer instructions (Caporaso *et al.*, 2012). Libraries were pooled and run on an Illumina MiSeq sequencer with a paired-end (2 × 250 bp) configuration. The run yielded 32.6 M reads that correspond to 8.4 Gbp.

Adapter and quality trimming of the 32.6 M reads were done with CLC Genomic Workbench 8.0.3 (https://www. qiagenbioinformatics.com/). Reads shorter than 100 bp after trimming and broken paired reads were discarded. The remaining 18.3 M paired reads were merged if they overlapped and had an alignment score higher than 20 (with mismatch cost = 2 and gap cost = 3), resulting in

a total of 8.64 M merged reads. Merged reads were aligned to SPT gene sequences of three different EhV genotypes: EhV99B1, EhV86 and EhV207 (accessions: FN429076, FJ531547 and FJ531555) with a 98% identity threshold using CLC Genomic Workbench 8.0.3; 77% of the reads were able to align to one of these reference sequences. To examine the variation of the amino acid sequences, each mapped read was then translated into all six reading frames with the longest read (>98 aa) not containing a stop codon being used for downstream analysis. All the translated proteins were clustered using CD-Hit (Li and Godzik 2006; Fu *et al.*, 2012) with an identity threshold of 100%. The 6.3 M translated reads clustered into 625,898 clusters, of which 445,414 (71%) were singletons (a cluster comprising of only one sequence). The largest cluster comprised of 1,313,528 reads; there were 45 clusters that had more than 5000 reads in each cluster. Raw SPT amplicon sequence reads from the *NA-VICE* have been submitted to NCBI's Sequence Read Archive (Bio-project ID: PRJNA385326; accession #: SRX2788856-SRX2788919).

Illumina MiSeq sequencing and bioinformatic analysis of environmental E. huxleyi-derived COI amplicons

PCR amplification and library preparation was conducted as follows. The environmental genomic DNA of 10–50 ng served as template for the specific amplification of a 276 bp-long region of *E. huxleyi* COI using the primer combination EhuxGoce-COIF and EhuxGoce-COIR and PCR reagents as described previously (Coolen 2011). The PCR conditions were modified after Coolen 2011 for subsequent Illumina MiSeq sequencing analysis after Caporaso *et al.* (Caporaso *et al.*, 2012). Each sample was amplified using the reverse primer that included a unique 12 base Golay barcode sequence (Caporaso *et al.*, 2012) to support pooling of the samples. The DNA concentration in the PCR products was quantified spectrofluorometrically using Quant-iT PicoGreen reagent (Life Technologies, Bedford, MA), and the quality of the PCR products were verified using an Agilent 2100 bioanalyzer and high-sensitivity DNA analysis kit. Equimolar amounts of the barcoded PCR products were pooled and purified using the AMPure XP PCR purification kit (Agencourt Bioscience Corp., Beverly, MA). Mixed and purified barcoded amplicons (library) of 400 ng was subject to Illumina MiSeq sequencing (2 × 250 bp chemistry) using the facilities of the W. M. Keck Center for Comparative and Functional Genomics, University of Illinois at Urbana–Champaign.

We used Illumina-utils (Eren *et al.*, 2013a, 2013b) (available from https://github.com/meren/illumina-utils) to demultiplex the raw Illumina sequencing data, and then to trim primers, merge partially overlapping reads and

perform quality filtering. Merged sequences were removed from the analysis if they contained more than three mismatches at the overlapping region. Mismatches at the overlapped regions of the remaining pairs were resolved by using the nucleotide with the higher *Q*-score. If neither of the nucleotides had a *Q*-score larger than Q10, we removed the pair from the analysis to minimize ambiguities downstream. We then used oligotyping (Eren *et al.*, 2013a, 2013b) to partition the resulting quality-filtered sequences into highly resolved units to identify *E. huxleyi* genotypes in our data set. Instead of relying on arbitrary sequence similarity thresholds to identify clusters of sequences, oligotyping uses Shannon entropy to utilize most highly variable nucleotide positions across amplicons to decompose a sequencing data set into 'oligotypes'. The reliance on high-entropy positions allows the accurate identification of distinct types within amplicon data even if they are only one nucleotide different from each other at the sequenced region, while minimizing the impact of noise by discarding nucleotide positions with minimum information content (Eren *et al.*, 2014). The oligotyping analysis of COI amplicons resulted in nine *E. huxleyi* genotypes. COI sequences from the *NA-VICE* were submitted to NCBI under accession # KY933659-KY933667.

Phylogenetic analysis of SPT and COI sequences

Multiple alignments of the full-length SPT proteins (870 aa) and the different components of the SPT protein (e.g. LCB2 [449 aa], LCB1 [343 aa] and linker region [78 aa]) encoded within the genomes of cultured EhVs were performed using ClustalW (pairwise alignment [gap opening = 10, gap extension = 0.6], multiple alignment [gap opening = 10, gap extension = 0.2], protein weight matrix [Gonnet], gap separation distance = 4 and delay divergent cutoff % of 30) and the Neighbor Joining method on MEGA6 (Tamura *et al.*, 2013) (bootstrap method [1000 replicates], Poisson model [substitution model] with a complete deletion of gaps and missing data treatment settings). Sequence accession numbers included in the analysis were: EhV84 (AEO97861), EhV86 (CAI65473.1), EhV88 (AEP14987.1), EhV164 (.1), EhV99B1 (JF429838.1) EhV201 (AET97936.1), EhV203 (AEO98267.1), EhV207 (AEP15594.1), EhV208 (AEP16020.1), EhV-156 (AHA55176.1), EhV202 (ACS28544.1), EhV18 (AHA54129.1) and EhV145 (AHA54613.1). The same settings were used for the multiple alignment and phylogenetic clustering of nine *E. huxleyi* genotypes from the NA-VICE cruise and 46 additional *E. huxleyi* sequences obtained from NCBI, using a 276-bp long fragment of the algal COI gene. The accession numbers of these additional sequences are shown before the genotype name of each *E. huxleyi* in the phylogenetic tree (Fig. S12).

Nonlinear population models

Three nonlinear ordinary differential equation models were used to explore contrasting assumptions about competition between fast- and slow-infecting EhVs (Fig. 4; Fig. S8; Fig. S9). Detailed descriptions of the models and the procedures used to fit model parameter values can be found in the Supporting Information.

Darwin ecosystem model

The Darwin ecosystem model resolves the cycling of carbon, phosphorus, nitrogen, silica, iron and oxygen through dissolved and particulate organic phases (Follows *et al.*, 2007; Ward *et al.*, 2012; Dutkiewicz *et al.*, 2013). A configuration of the Darwin ecosystem model was used to simulate interactions among four phytoplankton functional groups (diatoms, coccolithophores and representative classes of 'small' and 'large' algae), a single microzooplankton grazer and four representative virus groups specific to each phytoplankton type. Each modelled functional type is differentiated by key traits such as maximum growth rate, sinking rate and so on. All parameters are constrained within ranges reported in the literature and previous ecosystem models (Follows *et al.*, 2007; Bragg *et al.*, 2010; Ward *et al.*, 2012; Dutkiewicz *et al.*, 2015). Model parameterizations and traits related to viral infection of coccolithophores were informed by laboratory infection experiments with representative *E. huxleyi*-EhV pairs (see the Supporting Information). The class of viruses infecting coccolithophores were further divided into `fast' and `slow' types with traits consistent with laboratory observations (Fig. S8, Fig. S9). All biogeochemical and biological tracers were transported and mixed by the MIT general circulation model with $1° \times 1°$ horizontal resolution and 24 vertical levels (Wunsch and Heimbach 2007). The physical fields were consistent with remote sensing and *in situ* altimetric and hydrographic observations (ECCO-GODAE state estimates). Simulations were run for 10 years until the effects of initial conditions were removed and the model converged on reoccurring interannual biomass and productivity patterns.

Data availability

Amplicon sequence data have been deposited to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) and are available under the ID: PRJNA385326. Source code for mathematical modelling is available upon request.

References

Adilakshmi, T., and Laine, R.O. (2002) Ribosomal protein S25 mRNA partners with MTF-1 and La to provide a p53-mediated mechanism for survival or death. *J Biol Chem* **277**: 4147–4151.

Allen, M.J., Forster, T., Schroeder, D.C., Hall, M., Roy, D., Ghazal, P., and Wilson, W.H. (2006) Locus-specific gene expression pattern suggests a unique propagation strategy for a giant algal virus. *J Virol* **80**: 7699–7705.

Bidle, K.D. (2015) The molecular ecophysiology of programmed cell death in marine phytoplankton. *Ann Rev Mar Sci* **7**: 341–375.

Bidle, K.D., Haramaty, L., Barcelos E Ramos, J., and Falkowski, P. (2007) Viral activation and recruitment of metacaspases in the unicellular coccolithophore, Emiliania huxleyi. *Proc Natl Acad Sci U S A* **104**: 6049–6054.

Bidle, K.D., and Vardi, A. (2011) A chemical arms race at sea mediates algal host-virus interactions. *Curr Opin Microbiol* **14**: 449–457.

Bragg, J.G., Dutkiewicz, S., Jahn, O., Follows, M.J., and Chisholm, S.W. (2010) Modeling selective pressures on phytoplankton in the global ocean. *PLoS One* **5**: e9569.

Bratbak, G., Egge, J., Heldal, M., *et al.* (1993) Viral mortality of the marine alga Emiliania huxleyi (Haptophyceae) and termination of algal blooms. *Mar Ecol Prog Ser* **93**: 39–48.

Brown, C.M., and Bidle, K.D. (2014) Attenuation of virus production at high multiplicities of infection in Aureococcus anophagefferens. *Virology* **466–467**: 71–81.

Brussaard, C.P.D., Marie, D., and Bratbak, G. (2000) Flow cytometric detection of viruses. *J Virol Methods* **85**: 175–182.

Brussaard, C.P.D. (2004) Optimization of procedures for counting viruses by flow cytometry. *Appl Environ Microbiol* **70**: 1506–1513.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., *et al.* (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**: 1621–1624.

Collins, J.R., Edwards, B.R., Thamatrakoln, K., Ossolinski, J. E., DiTullio, G.R., Bidle, K.D., *et al.* (2015) The multiple fates of sinking particles in the North Atlantic Ocean. *Global Biogeochem Cycles* **29**: 1471–1494.

Coolen, M.J.L. (2011) 7000 years of Emiliania huxleyi viruses in the Black Sea. *Science* **333**: 451–452.

Cottrell, M.T., and Suttle, C.A. (1995) Dynamics of lytic virus infecting the photosynthetic marine picoflagellate *Micromonas pusilla*. *Limnol Oceanogr* **40**: 730–739.

Dutkiewicz, S., Scott, J.R., and Follows, M.J. (2013) Winners and losers: ecological and biogeochemical changes in a warming ocean. *Global Biogeochem Cycles* **27**: 463–477.

Dutkiewicz, S., Hickman, A.E., Jahn, O., Gregg, W.W., Mouw, C.B., and Follows, M.J. (2015) Capturing optically important constituents and properties in a marine biogeochemical and ecosystem model. *Biogeosciences* **12**: 4447–4481.

Eren, A.M., Maignien, L., Sul, W.J., Murphy, L.G., Grim, S.L., Morrison, H.G., and Sogin, M.L. (2013a) Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol Evol* **4**: 1111–1119.

Eren, A.M., Vineis, J.H., Morrison, H.G., and Sogin, M.L. (2013b) A filtering method to generate high quality short reads using illumina paired-end technology. *PLoS One* **8**: e66643.

Eren, A.M., Borisy, G.G., Huse, S.M., and Mark Welch, J.L. (2014) Oligotyping analysis of the human oral microbiome. *Proc Natl Acad Sci U S A* **111**: E2875–E2884.

Evans, C., Malin, G., Mills, G.P., and Wilson, W.H. (2006) Viral infection of Emiliania huxleyi (Prymnesiophyceae) leads to elevated production of reactive oxygen species. *J Phycol* **42**: 1040–1047.

Fischer, T., Elenko, E., McCaffery, J.M., DeVries, L., and Farquhar, M.G. (1999) Clathrin-coated vesicles bearing GAIP possess GTPase-activating protein activity in vitro. *Proc Natl Acad Sci U S A* **96**: 6722–6727.

Follows, M.J., Dutkiewicz, S., Grant, S., and Chisholm, S.W. (2007) Emergent biogeography of microbial communities in a model ocean. *Science* **315**: 1843–1846.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**: 3150–3152.

Fulton, J.M., Fredricks, H.F., Bidle, K.D., Vardi, A., Kendrick, B. J., Ditullio, G.R., and Van Mooy, B.A.S. (2014) Novel molecular determinants of viral susceptibility and resistance in the lipidome of Emiliania huxleyi. *Environ Microbiol* **16**: 1137–1149.

Hakomori, S. (2008) Structure and function of glycosphingolipids and sphingolipids: recollections and future trends. *Biochim Biophys Acta* **1780**: 325–346.

Han, G., Gable, K., Yan, L., Allen, M.J., Wilson, W.H., Moitra, P., *et al.* (2006) Expression of a novel marine viral single-chain serine palmitoyltransferase and construction of yeast and mammalian single-chain chimera. *J Biol Chem* **281**: 39935–39942.

Hanada, K. (2003) Serine palmitoyltransferase, a key enzyme of sphingolipid metabolism. *Biochim Biophys Acta* **1632**: 16–30.

Hunter, J.E., Frada, M.J., Fredricks, H.F., Vardi, A., and Van Mooy, B.A.S. (2015) Targeted and untargeted lipidomics of Emiliania huxleyi viral infection and life cycle phases highlights molecular biomarkers of infection, susceptibility, and ploidy. *Front Mar Sci* **2**: 1–12.

Laber, C.P., Hunter, J.E., Carvalho, F., Collins, J.R., Hunter, E.J., Schieler, B.M., *et al.* (2018) Coccolithovirus facilitation of carbon export in the North Atlantic. *Nat Microbiol* **3**: 537–547.

Lehahn, Y., Koren, I., Schatz, D., Frada, M., Sheyn, U., Boss, E., *et al.* (2014) Decoupling physical from biological processes to assess the impact of viruses on a mesoscale algal bloom. *Curr Biol* **24**: 2041–2046.

Li, W., and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.

Martínez Martínez, J., Schroeder, D.C., Larsen, A., Bratbak, G., and Wilson, W.H. (2007) Molecular dynamics of Emiliania huxleyi and cooccurring viruses during two separate mesocosm studies. *Appl Environ Microbiol* **73**: 554–562.

Martínez, J.M., Schroeder, D.C., and Wilson, W.H. (2012) Dynamics and genotypic composition of Emiliania huxleyi and their co-occurring viruses during a coccolithophore bloom in the North Sea. *FEMS Microbiol Ecol* **81**: 315–323.

Merrill, A.H. (2002) De novo sphingolipid biosynthesis: a necessary, but dangerous, pathway. *J Biol Chem* **277**: 25843–25846.

Michaelson, L.V., Dunn, T.M., and Napier, J.A. (2010) Viral trans-dominant manipulation of algal sphingolipids. *Trends Plant Sci* **15**: 651–655.

Monier, A., Pagarete, A., de Vargas, C., Allen, M.J., Read, B., Claverie, J.-M., and Ogata, H. (2009) Horizontal gene transfer of an entire metabolic pathway between a eukaryotic alga and its DNA virus. *Genome Res* **19**: 1441–1449.

Nissimov, J.I., Worthy, C.A., Rooks, P., Napier, J.A., Kimmance, S.A., Henn, M.R., *et al.* (2012) Draft genome sequence of four Coccolithoviruses: Emiliania huxleyi virus EhV-88, EhV-201, EhV-207, and EhV-208. *J Virol* **86**: 2896–2897.

Nissimov, J., Jones, M., Napier, J.A., Munn, C.B., Kimmance, S.A., and Allen, M.J. (2013) Functional inferences of environmental coccolithovirus biodiversity. *Virol Sin* **28**: 291–302.

Nissimov, J.I., Napier, J.A., Kimmance, S.A., and Allen, M.J. (2014) Permanent draft genomes of four new coccolithoviruses: EhV-18, EhV-145, EhV-156 and EhV-164. *Mar Genomics* **15**: 7–8.

Nissimov, J.I., Napier, J.A., Allen, M.J., and Kimmance, S.A. (2016) Intragenus competition between coccolithoviruses: an insight on how a select few can come to dominate many. *Environ Microbiol* **18**: 133–145.

Nissimov, J.I., Pagarete, A., Ma, F., Cody, S., Dunigan, D.D., Kimmance, S.A., and Allen, M.J. (2017) Coccolithoviruses: a review of cross-kingdom genomic thievery and metabolic thuggery. *Viruses* **9**: 52. https://doi.org/10.3390/v9030052.

Nissimov, J.I., Vandzura, R., Johns, C.T., Natale, F., Haramaty, L., and Bidle, K.D. (2018) Dynamics of transparent exopolymer particle production and aggregation during viral infection of the coccolithophore, Emiliania huxleyi. *Environ Microbiol* **20**: 2880–2897.

Pagarete, A., Lanzén, A., Puntervoll, P., Sandaa, R.A., Larsen, A., Larsen, J.B., *et al.* (2013) Genomic sequence and analysis of EhV-99B1, a new coccolithovirus from the Norwegian fjords. *Intervirology* **56**: 60–66.

Pagarete, A., Kusonmano, K., Petersen, K., Kimmance, S.A., Martínez Martínez, J., Wilson, W.H., *et al.* (2014) Dip in the gene pool: metagenomic survey of natural coccolithovirus communities. *Virology* **466–467**: 129–137.

Popendorf, K.J., Fredricks, H.F., and Van Mooy, B.A.S. (2013) Molecular ion-independent quantification of polar glycerolipid classes in marine plankton using triple quadrupole MS. *Lipids* **48**: 185–195.

Prozialeck, W.C., Fay, M.J., Lamar, P.C., Pearson, C.A., Sigar, I., and Ramsey, K.H. (2002) Chlamydia trachomatis disrupts N-cadherin-dependent cell-cell junctions and sequesters beta-catenin in human cervical epithelial cells. *Infect Immun* **70**: 2605–2613.

Roberts, K.P. (2002) A comparative analysis of expression and processing of the rat epididymal fluid and sperm-bound forms of proteins D and E. *Biol Reprod* **67**: 525–533.

Rose, S.L., Fulton, J.M., Brown, C.M., Natale, F., Van Mooy, B.A.S., and Bidle, K.D. (2014) Isolation and characterization of lipid rafts in Emiliania huxleyi: a role for membrane microdomains in host-virus interactions. *Environ Microbiol* **16**: 1150–1166.

Rosenwasser, S., Mausz, M.A., Schatz, D., Sheyn, U., Malitsky, S., Aharoni, A., *et al.* (2014) Rewiring host lipid metabolism by large viruses determines the fate of Emiliania huxleyi, a bloom-forming alga in the ocean. *Plant Cell* **26**: 2689–2707.

Ruiz, E., Oosterhof, M., Sandaa, R.-A., Larsen, A., and Pagarete, A. (2017) Emerging interaction patterns in the Emiliania huxleyi-EhV system. *Viruses* **9**: 61. https://doi.org/10.3390/v9030061.

Rütti, M.F., Richard, S., Penno, A., von Eckardstein, A., and Hornemann, T. (2009) An improved method to determine serine palmitoyltransferase activity. *J Lipid Res* **50**: 1237–1244.

Schatz, D., Shemi, A., Rosenwasser, S., Sabanay, H., Wolf, S.G., Ben-Dor, S., and Vardi, A. (2014) Hijacking of an autophagy-like process is critical for the life cycle of a DNA virus infecting oceanic algal blooms. *New Phytol* **204**: 854–863.

Schroeder, D.C., Oke, J., Malin, G., and Wilson, W.H. (2002) Coccolithovirus (Phycodnaviridae): characterisation of a new large dsDNA algal virus that infects Emiliana huxleyi. *Arch Virol* **147**: 1685–1698.

Sheyn, U., Rosenwasser, S., Ben-Dor, S., Porat, Z., and Vardi, A. (2016) Modulation of host ROS metabolism is essential for viral infection of a bloom-forming coccolithophore in the ocean. *ISME J* **10**: 1742–1754.

Sheyn, U., Rosenwasser, S., Lehahn, Y., Barak-Gavish, N., Rotkopf, R., Bidle, K.D., *et al*. (2018) Expression profiling of host and virus during a coccolithophore bloom provides insights into the role of viral infection in promoting carbon export. *ISME J* **12**: 704–713.

Sorensen, G., Baker, A.C., Hall, M.J., Munn, C.B., and Schroeder, D.C. (2009) Novel virus dynamics in an Emiliania huxleyi bloom. *J Plankton Res* **31**: 787–791.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**: 2725–2729.

Vardi, A., Van Mooy, B.A.S., Fredricks, H.F., Popendorf, K.J., Ossolinski, J.E., Haramaty, L., and Bidle, K.D. (2009) Viral glycosphingolipids induce lytic infection and cell death in marine phytoplankton. *Science* **326**: 861–865.

Vardi, A., Haramaty, L., Van Mooy, B.A.S., Fredricks, H.F., Kimmance, S.A., Larsen, A., and Bidle, K.D. (2012) Host-virus dynamics and subcellular controls of cell fate in a

natural coccolithophore population. *Proc Natl Acad Sci U S A* **109**: 19327–19332.

Van Mooy, B.A.S., and Fredricks, H.F. (2010) Bacterial and eukaryotic intact polar lipids in the eastern subtropical South Pacific: water-column distribution, planktonic sources, and fatty acid composition. *Geochim Cosmochim Acta* **74**: 6499–6516.

Ward, B.A., Dutkiewicz, S., Jahn, O., and Follows, M.J. (2012) A size-structured food-web model for the global ocean. *Limnol Oceanogr* **57**: 1877–1891.

Wilson, W.H., Schroeder, D.C., Allen, M.J., Holden, M.T.G., Parkhill, J., Barrell, B.G., *et al.* (2005) Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. *Science* **309**: 1090–1092.

Wunsch, C., and Heimbach, P. (2007) Practical global oceanic state estimation. *Physica D* **230**: 197–208.

Ziv, C., Malitsky, S., Othman, A., Ben-Dor, S., Wei, Y., Zheng, S., *et al.* (2016) Viral serine palmitoyltransferase induces metabolic switch in sphingolipid biosynthesis and is required for infection of a marine alga. *Proc Natl Acad Sci U S A* **113**: E1907–E1916.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1: Supplementary Information.