EVALUATING ECOSYSTEM FUNCTIONS FOR COCCOLITHOVIRUS INFECTION OF EMILIANIA HUXLEYI: BRIDGING MICRO, MESO, AND GLOBAL SCALES

by

CHRISTIEN PHILIP LABER

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Oceanography

Written under the direction of

Kay D. Bidle

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2018

ABSTRACT OF THE DISSERTATION

Evaluating Ecosystem Functions for *Coccolithovirus* Infection of *Emiliania huxleyi*: Bridging Micro, Meso, and Global Scales

by CHRISTIEN PHILIP LABER

Dissertation Director:

Kay Bidle

The coccolithophore *Emiliania huxleyi* is a globally distributed marine algal species that regularly forms large surface ocean blooms that can last from weeks to months. As a cosmopolitan species that also forms calcium carbonate plates called coccoliths, *E. huxleyi* plays a critical role in influencing both organic and inorganic carbon cycles. North Atlantic blooms of the algae are regularly infected by a double stranded DNA virus called *Coccolithovirus* (EhV). The ecological and biogeochemical influences of viral infection in marine algae are largely unknown, although studies largely suggest infection enhances chemical cycling within the microbial loop. This dissertation investigates *E. huxleyi*-EhV interactions within marine algal communities to further elucidate the role viruses play in influencing their host as well as the surrounding ecosystem. On a local scale, this was investigated by locating and interrogating mesoscale North Atlantic blooms of *E. huxleyi* with a comprehensive toolset using MODIS/AQUA satellite imagery, a suite of diagnostic lipid- and gene-based molecular biomarkers, *in situ* optical sensors, and sediment traps to show that EhV infections are coupled with particle aggregation, high zooplankton grazing, and enhanced downward

vertical fluxes of both particulate organic and particulate inorganic carbon from the upper mixed layer to the mesopelagic. The finding that viruses can stimulate vertical carbon flux through a mechanistic interplay with zooplankton grazers introduces novel complexities into microbial ecosystem interactions. These bloom communities are further investigated using optical absorption spectra, phytoplankton pigment composition, and flow cytometry, revealing EhV infection has little influence driving the variability in phytoplankton absorption compared to community photoacclimation within the blooms. It is further observed that photoacclimation driven absorption characteristics are discernible through different phytoplankton taxonomic compositions. On the global scale, the known biogeography of EhV infection and diversity of EhVs is expanded out of the North Atlantic Ocean using a combination of the Tara Oceans global metagenome database as well as targeted regional sampling throughout the Pacific Ocean. These results reveal that EhV infection is a global phenomenon tightly coupled to *E. huxleyi* production and is pervasive in bloom and non-bloom environments. Collectively, these studies show that EhVs are an influential regulator of carbon cycling by enhancing biological pump efficiency and through revealing their global distribution in the ocean.

Acknowledgements

I would like to thank my committee for their guidance throughout this dissertation. Foremost, I would like to thank Kay Bidle for the multitude of opportunities stimulating scientific and personal growth throughout my time under his advisement. His approach to both science and life are driven by creativity and inclusivity, and these tendencies have gratefully been imprinted in my conduct as well. I would like to thank Oscar Schofield for giving me sea legs early in my graduate career and stimulating my interest in ocean optics. The foundational experience I gained on my first Antarctic cruise provided both inspiration and my conviction that the sea holds a deep and rugged beauty. I would like to thank Emmanuel Boss and Debashish Bhattacharya for their critical input and advice throughout my PhD. I would not have the scientific wisdom required for this venture without their support and guidance.

I would like to thank the current and past members of the Bidle Lab for making the PhD experience not only of great value but also of great kinship. The atmosphere provided by such positive, collaborative group of people will be fondly remembered. Special thanks to Frank Natale and Liti Haramaty for all their assistance in sample processing and helping teach me important laboratory techniques. Thanks to Jozef Nissimov for so much aloha even when we nearly shared a bed. Thank you Kim Thamatrakoln, Chris Brown, Mansha Seth-Pasricha, Chris Johns, Ben Knowles, Dina AlRoumi, Chana Kranzler, Jason Latham, Maeve Eason Hubbard for experiences from NA-VICE to MESOHUX.

I would also like to thank all the scientists who helped me with sample collection throughout the Pacific Ocean and for providing the opportunities that made my science possible. Special thanks go to Yoshihiro Shiraiwa for facilitating and advising my EAPSI/JSPS summer fellowship and providing me with numerous connections that enriched my scientific and social experience in Japan. Thanks to the scientists lead by Dave Karl at University of Hawaii - Manoa, and Debora Iglesias-Rodriguez at University of California Santa Barbara for generously facilitating and assisting in sample collection throughout the Pacific. I would also like to thank Udi Zelzion for his extensive help with bioinformatics and understanding many concepts in genetics.

I am also grateful for all the friendships formed during my time at Rutgers. Thank you to the wonderful members of my GPO cohort whom which I grew as scientist and shared the ups and downs of the graduate student life. Thank you Filipa Carvalho, Nicole Couto, Nicole Waite, Jack McSweeney, Alex Lopez, Kevin Crum and Katie Harazin for this experience. Special thanks go to my cohort lab mate Brittany Schieler for the multitude of experiences, assistance, and empathy shared throughout our time at Rutgers and for all the help given after my premature departure from New Jersey. Thanks to Chris Free for being an invaluable friend and sharing great hikes, great beer, and great stories. Also thanks to Fernando Pareja-Roman for occasionally giving my brain a rest from science to discuss good music and ideas.

I would also like to give a very special thank you to my family for all of their support throughout my PhD as well as my entire life preceding. My parents taught me to be curious and to seek wisdom and beauty in every discipline of life and for that I am forever grateful. They also continually assist me with pursuing my passions in life as

well as provide moral support to find my way when difficulties pervade. I'd like to thank my brother for providing entertainment and light spirits over the years as well as heartfelt advice when needed. And thank you Cleo for being my counterpoint, with a melody that motivates my life every day.

During this PhD, I have been funded by Rutgers, the State University of New Jersey Graduate Program in Oceanography, grants from the National Science Foundation (OCE-1061876; OISE-1310730) and the Gordon and Betty Moore Foundation (GBMF3789) and the Japanese Society for the Promotion of Science Summer Fellowship. It is also acknowledged that Chapter 2 of this dissertation exists as a current publication in the journal Nature Microbiology (Laber *et al*. 2018). Chapters 3 and 4 are currently manuscripts in preparation for journal submission.

Table of Contents

List of Tables

List of Figures

Figure 2.4. Active EhV infection triggers TEP-facilitated aggregation, POC export, and remineralization in the mesopelagic. *(a) B_b spike signals revealed elevated particle aggregates at EI/EI_R (blue/green) below the euphotic zone compared to LI and PI (black and red). Centroids indicate the mean spike signal for 10m bins over an entire float deployment (dotted lines indicate ± 1 SE). (b) Particle dynamics at EI/EI_R were accompanied by high carbon remineralization rates in the mesopelagic. Dotted lines indicate 95% confidence intervals from linear regression. (c-d) Susceptible, EhV-infected cells were preferentially exported. (c) BLL (22:6/ 22:6)/(18:1/ 22:6) ratios for suspended biomass in the mixed layer (down to 30m), below the mixed layer, and in sediment traps (positioned at 50m, 150m, and 300m). Kruskal-Wallis ANOVA showed significantly higher ratios in sediment trap material. (d) Elevated PIC sinking flux was accompanied by increased sGSL sinking flux in the same sediment trap material. (e-f) Modeled geographic origins of sinking particles. MODIS/AQUA PIC images on 3 July (e) and 23 June (f). Tracks of deployed sediment traps shown in black for the EI/EI_R (e) and PI (f) deployments. The EI/EI_R (EI, blue star; EI_R, red star) and PI simulations (blue star) are within 10 km and 5 km of sediment trap trajectories, respectively. Dispersion simulations indicate that 99% of particles from the 50 m, 150 m and 300 m traps originated in circles of radius 2 km, 5 km, and 7 km, respectively. (g) TEP concentrations were low in the upper 50m at EI/EI_R and highest at LI as determined by ANOVA. (h) SEM images from EI at 10m and 150m respectively showed loose aggregates of cells/coccoliths compared*

Supplemental Figure 2.4. *Correlation analyses comparing profiling float optical backscatter (700nm) and chlorophyll fluorescence measurements with flow cytometry counts of specific groups of phytoplankton. Both analyses revealed that E. huxleyi abundance had a significant impact on these optical parameters. (a) Backscatter has a better relationship to E. huxleyi abundance than it does to (b) total phytoplankton abundance despite that E. huxleyi only makes up between 0.5-8% of the total cell count (d). (c) Coccolith concentrations also correlated with backscatter but not as well as E. huxleyi. Correlations (a-c) are anchored by the deeper measurements with generally low particle numbers having low backscatter. In shallow water measurements, where backscatter is decoupled with depth, the data in (a) show a tighter fit to the regression than in (b), suggesting that E. huxleyi cells contribute significantly to the backscatter signal within a mixed population. (d-h) comparison of the relative impacts of cell abundance for discernable subgroups of the community (via flow cytometry) on the chlorophyll:backscatter ratio. (d) E. huxleyi had the smallest range in abundance but was the only group that supported a significant relationship to the chlorophyll:backscatter ratio. We recognize that the (e) high fluorescence and (f) low fluorescence containing eukaryote groups are phylogenetically broader than E. huxleyi but they are distinguished by their similar chlorophyll fluorescence and side scatter properties. (g) Synechococcus was numerically dominant in many measurements but had a negligible influence on chlorophyll:backscatter. (h) Depth showed a small but positive*

Supplemental Figure 2.8. Relationship between total carbon flux and backscatter spike signal. *The measured particulate organic carbon (POC) flux rate measurements at 50 m (circles) and 150 m (triangles) was compared to the spike aggregate signal at those respective depths at EI, EI_R, and PI. Briggs et al. explored using the spike signal as an estimate for aggregate POC flux using equation [(2); rearranged below] using a 75 m d^{-1} sinking rate, terms that accounted for the ratio of POC to optical backscatter (B_b) signal [35400 mg C m^{-2}], and a correction for organic carbon loss in a particle during sinking [$(z/100m)^{-0.28}$], where z is the depth of the measurement (Briggs et al. 2011). The Briggs et al. estimates for the relationship between POC and the spike signal at 50 m and 150 m are shown in orange. Our data depart from this relationship between POC and the B_b spike signal, likely based on several factors. First, the significant contribution of PIC to the total particulate carbon (TPC) in our systems would change the relationship between POC and the B_b spike signal, as PIC would then account for a portion of the signal (PIC is assumed to inflate the B_b signal and thus lower the ratio). (b-c) The % contribution of POC to TPC varied among samples and stations, likely explaining some of the variability observed in the POC:B_b relationship (c). EI and EI_R commonly exhibited very low % POC within observations. However, using Eq. (2), a lower POC:B_b ratio would increase the slope observed in (a) rather than decreasing it to one comparable to our data. A*

Supplemental Figure 2.13. *Community and E. huxleyi specific growth and grazing rates derived from Landry-based (Landry & Hassett 1982) dilution experiments performed at EI (a, c, e, g) and EI_R (b, d, f, h). (A-B) Using the classic chlorophyll based measurement of growth rate, linear projection to community growth in the absence of grazers shows positive growth rates over the 24 h incubation period for both EI and EI_R. Grazing was relatively elevated at EI in comparison to EI_R. (c-d) While cell-specific community growth rates also showed generally positive growth rates in the absence of grazing, cell specific grazing pressure was higher at EI_R than EI. These differences may be explained by selective retention of phytoplankton during chlorophyll filtration (inefficient retention of small, < 0.7 μm, picophytoplankton such as Synechococcus onto GF-F filters) or by preferential grazing on particular prey species. (e-f) E. huxleyi specific grazing was elevated (almost three times higher) at both EI and EI_R in comparison to community grazing rates. (g-h) Some incubations were spiked with ammonia, phosphate, and silica to alleviate the possible impact of nutrient limitation on cells in dilution experiment. Statistically indistinguishable changes in cell concentration over 24 h for incubations with (Nut+; red) and without (Nut-; black) nutrients argue that resident populations were not nutrient limited (Student's T test, p = 0.38, 0.47, respectively for g and h T24). For (g); T0, n = 2; T24, n = 3 biological replicates. For (h); n = 3 biological replicates. Open circles indicate individual data points. Error bars represent ± 1 SD around the mean.....61*

Chapter 1: Introduction

Emiliania huxleyi is a cosmopolitan, coccolithophorid prymnesiophyte that grows globally in surface ocean waters with the exception of polar regions (Thierstein & Young 2004) and is well known for its production of calcite coccoliths that surround the cell and are continually shed into the water column. These chalky white plates are highly reflective and allow *E. huxleyi* blooms, which can exceed 10^5 km^2 with cell densities reaching 10^5 ml^{-1} (Turkoglu 2008), to be readily observed from space (Brown & Yoder 1994). As the most abundant coccolithophore species, it significantly contributes to the nearly one third of calcium carbonate production in the ocean by coccolithophores (Thierstein & Young 2004). Abundant coccolith formation and removal from the surface ocean counterbalances the organic carbon pump and can reduce or reverse the movement of CO_2 from the atmosphere into the ocean (Rost & Riebesell 2004).

Termination of *E. huxleyi* blooms is commonly accompanied by viral infection from the double-stranded DNA-containing *Coccolithoviruses* (EhVs) (Bratbak *et al.* 1993; Bratbak *et al.* 1996; Martínez *et al.* 2012). The lytic infection window of *E. huxleyi* cells by EhV is typically around 60-72 h, and can be as short as 48 hours, at which time upwards of 800-1000 free viruses are released into the environment (Bidle *et al.* 2007; Vardi *et al.* 2009; Bidle & Kwityn 2012). In dense coastal and open ocean blooms, infection spreads rapidly and can terminate large blooms in a period of weeks (Vardi *et al.* 2012; Lehahn *et al.* 2014). Traditionally, viral infection was thought to overturn a quarter of the carbon and release dissolved organic matter in the upper ocean, preventing trophic transfer to higher organisms and into the deep sea, instead fueling the microbial loop and increasing upper ocean respiration (Fuhrman 1999; Suttle 2005).

However, recent observations have documented intense aggregation of phytoplankton cells through the enhanced production of transparent exopolymeric (TEP) material during infection of natural populations (Vardi *et al.* 2012; Bidle 2015), which may serve to greatly stimulate particle aggregation and vertical sinking flux.

Historically, quantitative measurements of virus-induced lysis of natural phytoplankton populations have been problematic since it is been difficult to show that a specific virus can cause substantial lysis of its host at sea. This is rapidly and fundamentally changing. Recent work in several algal-virus systems has shed novel insight into elegant strategies of viral infection and cellular regulation of cell fate at the interface of specific cellular pathways—such as the molecular machinery involved in the reactive oxygen stress response, autophagy, and programmed cell death (PCD) in the case of *E. huxleyi*-EhV interactions (Bidle *et al.* 2007; Bidle & Vardi 2011; Schatz *et al.* 2014; Bidle 2015)—which not only reveals tantalizing aspects of viral replication and host resistance strategies, but also provides new diagnostic biomarkers for elucidating virus-mediated processes in the oceans (Vardi *et al.* 2009). Due to the collective insight from genomic analytics (Read *et al.* 2013) and the array of genetically diverse host (sensitive and resistant; naked and calcified) and virus strains in culture (Schroeder *et al.* 2002; Wilson *et al.* 2002; Bidle & Kwityn 2012; Johns *et al.* 2018; Nissimov 2018) the *E. huxleyi*-EhV host-virus system has emerged as one of the best model systems to investigate algal host-virus interactions and the subcellular processes mediating infection dynamics. Coccolithoviruses employ a sophisticated, co-evolutionary "arms race" to manipulate host lipid metabolism, alter glycosphingolipid (GSL) production, and regulate cell fate via autophagy and PCD [via ROS, metacaspase expression, and caspase activity

(Bidle *et al.* 2007; Vardi *et al.* 2009; Bidle & Vardi 2011; Vardi *et al.* 2012; Rosenwasser *et al.* 2014; Schatz *et al.* 2014). This mechanistic understanding has provided a unique suite of biomolecular proxies to diagnose and track infection in natural populations through the analysis of host and virus-specific genes and transcripts, as well as a suite of unique host- and virus-derived GSLs that are diagnostic of host dynamics, the state of host sensitivity, and active infection (Vardi *et al.* 2009; Vardi *et al.* 2012). These techniques have provided novel insight not only into the positive identification EhV infection, but have allowed researchers to assess the its ecological and biogeochemical impact through the coupling of oceanographic measurements.

The North Atlantic Viral Infection of Coccolithophores Expedition *(NA-VICE*; http://www.bco-dmo.org/project/2136) utilized a diverse toolset to identify EhV infection of *E. huxleyi* within North Atlantic blooms and investigate the dynamics of host growth and viral infection within the blooms, as well how this interaction regulates the flow of carbon and sulfur cycling within and exported from the surface ocean. Furthermore, the study was designed to test the efficiency of previously mentioned GSL biomarkers of infection as cornerstone molecules in the ocean driving host-virus interactions and the infection process in natural communities. Currently, studies from this project have revealed bloom dynamics of infected *E. huxleyi* populations and associated communities from multiple angles and scales, ranging from satellite observations of the complete life cycle of an infected coccolithophore bloom (Lehahn *et al.* 2014), the spread of viral particles through sea spray aerosols (Sharoni *et al.* 2015), high resolution genetic observations of infection dynamics throughout the water column (Sheyn *et al.* 2018), sinking particle flux dynamics and associated bacterial respiration and remineralization

rates (Collins *et al.* 2015), to investigating migrating zooplankton as transmission vectors for algal viruses (Frada *et al.* 2014).

Unfortunately, ship-based oceanographic approaches have a limited reach in molecular and microbial oceanography due to available ship time and coarse spatial resolution. *In situ* autonomous optical platforms have the potential to alleviate these problems with high sampling resolution (temporally and spatially) and long deployment times (weeks to years). Additionally, they may be deployed in advance of a study to identify areas of relevant interest, in contrast to searching blindly. Gliders and vertical profiling floats have been outfitted with bio-optical sensors and are increasingly providing detailed information on primary production and associated processes (Kahl *et al.* 2010; Alkire *et al.* 2012; Zhao *et al.* 2013). Identifying bio-optical signatures associated with EhV infection of *E. huxleyi* (given the aforementioned optical qualities of this organism) would provide a powerful complimentary tool for exploring *in situ* infection dynamics and subsequent impacts on ecosystem and biogeochemical cycling. Coupling bio-optical based observations with the aforementioned, diagnostic biomolecular markers of infection and cellular response mechanisms opens the possibility to interrogate these interactions and ecosystem response at an unprecedented range of scales and detail.

Pairing a detailed understanding of the ecological and biogeochemical influences of EhV infection with a more complete knowledge of the global biogeographic distribution of infection would also add context to the relevance of EhV infection to ocean chemical cycling. Despite over two decades of study since the initial discovery of viral particles associated with *E. huxleyi* blooms, exploration of the interaction between

this host and virus has been limited to the North Atlantic and marginal seas (Bratbak *et al.* 1996; Schroeder *et al.* 2003; Vardi *et al.* 2009; Coolen 2011). Meanwhile, *E. huxleyi*'s global distribution, with blooms in the South Atlantic and North Pacific that rival those of the North Atlantic, is unstudied regarding the interplay of host and virus (Balch *et al.* 2016; Ladd *et al.* 2016). Further, the diverse pan-genome of the algae raises critical questions about the distribution and diversity of its virus (Read *et al.* 2013). The importance of EhV infection in low host density populations is also relatively unexplored and requires further study.

The occurrence of monospecific blooms in the ocean is a relatively rare event, and therefore it is necessary to study elements of the community dynamics that surround any individual species population for a more comprehensive understanding of system in which it exists and interacts. One of the long-term goals of remote sensing and *in situ* optical oceanography has been developing the utility of ocean color as a tool for better understanding primary production of phytoplankton in the surface ocean (Bracher *et al.* 2017). This is largely through developing an understanding of the optical absorptive characteristics of phytoplankton in relation to taxonomy as well as photophysiological conditioning. Due to the multiple interactive variables that influence spectral absorption characteristics of phytoplankton communities (prominently community composition, light and nutrient availability), it has been necessary to study the impact of these variables in isolation as well as combined in order to inform algorithms that predict phytoplankton production and biomass.

This Ph. D. thesis explores EhV infection of *E. huxleyi* and its corresponding ecological and biogeochemical impact in the upper ocean. It expands our understanding

of *E. huxleyi*-EhV interactions in different ecological settings (bloom versus non-bloom) as well as different oceanic regions for a more complete understanding of the extent to which EhVs regulate the fate of fixed carbon associated with *E. huxleyi* populations. *In situ* autonomous ocean optics can provide high spatial resolution of discrete water column optical properties and, when combined with molecular methods of detecting EhV infection, may collectively provide accurate identification of *in situ* EhV infection in natural populations. These properties include scattering, fluorescence, and absorptive properties associated with infection.

Chapter 2 investigates the infection state of open ocean natural *E. huxleyi* blooms occurring within the North Atlantic to better understand how coupled EhV infection with surrounding ecosystem processes can increase the flux of carbon from the surface ocean into the mesopelagic. This study was conducted during the *North Atlantic Viral Infection of Coccolithophores Expedition*, a ~2000 nautical mile transect between Ponta Delgada, Azores to Reykjavik, Iceland from 15 June to 14 July 2012. Here, we utilize a diverse and comprehensive swath of oceanographic tools ranging from MODIS Aqua satellite products of chlorophyll *a* and particulate organic carbon, bio-optical vertical profiling floats, flow cytometry and cell staining techniques, to molecular lipid-based diagnostic markers of EhV infection and host susceptibility.

Chapter 3 further explores the infected natural bloom populations discussed in Chapter 2, but in the context of the phytoplankton communities present in the studied water masses and associated photoacclimation throughout the euphotic water column, and their influences on the *in situ* optical absorption spectra of phytoplankton. We utilize pigment concentrations and ratios to discriminate taxonomic composition of the blooms,

as well as explore the similarity of the blooms from a taxonomic standpoint. We also evaluate the photoacclimation states of these blooms based on pigmentation and phytoplankton absorption spectral characteristics. Finally, we identify that photoacclimation as the primary driver in regulating observed spectral absorption characteristics.

Chapter 4 expands the study of EhV infection from the North Atlantic into the global ocean with a combined suite of distinct genetic techniques to probe other oceanic systems for active infection of *E. huxleyi* by EhVs. We take advantage of the globally-distributed Tara Oceans metagenome assembly dataset to identify EhV-like genes associated with collected algal biomass throughout the global ocean, as well as conduct targeted EhV- and *E. huxleyi*-specific gene amplification from diverse bodies of water throughout the Pacific Ocean, including: an anomalous coccolithophore bloom in the California current near Santa Barbara Channel, California; the oligotrophic waters of Station ALOHA within the central North Pacific gyre; and coastal and offshore locations in the western Pacific around the Kuroshio current near Japan. We find evidence of EhV infection that is represented on a global scale within host populations across diverse and dynamic host populations, ranging from large *E. huxleyi* blooms (exceeding 10^3 cells ml^{-1}) to low concentration host populations (~10^2 cells ml^{-1}).

Collectively, these studies bring new insight into viral ecology by revealing novel pathways for virally mediated carbon biogeochemistry and uncovering the biogeography of ecosystems that encompass this host-virus interaction. The findings reveal virus infection as a pervasive ecosystem process that can facilitate export fluxes into the mesopelagic and enhance biological pump efficiency and reveal layers of carbon flux

biogeochemistry the quantitative contributions of which will help the decades-long challenge of resolving ecosystem influences on carbon export. They also reveal that virus infection of this cosmopolitan coccolithophore is pervasive across diverse ocean basins and host population densities. Results of this PhD thesis present variety of key, open questions and future opportunities to elucidate the infection strategies used by EhVs at the range of host cell densities found across the global oceans and the ecological and biogeochemical impacts of viruses on Earth's carbon cycle, including heretofore unappreciated linkages with vertical carbon flux and grazing ecosystem pathways that were traditionally thought to be independent and competing.

Chapter 2: Coccolithovirus Facilitation of Carbon Export in the North Atlantic

2.1 ABSTRACT

Marine phytoplankton account for approximately half of global primary productivity (Field *et al.* 1998), making their fate an important driver of the marine carbon cycle. Viruses are thought to recycle more than a quarter of oceanic photosynthetically-fixed organic carbon (Weitz *et al.* 2015), which can stimulate nutrient regeneration, primary production, and upper ocean respiration (Weitz *et al.* 2015) via lytic infection and the 'virus shunt'. Ultimately, this limits the trophic transfer of carbon and energy to both higher food webs and the deep ocean (Weitz *et al.* 2015). Using a combination of MODIS/AQUA satellite imagery, a suite of diagnostic lipid- and gene-based molecular biomarkers, *in situ* optical sensors, and sediment traps, we show that coccolithovirus infections of mesoscale (~100 km) *Emiliania huxleyi* blooms in the North Atlantic are coupled with particle aggregation, high zooplankton grazing, and greater downward vertical fluxes of both particulate organic and particulate inorganic carbon from the upper mixed layer. Our analyses captured blooms in different phases of infection (early, late and post) and revealed the highest export flux in 'early-infected blooms' with sinking particles being disproportionately enriched with infected cells and subsequently remineralized at depth in the mesopelagic. Our findings reveal viral infection as a previously unrecognized ecosystem process enhancing biological pump efficiency.

2.2 INTRODUCTION

Coccolithophores, calcified photosynthetic protists, account for ~20% of marine primary productivity (Rousseaux & Gregg 2013), 12% (0.2 Gt y^{-1}) of global $CaCO_3$ export flux, and half of the $CaCO_3$ reservoirs reaching deep sea sediments (Berelson *et al.* 2007). The most abundant and globally-distributed coccolithophore *Emiliania huxleyi* forms massive blooms in the oceans that can exceed 10^5 km^2 with cell densities reaching 10^5 cells ml^{-1} (Tyrrell & Merico 2004). The termination of dense coastal and open ocean North Atlantic blooms of *E. huxleyi*, which can last over a period of days to weeks, is commonly attributed to infection by double-stranded DNA-containing coccolithoviruses (EhVs) (Bratbak *et al.* 1993; Vardi *et al.* 2012; Lehahn *et al.* 2014). The interaction between *E. huxleyi* and EhVs is mediated by a lipid-based, co-evolutionary arms race centered around the rewiring of host lipid metabolism, glycosphingolipid production and the activation and recruitment of host's oxidative stress, autophagy, and programmed cell death (PCD) pathways, a process that takes between 72-96 h (Bidle *et al.* 2007; Vardi *et al.* 2009; Schatz *et al.* 2014; Ziv *et al.* 2016). This mechanistic insight has led to the development of functional biomarkers (Vardi *et al.* 2009; Vardi *et al.* 2012; Fulton *et al.* 2014), which can now be used to diagnose the level of active infection in natural oceanic populations and quantitatively assess its ecosystem and biogeochemical impacts — a capability that has been lacking in virus ecology.

A unique suite of structurally distinct polar lipids — glycosphingolipids (GSLs) and betaine-like lipids (BLLs) — represent distinct functional aspects of the *E. huxleyi*-EhV infection process (Vardi *et al.* 2009; Vardi *et al.* 2012; Fulton *et al.* 2014; Hunter *et al.* 2015). Host-specific GSLs (hGSLs) are both abundant and constitutively produced in

E. huxleyi cells, providing more taxonomic resolution than the 19'-hexanoyloxyfucoxanthin (19'-hex) pigment. Viral glycosphingolipids (vGSLs) are critical to and diagnostic of successful infection (Vardi *et al.* 2009; Ziv *et al.* 2016), triggering the production of reactive oxygen species, autophagy, PCD and eventual lysis of *E. huxleyi* cells (Bidle *et al.* 2007; Vardi *et al.* 2009; Schatz *et al.* 2014). EhV infection also specifically triggers an exponential increase in BLL(22:6/22:6), which becomes the major BLL species in infected cells (Hunter *et al.* 2015). At the same time, sialic acid GSLs (sGSLs) provide a diagnostic of population susceptibility to EhV infection, having only been detected in susceptible *E. huxleyi* strains (Fulton *et al.* 2014).

The manner in which phytoplankton die ultimately determines the flow of photosynthetically-fixed organic matter (and associated elements) and relative coupling through three main ecosystem pathways — transfer to higher trophic levels via grazing, recycling through the microbial loop, and downward export via sinking — with each having distinct biogeochemical consequences (Bidle 2015). Pairing the aforementioned diagnostic biomarkers with high resolution *in situ* measurements from observational sensing platforms stands to elucidate broader ecosystem and biogeochemical consequences of infection. The impact of virus infection on optical backscatter (B_b), chlorophyll *a* (Chl *a*), dissolved oxygen (DO), and chromophoric dissolved organic matter (CDOM) remain unexplored as possible signals and indicators of acute viral infection events. EhV infection is thought to alter B_b during infection as the host sheds and produces free coccoliths, while leaking and lysing cells may release dissolved organic matter (DOM) as respiratory substrates and components of CDOM (Castillo *et al.* 2010). At the same time, EhV infection enhances cellular production of transparent

exopolymer particles (TEP), sticky, particle-aggregating polysaccharide matrices (Passow & Alldredge 1999; Vardi *et al.* 2012), and can trigger preferential grazing responses (Evans & Wilson 2008). The particle production associated with these pathways (aggregates and fecal pellets from grazers) together with ballasting provided by $CaCO_3$ coccoliths (ρ = 2.8 g cm^{-3}) provide mechanisms (Briggs *et al.* 2011; Frada *et al.* 2014) by which viruses may contribute to downward export instead of fueling the futile carbon cycle through rapid upper water column remineralization via the microbial loop (Azam 1998).

We tested the hypothesis that coccolithovirus infection of *E. huxleyi* facilitates downward vertical flux of both particulate organic carbon (POC) and particulate inorganic carbon (PIC) during the *North Atlantic Viral Infection of Coccolithophores Expedition* [*NA-VICE*; see Methods; (Figure 2.1). Host-virus interactions and associated ecosystem dynamics were interrogated over 4-5 d using a Lagrangian mode sampling strategy within and below the mixed layer (down to 300 m) and through a combination of water samples obtained via traditional CTD casts and deployments of both surface-tethered sediment traps and vertically migrating *in situ* optical profilers (see Methods). The timing of occupation and field sampling strategy were specifically designed to take advantage of the well-documented ~4 d infection window and to interrogate its potential impact on ecosystem dynamics and potential export of biomineral rich particles with high sinking rates (Bidle *et al.* 2007; Vardi *et al.* 2009; Vardi *et al.* 2012). Our post-cruise analyses identified three physically and biologically distinct water features containing *E. huxleyi* blooms at different phases of EhV infection —early infection (EI), late infection (LI), and post infection (PI) (see Methods and details below for how these phases were

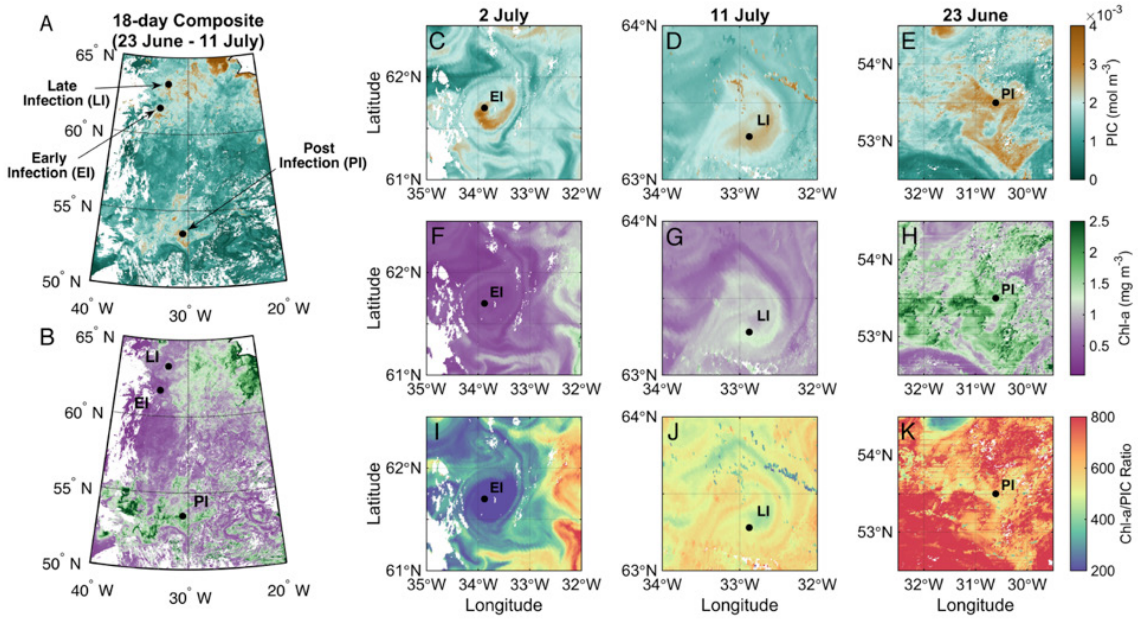


Figure 2.1. AQUA/MODIS 1 km resolution derived near-surface Particulate Inorganic Carbon (PIC) and Chlorophyll *a* (Chl *a*) concentration images of water mass features containing *E. huxleyi* blooms in the North Atlantic. (a-b) Location and overview of PIC (a) and Chl *a* (b) at water masses sampled in the North Atlantic (Early Infection, EI; Late Infection, LI; Post Infection, PI). EI populations were occupied from 30 June -5 July and again from 7-11 July as EI_R. LI populations were occupied on 6 July and 12 July, with regular profiling float measurements taken continuously between occupations. PI populations were occupied from 23-27 June. Note that PIC and Chl *a* scales are on the far right. (c-e) Satellite PIC measurements showed elevated PIC at each station during occupation. (f-h) In contrast, Chl *a* showed more variability between stations; it was lowest at EI and highest at PI. (i-k) The resulting Chl *a*:PIC ratios associated with resident populations were quite different at each water mass (low ratios at EI, intermediate at LI, and high at PI), consistent with differences in physiological state, phytoplankton composition, or photoacclimation. For a-b, $1km^2$ pixels are from max 18 day average, min 1 day depending on cloud cover. For c-k, $1km^2$ pixels are from single day observations. One observation per day.

diagnosed). EI and LI populations were within distinct anticyclonic mesoscale eddy features with sampling efforts following the clockwise rotation pattern; the PI water mass coherently migrated south during sampling and began mixing with a fresher water mass at the end of our occupation (Supplementary Figures 2.1 and 2.2). A second visit to EI (EI_R), two days later, followed a more central trajectory within the eddy.

2.3 RESULTS AND DISCUSSION

Remote sensing analysis of these three features showed elevated PIC signals indicative of coccolithophore blooms, but the satellite derived Chl *a* concentration varied widely; Chl *a* was lowest at EI, intermediate at LI, and relatively high at PI (Figure 2.1). This made for disparate Chl *a*:PIC signature ratios (i.e., low at EI, intermediate at LI and high at PI), features consistent with differences in physiological state, phytoplankton composition, and/or photoacclimation. Meanwhile, *E. huxleyi* cell densities ranged from ~1-4 x 10^3 cells ml^{-1} at EI and EI_R, similar to other large scale *E. huxleyi* blooms reported in the North Atlantic and elsewhere (Supplementary Figure 2.3). Depth integrated *E. huxleyi* cell abundances to 150 m revealed larger cell inventories at EI and EI_R (7×10^5 and 8×10^5 cells m^{-2}) than LI and PI (5×10^5 and 2×10^5 cells m^{-2}, respectively).

We deployed vertical profiling floats equipped with BioOptical Sensor System (BOSS)-mini packages (see Methods) to characterize the dynamic *in situ* biological and chemical signals associated with different stages of EhV infection. Observations spanning the euphotic zone depths generally affirmed the satellite Chl *a* data, revealing relatively low Chl *a* at EI and EI_R deployments, increasing through the deployment at LI and highest at PI (Figure 2.2a-d). B_b measurements showed an opposite trend (Figure 2.2e-h), influenced by the abundances of coccolithophores and shed coccoliths (Supplementary Figures 2.4 and 2.5); it was highest at EI and EI_R, decreasing rapidly during LI, and was relatively low at PI. Notably, respective *in situ* Chl *a*:B_b ratios also corroborated and complemented the satellite-derived Chl *a*:PIC with a ~4-fold difference in Chl *a*:B_b between the EI and PI states (Figure 2.2i-l). We suggest that the lower Chl *a*:B_b ratios associated with EI/EI_R communities were due to the abundance of

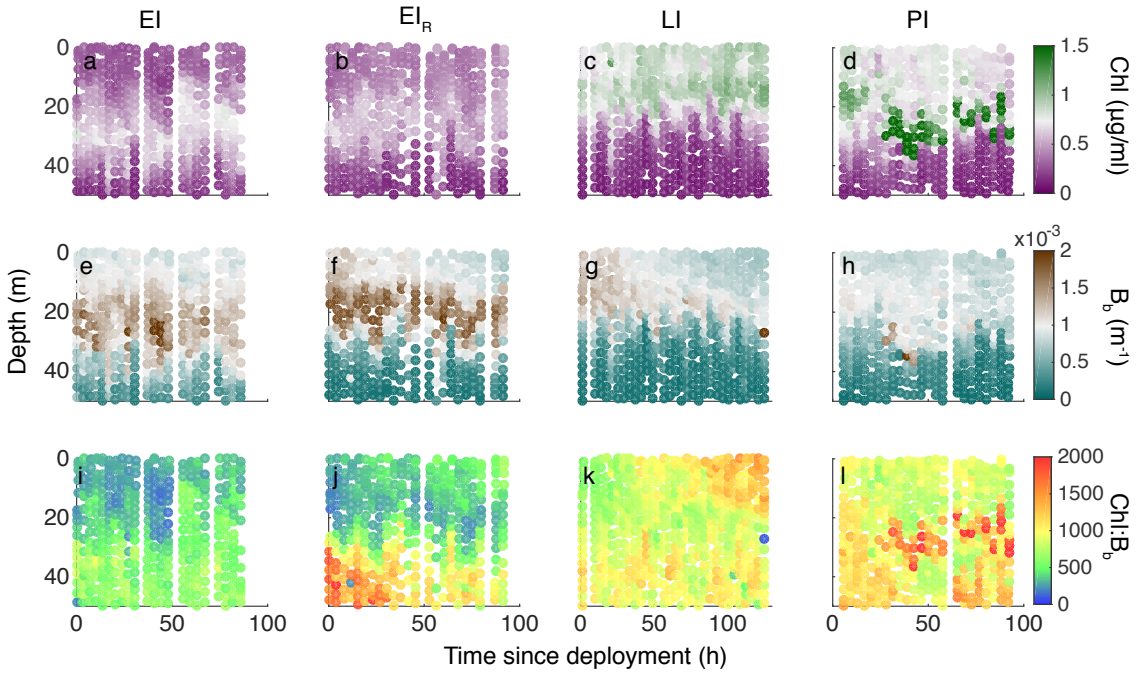


Figure 2.2. Natural populations at different stages of EhV infection have distinct *in situ* optical properties. (a-d) Relatively low Chlorophyll (Chl) fluorescence is observed in the water column at EI/EI$_R$ compared to the rapidly increasing Chl at LI and high signal observed at PI. (e-h) Optical backscatter (B$_b$) (700nm) observes the opposite trend between stations, highest at EI/EI$_R$, rapidly decreasing over the observation period at LI, and low at PI. (i-l) These patterns drive an increasing Chl:B$_b$ signal from early infected blooms (EI) to post bloom (PI) populations, similar to satellite Chl *a*:PIC observations. For each data point, $n = 1$ technical replicate.

coccolithophores and that this is a characteristic signature of an intense *E. huxleyi* bloom. Even though pigment data suggest that *E. huxleyi* and closely related haptophytes contributed 50-60% of the total Chl *a* signal on average (Supplementary Figure 2.6), the high scattering characteristics of coccoliths provided a more robust influence on the Chl *a*:B$_b$ ratio than the other eukaryotic phytoplankton groups or the numerically dominant picocyanobacteria (Supplementary Figure 2.4).

Depth-integrated measurements of different GSL parameters – sGSL and vGSL inventories, and vGSL:hGSL and vGSL:sGSL ratios - revealed that *E. huxleyi* were at different stages of infection (Figure 2.3a-d). These diagnostics are grounded by fine-scale observations of the parameters at high temporal resolution in well-characterized events of

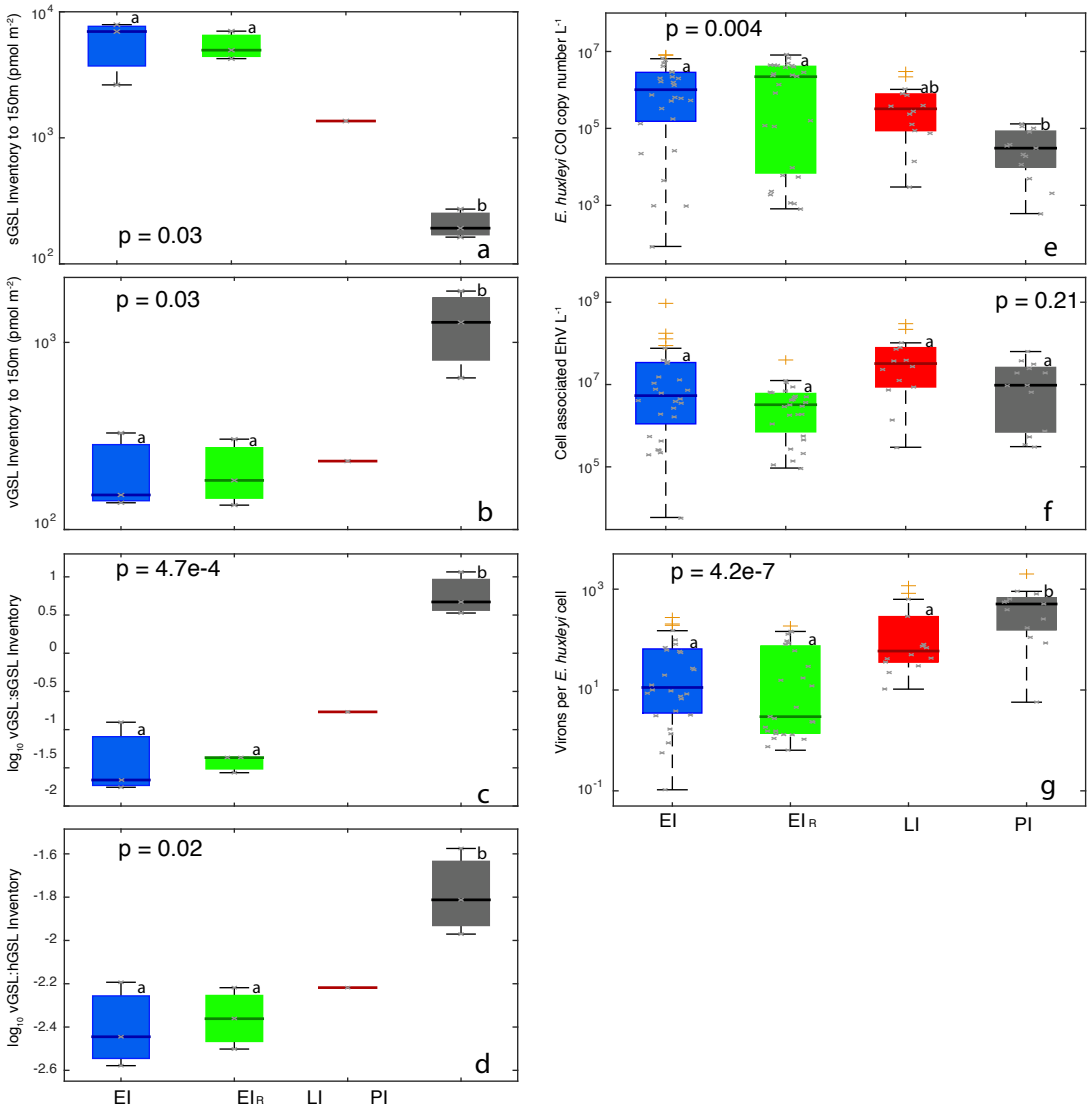


Figure 2.3. Diagnosis of virus infection using lipid- and DNA-based biomolecular proxies. Box and whisker plots of (a) sialic acid glycosphingolipid (sGSL) and (b) viral glycosphingolipid (vGSL) inventories [down to 150m; EI (blue), EI_R (green), LI (red), and PI (grey)] and (c) vGSL:sGSL and (d) vGSL:host glycosphingolipid (hGSL) inventory ratios at each station revealed distinct stages of infection. (e) Corresponding cytochrome oxidase- (COI-) based qPCR results revealed that *E. huxleyi* abundance was higher for EI/EI_R than in PI, while (f) Major capsid protein (MCP) gene copy number representing coccolithovirus (EhV) abundance was similar across stations. (g) EhVs per *E. huxleyi* cell, expressed as the ratio of MCP:COI gene copy numbers, were highest at PI. Importantly, our GSL and qPCR analyses were based on independent macromolecular pools and were extracted from biomass collected onto 0.8 µm pore-size filters, thereby reflecting EhVs associated with (and replicating within) *E. huxleyi* cells. For all plots, upper and lower box bounds denote 25% and 75% quantiles around the median (thick line). Vertical capped lines indicate max and min data values. Outliers indicated with (+). Letters above boxes denote statistically different groups based on ANOVA. For a-d: EI, n = 3; EIR, n = 3; LI, n = 1; PI, n = 3 station replicates; LI excluded from statistical analysis, given lack of replicate inventories. Gray x indicate individual data points. For e-g: EI, n = 28; EIR, n = 28; LI, n = 14; PI, n = 13 station replicates.

EhV-induced bloom termination in both laboratory and mesocosm settings (Vardi *et al.* 2009; Vardi *et al.* 2012; Fulton *et al.* 2014). Both EI and EI_R exhibited low vGSL:sGSL and vGSL:hGSL concentration ratios, indicating a resident *E. huxleyi* population that was highly sensitive (represented by high sGSL inventories) and undergoing early infection (as indicated by the low yet detectable inventories of vGSLs). PI had a significantly lower concentration of sGSLs in the upper 150 m compared to EI/EI_R, indicating a greatly reduced population of sensitive cells. It also had the highest vGSL:sGSL and vGSL:hGSL ratios, consistent with a sensitive *E. huxleyi* population that had been terminated by a mature EhV infection. The LI population had intermediate hGSL, sGSL and vGSL characteristics, consistent with cells transitioning between infected to post-infection stages.

Independent, DNA-based, qPCR analysis corroborated the above lipid-based observations of EhV infection (Figure 2.3e-g). COI-based *E. huxleyi* numbers ranged from 9×10^1 to 8×10^6 cells L^{-1} while MCP-based EhV abundances associated with *E. huxleyi* cells ranged from 6×10^3 to 9×10^8 virions L^{-1}. The highest *E. huxleyi* COI copy numbers were at EI and EI_R (Figure 2.3e), consistent with the high *E. huxleyi* abundances measured via flow cytometry. CHEMTAX pigment analyses corroborated that the contribution of *E. huxleyi* to total phytoplankton Chl *a* biomass was higher during the EI stages (20%) than LI (12%) and PI (15%), while dinoflagellates, prasinophytes, chlorophytes and diatoms displayed increasing contribution to the communities at LI and PI (Supplementary Figure 2.6). Total cell-associated, MCP copy numbers were similar across stations (Figure 2.3f), which appeared to be largely driven by host biomass. MCP:COI ratios were highest in the PI population indicative of more infected cells, with

relatively low COI-based host abundances and a more advanced stage of lytic infection (Figure 2.3G). *E. huxleyi* cell mortality, taken as operationally dead cells with compromised membranes, was greatly elevated at PI (85%) and LI (66%), compared to EI (42%) and EI_R (33%), further supporting discrete stages of infection (Supplementary Figure 2.7).

Sediment trap measurements in conjunction with a 'spike' transformation of the profiling float B_b profiles (Briggs *et al.* 2011) quantified particle flux dynamics via the removal of POC and PIC from the surface ocean in relation to EhV infection signatures. Elevated POC and PIC fluxes (maxima of 109 and 249 mg C m^{-2} d^{-1} at 50m, respectively) corresponded to the EI and EI_R populations during *NA-VICE* (Collins *et al.* 2015), far exceeding flux rates at PI (59 and 7 mg C m^{-2} d^{-1}, respectively) and other stations without *E. huxleyi* blooms. The B_b spike transformation analysis, previously used to discern sinking aggregated phytoplankton and detrital matter (Briggs *et al.* 2011), corroborated sediment trap flux measurements (Figure 2.4a; Supplementary Figure 2.8). Higher spike signals were observed below 50 m at EI and EI_R compared to LI and PI, consistent with greater aggregate sinking during early infection and dissipating following virus-induced termination. Notably, the EI and EI_R populations were associated with some of the lowest recorded POC:PIC export ratios (2.3 and 4.3 at 50m, respectively) in the North Atlantic, driven by the extremely high fluxes of PIC (Collins *et al.* 2015).

Vertical inventories of DO measured from *in situ* optical profilers provided a platform to investigate the biogeochemical fate and rates of carbon remineralization (C_{remin}) of the POC exported below the mixed layer (>50 m)(see Methods). We observed distinct oxygen minimum layers below the mixed layer at each station, which were

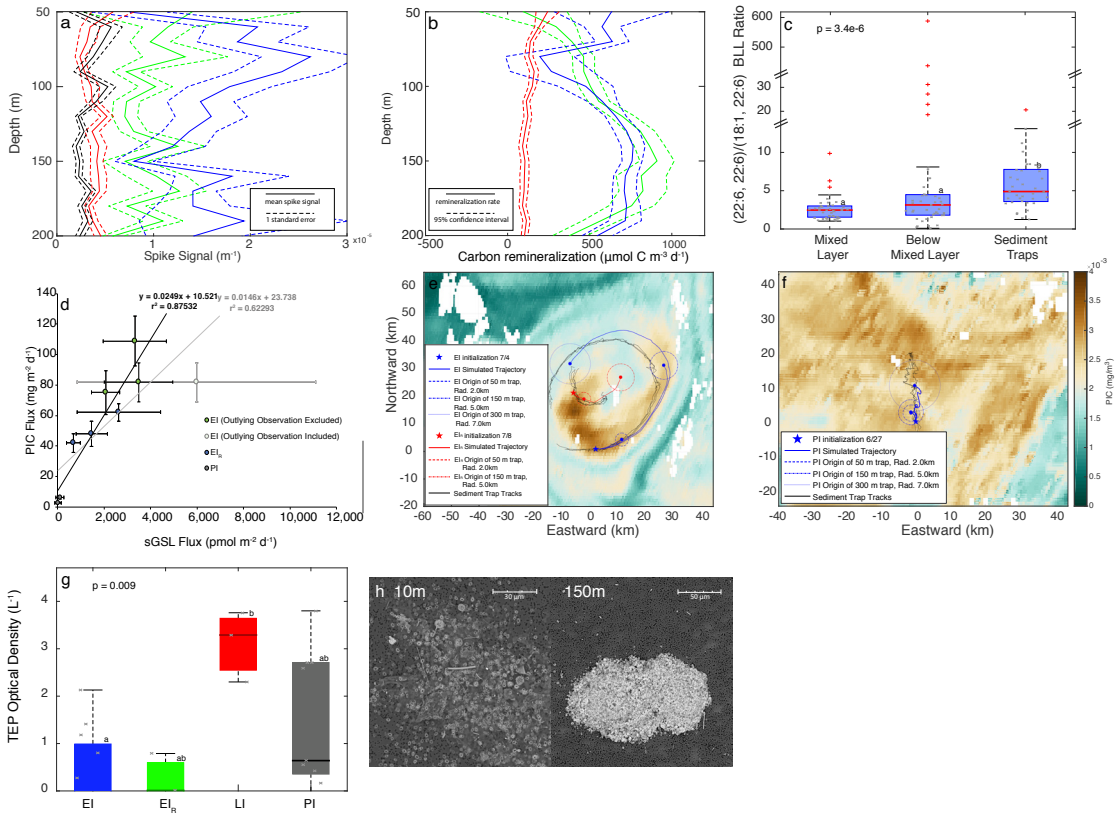


Figure 2.4. Active EhV infection triggers TEP-facilitated aggregation, POC export, and remineralization in the mesopelagic. (a) B_b spike signals revealed elevated particle aggregates at EI/EI_R (blue/green) below the euphotic zone compared to LI and PI (black and red). Centroids indicate the mean spike signal for 10m bins over an entire float deployment (dotted lines indicate ± 1 SE). (b) Particle dynamics at EI/EI_R were accompanied by high carbon remineralization rates in the mesopelagic. Dotted lines indicate 95% confidence intervals from linear regression. (c-d) Susceptible, EhV-infected cells were preferentially exported. (c) BLL (22:6/ 22:6)/(18:1/ 22:6) ratios for suspended biomass in the mixed layer (down to 30m), below the mixed layer, and in sediment traps (positioned at 50m, 150m, and 300m). Kruskal-Wallis ANOVA showed significantly higher ratios in sediment trap material. (d) Elevated PIC sinking flux was accompanied by increased sGSL sinking flux in the same sediment trap material. (e-f) Modeled geographic origins of sinking particles. MODIS/AQUA PIC images on 3 July (e) and 23 June (f). Tracks of deployed sediment traps shown in black for the EI/EI_R (e) and PI (f) deployments. The EI/EI_R (EI, blue star; EI_R, red star) and PI simulations (blue star) are within 10 km and 5 km of sediment trap trajectories, respectively. Dispersion simulations indicate that 99% of particles from the 50 m, 150 m and 300 m traps originated in circles of radius 2 km, 5 km, and 7 km, respectively. (g) TEP concentrations were low in the upper 50m at EI/EI_R and highest at LI as determined by ANOVA. (h) SEM images from EI at 10m and 150m respectively showed loose aggregates of cells/coccoliths compared to densely aggregated particles entrapped in a polysaccharide-like substance. (c,g) box bounds denote 25% and 75% quantiles around the median (thick line). Vertical capped lines indicate max and min data values. Outliers indicated with (+). For (c) n = 41, 43, 37 environmental replicates, respectively. For (d), error bars signify 1 SD around the mean (centroid). For e-f, satellite overlay pixels are n = 1. For (g): EI, n = 12; EI_R, n = 3; LI, n = 3; PI, n = 9 station replicates. Negative optical density measurements not illustrated. For (c,g), letters above boxes denote statistically different groups. Images in (h) are singular qualitative descriptors. Scale bars are included for size reference.

accompanied by steady decreases in sub-mixed layer water column DO concentrations for EI, EI_R, and LI populations (Supplementary Figure 2.9). EI and EI_R had up to nine-fold higher C_{remin} than LI with C_{remin} maxima in the mesopelagic between 125-200m (Figure 2.4b) and similar depth-integrated results comparable to direct observations (Supplementary Table 2.1). C_{remin} calculations at PI were not possible due to mixing of water masses. Elevated CDOM did not coincide with the high C_{remin} rates (Supplementary Figure 2.10), arguing against it supporting the respiration in the water column; other forms of DOM were not accounted for in our measurements. Our findings link the enhanced flux rates at EI and EI_R with greatly elevated C_{remin} due to the respiration of transient sinking particles in the mesopelagic (Collins *et al.* 2015).

Lipidomics of sinking particles collected in net traps directly supported EhV infection as a contributing mechanism to enhanced export. We observed significantly higher BLL(22:6/ 22:6) to BLL(18:1/ 22:6) ratios (Hunter *et al.* 2015) in sediment trap material than in water column samples from both above and below the mixed layer (Figure 2.4c; Supplementary Figure 2.11), indicating a higher proportion of infected cells in sinking particles. No significant difference was observed for water column samples in or below the mixed layer. Statistically significant correlations between sGSLs and PIC flux confirmed that susceptible cells dominated the coccolith/coccolithophore export (Figure 2.4d; Supplementary Table 2.2); PIC fluxes did not correlate with hGSLs, a marker for all *E. huxleyi* cells. The disproportional enrichment of infected cells in exported material provides strong evidence for EhV-induced carbon flux. Notably, downwelling-induced vertical transport did not appear to account for observed fluxes of infected particles (Denman & Gargett 1995), as it would serve to homogenize rather than decouple the

observed lipid signals. Indeed, downwelling-driven Chl injection into deep water was not observed in our data. In fact, expression profiling of genes upregulated during infection also showed that active infection at EI and EI_r occurred below the pycnocline with the degree of infection correlating with depth (Sheyn *et al.* 2017).

We quantitatively addressed the geographic source of trap material using the available ADCP data from the *R/V Knorr* along with the transit trajectories of the surface tethered traps (at 50, 150 and 300 m; see Methods). Lagrangian dispersion simulations indicated that the footprint containing 99% of particles from the 50 m, 150 m and 300 m traps originated in circles of radius 2 km, 5 km, and 7 km, respectively, at EI, EI_R and LI (Figure 2.4e-f). This demonstrates the source material for particles transiting to sediment traps was likely derived from the infected bloom above and within the local water mass boundaries, as opposed to lateral advection from adjacent water masses. Further, the respective BLL and sGSL:hGSL ratios were statistically indistinguishable with trap depth, indicating that the dominant signal derived from the same, rapidly sinking source material rather than a separate population that had been infected prior to our occupation.

While we recognize that sinking oceanic particles can be complex and can derive from various sources, one mechanism that could induce vertical transport and export of infected cells is TEP-based particle aggregation. With a buoyant density of 0.70 to 0.84 g cm^{-3}, TEP can lead to 'ascending particles' (Azetsu-Scott & Passow 2004) and enhance retention of particulate carbon in the upper ocean. However, association with denser cellular material such as $CaCO_3$-containing coccoliths may form large, dense aggregates with higher sinking rates, as has been observed previously with dinoflagellate thecal plates, where sinking rates increased exponentially with aggregate size (Alldredge *et al.*

1998). Lab-based experimental data indeed demonstrated a mechanistic connection between viral infection and TEP production and aggregation, whereby enhanced TEP production and aggregate formation was observed during the early phase of EhV infection of a calcified, *E. huxleyi* strain, prior to cell lysis (Supplementary Figure 2.12). Mixed layer TEP concentrations were significantly higher for LI compared to EI populations, corroborating previous studies showing TEP accumulation as characteristic of the late-stages of infection (Vardi *et al.* 2012) (Figure 2.4g). While *in situ* TEP concentrations were low at EI and EI_R, deck-board bottle incubations of these infected populations revealed enhanced TEP production over a three-day period (Supplementary Table 1.3), suggesting that, in the absence of export, these populations were producing elevated TEP. *In situ*, large (>200um), densely-packed aggregates of coccoliths and calcified cells enveloped in polysaccharide-like film were observed for EI at 150m with scanning electron microscopy (SEM), while looser aggregates dominated the particle field for samples obtained at shallower depths in the mixed layer (upper ~20 m) (Figure 2.4h). Taken together, these data support the notion that EhV infection of EI populations led to enhanced TEP production and larger particle aggregation, thereby facilitating sinking and carbon export. The LI and PI stations, on the other hand, were characterized by material that remained in the mixed layer. Late stage TEP accumulation may be due to limited ballasting from remaining liths and calcified cells. We acknowledge that other cohabitating algae may have also contributed TEP in these blooms (Alldredge *et al.* 1998), but our data collectively suggest infected *E. huxleyi* cells were a strong contributor at EI and EI_R.

The coupling of infection with grazing and fecal pellet production may have concurrently contributed to the particle composition and enhanced export fluxes observed at EI and EI_R. High *E. huxleyi* specific microzooplankton grazing rates of 0.34 d^{-1} and 0.90 d^{-1} were measured for infected cell populations at EI and EI_R (Chl *a* based grazing rates were 0.49 d^{-1} and 0.13 d^{-1}; Supplementary Figure 2.13), indicating that grazing was indeed an important contributor to *E. huxleyi* mortality in the upper mixed layer, with comparable rates to previous studies in the North Atlantic (Morison & Menden-Deuer 2015). Ecosystem interactions between grazing and infection are yet unresolved, as infection can manifest in enhanced (Evans & Wilson 2008) or deterred grazing (Vermont *et al.* 2016). The aforementioned infection-induced aggregation would serve to increase accessibility of small *E. huxleyi* cells (5 μm) to grazing by large, macrozooplankton (Passow & Alldredge 1999) and serve to further facilitate flux through fecal pellet production with sinking rates of >100 m d^{-1} (Harris 1994). Taken together with our previous observations of infected *E. huxleyi* cells (via EhV-derived MCP genes) within the gut content of >80% of copepods examined at EI and PI (Frada *et al.* 2014), these findings are supportive of intense grazing on and preferential export of infected cells by both micro- and macro-zooplankton during NA-VICE. To date, this interactive coupling between infection and grazing ecosystem pathways and their collaborative impact on carbon export has been largely unexplored, having traditionally been seen as distinct mortality and removal mechanisms.

We found no evidence that nutrient limitation was a contributing factor to mortality, TEP production, aggregation, or enhanced export at these stations. Nutrient additions (N, P and Si at 6.0, 0.38 and 6.0 μM, respectively) to microzooplankton

dilution experiments yielded statistically indistinguishable changes in cell abundance relative to non-nutrient addition incubations (Supplementary Figure 2.13). We also found only very weak relationships between the photochemical quantum yield of PSII (F_v/F_m) and nutrient concentrations (Supplementary Figure 2.14), arguing against any significant impact of nutrient stress on photophysiology.

2.4 CONCLUSIONS

Our findings provide multi-layered evidence that active coccolithovirus infection of *E. huxleyi* stimulates coupled TEP production, particle aggregation and high zooplankton grazing within mesoscale, open-ocean blooms in the North Atlantic and greater downward vertical fluxes and carbon export. Further, the integration of our biomolecular, *in situ* optics, and satellite hindcasting uniquely contextualized these findings across three distinct bloom snapshots (with separate Chl and PIC histories) and layered them into a conceptual model of an entire *E. huxleyi* bloom and EhV-induced bust cycle from early-June to mid-July 2012 (Lehahn *et al.* 2014) (Figure 2.4; Supplementary Figure 2.5; Supplementary Figure 2.15). The viral shunt observed in bacterioplankton and phytoplankton systems (Weitz *et al.* 2015) provides a conduit whereby carbon is remineralized in the surface ocean, but infection of ballasted, biomineral-containing phytoplankton (like *E. huxleyi*) appears to stimulate vertical export flux, an idea consistent with recent observations that variability in carbon export to 150 m is best explained by viral gene abundance (Guidi *et al.* 2016). The widespread distribution of *E. huxleyi* and the fact that investigations regularly find EhV production associated with mesoscale bloom termination across the North Atlantic serves as important temporal and spatial context for our findings. Notably, similar investigations

have yet to be performed in the Pacific Ocean or the Southern Ocean calcite belt where large-scale *E. huxleyi* blooms occur. While the relative contributions of transport mechanism(s)—via sinking of TEP-induced aggregates and/or via a coupling with grazing and packaging of infected cells in fecal pellets—remain unresolved, our results highlight virus infection as a pervasive ecosystem process that can facilitate export fluxes into the mesopelagic and enhance biological pump efficiency. Our work reveals layers of carbon flux biogeochemistry the quantitative contributions of which will help the decades-long challenge of resolving ecosystem influences on carbon export.

2.5 MATERIALS AND METHODS

2.5.1 North Atlantic field sampling. The *North Atlantic Virus Infection of Coccolithophores Expedition* (*NA-VICE*; KN207-03; 13 June – 16 July 2012) traversed a ~2000 nautical mile transect from Ponta Delgada, Azores to Reykjavik, Iceland aboard the *R/V Knorr* during which time water and sediment samples were collected at 30 discrete station locations (see: http://www.bco-dmo.org/project/2136). Water was collected using 10-liter Niskin bottles mounted on a 24-position rosette equipped with a Seabird SBE conductivity-temperature-depth (CTD) profiler. Water from each of six depths (mostly within the upper mixed layer and extending down to 150 m) was filtered onto various membrane filters to collect host cell biomass and associated viruses for lipid and nucleic acid analyses (see below for details). All filters were snap-frozen in liquid nitrogen and stored at -80°C until processed. Analytical flow cytometry of water collected during CTD casts was used to enumerate both *E. huxleyi* cells and EhVs, as well as perform diagnostic staining analyses (see below for details). Surface-tethered net

and cylindrical sediment traps were simultaneously deployed (50, 150 and 300 m). An autonomous optical profiler was deployed and programmed to sample down to 300 m (see below for details).

This study focused on three multi-day (4 d) process stations (EI, EI_R, and PI) conducted during the cruise where Lagrangian sampling methods were employed. EI was re-occupied a second time for 4 d as EI-Redux (EI_R). The ship tracked coherent water features by maintaining position with deployed, drifting sediment traps. Profiling floats were also deployed for the duration of each Lagrangian observation. The fourth location (LI) was visited for one day with multiple CTD casts as well as the deployment of an optical profiling float that profiled the feature for five days. LI was sampled over 6 days using the optical profiling float with CTD cast measurements and water samples taken at deployment and recovery and remote sensing imagery taken from year days 150-200 (Lehahn *et al.* 2014). Measurements of LI were complemented by month-long satellite observations of the bloom, observing the entire bloom cycle from initiation to termination (Lehahn *et al.* 2014). Ship based sampling was also conducted when the vessel returned to LI to recover the float, but with overboard hose pumped sampling due to a damaged CTD rosette at the end of the cruise.

2.5.2 Satellite remote sensing. We used 1 km resolution Level-2 MODIS/AQUA ocean color imagery [Particulate Inorganic Carbon (PIC), near-surface chlorophyll *a* (Chl *a*), Remote Sensing Reflectance centered at 555 nm (Rrs 555), and Remote Sensing Reflectance centered at 547 nm (Rrs 547)] daily downloaded from the NASA Ocean Color Web (https://oceancolor.gsfc.nasa.gov/) and further processed (geographically re-

sampled, re-projected over the target areas and statistically analyzed) aboard *R/V Knorr* using the SeaDAS V6.x software. The standard operational algorithms included as part of the standard Level-2 and Level-3 OC suite products provided by NASA for each of the geophysical parameters were used (https://modis.gsfc.nasa.gov/data/dataprod/). The main aim was to locate *E. huxleyi* populations based on the available satellite information. These also included daily AVISO altimetry maps (http://www.aviso.altimetry.fr) with information on mean geostrophic currents. Areas of research included ocean cyclonic (upwelling in the center, high satellite-derived Chl *a* concentrations initially, low satellite-derived PIC, low Rrs 555 and low Rrs 547) and anticyclonic (downwelling in the center, low satellite-derived Chl *a* concentrations initially, high satellite-derived PIC, high Rrs 555 and high Rrs 547) structures in the water.

2.5.3 In Situ Optical Profiling. Two profiling floats equipped with BioOptical Sensor System(BOSS)-mini (Satlantic) (Float #8493 and 8502) were used on multiple deployment operations during the cruise and recovered after leaving each station. Floats were deployed at each of the process stations and programmed to profile from 300 m to the surface; actual profiling depth varied from 200-300m. Deployments lasted between 91 and 123 h with ~4 h profile cycles and ~2.5 m vertical resolution. Deployments were accompanied by shipboard sampling and sediment traps along the Lagrangian path, except at LI where the deployment was only anchored by ship sampling at the beginning and end of the deployment. Floats were equipped with an ECO Triplet Puck [Chl Fluor, B_b (700 nm), CDOM], Seapoint turbidity meter, oxygen optode (AADI), multispectral radiometer [PAR, downward irradiance (E_d) E_d (412, 443, 490)], and CTD.

Fluorescence data was calibrated to ship-based Chl *a* measurements using night-time float profiles (between 2300 h and 0400 h) and adjacent CTD casts (μg l^{-1}). Daytime profiles exhibited significant non-photochemical quenching (NPQ) in the upper water column, which reduces the observed Chl *a* measurements. To account for observed NPQ, fluorescence (Fluo) and E_d(443) nm data from the top 20 m were binned in 5m intervals for individual profiles. Binned Fluo data was interpolated between night casts for predicted day cast Fluo values ($Fluo_p$) in the absence of NPQ. The ratio of Fluo:$Fluo_p$ was fit in a two-term exponential model against E_d(443). Individual Fluo data were NPQ corrected using the E_d(443) measurement from the associated data burst.

B_b was initially calibrated according to manufacturer's instructions. Dark offsets for B_b were then accounted for by subtracting the lowest value observed during individual deployments from the data. B_b transformation into the Spike signal was performed in accordance with Briggs et al. (Briggs *et al.* 2011). The spike transformation removes the background signal by running minimum and maximum filters along profiles, leaving only the anomalously high spikes in the signal. Similar B_b spikes have been interpreted as aggregated phytoplankton and detrital matter in the North Atlantic following an intense bloom (Briggs *et al.* 2011). B_b spikes were averaged in 10 m, 12 h bins. Spike changes over time were negligible so bins were modified to include all data at 10 m depth intervals over the duration of the deployments. Standard error around the mean of these bins was calculated from the individual data values within the bin, which conservatively contained at least 50 values.

Salinity and temperature corrections for DO were provided by the sensor operation manual. DO % saturation was calculated with SW_SAT02 MATLAB script

(Morgan 1998). Oxygen utilization measurements were performed similar to Martz et al. (Martz *et al.* 2008). Data were separated into 10 m depth bins between 50-250 m before calculating the linear slope of oxygen concentration vs. time over the deployment periods. Oxygen utilization was converted to carbon remineralization (C_{remin}) using the ratio $C_{org}:O_2 = 0.688$ (Anderson & Sarmiento 1994). Data above 50 m were excluded; this depth conservatively represented the bottom of the euphotic zone and was well below the mixed layer depth at each station (Supplemental Figure 2.6). PI was excluded from C_{remin} analysis because of float movement and mixing into another water mass, which influenced ΔDO, as observed by an abrupt change in depth horizon salinity >0.01ppt over the deployment. It was assumed that oxygen production and diffusion had a negligible influence on this analysis based on the constraints implemented and the relatively short timescales over which the analysis was performed. For comparison, C_{remin} was also determined along isopycnals to examine the possible influences of internal waves; this approach yielded similar results our depth-binned analysis (not shown).

CDOM measurements from the two floats were calibrated to each other by subtracting the difference between the floats' average measurements between 100-150 m when both floats were profiling the EI eddy. Average CDOM measurements were calculated for each profile between the pycnocline, identified as the depth interval with the greatest change in density, and 100 m over the duration of deployments to identify changes in CDOM concentration over time. CDOM accumulation rate profiles were achieved similar to the oxygen utilization measurements, separating data into 10 m depth bins for the deployment period before calculating the linear slope of CDOM concentration vs. time.

2.5.4 Sediment Trap deployments. Vertically sinking, particulate carbon fluxes were measured at 50, 150, and 300 m according to Collins et al. (Collins *et al.* 2015) and using surface-tethered cylindrical sediment traps [0.0125 m^2 cross-sectional area; materials and construction as described in McDonnell and Buesseler (McDonnell & Buesseler 2012)]. A mooring consisting of four traps at each depth, a surface buoy, wave-action mitigation bungee cord, and several floats, was deployed at each process station and allowed to drift for 3-5 days. The quasi-Lagrangian behavior of the mooring during each deployment was confirmed by comparison of positional data obtained from an Argos satellite beacon mounted on the surface buoy with shipboard Acoustic Doppler Current Profiler (ADCP) data from the *R/V Knorr*, which trailed the mooring at a range of 1-2 nautical miles. Traps were prepared, deployed, and recovered as described in McDonnell and Buesseler (McDonnell & Buesseler 2012). Traps were then sampled for particulate carbon in accordance with McDonnell and Buesseler (McDonnell & Buesseler 2012), except that the screened brine suspension (350 μm pore size to exclude macrozooplankton) was filtered onto a series of pre-combusted, 47 mm GF/F filters (0.7 μm nominal pore size). Field and analytical blanks were collected at each station. Filters were immediately frozen in liquid nitrogen and then stored at -80°C. Analysis of POC and PIC in the frozen samples was as described in Collins et al (Collins *et al.* 2015).

Particulate material was also collected for biomolecular and microscopic analysis from the appropriate depth using a series of surface-tethered, large-diameter net traps with a detachable 0.2 μm mesh cod end (Peterson *et al.* 2005), also at 50, 150, and 300 m. These traps were allowed to drift with a surface mooring for approximately 24 hours

in the same eddy feature as the corresponding cylindrical sediment traps. The exact deployment time was controlled by use of a remote acoustic release, which allowed us to close the traps prior to recovery. Upon recovery, the particle material in the cod end was homogenized by gentle shaking, then quantitatively split into fractions using an eight-way rotating electric splitter (Lamborg *et al.* 2008); aliquots were taken from one or more of these fractions, screened to 350 μm to exclude macrozooplankton, and dispensed quantitatively into replicate BOD bottles for incubation

2.5.5 Lipid analysis. Total lipid extracts were extracted from frozen biomass collected onto 0.2 μm pore-size, Durapore membrane filters (GVWP-type; Millipore) by a modified Bligh and Dyer extraction (Bligh & Dyer 1959; Popendorf *et al.* 2013). An internal standard of 2,4-dinitrophenylmodified phosphatidylethanolamine (DNP-PE) was added during extraction. Total lipid extracts were analyzed by normal phase high performance liquid chromatography – tandem mass spectrometry (HPLC-MS^2) using an Agilent 1200 HPLC coupled to a Thermo Scientific TSQ Vantage triple quadrupole MS. Chromatography and mass spectrometry conditions were as described by Popendorf et al. (Popendorf *et al.* 2013). Characteristic retention time and MS^2 fragmentation spectra were used to identify lipid classes, which were then quantified by peak area for a given MS^2 mass chromatogram as in previous studies (Vardi *et al.* 2012; Fulton *et al.* 2014).

For sGSL, hGSL and vGSL instrument response was accounted for by comparison to external standard response factor calibrations, acquired from a dilution series of a soy glucocerebroside (Avanti Polar Lipids, Alabaster, AL, USA) standard run immediately prior to analysis (Hunter *et al.* 2015).

Deviations from previous published methodology (Vardi *et al.* 2012; Fulton *et al.* 2014) were as follows:

For vGSL, selected reaction monitoring (SRM) mode was used for the 734.7 to 572.7 transition, representing the environmental vGSL lipid species as identified in Vardi et al. (2009) (Vardi *et al.* 2009). This SRM was utilized instead of the neutral loss or precursor ion scanning modes applied in previous investigations (Fulton *et al.* 2014; Hunter *et al.* 2015), due to its greater sensitivity and the low abundance of the target species. The vGSL SRM data was then corrected for instrument response based on a calibration curve from the 714.7 to 534.7 SRM transition, which represents the predominant component of the soy glucocerebroside standard. These transitions represent the neutral loss of 162 Da corresponding to a characteristic cleavage of the glycoside bond (Vardi *et al.* 2009).

Host hGSL was characterized based upon its characteristic long chain base derived fragment of 257 Da and diagnostic retention time (Vardi *et al.* 2012). Two chromatographic peaks in the retention time window were observed in the MS^2 mass chromatogram for a precursor ion scan of 257 Da (P257). The first eluted at 10 minutes in line with hGSL in cultured *E. huxleyi* (Hunter et al. 2015); the second eluted at 12 minutes, co-eluting with vGSL (Vardi *et al.* 2009; Hunter *et al.* 2015). The 10-minute pure hGSL peak was cross calibrated with its counterpart in the MS^2 mass chromatogram of a neutral loss of 180 Da (NL180). This response factor was used to quantify the hGSL in the 12-minute peak by converting the unambiguous observed P257 peak to the equivalent value in NL180. The NL180 equivalent could then be quantified relative to the standard response factor (unavailable for the P257 mass chromatograms as the standard does not undergo this fragmentation). All quantities were corrected for recovery based

upon the DNP-PE internal standard. BLL species were quantified in terms of relative abundance and are presented as an internally normalized ratio (Hunter *et al.* 2015).

2.5.6 Filtration, DNA extraction and subsequent Quantitative PCR Analysis. Water was first pre-filtered through a 200 μm mesh to eliminate large macrozooplankton. Cell biomass (and associated viruses) was then collected from ~3-5 liters of seawater by filtration onto large (142 mm diameter; 0.8 μm pore-size PC filters (Millipore) to minimize clogging of the pores and the collection of extracellular 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 μm and 400 μm diameter beads) (OPS diagnostics, Libanon, 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 5 ml buffer was incubated for 1 hour at 50 °C in the presence of 100 μg Proteinase K, followed by standard phenol/chlorophorm/isoamylalcohol 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).

E. huxleyi and EhV genetic markers (COI vs. MCP) in the filtered POM were quantified by SYBR®Green-based qPCR assays using primers and PCR conditions after Coolen (Coolen 2011). Using the same DNA extraction method, genomic DNA was extracted from FACS-counted cells vs. particles of *E. huxley* strain 1216 and EhV strain 86, respectively. To calibrate the COI and MCP-specific qPCRs, ten-fold dilution standard series of this genomic DNA were subjected to qPCR alongside the samples.

2.5.7 Photophysiology. The photochemical quantum yield (F_v/F_m) of natural populations was measured using Fluorescence Induction and Relaxation (FIRe) instrument (Satlantic). Discrete samples were collected from Niskin bottles and measured on the FIRe immediately after a 10 min dark adaption (Kolber *et al.* 1998). Instrument settings in FireView 1.0.0 were as follows: Gain, 2400; Samples, 10; Sample Delay, 1000 ms; PAR, 0. Analysis in FirePro 1.3.3 was completed with 240 iterations. Because samples were collected at different times of day, we acknowledge results may be biased by diel periodicity of fluorescence quenching.

2.5.8 Staining and flow Cytometry. Phytoplankton cell abundance and diagnostic staining were performed with a Guava flow cytometer (Millipore). After water collection, samples were immediately prepared for analysis in duplicates. The instrument was set at medium flow rate (0.59 μl s^{-1}). Triggering was set to collect all events above chlorophyll fluorescence (692 nm) threshold value of 1, with minimum gain for all

detectors except yellow fluorescence, set at 23.63. Samples were analyzed with GuavaSoft InCyte 2.2.2. Synechococcus was gaited using red (threshold immediately above noise) and yellow (threshold at 2×10^1 rfu) fluorescence with an oblique gate that encompassed the range of the population. Low fluorescence eukaryotes were gaited between red fluorescence 2×10^0 and 1×10^2 rfu and side scatter events above the axis, excluding events gaited as Synechococcus. High fluorescence eukaryotes were gaited from red fluorescence 1×10^2 and above as well as forward scatter events above the axis, excluding the *E. huxleyi* population.

Populations of *E. huxleyi* were identified by forward scatter (FSC), side scatter (SSC), and Chl fluorescence (692 nm) readings using gates based on properties of cultured strains. *E. huxleyi* were enumerated by first gating events between 4×10^1 - 5×10^2 FSC and $5 \times 10^2 - 1 \times 10^4$ SSC, with the elevated SSC from attached coccoliths being the primary characteristic separating this population from other high fluorescence eukaryotes. A second gate was placed setting a minimum red fluorescence threshold at 1×10^2 rfu.

For live/dead cell analysis, samples were stained with 5 μM SYTOX Green (1:1000 dilution of stock) in the dark for 10 min. The green fluorescence (520 nm) threshold for SYTOX positive cells was set at 10 fluorescence units, based on the lower boundary of fluorescence for positive control, heat treated samples. The average of the percent positive cells at each station, excluding samples collected at 150m, was determined for each water mass.

2.5.9 Float Optics and Phytoplankton Abundance Correlations. B_b- and Chl:B_b-flow cytometry correlations were conducted using CTD casts and float profiles where the ship and float were within 2km of each other during sampling/profiling. B_b values were an average of the three data point in the profile with depth most close to the CTD sampling depth compared to in the scatter plot. Phytoplankton were gaited according to the following criteria: *E. huxleyi* population was separated from other phytoplankton in a distinct event cluster based on FSC and high SSC and separated from non-fluorescent particles with a minimum Chl fluorescence threshold. High fluorescence eukaryotes included all events above a Chl fluorescence threshold of 100 rfu, where there was commonly a clear separation between high and low fluorescence populations. Synechococcus was gaited based on elevated fluorescence at 585 indicating presence of phycoerythrin. Low fluorescence eukaryotes had a Chl fluorescence below 100 rfu.

2.5.10 Grazing Dilution Experiments. We used a modified version of the standard dilution technique (Landry & Hassett 1982) to evaluate phytoplankton cell loss due to microzooplankton grazing, where encounter rates of predator-prey is inversely proportional to the dilution rate. Two dilution experiments were performed in near surface waters characterized by *E. huxleyi* populations in the Early Infection (EI) and Early Infection-redux (EI_R) stages (July 2 and July 11, 2017). Seawater was collected from approximately 10 m and the 0.45 μm were prepared for the dilution bottles that were then filled with 25, 50, 75 and 100% natural seawater. The natural seawater additions for the incubations were prescreened through a 200 μm Nitex mesh to remove large macrozooplankton grazers (e.g. copepods). Aqueous macronutrients (N,P and Si)

were added to all 2.7L polycarbonate dilution bottles at levels of approximately 6, 0.38 and 6 μM, respectively. The bottles were incubated at approximately 40% surface irradiance at sea surface temperature for approximately 24 hr. Chlorophyll specific growth rates were calculated relative to the initial chl values at time zero. Flow cytometry based cell counts were used to calculate cell specific growth rates for comparison as well as *E. huxleyi* specific growth rates, similar to (Evans *et al.* 2003).

2.5.11 Microscopy. SEM images of coccolithophores and particle aggregates were taken with a Phenom ProX desktop scanning electron microscope. Samples were collected on 0.4 μm pore-size, Isopore filters and immediately desiccated with silica gel until visualization on SEM. Filters were quartered then loaded directly to specimen mount for imaging. Images taken at 1300x and 2200x magnification with 2048 x 2176 resolution, 10 kV, 6 s exposure. For coccolith counts, concentrations were calculated using the average coccolith count per image field of view and accounting for volume filtered, field of view, and filtration area.

2.5.12 Nutrient analysis. Seawater was filtered through pre-combusted, GF/F filters and filtrate was stored at -80° C in acid-washed Falcon tubes until analysis. Analyses of orthophosphate, total inorganic nitrogen (nitrate and nitrite, ammonia) and silicate were performed on thawed samples at the Rutgers' Nutrient Analysis Facility in the Department of Environmental Sciences (http://nutrientlab.rutgers.edu/) using standard facility protocols and a Lachat QuickChem 8500 Nutrient Analyzer Flow Injection Analysis System (Lachat Instruments, a Hach Company Brand). Concentrations were

determined using Lachat Instruments' Methods using facility Standard Operating Protocols (SOPs; found at: ftp://boardwalk.marine.rutgers.edu/bidle/Protocols/). Detection limits for $NO_{2/3}$, NH_4, and P were 0.1 µM.

2.5.13 Pigment analysis. Standard Chl *a* extraction and fluorescence protocols were followed. Triplicate aliquots of 280 mL were filtered through 25 mm GF/F filters and extracted in 90% acetone for 24 h at -20°C. The filtrate was then analyzed on a Turner 10-AU fluorometer that had been calibrated with pure Chlorophyll a (Sigma Chemicals) using the non-acidification technique (Welschmeyer 1994). Water samples (2-4 L) were also collected from CTD casts and filtered onto 25 mm Whatman GF/F glass fiber filters and stored in liquid nitrogen until on-shore analysis. Following extraction in 90% acetone, samples were analyzed on an Agilent 1100 HPLC (High Performance Liquid Chromatography) system with diode array and fluorescence detection. The elution gradient utilized was a modified version of the Zapata et al. (Zapata *et al.* 2000) method. The exact elution gradient and protocols have been described in detail elsewhere (DiTullio & Geesey 2002). The HPLC pigment concentrations were then analyzed using the iterative matrix factorization program CHEMTAX (Mackey *et al.* 1996). Output values of % contribution phytoplankton taxa to Chl *a* were averaged for each of the four station. Averages represent samples collected between 0-50m.

2.5.14 TEP and protein analysis of field samples. Seawater was filtered onto a 0.4 µm pore-size, Durapore filter and stained with 1ml 0.02% Alcian Blue before freezing at -20°C. For analysis, filters were soaked for 2 h in 2 ml of 80% sulfuric acid before

removing the filter and measuring the optical density (OD; 787 nm) of extracted solution in a 1 cm cuvette on an Agilent 8453 spectrophotometer. ODs were calculated using the volume filtered for individual samples. Blank measurements were prepared with 0.2 µm-filtered seawater.

Protein concentration was measured on biomass from a known volume of filtered seawater (collected onto 0.8 µm pore-size filters; Millipore RAWG04700), which had been kept frozen at -80°C until time of analysis. Biomass was scrapped from each filter into LDS lysis buffer (140mM Tris base, 105 mM Tris HCl, 0.5 mM EDTA, 2% LDS, 10% glycerol) with protease inhibitor cocktail (Sigma, P2714), frozen in liquid nitrogen, and sonicated three times (Misonix; power setting of 2-3), followed by centrifuging (10,000x g, 10 min, 4°C). The supernatant was transferred to a fresh tube and protein concentration was quantified using BioRad DCTM protein assay kit (BioRad 5000111), following the manufacturer's protocol. In short, 25 µl of proprietary reagent A was mixed with 5µl of each sample in duplicate within microtiter plate wells followed by addition of 200µl proprietary reagent B. OD at 750 nm was read on a microtiter plate reader (Molecular Devices, SpectraMax M3) after a 15 min of incubation at 37°C. Known concentrations of bovine serum albumin were used as standards.

2.5.15 Host/virus abundances and TEP analysis in lab-based experiments. Experiments used *E. huxleyi* strain DHB607 in triplicate in f/2-Si media. *Emiliania huxleyi* virus 99B1 (EhV99B1) (Nissimov *et al.* 2012) was added to exponentially growing cultures at a virus:host ratio of 5:1, based on the total number of virus like particles (VLPs) as detected by flow cytometry (FC) in a primary EhV lysate stock. Virus

free controls were set by adding appropriate f/2-Si volume into exponentially growing cultures. Cultures were monitored daily for host and virus abundances, bulk TEP production, and particle aggregation over a 4 d period.

The total number of virus like particles (VLPs) throughout the infection experiments were counted using SYBR Gold stain and an Influx Mariner 209S flow cytometer (FC) based on the protocol developed by (Brussaard *et al.* 2010). Briefly, a 1 mL subsample of each virus stock or infected culture was fixed with glutaraldehyde (final concentration of 0.5%), incubated in the dark at 4°C for 15-30 min, snap frozen in liquid nitrogen and stored at -80°C until processed. Samples were thawed at room temperature and diluted 50-fold in Tris-EDTA (TE) buffer (1M TRIS pH-8, 0.5M EDTA, and MilliQ) in 3 ml flow cytometery tubes. SYBR Gold was diluted into the TE buffer at a 20,000:1 ratio of the commercial stock. The subsamples were agitated and incubated at 80°C for 10 min, followed by 5 min at room temperature in the dark. Each sample was analyzed for 1 min on the InFlux at a variable flow rate in μL min^{-1}. The flow rate was adjusted for each set of measurements. Host cell abundance measurements were performed using an Accuri C6 flow cytometer (BD Biosciences), where 200 μL of the thawed samples was distributed into 96 well plates and the cells were counted using chlorophyll red fluorescence (692 nm) *vs* side scatter (SSC).

TEP samples for OD quantification with AB stain were prepared as described above with the following exceptions: 3 mL of each experimental treatment was filtered daily in triplicates and stained with 500 μL AB. OD (787 nm) of the AB-stained TEP was measured on a Molecular Devices SpectraMax M3 plate reader with OD values corrected

for f/2-Si media blanks and plotted against a xanthum gum (XG) standard curve of known concentrations.

TEP particles (including the particle size spectrum) associated with infected and virus free cells were also visualized and characterized daily with a FlowCAM and VisualSpreadsheet© software (version 4.0.27). Prior to visualization, subsamples (500 µl) were stained with 200 µl of 0.02% AB and 20 µl of acetic acid at RT for 10 minutes and applied to the FlowCam sample intake. A 10X objective lens was used in the instrument's auto-image mode setting (20-30 frames per second) with the threshold set to image particles that ranged from 7-100 µm with a minimum distance between particles of 1 µm. Image collection of particles was restricted to those above 7 µm in diameter based on the known size range of *E. huxleyi* cells (i.e. 3-8 µm) and an upper size cutoff of 100 µm based on limitations of the size of the flow cell used. The flow rate was adjusted to 0.08-0.1 ml min^{-1} to prevent duplicate images and auto-image mode was set to stop after imaging 100 µl volume. Files were analyzed manually for quality control and identification of duplicate images. Particles were arbitrarily "binned" into 12 particle size groups that spanned 5 µm in diameter (i.e. 7-11, 12-16, 17-21, 22-26, 27-31, 32-36, 37-41, 42-46, 47-51, 52-56, 57-61, and 62-66 µm). Once particles were imaged they were sorted based on the average blue to green ratio of each image and their area based diameter (ABD), a FlowCAM image analysis setting to calculate particle diameter based on the measured area of the visualized particle (FlowCAM Manual, 2011).

2.5.16 Measurements of PIC and POC. Particulate matter from 0.5 – 2 liters of seawater was filtered in duplicate onto pre-combusted, GF/F filters under light vacuum

and snap frozen in liquid nitrogen. Filters were placed into numbered, etched glass petri dishes and baked at 60°C for drying. One filter, designated for POC, from each set was placed into a glass desiccator with concentrated HCl fumes for 24 h to dissolve PIC. The other filter was untreated and represented total carbon (TC). All filters were trimmed, packed into tin boats, and combusted in a CNS Elemental Analyzer (Carlo Erba NA 1500). PIC was determined by taking the difference between TC and acidified, POC sample. Due to instrument failure, 2 samples were excluded from analysis. For resolution of *E. huxleyi* DHB607 culture PIC quotas, exponentially growing cells (`5 x 10^5 cells ml^{-1}) were filtered onto triplicate GF/F filters and prepared for POC/PIC analysis as above.

2.5.17 General Statistics. Differences among stations were determined with one-way ANOVA in the MATLAB Statistics Toolbox for the following data sets: GSL inventories and ratios, MCP and COI gene copy number, TEP OD, *E. huxleyi* cell inventories, % live/dead *E. huxleyi,* dissolved nutrients, PIC, and Coccoliths. BLL lipid data was non-normal and therefor a Kruskal-Wallis test was applied. For all boxplots, outliers were greater than ± 2.7 standard deviations from the mean.

2.6.18 Sediment Trap Particle Origins. As the sampling of sediment traps is dependent on both vertical and horizontal physical processes, it is useful to simulate the trajectories of particles captured in sediment traps (Siegel *et al.* 2008) in order to determine whether particles originated in the region of interest, as opposed to adjacent water masses with differing biological processes. In order to simulate particle trajectories, we solve the following system of equations:

$$d(\vec{X}(t) + \vec{X}'(t))/dt = \vec{u}(t) + \vec{u}'(t) \quad (1)$$

Where $\vec{X}$ is the low-frequency trajectory of a simulated particle, $\vec{X}'$ is the high – frequency trajectory associated with horizontal dispersion, $\vec{u}$ is the low- frequency water velocity, and $\vec{u}'$ is the high-frequency velocity associated with dispersion. In order to solve these equations, we need to specify $\vec{u}(t)$, $\vec{u}'(t)$, and $\vec{X}_i = \vec{X}(t_i)$, where t_i is the initialization time.

The *R/V Knorr* was equipped with a 75 kHz Acoustic Doppler Current profiler (ADCP), continuously measuring current profiles to approximately 600 m depth with 8m resolution. The *R/V Knorr* was often on–station near the drifting sediment traps, providing sufficient current profile measurements to define the horizontal components of $\vec{u}(t)$ in most cases, the exception being EI_R (see below). The vertical component of $\vec{u}(t)$ is defined by the descent rate of particles that ultimately end up in the sediment trap. A value of 100 m d^{-1} was chosen to represent the average descent rate of particles (Harris 1994; Ploug *et al.* 2008) .

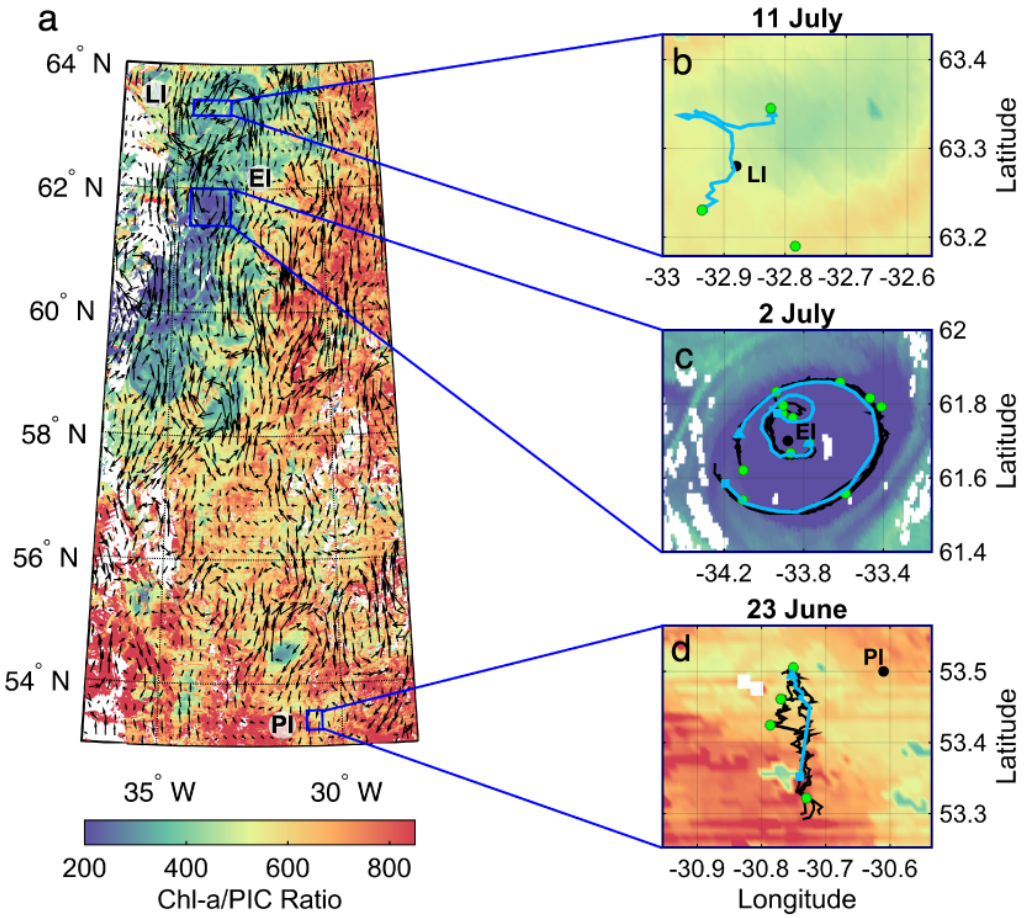
There were no direct measurements of $\vec{u}'(t)$. The horizontal component was modeled as a probability distribution function (PDF; i.e. a Gaussian distribution with a mean μ and standard deviation σ). The $\vec{u}'(t)$ PDF was then used in a Monte Carlo simulation (n=20000) to generate a set of trajectories representing the PDF of $\vec{X}'$ associated with dispersion. The mean (μ) is associated with the low-frequency velocities already obtained from ADCP data and thus assumed to be zero. The standard deviation

(σ) is associated with unresolved velocity scales and drifter motions not associated with the background velocity measured by the ADCP. We estimate this by comparing velocities measured by the ADCP and velocities derived from the drifting sediment traps. We assume the standard deviation of this velocity difference to be σ, and is calculated to be ~20cm/s. Note that this value includes any error in the ADCP measurements and the drifter locations so is an upper bound on σ, assuming these errors are uncorrelated with $\vec{u}'(t)$, which is likely.

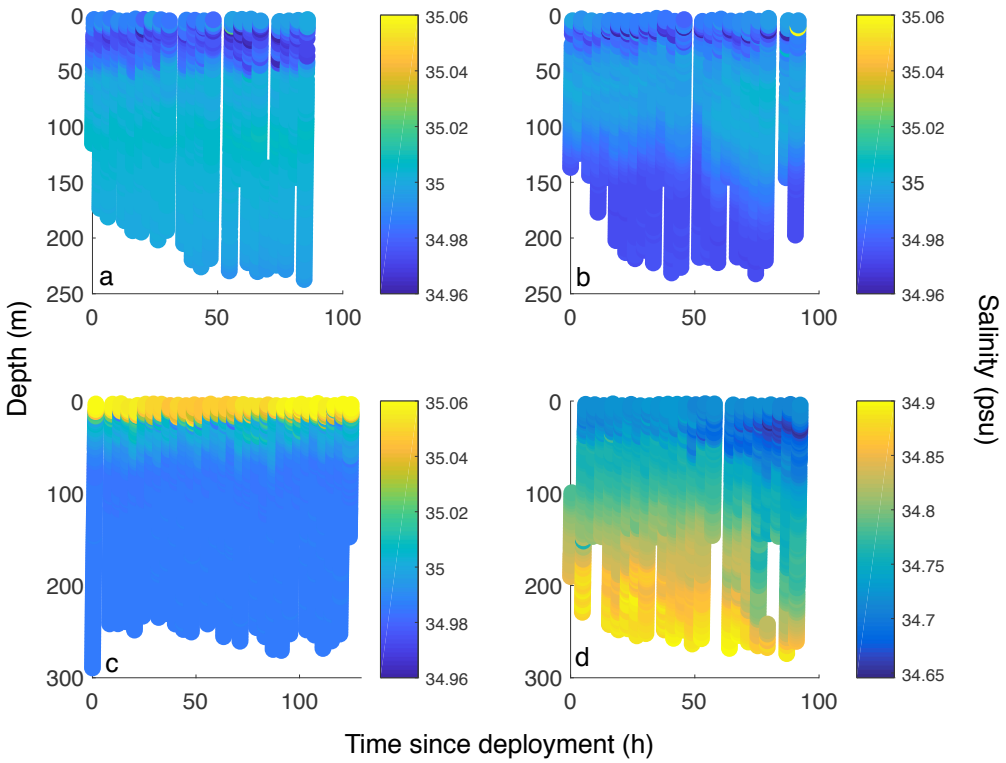
Since the objective was to determine the potential origin of particles in the sediment trap, it was desirable to define $\vec{X}_i$ and t_i as the location and time where traps were recovered for processing. We chose a representative trap recovery time which allows sufficient ADCP data to perform the simulations. This was possible for EI and PI, however, EI_R is a shorter record with more gaps, so the $\vec{X}_i = \vec{X}(t_i)$ chosen allows for the longest simulation time. Initialization depths where 50m, 150m, and 300m, representing the depths of the drifting sediment traps.

Three simulations were carried out for EI and PI, one for each initialization depth. Simulations were ended after 7.2 h, 31.2 h, and 64.8 h, the time it took for particles to reach 20 m depth (the depth corresponding to E. huxleyi maximum abundance) from 50 m, 150 m, and 300 m, respectively. The available ADCP data only allowed for 2 simulations during the EI_R time period, one each for the 50 m and 150 m traps.

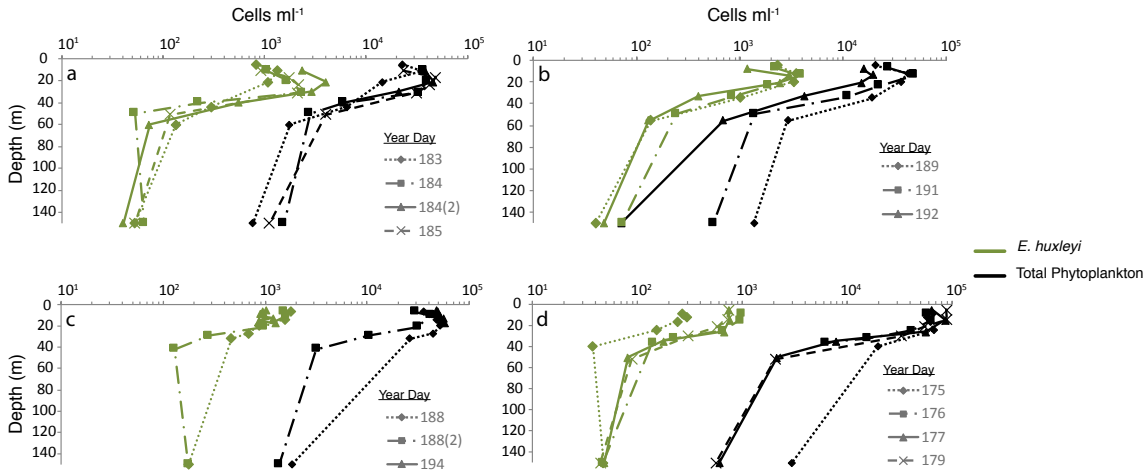
2.6 SUPPLEMENTAL FIGURES



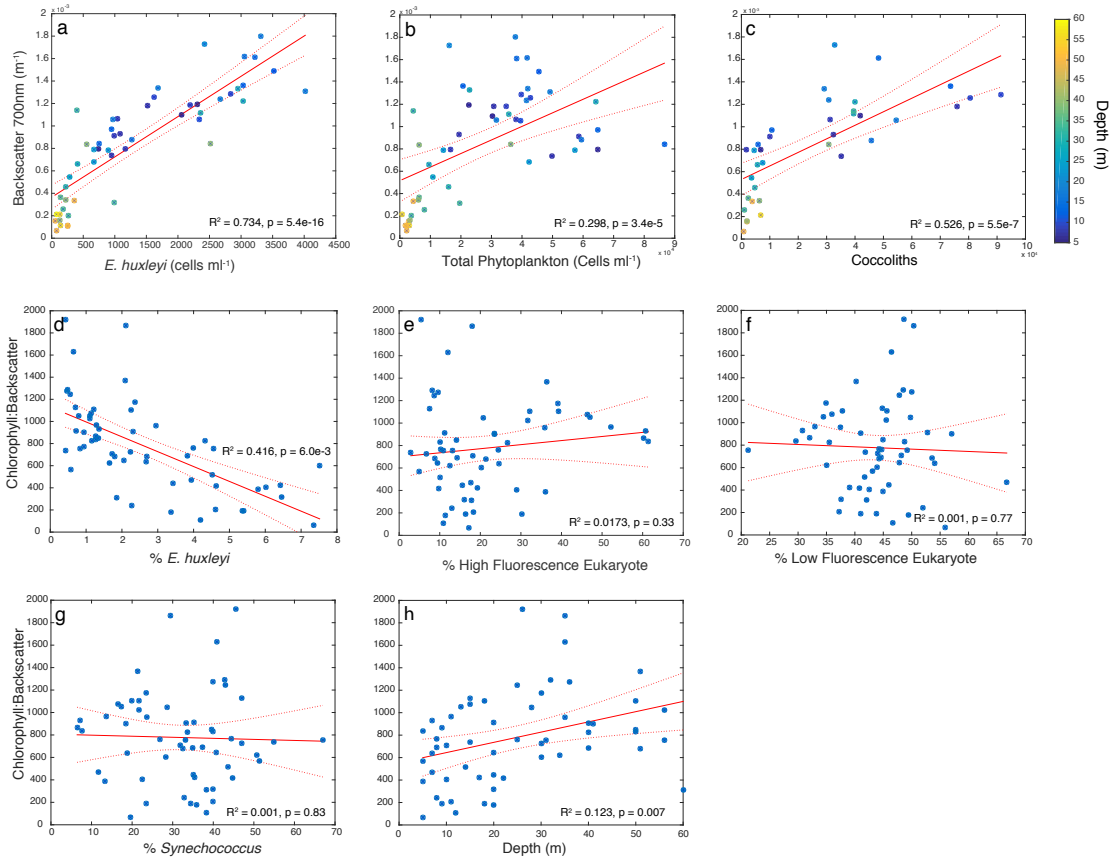
Supplemental Figure 2.1. Landscape of surface currents across the NA-VICE study region. (a) 18-day (23 June – 11 July) composite of Chlorophyll *a* (Chl *a*)/particulate inorganic carbon (PIC) ratio over the study region with an 8-day moving average altimetry overlay centered on June 23. Altimetry reveals the anticyclonic structure of EI and LI and the southward movement of PI. Profiling floats (blue) and sediment traps (black) deployed at LI (b), EI/ EI_R (c) and PI (d) closely followed the observed sea surface movement. Green markers indicate CTD cast locations. (a) The float and sediment traps at EI were deployed in the western side of the eddy. The float and traps traveled clockwise around the eddy, forming an almost complete circle upon recovery. Float and traps were deployed at EI_R near the center of the eddy feature and traveled with tighter clockwise movement before recovery. (b) The profiling float deployed at LI the traveled west before looping back south, remaining in the eddy for the duration of the deployment. No sediments traps were deployed at LI. (c) Float and sediment trap deployments at PI experienced coherent movement south of the deployment location. Triangles indicate float deployment location; squares indicate recovery. Note that color images in (b-d) are for illustrative purposes and correspond to satellite imagery for a specific day, while tracks of floats and sediment traps cover the 4-day deployment period. For (a), Chl *a*/PIC layer pixels are $n = 18$, other than where covered by cloud. For (b-d), single day satellite products are $n = 1$.



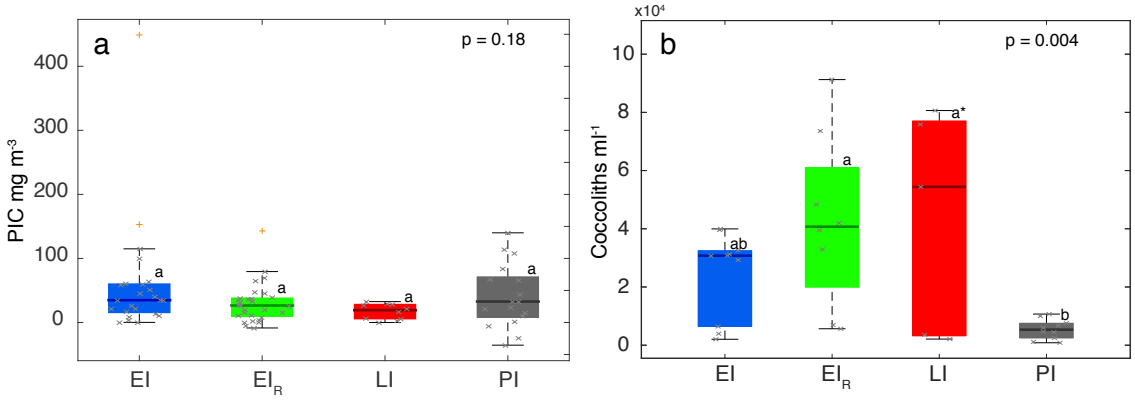
Supplemental Figure 2.2. Salinity profiles for EI (a), EI_R (b), LI (c), and PI (d) stations during float deployments. Salinity remained similar for individual depths at all stations except for PI, where decreases in salinity by 0.1 psu were observed below the surface layer between 50-74 h. These observations are consistent with mixing water masses. The consistency of flow cytometry data, along with lipid- and gene-based measurements, before and after this incursion suggests that Lagrangian sampling was maintained in the surface waters. Color bars all scaled similarly except for D. For each data point, n = 1 technical replicate.



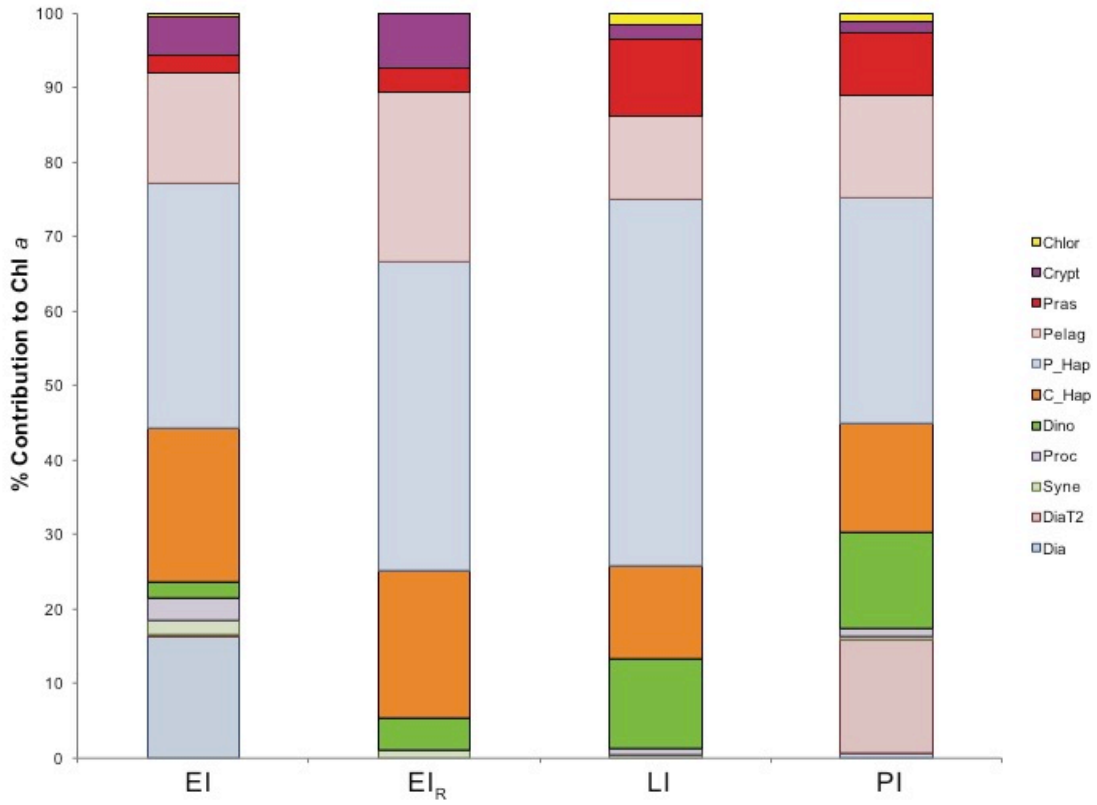
Supplemental Figure 2.3. Flow cytometry counts of *E. huxleyi* (green) and all other phytoplankton (black) at EI (a), EI_R (b), LI (c), and PI (d). Numbers associated with individual series indicate the year-day of the cast; parentheses represent days for multiple casts. The highest total phytoplankton concentrations were observed at PI at 1×10^5 cells ml^{-1}, which extends from the surface down to 30 m. Stations EI/EI_R had a subsurface peak in cell abundance between 20-40 m at 5×10^4 cells ml^{-1}. Station LI also had a subsurface maximum in cell abundance of 6×10^4 cells ml^{-1} between 10-30 m. *E. huxleyi* abundance was most concentrated at EI/EI_R with 4×10^3 cells ml^{-1} and least concentrated at PI, ranging from 3×10^2 to 1×10^3 cells ml^{-1} at the depths of highest abundance. (a-d), for each sample, $n = 2$ technical replicates.



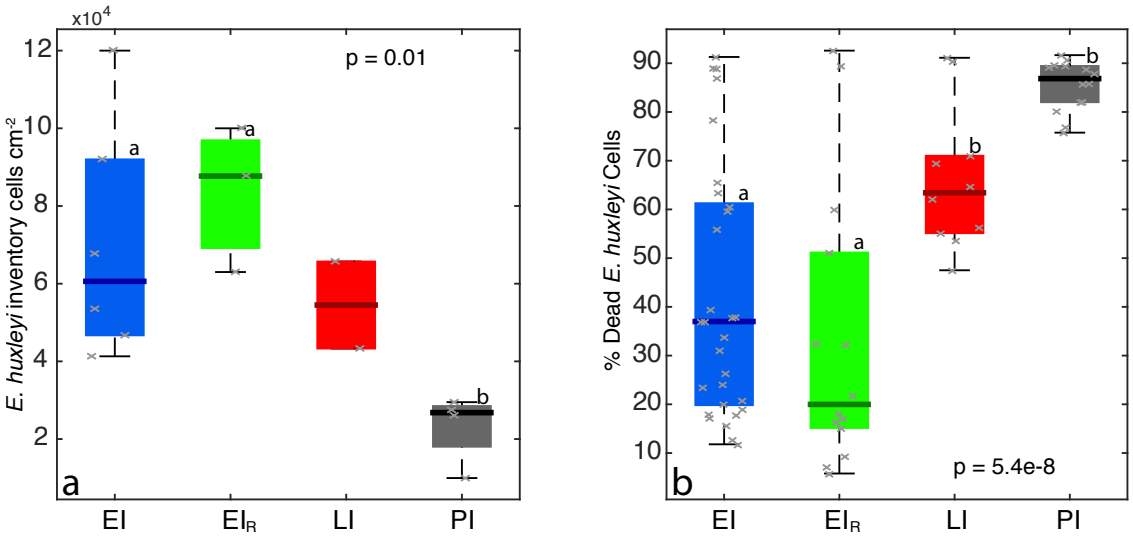
Supplemental Figure 2.4. Correlation analyses comparing profiling float optical backscatter (700nm) and chlorophyll fluorescence measurements with flow cytometry counts of specific groups of phytoplankton. Both analyses revealed that *E. huxleyi* abundance had a significant impact on these optical parameters. (a) Backscatter has a better relationship to *E. huxleyi* abundance than it does to (b) total phytoplankton abundance despite that *E. huxleyi* only makes up between 0.5-8% of the total cell count (d). (c) Coccolith concentrations also correlated with backscatter but not as well as *E. huxleyi*. Correlations (a-c) are anchored by the deeper measurements with generally low particle numbers having low backscatter. In shallow water measurements, where backscatter is decoupled with depth, the data in (a) show a tighter fit to the regression than in (b), suggesting that *E. huxleyi* cells contribute significantly to the backscatter signal within a mixed population. (d-h) comparison of the relative impacts of cell abundance for discernable subgroups of the community (via flow cytometry) on the chlorophyll:backscatter ratio. (d) *E. huxleyi* had the smallest range in abundance but was the only group that supported a significant relationship to the chlorophyll:backscatter ratio. We recognize that the (e) high fluorescence and (f) low fluorescence containing eukaryote groups are phylogenetically broader than *E. huxleyi* but they are distinguished by their similar chlorophyll fluorescence and side scatter properties. (g) *Synechococcus* was numerically dominant in many measurements but had a negligible influence on chlorophyll:backscatter. (h) Depth showed a small but positive correlation with chlorophyll:backscatter likely influenced by photoacclimation of cells to lower light at depth. For all panels, solid lines indicate linear regression line; dotted lines indicate 95% confidence intervals. For (a-b), n = 51 observations; (c-g), n = 57 observations.



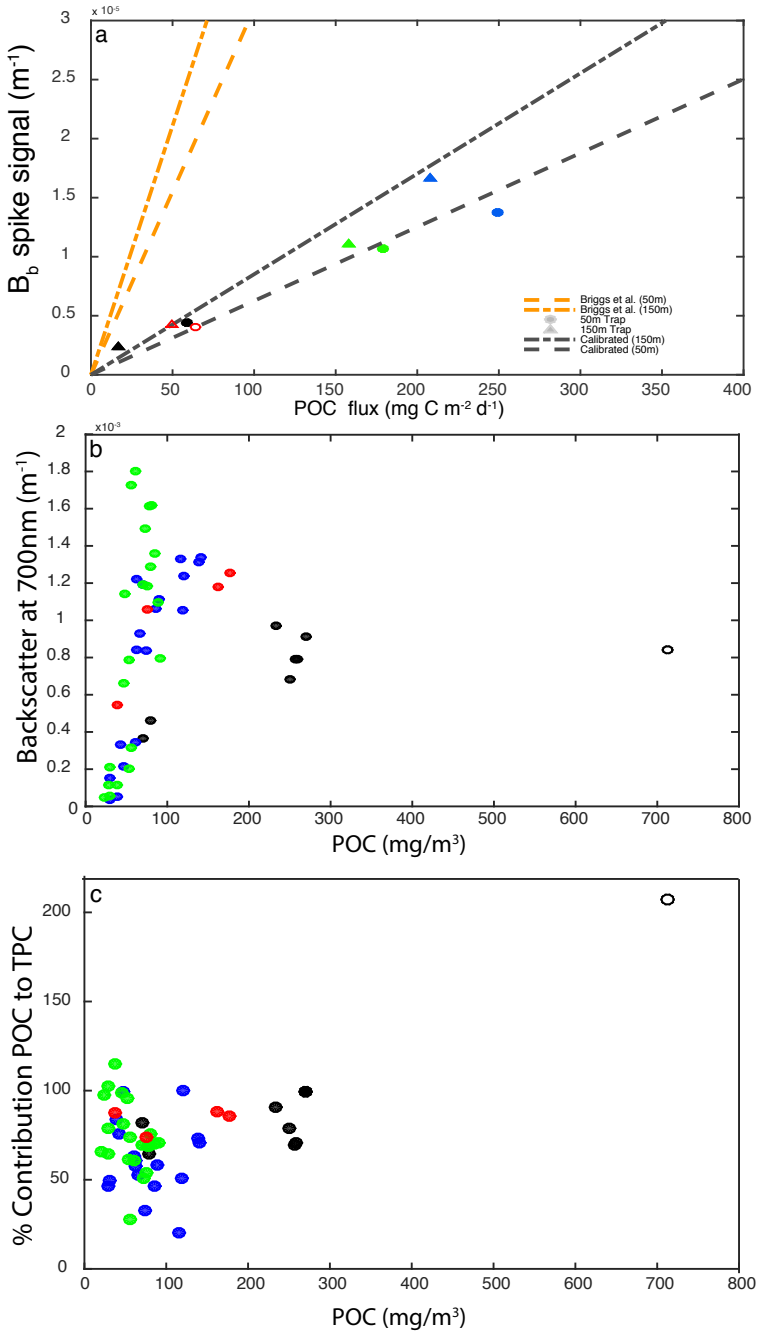
Supplemental Figure 2.5. (a) Particulate inorganic carbon (PIC) concentrations in the upper 50 m at EI (blue), EI_R (green), LI (red), and PI (gray). No significant differences were observed between stations, with all showing elevated concentrations of PIC. This corroborates with satellite PIC measurements, with elevated PIC observed at each of the water masses (see Figure 2.1). (b) Free coccolith counts revealed a significantly greater number at EI_R and LI compared to PI. This is in line with the relatively low optical backscatter (B_b) observed at PI (Figure 2.3). Seeing as calcite-specific scattering is highest for coccolith size particles (Balch *et al.* 1996), higher free coccolith concentrations may inflate the B_b signal while PIC concentrations are comparable to areas with lower B_b. *Coccolith samples at LI were only collected at the initial station occupation and not final occupation and, therefore, only reflect the initial optical properties. Box bounds denote 25% and 75% quantiles around the median (thick line). Vertical capped lines indicate max and min data values. Outliers indicated with (+). Letters above boxes denote statistically different groups based on ANOVA. Gray x indicate individual data points. For (a): EI, n = 24; EI_R, n = 29; LI, n = 10; PI, n = 17 station replicates. For (b): EI, n = 10; EI_R, n = 8, LI, n = 5; PI, n = 10 station replicates.



Supplemental Figure 2.6. The relative contribution of different phytoplankton taxa to total chlorophyll *a* (% contribution to Chl *a*) at EI, EI_R, LI, and PI, as assessed by HPLC-based CHEMTAX. The percent contribution value is normalized to the averaged pigment samples from 0-60 m for each bloom feature. LI and PI stations feature a higher Chl *a* contribution from prasinophytes, chlorophytes, diatoms, and dinoflagellates, while coccolithophore-like haptophytes are less abundant, which may indicate taxa succession. [Chlor –chlorophytes; Crypt – cryptophytes; Pras – prasinophytes; Pelag – pelagophytes; P_Hap – *Phaeocystis* like haptophytes; C_Hap – coccolithophore like haptophytes; Dino – dinoflagellates; Proc – *Prochlorococcus*; Syne – *Synechococcus*; DiaT2 (Type 2) and Dia –diatoms]. In order EI-PI, n = 20, 26, 12, 14 station replicates.



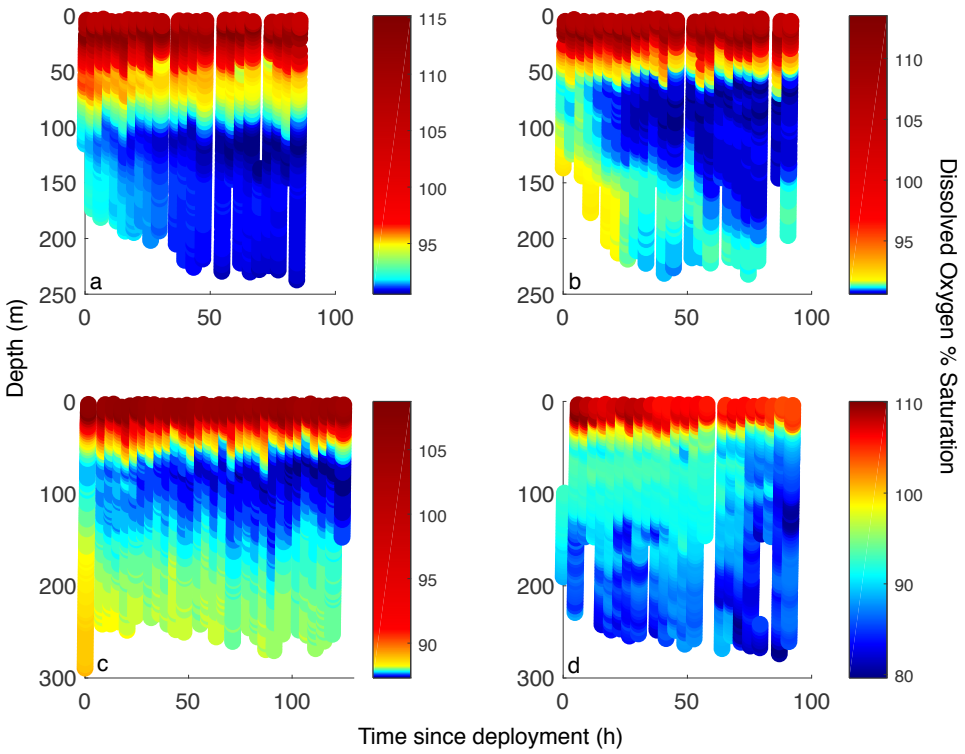
Supplemental Figure 2.7. Box and whisker plots of water column inventory of *E. huxleyi* cells and the corresponding fraction of dead cells. (a) Water column inventory of *E. huxleyi* cells per cm^2 from the surface down to 150 m at EI (blue), EI_R (green), LI (red), and PI (black), using flow cytometry data from Supplemental Figure 2.3. Inventories were most abundant at EI/EI_R and least abundant at PI ($p<0.05$). (b) Fraction of *E. huxleyi* cells positively stained with SYTOX Green, indicating the percentage of dead cells within each feature. SYTOX Green is a charged, DNA intercalating stain that only stains the DNA of operationally dead cells (i.e., those with compromised membranes). Mean mortality was lowest at EI_R (33%) and EI (42%), and progressively increased at LI (66%) and PI (85%), where most *E. huxleyi* cells were dead ($p<0.001$). Box bounds denote 25% and 75% quantiles around the median (thick line). Vertical capped lines indicate max and min data values. Letters above boxes denote statistically different groups based on ANOVA. For (a), n = 6, 3, 2, 4 station replicates. Gray x indicate individual data points. No statistic on LI with n = 2. For (b), n = 29, 14, 10, 14 station replicates, respectively.



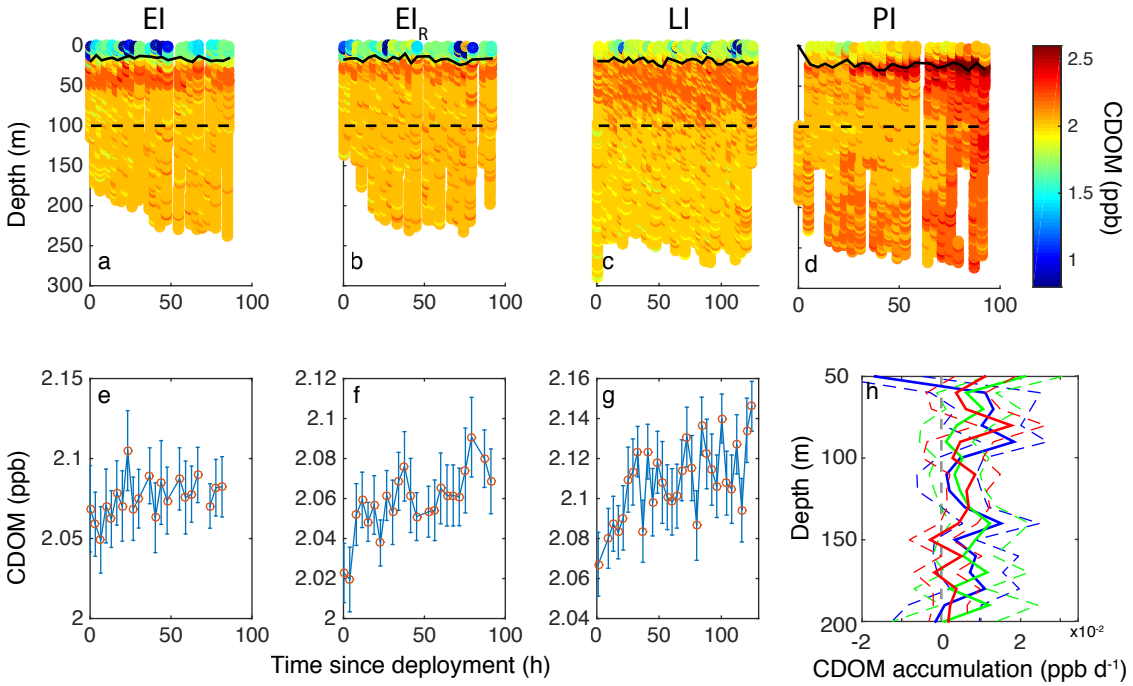
Supplemental Figure 2.8. Relationship between total carbon flux and backscatter spike signal. The measured particulate organic carbon (POC) flux rate measurements at 50 m (circles) and 150 m (triangles) was compared to the spike aggregate signal at those respective depths at EI, EI_R, and PI. Briggs et al. explored using the spike signal as an estimate for aggregate POC flux using equation [(2.2); rearranged below] using a 75 m d^{-1} sinking rate, terms that accounted for the ratio of POC to optical backscatter (B_b) signal [35400 mg C m^{-2}], and a correction for organic carbon loss in a particle during sinking [$(z/100m)^{-0.28}$], where z is the depth of the measurement (Briggs *et al.* 2011). The Briggs et al. estimates for the relationship between POC and the spike signal at 50 m and 150 m are shown in orange. Our data depart from this relationship between POC and the B_b spike signal, likely based on several factors. First, the significant contribution of PIC to the total particulate carbon (TPC) in our systems would change the

relationship between POC and the B_b spike signal, as PIC would then account for a portion of the signal (PIC is assumed to inflate the B_b signal and thus lower the ratio). (b-c) The % contribution of POC to TPC varied among samples and stations, likely explaining some of the variability observed in the POC:B_b relationship (c). EI and EI_R commonly exhibited very low % POC within observations. However, using Eq. (2.2), a lower POC:B_b ratio would increase the slope observed in (a) rather than decreasing it to one comparable to our data. A second possibility is that differences in the calibration in B_b sensor between our profiling floats and the platform used in the Briggs et al. study. We tested this by using an average ratio of POC:B_b from our data (175443 mg C m^{-2}) in Eq. (2.2). The result is shown above with the calibrated 50 m and 150 m regression lines (gray). Of note, the data not only follow the calibrated regressions more closely, but the 150 m data also depart from the 50 m as predicted by Eq. (2.2). We then used the calibrated Eq. (2.2) to estimate the POC flux at LI, where no sediment traps were deployed, based on the B_b spike signals observed at 50 m and 150m, yielding estimates of 64 and 49 mg C m^{-2} d^{-1}, respectively (a; hollow red shapes). For (a-c) scatter plots are colored by station with EI (blue), EI_R (green), LI (red), and PI (black). In (b-c), hollow data outlier with extreme POC value was removed from analysis; n = 51 observations.

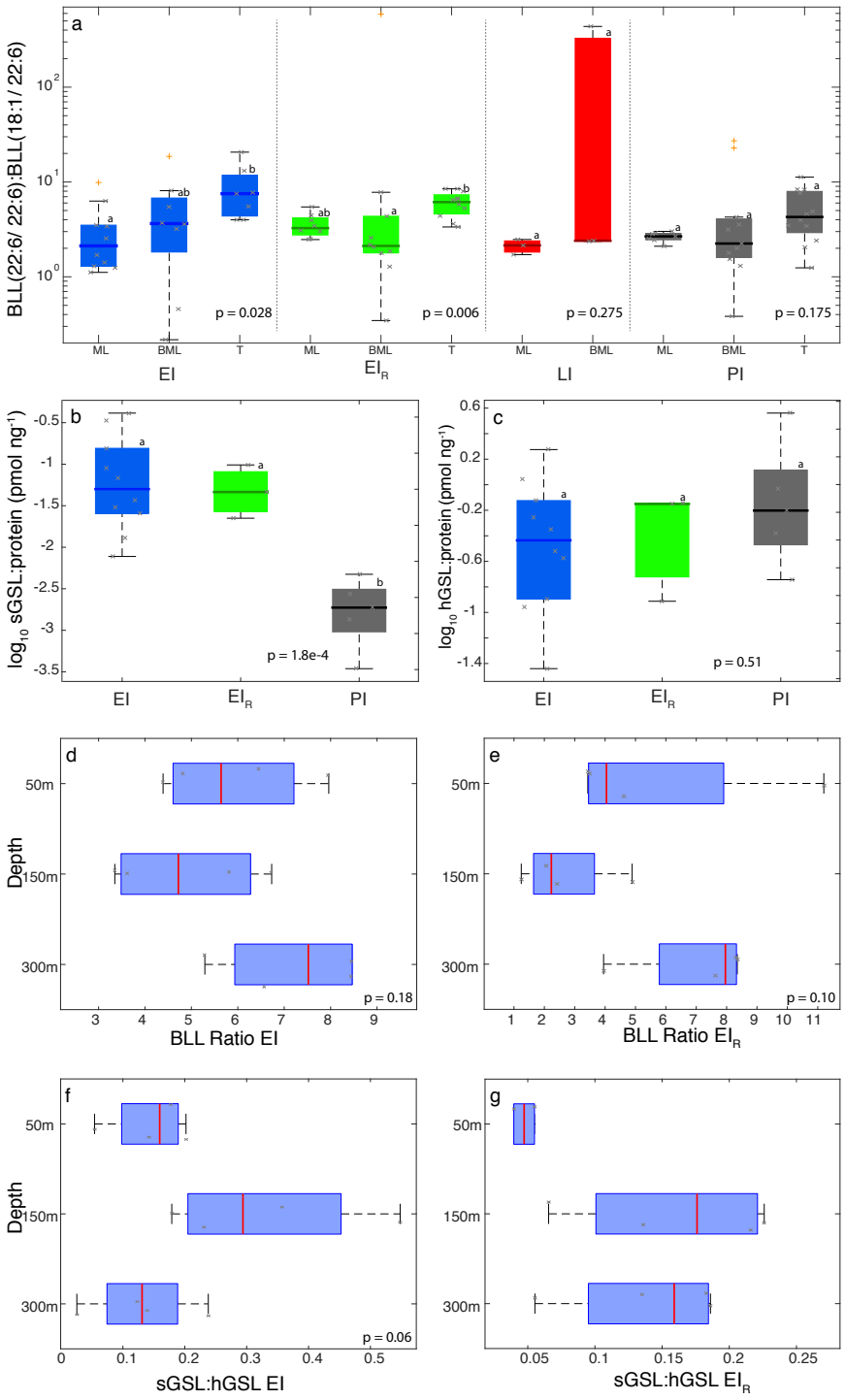
$$B_b \text{ spike signal} = \text{aggregate POC flux}/(35400 \text{ mg C m}^{-2} \times 75 \text{ m d}^{-1} \times (z/100\text{m})^{-0.28}) \quad \textbf{(2.2)}$$



Supplemental Figure 2.9. Dissolved oxygen concentrations (DO; % saturation) measured during profiling float deployments for EI (a), EI_R (b), LI (c), and PI (d) water masses. Sub-surface minimum layers in DO were observed between 50 and 150 m for EI, EI_R, and LI. Additionally, steady decreases were observed over the observation time at depths below the euphotic zone (~50 m across all stations). Color bars at EI, EI_R, and LI are adjusted to highlight the small sub-euphotic decreases in DO over the deployment periods. For individual data points, *n* =1 technical replicate.

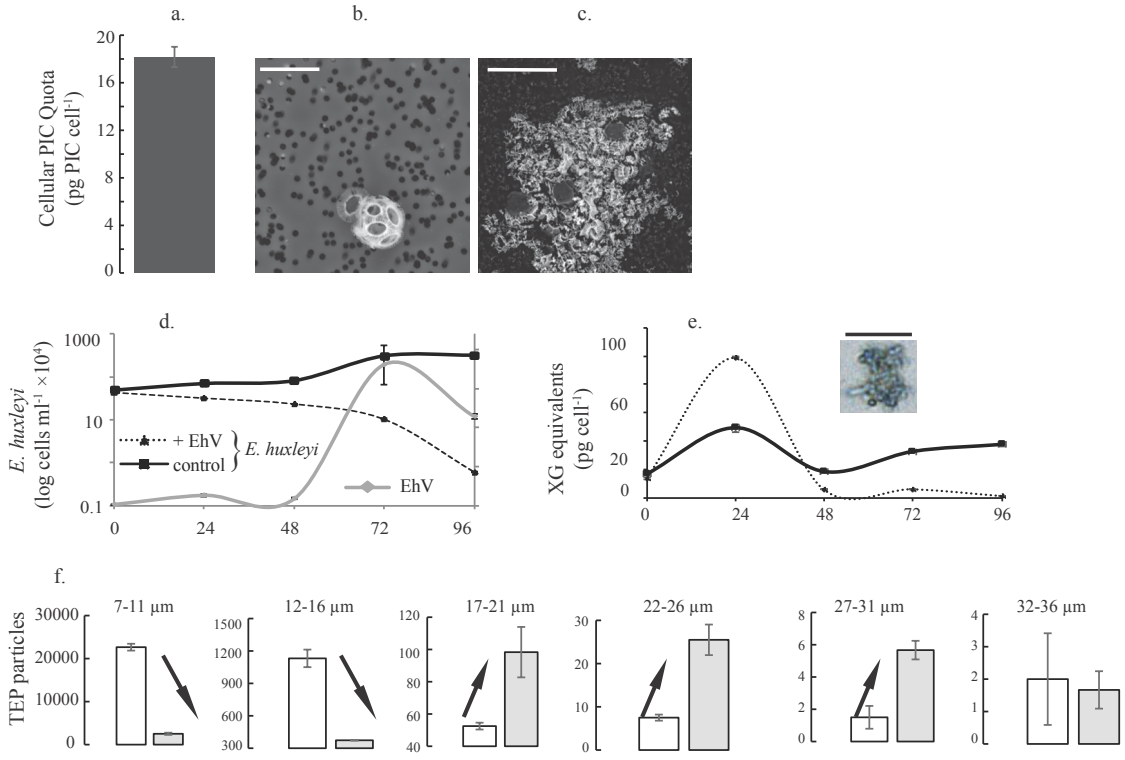


Supplemental Figure 2.10. Water column profiles of fluorescent chromophoric dissolved organic matter (CDOM) at EI (a), EI_R (b), LI (c), and PI (d) over periods of profiling float deployments. Solid black lines indicate the depth of the pycnocline, as measured by the maximum change in density over change in pressure. Dashed black lines indicate the 100 m depth horizon. (e-g) Changes in the average CDOM concentration from the pycnocline to 100 m (error bars denote ± 1 SE). Virus infection and cell lysis has previously been linked to releases of CDOM, but accumulation is not observed in all studies (Balch *et al.* 2002; Rochelle-Newall *et al.* 2004). The elevated fluorescent CDOM in the photobleached surface water observed at LI and PI, compared to EI and EI_R, suggests recent production of CDOM at those features. Vertical mixing between the surface and sub-surface layer is likely not an explanation for higher surface CDOM at LI because of increases in the subsurface layer CDOM (g); rather, it would show decreases if it was diluted by the surface water. Due to mixing sub-surface water masses, PI was not included in this analysis. (h) CDOM accumulation rates are small and positive at EI (blue), EI_R (green) and LI (red). Dashed lines indicate upper and lower 95% confidence intervals. Though it is unclear what processes are contributing to CDOM production—whether phytoplankton can directly exude CDOM during lysis (Castillo *et al.* 2010) or there is a necessary intermediate after viral lysis and release of DOM to produce CDOM (Balch *et al.* 2002)— it appears CDOM production may occur throughout infected blooms. For data points in (a-d), n = 1 technical replicate. For e-g, min n = 19, max n = 31 data points contribute to each average.

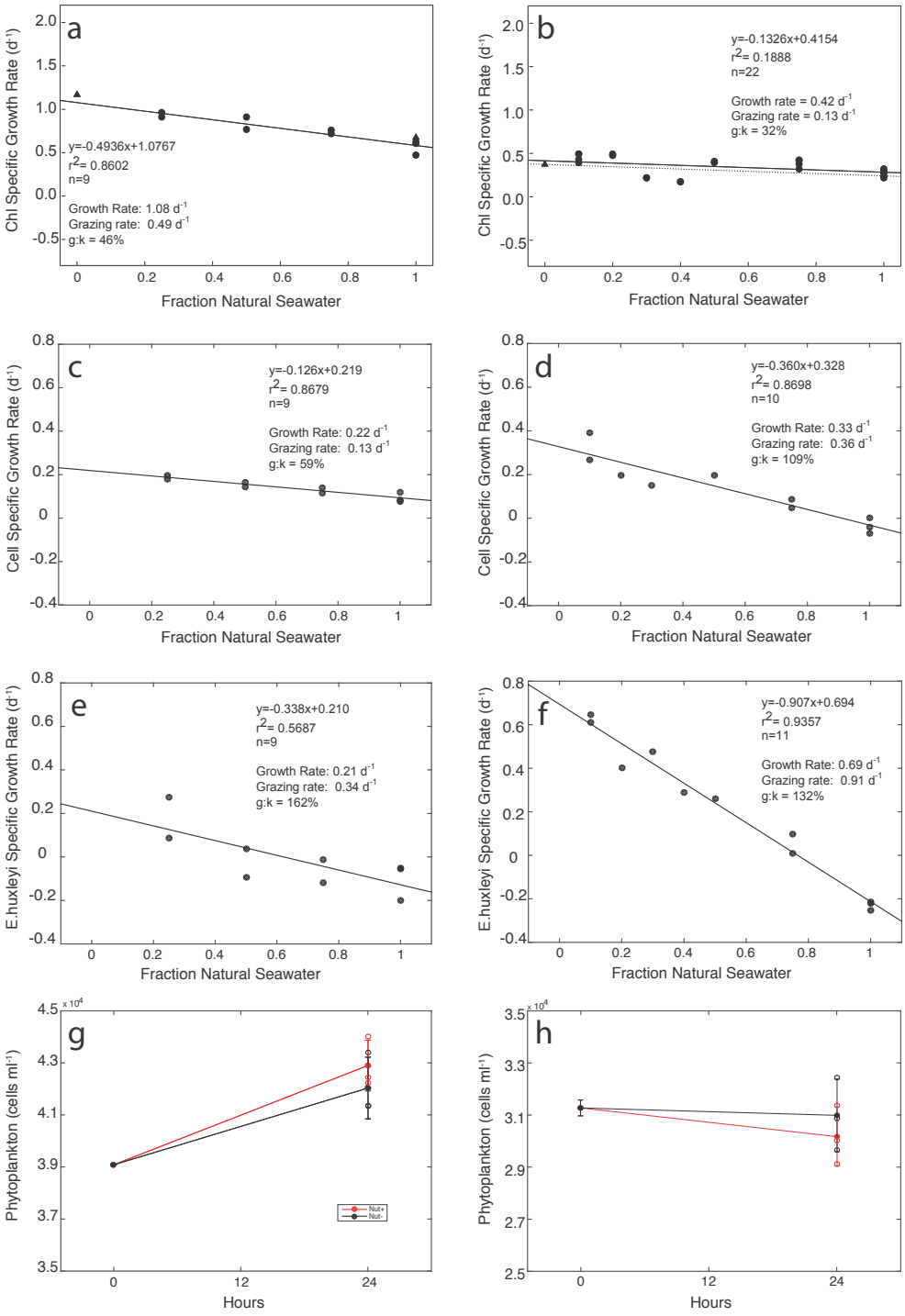


Supplemental Figure 2.11. Observations of infectivity lipid ratios at individual stations. (a) The betaine-like lipid BLL(22:6/ 22:6) to BLL(18:1/ 22:6) ratio at individual stations had median values that were consistently higher in sediment trap material than in water column material. Significant differences between the mixed layer and trap material ratios were observed at EI, while at EI_R, only the water below the mixing layer differed significantly from the trap material (Kruskal-Wallis ANOVA). There were no significant differences between mixed layer (ML), below mixed layer (BML), and trap (T) ratios at PI. No

traps were deployed at LI and there was no difference above and below the mixing layer depth. (b-c) The biomass-normalized ratio of sialic acid glycosphingolipid (sGSL) (sGSL:protein) was greatly elevated in trap material at EI and EI_R compared to PI (ANOVA). These differences were not observed for biomass-normalized ratios of host glycosphingolipid (hGSL) demonstrating the preferential enrichment of sensitive, *E. huxleyi* cells in the traps at EI and EI_R. For a-b, letters above boxes denote statistically different groups. (d-e) BLL ratios and (f-g) sGSL:hGSL ratios in sediment trap material collected at 50 m, 150 m, and 300 m were statistically indistinguishable between depths (ANOVA). Taken together, data from panels (d-g) are consistent with sinking particles being derived from the same infected surface source populations at EI and EI_R with high sinking rates. For all plots, upper and lower box bounds denote 25% and 75% quantiles around the median (thick line). Vertical capped lines indicate max and min data values. Outliers indicated with (+). Gray x indicate individual data points. For (a), n = 10, 8, 7,8 10, 12, 3, 3, 7, 11, 12 environmental replicates; For (b-c), n = 10, 3, 5 station replicates; for (d-f), n = 4, 4, 4 environmental replicates; For (g), n = 2, 4, 4 environmental replicates, respectively. No statistic on (g) with n = 2.

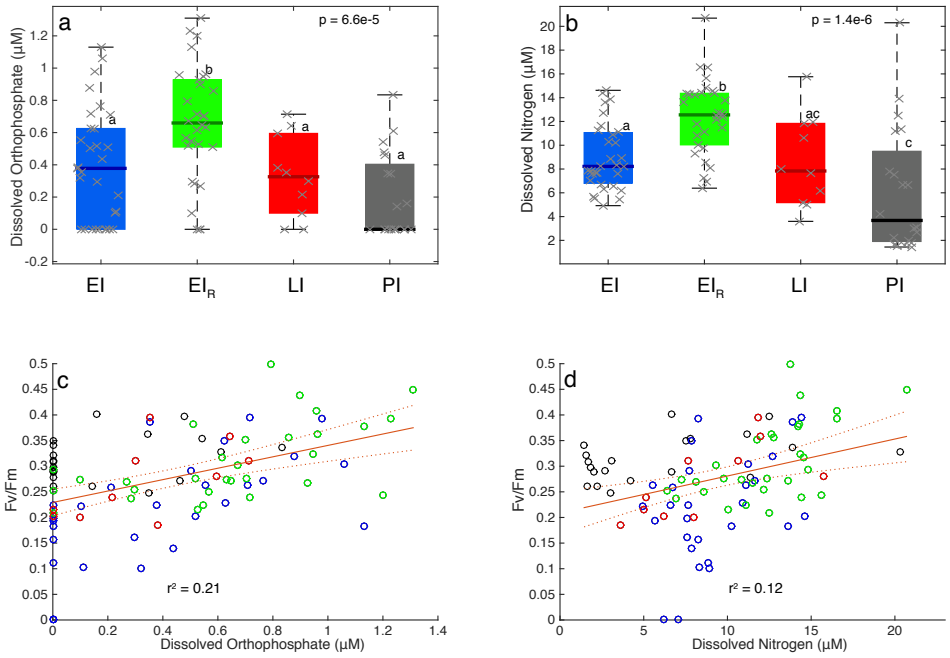


Supplemental Figure 2.12. EhV infection of calcified *E. huxleyi* triggers transparent exopolymeric particle (TEP) and particle production. (a) Cell-associated particulate inorganic carbon quota for *E. huxleyi* DHB607. Error bars denote ± SD around the mean. (b-c) Scanning electron microscopy (SEM) images of calcified DHB607 as a single cell (b) and aggregated particle during virus infection (c); scale bars correspond to 5 µm and 8 µm, respectively. (d) Infection dynamics of host and virus abundance for infected and uninfected DHB607 cells. Average host abundance (± SD from biological triplicates) during infection with EhV99B1 (dotted lines) and for virus free control (solid lines), along with corresponding EhV99B1 abundance for infected treatment, as measured by flow cytometry (see Methods). (e) Comparative dynamics of TEP production for data shown in panel (d). Plots show the dynamics of TEP production (average ± SD) for DHB607 cells in the presence (dotted lines) and absence (solid lines) of EhV99B1 (left panels) over 96 h infection period. TEP concentrations are expressed in xanthum gum (XG) equivalents. Note the increse in cellular TEP production during early infection dynamics, which is not due to cell lysis (which was observed 72 hours post infection, hpi). *Inset*: FlowCam imaged particle from EhV99B1-infected DHB607 culture (scale bar= 30 µm). (f) Patterns in the number of aggregates within arbitrarily-defined size classes for *E. huxleyi* DHB607 during infection with EhV99B1. Alcian Blue-stained TEP-containing aggregates (like those pictured in inset of panel (e) were imaged by FlowCAM and sorted based on area based diameter, which is automatically calculated by the FlowCam software. Size bin ranges are indicated above each panel. Bars represent the average number of aggregates (± SD for biological triplicates) within each size range at the end of the experiment for uninfected control (white) and EhV99B1-infected (gray) cultures. Arrows indicate the respective shift in aggregate number. Note the increase in larger aggregates in infected cells relative to control. For a, d, e, and g, n = 3 biological replicates.

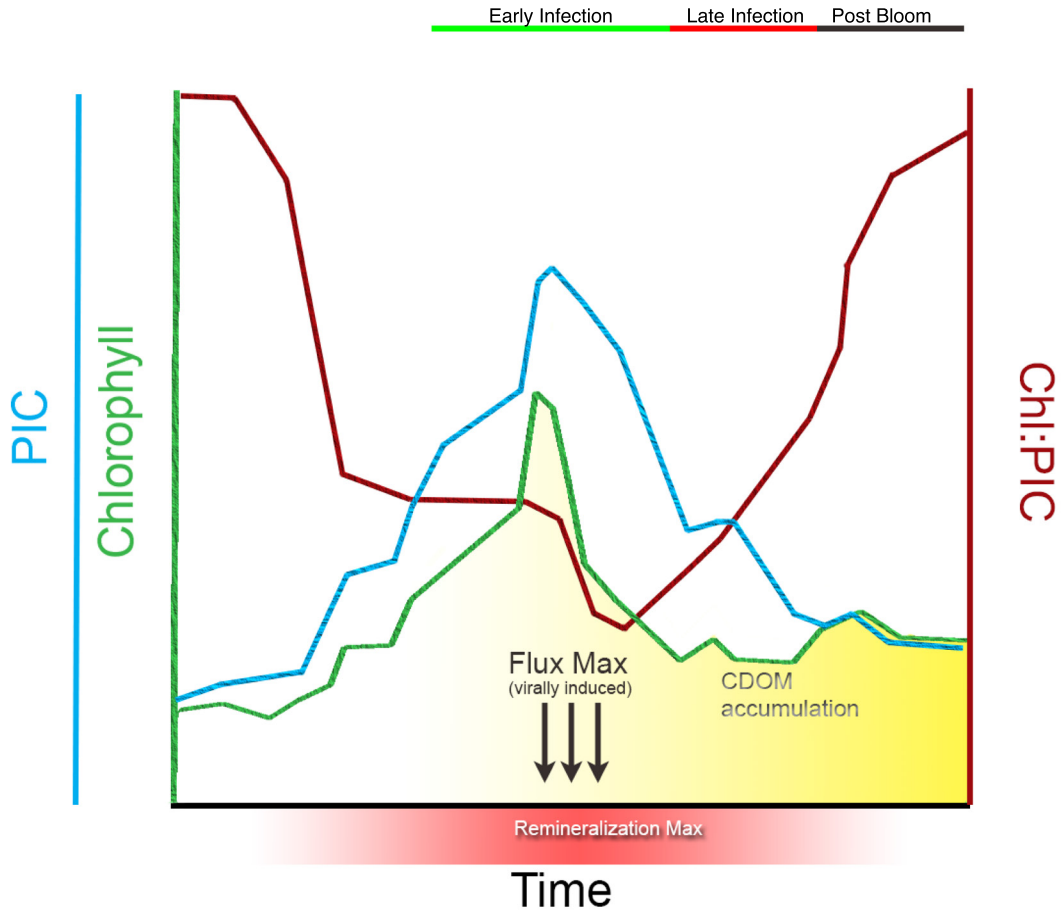


Supplemental Figure 2.13. Community and *E. huxleyi* specific growth and grazing rates derived from Landry-based (Landry & Hassett 1982) dilution experiments performed at EI (a, c, e, g) and EI_R (b, d, f, h). (A-B) Using the classic chlorophyll based measurement of growth rate, linear projection to community growth in the absence of grazers shows positive growth rates over the 24 h incubation period for both EI and EI_R. Grazing was relatively elevated at EI in comparison to EI_R. (c-d) While cell-specific community growth rates also showed generally positive growth rates in the absence of grazing, cell specific grazing

pressure was higher at EI_R than EI. These differences may be explained by selective retention of phytoplankton during chlorophyll filtration (inefficient retention of small, < 0.7 μm, picophytoplankton such as *Synechococcus* onto GF-F filters) or by preferential grazing on particular prey species. (e-f) *E. huxleyi* specific grazing was elevated (almost three times higher) at both EI and EI_R in comparison to community grazing rates. (g-h) Some incubations were spiked with ammonia, phosphate, and silica to alleviate the possible impact of nutrient limitation on cells in dilution experiment. Statistically indistinguishable changes in cell concentration over 24 h for incubations with (Nut+; red) and without (Nut-; black) nutrients argue that resident populations were not nutrient limited (Student's T test, $p = 0.38$, 0.47, respectively for g and h T24). For (g); T0, $n = 2$; T24, $n = 3$ biological replicates. For (h); $n = 3$ biological replicates. Open circles indicate individual data points. Error bars represent ± 1 SD around the mean.



Supplemental Figure 2.14. (a) Orthophosphate and (b) nitrogen concentrations in the upper 50m at EI (blue), EI_R (green), LI (red), and PI (gray). EI/EI_R and LI concentrations were similar to or higher than those observed during an *E. huxleyi*-dominated mesocosm bloom experiment (Kimmance *et al.* 2014), suggesting nutrients were not limiting growth. PI had significantly lower dissolved nitrogen concentrations but also had the highest phytoplankton community abundance (Figure 2.S1). Only weak correlations were observed between photosynthetic quantum yield of photosystem II (F_v/F_m) and respective concentrations of (c) dissolved orthophosphate or (d) dissolved nitrogen at all stations. It was noteworthy that F_v/F_m measurements for PI populations were all >0.2, suggesting the resident phytoplankton community encountered limited nutrient stress. For (a-b), box bounds denote 25% and 75% quantiles around the median (thick line). Vertical capped lines indicate max and min data values. Gray x indicate individual data points. Letters above boxes denote statistically different groups based on ANOVA. In order, sample size is n = 31, 30, 10, 20 station replicates. For (c-d), dashed lines represent 95% confidence bounds around the regressions.



Supplemental Figure 2.15. Conceptual model showing the expected dynamics of an infected *E. huxleyi* bloom, based on satellite hindcasting over the lifespan of the LI population and our collective *in situ* observations at all three stations. The LI station had a chlorophyll (Chl) peak four times higher than EI/EI_R and two times higher particulate inorganic carbon (PIC) reflectance, consistent with a very large coccolithophore bloom. With atmospheric conditions allowing for comprehensive satellite imagery, this bloom served as a scaffold upon which to layer and interpret the different stages of EhV infection with EI/EI_R representing the peak of the bloom, LI representing late stages of termination, and PI following the termination as Chl recovers. Through the model bloom initiation seen above, increasing Chl and PIC (with a greater change in magnitude of PIC) drive down the Chl:PIC ratio, as observed by remote and *in situ* sensing instruments. Induced production of aggregated particles (via TEP production) during early infection coupled with enhanced grazing of infected cells become prominent controls on *E. huxleyi* accumulation, in turn increasing particle flux from the surface ocean, becoming maximal around the peak of the bloom. The high particle flux is accompanied by elevated remineralization of the sinking particulate matter in the mesopelagic. Following the peak of the bloom, the Chl *a*:PIC ratio decreases further, with the removal of PIC lagging behind decreases in Chl, likely due to enhanced presence of high-scattering, free coccoliths before the ratio increases into late infection upon phytoplankton succession. Flux rates relax through the termination of the bloom. CDOM accumulation rate increases through the termination of the bloom and into the post-bloom, with succeeding phytoplankton increasing Chl at the as PIC continues to diminish.

2.8 SUPPLEMENTAL TABLES

Supplemental Table 2.1. Depth integrated water column respiration from 50-150 m derived from optical measurements of dissolved oxygen utilization and direct bottle based measurements of microbial community respiration (Collins *et al.* 2015). The statistical uncertainty of direct measurements was calculated using the bootstrap method described in Collins et al.[1]. Optical measurements closely follow the direct observation measurements at EI and EI_R. Confidence intervals represent ± 1 SD of the mean.

50-150m	**Optically derived Depth-integrated water column respiration**	**Direct observation Depth-integrated water column respiration**
Station	**(mg C m^{-2} d^{-1})**	**(mg C m^{-2} d^{-1})**
EI	669.0	887.9 ± 456.2
EI_R	665.9	835.0 ± 629.0
LI	165.8	--
PI	--	465.2 ± 182.1

Supplemental Table 2.2. Correlations for PIC with GSL species. sGSL explain 65% of the variability in PIC sinking flux rate while hGSL and vGSL had negligible explanatory power. This suggests that material from susceptible *E. huxleyi* makes up a significant fraction of the coccolithophore and coccolith export. P values obtained by ANOVA.

	sGSL:PIC*	**hGSL:PIC**	**vGSL:PIC**
Regression Coefficient	0.019	1.639E-04	0.009
Intercept	18.307	45.594	46.975
R Square	0.650	0.008	0.002
P-value	1.47E-08	0.604	0.791

*with the removal of one highly leveraging outlier

Supplemental Table 2.3. **Deck-board seawater incubations.** Flow cytometery measurements of *E. huxleyi* and total phytoplankton populations measured over 2-3 d in seawater collected at 8 m depth from EI_R. TEP concentrations were also measured at each time point. Control incubations were maintained at 20% surface PAR while the low light incubation was maintained at 1% surface PAR.

Incubation	Date	Group	hours	*E. huxleyi* Average ± 1 SD (cells ml^{-1})		Total Phytoplankton Average ± 1 SD (cells ml^{-1})		TEP Average ± 1 SD (OD L^{-1})	
	7/10/12	Control	20	1422.06	± 164.06	29079.49	± 29.64	-3.82	± 0.47
EI_R 1	7/11/12		41	1651.64	± 41.08	34123.73	± 1180.66	-6.03	± 4.90
	7/12/12		66	1871.93	± 203.13	24402.52	± 1284.64	33.51	± 4.45
		Low							
	7/12/12	Light	31	807.96	± 64.14	25773.77	± 1084.19	4.61	-
EI_R 2	7/13/12		53	924.52	± 12.20	24137.39	± 743.47	50.58	± 1.23
	7/12/12	Control	31	1190.53	± 51.29	29903.26	± 1514.31	12.12	-
	7/13/12		53	1469.73	± 273.49	27766.18	± 884.03	41.27	± 7.14

Chapter 3: Photoacclimation in North Atlantic Natural Bloom Communities

3.1 ABSTRACT

The optical absorption properties of seawater can be used to assess details of phytoplankton community structure. Given the range of EhV infection states identified in North Atlantic blooms of *E. huxleyi* in Chapter 2, we first tested whether infection dynamics can be resolved from spectral absorption. While absorption characteristics of infected *E. huxleyi* cultures differ from control cultures, this signal was unresolvable in the intense photoacclimation gradient and mixed assemblage within natural North Atlantic communities. Therefore, we investigated the photoacclimation properties of the North Atlantic bloom communities and assess how much of the observed variability in spectral absorption characteristics reflects the taxonomic diversity and/or physiological variability of the populations. Characterizing population structure, spectral absorption, and pigmentation, we found that variations in absorption characteristics were largely due to the photoacclimation state of resident phytoplankton populations. Spectral suppression and absorption per cell increased with depth while chlorophyll specific absorption decreased with depth. Absorption characteristics were also observed to adjust to the light field, with absorption ratio of 440 nm to 490 nm decreasing at a water mass with a low irradiance ratio 444 nm to 490 nm. Differences in the photoprotective to photosynthetic carotenoid ratios were largely driven depth with little difference between visited water masses. Species composition may drive some spectral characteristics but are difficult to separate from depth. These field and culture studies reveal the profound influence environmental factors have on spectral absorption characteristics of phytoplankton.

3.2 INTRODUCTION

Phytoplankton play a critical role in driving the carbon cycle and contribute to up to 50% of global primary production (Field *et al.* 1998). The diversity that exists between phytoplankton taxa and the environmental conditions that structure their physiology make it necessary to know more than the amount of chlorophyll to estimate productivity. As a result, photoacclimation (short term, environmentally driven changes in physiology within a community) and photoadaptation (physiological characteristics of individual taxa) within phytoplankton populations have become important components to understand dynamics of primary productivity in natural communities (Moore *et al.* 2006).

Environmental parameters can play a strong role in regulating phytoplankton photoadaptation by influencing photophysiology. High light conditions increase the cellular concentrations of photoprotective carotenoids to prevent damage to light harvesting reaction centers. In low light, phytoplankton will increase the concentration of chlorophyll and photosynthetic carotenoids to maximize utility of the light field. These adaptations can alter the spectral absorption of the phytoplankton. Nutrient limitation can also impact absorption characteristics and has been reported as a function of growth phase. Nitrogen limitation in particular can impact photophysiology and increase the Chl *a* specific absorption spectra (Berges *et al.* 1996; Henriksen *et al.* 2002).

The pelagic open ocean contains extensive variability, as well as strong gradients in environmental characteristics that influence photoadaptation of community members. Factors such as light attenuation in the water column due to absorption by seawater, as well as dissolved and particulate constituents are coupled with variability in surface

incident light due to changes in cloud cover, which bring phytoplankton to extremes in light conditions throughout the water column (Gardner *et al.* 1993). Displacement in the water column over short time scales through vertical mixing processes further regulates the photosynthetic and photoprotective requirements of cells throughout the water column (Huisman *et al.* 2006).

Photosynthetic and photoprotective pigment composition of phytoplankton communities largely control the community spectral absorption characteristics. The molecular structure of each pigment endows unique absorption characteristics, which are collectively used to optimize interaction with light in the specified phytoplankton niche (Bidigare *et al.* 1987b). Absorption characteristics of phytoplankton are also controlled by cell size and pigment concentration, creating an effect known as pigment packaging (Duysens 1956; Morel & Bricaud 1981) or intracellular self-shading. Increasing cell size or pigment concentration reduces the absorption per unit of pigment, effectively flattening the absorption spectrum by dampening absorption where the coefficient is largest.

The great diversity of phytoplankton taxa possess a wide variety of pigments which are generally more similar within individual phyla (Hoepffner & Sathyendranath 1991). These suites of pigments produce absorption spectra that differ between major phyla and may be used to identify species assemblage in nature (Millie *et al.* 2002). Given the number of factors to that can contribute to phytoplankton spectral absorption in the ocean, differentiating between contributing factors is critical contextual information and has been a work in progress for several decades (Bracher *et al.* 2017). It is still unclear however what the relative importance of factors contributing to the

phytoplankton absorption signal can be. It has been argued that environmental acclimation may have less importance in absorption characteristics than species composition in the natural environment (Stæhr *et al.* 2002); however, acclimation has been shown to have a strong influence on spectral characteristics (Eisner *et al.* 2003).

We explored the factors that contribute to the phytoplankton spectral absorption and photoacclimation in several open ocean North Atlantic blooms during the *North Atlantic-Viral Infection of Coccolithophores Expedition* (NA-VICE). This cruise was designed to investigate the host-virus infection dynamics and ecosystem function of the coccolithophore *Emiliania huxleyi*, thus poising us within several similar blooms dominated by haptophytes. Initially, the influence of EhV infection on water column spectral absorption was explored. However, in determining that this signal was not detectable in the North Atlantic natural communities, we instead tested the ability to observe photoaclimation state among other interacting variables influencing phytoplankton absorption spectra including nutrient limitation and taxonomic composition, focusing on four previously characterized locations [(Laber *et al.* 2018); also see Chapter 2].

3.3 MATERIALS AND METHODS

3.3.1 The North Atlantic Virus Infection of Coccolithophores Expedition. (*NA-VICE*; KN207-03; 13 June – 16 July 2012) traversed a ~2000 nautical mile transect from Ponta Delgada, Azores to Reykjavik, Iceland aboard the *R/V Knorr* during which time water column samples were collected at 6 depths for 15 discrete station locations. Water was collected using 10-liter Niskin bottles mounted on a 24-position rosette equipped with a

Seabird SBE conductivity-temperature-depth (CTD) profiler and processed on board for analyses.

3.3.2 Biomass collection for spectral absorption and chlorophyll. Water from three depths (above, below, and at the deep chlorophyll maximum) within the euphotic zone was filtered onto 0.7 μm nominal pore-size, GF/F glass fiber membrane filters to collect biomass for bulk particle spectral absorption. Samples were filtered until biomass was observable as a light green color on the filter, generally 1-3 L. Blank reference filters were wetted with 0.2 μm pore-size filtered seawater (FSW) and treated similar to biomass samples. Filters were placed in tissue capsules, snap-frozen in liquid nitrogen, and stored at -80°C until processed. Total chlorophyll a samples were collected filtering 0.28 L seawater onto GF/F filters and measured using the non-acidified fluorometric method of JGOFS (Joint Global Ocean FluxStudy), with a Turner Designs TD700 fluorometer.

3.3.3 Phytoplankton cell abundance. Phytoplankton cell abundance was performed with a Guava flow cytometer (Millipore) equipped with a 488 nm laser. After water collection, duplicate samples were immediately prepared for analysis; duplicates were averaged for final concentrations. The instrument was set at medium flow rate (0.59 μl s^{-1}). Triggering was set to collect all events above chlorophyll fluorescence (692 nm) threshold value of 1, with minimum gain for all detectors except yellow fluorescence, set at 23.63. Samples were analyzed with GuavaSoft InCyte 2.2.2. *Synechococcus* was gaited using red (threshold immediately above noise) and yellow (threshold at 2 x 10^1

rfu) fluorescence with an oblique gate that encompassed the range of the population. Eukaryotes were gaited with a red fluorescence minimum threshold at 2 x 10^0 and side scatter events above the axis, excluding events gaited as *Synechococcus*.

3.3.1.4 HPLC Pigment. Water samples (0.5-1.5 L at each depth) were collected from CTD casts and filtered onto 25 mm Whatman GF/F glass fiber membrane filters and stored in liquid nitrogen until on-shore analysis. Following extraction in 90% acetone, samples were analyzed on an Agilent 1100 HPLC (High Performance Liquid Chromatography) system with diode array and fluorescence detection. The elution gradient utilized was a modified version of the Zapata et al. (2000) method. The exact elution gradient and protocols have been described in detail elsewhere (DiTullio & Geesey 2002). The HPLC pigment concentrations were then analyzed using the iterative matrix factorization program CHEMTAX (Mackey *et al.* 1996). For comparison of the taxonomic contribution to Chl (taken as % Chl) between process stations, a multivariate analysis of variance was conducted with canonical variates analysis. To compare the ratio of photoprotective carotenoids to photosynthetic carotenoids as a metric for photoadaptive state, the sum of photoprotective carotenoids (PPC) (alloxanthan, β-carotine, diadinoxanthin, diatoxanthan, lutein, and neoxanthan) was divided by the sum of the concentrations of photosynthetic carotenoids (PSC)(19-butanoyloxyfucoxanthin, 19'-hexanoyloxyfucoxanthin, cis-fucoxanthan, fucoxanthin, and peridinin).

3.3.5 Spectral analysis. Samples were processed on an Aminco DW 2000 spectrophotometer using the quantitative filter technique (QFT) to determine pigmented

absorption spectra of phytoplankton $a_{\Phi}(\lambda)$. The spectrophotometer was set up to with a split beam to collect absorption measurements between 350-750 nm at 1 nm resolution. After thawing, samples were kept hydrated and refrigerated in the dark until processing. Reference and sample GF/F filters were mounted on a quartz glass slip in the beam paths and kept hydrated during spectral analysis with filtered seawater. Initial absorption spectra were collected for particle absorption $a_p(\lambda)$, after which filters were washed with methanol. An initial rinse was performed by allowing the sample to stand in 5 ml methanol for 1 min before filtering out methanol and extracted pigments. A longer wash (1 h) was performed in 10 ml methanol to completely remove pigmentation from the filter before rinsing filters in triplicate with 20 ml 0.2 μm FSW. A second spectral scan was performed on the filters to collect the absorption of detrital matter $a_d(\lambda)$ on the filter. $a_{\Phi}(\lambda)$ was calculated by subtracting $a_d(\lambda)$ from $a_p(\lambda)$. All filter blanks were treated like the samples to account for the influence of methanol washing on the optical characteristics of the filters. Only the reference filter was not manipulated (other than hydration) through sample processing.

To determine $a_p(\lambda)$ from the optical density of sample filters $OD_{fp}(\lambda)$, a path length correction was applied using the following equation

$$a_p(\lambda) = \frac{2.303 A_f}{\beta V_f}\left[\left[OD_{fp}(\lambda) - OD_{bf}(\lambda)\right] - OD_{null}\right] \quad \textbf{(3.1)}$$

where β is the scaling factor accounting for the added path length due to scattering in the filter, A_f is the area of the filter covered in particles, V_f is the volume of water filtered, OD_{bf} is the optical density of the blank filter, and OD_{null} is the wavelength residual

correction from where particle absorption is minimal, the average $OD_{fp}(\lambda)$ between 735-745 nm. Here we use β reported in Nelson (1998) as others provided an overestimation of $a^*(676)$,

$$\beta = [C_1 + C_2[OD_{fp}(\lambda) - OD_{null}(\lambda)]]^{-1} \quad \textbf{(3.2)}$$

where C_1 and C_2 are coefficients of least squares regression fits of measured data, 0.0 and 0.277 respectively for mixed cultures. And finally, $a_\Phi(\lambda)$ is obtained by subtracting $a_d(\lambda)$ from $a_p(\lambda)$.

$$a_\Phi(\lambda) = a_p(\lambda) - a_d(\lambda) \quad \textbf{(3.3)}$$

Following the computation of $a_\Phi(\lambda)$, spectra were smoothed using a 7 nm smoothing function and then interpolated to 1 nm resolution between 350 and 750 nm. Absorption was normalized to Chl *a* concentration, phytoplankton cell counts, and the red peak at 676 nm. Seven absorption spectra were thrown out when normalized to Chl *a* due to the extremely high $a^*(676)$, well above the theoretical limit. These were likely due to errors in sample processing.

3.3.6 Water Column Structure. We utilized measurements of temperature, salinity, downward irradiance (*Ed*) at 444 nm and 490 nm, and photosynthetically active radiation (PAR) collected by profiling floats equipped with BioOptical Sensor System(BOSS)-mini (Satlantic) to observe physical and irradiance structure of the water column at the

major process stations visited throughout the cruise. Two profiling floats (Float #8493 and #8502) were used on multiple deployment operations and recovered after leaving each station. Floats were programmed to profile from 300 m to the surface; actual profiling depth varied from 200-300 m. Deployments lasted between 91 and 123 h with ~4 h profile cycles and ~2.5 m vertical resolution.

Estimates of pycnocline depth were made using raw salinity, temperature, and pressure data from the floats to calculate density anomaly using MATLAB swstate function (http://gb11.whoi.edu/globec-dir/CTDSoftware/Oceans/swstate.m) based on (Millero & Poisson 1981). The location of the pycnocline was calculated as the depths over which the greatest change in density was observed in the upper 40 m of the water column. These depths were averaged for a single pycnocline depth at each process station.

Estimates of the euphotic depth were computed using the float onboard PAR sensor. Because % surface irradiance is usually calculated using parallel surface and depth measurements for maximum precision, we instead normalized the PAR measurements from individual profiles to the PAR measurement made closest to the surface from the respective profile.

Measurements of *Ed(*444) and *Ed(*490) were made using calibration factors provided by the float manufacturer. Ratios were constructed from parallel *Ed* values. Gausian kernel regression models, bandwidth of 2, were based only on profiles taken with adequate light, between 0400 and 1100 local time.

3.3.7 Culture Experiments - growth conditions. Culture experiments exploring the plasticity of absorption spectra for *E. huxleyi* were conducted in three sets of experiments: one using strain CCMP1516 to explore the interaction of light and viral infection; one using strain CCMP374 to explore calcium limitation and viral infection; and one using strain CCMP 374 to explore nitrogen limitation. The light and infection experiment was conducted at 50 and 250 μE and f/2 – Si media. Cultures were acclimated to light conditions in exponential growth for one week before starting the experiment. At the beginning of the experiment, 200 ml cultures were prepared in duplicate at 4 x 10^4 cells ml^{-1} and grown for two days in respective light conditions before inoculating virus strain EhV 86 at virus:host ratio of 5:1 (< 1 ml) into the infected cultures. Control cultures received similar volumes of FSW. Cultures were grown for 72 h and sampled daily.

Ca^{2+} limitation-infection experiments were conducted in NEPC artificial seawater media with 10 μM Ca^{2+} in control cultures and 0.1 μM Ca^{2+} in calcium limited cultures. Cultures were prepared and sampled in a similar manner to aforementioned light-infection experiment, inoculating infection cultures after two days of growth.

The nitrogen limitation experiment, not paired with infection, was conducted in NEPC media prepared with replete (300 μM) or limited (25 μM) NO_3. N limited cultures were acclimated to the N limited media 1 week before the experiment. At T_0, control and N limited triplicate cultures were diluted to 2 x 10^5 cells ml^{-1}. Sampling was conducted similar to the light and Ca limitation experiments.

3.3.8 Cell abundance. Counts of cell abundance in lab experiments were collected on a Beckman Coulter Multisizer 3 coulter counter. Samples were diluted 50x in FSW and run for 30 sec. FSW blanks were subtracted from particle counts. Cells were identified as particles occurring within the 2.5-7.5 nm size range.

3.3.9 Virus abundance. The total number of virus like particles (VLPs) throughout the infection experiments were counted on an Influx Mariner 209S flow cytometer using SYBR Gold stain based on the protocol developed by (Brussaard *et al.* 2010). A 1 mL subsample of each virus stock or infected culture was fixed with glutaraldehyde (final concentration of 0.5%), incubated in the dark at 4°C for 15 min, snap frozen in liquid nitrogen and stored at -80°C until processed. Samples were thawed at room temperature and diluted 50-fold in Tris-EDTA (TE) buffer (1M Tris pH-8, 0.5M EDTA, and MilliQ) in 3 ml flow cytometery tubes. SYBR Gold was diluted into the TE buffer at a 20,000:1 ratio of the commercial stock. The subsamples were agitated and incubated at 80°C for 10 min, followed by 5 min at room temperature in the dark. Each sample was analyzed for 1 min on the InFlux at a variable flow rate in μL min^{-1}. The flow rate was adjusted for each set of measurements.

3.3.10 Absorption Spectra. After determining culture cell concentrations, biomass for spectral absorption was collected on GF/F filters. Variable volumes of biomass were filtered so that 2.5 x 10^6 cells were retained on the filter. This cell number was optimized to minimize pigment shading on the filter while providing enough biomass to minimize

concern for signal noise. Samples were processed and analyzed similar to those taken at sea.

3.4 RESULTS AND DISCUSSION

3.4.1 North Atlantic Phytoplankton Blooms

Temperature profiles at the four major sampling locations (for consistency with Chapter 2, we refer to these communities as EI, EI_R, LI, PI) revealed a structured water column at each location (Figure 3.1). Temperatures were ~11°C at the surface of EI, EI_R, and LI, and 13°C at PI. The strongest thermoclines were present at EI_R and LI, while the weakest was at PI. We also used temperature and salinity measurements (See Chapter 1, Supplemental Figure 2) to calculate the position of the pycnocline at each location (respectively 15 m, 17 m, 19 m, and 25 m at EI, EI_R, LI, and PI). As salinity did not vary greatly within the water column (< 0.25 psu at all locations), water column structure was largely controlled by the temperature gradient.

Profiles of the percent surface photosynthetically active radiation (sPAR) at each of the locations is described in Figure 3.2. Light attenuated exponentially with depth at each location, however the light attenuated faster at PI than EI and EI_R. The depth of the euphotic layer, ~1% of the sPAR, was between 35-45 m at EI and EI_R, 30-40 m at LI, and 25-35 m at PI. Below these respective depths, the degree of light attenuation decreased, likely due to a decrease in the amount of light absorbing biomass below these depths (See Chapter 2, Figure 2.2 for Chl *a* and backscatter profiles at each location; these signals greatly decreased below the respective depths and were proxies for phytoplankton

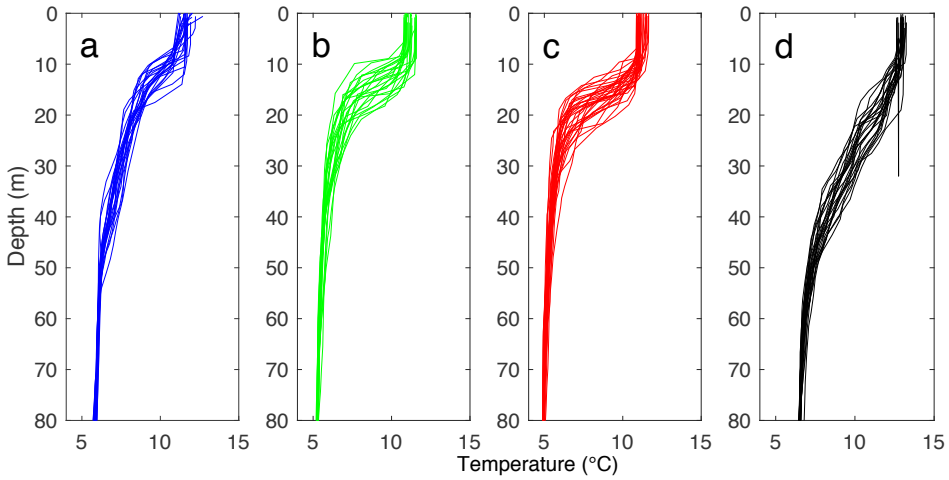


Figure 3.1 Temperature profiles from vertical profiling floats deployed at a) EI b) EI_R c) LI and d) PI. Each line represents one profile of the water column.

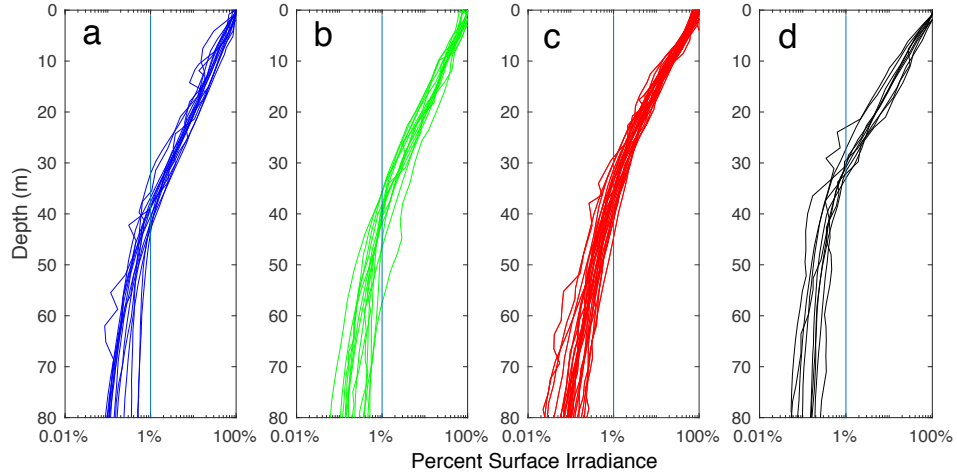


Figure 3.2 Irradiance profiles of percent surface PAR from vertical profiling floats deployed at a) EI b) EI_R c) LI and d) PI. Each line represents one profile of the water column. The 1% surface light level is denoted with a vertical blue line.

biomass.) The difference in light attenuation between stations can be explained by the differences in total phytoplankton concentrations at each station (See Chapter 2, Supplementary Figure 2.3). Total phytoplankton concentrations are up to twice as dense at PI compared to EI and EI_R.

Of note, these profiles are estimated from profiling float irradiance measurements. Therefore simultaneous sPAR and at depth PAR were not collected.

Rather they are based on the sPAR once the floats reached the surface. This likely increased the variability of the % sPAR measurements due to variations in incident light over the duration of individual profiles, thus explaining some of the variability in the profiles. Ship CTD PAR measurements were not used due to the impact of ship shadow on the PAR profiles.

PI was also differentiated from the other stations when looking at the relative attenuation of downward light irradiance at 440 *Ed*(440) to 490 *Ed*(490) (*Ed*440:*Ed*490) throughout the water column (Figure 3.3). Generally, the ratio decreased with depth at

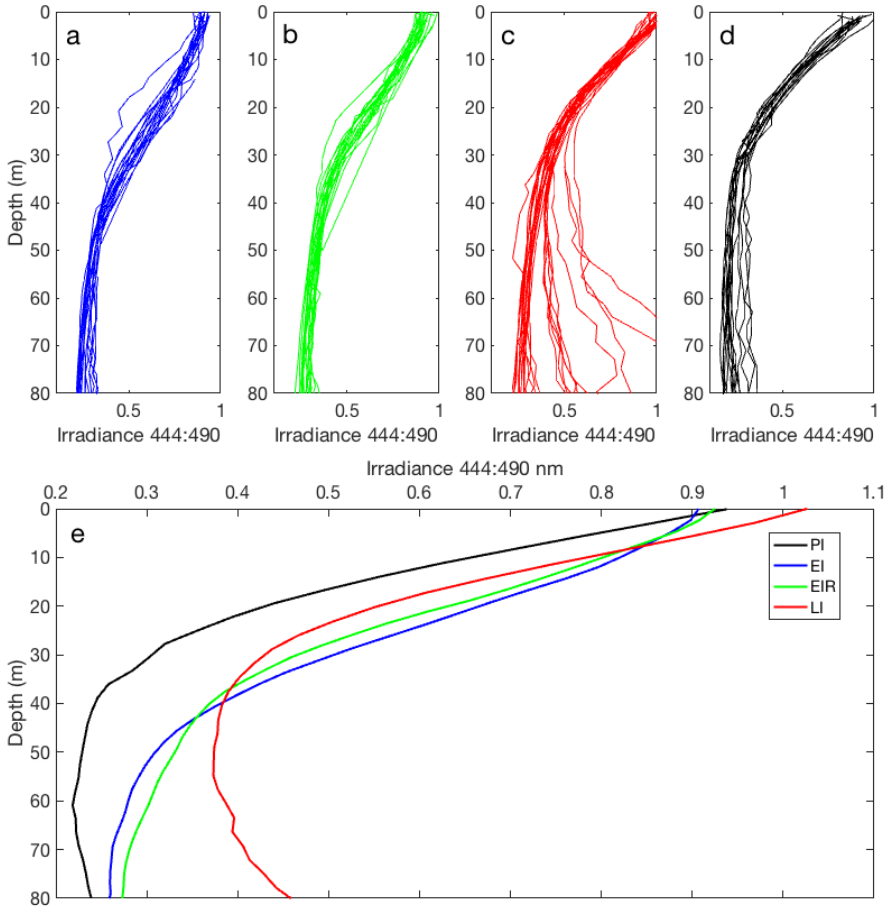


Figure 3.3 Irradiance ratios comparing the incident light irradiance of 440:490 nm wavelengths with depth from vertical profiling floats deployed at a) EI b) EI_R c) LI and d) PI. Each line represents one profile of the water column. The 1% surface light level is denoted with a vertical blue line. e) Gausian kernel regressions of all profiles at EI, EI_R, LI and PI to compare differences in the ratios with depth.

all stations, revealing that there was relatively more *Ed*(490) available at depth than *Ed*(440). Inflection points where the ratios became more consistent with depth were observed at different depths for each station, occurring lowest at EI (45 m), and EI_R (35 m), and higher in the water column at LI and PI (25 m). Station PI observed the greatest change and decrease in *Ed*440:*Ed*490 with depth, being 0.3 at 30 m in comparison to 0.5 at EI and EIR. This shows a substantial difference in the spectral characteristics of the light field observed between stations in addition to the differences in attenuation observed by PAR.

A total of 77 phytoplankton absorption spectra [$a_\Phi(\lambda)$] were collected in the North Atlantic, largely from the major sampling locations. Normalizing the $a_\Phi(\lambda)$ to fluorescence measured Chl *a* for the spectral Chl specific absorption [$a_\Phi^*(\lambda)$], the range of a_Φ^* at 440 nm [a_Φ^* (440)] for all spectra spanned over an order of magnitude (0.03-0.44 m^2 (mg Chl *a*)$^{-1}$ (Figure 3.4a). Identifying the spectra by depth ranges within the

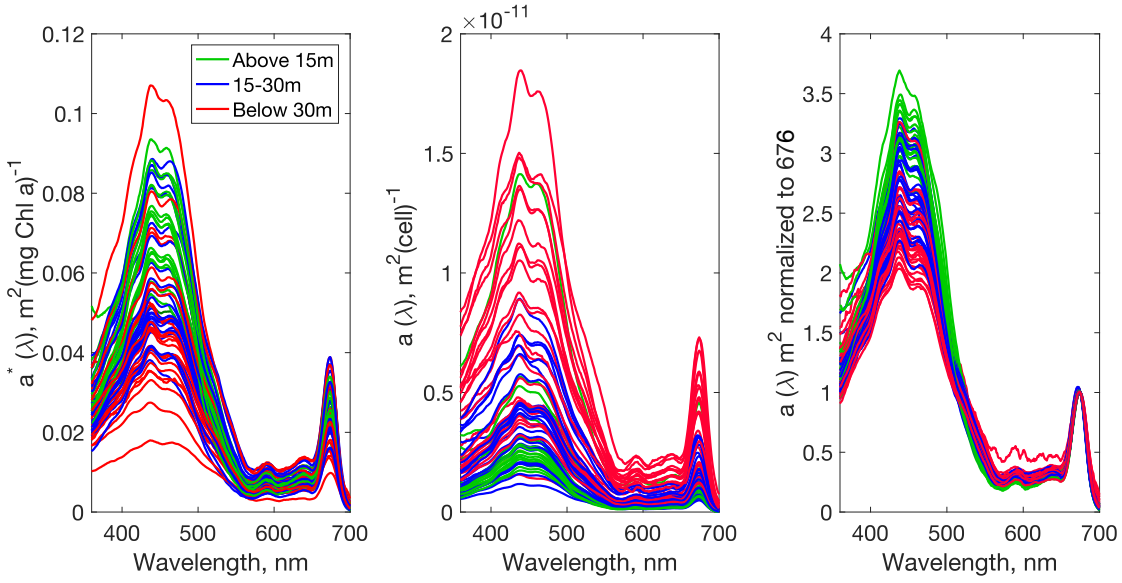


Figure 3.4 Spectral Absorption normalized to a) Chl a, b) eukaryotic phytoplankton cell concentration and c) chl a peak at 676 nm. Spectral color indicates depth of biomass sampling: green < 15 m, blue 15-29 m, and red 30-50 m.

euphotic zone, $a_{\Phi}^*(\lambda)$ collected at shallow (<15m) depths, exhibited the highest absorption per unit Chl in the 400-500 nm with the exception of several large absorption values from deep (30-50 m) and mid (15-29 m) water column spectra. Mid-water column spectra tended to have flatter $a_{\Phi}^*(400\text{-}500)$ than shallow -water column spectra, while the deep-water spectra tended to be the most suppressed. Six spectra that exhibited high $a_{\Phi}^*(676)$, well above the established theoretical limit of 0.033 m^2 mg Chl a^{-1}, are likely a result of experimental error (Bricaud *et al.* 2004), and are excluded from Chl *a* normalized analysis. Otherwise, the suppressed absorption observed in the mid- and deep-water column spectra are likely due to a greater amount of pigment packaging within the cells, with packaging decreasing the overall efficiency of absorption per unit Chl.

Normalized to eukaryote cell concentration, the opposite trend is observed, with deep-water spectra exhibiting the highest absorption per cell, mid-water column spectra having intermediate $a_{\Phi}(\lambda)$ $cell^{-1}$, and the shallow-water spectra having the lowest $a_{\Phi}(\lambda)$ $cell^{-1}$ (Figure 3.4b). Cells acclimated to high light conditions are likely to have a lower cell pigment quota than those growing at depth and removed from the light, where more pigment is needed to efficiently harvest light. These observations suggest there is a limitation to the mixing within the water column.

The $a_{\Phi 676}(\lambda)$ normalized to the Chl red peak at 676 nm [$a_{\Phi 676}(\lambda)$] is highest for shallow-water spectra and lowest for deep-water spectra between 400-515 nm, but the relationship flips at 515 nm, where deep-water spectra are elevated relative to shallow-water spectra (Figure 3.4c). This is consistent with the cell normalized spectra in that a higher pigment quota for cells growing at depth would increase the pigment packaging

within cells, with enhanced packaging leading to a suppression of absorption at the wavelengths of highest absorption efficiency.

To observe the average changes in $a_\Phi(\lambda)$ over the water column, we took the difference between the respectively normalized deepest spectra and shallowest spectra collected in each profile of the water column (Figure 3.5). The average difference $a_\Phi^*(\lambda)$ was characterized by negative minimum peaks at 440 and 456nm (Figure 3.5a). Above 540 nm, the difference was steadily slightly positive but dipped for a second smaller minimum peak at 676. For spectra normalized to cell concentration, the average difference $a_\Phi(\lambda)$ cell^{-1} exhibited a maximum peak at 440 nm before approaching zero at 560 nm, with a second small peak at 676 nm (Figure 3.5b). The difference remained positive over the entire spectrum. The average difference $a_{\Phi 676}(\lambda)$ was shaped similarly to the average difference-$a_\Phi^*(\lambda)$ (Figure 3.5c), presenting a degree of consistency in the changes in absorption characteristics with depth.

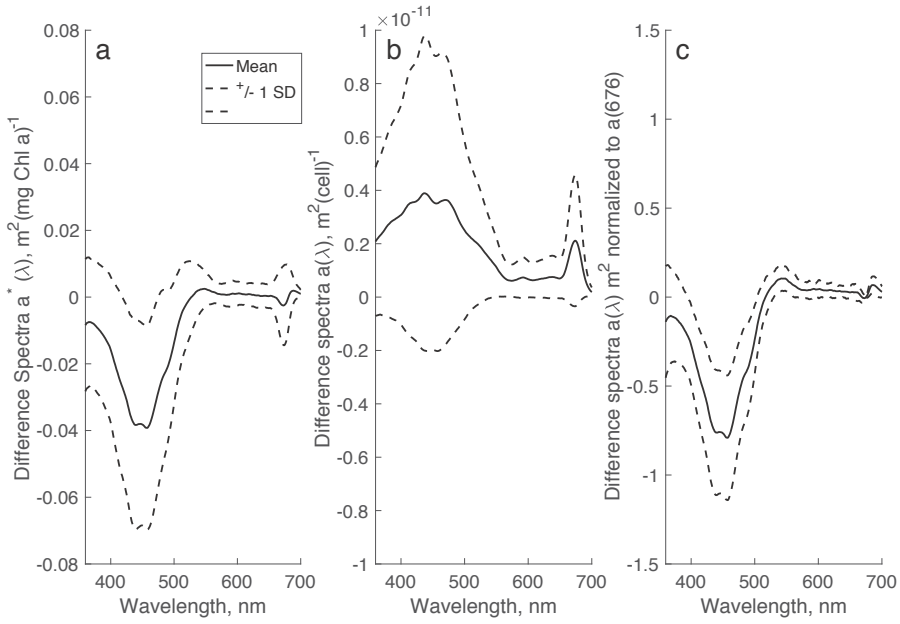


Figure 3.5 Difference absorption spectra representing the average difference between the deepest and shallowest $a_\Phi^*(\lambda)$ from individual profiles of the water column for spectra normalized to a) Chl a; b) eukaryote phytoplankton cell concentration; c) 676 nm. Dotted lines indicate ±1SD.

The suppression of absorption peaks induced by increased pigment packing within cells is apparent in the relationship between depth and the pigment absorption ratio at $a_{\Phi}(440)$ and $a_{\Phi}(676)$ ($a_{\Phi 440:676}$) (Figure 3.6a). For $a_{\Phi}(\lambda)$ near the surface, $a_{\Phi 440:676}$ was ~1.5x higher than those approaching 40 m and decreased linearly with depth. A similar relationship also manifested in the production of the photoprotective pigment diadinoxanthin (Ddx) with depth. Ddx:Chl *a* ratio values were highest in the 10 and 20 m depth bins and decreased exponentially with depth (Figure 3.7). Because Ddx:Chl *a* represents long term acclimation to light levels (hours-days), these values represent long term photoacclimation (Bidigare *et al.* 1987a). The Nitrate+nitrite (NO_3+NO_2) concentration also explained a smaller amount of the variability in $a_{\Phi 440:676}$, which increased with decreasing NO_3+NO_2 (Figure 3.6b). Seeing as N limitation can induce changes in $a_{\Phi}(\lambda)$ (Berges *et al.* 1996), this may indicate a small but detectable amount of N limitation influencing spectral characteristics within our observations.

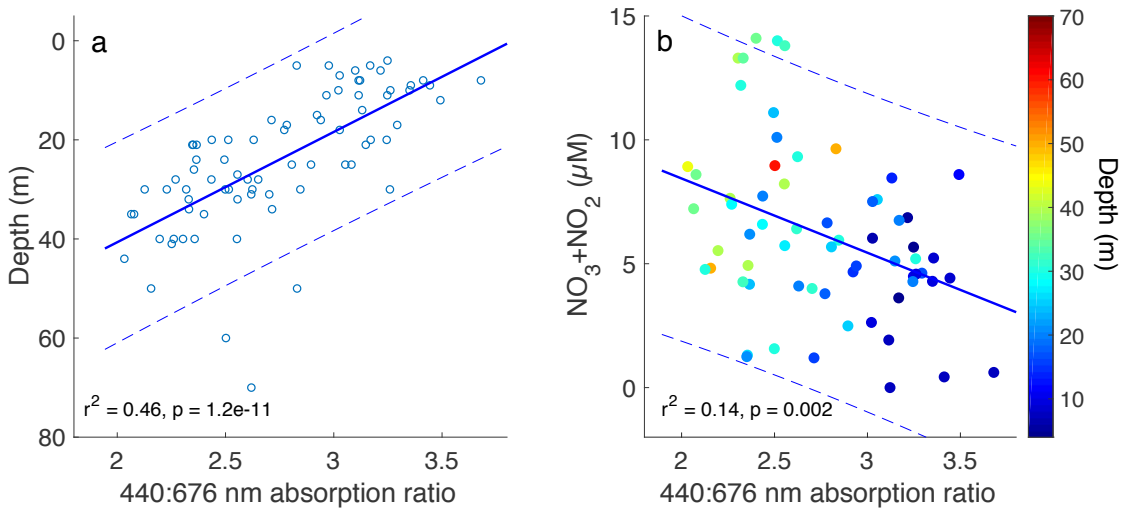


Figure 3.6 Correlation between the ratio of phytoplankton absorption peaks at 440:676 nm with a) depth (r^2 = 0.46, p = 1.26e^{-11}_ and b) NO_3+NO_2 (r2 = 0.14, p = 0.002). Solid line indicates linear regression. Dotted lines indicate 95% confidence intervals. Color overlay on (b) represents depth of observation.

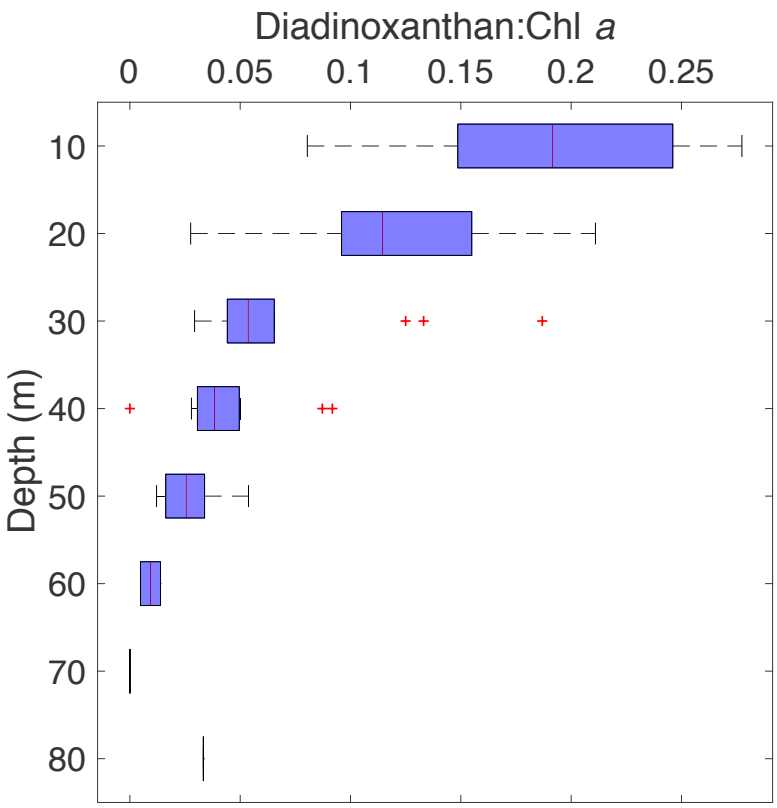


Figure 3.7 Depth binned water column diadinoxanthan concentration normalized to Chl *a*. Bins are 10 m. Box bounds denote 25% and 75% quantiles around the median (centroid). Vertical capped lines indicate max and min data values. Outliers indicated with (+).

Because community structure can similarly influence the qualities of $a_{\Phi}(\lambda)$, we looked at the difference in community structure among the process stations visited throughout the cruise. We applied a MANOVA with canonical variates analysis to test for differences among the communities based on the contribution of taxonomic groups to total Chl *a* and plotted the first two canonical axes to observe the relative similarity (Figure 3.8). Each community occupied a distinct region in the variate space with minimal overlap between stations. EI and EI_R shared canonical space and were most separated from the other communities on canonical axis 1 (c1) and from each other on canonical axis 2 (c2). PI was most separated from the other groups on c1 but has a greater

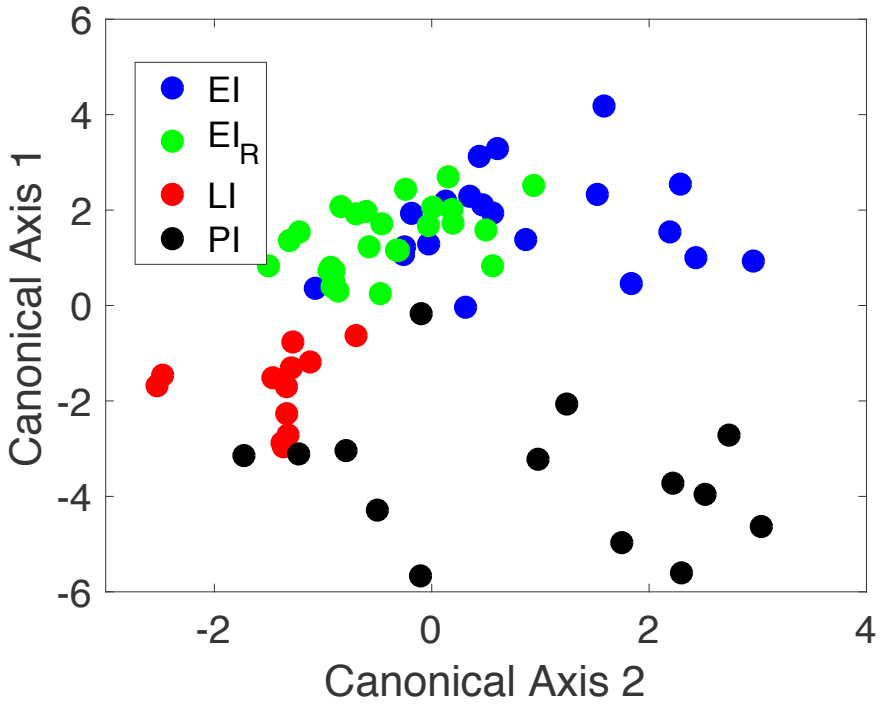


Figure 3.8 Canonical variance analysis of the taxonomic contribution to Chl *a* among process stations: EI – blue, EI_R – green, LI – red, PI - black. Composition is estimated by CHEMTAX analysis of pigment ratios.

spread along c2. The differences observed along these axes can likely be ascribed to: the presence of diatoms at EI and PI, pulling these communities along c2; increases in prasinophytes and chlorophytes at LI and PI; decreases in cryptophytes, moving these communities down c1 (See Chapter 2, Supplemental Figure 2.6 for average taxonomic assemblages at each station).

These differences in community structure between stations do not appear to have a significant impact on $a_\Phi(\lambda)$ when comparing characteristics within and below the mixed layer. Looking at indicators of pigment packaging, a_Φ*(440) was similar between stations (Figure 3.9a). Also, while median values were consistently lower below the mixed layers, no statistical differences were observed. In fact, at EI and EI_R a_Φ*(440)

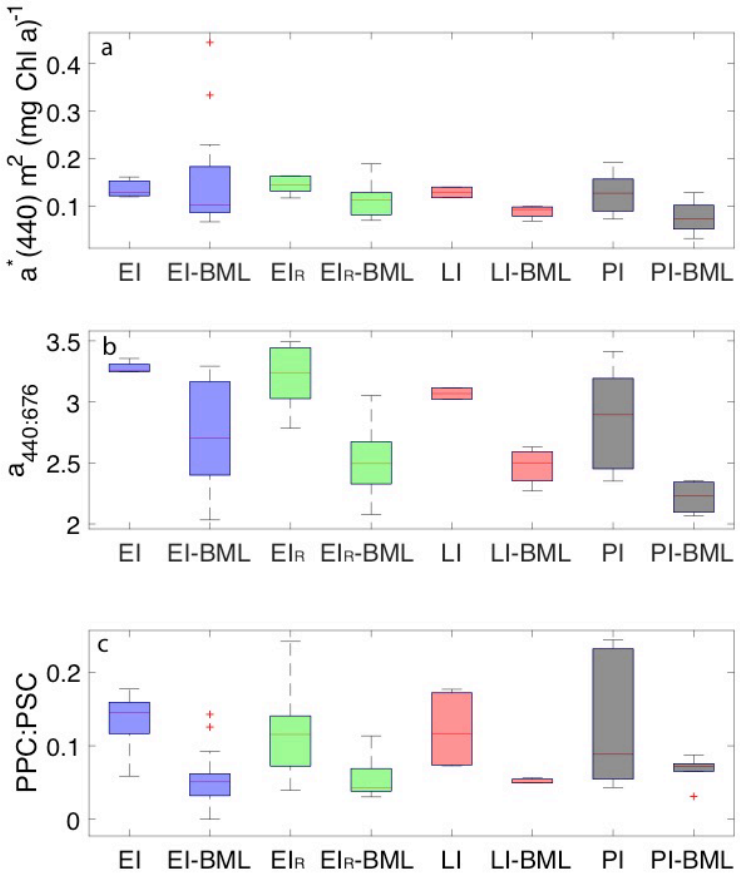


Figure 3.9 Comparison of a) Chl *a* specific absorption at 440 nm [$a^*(440)$], b) the 440 to 676 ($a_{440:676}$) absorption peak ratio, and c) the ratio of photoprotective carotenoids to photosynthetic carotenoids (PPC:PSC)between the locations EI, EI_R, LI and PI. Each location was grouped into observations collected within and below (BML) the surface mixed layer. Box bounds denote 25% and 75% quantiles around the median (centroid). Vertical capped lines indicate max and min data values. Outliers indicated with (+).

have higher values binned below the mixed layer, leading to a wider range in $a_\Phi^*(440)$. The similarity across stations suggests that the differences in community composition were not great enough to effect $a_\Phi^*(\lambda)$. Similarly, $a_{\Phi 440:676}$ does not differ between locations, though variance was greater at PI than the other locations (Figure 3.9b). However, values were consistently lower below the mixed layer than above. This

suggests that cells below the mixed layer are experiencing greater pigment packaging. The variance observed in the mixed layer at PI is likely due to the deeper mixing layer depth (MLD) (25 m) compared to EI, EI_R, and LI (15 m, 17 m, and 19 m, respectively). Likewise, the ability of deeper waters to mix closer to the surface at EI and EI_R (shallower) likely raised $a_{\Phi 440:676}$ values here.

Comparing the ratio of photoprotective carotenoids to photosynthetic carotenoids in a similar manner, values ranged between 0 and 0.25 across stations (3.9c). Significant differences were observed at EI and EI_R above and below the mixed layer depth, but not at LI and PI. These decrease in PPC:PSC below the mixed layer help explain the aforementioned $a_{\Phi 440:676}$, suggesting that the pigment packaging seen at depth is due to photoacclimation and not changes in taxonomic assemblage with depth. Interestingly, there was generally a greater range in PPC:PSC ratios observed in the mixed layer compared to below. This suggests that below the mixed layer, there is little need to produce photoprotective pigments beyond basal levels, whereas in the mixed layer, active modulations of PPC:PSC ratios are utilized to optimize photoprotective and photosynthetic capacity of cells.

There were differences in the ratio of phytoplankton absorption at 440 and 490 nm ($a_{\Phi 440:490}$) between stations (Figure 3.10). Generally, $a_{\Phi 440:490}$ decreased with depth, which can be interpreted as an increase in the photosynthetic carotenoids as light becomes more limited with depth relative to Chl *a*. However, the $a_{\Phi 440:490}$ was also lower at PI stations than EI and EI_R for a given depth, revealing relatively higher absorption at 490 nm. Both of these observations reflect the changes in the *Ed*(444):*Ed*(490) seen throughout the water column and between stations. With a light field that becomes

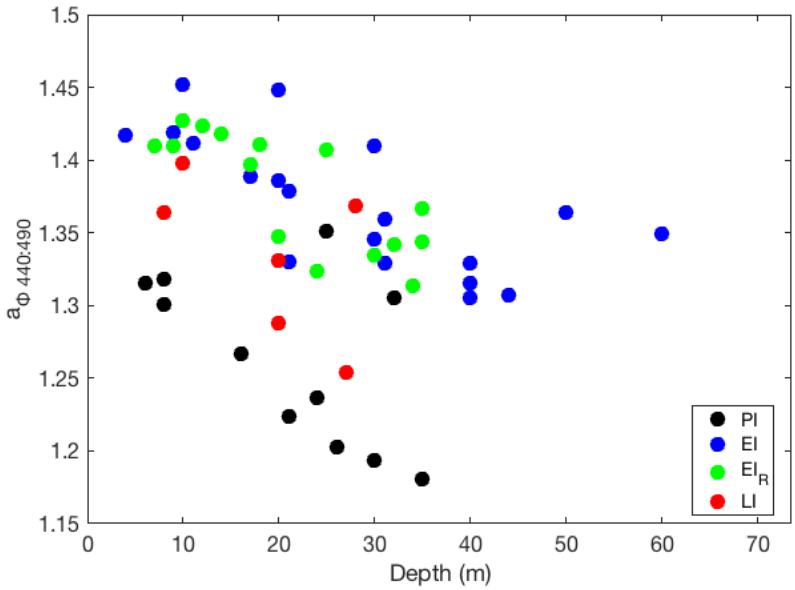


Figure 3.10 Relationship between the ratio of phytoplankton absorption at 440 to 490 nm with depth at stations EI (blue) EI_R (green) LI (red) and PI (black).

relatively inflated in *Ed*(490) with depth (like that seen at PI), phytoplankton are preferentially taking advantage of this part of the light spectrum by increasing the absorptive capacity in this range.

The influence of community assemblage on $a_{\Phi 440:676}$ independent of stations of origin was also explored utilizing principle component analysis to maximize the variance in the data among taxonomic % contribution to Chl *a* (Figure 3.11). The first two principle components explained 45% and 28% of the total variability in the data. On these two principle components, haptophytes, diatoms, type 2 diatoms, and pelagophytes eigenvectors had the strongest pull. First, a color overlay of $a_{\Phi 440:676}$ was added over the component scores for each observation (Figure 3.11a). This revealed higher values of $a_{\Phi 440:676}$ tending to the positive directions of pc1 and pc2. This suggests there may be a relationship between the relative proportions of taxonomic groups and $a_{\Phi 440:676}$. However, we also overlaid the observation depths in the principle component space and

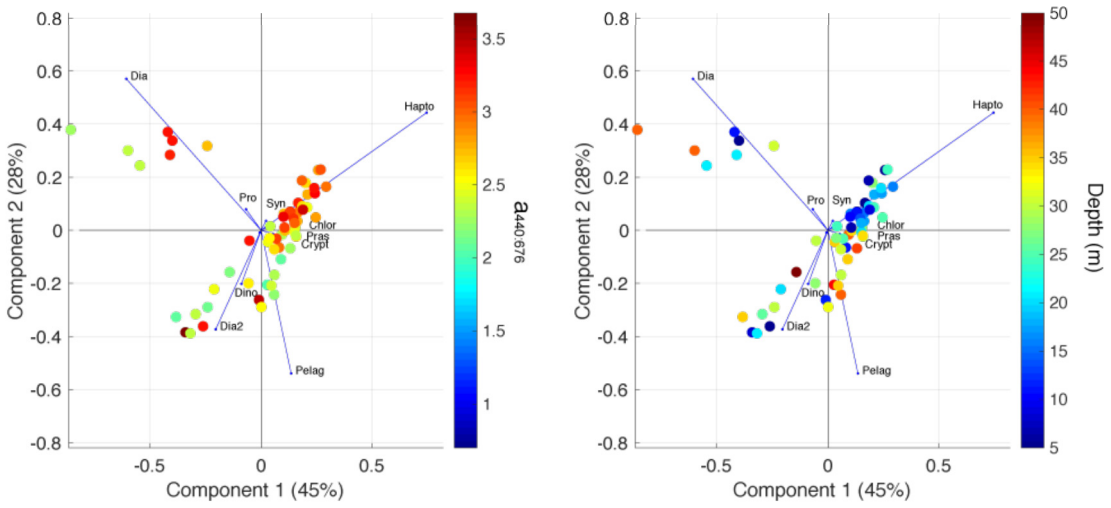


Figure 3.11 Principle component analysis showing the first two principle components maximizing variance in observations of taxaonomic % contribution to total Chl a throughout the study. Taxonomic grouping includes Diatoms (Dia), Type 2 diatoms (Dia2), pelagophytes (Pelag), haptophytes (Hapto), *Prochlorococcus* (Pro), *Synechococcus* (Syn), chlorophytes (Chlor), prasinophytes (Pras), cryptophytes (Crypto), and dinoflagellates (Dino). Percentages represent the amount of variability explained along the component axis. a and b show same PCA but with a) Observation component scores are overlayed with $a_{440:676}$ of the measurements. b) Observation component scores are overlaid with depth of measurements.

found a similar trend with decreasing depth tending in the positive directions of pc1 and pc2, suggesting changes in composition with depth, and that the trend between taxonomic composition *and* $a_{\Phi 440:676}$ occurs at a depth gradient (Figure 3.11b). This reveals the difficulty in separating the two variables (depth and composition) to measure the impact of composition on $a_{\Phi 440:676}$ by itself.

3.4.2 Lab Experiments

Given the dominance of haptophytes within the blooms investigated in the North Atlantic, as well as the intense coccolithovirus infection within these blooms (Laber *et al.* 2018), we conducted a series of growth experiments to observe the plasticity of $a_{\Phi}(\lambda)$ for the individual haptophyte species, *E. huxleyi,* when grown under an array of different environmental stresses including light limitation, nutrient limitation, and infection by

EhVs. These stresses were chosen based on their known impacts on cellular pigment quotas and ratios (Stolte *et al.* 2000; Llewellyn *et al.* 2007).

Environmental stresses produced differing independent and mixed effects on cell growth rates. Looking at light limitation, EhV infection, and the combination, light limitation at 50 µE compared to 250 µE decreased growth rates over the growth period while EhV infection also decreased the net growth rate and a negative net growth rate by 72 hours post infection (hpi) (Figure 3.12a). The combination of light limitation and EhV infection resulted in an earlier departure between the control and infected groups at 50 µE (24 hpi), relative to the 250 µE groups. For calcium limitation, growth rates did not differ between the 10 µM and 0.1 µM Ca^{2+} groups (Figure 3.12b); however, the combined effect of EhV infection and Ca^{2+} limitation resulted in cultures having the lowest net growth and crashing by 96 hpi. Nitrogen limitation reduced cell growth rate compared to N replete cultures; however, net growth remained positive throughout the growth period (Figure 3.12c).

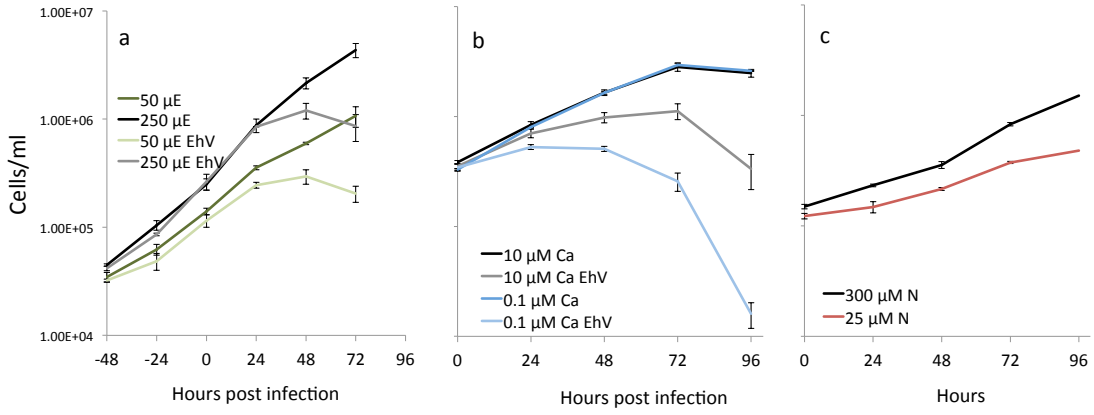


Figure 3.12 Dynamics of *E. huxleyi* cell growth under various conditions: a) *E. huxleyi* strain 374 under high light (250 µE) and low light (50 µE) with and without EhV infection (EhV). Hours post infection (hpi) indicate the time elapsed since culture inoculation with virus. Observations started at -48 hpi. b) *E. huxleyi* strain 1516 with calcium replete (10 µM Ca^{2+}) and calcium limited (0.1 µM Ca^{2+}) media, with and without EhV infection. c) *E. huxleyi* strain 1516 with nitrogen replete (300 µM N) and nitrogen limited (25 µM N) media. Hours indicates time since cultures introduced to new media.

We compared the $a_{\Phi}(\lambda)$ for each experimental group after 72 hours of growth, with $a_{\Phi}(\lambda)$ exhibiting a range of absorbance values when normalized to the cell concentration (Figure 3.13a). While spectral shapes were largely retained across experimental groups, differences in absorption magnitude and peak suppression were

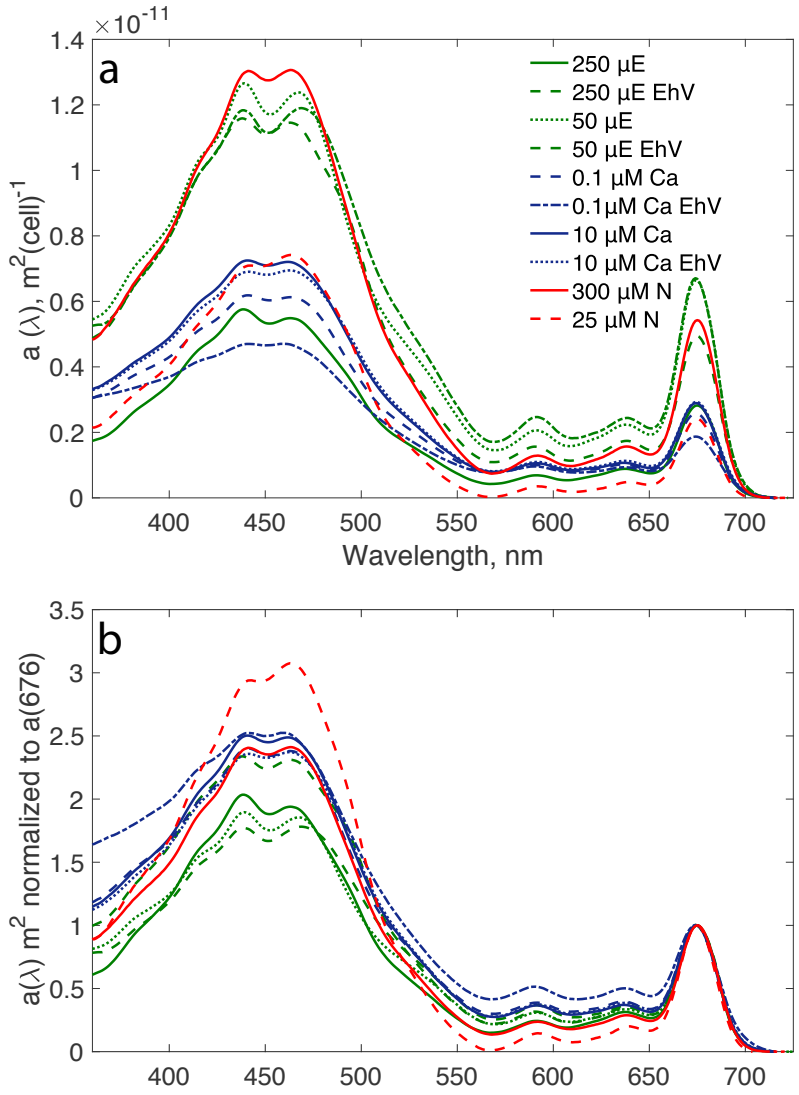


Figure 3.13 Phytoplankton absorption spectra between 400 and 700 nm collected at 72 hours of growth in respective conditions. a) Normalized to cell concentration for cell specific absorption. b) Normalized to the Chl *a* red absorption peak at 676 nm. For both a) and b) High and low light with mixed infection experiment is show in green. Calcium replete and limited with mixed infection experiment is shown in blue. Nitrogen replete and limited experiment is shown in red.

observed with the responses to environmental stresses both increasing and decreasing absorption. Low light, viral infection, and a combination of the two, increased $a_{\Phi}(\lambda)$ cell^{-1} by a factor of two relative to the high light control. Normalized to 676 nm, there was less difference between spectra, with only the 50 μE cultures having a relatively greater $a_{\Phi}(440)$ peak (Figure 3.13b). It was previously shown that total cell pigment can double when grown at 50 μE as opposed to 200 μE (Stolte *et al.* 2000), which may explain the differences in the $a_{\Phi}(\lambda)$ cell^{-1}. If pigment concentrations are different per cell, the relative similarity in magnitude between light regimes would likely be explained by either differences in cell size or possibly the method of photophysiological adaptation (increasing the antenna size for photosystem II reaction centers or increasing the number of reaction centers) (Falkowski & Raven 1997).

Calcium limitation had little influence on absorption by itself (Figure 3.13). Both magnitude and spectral shape were maintained between 0.1 and 10 μM Ca^{2+} regimes. This is not surprising, as the efficiency of photosynthesis does not change when growing *E. huxleyi* under Ca^{2+} limiting conditions (0.1 μM) (Trimborn & Langer 2007).

EhV infection had mixed influences on absorption (Figure 3.13). The $a_{\Phi}(\lambda)$ cell^{-1} was dramatically increased comparing infected and non-infected 250 μE group, doubling at $a_{\Phi}(440)$ from 0.6 to 1.2 m^2 cell^{-1}. There was little difference at 50 μE, however. Interestingly, when combined with Ca^{2+} limitation, the two acted to suppress $a_{\Phi}(\lambda)$ cell^{-1}, suggesting a lower cellular pigment content.

The greatest differences occurred between the N replete and limited cultures (Figure 3.13). N-limited absorption per cell decreased by a factor of two relative to the N replete, while $a_{\Phi}(676)$ normalized spectra increased by 0.5 at $a_{\Phi}(440)$. Indeed, N

limitation has previously been shown to increase the $a_{\Phi}^{*}(\lambda)$ in other taxa (Berges *et al.* 1996). This is likely a coupled relationship, with lower total pigments in N limited cells increasing $a_{\Phi}^{*}(\lambda)$, understanding that pigment packaging suppresses absorption peaks, and decreasing $a_{\Phi}(\lambda)$ cell^{-1}, given there is less pigment (Stolte *et al.* 2000).

Of note, we used *E. huxleyi* CCMP 374 for the light/infection experiment and CCMP 1516 for the Ca/infection and N limitation experiments. These two strains have been extensively used in laboratory EhV infection experiments. Strain 1516 has also previously been observed to undergo shifts in cell pigment content and ratios throughout the infection cycle (Llewellyn *et al.* 2007). Cell pigment quotas are known to be variable between strains, varying by as much as an order of magnitude when grown under similar conditions (Stolte *et al.* 2000). This helps explain the large difference in $a_{\Phi}(\lambda)$ cell^{-1} observed between control groups in each experiment.

These experiments exemplify the inherent variability in spectral characteristics within an individual phytoplankton species. Understanding this, open ocean communities that share similar taxonomic composition have the potential for significant variations in $a_{\Phi}(\lambda)$. Differences in these $a_{\Phi}(\lambda)$ will likely be useful in interpreting physiological stresses/adaptations of phytoplankton within these blooms. As changes in taxonomic composition likely can influence $a_{\Phi}(\lambda)$ to a similar or greater degree, having knowledge of composition, as well as other variables such as nutrient concentrations the increases the utility of $a_{\Phi}(\lambda)$ in assessing physiological states.

3.5 CONCLUSIONS

Photoacclimation states and spectral absorption properties had similar characteristics between the water masses observed. The similar water column structures between the stations likely contributed to this, with the least similar location being PI, having the weakest thermocline, as well as the greatest light attenuation with depth. While community structure did vary between stations, the differences were not great enough to have a dominant influence over $a_{\Phi}(\lambda)$ in the water column. Rather photoacclimation through increased pigment concentrations and pigment packaging explained the differences in the absorption parameters explored. It is evident through both the field and culture studies that environmental factors have a profound influence on spectral characteristics and therefore must be taken into consideration when using $a_{\Phi}(\lambda)$ to evaluate phytoplankton diversity and primary production.

Chapter 4: Expanding the Known Biogeography of Coccolithovirus Infection to the Global Ocean

4.1 ABSTRACT

Coccolithovirus infection of *Emiliania huxleyi* has become a prominent model system for exploring viral infection in phytoplankton, but exploration of EhV infection in naturally occurring *E. huxleyi* populations has thus far been limited to the North Atlantic and marginal seas. Here, we investigate the occurrence of EhV infection throughout the global ocean with specific focus on the North Pacific. We use the Tara Oceans metagenome assembly database to identify associated *E. huxleyi*- and EhV-like genes in each major ocean body and additionally explore an individual location, Station 82 of the Tara Oceans study in the South Atlantic, where a mesoscale coccolithophore bloom harbored prominent EhV assembly sequences. Extensive diversity (30-96% similarity to known EhV genes) was observed among EhV-like genes throughout the global dataset. We also investigated the genetic diversity of EhV-derived Major Capsid Protein (MCP) and Serine Palmitoyltransferase (SPT) amplicons associated with (and replicating within) *E. huxleyi* biomass in diverse regions of the North Pacific (California upwelling zone near Santa Barbara, Station ALOHA, and off-shore Japan) and compare them to representatives of North Atlantic EhVs. MCP sequences were found to be conserved and similar to the North Atlantic counterparts while SPT sequences were divergent, likely due to the functional role SPT plays in regulating different infection strategies among EhVs. Finally, culture-based infections indicated susceptibility of North Pacific hosts to North Atlantic EhVs, indicating compatibility across ocean basins. These results reveal that *E. huxleyi* is not only commonly

infected by EhVs throughout the global ocean but that infection occurs at a wide range of host population densities and with genetically diverse EhVs.

4.2 INTRODUCTION

The cosmopolitan phytoplankton *Emiliania huxleyi* is a globally abundant species with an influence on both organic and inorganic carbon cycling (Tyrrell & Merico 2004). Due to the production of calcium carbonate coccoliths that surround the cell, blooms of the coccolithophore can be identified by satellite based on high reflectance of the coccoliths (Brown & Yoder 1994). Open ocean blooms in the North Atlantic can reach concentrations of 10^4 while coastal and marginal seas can reach 10^5 cells/ml (Holligan *et al.* 1993; Turkoglu 2008). In the South Atlantic extending off the southern tip of Argentina a region known as the Great Calcite Belt also produces large blooms of *E. huxleyi* (Balch et al. 2016). Since 1997, the Bering Sea has also regularly experienced extensive summer blooms spanning as much as 200,000 km^2 over a two month time period (Ladd *et al.* 2016). Meanwhile, in areas of the ocean such as the western North Pacific, concentrations have been observed between 1-10^2 cells/ml (Hagino *et al.* 2005; Minoshima *et al.* 2007). Even waters in oligotrophic subtropical gyres host low concentrations of *E. huxleyi* over the annual period (Cortés *et al.* 2001). The only regions of the ocean in which *E. huxleyi* is not readily observed are in extreme latitudes of the Arctic and Antarctic (Paasche 2001).

E. huxleyi populations are commonly susceptible to infection by a double stranded DNA virus called coccolithovirus. These viruses were discovered in blooms of *E. huxleyi* in the English Channel and are thought to only infect *E. huxleyi*, as suggested by

the very involved method of infection they use to replicate and propagate from the cell (Bratbak *et al.* 1993; Mackinder *et al.* 2009). This involves the use of a pathway transferred between host and virus as well as delaying and precisely triggering an apoptotic pathway for cell lysis (Bidle *et al.* 2007; Monier *et al.* 2009; Vardi *et al.* 2009). In lab cultures, infection can terminate populations in as little as 72-96 hours, while mesocosm experiments and blooms in the ocean terminate within days to weeks after acute infection is observed (Bidle *et al.* 2007; Vardi *et al.* 2012). Infected blooms are frequently observed in different parts of the North Atlantic, from Norwegian Fjords to waters north of Bermuda, as well as in marginal seas (Vardi *et al.* 2009). In the Black Sea, over 7000 years of host-virus interaction have been observed through the amplification of ancient DNA preserved in layers of sediment (Coolen 2011). Astoundingly, this interaction has not been described elsewhere in the world ocean, most notably the Pacific Ocean from which the type *E. huxleyi* host strain (CCMP1516) was isolated and for which a complete genome was recently completed (Read *et al.* 2013).

The long standing hypothesis that '*everything is everywhere'* in the microbial world, first introduced nearly a century ago, suggests that even when microbes are not numerally abundant, transportation and the ability to sit out unfavorable conditions has led to the potential that microbial communities are very similar on a global scale (Bass & Boenigk 2011). However, biogeographic patterns in microbes are regularly observed, with natural boundaries posing limits to connectivity between hospitable waters. For coccolithovirus, with a host that is globally distributed and abundant, it is feasible that the virus would share a common biogeography despite the current lack of observations outside the North Atlantic.

In addition to local transport through infected populations, multiple vectors also exist for extracellular transportation of coccolithoviruses. They can be incorporated into aerosols for atmospheric transport and deposition into new bodies of water (Sharoni *et al.* 2015) and also have the capacity to stay virulent after ingestion and incorporation into gut content of copepods with an increased half-life of infectivity, increasing transmission potential (Frada *et al.* 2014). Furthermore, ship discharging of ballast waters after trans-ocean voyages provides a bypass for natural geographic boundaries to propagation, introducing viruses to new regions (Drake *et al.* 2007). However, throughout vast swaths of the ocean such as the gyres, where *E. huxleyi* concentrations are low, cell densities may not be high enough to transmit viruses effectively within the resident populations due to low encounter rates (Wilcox & Fuhrman 1994). Because very little is known about EhV infection dynamics at low cell densities, it is difficult to gauge whether/how infection occurs in these scenarios.

Viral infection is estimated to overturn over a quarter of photosynthetically fixed carbon in the ocean creating a significant biogeochemical pathway (Weitz *et al.* 2015). It has recently been shown virally infected *E. huxleyi* blooms in the North Atlantic can contribute to the downward vertical flux of carbon out of the sunlit surface ocean into the mesopelagic (Laber et al 2018). Given the significance that this interaction can have on local biogeochemistry in the North Atlantic, it is important to how the global population of *E. huxleyi* is impacted by coccolithovirus infection.

DNA-based molecular detection of viruses provides a reliable method of probing environmental samples for specific viruses as well as discerning the diversity that exists within a population (Larsen *et al.* 2008). While there is no universal 16S or 18S type

molecular sequence for identifying viral phylogeny, sequences have been identified within viral groups that can be used to discern species level differences. For Phycodnaviridae, the structurally conserved major capsid protein gene has been identified as a good marker for genetic distance within the group, and can be used to target specifically for Coccolithovirus (Schroeder *et al.* 2003; Larsen *et al.* 2008). Amplification and sequencing of MCP genes in a geographic context would allow observation of the connectivity and genetic distance between viral populations in different regions of the ocean.

Genetic markers can also target functional differences in viruses. During infection of *E. huxleyi*, the virally encoded production of serine palmytoyltransferase provides a rate-limiting step in the vital production of viral glycosphingolipids, which regulate the induction of programmed cell death (Ziv *et al.* 2016). Controlling the expression of this pathway is critical to regulating infection in the environment and structural differences in the SPT have been shown to influence infection dynamics. In particular, the placement and variations of a linker region between two subunits of the protein impacts the lytic infection dynamics (Nissimov, in prep). Cataloguing the diversity and distribution of the SPT linker stands to provide key insight into coccolithovirus competitive ecology and infection strategies in different environments.

While probing discrete samples for targeted DNA sequences provides a precise way of confirming presence, the presence of large metagenome databases such as that from the Tara Oceans global ocean survey provide an opportunity to explore a large swath of genetic data from across the ocean with minimal investment in sampling (Brum *et al.* 2015; Pesant *et al.* 2015; Sunagawa *et al.* 2015). The Tara Oceans dataset hosts

metagenome data from samples collected between 2009-2013 at 210 locations across the global ocean. Further, the assembled gene sequence dataset provides an accessible product for comparison to reference *E. huxleyi* and EhV genomes and for identification of relevant genetic information.

In this study, we investigate the global distribution of coccolithovirus infection and in the EhV genome by using direct EhV infection experiments, targeted EhV gene amplification and the Tara Ocean metagenome database. Our targeted gene amplification is concentrated on regions in the North Pacific Ocean that represent very different environments for *E. huxleyi*, ranging from a coccolithophore bloom in the Santa Barbara Channel to the oligotrophic North Pacific Gyre at Station ALOHA. Our hypothesis is that coccolithovirus infection is a globally important process that impacts *E. huxleyi* populations in diverse surface ocean environments and across coccolithophore abundances that span orders of magnitude.

4.3 MATERIALS AND METHODS

4.3.1 Retrieval and analysis of Tara Oceans sequences. Tara Oceans metagenome assemblies were downloaded from the European Nucleotide Archive (https://www.ebi.ac.uk/ena/about/tara-oceans-assemblies). These assemblies were used instead of the Tara Oceans Viromes database for several reasons. *First*, there is better geographic coverage. Also the size fractionation (larger than 0.22 μm) was more appropriate for capturing cell associated viruses rather than free viruses. EhVs are also ~0.2 μm in size, which would result in a large portion being discarded, as TOV only examined the fraction < 0.22 μm. Notably, the Tara Oceans samples collected material

between 0.22 - 3.0 μm, which would have discarded a fraction of *E. huxleyi* (2-6 μm) cells from collection. Detailed information on the Tara Oceans sampling strategy can be found in Sunagawa et al (2015). Analyses used all 243 assemblies collected at 68 locations across the global ocean to probe for genes associated with both *E. huxleyi* and EhVs.

To identify the assembled genes that were similar to EhVs, a total of 7,069 protein sequences queried as *Emiliania huxleyi* virus (EhV) were downloaded from NCBI for comparison to the Tara Oceans dataset. Redundancy among these proteins was reduced by clustering the results to 98% similarity, lowering the number of target proteins to 1,941. These proteins were then blasted (tBLASTn) against the Tara Oceans database with a sequence identity cutoff at 30%. A reciprocal blast was performed on the returned contigs against the NCBI non-redundant protein database. Only contigs matching EhV genes as the top BLAST hit passed this step of the filter.

4.3.2 Field sample collection. Biomass samples around Japan were collected on cruises WK13-07, MR13-06, UM13-08, OM13-08 (aboard R/Vs Wakataka-maru, Ushio-maru, Oshoro-maru, respectively), as well as with small boat sampling. WK13-07 collected samples from two stations in both Oyashio and Kuroshio current areas. UM13-08 collected samples from two stations in the Sea of Japan, off the western coast of Hokkaido. OM13-08 collected samples from two stations in the Oyashio/Kuroshio current mixing region off the east coast of Japan. In general, between 1-4 L of water was collected between 10 m and 60 m, filtered onto 0.2 μm pore-size Sterivex filters with a

Masterflex L/S with type 25 tubing, snap frozen in liquid nutrient, and stored at -80°C until processed.

Biomass at Station ALOHA (22°45'N, 158°00'W) was collected on KM15-04 on 26 April 2-6 2015 aboard the *R/V Kilo Moana*. A total of 18 samples were collected twice daily for three days from CTD casts at three depths corresponding to 1%, 22% and 55% surface light irradiance. Four liters of seawater was filtered with peristaltic pumping onto 0.45 μm pore-size Sterivex filter and snap frozen in liquid nitrogen before being stored at -80°C.

Biomass from the Santa Barbara Channel (SBC) was collected on two dates (10 June and 18 June 2015) corresponding to coccolithophore bloom and post-coccolithophore bloom scenarios, respectively. Samples on 10 June were collected via small boat operations with bucket sampling of the surface waters at 16 locations throughout the SBC basin. Samples on 18 June were collected as part of the 'Plumes & Blooms' cruise aboard the *R/V Shearwater* with 5 L. Niskin bottles at 1 m and 50 m from two locations in SBC. Volume filtered ranged between 100-650 ml was filtered onto 1.2 μm pore-size Isopore filters with gentle vacuum, snap frozen with liquid nitrogen and stored at -80°C until processed.

Additional samples were collected at three stations (59, 79, 86) on the MR13-06 cruise (08/28/13-10/20/13) aboard the *R/V Mirai* in the Bering Strait and Chukchi Sea. Between 10-20L of seawater was filtered onto 47 mm GF/F filters under gentle vacuum and immediately frozen at -80°C until processed.

4.3.3 DNA extractions. DNA extractions from samples collected onto Sterivex filters were conducted using the PowerWater Sterivex DNA Isoaltion Kit using manufacturer instructions (MoBio Laboratories, Inc).

DNA extractions of biomass collected onto GF/F filters was conducted by transferring each filter to a 15 ml Falcon tube, adding 2 ml lysis buffer [100 mM Tris (pH = 8), 40 mM EDTA, 100 mM NaCl, 1% SDS] and heating at 70°C for 5 min prior to adding 200 µl silica beads and vortexing for 30 sec. Vortexing was repeated three times before transferring lysate and filter to a syringe barrel and pressing the lysate into a new 15 ml Falcon tube and adding 903 µl 0.7 M NaCl (0.2 M final concentration) and 323 µl 10% CTAB (1% final concentration) and mixing. Contents were distributed to fresh Eppendorf tubes before performing equal volume Phenol:Chlorophorm:Isoamylalcohol (25:24:1) extraction followed by equal volume Chloroform:Isoamyl Alcohol (24:1) extraction. The aqueous phase was transferred to a new tube and precipitated with 1 volume of 95% Ethanol and 0.2 volumes of 5.25 M ammonium acetate over night at -20° C. DNA was pelleted by centrifugation [13200 RPM, 16,000 xg at 4°C] for 30 min. Pellets were washed with 70% ethanol, dried and resuspended in TE buffer.

Biomass collected onto Isopore filters was extracted by first cutting filters into pieces and placing them into a 2 ml Eppendorf tube with 800 µL of GTE buffer (50 mM Glucose, 25 mM Tris-Cl, 10 mM EDTA, ph = 8), 100 µl 2 mg/mL lysozyme and 100 µL 0.5 EDTA. Lysates were incubated at room temperature for 2 h under gentle shaking. 200 µL 10% SDS was added to the lysate and incubated for an additional 10 min. The sample was mixed with 840 µL phenol followed by centrifugation (10 min; 16,000 x g; RT). The aqueous layer was then transferred to a new tube and cleaned with an equal

volume Chloroform/Isoamylalcohol 24:1 and centrifuged again for 10 min. The aqueous phase as transferred to a fresh tube and one-half volume of 7.4 M ammonium acetate was added and incubated for 30 min before centrifuging at max speed (16,000 x g;15 min; RT). The supernatant was treated with 2.5 volumes of 100% ethanol and precipitated overnight at 4°C. Precipitated DNA was recovered by centrifugation for 30 min (16,000 x g; 4°C) before discarding supernatant and washing pellet with 300 µL ice-cold 70% ethanol. Pellets were dried and suspended in TE buffer.

4.3.4 PCR amplification. PCR reactions (25 µL) were prepared with 1.25U GoTaq G2 DNA polymerase, 1x GoTaq Reaction Buffer (both from Promega), 1.5mM MgCl, 0.2mM each dNTP, and nucleic acid and nuclease free water (Ambion). Forward and Reverse primer concentrations were 0.2 µM, 0.6 µM, and 0.4 µM for MCP (FW primer: 5'-GTC TTC GTA CCA GAA GCA CTC GCT-3'; RV primer: 5'-ACG CCT CGG TGT ACG CAC CCT CA-3'), SPT (FW primer: 5'-GAR CAT ACA CGI RAG CAA CTT-3'; RV primer: 5'-CGG TCC ACA TGT ACC ACA-3'), and COI (FW primer: 5'-TCG GAA TTG TTT CTC ACA-3'; RV primer: 5'-AAT CCT ACC GCA AAA AGC-3') primers, respectively. Template DNA concentration varied between 10-100 ng but were always added to the reaction mixture in a 2µL volume. Positive and negative controls were prepared and run for each PCR amplification series. MCP and SPT amplifications used a fresh EhV86 lysate (50 ml; 1 x 10^8 virus/ml; extracted from a culture infection of *E. huxleyi* strain CCMP 1516 after 96 hr infection; 0.45 µm pore-size filtered to remove debris) as a positive control, while COI used *E. huxleyi* strain CCMP1516.

Settings on an Mastercycler epGradient thermocycler (Eppendorf) consisted of the following cycles for the different primer sets: COI— initial denaturation at 94°C for 3 min; 30 cycles of denaturation at 94°C for 30 sec, annealing at 55°C for 45 sec, and extension at 74°C for 45 sec; final extension at 74°C for 3 min. MCP — initial denaturation at 94°C for 3 min; 40 cycles of denaturation at 94°C for 45 sec, annealing at 54°C for 30 sec, and extension at 72°C for 30 sec; final extension at 72°C for 1 min. SPT —denaturation at 95°C for 5 min; 40 cycles of denaturation at 95°C for 30 sec, annealing at 53°C for 1 min, and extension at 74°C for 1.5 min; final extension at 74°C for 5 min.

Following PCR amplification, DNA bands were resolved and visualized by gel electrophoresis (1.5% agarose, w/vol) with Ethidium Bromide staining (1:20,000) and UV excitation. Sizing used a 100 bp molecular weight ladder. DNA bands of expected lengths were excised and gel purified with a QIAquick Gel Extraction Kit (Qiagen) according to manufacturer's protocols. Purified amplicons were used as a template in a second round of PCR with Illumina Nextera primer overhang sequences to the beginning of respective MCP, SPT, and COI forward and reverse sequences; Forward overhang: 5'TCG TCG GCA GCG TCA GAT GT GTA TAA GAG ACA G-[locus-specific sequence]; Reverse overhang: 5' GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G-[locus-specific sequence]. PCR conditions were similar to the first round of amplification for the respective primers. Amplicons were separated and visualized with an agarose gel with the bands being excised and purified as above. Purified amplicons were used in Illumina library preparation (details below)

4.3.5 Illumina library construction. Illumina library construction was performed according to manufacturer's protocol (https://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html) with one modification, using NEBNext High Fidelity 2X PCR Master Mix for the Index PCR reactions. Library concentrations were measured using Qubit dsDNA Broad Range kit. Eleven libraries were randomly selected for analysis on the Agilent Bioanalyzer 2100, High Sensitivity DNA chip. Libraries were then diluted to a final concentration of 4 nM in 10 mM Tris HCl (pH 8.5) and pooled. Sequencing was conducted in two runs, the first loaded onto the MiSeq at 8 pM final concentration, while the second was loaded at 6 pM final concentration for optimal cluster density on the flow cell. Both runs were loaded with 10% PhiX control spiked in (for sequence heterogeneity). 500-cycle run, paired end 250x250.

4.3.6 Bioinformatic analysis. Sequencing adapters, PCR primers and low-quality reads were trimmed from the raw sequencing reads using CLC Genomics Workbench (QIAGEN, Inc). 225 MCP reference sequences were obtained from the National Center for Biotechnology Information (NCBI) website [https://www.ncbi.nlm.nih.gov/] nucleotide database, by collecting all 'Coccolithovirus major capsid protein' nucleotide sequences available. Redundancy in the MCP reference set was reduced by clustering at 100% identity with CD-HIT (Fu *et al.* 2012), which resulted in 125 unique MCP sequences. Trimmed reads from each sample were independently mapped to the MCP reference set, while applying a 94% sequence identity threshold over 95% of the read length. The threshold was set at 94% to take advantage of the known diversity within

EhV MCP sequences. Similarly, trimmed reads from the SPT-like samples were mapped against 5 non-redundant EhV SPT references with the criteria of 70% sequence similarity over at least 70% of the read length, chosen thresholds were based on both database known diversity among EhV SPT sequences and unpublished data.

Because of the high sequence similarity between both *E. huxleyi* and *Gephyrocapsa oceanica* COI genes, mappings were conducted against 48 COI sequences from both algae collected from the NCBI. The rational was that accounting for both *E. huxleyi* and *G. oceanica*, we could better discriminate sequences that uniquely mapped to *E. huxleyi*, therefor better distinguishing between the two using the COI gene and correctly identifying the source of DNA in our samples. Validated mapped COI reads had to pass a 98% sequence identity over 99% of read length threshold. Following this a second filter was placed, allowing only reads that uniquely mapped to an individual *E. huxleyi* or *G. oceanica* strain to pass.

To explore the diversity among mapped reads for MCP and SPT, individual reads were grouped into clusters at a 98% similarity. Sequences that fit into multiple existing clusters were added to the cluster with the representative of highest similarity. The top 25 clusters with the most recruits were used for the diversity analysis. Dendrograms were prepared for cluster comparison as well as comparison to known EhV sequences in MEGA7 (v7.0.26). Nucleotide based dendrograms were made with a bootstrap test (500 replicates) and evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura *et al.* 2004). Amino acid level dendrograms were also made with a bootstrap test (500 replicates) and evolutionary distances were computed using the Poisson correction method (Zuckerkandl & Pauling 1965).

To explore biogeographic distributions of clusters, we used only clusters with >100 recruits. Clusters were then identified and grouped by whether they contained reads uniquely from one North Pacific region (broken into samples from SBC, ALOHA, and JPN), shared among two regions, or present in all three regions.

4.3.7 Lab-based Infection Experiments. Two strains of *E. huxleyi*: MR70N (isolated in the Arctic Ocean, north of the Chukchi sea) and NIES-1311 (isolated from the Bering Sea) were tested for susceptibility to three different strains of virus: EhV1 and EhV86 both isolated from the English Channel; and EhV163 which was isolated from a Norwegian fjord). Cultures of MR70N and NIES-1311 were started at 1-2 x 10^5 cells ml^{-1} in F/2 media and incubated at 15°C 16:8 light dark cycle. After cultures reached 0.5-1 x 10^6 cells ml^{-1}, flasks were inoculated with either virus (EhV1, EhV86, or EhV163) at an MOI of 1:1 (virus:host) or a similar volume of media, making triplicates of each experimental group. The MR70N experiment was sampled daily for 6 d after inoculation with virus. NIES-1311 cultures were sampled daily for 3 d after viral inoculation. Cell counts were made using an automated cell counter.

4.3.8 Virus Abundance. After homogenizing cultures with gentle shaking, a 40 µl aliquot was preserved with 0.5% glutaraldehyde (final concentration) in a 2 ml microcentrifuge tube and refrigerated at 4°C for 15 min before freezing at -80°C for storage. To count viruses, 0.96 ml of filtered TE with SYBR gold stain (1:20,000) was added to samples and heated at 80°C for 10 min. After cooling, EhVs were counted on an BD Biosciences Influx Mariner flow cytometer with BD FACS software using 488 nm

excitation and 542nm fluorescence emission and forward scatter. Event gates were drawn in FlowJo v9.9.3 using a culture population of EhV86 to draw the boundaries of the gate.

4.4 RESULTS AND DISCUSSION

EhV infection of *E. huxleyi* is a process that has thus far been limited to observations throughout the North Atlantic and marginal seas. Our analyses using global metatranscriptomic datasets and targeted amplicon gene sequencing of EhV-derived genes in diverse regions of the North Pacific, both associated with host cellular biomass, suggests that it is a global phenomenon, impacting *E. huxleyi* cell fate throughout the ocean. Further, this interaction may include a large degree of diversity on the genetic level that manifests in structural differences among viruses.

4.4.1 Tara Oceans Assemblies. *E. huxleyi*-like genes were abundantly distributed in the Tara Oceans gene assembly dataset. Of the 77 globally-distributed stations with metagenome assemblies in the Tara Oceans database, 58 stations contained gene assemblies that matched (total of 617 gene assemblies) *E. huxleyi* genes with protein alignments of > 90% similarity and returned *E. huxleyi* as the top hit in blast searches (Figure 4.1a). Given its global and cosmopolitan distribution, it is not surprising to see a large number of assembly recruits to the *E. huxleyi* reference genomes. When normalized for the number of samples at each station, the South Atlantic off the coast of Argentina and near the Southern Ocean had the highest density of sequences per sample with 36 and 23 sequences per sample, respectively. The North Atlantic also contained stations with elevated *E. huxleyi* representation, including Station (St) 148 in the west and St150 in the

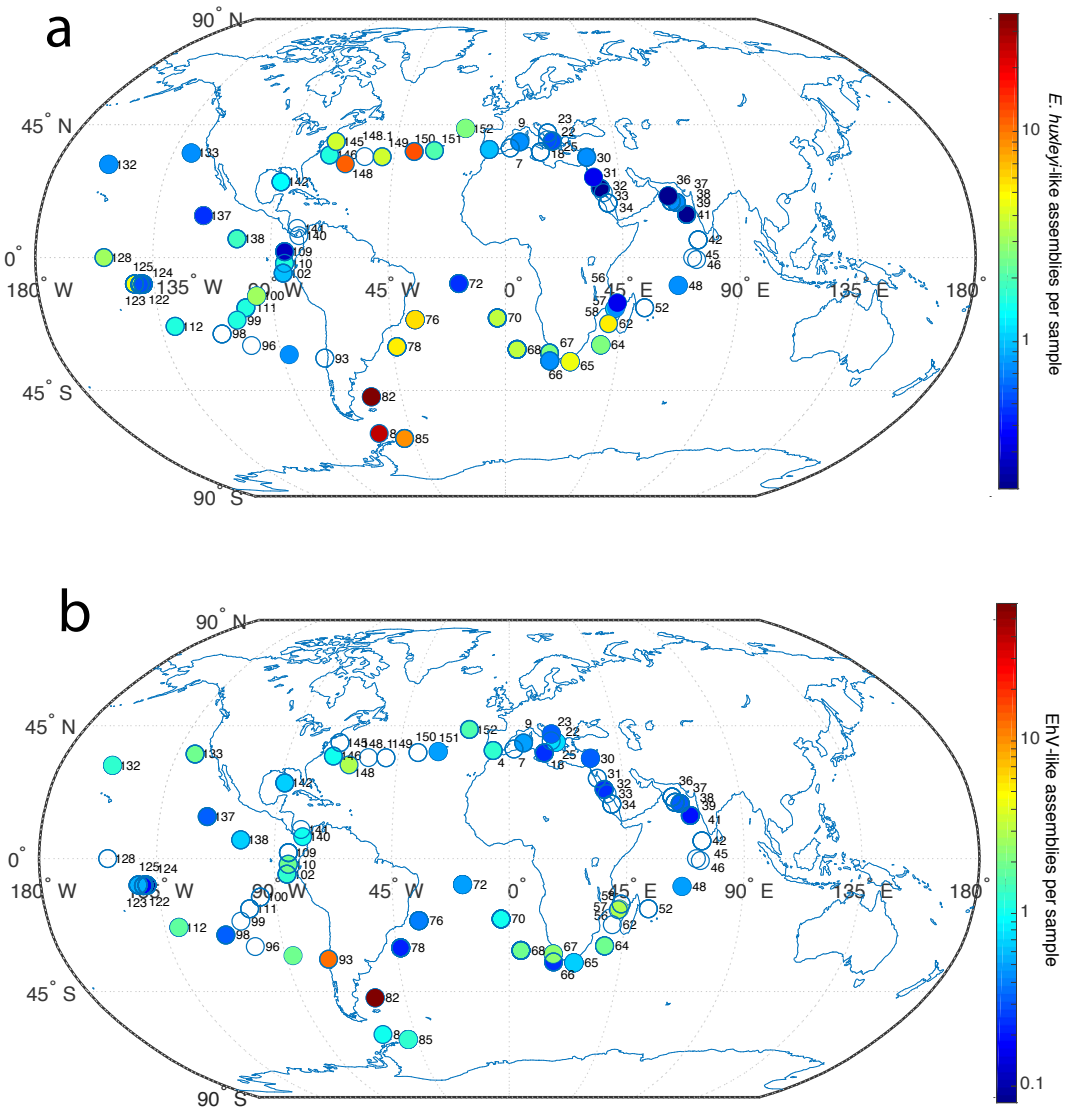


Figure 4.1. a) Distribution of *E. huxleyi* gene assemblies collected from the Tara Oceans study. Individual circles represent stations where samples were collected for metagenome analysis and assembly construction. Open circles represent stations without *E. huxleyi* genes present in assemblies. The color bar provides a visualization of the density of gene assembly representation (normalized per sample) with values plotted on log 10 scale. b) Distribution of EhV-like gene assemblies collected from the Tara Oceans study. Individual circles represent stations where samples were collected for metagenome analysis and assembly construction. Open circles represent stations without *E. huxleyi* genes present in assemblies. The color bar provides a visualization of the density of gene assembly representation (normalized per sample) with values.

central North Atlantic with 11 and 12 hits sample^{-1}, respectively. The North and South Pacific both observed lower representation of *E. huxleyi* among stations with a maximum of 7 hits sample^{-1} at St123. The Indian Ocean also observed relatively low representation

of *E. huxleyi* with the exception of St62 that averaged 4.2 hits sample^{-1} west of Madagascar.

Given the known diversity within the *E. huxleyi* 'pan-genome' (Read *et al.* 2013), we used stringent criteria for successful mapping (>90% identity) that may have overlooked many *E. huxleyi* assemblies in the database. The ability to capture the relative differences in abundance between stations was still achieved by establishing the 90% threshold, however, and with less risk of falsely recruiting non-*E. huxleyi* genes. Given similar sequence processing across stations (Sunagawa *et al.* 2015), these values accurately reflect the relative abundances at the time of sampling.

A total of 198 unique gene sequences within this collection were identified from the 617 recruited sequences. The top twenty most represented sequence matches made up 52% of the total number and were comprised mainly of chloroplast associated genes (Table 4.1). The top hit, a photosystem (PS) I subunit VII, was observed 71 times over the entire assembly dataset, with an average identity of 99.8%. The number of sequences dropped rapidly for the second and third most representative genes, a cytochrome b6/f complex subunit IV (46 times) and ribosomal protein L14 (25 times), respectively, with representative sequence counts decreasing more gradually further down the rank order. The average percent identity values for assembly alignments varied throughout the entire data set but ranged 90%-100%. Notably, hypothetical proteins made up nine of the top twenty identified genes and the majority of *E. huxleyi* genes identified in the assembly dataset (110) were only observed once (data not shown).

Chloroplast genes were particularly abundant in the *E. huxleyi* hits, which may have been due to several reasons. First, the *E. huxleyi* pan-genome includes considerable

Table 4.1 List of top 20 *E. huxleyi* gene hits identified in assemblies from the Tara Oceans metagenomes assembly database. Highlighted in green are genes associated with chloroplasts. Hit is the reference gene that assemblies were identified as given our constraints. Count is the total number of times a given gene was identified throughout the dataset. Average % ID is the average amino acid similarity between a given gene and assembly alignments.

Hit	Count	Average %ID
>YP_277346 photosystem I subunit VII (chloroplast) [Emiliania huxleyi].	71	99.8264789
>YP_277310 cytochrome b6/f complex subunit IV (chloroplast) [Emiliania huxleyi].	46	96.5826087
>YP_277411 ribosomal protein L14 (chloroplast) [Emiliania huxleyi].	25	92.754
>XP_005776965 hypothetical protein EMIHUDRAFT_354368 [Emiliania huxleyi CCMP1516].	20	98.328
>YP_277372 photosystem II protein D2 (chloroplast) [Emiliania huxleyi].	20	96.161
>YP_277313 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (chloroplast) [Emilian	18	93.87
>YP_277423 ribosomal protein S12 (chloroplast) [Emiliania huxleyi].	18	93.0816667
>AEI29582 elongation factor Tu (plastid) [Emiliania huxleyi].	16	93.445
>AAR30308 photosystem I P700 chlorophyll a apoprotein A2, partial (chloroplast) [Emiliania huxley	14	95.4378571
>XP_005783192 hypothetical protein EMIHUDRAFT_442340 [Emiliania huxleyi CCMP1516].	9	92.5733333
>XP_005768399 hypothetical protein EMIHUDRAFT_356218 [Emiliania huxleyi CCMP1516].	8	96.065
>YP_277312 photosystem I P700 chlorophyll a apoprotein A2 (chloroplast) [Emiliania huxleyi].	7	95.6057143
>YP_277398 photosystem I subunit II (chloroplast) [Emiliania huxleyi].	7	94.9014286
>XP_005786081 hypothetical protein EMIHUDRAFT_418014 [Emiliania huxleyi CCMP1516].	6	92.8916667
>YP_277371 photosystem II 44 kDa protein (chloroplast) [Emiliania huxleyi].	6	96.9983333
>YP_277387 photosystem II protein V (chloroplast) [Emiliania huxleyi].	6	97.2683333
>XP_005759979 hypothetical protein EMIHUDRAFT_438587 [Emiliania huxleyi CCMP1516].	5	93.648
>XP_005768679 hypothetical protein EMIHUDRAFT_66170 [Emiliania huxleyi CCMP1516].	5	97.178
>XP_005777144 hypothetical protein EMIHUDRAFT_450596 [Emiliania huxleyi CCMP1516].	5	92.118
>XP_005783442 hypothetical protein EMIHUDRAFT_415280 [Emiliania huxleyi CCMP1516].	5	92.022

diversity and chloroplast genes may represent more conservative genes (Read *et al.* 2013). Second, *E. huxleyi* cells can contain two plastids with plastid genomes containing up to hundreds of ptDNA copies, multiplying the genetic content of plastids relative to nuclear DNA (Smith *et al.* 2011). Third, it could be due to experimental bias and the differences in GC content between nuclear DNA (65%) and plastid DNA (37%) (Puerta *et al.* 2005).

Although EhV-like genes were present at a lower number of stations than *E. huxleyi* hosts, they were detected among cell biomass within every ocean basin examined. Gene assemblies from 45 of the 77 stations across the global dataset contained sequences that matched EhV protein sequences above a 30% alignment identity threshold and returned an EhV as the top hit on a reciprocal blast search (Figure 4.1b), comprising a

total of 336 gene assemblies. Similar to the *E. huxleyi* sequence distribution, the South Atlantic contained the station with the highest sequence density representation, St82, with an average of 63 sequences sample^{-1} and a total of 188 sequence assemblies and was the same station with the most abundant *E. huxleyi* sequence assemblies. The South Atlantic was also the only ocean basin in which all eight stations (St66-82), spanning from the African to South American margins, contained at least one EhV sequence.

EhV-like genes were generally present at many of the same stations as *E. huxleyi*-like genes, overlapping at 33 stations, while *E. huxleyi*-like genes appeared independently at 19 stations. EhV genes did appear in the absence of *E. huxleyi*-like at six stations in the South and Equatorial Pacific (St140, St93, St98) and in the Mediterranean (St18, St23). The South Pacific Station 93, off the coast of Chile, had a relatively high number of EhV sequences with an average of 12 and total of 24. However, this station did not have any *E. huxleyi* sequences. Assuming that these are EhV genes, possible scenarios whereby EhVs can be observed at very low to no host abundance include incorporation into wind blow aerosols (Sharoni *et al.* 2015) or incorporation into the gut contents and resulting fecal pellets of copepods (Frada *et al.* 2014) could provide transport away from *E. huxleyi* populations. Given the large size of EhVs, 0.2 μm diameter, they are on the threshold of being collected in sample biomass according to the Tara Oceans sampling protocols, biomass > 0.22 μm, without actively infecting a host cell. It is also conceivable that *E. huxleyi* were present at very low abundance but not detectable by metagenome sampling methods.

Relative to the rest of the ocean, *E. huxleyi*- and EhV-like recruits were most prominent in the South Atlantic Ocean, surprisingly more so than the North Atlantic,

where a majority of coccolithophore studies have been conducted (Wilson *et al.* 2002; Vardi *et al.* 2009). However, the area of the South Atlantic where host and virus were most abundant was part of an area known as the Great Calcite Belt (Balch *et al.* 2016). Coccolithophore blooms are well documented in this area and can extend from the coast of South America east, parallel to the Southern Ocean.

One factor that could have greatly impacted the distribution and relative abundance of both host and virus hits was the seasonality of sampling. *E. huxleyi* blooms often occur in the late spring following the diatom blooms once surface stratification has started to occur (Tyrrell & Merico 2004). The North Atlantic transect was made early in the year (January-March, 2012) (Tara Oceans Consortium & Tara Oceans Expedition 2015), prior to the annual North Atlantic spring bloom. If the transect had been made several months later, it is likely that both *E. huxleyi* and EhV sequences would have been relatively more abundant than they were in the winter. Likewise, the transect through the Red Sea and Indian Ocean, which exhibited the lowest abundance of host and virus sequences, may have been different if sampling had occurred within another season, such as in blooming regions that correspond to summer or winter monsoon events (Lévy *et al.* 2007).

The identified EhV sequences from the Tara Oceans dataset represented 208 different genes/homologs that are present in the 13 currently sequenced EhV genomes, with eleven of the top twenty being hypothetical proteins (Table 4.2). Other commonly identified genes included a zinc finger protein and DNA-directed RNA polymerase subunit B genes. There was less of a difference in rank order between the most commonly identified gene, a hypothetical protein with 13 representatives, and the

Table 4.2 List of top 20 EhV gene hits identified in assemblies from the Tara Oceans metagenomes assembly database. Hit is the reference gene that assemblies were identified as given our constraints. Count is the total number of times a given gene was identified throughout the dataset. Average % ID is the average amino acid similarity between a given gene and assembly alignments.

Hits	Count	Average % ID
>AEO97675 hypothetical protein ENVG_00281, partial [Emiliania huxleyi virus 84].	13	47.12
>AEP16139 hypothetical protein ERVG_00264, partial [Emiliania huxleyi virus 208].	10	46.47
>AEP15697 zinc finger protein [Emiliania huxleyi virus 207].	9	58.03
>AET42512 hypothetical protein EXVG_00327 [Emiliania huxleyi virus 202].	8	58.78
>AEO98258 hypothetical protein ELVG_00209 [Emiliania huxleyi virus 203].	7	52.03
>AEP15111 DNA-directed RNA polymerase subunit B [Emiliania huxleyi virus 88].	7	44.59
>CAZ69379 putative endonuclease [Emiliania huxleyi virus 99B1].	7	57.02
>AEO98384 deoxyuridine 5'-triphosphate nucleotidohydrolase [Emiliania huxleyi virus 203].	6	51.74
>AET42789 hypothetical protein EXVG_00140 [Emiliania huxleyi virus 202].	5	53.94
>CAZ69617 hypothetical protein [Emiliania huxleyi virus 99B1].	5	50.70
>YP_294064 hypothetical protein EhV307 [Emiliania huxleyi virus 86].	5	43.53
>AET42590 fatty acid desaturase [Emiliania huxleyi virus 202].	4	63.29
>AEO98072 hypothetical protein ELVG_00397 [Emiliania huxleyi virus 203].	3	59.17
>AEO98334 hypothetical protein ELVG_00033 [Emiliania huxleyi virus 203].	3	56.08
>AEP15315 zinc finger protein [Emiliania huxleyi virus 88].	3	63.71
>AEP16007 DNA-dependent RNA polymerase II largest subunit [Emiliania huxleyi virus 208].	3	69.17
>AET42561 major capsid protein [Emiliania huxleyi virus 202].	3	52.81
>AET42698 hypothetical protein EXVG_00049 [Emiliania huxleyi virus 202].	3	47.73
>AET42704 hypothetical protein EXVG_00055 [Emiliania huxleyi virus 202].	3	78.09

remaining genes, indicating a more even representation of EhV-like sequences. Over representation of a single gene may have indicated that the recruited gene was from a sequence in another abundant organism that shared similarities to an EhV gene. However, this was not observed. Notably, average percent protein level identifications ranged between 44-78% for these top 20 hits, much lower than what was observed for *E. huxleyi* hosts. There was an even greater range in sequence similarity scores to known EhV proteins (Figure 4.2a), which spanned from 30% (the minimum criteria) to 96% and displayed a bimodal distribution of alignment percent identity with the largest peak at the 50-55% bin (with 67 hits) and the other peak between the 85-95% bins (including 73 hits between the two bins). This distribution changes, however, when separating the data from the South Atlantic, especially Station 82, which contained roughly half of the total global sequence hits. Excluding Station 82, the global collection of sequences have a normal distribution with a peak in the 50-55% identity range and no longer contain hits

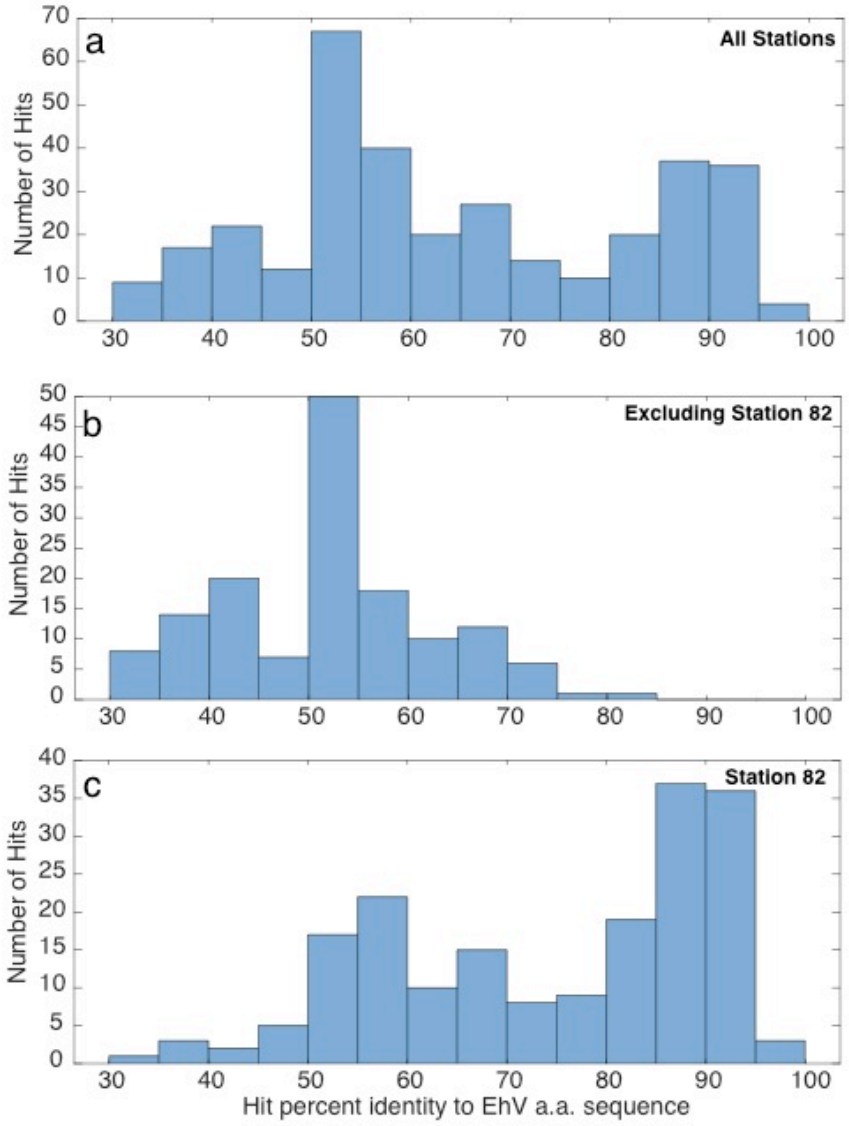


Figure 4.2. Histogram of the EhV amino acid (a.a.) percent identity for gene assembly alignments scored against reference EhV proteins. (a) Histogram of sequence identities across all stations in the Tara Oceans assembly dataset. (b) All stations excluding the anomalous Station 82. (c) Assembled sequences from Station 82 only. Bins are in 5% intervals. Minimal percent identity for an accepted hit was 30%.

above 85% (Figure 4.2b). A bimodal distribution is still observed when looking specifically at Station 82, but the distribution is skewed towards the higher percent identity, showing that the aligned assemblies resemble EhVs more closely here than the global average (Figure 4.2c).

It is difficult to conclude whether the EhV-like assemblies that recruited to the EhV reference genomes are in fact derived from EhVs based solely on their presence, because the vast diversity of microbial DNA could contain sequences that are similar to EhVs but derived from other organisms. The percent identity threshold for recruitment was set adequately low (30%) to catch divergent EhV sequences. Of course, this comes with the tradeoff off increasing the likelihood that unrelated, unknown but related genes will be counted as EhV sequences. Indeed, some genes that are very conservative, such as MCP, should have a higher threshold for identity, as recruits below this threshold are not likely to actually be this gene. On the other hand, genes such as SPT, should have a lower threshold, as it is established that there is an abundance of known genetic diversity and it is more likely that recruits with lower identities will be the gene of interest. However, when searching across the entire genome, setting individual gene constraints is not practical, because it is unknown how conservative many of the genes and 'hypothetical genes' in the EhV genome are, and therefore a generally lower threshold is acceptable.

To help address the likelihood that genes from the Tara Oceans dataset were indeed EhV genes, we compared twenty random homologues from distinct gene families within the Tara Oceans assemblies for a conservative estimate on variability in gene structure. This analysis was considered conservative because we compared homologs among EhVs isolated from blooms in the North Atlantic and marginal seas; diversity is likely greater on the global scale. This analysis showed that there is a large range in the similarity of genes among reference EhV genomes (Table 4.3). Some genes, such as DNA-directed RNA polymerase subunit B, are very conservative among homologs and

Table 4.3 Homolog similarity comparison for 20 randomly selected EhV genes with hits in the Tara Ocean assembly database. Gene homologs among thirteen EhV genomes were compared for these genes. Query Sequence ID refers to the assembly that aligned to the hit EhV gene. Hit ID is the gene name that the assembly sequence aligned with. Genome refers to the EhV strain that the reference gene came from. Percent identity is the similarity of the assembly to the reference gene. Pertaining to the reference gene, the table describes the a.a. sequence length, number of EhV strains that contain homologs, the highest and lowest percent identity to the reference gene among homologs, and the highest and lowest E value scored among homologs with respect to the reference gene. Self identity is not shown within the table.

Query Sequence ID	Hit Accession	Hit ID	Genome	Percent Identity	a.a. length	% strains with Homolog	Highest Homolog Identity	Lowest Homolog Identity	Lowest E value	Higest E value
ENA|CESI01141127|CESI01141127.1	gi|558480880|gb|AHA54446.1|	hypothetical protein	EhV18	87.98	233	100%	100	83	2.00E-177	1.00E-132
ENA|CESI01241086|CESI01241086.1	gi|558481062|gb|AHA54627.1|	putative lectin protein	EhV145	82.59	913	100%	57	36	0	2.00E-97
ENA|CESO01168118|CESO01168118.1	gi|558481691|gb|AHA55255.1|	stimulates sugar fermentation	EhV156	64.65	291	92%	100	69	0	1.00E-113
ENA|CESI01171739|CESI01171739.1	gi|558482257|gb|AHA55820.1|	hypothetical protein	EhV164	90	147	100%	100	95	1.00E-103	9.00E-99
ENA|CEUS01379454|CEUS01379454.1	gi|73852780|ref|YP_294064.1|	hypothetical protein	EhV86	59.23	410	92%	100	41	0	2.00E-154
ENA|CEUS01285982|CEUS01285982.1	gi|73852819|ref|YP_294103.1|	hypothetical protein	EhV86	64.65	125	92%	100	82	3.00E-89	8.00E-62
ENA|CERS01029622|CERS01029622.1	gi|283481648|emb|CAZ69764.1|	polymerase II subunit precursor	EhV99B1	34.86	1507	100%	99	93	0	7.00E-154
ENA|CESI01063598|CESI01063598.1	gi|347482379|gb|AEO98320.1|	hypothetical protein	EhV203	32.52	660	100%	100	81	0.00E+00	0.00E+00
ENA|CEOU01067597|CEOU01067597.1	gi|347482443|gb|AEO98384.1|	deoxyuridine 5'-triphosphate nucleotidohydrolase	EhV203	50	148	100%	100	85	1.00E-108	9.00E-93
ENA|CESU01177357|CESU01177357.1	gi|347482131|gb|AEO98072.1|	hypothetical protein	EhV203	55.26	468	100%	100	74	0	2.00E-124
ENA|CESO01210385|CESO01210385.1	gi|347482002|gb|AEO97988.1|	mRNA capping enzyme, beta chain.	EhV84	57.75	230	100%	100	83	2.00E-178	8.00E-146
ENA|CESU01187010|CESU01187010.1	gi|347481894|gb|AEO97880.1|	hypothetical protein	EhV84	63.08	447	100%	100	90	0	0.00E+00
ENA|CEUB01003139|CEUB01003139.1	gi|347600491|gb|AEP14978.1|	GIY-YIG catalytic domain.	EhV88	40.76	331	100%	80	33	6.00E-79	4.00E-27
ENA|CEVR01172203|CEVR01172203.1	gi|347600624|gb|AEP15111.1|	DNA-directed RNA polymerase subunit B	EhV88	44.61	1156	100%	99	94	0	0.00E+00
ENA|CEWG01307690|CEWG01307690.1	gi|347601211|gb|AEP15697.1|	Zinc finger, C3HC4 type (RING finger).	EhV207	54.72	390	100%	100	71	0.00E+00	3.00E-143
ENA|CESI01086566|CESI01086566.1	gi|356927888|gb|AET42678.1|	hypothetical protein	EhV202	93.25	1216	100%	100	84	0	0.00E+00
ENA|CESI01077160|CESI01077160.1	gi|356927934|gb|AET42724.1|	Bacterial lipocalin	EhV202	54.55	202	100%	100	83	0.00E+00	5.00E-51
ENA|CESI01307700|CESI01307700.1	gi|356927763|gb|AET42553.1|	hypothetical protein	EhV202	94.72	530	100%	100	80	0	0.00E+00
ENA|CESU01111290|CESU01111290.1	gi|356927842|gb|AET42632.1|	hypothetical protein	EhV202	88.97	586	100%	100	69	0.00E+00	0.00E+00
ENA|CESI01117862|CESI01117862.1	gi|356927657|gb|AET42447.1|	hypothetical protein	EhV202	92	170	100%	100	62	6.00E-130	2.00E-78

are present in all strains with a minimum homolog identity of 94%. So it is unlikely that an assembly sequence with a low % identity originated from an EhV. Other gene homologs, such as the putative lectin proteins, were found to be very dissimilar. Given

these empirical observations, it is justified to reason that an assembly with a low ID score derived from EhVs.

Tara Station 82 could also be used to help address the issue of whether low identity hits are from EhVs. Through satellite observation, it is evident that a coccolithophore bloom was present at St82 during sample collection. Aqua MODIS-derived calcite concentration revealed a large patch of elevated calcite reflectance stretching 1300 km off the coast of Argentina in an 8-day composite image of the area between 3-10 December 2010 (Figure 4.3). At the highest concentrations, calcite was measured over two orders of magnitude higher than outside the patch, indicative of a dense coccolithophore bloom. Given Tara Oceans Station 82 contained a relatively high number of both *E. huxleyi* and EhV sequences and that many of the EhV sequences had high identity scores (85-95% identity), we can confidently conclude that an EhV-infected bloom was present during sampling. It is noteworthy that this is the first observation of this host-virus interaction in the Southern Calcite Belt. At the same time, Tara Oceans Station 82 also possessed an elevated number of EhV sequences with 50-60% identity (Figure 4.2c). In fact, one third of the total EhV sequences in the global dataset with this range of identity percentages originated from this station, with a disproportionally high number in the 30-50% range. It suggests that the EhV pan-genome is indeed diverse and that sequences at other stations, which fall within that range of percent identifies, do indeed derive from EhVs.

An important issue to consider is the large difference between the histogram of identities produced by sequences obtained at Tara Oceans Station 82 compared to those recovered at the rest of the stations. No other station had sequences that scored above

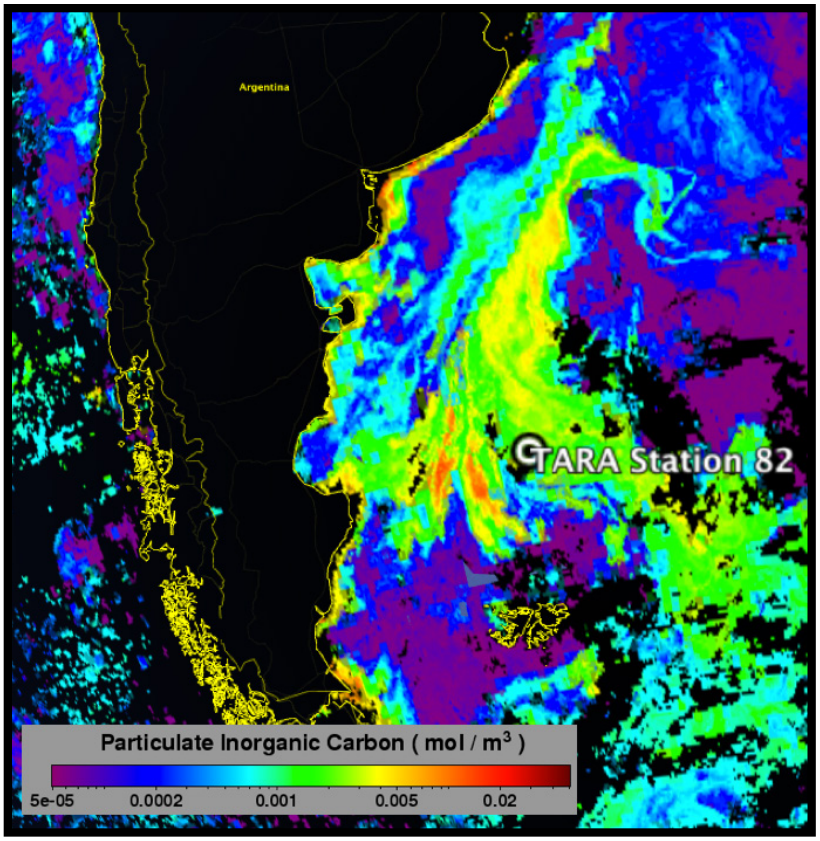


Figure 4.3 MODIS-Aqua 8-day composite satellite product of sea surface particulate inorganic carbon on 3-10 December 2010, off the coast of Argentina. Location of Tara Oceans Station 82 is denoted with the white target. Area with cloud cover and no data available are colored black. Black area within yellow trim represents continent. Color bar is on log scale.

90% identity, in contrast to the large number present at Tara Oceans Station 82. In fact, only one other station (on the coast of Chile; Station 93) had sequences with 75-85% identity to EhVs. If we reason that EhVs throughout the global ocean were similar to those observed at Tara Oceans Station 82, then this should be reflected in the global identity histogram. In contrast, their comparative difference suggests that EhVs are quite diverse and can differ from reference strains currently in the database. One contributor to the observed differences could be that EhVs that are successful in bloom scenarios (like those from Tara Oceans Station 82 or those isolated previously isolated from declining blooms and currently in culture collections) are genotypically different from those

observed in more dispersed, non-bloom populations. This is a difficult issue to resolve as very little is known about infection at low cell densities of *E. huxleyi*, but one well worth considering.

4.4.2 Targeted gene sequence proxies. Having demonstrated the representation of *E. huxleyi* and EhVs in the global context, we wanted to use a targeted gene approach to address the incidence and diversity of EhVs among populations of *E. huxleyi* in diverse regions of the Pacific Ocean basin – a coccolithophore bloom in the Santa Barbara Channel, oligotrophic waters of Station ALOHA, and assorted waters around Japan – and into the Chukchi Sea (Figure 4.4). We chose EhV-encoded genes that explored diversity

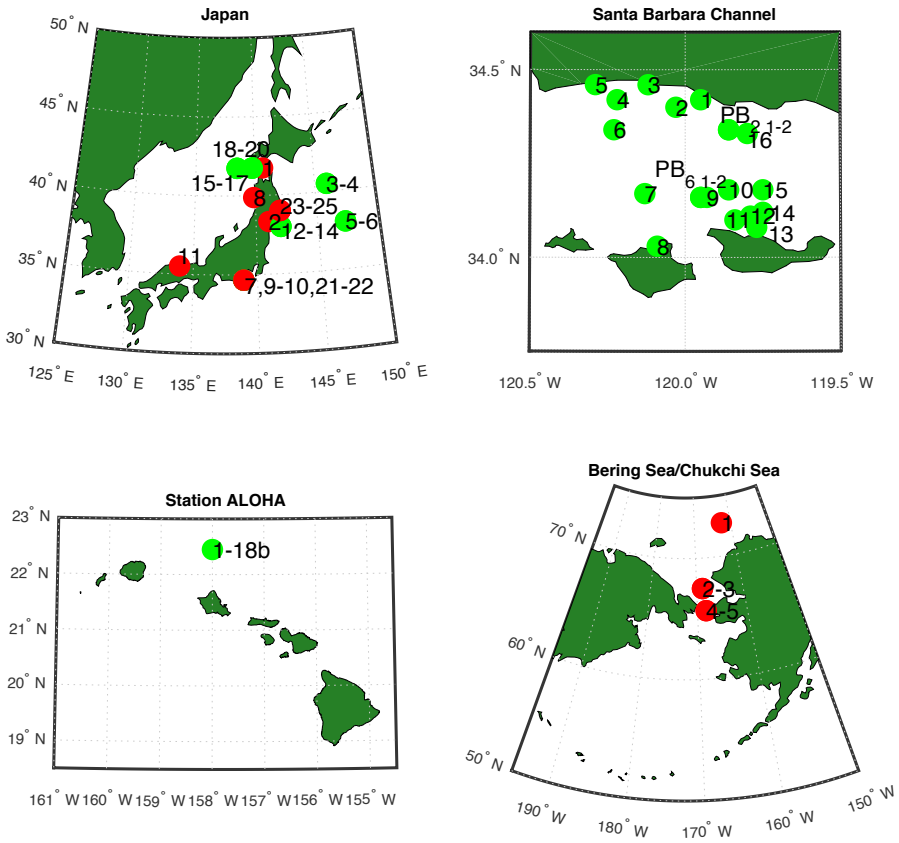


Figure 4.4. Sampling locations for targeted COI, MCP, and SPT gene amplifications around Japan, Santa Barbara Channel, Station ALOHA, and the Bering/Chukchi Sea. Numbers correspond to the sample numbers from the respective areas. Bright green symbols indicate locations where EhV-like genes were successfully amplified. Red symbols indicate no amplification.

in both a structural, conservative gene –major capsid protein (MCP) – and a functional, non-conservative gene–serine palmitoyl transferate (SPT). We also used a conservative, host genetic marker– cytochrome oxidase c subunit 1 (COI) – to establish the presence/absence of *E. huxleyi*. A total of 72 samples (66 unique samples) were collected between 2013 and 2015 (Table 4.4).

Twenty samples were collected from the Santa Barbara Channel during an intense coccolithophore bloom, which is uncommon in these waters. While low concentrations of *E. huxleyi* are regularly present, large blooms are not regularly documented (Rodriguez-Iglesias, personal communication). Samples were primarily collected from the surface during the bloom; however, subsurface samples were collected during post-bloom dynamics. PCR amplification of the partial *E. huxleyi* COI-like gene was successful in 20 of 20 samples, both during and following the bloom, with >99.5% of corresponding Illumina sequences mapping individual *E. huxleyi* strains sequence, instead of multiple strains and/or *G. oceanica*, ranged between 51-80%. This boosted confidence that the COIs were *E. huxleyi* derived and represented an *E. huxleyi* bloom.

EhV-derived MCP amplicons were also obtained in all twenty samples. However, subsequent analysis and mapping the associated Illumina sequences revealed a large range (<1% to 92% for SBC_10 and SBC_11, respectively) in the percent of reads that mapped to known EhV MCP sequences, with a minimum of 223 mapped reads in SBC_10. EhV derived SPT-like amplicons were amplified in 18 of 20 samples, excluding SBC_11 and SBC_PB_2_1. The range in the percent of sequences mapping to known EhV SPTs was also high, varying between <1% and 99%. Notably, 10 of the 18 samples had a <10% mapping efficiency. Several of these samples had as few as 8 reads

Table 4.4 PCR amplification table and sequencing read chart. For samples collected from the Santa Barbara Channel (SBC), Station ALOHA (ALOHA), regionally around Japan (JPN), and Bering Sea (BS) successful PCR amplification of *E. huxleyi* COI-, EhV MCP-, and EhV SPT-like genes from individual samples collected are denoted with an 'x'. 'No. mapped' indicates the number of sequences that successfully mapped to the respective gene for a given sample and '% mapped' indicates the fraction of the total number of reads that the mapped reads represent.

Sample Name	Date	Location			Host COI				EhV MCP			EhV SPT		
		Latitude (decimal)	Longitude (decimal)	Depth (m)	PCR Band	Number maped	Percent mapped	Unique E. hux	PCR Band	Number mapped	Percent mapped	PCR Band	Number mapped	Percent mapped
SBC_1	6/10/15	34.42	-119.95	surface	x	148618	100%	80%	x	51085	38%	x	8	0%
SBC_2	6/10/15	34.40	-120.03	surface	x	75379	100%	76%	x	171000	87%	x	12	0%
SBC_3	6/10/15	34.46	-120.12	surface	x	136447	100%	69%	x	128667	91%	x	94838	99%
SBC_4	6/10/15	34.42	-120.22	surface	x	146607	100%	68%	x	74311	53%	x	13018	16%
SBC_5	6/10/15	34.46	-120.29	surface	x	116947	100%	69%	x	70252	49%	x	43609	60%
SBC_6	6/10/15	34.34	-120.23	surface	x	139261	100%	51%	x	109385	80%	x	11	0%
SBC_7	6/10/15	34.17	-120.13	surface	x	156874	100%	68%	x	4061	5%	x	6	0%
SBC_8	6/10/15	34.03	-120.09	surface	x	157851	100%	53%	x	1464	2%	x	57752	96%
SBC_9	6/10/15	34.16	-119.93	surface	x	166399	100%	62%	x	445	0%	x	26523	28%
SBC_10	6/10/15	34.18	-119.86	surface	x	146810	100%	73%	x	223	0%	x	30	0%
SBC_11	6/10/15	34.10	-119.84	surface	x	140765	100%	67%	x	131310	92%			
SBC_12	6/10/15	34.11	-119.79	surface	x	89155	100%	70%	x	58782	52%	x	210	0%
SBC_13	6/10/15	34.08	-119.77	surface	x	147556	100%	71%	x	1530	2%	x	97	4%
SBC_14	6/10/15	34.12	-119.75	surface	x	275698	100%	69%	x	32849	41%	x	65581	38%
SBC_15	6/10/15	34.18	-119.75	surface	x	122396	100%	69%	x	3679	6%	x	24	0%
SBC_16	6/10/15	34.33	-119.80	surface	x	137230	100%	68%	x	37957	33%	x	5104	32%
SBC_PB_2_1	6/18/15	34.34	-119.86	1	x	94387	100%	69%	x	50522	73%			
SBC_PB_2_2	6/18/15	34.34	-119.86	50	x	88016	100%	80%	x	67213	73%	x	13	0%
SBC_PB_6_1	6/18/15	34.16	-119.95	1	x	119920	100%	72%	x	62462	57%	x	599	2%
SBC_PB_6_2	6/18/15	34.16	-119.95	50	x	107247	100%	74%	x	76470	76%	x	38082	80%
ALOHA_1	4/7/15	22.45	-158.00	14										
ALOHA_2	4/7/15	22.45	-158.00	35	x	105632	100%	72%	x	26534	54%	x	18774	36%
ALOHA_3	4/7/15	22.45	-158.00	107	x	105571	97%	67%	x	20928	63%	x	60281	99%
ALOHA_1b	4/7/15	22.45	-158.00	14										
ALOHA_2b	4/7/15	22.45	-158.00	35										
ALOHA_3b	4/7/15	22.45	-158.00	107	x	93782	100%	59%	x	45271	81%	x	48221	84%
ALOHA_4	4/8/15	22.45	-158.00	14										
ALOHA_5	4/8/15	22.45	-158.00	35										
ALOHA_6	4/8/15	22.45	-158.00	107	x	99824	100%	77%	x	58647	80%	x	37885	87%
ALOHA_7	4/8/15	22.45	-158.00	14										
ALOHA_8	4/8/15	22.45	-158.00	35										
ALOHA_9	4/8/15	22.45	-158.00	107	x	131401	100%	59%	x	24425	74%	x	50990	82%
ALOHA_10	4/9/15	22.45	-158.00	14										
ALOHA_11	4/9/15	22.45	-158.00	35										
ALOHA_12	4/9/15	22.45	-158.00	107										
ALOHA_13	4/9/15	22.45	-158.00	14										
ALOHA_14	4/9/15	22.45	-158.00	35										
ALOHA_15	4/9/15	22.45	-158.00	107										
ALOHA_16	4/9/15	22.45	-158.00	14										
ALOHA_17	4/9/15	22.45	-158.00	35										
ALOHA_18	4/9/15	22.45	-158.00	107										
ALOHA_18b	4/9/15	22.45	-158.00	107	x	52580	100%	82%	x	5559	24%	x	16348	97%

Table 4.4 continued

Sample Name	Date	Location			Host COI				EhV MCP			EhV SPT		
		Latitude (decimal)	Longitude (decimal)	Depth (m)	PCR Band	Number maped	Percent mapped	Unique E. hux	PCR Band	Number mapped	Percent mapped	PCR Band	Number mapped	Percent mapped
JPN_1	7/26/13	41.74	140.69	surface										
JPN_2	7/12/13	38.33	141.04	surface										
JPN_3	7/18/13	40.50	146.00	10	x	81986	100%	88%	x	117158	73%	x	34992	26%
JPN_4	7/18/13	40.50	146.00	50	x	31364	100%	86%	x	688	1%	x	27470	31%
JPN_5	7/20/13	38.01	147.25	10								x	1612	2%
JPN_6	7/20/13	38.01	147.25	40								x	16425	12%
JPN_7	7/29/13	34.66	138.93	surface										
JPN_8	8/3/13	39.88	139.86	surface										
JPN_9	7/29/13	34.66	138.93	surface										
JPN_10	7/30/13	34.66	138.93	surface										
JPN_11	7/9/13	35.51	133.93	surface										
JPN_12	8/23/13	37.99	142.00	10	x	33010	99%	37%	x	218	0%	x	19	0%
JPN_13	8/23/13	37.99	142.00	20	x	83579	100%	39%	x	118	0%	x	2824	2%
JPN_14	8/23/13	37.99	142.00	30					x	211	0%	x	3559	2%
JPN_15	8/3/13	41.75	138.51	20	x	95341	96%	99%				x	3	0%
JPN_16	8/3/13	41.75	138.51	30								x	26	0%
JPN_17	8/3/13	41.75	138.51	40	x	38817	99%	99%	x			x	35	0%
JPN_18	8/4/13	41.75	139.79	40	x	77554	100%	99%	x	91	0%	x	32	0%
JPN_19	8/4/13	41.75	139.79	50	x	26611	99%	95%	x	38	0%	x	9991	36%
JPN_20	8/4/13	41.75	139.79	60	x	83065	100%	79%	x	128	0%	x	3	0%
JPN_21	7/30/13	34.66	138.93	10								x	12	0%
JPN_22	7/30/13	34.66	138.93	20										
JPN_23	8/25/13	39.00	142.00	30										
JPN_24	8/25/13	39.00	142.00	40										
JPN_25	8/25/13	39.00	142.00	50										
BS_1	9/28/13	72.73	-162.47	surface										
BS_2	10/4/13	67.49	-168.74	surface										
BS_3	10/4/13	67.49	-168.74	surface										
BS_4	10/4/13	65.65	-168.25	surface										
BS_5	10/4/13	65.65	-168.25	surface										

mapped. Lowering the threshold from 70% to 60%, yielded few new recruits, confirming our confidence in the 70% identity threshold. It is possible that environmental DNA with similar compliments to our MCP and SPT primers may have been inadvertently primed and, if relatively abundant, drowned out the actual signal. The large variability could suggest patchiness in the ecosystem and infection dynamics. Nonspecific priming/amplification of environmental DNA by the chosen primer sets may have led to lower percentage of mapped sequences, but this seems unlikely, given that the same PCR conditions were used for all samples. Nonetheless, the variability in recruitment suggests differences in the community DNA composition. It is apparent that this coccolithophore

bloom was undergoing infection by EhVs given that the aforementioned EhV genes were detected with biomass collected onto 1.2 μm pore-size filters during the development of the bloom. With this pore size, these filters would contain viruses associated with (and presumably replicating in) *E. huxleyi* cells while free viruses would pass through. EhV-derived MCP and SPT sequences found in the aftermath of the bloom (the following week) both at the surface and 50 m indicate infection was still progressing despite the passing of the bloom.

Only six samples successfully amplified *E. huxleyi* COI-like sequences at Station ALOHA over the week of sampling, all of which were collected at 35 m and 107 m (Table 4.4), suggesting abundance was low during sampling. COI mapping was also high at this location, with a minimum of 97% recruiting to reference sequences. The range for reads that uniquely mapped to *E. huxleyi* strains was from 59-77%, showing that the population was largely comprised of *E. huxleyi*. Every sample that yielded COI amplicon sequences also yield MCP and SPT amplicon sequences. Notably, with the exception of ALOHA_2, each of these samples was collected at the lowest depth sampled, 107 m, where light was ~1% of the surface irradiance. All of these samples had relatively high mapping, the lowest being at 24% from >5,000 sequence reads. EhV SPT-like sequences were successfully amplified from the same samples as MCP. Read mapping was also high for SPTs, the lowest being at 36%, but all others being above 80%, all with >10,000 mapped reads. The relatively high values and low range in % mapping (other than ALOHA_2) for both MCP and SPT of sequence reads compared to SBC bloom samples suggested this EhV population was relatively homogenous, as might be expected in an oligotrophic open ocean environment (Hewson *et al.* 2006). Given that cell densities

were likely very low – historically, Station ALOHA experiences the most abundant *E. huxleyi* populations in March and August/September, only reaching concentrations of 2 x 10^4 cells L^{-1} (Cortés *et al.* 2001) – unlike many bloom conditions observed in the North Atlantic, our findings indicate that EhV infection still occurs at low host densities. Notably, having low host density would greatly impact host/virus contact rates, and it makes one question whether infection at these densities would proceed in a classic lytic fashion as observed in blooming populations, or if infection can be a chronic within low host density (Mackinder *et al.* 2009).

Less than half of the samples collected in waters near Japan successfully yielded COI, SPT and MCP amplicon sequences, suggesting limited host/virus activity at sampling sites (Table 4.4). It did encompass a much larger and more diverse geographic region than SBC and ALOHA sampling, as it included both coastal and open ocean samples spanning multiple prevailing current regions including the Kuroshio and Oyasho Currents. Thus, each station sampled could arguably represent independent environments. Eight of 25 samples successfully yielded *E.huxleyi/G.oceanica* COI-like sequences all with a >99% identity. The percent reads that uniquely mapped to *E. huxleyi* were generally high (above 80%) with the exception of two samples that came from the same location, JPN 12 and 13, both with low unique mapping, 37% and 39%, respectively. This confirmed that, though somewhat limited, *E. huxleyi* populations were present throughout the region.

Locations that successfully yielded COI amplicon sequences also yielded MCP and SPT amplicon sequences, paralleling the paired host-virus observations found at Station ALOHA. EhV MCP-like sequences were amplified from 8 samples and had

significant station overlap with the COI-like amplifications, both being amplified at the same six stations. However, mapping to known EhV MCPs was very low, with only one station, JPN_3, having high read mapping, at 73% and >100,000 mapped reads. All other samples expressed <1% mapping and <1,000 mapped reads. EhV SPT-like sequences were amplified from 14 samples, many of which overlapped with successful MCP and COI amplifications. In fact, no station had successful MCP or COI amplifications without also having an SPT amplification. Mapping for SPTs was also generally low, with only four samples recruiting >10% of their reads. The very low MCP and SPT % mapping suggested the majority of the DNA amplified with these primers was not the genes of interest. With such low mapping (<1%) in many of the locations around Japan, and the fact that the sample biomass included both free and host-associated viruses, it is difficult to confidently conclude that infection was occurring at many of these locations. However, some sequenced amplicons were identified as EhV DNA, particularly at station JPN 3-4, which had coincident COI, MCP, and SPT amplicon sequences all representing >30% of the reads mapped to reference host and virus genomes. Sample collected through the Bering Strait and Chukchi Sea did not amplify *E. huxleyi* COI-, EhV MCP-, or SPT-like sequences.

To explore the diversity of the sequences amplified throughout the Pacific in comparison to known MCP and SPTs, we clustered all mapped sequences with a 98% identity threshold. Clusters were then phylogenetically compared to sequences from 13 EhV isolated reference strains with sequenced genomes using the top 25 clusters with the most recruits. Clustering was performed on the gene level; however, phylogenic distance

was measured at both the gene and protein level to observe both genetic variation and structural differences between clusters and cultured EhV strains.

At the gene level, MCP clusters from this study were very similar to reference sequence counterparts (Figure 4.5). Clusters were generally interspersed among the reference sequences and not weighted towards being similar to any particular EhV strains. EhV 201 hosted the largest number of close relatives with the least genetic distance between them. In contrast, EhVs 202, 208, 207, 84, and 88 did not have relatives with particularly close genetic distance. Looking at the protein level tree, there is structurally little difference between MCPs (Figure 4.6). Bootstraps display a low level of confidence in the structure of the tree as a result of the high similarity among clusters and isolates. Indeed, in the protein sequence alignments, most clusters were identical or had single amino acid substitutions over the length of the alignment (not shown). From this, it is clear that the dominant novel MCP clusters throughout the Pacific are phylogenetically very similar to their North Atlantic counterparts, with many clusters grouping in with individual North Atlantic strains and translating nearly identical amino acid sequences between the Pacific and Atlantic.

Phylogenetic analysis revealed a greater evolutionary distance among retrieved SPT gene sequences than for MCP gene sequences (Figure 4.7). Unlike MCPs, SPT sequences from EhV reference strains were most similar to each other and distant from all of the novel, environmental SPT clusters. The one exception was the presence of six clusters with a close relationship to the sequences from EhV18, EhV156, and EhV202, which were identical across the 298 bp linker region of the SPT. The unique environmental SPT clusters formed three groups the largest containing 15 clusters with

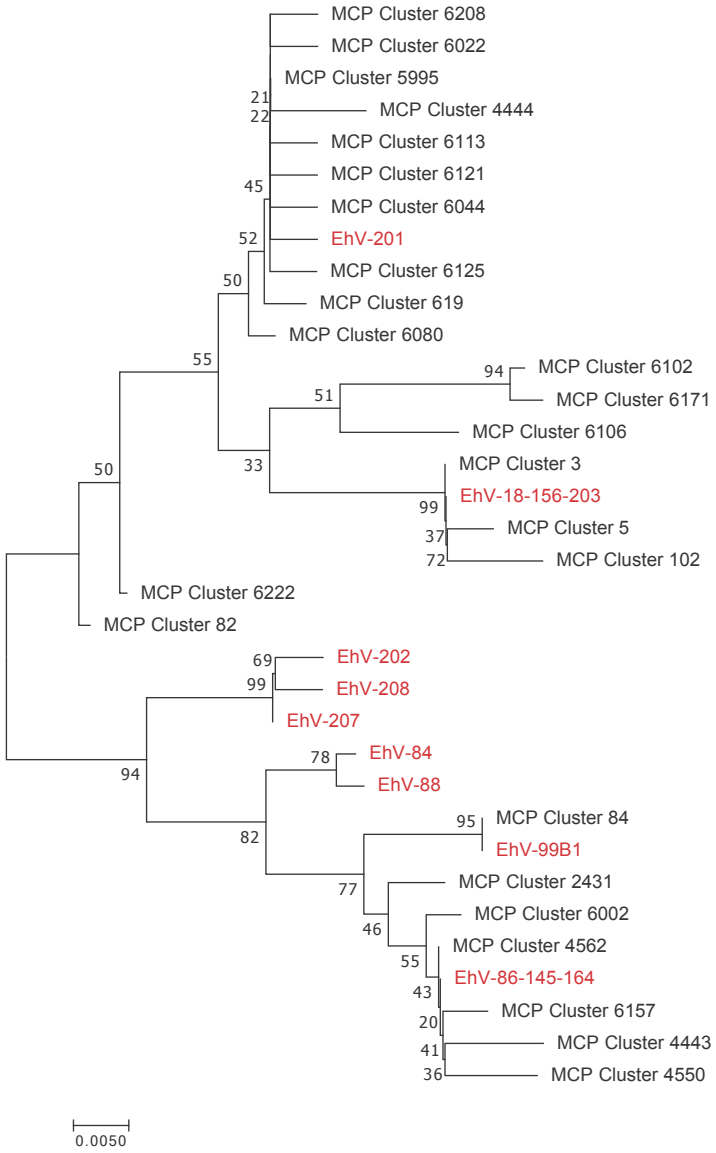


Figure 4.5 Phylogenic analysis of Pacific MCP gene clusters based on nucleotide sequences. Shown clusters are the 25 most abundantly represented clusters among all samples. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.25041180 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The analysis involved 34 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 237 positions in the final dataset. Known EhV MCP sequences are shown in red with strain numbers separated by hyphens when identical. Bootstraps display a relatively high level of confidence in the structure of the tree.

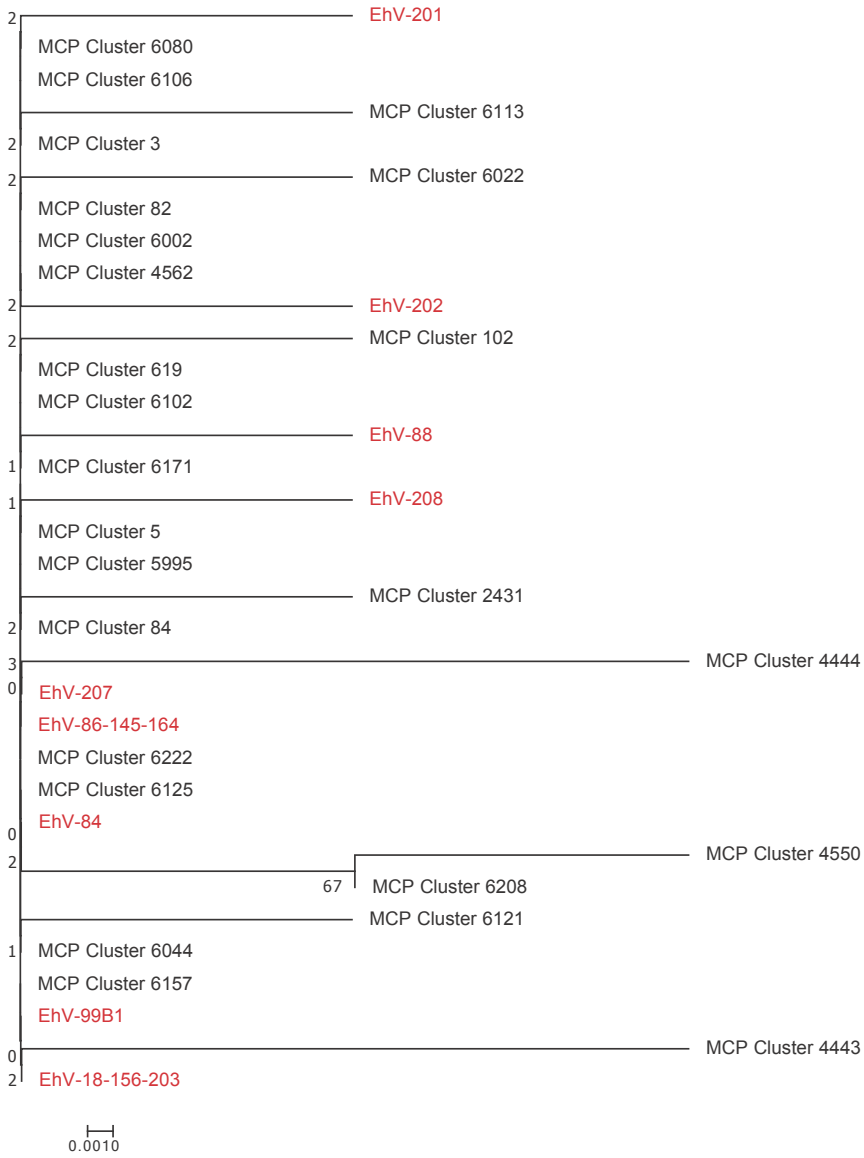


Figure 4.6 Phylogenic analysis of Pacific MCP gene clusters based on amino acid sequences. Shown clusters are the 25 most abundantly represented clusters among all samples. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.19484675 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved 34 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 78 positions in the final dataset. Known EhV MCP sequences are shown in red with strain numbers separated by hyphens when identical.

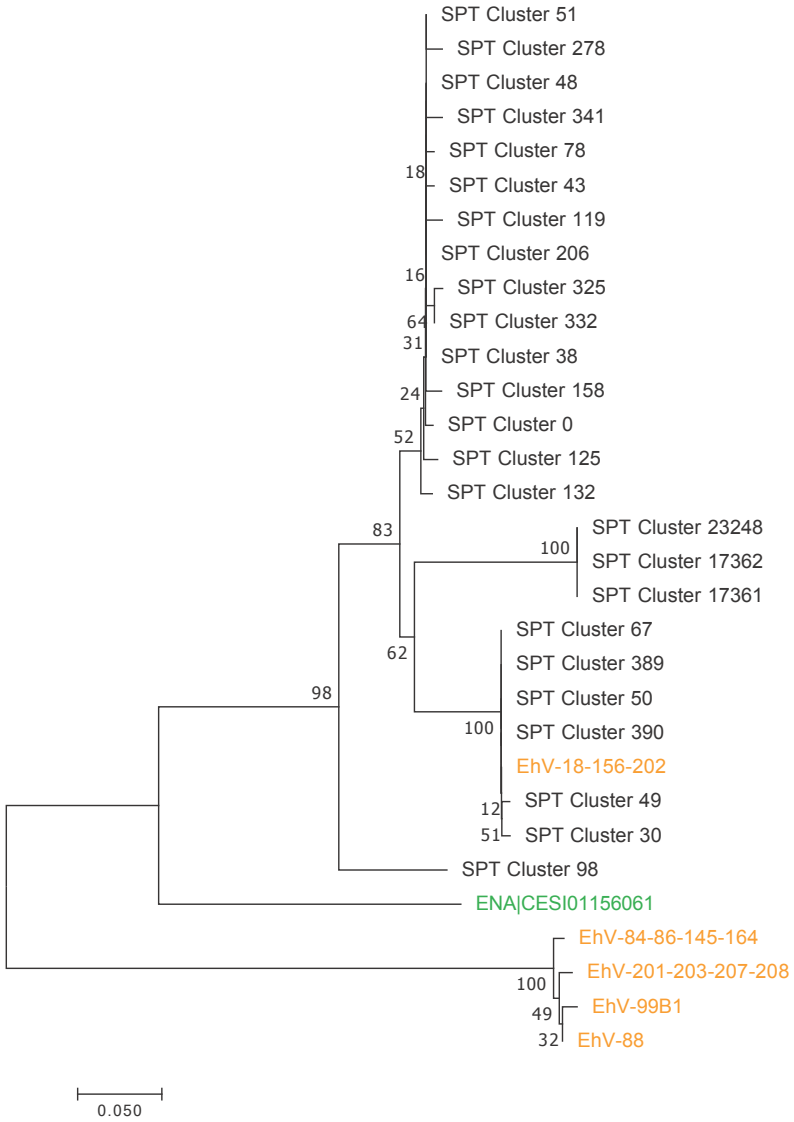


Figure 4.7. Phylogenic analysis of Pacific SPT gene clusters based on nucleotide sequences. Shown clusters are the 25 most abundantly clusters among all samples. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 1.08898006 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The analysis involved 31 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 211 positions in the final dataset. Known EhV SPT sequences are shown in orange with strain numbers separated by hyphens when identical. The SPT sequence extracted from the Tara Ocean metagenome assembly is shown in green.

low bootstrapping scores between them, the second containing 3 clusters, and the third clustering with EhV18/EhV156/EhV202. SPT Cluster 98 was relatively different from all other clusters, with many nucleic acid substitutions that manifested in amino acid substitutions (not shown). Of particular interest, Tara Oceans Station 82 contained an assembly that aligned with the a complete 322 bp SPT linker region (Figure 4.8). This assembly matched at 82% identity to SPTs of EhV18 and EhV156 and was also quite different from all other SPT sequence clusters and those obtained from EhV isolate reference strains, containing four insertions (all less than 12bp) toward the middle of the sequence.

A protein phylogenetic tree was also examined to see whether clusters organized differently at the amino acid level; however, the protein phylogeny looked similar to the

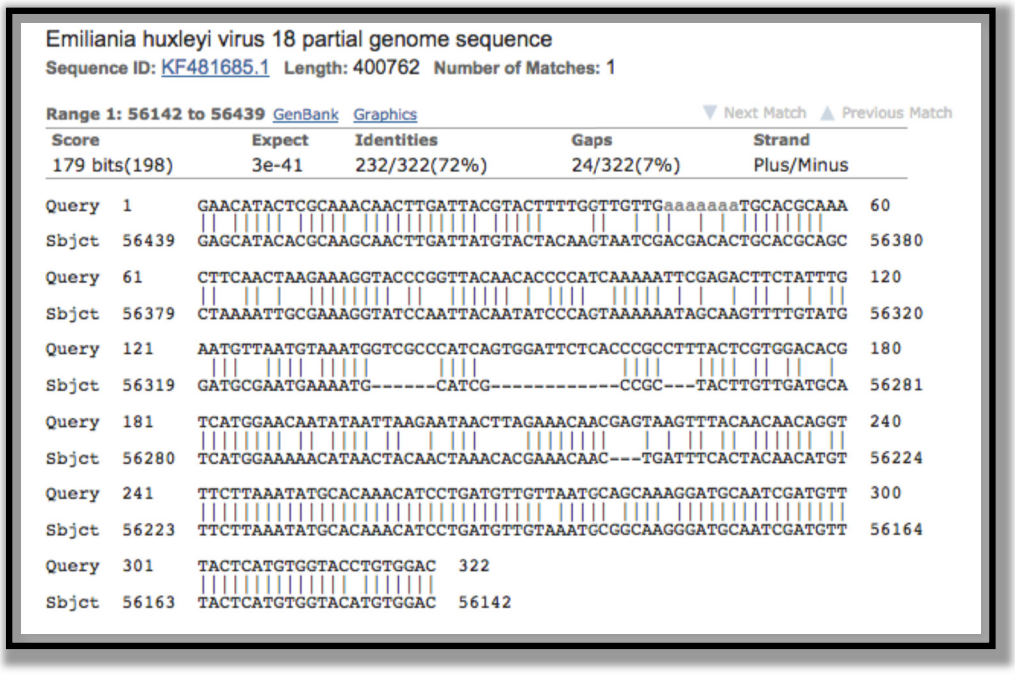


Figure 4.8. Alignment between the EhV18-encoded SPT gene linker region (Sbjct) and a partial gene assembly (Query) constructed from metagenome data collected at the Tara Ocean Station 82.

gene-based tree (Figure 4.9). One exception is that cluster 341 showed greater evolutionary distance at the protein level. Nodes separating the three major cluster groups were also organized such that the large group and the EhV18/EhV156/EhV202 group were more similar. The differences observed here may likely have an influence on the function of the SPT protein among EhV variants. The structure of the SPT protein is suggested to play a critical role in enzyme catalytic efficiency and regulating downstream viral glycosphingolipid production (Nissimov 2013) Given the structural differences observed in clusters from this study, it is evident that specific infection strategies may vary widely throughout the ocean.

There was a clear biogeographic organization to COI, MCP and SPT gene clusters, which provided geospatial context. Only 3% of the clusters were shared among all three regions. Meanwhile, all three regions contained MCP clusters that were uniquely expressed within the respective regions (Table 4.5). Santa Barbara Channel had the largest number of unique clusters with 113 and Japan had the smallest number with four, while ALOHA hosted 25 unique clusters. Looking at this on the local level with respect to the total number of clusters found in each region, unique clusters made up half of the total clusters in SBC. They also made up nearly half of the clusters at ALOH (42%); however, JPNs unique clusters only made up 4% of all clusters observed there. Overall, unique clusters from SBC made up 44% of all clusters while unique reads from ALOHA made up 10% and JPN only made up 2%. The remaining clusters were shared between two of the regions. The majority of these remaining clusters (31%) were shared between SBC and JPN, while a smaller portion (10%) were shared between ALOHA and SBC. There were no MCP clusters that were only shared between ALOHA and JPN.

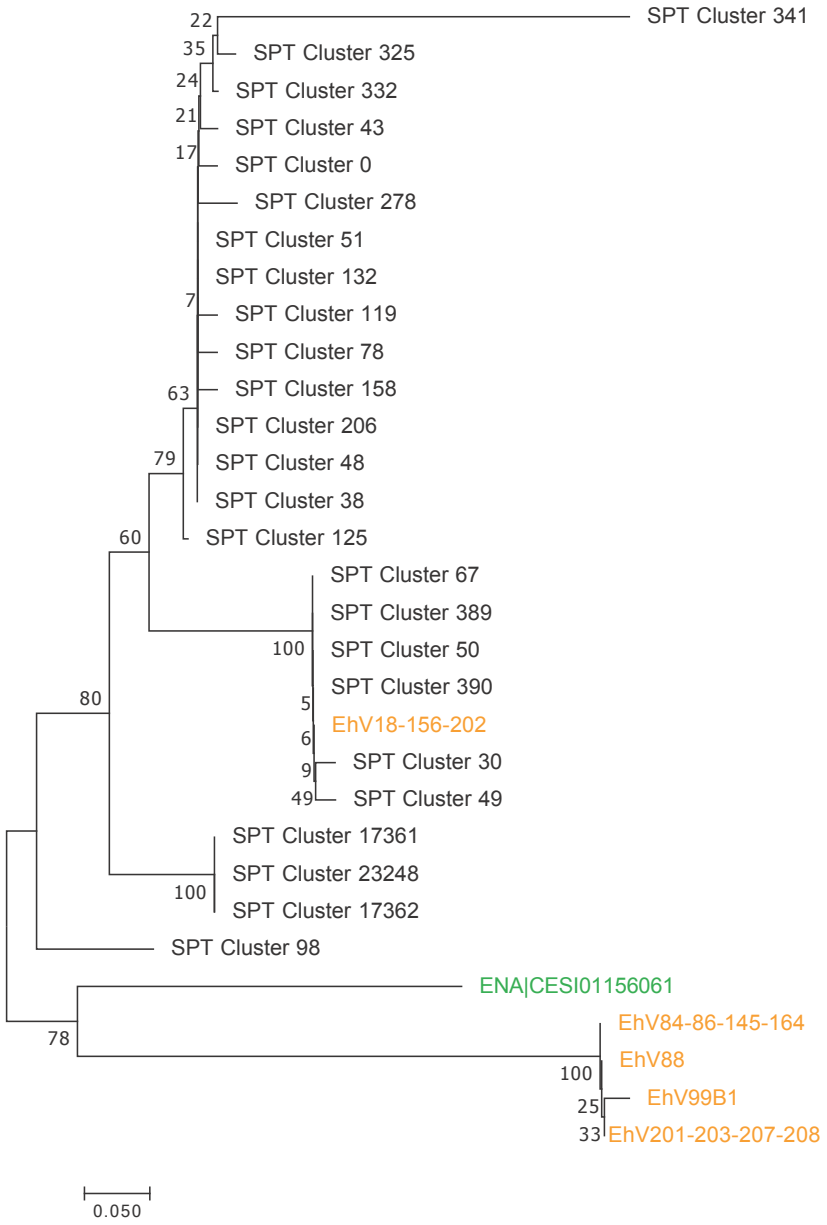


Figure 4.9 Phylogenic analysis of Pacific SPT gene clusters based on amino acid sequences. Shown clusters are the 25 most abundantly represented clusters among all samples. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 1.69114316 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved 31 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 68 positions in the final dataset. There were a total of 211 positions in the final dataset. Known EhV SPT sequences are shown in orange with strain numbers separated by hyphens when identical. The SPT sequence extracted from the Tara Ocean metagenome assembly is shown in green.

The regional appearances of SPT clusters reflected those observed for MCP clusters with some differences. Only 8% of clusters had representatives from all three regions. Similarly, SPT richness was highest at SBC with 184 clusters, followed by JPN with 145 and ALOHA with only 74 clusters (Table 4.5). However, no unique SPT clusters were observed in SBC. At ALOHA, unique SPT clusters made up 15% of the regional clusters and 5% of total clusters. Meanwhile, JPN also had few, representing only 3% of the regional clusters and 2% of total. Instead, a majority of clusters were shared between two regions. The clusters shared between only SBC and JPN made up 61% of total clusters, while SPT clusters shared between only ALOHA and SBC made up 22% of total clusters. As with MCP, few (<1%) clusters were shared exclusively between ALOHA and JPN.

Overall, SBC and JPN amplicons had a larger pool of shared MCP and SPT sequence clusters than ALOHA did with either sequence type. Meanwhile, Station ALOHA contained the largest number of unique SPT sequence clusters. Given the

Table 4.5 Distribution of gene clusters throughout the Pacific Ocean. Samples are divided into three regions, ALOHA (Station ALOHA), SBC (Santa Barbara Channel), and JPN (waters surrounding Japan). For individual regions, unique clusters are examined within the context of the region as well as throughout the ocean basin. Unique cosmopolitan clusters are those observed among all three regions. % Shared clusters are those shared exclusively with another region.

	MCP				SPT			
	ALOHA	SBC	JPN	Cosmopolitan	ALOHA	SBC	JPN	Cosmopolitan
Clusters	59	226	91	255	74	184	145	201
Unique Clusters	25	113	4	8	11	0	5	17
% Unique Local	42	50	4		15	0	3	
% Unique Total	10	44	2	3	5	0	2	8
Shared with SBC	26				45			
% Shared with SBC	10				22			
Shared with JPN		79				122		
% Shared with JPN		31				61		
Shared with ALOHA			0				1	
% Shared with ALOHA			0				0	

respective proximity of SBC and JPN regions to respective eastern and western boundary currents in the North Pacific gyre and ALOHA's placement in the oligotrophic gyre center, the observed sequence similarities reflect a physical connectivity of viral ecotypes between these component currents and environments. Further, the greater cluster similarity between SBC and ALOHA compared to JPN and ALOHA may indicate a North Pacific circulatory pattern with more water mass connectivity between the California current region and the central North Pacific gyre. Though transport dynamics of open ocean microbe populations are largely unexplored, South Atlantic Agulas rings circulated from ocean waters in the Indian Ocean have been identified as a mechanism injecting plankton populations between ocean basins, revealing semi-regular microbial transport from the Indo-Pacific to South Atlantic (Villar *et al.* 2015). Similarly, our results may reflect transport dynamics observed in the North Pacific.

4.4.3 Culture Infections. With the Tara Oceans assemblies and targeted MCP and SPT amplicon sequencing both indicating pervasive EhV infection of *E. huxleyi* throughout the Pacific, we wanted experimentally explore the dynamics of infection in Pacific *E. huxleyi* strains and directly test whether these strains of *E. huxleyi* would be functionally infected by EhVs isolated from the North Atlantic. These lab-based infections revealed a mixture of resistance and susceptibility between North Pacific host and North Atlantic virus strains. For *E. huxleyi* 70N, which was isolated from the Chukchi Sea, net host cell growth occurred over the duration of the experiment, with cell concentrations reaching 3.14×10^6 cells ml^{-1} at 144 hours post infection (hpi; Figure 4.10a). For experimental groups exposed to EhV1, EhV86 and EhV163, growth remained similar to the control up

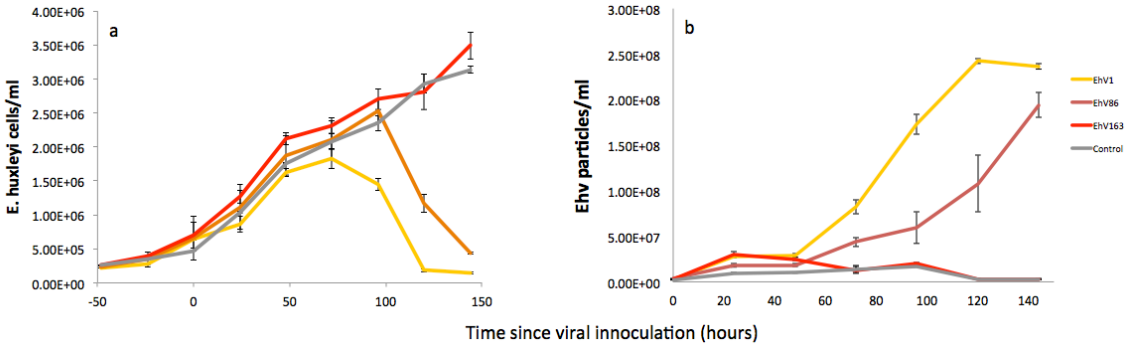


Figure 4.10 Dynamics of (a) host and (b) viral abundance during infection of *E. huxleyi* MR70N with EhV strains 1, 86, and 163. Time scale is based around T_0, when virus was inoculated into cell cultures to initiate infection. For each experimental group, n = 3. Error bars denote 1 SE.

to 48 hpi. Afterwards, EhV1-infectedcells had lower net growth rate by 72 hpi and decreasing growth by 96 hpi. A large reduction in cell concentration occurred between 96-120 hpi and continued to decrease by 144 hpi. EhV86 infections displayed host cell growth until 96 hpi without notable signs of infection; cell concentrations sharply decreased between 96-120 hpi and continued to decrease rapidly by 144 hpi. EhV163 infections neither crashed throughout the duration of the experiment nor displayed any decreased in growth rate, ending the experiment with a cell concentration higher than the control group.

Accompanying virus production in the uninfected control group over the course of the experiment was negligible, with no detectable counts above the instrument noise (Figure 4.10b). EhV1 production was observed as early as 24 hpi, with production slowing after 48 h and slowly increasing up to 2.42 x 10^8 virus ml^{-1} by 120 hpi before leveling out by 144 hpi. EhV86 production was also observed with an initial increase in viral particles 24 hpi with a similar lag in production between 24-48 h. Viral concentrations increased again between 48-144 hpi to 2.0 x 108 virus ml^{-1}, but not as

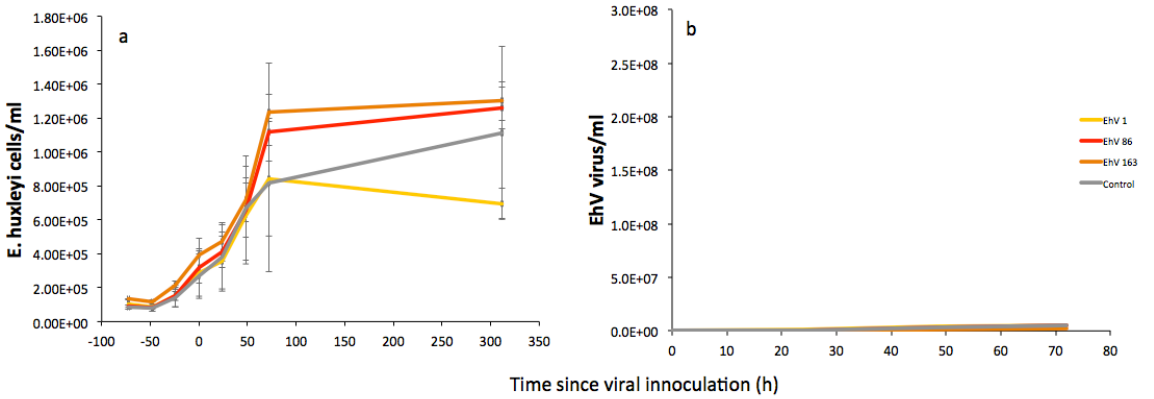


Figure 4.11 Dynamics of (a) host and (b) viral abundance during infection of *E. huxleyi* MR1311 with EhV strains 1, 86, and 163. Time scale is based around T_0, when virus was inoculated into cell cultures to initiate infection. For each experimental group, n = 3. Error bars denote 1 SE.

rapidly as EhV1. EhV163 infection was characterized by an initial increase in viral particles between 0-24 hpi after which no further increases were observed; concentrations were largely similar to the control group by 72 h.

Host growth and infection dynamics for *E. huxleyi* NIES 1311 varied significantly from those of 70N (Figure 4.11a). Growth remained similar for all four treatment groups. Notably high standard errors were observed throughout the experiment for all treatment groups. High standard errors were also observed at T_0, suggesting the error was not a result of some cultures infecting and others not. A final measurement of cell abundances 13 d after inoculating with virus hinted at mild infection by EhV1, as host cells were less abundant than the other two infected groups. However, no virus production was observed in any of the treatment groups for NIES 1311 throughout the incubation period (Figure 4.11b), arguing that infection did not occur. Similar to NIES 1311, *E. huxleyi* strains from the North Atlantic can express resistance to EhV infection, a trait attributed to suppressed sialic acid glycosphingolipid production (Fulton *et al.* 2014).

E. huxleyi 1516, which was isolated in the Pacific off the coast of Chile and whose genome was completely sequenced (Read *et al.* 2013), is also susceptible to EhVs (Mackinder *et al.* 2009), so it is not surprising that other Pacific strains can succumb to infection. This experimental work provides additional evidence that EhVs are capable of infecting hosts from the Pacific basin. Divergence in regional host diversity clearly does not act as a barrier to infection. As certain host strains (CCMP373, CCMP379) isolated from the Atlantic routinely display resistance to infection by the EhV strains explored here, selective resistance in NIES1311 was likely related to host strain physiology (Schroeder *et al.* 2002; Bidle *et al.* 2007; Bidle & Kwityn 2012) independent of region of isolation.

4.5 CONCLUSIONS

Our collective analyses of the global Tara Oceans metagenomics survey, targeted amplicon gene sequencing the Pacific Ocean basin, and culture infection experiments provide direct evidence of EhV infection on a global scale, not only in the North Atlantic. As a global phenomenon, EhVs likely have a notable influence on the fate of *E. huxleyi* populations and associated ecosystem dynamics and carbon biogeochemistry. It also gives credence to a derivative of the idea that 'everything is everywhere' where 'infection is everywhere' among a microbial host-virus system. While genetic differences are observed among locations, the ecological and evolutionary pressures of viral infection are persistent across ocean basins, though similarity of infection dynamics needs to be investigated.

We recently showed that EhV infection facilitates carbon export flux from the surface ocean into the mesopelagic in North Atlantic blooms of *E. huxleyi*, a process that involves the interaction of virally induced particle aggregation coupled with enhanced grazing of infected host cells, lubricating the movement of infected host material out of the surface ocean. It has also been shown that EhV lysis and associated grazing can enhance the cleavage of dimethylsulfoniopropionate (DMSP), which alongside its breakdown product dimethylsulphide (DMS) is associated with is associated with *E. huxleyi* blooms (Malin *et al.* 1993; Evans *et al.* 2007). It is likely that the influence of infection on the surrounding environment acts differently through low-density population than in dense bloom conditions (Bratbak *et al.* 1995), though studying these conditions can prove difficult. Regardless, the presence of EhVs on a global scale reveals that they are intimately connected with the fate of *E. huxleyi* biomass at scales not previously recognized, impacting both the global PIC and POC cycles.

Chapter 5: Conclusions

The known roles and diversity of marine viruses are rapidly expanding in the ocean system (Brum *et al.* 2015; Roux *et al.* 2016; Philosof *et al.* 2017). At 10^9 to 10^{10} viruses l^{-1}, they play a critical role in regulating primary production, accounting for the turnover of 25% of the primary production in the ocean (Fuhrman 1999). Over the past several decades *E. huxleyi*-EhV host-virus interactions have been extensively studied, expanding the knowledge of infection strategy and dynamics, making it a model for studying infection in marine algae (Schroeder *et al.* 2003; Mackinder *et al.* 2009; Bidle & Vardi 2011). This dissertation has further elucidated the intimate relationship observed in this host virus system and identifies ecosystem processes and biogeochemical pathways stimulated by this interaction in natural phytoplankton communities.

In Chapter II, North Atlantic blooms of *E. huxleyi* encompassed in distinct mesoscale eddies were interrogated to test the hypothesis that viral infection stimulates the movement of carbon out of the surface ocean and into the mesopelagic. The multifaceted sampling approach revealed these populations experiencing distinct stages of infection and bloom termination, ranging from early infection bloom, late infection and bloom decline, to a post-infection terminated bloom. Optical and sediment trap measurements of carbon flux and remineralization rates below the euphotic zone revealed the highest flux rates occurring below the early infection bloom, which was coupled with enhanced organic carbon remineralization. Sinking material collected by sediment traps at 50m, 150m, and 300m, contained enhanced representation of virally infected cellular material relative to water column measurements, suggesting that infected cells were preferentially being removed from the surface ocean. This observation is explained by

the coupled interaction of cell and particle aggregation by the production of TEP ballasted the calcium carbonate coccoliths, and the selective grazing of *E. huxleyi* by microzooplankton. These results suggest that through complex ecosystem interactions, EhV infection enhances carbon flux out of the surface ocean, revealing a novel pathway for viral biogeochemistry.

Chapter III further explored these North Atlantic blooms but from the perspective of the total phytoplankton community. The utility of phytoplankton absorption spectra was investigated to assess environmental and photoacclimation state within mixed phytoplankton communities. Photoacclimation played a critical role in controlling cell pigment quotas and composition. Variations in the explored absorption characteristics were largely due to photoacclimation state with limited influence by differences in taxonomic composition. Combined with laboratory studies of the plasticity of spectral absorption for an individual species, given variable environmental conditions and stresses including light acclimation, nutrient limitation, and viral infection, it is apparent that environmental conditions can play a significant role in shaping phytoplankton absorption spectra characteristics in addition to taxonomic assemblage.

In Chapter IV, the study of EhV infection was expanded from the North Atlantic to the global ocean using genetic tools to detect the specific presence of EhVs in diverse oceanographic water masses. Tara Ocean metagenome assemblies confirmed the presence of EhV-like genes in every major ocean basin and revealed the potential diversity present in the EhV genome, largely supported by abundant and diverse sequences discovered in a large coccolithophore bloom in the South Atlantic. Targeted sampling throughout the North Pacific confirmed the presence of EhVs in a dynamic

range of surface ocean ecosystems including the oligotrophic central gyre as well as bloom and non-bloom conditions in and around boundary current regions. The MCP and SPT genes used to identify the EhVs revealed conservative and non-conservative tendencies, respectively, between these genes throughout the ocean, which may indicate functional differences in the SPT partial gene sequence.

It is evident through these studies that EhV infection is a process that is widespread throughout the ocean. It is therefore likely that seasonal blooms occurring in the areas such as the Bering Sea, and the southern calcite belt observe similar infection dynamics as those in the North Atlantic. With EhV infection acting as a mechanism to enhance vertical flux, the interplay between host and virus may have an interesting impact on the biological pump. Considering that coccolithophores also contribute to the carbonate counter pump in addition to the organic carbon pump (Rost & Riebesell 2004), widespread viral infection may tip the balance of the rain ratio towards the export of organic carbon and away from inorganic carbon in two ways. First through suppressing population growth through cell lysis (thus preventing the formation of $CaCO_3$), and second through the ecosystem processes that lead to preferential removal of EhV infected cell material.

This dissertation broadens current knowledge on how viral infection can shape the surrounding ecosystem and influence ocean biogeochemistry and also explores photoacclimation and phytoplankton absorption characteristics in communities surrounding these processes. It is evident that the effects of host-virus interactions are not limited to the host-virus system and may ripple through the ecosystem in profound ways. For host-virus systems with broad biogeographic boundaries, it will be necessary

to investigate how differences in ecosystem and population dynamics impact the environmental effect of infection. As the most abundant and diverse entities in the ocean, the versatility of their functions is not likely small.

References

1.
Alkire, M.B., D'Asaro, E., Lee, C., Jane Perry, M., Gray, A., Cetinić, I. *et al.* (2012). Estimates of net community production and export using high-resolution, Lagrangian measurements of O_2, NO_3^-, and POC through the evolution of a spring diatom bloom in the North Atlantic. *Deep Sea Research Part I: Oceanographic Research Papers*, 64, 157-174.
2.
Alldredge, A.L., Passow, U. & Haddock, H. (1998). The characteristics and transparent exopolymer particle (TEP) content of marine snow formed from thecate dinoflagellates. *Journal of Plankton Research*, 20, 393-406.
3.
Anderson, L.A. & Sarmiento, J.L. (1994). Redfield ratios of remineralization determined by nutrient data analysis. *Global biogeochemical cycles*, 8, 65-80.
4.
Azam, F. (1998). Microbial control of oceanic carbon flux: the plot thickens. *Science*, 280, 694-696.
5.
Azetsu-Scott, K. & Passow, U. (2004). Ascending marine particles: Significance of transparent exopolymer particles (TEP) in the upper ocean. *Limnology and Oceanography*, 49, 741-748.
6.
Balch, W.M., Bates, N.R., Lam, P.J., Twining, B.S., Rosengard, S.Z., Bowler, B.C. *et al.* (2016). Factors regulating the Great Calcite Belt in the Southern Ocean and its biogeochemical significance. *Global Biogeochemical Cycles*, 30, 1124-1144.
7.
Balch, W.M., Kilpatrick, K.A., Holligan, P., Harbour, D. & Fernandez, E. (1996). The 1991 coccolithophore bloom in the central North Atlantic. 2. Relating optics to coccolith concentration. *Limnology and Oceanography*, 41, 1684-1696.
8.
Balch, W.M., Vaughn, J.M., Novotny, J.F., Drapeau, D.T., Goes, J.I., Booth, E. *et al.* (2002). Fundamental changes in light scattering associated with infection of marine bacteria by bacteriophage. *Limnology and oceanography*, 47, 1554-1561.
9.
Bass, D. & Boenigk, J. (2011). Everything is everywhere: a twenty-first century de-/reconstruction with respect to protists. *Biogeography of microscopic organisms: Is everything small everywhere*, 88-110.

10.
Berelson, W., Balch, W., Najjar, R., Feely, R., Sabine, C. & Lee, K. (2007). Relating estimates of CaCO3 production, export, and dissolution in the water column to measurements of CaCO3 rain into sediment traps and dissolution on the sea floor: A revised global carbonate budget. *Global Biogeochemical Cycles*, 21.
11.
Berges, J.A., Charlebois, D.O., Mauzerall, D.C. & Falkowski, P.G. (1996). Differential effects of nitrogen limitation on photosynthetic efficiency of photosystems I and II in microalgae. *Plant Physiology*, 110, 689-696.
12.
Bidigare, R.R., Smith, R., Baker, K. & Marra, J. (1987a). Oceanic primary production estimates from measurements of spectral irradiance and pigment concentrations. *Global Biogeochemical Cycles*, 1, 171-186.
13.
Bidigare, R.R., Smith, R.C., Baker, K.S. & Marra, J. (1987b). Oceanic primary production estimates from measurements of spectral irradiance and pigment concentrations. *Global biogeochemestry Cycles*, 1, 171-186.
14.
Bidle, K.D. (2015). The Molecular Ecophysiology of Programmed Cell Death in Marine Phytoplankton. *Annual Review of Marine Science*, 7, null.
15.
Bidle, K.D., Haramaty, L., e Ramos, J.B. & Falkowski, P. (2007). Viral activation and recruitment of metacaspases in the unicellular coccolithophore, Emiliania huxleyi. *Proceedings of the National Academy of Sciences*, 104, 6049-6054.
16.
Bidle, K.D. & Kwityn, C.J. (2012). Assessing the role of caspase activity and metacaspase expression on viral susceptibility of the coccolithophore, Emiliania huxleyi (Haptophyta). *Journal of Phycology*, 48, 1079-1089.
17.
Bidle, K.D. & Vardi, A. (2011). A chemical arms race at sea mediates algal host–virus interactions. *Current opinion in microbiology*, 14, 449-457.
18.
Bligh, E.G. & Dyer, W.J. (1959). A rapid method of total lipid extraction and purification. *Canadian journal of biochemistry and physiology*, 37, 911-917.
19.
Bracher, A., Bouman, H.A., Brewin, R.J., Bricaud, A., Brotas, V., Ciotti, A.M. *et al.* (2017). Obtaining phytoplankton diversity from ocean color: a scientific roadmap for future development. *Frontiers in Marine Science*, 4, 55.
20.
Bratbak, G., Egge, J.K. & Heldal, M. (1993). Viral mortality of the marine alga Emiliania huxleyi (Haptophyceae) and termination of algal blooms. *Marine Ecology Progress Series*, 93.

21.
Bratbak, G., Levasseur, M., Michaud, S., Cantin, G., Fernández, E., Heimdal, B.R. *et al.* (1995). Viral activity in relation to Emiliania huxleyi blooms: a mechanism of DMSP release? *Marine Ecology Progress Series*, 133-142.
22.
Bratbak, G., Wilson, W. & Heldal, M. (1996). Viral control of< i> Emiliania huxleyi</i> blooms? *Journal of Marine Systems*, 9, 75-81.
23.
Bricaud, A., Claustre, H., Ras, J. & Oubelkheir, K. (2004). Natural variability of phytoplanktonic absorption in oceanic waters: Influence of the size structure of algal populations. *Journal of Geophysical Research: Oceans*, 109.
24.
Briggs, N., Perry, M.J., Cetinić, I., Lee, C., D'Asaro, E., Gray, A.M. *et al.* (2011). High-resolution observations of aggregate flux during a sub-polar North Atlantic spring bloom. *Deep Sea Research Part I: Oceanographic Research Papers*, 58, 1031-1039.
25.
Brown, C.W. & Yoder, J.A. (1994). Coccolithophorid blooms in the global ocean. *Journal of Geophysical Research: Oceans (1978–2012)*, 99, 7467-7482.
26.
Brum, J.R., Ignacio-Espinoza, J.C., Roux, S., Doulcier, G., Acinas, S.G., Alberti, A. *et al.* (2015). Patterns and ecological drivers of ocean viral communities. *Science*, 348, 1261498.
27.
Brussaard, C.P., Payet, J.P., Winter, C. & Weinbauer, M.G. (2010). Quantification of aquatic viruses by flow cytometry. *Manual of aquatic viral ecology*, 11, 102-107.
28.
Castillo, C.R., Sarmento, H., Alvarez-Salgado, X.A., Gasol, J.M. & Marraséa, C. (2010). Production of chromophoric dissolved organic matter by marine phytoplankton. *Limnology and Oceanography*, 55, 446-454.
29.
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 Biogeochemical Cycles*, 29, 1471-1494.
30.
Coolen, M.J.L. (2011). 7000 years of Emiliania huxleyi viruses in the Black Sea. *Science*, 333, 451-452.
31.
Cortés, M.Y., Bollmann, J. & Thierstein, H.R. (2001). Coccolithophore ecology at the HOT station ALOHA, Hawaii. *Deep Sea Research Part II: Topical Studies in Oceanography*, 48, 1957-1981.
32.
Denman, K. & Gargett, A. (1995). Biological-physical interactions in the upper ocean: the role of vertical and small scale transport processes. *Annual Review of Fluid Mechanics*, 27, 225-256.

33.
DiTullio, G. & Geesey, M.E. (2002). Photosynthetic pigments in marine algae and bacteria. *Encyclopedia of environmental microbiology*.
34.
Drake, L.A., Doblin, M.A. & Dobbs, F.C. (2007). Potential microbial bioinvasions via ships' ballast water, sediment, and biofilm. *Marine pollution bulletin*, 55, 333-341.
35.
Duysens, L. (1956). Energy transformations in photosynthesis. *Annual review of plant physiology*, 7, 25-50.
36.
Eisner, L.B., Twardowski, M.S., Cowles, T.J. & Perry, M.J. (2003). Resolving phytoplankton photoprotective: photosynthetic carotenoid ratios on fine scales using in situ spectral absorption measurements. *Limnology and Oceanography*, 48, 632-646.
37.
Evans, C., Archer, S.D., Jacquet, S. & Wilson, W.H. (2003). Direct estimates of the contribution of viral lysis and microzooplankton grazing to the decline of a Micromonas spp. population. *Aquatic Microbial Ecology*, 30, 207-219.
38.
Evans, C., Kadner, S.V., Darroch, L.J., Wilson, W.H., Liss, P.S. & Malin, G. (2007). The relative significance of viral lysis and microzooplankton grazing as pathways of dimethylsulfoniopropionate (DMSP) cleavage: an Emiliania huxleyi culture study. *Limnology and Oceanography*, 52, 1036-1045.
39.
Evans, C. & Wilson, W.H. (2008). Preferential grazing of Oxyrrhis marina on virus infected Emiliania huxleyi. *Limnology and Oceanography*, 53, 2035-2040.
40.
Falkowski, P.G. & Raven, J. (1997). An introduction to photosynthesis in aquatic systems. *Aquatic Photosynthesis. Blackwell Science*, 1-32.
41.
Field, C.B., Behrenfeld, M.J., Randerson, J.T. & Falkowski, P. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science*, 281, 237-240.
42.
Frada, M.J., Schatz, D., Farstey, V., Ossolinski, J.E., Sabanay, H., Ben-Dor, S. *et al.* (2014). Zooplankton may serve as transmission vectors for viruses infecting algal blooms in the ocean. *Current Biology*, 24, 2592-2597.
43.
Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28, 3150-3152.
44.
Fuhrman, J.A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature*, 399, 541-548.

45.
Fulton, J.M., Fredricks, H.F., Bidle, K.D., Vardi, A., Kendrick, B.J., DiTullio, G.R. *et al.* (2014). Novel molecular determinants of viral susceptibility and resistance in the lipidome of Emiliania huxleyi. *Environmental microbiology*, 16, 1137-1149.
46.
Gardner, W.D., Walsh, I.D. & Richardson, M.J. (1993). Biophysical forcing of particle production and distribution during a spring bloom in the North Atlantic. *Deep Sea Research*, 40, 171-195.
47.
Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S. *et al.* (2016). Plankton networks driving carbon export in the oligotrophic ocean. *Nature*.
48.
Hagino, K., Okada, H. & Matsuoka, H. (2005). Coccolithophore assemblages and morphotypes of< i> Emiliania huxleyi</i> in the boundary zone between the cold Oyashio and warm Kuroshio currents off the coast of Japan. *Marine Micropaleontology*, 55, 19-47.
49.
Harris, R. (1994). Zooplankton grazing on the coccolithophore Emiliania huxleyi and its role in inorganic carbon flux. *Marine Biology*, 119, 431-439.
50.
Henriksen, P., Riemann, B., Kaas, H., Sorensen, H.M. & Sorensen, H.L. (2002). Effects of nutrient-limitation and irradiance on marine phytoplankton pigments. *Journal of Plankton Research*, 24, 835-858.
51.
Hewson, I., Steele, J.A., Capone, D.G. & Fuhrman, J.A. (2006). Temporal and spatial scales of variation in bacterioplankton assemblages of oligotrophic surface waters. *Marine Ecology Progress Series*, 311, 67-77.
52.
Hoepffner, N. & Sathyendranath, S. (1991). Effect of Pigment Composition on Absorption Properties of Phytoplankton. *Marine Ecology Progress Series*, 73, 11-23.
53.
Holligan, P.M., Fernández, E., Aiken, J., Balch, W.M., Boyd, P., Burkill, P.H. *et al.* (1993). A biogeochemical study of the coccolithophore, Emiliania huxleyi, in the North Atlantic. *Global Biogeochemical Cycles*, 7, 879-900.
54.
Huisman, J., Thi, N.N.P., Karl, D.M. & Sommeijer, B. (2006). Reduced mixing generates oscillations and chaos in the oceanic deep chlorophyll maximum. *Nature*, 439, 322-325.
55.
Hunter, J.E., Frada, M.J., Fredricks, H.F., Vardi, A. & Van Mooy, B.A. (2015). Targeted and untargeted lipidomics of Emiliania huxleyi viral infection and life cycle phases highlights molecular biomarkers of infection, susceptibility, and ploidy. *Frontiers in Marine Science*, 2, 81.

56.
Johns, C.T., Nissimov, J.I., Natale, F., Knapp, V., Mui, A.F., H., Mooy, B.A.S.V. *et al.* (2018). The mutual interplay between calcification and coccolithovirus infection. *Environ Microbiol*, (In Review).
57.
Kahl, L.A., Schofield, O. & Fraser, W.R. (2010). Autonomous gliders reveal features of the water column associated with foraging by Adélie penguins. *Integrative and comparative biology*, icq098.
58.
Kimmance, S.A., Allen, M.J., Pagarete, A., Martínez, J.M. & Wilson, W.H. (2014). Reduction in photosystem II efficiency during a virus-controlled Emiliania huxleyi bloom. *Marine Ecology Progress Series*, 495, 65-76.
59.
Kolber, Z.S., Prášil, O. & Falkowski, P.G. (1998). Measurements of variable chlorophyll fluorescence using fast repetition rate techniques: defining methodology and experimental protocols. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1367, 88-106.
60.
Laber, C.P., Hunter, J.E., Carvalho, A.F., Collins, J.R., Hunter, E., Schieler, B. *et al.* (2018). Coccolithovirus facilitation of carbon export in the North Atlantic. *Nat Microbiol*.
61.
Ladd, C., Salo, S. & Eisner, L. (2016). Coccolithophores in the Bering Sea. In: *EBS Ecosystem Considerations 2016* (eds. Zador, S & Siddon, E). National Oceanic and Atmospheric Administration
National Marine Fisheries Service
Alaska Fisheries Science Center Seattle, WA, USA, pp. 87-89.
62.
Lamborg, C., Buesseler, K., Valdes, J., Bertrand, C., Bidigare, R., Manganini, S. *et al.* (2008). The flux of bio-and lithogenic material associated with sinking particles in the mesopelagic "twilight zone" of the northwest and North Central Pacific Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, 55, 1540-1563.
63.
Landry, M. & Hassett, R. (1982). Estimating the grazing impact of marine micro-zooplankton. *Marine biology*, 67, 283-288.
64.
Larsen, J.B., Larsen, A., Bratbak, G. & Sandaa, R.A. (2008). Phylogenetic analysis of members of the Phycodnaviridae virus family, using amplified fragments of the major capsid protein gene. *Applied and environmental microbiology*, 74, 3048-3057.
65.
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. *Current Biology*, 24, 2041-2046.

66.
Lévy, M., Shankar, D., André, J.M., Shenoi, S., Durand, F. & de Boyer Montégut, C. (2007). Basin‐wide seasonal evolution of the Indian Ocean's phytoplankton blooms. *Journal of Geophysical Research: Oceans*, 112.
67.
Llewellyn, C.A., Evans, C., Airs, R.L., Cook, I., Bale, N. & Wilson, W.H. (2007). The response of carotenoids and chlorophylls during virus infection of *Emiliania huxleyi* (Prymnesiophyceae). *Journal of Experimental Marine Biology and Ecology*, 344, 101-112.
68.
Mackey, M., Mackey, D., Higgins, H. & Wright, S. (1996). CHEMTAX-a program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton. *Marine Ecology Progress Series*, 144, 265-283.
69.
Mackinder, L.C., Worthy, C.A., Biggi, G., Hall, M., Ryan, K.P., Varsani, A. *et al.* (2009). A unicellular algal virus, Emiliania huxleyi virus 86, exploits an animal-like infection strategy. *Journal of general virology*, 90, 2306-2316.
70.
Malin, G., Turner, S., Liss, P., Holligan, P. & Harbour, D. (1993). Dimethylsulphide and dimethylsulphoniopropionate in the Northeast Atlantic during the summer coccolithophore bloom. *Deep Sea Research Part I: Oceanographic Research Papers*, 40, 1487-1508.
71.
Martínez, J.M., Schroeder, D.C. & 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 microbiology ecology*, 81, 315-323.
72.
Martz, T.R., Johnson, K.S. & Riser, S.C. (2008). Ocean metabolism observed with oxygen sensors on profiling floats in the South Pacific. *Limnology and Oceanography*, 53, 2094.
73.
McDonnell, A.M. & Buesseler, K.O. (2012). A new method for the estimation of sinking particle fluxes from measurements of the particle size distribution, average sinking velocity, and carbon content.
74.
Millero, F.J. & Poisson, A. (1981). International one-atmosphere equation of state of seawater. *Deep Sea Research Part A. Oceanographic Research Papers*, 28, 625-629.
75.
Millie, D., Schofield, O., Kirkpatrick, G., Johnsen, G. & Evens, T. (2002). Using absorbance and fluorescence spectra to discriminate microalgae. *European Journal of Phycology*, 37, 313-322.

76.
Minoshima, K., Kawahata, H. & Ikehara, K. (2007). Changes in biological production in the mixed water region (MWR) of the northwestern North Pacific during the last 27 kyr. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 254, 430-447.
77.
Monier, A., Pagarete, A., de Vargas, C., Allen, M.J., Claverie, J.-M. & Ogata, H. (2009). Horizontal gene transfer of an entire metabolic pathway between a eukaryotic alga and its DNA virus. *Genome research*, 19, 1441-1449.
78.
Moore, C.M., Suggett, D.J., Hickman, A.E., Kim, Y.-N., Tweddle, J.F., Sharples, J. *et al.* (2006). Phytoplankton photoacclimation and photoadaptation in response to environmental gradients in a shelf sea. *Limnology and Oceanography*, 51, 936-949.
79.
Morel, A. & Bricaud, A. (1981). Theoretical Results Concerning Light-Absorption in a Discrete Medium, and Application to Specific Absorption of Phytoplankton. *Deep-Sea Res*, 28, 1375-1393.
80.
Morgan, P., Pender, L. (1998). sw_satO2. In: *CSIRO MatLAB Seawater Library*. The CSIRO Division of Oceanography http://www.marine.csiro.au/~morgan/seawater.
81.
Morison, F. & Menden‐Deuer, S. (2015). Early spring phytoplankton dynamics in the subpolar North Atlantic: The influence of protistan herbivory. *Limnology and Oceanography*, 60, 1298-1313.
82.
Nelson, N., Siegel, D. & Michaels, A. (1998). Seasonal dynamics of colored dissolved material in the Sargasso Sea. *Deep Sea Research Part I: Oceanographic Research Papers*, 45, 931-957.
83.
Nissimov, J.I. (2013). Ecological and functional biodiversity in a marine algal-virus system: genotypes, phenotypes and their ecological significance. University of Nottingham.
84.
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. *Journal of virology*, 86, 2896-2897.
85.
Nissimov, J.I.G., R.; Johns, C. T.; Natale, F.; Haramaty, L.; Bidle, K. D. (2018). Dynamics of transparent exopolymeric particle production and cell aggregation during viral infection of the coccolithophore, Emiliania huxleyi. *Environ Microbiol*, (In revision).

86.
Paasche, E. (2001). A review of the coccolithophorid Emiliania huxleyi (Prymnesiophyceae), with particular reference to growth, coccolith formation, and calcification-photosynthesis interactions. *Phycologia*, 40, 503-529.
87.
Passow, U. & Alldredge, A.L. (1999). Do transparent exopolymer particles (TEP) inhibit grazing by the euphausiid Euphausia pacifica? *Journal of Plankton Research*, 21, 2203-2217.
88.
Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G. *et al.* (2015). Open science resources for the discovery and analysis of Tara Oceans data. *Scientific data*, 2, 150023.
89.
Peterson, M.L., Wakeham, S.G., Lee, C., Askea, M.A. & Miquel, J.C. (2005). Novel techniques for collection of sinking particles in the ocean and determining their settling rates. *Limnology and Oceanography: Methods*, 3, 520-532.
90.
Philosof, A., Yutin, N., Flores-Uribe, J., Sharon, I., Koonin, E.V. & Béjà, O. (2017). Novel abundant oceanic viruses of uncultured marine group II Euryarchaeota. *Current Biology*, 27, 1362-1368.
91.
Ploug, H., Iversen, M.H., Koski, M. & Buitenhuis, E.T. (2008). Production, oxygen respiration rates, and sinking velocity of copepod fecal pellets: direct measurements of ballasting by opal and calcite. *Limnology and Oceanography*, 53, 469-476.
92.
Popendorf, K.J., Fredricks, H.F. & Van Mooy, B.A. (2013). Molecular ion-independent quantification of polar glycerolipid classes in marine plankton using triple quadrupole MS. *Lipids*, 48, 185-195.
93.
Puerta, M.V.S., Bachvaroff, T.R. & Delwiche, C.F. (2005). The complete plastid genome sequence of the haptophyte Emiliania huxleyi: a comparison to other plastid genomes. *DNA Research*, 12, 151-156.
94.
Read, B.A., Kegel, J., Klute, M.J., Kuo, A., Lefebvre, S.C., Maumus, F. *et al.* (2013). Pan genome of the phytoplankton Emiliania underpins its global distribution. *Nature*, 499, 209-213.
95.
Rochelle-Newall, E., Delille, B., Frankignoulle, M., Gattuso, J.-P., Jacquet, S., Riebesell, U. *et al.* (2004). Chromophoric dissolved organic matter in experimental mesocosms maintained under different pCO2 levels. *Marine Ecology Progress Series*, 272, 25-31.

96.
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. *The Plant Cell*, 26, 2689-2707.
97.
Rost, B. & Riebesell, U. (2004). Coccolithophores and the biological pump: responses to environmental changes. In: *Coccolithophores*. Springer, pp. 99-125.
98.
Rousseaux, C.S. & Gregg, W.W. (2013). Interannual variation in phytoplankton primary production at a global scale. *Remote Sensing*, 6, 1-19.
99.
Roux, S., Brum, J.R., Dutilh, B.E., Sunagawa, S., Duhaime, M.B., Loy, A. *et al.* (2016). Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature*, 537, 689.
100.
Schatz, D., Shemi, A., Rosenwasser, S., Sabanay, H., Wolf, S.G., Ben‐Dor, S. *et al.* (2014). Hijacking of an autophagy‐like process is critical for the life cycle of a DNA virus infecting oceanic algal blooms. *New Phytologist*, 204, 854-863.
101.
Schroeder, D.C., Oke, J., Hall, M., Malin, G. & Wilson, W.H. (2003). Virus succession observed during an Emiliania huxleyi bloom. *Applied and Environmental Microbiology*, 69, 2484-2490.
102.
Schroeder, D.C., Oke, J., Malin, G. & Wilson, W.H. (2002). Coccolithovirus (Phycodnaviridae): characterisation of a new large dsDNA algal virus that infects Emiliana huxleyi. *Archives of virology*, 147, 1685-1698.
103.
Sharoni, S., Trainic, M., Schatz, D., Lehahn, Y., Flores, M.J., Bidle, K.D. *et al.* (2015). Infection of phytoplankton by aerosolized marine viruses. *Proceedings of the National Academy of Sciences*, 112, 6643-6647.
104.
Sheyn, U., Rosenwasser, S., Lehahn, Y., Barak-Gavish, N., Bidle, K., Koren, I. *et al.* (2017). Dual gene expression profiling reveals active infection below the pycnocline during a coccolithophore bloom in the North Atlantic. *The ISME journal*, in press.
105.
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. *The ISME journal*, 1.
106.
Siegel, D.A., Fields, E. & Buesseler, K.O. (2008). A bottom-up view of the biological pump: Modeling source funnels above ocean sediment traps. *Deep Sea Research Part I: Oceanographic Research Papers*, 55, 108-127.

107.
Smith, D.R., Crosby, K. & Lee, R.W. (2011). Correlation between nuclear plastid DNA abundance and plastid number supports the limited transfer window hypothesis. *Genome biology and evolution*, 3, 365-371.
108.
Stæhr, P.A., Henriksen, P. & Markager, S. (2002). Photoacclimation of four marine phytoplankton species to irradiance and nutrient availability. *Marine Ecology Progress Series*, 238, 47-59.
109.
Stolte, W., Kraay, G.W., Noordeloos, A.A.M. & Riegman, R. (2000). Genetic and physiological variation in pigment composition of Emiliania huxleyi (Prymnesiophyceae) and the potential use of its pigment ratios as a quantitative physiological marker. *Journal of Phycology*, 36, 529-539.
110.
Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G. *et al.* (2015). Structure and function of the global ocean microbiome. *Science*, 348, 1261359.
111.
Suttle, C.A. (2005). Viruses in the sea. *Nature*, 437, 356-361.
112.
Tamura, K., Nei, M. & Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 11030-11035.
113.
Tara Oceans Consortium, C. & Tara Oceans Expedition, P. (2015). Registry of all stations from the Tara Oceans Expedition (2009-2013). PANGAEA.
114.
Thierstein, H.R. & Young, J.R. (2004). *Coccolithophores: from molecular processes to global impact*. Springer.
115.
Trimborn, S. & Langer, G. (2007). Effect of varying calcium concentrations and light intensities on calcification and photosynthesis in Emiliania huxleyi. *Limnology and Oceanography*, 52, 2285-2293.
116.
Turkoglu, M. (2008). Synchronous blooms of the coccolithophore Emiliania huxleyi and three dinoflagellates in the Dardanelles (Turkish Straits System). *Journal of the Marine Biological Association of the UK*, 88, 433-441.
117.
Tyrrell, T. & Merico, A. (2004). Emiliania huxleyi: bloom observations and the conditions that induce them. In: *Coccolithophores*. Springer, pp. 75-97.
118.
Vardi, A., Haramaty, L., Van Mooy, B.A.S., Fredricks, H.F., Kimmance, S.A., Larsen, A. *et al.* (2012). Host–virus dynamics and subcellular controls of cell fate in a natural coccolithophore population. *Proceedings of the National Academy of Sciences*, 109, 19327-19332.

119.
Vardi, A., Van Mooy, B.A., Fredricks, H.F., Popendorf, K.J., Ossolinski, J.E., Haramaty, L. *et al.* (2009). Viral glycosphingolipids induce lytic infection and cell death in marine phytoplankton. *Science*, 326, 861-865.
120.
Vermont, A., Martínez Martínez, J., Waller, J., Gilg, I., Leavitt, A., Floge, S. *et al.* (2016). Virus infection of Emiliania huxleyi deters grazing by the copepod Acartia tonsa. *Journal of Plankton Research*, 38, 1194-1205.
121.
Villar, E., Farrant, G.K., Follows, M., Garczarek, L., Speich, S., Audic, S. *et al.* (2015). Environmental characteristics of Agulhas rings affect interocean plankton transport. *Science*, 348, 1261447.
122.
Weitz, J.S., Stock, C.A., Wilhelm, S.W., Bourouiba, L., Coleman, M.L., Buchan, A. *et al.* (2015). A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. *The ISME journal*, 9, 1352.
123.
Welschmeyer, N.A. (1994). Fluorometric analysis of chlorophyll a in the presence of chlorophyll b and pheopigments. *Limnology and Oceanography*, 39, 1985-1992.
124.
Wilcox, R.M. & Fuhrman, J.A. (1994). Bacterial viruses in coastal seawater: lytic rather than lysogenic production. *Marine Ecology Progress Series*, 35-45.
125.
Wilson, W.H., Tarran, G.A., Schroeder, D., Cox, M., Oke, J. & Malin, G. (2002). Isolation of viruses responsible for the demise of an Emiliania huxleyi bloom in the English Channel. *Journal of the Marine Biological Association of the UK*, 82, 369-377.
126.
Zapata, M., Rodríguez, F. & Garrido, J.L. (2000). Separation of chlorophylls and carotenoids from marine phytoplankton: a new HPLC method using a reversed phase C8 column and pyridine-containing mobile phases. *Marine Ecology Progress Series*, 195, 29-45.
127.
Zhao, J., Hu, C., Lenes, J.M., Weisberg, R.H., Lembke, C., English, D. *et al.* (2013). Three-dimensional structure of a Karenia brevis bloom: Observations from gliders, satellites, and field measurements. *Harmful Algae*, 29, 22-30.
128.
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. *Proceedings of the National Academy of Sciences*, 113, E1907-E1916.
129.
Zuckerkandl, E. & Pauling, L. (1965). Evolutionary divergence and convergence in proteins. *Evolving genes and proteins*, 97, 97-166.