Functional genomics of coccolithophore viruses

Antonio Joaquin Pereira Pagarete

Avertissement

THESE DE DOCTORAT

Spécialité
Océanologie biologique (Ecole doctorale Sciences de l'Environnement de l'Ile de France)

Functional Genomics of Coccolithophore Viruses

Présentée par
António Joaquim Pereira PAGARETE

Pour obtenir le grade de
DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE – PARIS VI

soutenue à Roscoff, le 5 Mars de 2010

devant le jury composé de :

Prof. François Lallier, Examinateur
Dr. Colomban de Vargas, Directeur de thèse
Dr. Michael Allen, Directeur de thèse
Dr. William Wilson, Directeur de thèse
Dr. Filomena Caeiro, Rapporteur
Dr. Stéphan Jacquet, Rapporteur
Dr. Hiroyuki Ogata, Examinateur

CNRS UPMC INSU
Station Biologique Roscoff
UPMC PARIS UNIVERSITAS
cnrs
MARIE CURIE ACTIONS

THESE DE DOCTORAT

Spécialité

Océanologie biologique (Ecole doctorale Sciences de l'Environment de l'Ile de France)

Présentée par

António Joaquim Pereira PAGARETE

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE – PARIS VI

Functional Genomics of Coccolithophore Viruses

soutenue le 5 Mars de 2010

devant le jury composé de :

Prof. François Lallier, Examinateur
Dr. Colomban de Vargas, Directeur de thèse
Dr. Michael Allen, Directeur de thèse
Dr. William Wilson, Directeur de thèse
Dr. Filomena Caeiro, Rapporteur
Dr. Stéphan Jacquet, Rapporteur
Dr. Hiroyuki Ogata, Examinateur

Cover page

The two major actors of this PhD thesis, *Emiliania huxleyi* and coccolithovirus…
(illustration by Miguel Frada and Glynn Gorick)

Dedicado à minha Mãe e ao meu Pai

For my mother and father

Acknowledgements

I have always considered words to be a rather weak mean of showing our acknowledgement for someone's help. However, now that I'm sitting down in my living room, facing the shaded blue waters of Roscoff's shores with this thesis manuscript lying alongside me, I finally see myself confronted with the end of this very intense and challenging period of my life, and realize that the time has come to (try to) thank a few of you that were fundamental for the outcome of this quest…

I would like to start by acknowledging the extraordinary opportunity I've been given to enter the plankton virology realm. In that regard two persons, Dr. Colomban de Vargas and Dr. William Wilson, were crucial for the creation and frame-working of the project here presented. I sincerely thank them for the trust deposited in me and all the advice and discussion provided. The precious chance they gave me will always be treasured, and hopefully develop into many and prosperous years of scientific collaboration and achievements.

The development of a PhD programme is rarely a straightforward and peaceful relaxing time. Its inherent nature of permanent quest to push forward the barriers of knowledge... often leads to frustrations and disillusions which in the end reveal to be the most challenging mission facing a PhD student. This thesis was no exception, but fortunately I did not have to face it on my own. One person was always there with me, someone for whom words will never be enough to explain my gratefulness. During these past three years I have worked in close collaboration with Dr. Michael Allen (even if most of the time with a sea separating us...) in order to produce the scientific progresses here presented. This works are also his. More than "just" an extraordinary scientific mentor, who used of all his engagement to teach me and make me go beyond my limits... Mike became a real friend, always honest and with a word of wisdom to guide me through... For that, and much more that words will never tell, I want to express here my profound gratitude to Mike, the funniest and most dedicated scientist and friend: *thank you Big Cook*...

During the course of this thesis I developed most of my works at the Station Biologique de Roscoff. This is an experience that I will always carry with me wherever my feet take in the future. Immerged in Bretagne, a region of singular and touching cultural traits, we find this scientific institution of excellence. There I had the chance to learn and grow with many most valuable colleagues. They are so numerous, and their influences in my life so varied, that I would certainly risk to be unfair if I tried to point out a few among them. Hence,… à vous tous, mes chers collègues Roscovites, je vous laisse ici mon grand remerciement.

The companionship of my closest friends revealed of particular importance to help me face with a smile the adversities imposed by the thesis. Here is my acknowledgement to my "newly found" friends that helped me through this quest... Miguel, Ian, Robert, Mahdi, Yoshi, Joana, Jan, Christian,

Matthieu, Maxime, Adam, Andres and Blanche, Yoan and Sophie, Christophe, François... and also to my friends of long date… Bruno e Carla Lopes, Adriano Pimentel, Carla Coutinho, Cristina Matos, Diana Reis, Mario Vargas, Joana e Francisco Megre... Thank you for always being there.

This thesis is also the accomplishment of a long period of formation that started in my earliest years. I could never have done it without the engagement of numerous teachers who deposited their trust in me. To all of them I leave here my acknowledgment. In special, I cannot pass this occasion without mentioning a few exceptional of these persons that had an outstanding contribution to help build what I am today: Teresa Oom, Professor Manuel João Costa, Professor Carlos Brito, and Professor Steph Menken. Not only you are brilliant at what you do, as you also have the wisdom to pass it on to the generations to come. I'm only writing these words because I had the rare chance one day of crossing your ways. You are my role models and what I seek to achieve. To you an immense word of gratitude, this thesis is also yours…

Este texto nunca estaria completo sem uma menção muito especial à minha família. Agradeço-vos a todos pela confiança que depositam em mim e pela força dos valores humanísticos que sempre me transmitiram. Seria injusto nao mencionar em particular aqueles que sempre me acompanharam mais de próximo... ao meu sábio Pai, Joaquim Pagarete, à minha Mãe poeta e sonhadora, Carmelinda Pereira, à abnegada e sensível Tia Rosa, à minha maravilhosa metade, a minha mana Margarida Pagarete, e ainda à minha querida "segunda" irmã Marta Jacob... jamais alguma palavra poderá corresponder ao sentimento de profunda gratidão que nutro por vós.

Finalement… un mot pour ma compagne, Constance, la plus résistante, tendre et généreuse parmi toutes les femmes… Quand le monde se transforme dans quelque chose d'hostile, et tout ce qu'on cherche c'est à se cacher dans un coin sombre et oublié… dans ces moments je me suis toujours trouvé accompagné de Constance, qui ne désiste jamais. Tu m'as appris que le mot amour ne se quantifie pas, vu qu'il est magnanime. Quand je pense à toi, Constance, le mot merci prend cette même magnitude. Pas un petit ni un grand… juste un vrai merci, mon amour.

Thank you very much, Merci beaucoup, Sincero obrigado,

António

Abstract

Emiliania huxleyi Virus (EhV) is a giant nucleo-cytoplasmic double stranded DNA virus that belongs to the Phycodnavirus family. It has the capacity to infect *Emiliania huxleyi*, the most abundant coccolithophore in today's oceans. Population dynamics of these eukaryotic microalgae is clearly controlled by the severe lytic action of EhV. After an extended bibliographic review on the current knowledge existing on these viruses, we present a series of bioinformatic and experimental analyses conducted to unveil important functional genomic features of the EhV. Evidence for the transfer of 29 genes between *E. huxleyi*'s and the EhV genomes is presented. In particular, we investigate the origin of seven genes involved in the unique viral sphingolipid biosynthesis pathway (SBP) encoded in EhV genome. This is the first clear case of horizontal gene transfer of multiple functionally-linked enzymes in a eukaryotic host-virus system. We then focus on a field *E. huxleyi*/EhV system from a mesocosm experiment in Norway. The dynamics of expression for two of the most important homologous, host and virus, genes of this pathway, serine palmitoyl transferase and dihydroceramide desaturase is investigated. Three defined transcriptional stages are reported during the bloom, with the coccolithovirus transcripts taking over and controlling the SBP. Finally, host and virus global transcript abundance occurring along the mesocosm experiment was investigated. The majority of the genes that significantly increased in abundance from pre to post viral takeover corresponded to viral sequences for which there is so far no match in the protein databases. Nonetheless, novel transcription features associated with EhV infection were discovered, namely the utilization of genes potentially related to genetic information processing, posttranslational control, intracellular trafficking mechanisms, and control of programmed cell death. As a conclusion, the entire dataset analysed herein is discussed, followed by the potential implications of these findings and future research perspectives in the field of plankton virology.

Key words

Virioplankton, coccolithovirus, coccolithophore, *Emiliania huxleyi*, horizontal gene transfer (HGT).

Résumé

La découverte de la diversité de virus marins est encore dans une phase embryonnaire. Au fur et à mesure que cette prospection se réalise, on prend conscience sur le rôle major joué par les virus en tant que régulateurs de la dynamique populationnel de ses hôtes, et finalement dans l'évolution de la vie cellulaire.

Le Emiliania huxleyi virus (EhV) est un virus géant nucléo-cytoplasmique possédant un génome en double hélice d'ADN. Il appartient à la famille des virus algaux, les Phycodnaviridae. Ce virus a la capacité d'infecter *Emiliania huxleyi*, le coccolithophore le plus abondant dans les océans modernes. Quand les conditions sont adéquates, cette micro-algue eucaryote a la capacité rare et extraordinaire de former des efflorescences ou blooms océaniques très étendus. Ces phases d'efflorescence se terminent du fait de l'action extrêmement lytique d'EhV. Après une revue bibliographique portant sur les connaissances actuelles concernant ces virus, il sera présenté dans ce manuscrit une série d'approches aient comme objective major l'étude de la génomique fonctionnelle des EhVs et à mieux comprendre leur stratégie d'infection.

Nous avons montré tout d'abord sur une base phylogénétique le transfert de 29 gènes entre le génome d'*Emiliania huxleyi* et d'EhV. Parmi ceux, nous nous sommes en particulier concentrés sur sept gènes de EhV impliqués dans une unique voie virale de biosynthèse des sphingolipides (SBP). Notre étude a montré le premier cas patent, dans un système de virus et phytoplancton eucaryotes, de transfert horizontal de multiples gènes d'enzymes liées fonctionnellement. Les patrons de conservation des séquences de ces gènes et des protéines respectives corrobore leur fonctionnalité, à la fois chez *E. huxleyi* et chez EhV. Nous avons étudié les possibilités du sens de ce transfert de gènes. Le sens virus-hôte suggère l'existence d'anciens virus qui contrôleraient des voies métaboliques complexes, ce que leur permettrai d'infecter des cellules eucaryotes primitives. A l'opposé, le sens hôte-virus parait être l'hypothèse la plus parcimonieuse, due à a la fois à la présence de cette vois métabolique parmi la grande généralité d'organismes eucaryotes, et aussi à la position phylogenetic présenter par ces gènes viraux. La façon très discriminé dont ils sont placés dans le génome de EhV, suggère que leur acquisition s'est procédé en série, ce qui a peut être été une stratégie d'un ancêtre de EhV pour rester au plus près de son hôte dans la course à l'évolution.

Nous nous sommes dès lors concentrés sur le système naturel *E. huxleyi* / coccolithovirus, des fjords Norvégiens, pour étudié la dynamique de l'expression de gènes homologues chez le virus et son hôte, pour deux des plus importantes enzymes de cette voie métabolique (SBP), la sérine palmitoyl transférase et la dihydroceramide désaturase. Cette étude transcriptomique a permis de définir trois étapes au cours de la formation et de la disparition des blooms de *E. huxleyi*, pendant lesquelles on registre une activation progressive des transcrits de coccolithovirus, culminant avec leur contrôle de la SBP au cours des étapes 2

et 3. Ces phases sont en accord avec l'hypothèse supposant que les sphingolipides viraux sont impliqués dans la synchronisation et le processus physique de relargage de virions par l'hôte. Cette étude donne une vision unique des interactions des transcriptomes des chaines métaboliques homologues entre le virus et son hôte durant les stades de développement des blooms d'*E. huxleyi* océaniques.

Cherchant à élargir au reste du génome notre connaissance sur les interactions hôtes-virus, nous avons utilisé la technique de biopuce ou puce à ADN pour réaliser la première étude transcriptomique globale entre un hôte (*E. huxleyi*) et un virus (EhV) au sein d'une communauté océanique naturel. Nos résultats montrent que durant les efflorescences d'*E. huxleyi* il y a un épisode synchrone de dominance virale qui est clairement visible à travers les signaux transcriptomiques qui en résultent. Parmi les 279 gènes fortement induits entre la pre et la post dominance virale, la majorité (52%) correspondent à des séquences virales pour lesquelles il n'y a actuellement pas de correspondance dans les bases de données de protéines. En parallèle, les gènes annotés de *E. huxleyi* et de EhV (dont la quantité de transcrits augmentent significativement entre la pre et la post dominance virale) correspondent d'une part à des fonctions attendues impliquées dans le transfert de l'information génétique, mais aussi, et de manière plus surprenante, à certaines gènes probablement impliqués dans le contrôle post-transitionnel, dans les mécanismes de déplacement intracellulaires, ou même dans le contrôle de l'apoptose. Bien que nous soyons loin de pouvoir donner une explication définitive à propos du rôle joué par ces gènes, nos résultats (1) indiquent qu'une stratégie complexe d'infection est utilisée par les coccolithovirus, qui à la fois se confronte avec et demande l'utilisation équilibré d'une complexe machinerie cellulaire de l'hôte ; et (2) donnent des éclaircissements précieux quant aux mécanismes des processus d'infection des coccolithovirus et des fonctions génomiques qui y sont associées.

Du fait que les virus de microalgues disponibles actuellement dans les laboratoires restent assez rares en comparaison de leur diversité gigantesque au sein des océans, une partie significative du travail de thèse a été consacrée aux essais d'isolation de nouveaux virus de coccolithophores. Un bref résumé de ce travail incluant une analyse des contraintes qui y sont trouvées est présenté à la suite des travaux scientifiques précédents.

Pour conclure, la dernière partie de cet ouvrage est consacrée à une discussion générale portant sur les résultats précédemment présentés, suivie d'une analyse de leurs implications potentielles et des futures perspectives de recherche dans le domaine de la virologie planctonique.

Mots clés

Virioplankton, coccolithovirus, coccolithophore, *Emiliania huxleyi*, transfer horizontal de gènes.

Table of Contents

Avant propos...

Marine virology is a recent field of research that has been gaining an increasingly important role in science during these past two decades. Modern optical, cell cytometry, and molecular biological techniques allowed the discovery of an immense unknown world of marine viruses, which is now considered the most abundant and diverse biological realm, representing by far the largest reservoir of genetic variability on Earth. Moreover, this viral existence is far from being innocuous for "cellular" life. Viruses have a significant impact in their hosts dynamics, particularly in the oceans, where viruses are amongst the major causes of microbial death. However, and although the influence of viruses on marine geochemical cycles (via the regulation of host populations), and cellular evolution (via lateral gene transfer), is starting to be recognized, we are still in an embryonic state regarding the comprehension of viral function in the oceans, and the amplitude of their impact in the evolution of life.

The few models of isolated microalgae viruses currently available have assumed major prominence along the path to understand the nature of the host/virus interactions occurring in the oceans, and the impact viruses have on the development of microbial life. Presenting unexpectedly large genome, a particular group of nuclear-cytoplasmic large DNA viruses (NCLDVs) has been recurrently found to infect different microalgae species. These viruses have been classified under a common family, the Phycodnaviridae. Numerous studies have now demonstrated that these viruses have a very important impact in the development of their host population dynamics.

Emiliania huxleyi virus (EhV) is a giant virus that belongs to the family of Phycodnaviruses. It has the capacity to infect *Emiliania huxleyi*, the most abundant coccolithophore in today's oceans. When the conditions are suitable, these microalgae have a rare and extraordinary capacity to form extensive oceanic blooms. By the end of the *E. huxleyi* blooms, very high concentrations of EhVs are found in the waters, and we now know that these viruses are the major cause for the termination of the bloom.

In the present work we will describe a series of approaches, conducted during a three year PhD programme, with the major aim of unveiling unknown functional genomic features of these viruses, and better understand their strategy of infection. One of the central topics here debated will be horizontal gene transfer (HGT) between viruses and their hosts. In that regard, the analysis will mostly focus on the particularly interesting and also unique viral *de novo* sphingolipid biosynthesis pathway present in the EhV genome. Origin and utilization of

these genes will be analysed in the light of an ecological but also evolutionary perspectives. Further on, new wide host and virus transcriptomic approaches, involving microarray techniques, will be presented. The potential novelties regarding gene use during EhV infection and consequent implications on the mechanisms of infection will be discussed.

We hope that the findings presented in this PhD thesis reveal consistency and may be a seed for future research. Above all, we nourish the will that they constitute a useful grain of sand in the unending construction of the pyramid of knowledge...

Chapter 1.
Introduction

1. Virus – life's lubricant

Nowadays the presence of viruses in all ecosystems is fully recognized by the scientific community. Curiously though, while the viral basic function as mortality agents for their host populations is generally accepted (but still not understood in its full extent), few biologists so far realize the immense role viruses assume regulating life's meanders. The majority of known viruses are persistent and not pathogenic (Dimmock, Easton, and Leppard, 2007; Mindell and Villarreal, 2003; Villarreal, 2005), they have their own ancient evolutionary history, possibly dating to the very origins of cellular life (Forterre, 2006a; Forterre, 2006b), and they represent by far the largest reservoir of genetic information in the Earth's biosphere (Suttle, 2005a).

Evolutionary biology has generally failed to consider the contribution that viruses have made to the evolution of life. Some of the reasons are historical, but mainly this is due to the view that viruses do not represent living entities and thus cannot be significant components of or contributors to the tree of life (Moreira and Lopez-Garcia, 2009). This controversy only serves to avoid the real issue. Regardless of their position in the tree of life viruses have an enormous impact on the evolution and ecology of their hosts. Modern evolutionary biology should acknowledge that viruses are ancient biological forms, their numbers are vast, and their role in the fabric of life is fundamental.

The subject of this thesis regards the study of large DNA viruses that infect a group of phytoplanktonic eukaryotic organisms, the coccolithophores. Hence during the course of this Introduction we will mostly focus on the different aspects related to the Phycodnaviruses, from both an ecological and evolutionary perspectives. Cyanophages will also be mentioned for their host's are also unicellular photoautotrophs and inhabit the same niche as eukaryotic phytoplankton. It is worth noting that, despite escaping the ambit of this introduction, to date many other viruses have already been identified which infect such different marine organisms as heterotrophic bacteria (Breitbart, Middelboe, and Rohwer, 2008), cnidaria (Wilson et al., 2005a), molluscs (Renault and Novoa, 2004), crustaceans (Sukhumsirichart et al., 2002), fish

(Essbauer and Ahne, 2001), turtles (Greenblatt et al., 2005), and of course mammals (Van Bressem, Van Waerebeek, and Raga, 1999).

1.1. Virus – a definition

In its most simple form a virus (from the Latin virus meaning *toxin* or *poison*) can be defined as a microscopic non-cellular infectious organism, possessing a structure consisting of a core of DNA or RNA surrounded by a protein coat, that requires a living cell to replicate. Viruses can occur in four basic combinations (single stranded (ss) or double stranded (ds), RNA or DNA viruses). They infect organisms from all kingdoms of life: eukarya, bacteria and archaea (Koonin, Senkevich, and Dolja, 2006). The current virus database contains about 4000 viral species, corresponding to about 30.000 virus strains and subtypes (Dimmock, Easton, and Leppard, 2007). Analysis of the current collection suggests that ssRNA viruses are the most diverse types, followed by dsDNA viruses, dsRNA viruses, and finally ssDNA viruses (Villarreal, 2005). The real overall diversity of viruses is however hard to estimate since so many have not yet been characterized. It is also highly likely to be biased due to sampling limitations, as scientists have historically focused their studies on the viruses of *Escherichia coli*, humans, and domesticated animals and plants. Relatively unstudied habitats are still known to exist. These are anticipated to have populations of virus types not discovered so far. The clearest example of the incipiency of our knowledge regards marine viruses. Recent estimates point to an outstanding number of 10^{30} virus existing in the oceans (Suttle, 2005b), from which the vast majority still wait discovery and classification (Suttle, 2005a) (see below for developments).

The fact that all viruses share a basic overall structure – a protein coat enclosing a nucleoprotein filament – suggests, at the very least, a common mechanism for their appearance. However this remains an enigmatic and controversial subject. Three main hypothesis have been proposed to explain viral origin: (1) they are relics of pre-cellular life forms (Prangishvili, Stedman, and Zillig, 2001); (2) they are derived by reduction from unicellular organisms (via parasitic-driven evolution) (Bandea, 1983; Forterre, 2003); (3) they originated from fragments of genetic material that escaped the control of the cell and became parasitic (Hendrix et al., 2000). The most recent studies suggest that RNA viruses are probably the most ancient, having originated in the nucleoprotein world by escape or reduction from RNA cells (Makeyev and Grimes, 2004; Weiner and Maizels, 1994).

Regarding DNA viruses, it is believed that they are more recent, and (at least some of them) might have originated from RNA viruses (Forterre, 2006a, and references therein; Gorbalenya, Koonin, and Wolf, 1990). Independently of their origin, more and more studies attribute to viruses major roles in fundamental evolutionary questions as diverse as the origin of DNA, the origin of the three modern cellular domains, the origin of the eukaryotic nucleus, or the evolution of mitochondria and chloroplasts (for reviews check Claverie, 2006; Forterre, 2006a; Villarreal, 2005).

1.2. Plankton Viruses – abundance and host mortality

"*The concentration of bacteriophages in natural unpolluted waters is in general believed to be low, and they have therefore been considered ecologically unimportant. Using a new method for quantitative enumeration, we have found up to* 2.5×10^{8} *virus particles per millilitre in natural waters. These concentrations indicate that virus infection may be an important factor in the ecological control of planktonic micro-organisms, and that viruses might mediate genetic exchange among bacteria in natural aquatic environments.*" (*from* Bergh et al., 1989).

When, some twenty years ago, Bergh and his colleagues resumed their newest discovery in the paragraph transcribed above, few would have predicted that high viral abundances in seawater would gain such a profound influence on our understanding of biological oceanographic processes, evolution and geochemical cycling. A recent extraordinary extrapolation of those numbers, which takes into account the average amount of viruses (3×10^{9} per l) and the total volume of the oceans (1.3×10^{21} per l), predicts that the ocean waters can contain around 10^{30} viruses (Suttle, 2005b). This implies that, after bacteria, viruses represent the second largest carbon reservoir in the planet.

Numerous studies have demonstrated that in the oceans the composition and abundance of the viral community is directly related to the dynamics of the microbial plankton (comprising hetero and auto trophic bacteria and protists) (for extensive reviews check Breitbart et al., 2007; Fuhrman, 1999; Suttle, 2005a; Suttle, 2005b; Wommack and Colwell, 2000). In general virioplankton abundance varies with depth (Hara et al., 1996), along trophic gradients (Noble and Fuhrman, 2000), and during the course of phytoplankton blooming events (Brussaard et al., 2004b; Castberg et al., 2002).

The majority of the virioplankton consists of bacteriophages, and their abundance (on average around 10^{10} l^{-1}) follows the same general pattern as bacteria (Maranger, Bird, and Juniper, 1994; Wommack et al., 1992). This claim is supported by observations such as the ability of changes in bacterial abundance to predict changes in viral abundance (Hara et al., 1996), the greater abundance of bacteria over all of other planktonic hosts (Boehme et al., 1993), and the predominance of viruses within the virioplankton with bacteriophage-sized genomes (Wommack et al., 1999). Moreover, phages are estimated to be responsible for about 10-50% of the total bacterial mortality in surface waters (Fuhrman and Noble, 1995; Steward, Smith, and Azam, 1996; Suttle, 1994; Weinbauer et al., 1995).

The data relating to the abundance and impact of eukaryotic phytoplankton viruses (herein referred as algal viruses) is not as extensive as for marine bacteriophages. Nevertheless, evidence is also accumulating that viruses assume a clear role in the control of eukaryotic phytoplankton dynamics. Algal viruses have now been isolated from many geographic locations, including both freshwater and marine environments, and ranging from oligotrophic to eutrophic ecosystems, and even sediments (Brussaard et al., 2004b; Castberg et al., 2002; Cottrell and Suttle, 1991; Jacobsen, Bratbak, and Heldal, 1996; Lawrence, Chan, and Suttle, 2001; Nagasaki and Yamaguchi, 1997; Sandaa et al., 2001; Suttle and Chan, 1995). Most of the algal-virus systems in culture today correspond to large double stranded DNA viruses, which belong to the Phycodnaviridae (for an extensive review check Brussaard, 2004a). Although not as numerous yet as their DNA counterparts, RNA algal viruses have also been isolated and described (Tai et al., 2003; Tomaru et al., 2004).

The Phycodnaviridae are a diverse group of viruses, but their common ancestry is clear at the molecular level. Since the discovery that the DNA pol gene is highly conserved within this group, it became possible to design PCR primers that theoretically cover the majority of the phycodnaviruses (Short and Suttle, 1999). Using these tools several studies have demonstrated the wide distribution of the Phycodnaviridae in all studied aquatic environments (Clasen and Suttle, 2009; Short and Suttle, 2002; Short and Suttle, 2003). More recently, new metagenomic data have corroborated those results (Monier, Claverie, and Ogata, 2008; Monier et al., 2008).

Algal viruses have often been associated with the termination of phytoplankton blooms (Bratbak, Egge, and Heldal, 1993; Brussaard et al., 1996b; Castberg et al., 2001; Jacquet et al., 2002; Nagasaki et al., 1994), however there is growing evidence that, by limiting host population size, these viruses can also play a significant role in preventing the development of bloom events (Larsen et al., 2001; Suttle and Chan, 1994; Tomaru et al., 2007). A

considerable decrease in photosynthetic rate was demonstrated by researchers adding natural virus concentrates to algal populations, suggesting the potential for a reciprocal viral control of global primary productivity (Suttle, 1992; Suttle, Chan, and Cottrell, 1990). Reports of viral lysis rates of phytoplankton in the field are still limited. There is evidence though that viral lysis is responsible for massive cell mortality (rates up to 0.3 d^{-1}), particularly during the decline of algal blooms (Brussaard et al., 1996a; Brussaard et al., 1995), but also in oligotrophic ecosystems (Agusti and Duarte, 2000; Agusti and Sanchez, 2002).

1.3. Virioplankton as catalysts of global nutrient cycles

Viruses are constantly and actively influencing the marine microbial loop (Azam et al., 1994). Lytic infection of the primary producers converts cells into viruses plus cellular debris. This debris is made up of dissolved molecules (monomers, oligomers and polymers) plus colloids and cell fragments (Shibata et al., 1997), most of which is operationally defined as dissolved and particulate organic matter (P-D-OM). Most or all of the lysis products, which contain substantial amounts of major nutrients (C, N, P) and trace nutrients (e.g. Fe), will eventually become available to bacteria (Bratbak et al., 1990; Gobler et al., 1997; Middelboe et al., 2003; Poorvin et al., 2004; Proctor and Fuhrman, 1990). This will provoke an increase in bacterial production and respiration, and reduce protist and animal production, an effect called the "*viral shunt*" (Fig. 1). This sequestration of materials in viruses, bacteria and dissolved matter may lead to better retention of nutrients in the euphotic zone in virus-infected systems, because more material remains in small non-sinking forms (Shibata et al., 1997). On the other hand reduced viral activity may result in more material in larger organisms, which either sink themselves or as detritus, transporting carbon and inorganic nutrients from the euphotic zone to the deep sea (Fuhrman, 1999; Suttle, 2005b).

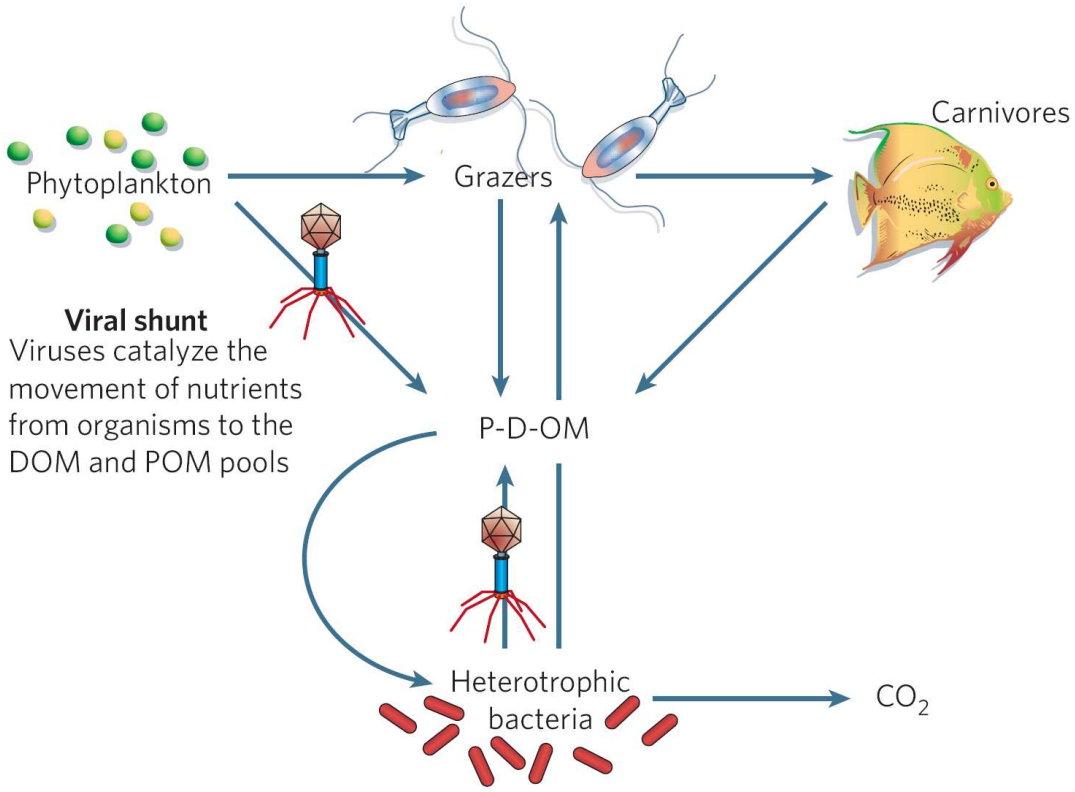

Figure 1. The *"viral shunt"*. Energy, in the form of fixed carbon, is provided to the marine environment via photosynthesis by the primary producers. The fixed carbon, or photosynthate, supports new biomass and respiration of the primary producers. In turn, the primary producers are consumed by grazers (copepods, fish, etc.), who are eaten by bigger predators. A significant amount of photosynthate is also released as particulate and dissolved organic matter (P-D-OM), which supports heterotrophic microbial growth (both bacteria and archaea). The viruses and protists kill similar proportions of the microbes, and the lysed cells then join the P-D-OM pool, which feeds more heterotrophic microbes. The result is more carbon respired, thereby increasing the trophic transfer efficiency of nutrients and energy through the marine foodweb. *Adapted from Suttle (2005*).

1.4. Viral influence in phytoplankton community composition

It is evident, from their effect on algal blooms and cyanobacteria, that viruses are also in a unique position to influence community species composition. Even if viruses were to cause only a small proportion of the mortality of a group of organisms, they could still have a profound effect on the relative proportions of different species or strains in the community (Hennes, Suttle, and Chan, 1995; Waterbury and Valois, 1993). Considering that viral infection is density dependent and that the majority of marine viruses appear to have narrow host specificity, then a particular species or strain becomes more susceptible to infection as its

density increases. This may help explaining Hutchinson's "*paradox of plankton*" on the coexistence of unexpected phytoplankton diversity (Hutchinson, 1961). Competition theory would predict just one or a few competitive winners, however viral activity probably assists because the competitive dominants become particularly susceptible to infection, whereas rare species are relatively protected (Fuhrman and Suttle, 1993). With this "*killing the winner*" strategy (Thingstad, 2000) viruses become a driving force for community composition and succession, both at the interspecific (Brussaard et al., 2005; Castberg et al., 2001; Larsen et al., 2001) and intraspecific (Martinez-Martinez et al., 2006; Muhling et al., 2005; Tarutani, Nagasaki, and Yamaguchi, 2000) levels.

1.5. Viruses and genetic exchange

Virus-host interaction is often promiscuous at the genetic level, a situation that creates a different opportunity for marine viruses to affect genetic exchange in the oceanic realm. This can happen between virus and cellular organisms (direct hosts or not), and among different viruses (especially in situations of co-infection). Recognizing the magnitude and characteristics of horizontal gene transfers (HGT) in the oceans is important from an ecological point of view, and in our case especially important when trying to incorporate viral impact factors in models that try predict phytoplankton dynamics.

HGT can happen during the course of both lysogenic and lytic viral infections. A persistent virus has its genome incorporated in the genome of its host "waiting" for a *stimulus* that will trigger a lytic infection. At that moment new virions are formed and passed onto new host cells. To present date, and to the author's knowledge, plankton viruses with lysogenic strategies have only been documented in marine phages. The occurrence of lysogeny in freshwater filamentous cyanobacteria has been known for more than 35 years (Padan, Shilo, and Oppenhei.Ab, 1972), but only now are we starting to understand the real magnitude of this phenomenon. The generalized occurrence of lysogeny involving marine phages has been extensively documented (Jiang and Paul, 1996; Jiang and Paul, 1998a; Jiang and Paul, 1998b; McDaniel, delaRosa, and Paul, 2006; Weinbauer and Suttle, 1996; Weinbauer and Suttle, 1999). Recent estimates point to roughly half of marine bacterial isolates containing prophages (Paul, 2008).

HGT can also occur between virus and host in the course of lytic infections. Such situations are usually denounced by close phylogenetic identity between host and virus

homologous genes, confirming that either the viruses "stole" the genes from its host, or vice-versa. Evidence for this type of "direct" HGT is becoming more and more abundant with the progressive sequencing of genomes belonging to marine organisms and their respective viruses. As for lysogeny, it was with phages and their prokaryote hosts that the first evidence started to appear. One of the clearest and most interesting examples regards the cyanophages and their photosynthesis genes. Cyanophages infect the abundant cyanobacterial genera, *Synechococcus* and *Prochlorococcus*. Sequencing of these viral genomes showed that they commonly carry genes involved in photosynthesis (Lindell et al., 2004; Mann et al., 2005; Millard et al., 2004). These genes include the highlight-inducible (*hli*) gene, as well as *psbA* and *psbD*, which encode the photosystem II (PSII) core reaction-centre proteins D1 and D2, respectively (Sullivan et al., 2005; Sullivan et al., 2006). The D1 protein is of particular interest because it is the most labile protein in PSII and the most likely to be rate limiting. During the lytic cycle, most of the host's transcription and translation is shut down by phage, which replaces like for like function with its own virally encoded proteins. Because phage must maintain the proton motive force if they are to lyse the host, they need to prolong photosynthesis during the infection cycle. Thus, the cyanophage-encoded D1 proteins are expressed during the infection cycle, countering the virally induced decline in host gene expression (Clokie et al., 2006; Lindell et al., 2005). It is thought that by encoding *psbA* and other genes involved in photosynthesis, phages manipulate their host systems to generate the energy necessary for viral production. Still concerning cyanophages Sullivan and co-workers (2005) have also demonstrated the presence of an aldolase family gene (*talC*), that could facilitate alternative routes of carbon metabolism during infection; and phosphate-inducible genes (*phoH* and *pstS*), that are likely to be important for phage and host responses to phosphate stress, a commonly limiting nutrient in marine systems.

Regarding eukaryotic phytoplankton, examples of direct HGT are also starting to appear. Sequencing of the nucleo-cytoplasmic large DNA virus (NCLDV) *Emiliania huxleyi* Virus (EhV) revealed the presence of some unexpected genes. The most striking example is a unique sphingolipid biosynthesis pathway (SBP) (Wilson et al., 2005b), which was later concluded to be imported from its host *Emiliania huxleyi* (Monier et al., 2009). Sphingolipids are membrane lipids present in all eukaryotes and some prokaryotes. The SBP can ultimately lead to the production of ceramide, a central molecule often involved in signal transduction and control of cell death, namely apoptosis mechanisms (Hannun, 1996; Hannun and Obeid, 1995; Hannun and Obeid, 2002; Pettus, Chalfant, and Hannun, 2002). Other examples of viral control of host apoptosis have already been documented (McLean et al., 2008; Roulston,

Marcellus, and Branton, 1999). When the new viral SBP was discovered hypotheses were immediately drawn on the possibility of EhV using its own virally encoded SBP to control the death of its host. An important part of this thesis works focused precisely on trying to explain the origin and function of this EhV metabolic pathway.

Growing evidence of HGT events involving viruses and cellular organisms other than their direct hosts, so called indirect transfer, is also accumulating. The most notable examples regard bacterial-like genes present in protist and metazoan viruses (Dunigan, Fitzgerald, and Van Etten, 2006; Iyer et al., 2006; Suzan-Monti, La Scola, and Raoult, 2006). Possible explanations for the mechanisms involving this type of genetic transfer are still rudimentary. A recent study from Fillée et al. (2008) has provided some clues. Partial results suggest that indirect HGT seems to be more frequent in viruses whose eukaryotic hosts graze on bacteria. *Chlorella* and Mimivirus (whose hosts feed on bacteria), and EhV and EsV (which infect free leaving microalgae that do not graze on bacteria) show marked variation in bacterial-like genes. While there is a general increase in bacterial gene number with genome size, the strongest dichotomy appears between the *Chlorella* Phycodnaviruses and Mimivirus, which are considerably enriched for bacterial genes, in contrast to Phycodnaviruses EhV86 and EsV-1 which are not. Moreover, very few mobile genetic elements (MGE) of bacterial origin could be found in these latter two algal viruses (Filee, Pouget, and Chandler, 2008).

The development of new metagenomic sequencing techniques has brought the study of HGT to a new level. A considerable portion of the genes present in the viromes analysed so far share very close homology with genes found in both eukaryotic and prokaryotic databases. A metagenomic study of 9 biomes, in which 42 distinct viromes were characterized, found that all the functional diversity present in the microbial metagenomes was also present in the viromes (Dinsdale et al., 2008). A striking example was the totally unexpected discovery of motility related genes present in the viromes. It also became clear that the acquisition of these proteins by the viral community was not random. For example, in the viromes, flagellar biosynthesis protein *FlhA*, the chemotaxis response regulator proteins *CheA* and *CheB* and deacylases were overrepresented when compared to their presence in the microbial genomes. In another study Sharon et al. (2007) reported that up to 60% of the *psbA* genes in surface water are of phage origin. Moreover, phage genes were shown to be undergoing an independent selection for distinct *D1* proteins, and also different viral *psbA* genes are being expressed in the environment. Recently, it was demonstrated that photosystem 1 gene cassettes are also present in cyanophage genomes (Sharon et al., 2009). Regarding eukaryotic hosts Monier et al. (2007) analysed a large dataset of Large Eukaryotic DNA Virus genomes

and reported the presence of many genes putatively associated with the control of host defence systems, such as innate/adaptive immune systems or apoptosis pathways.

All this evidence adds further credence to the idea that viral communities represent reservoirs of genetic diversity, with viruses themselves serving as potential vectors of genetic information among host communities and ecosystems. HGT are rather rare events on an individual scale, but analysed on a global planetary scale this phenomenon assumes a totally different magnitude. Fuhrman (1999) proposed an exercise to infer global oceanic HGT frequency involving marine bacteria. Considering the great abundance of potential cellular hosts (typical bacterial abundances, for example, are around 10^9 l^{-1} in the euphotic zone) and huge volume of the sea (~3.6×10^7 km^3 in the top 100 m), coupled with generation times on the order of a day, implies that an event with a probability has low as 10^{-20} per generation would be occurring about a million times per day.

On the other hand the relevance of HGT between virus and their hosts is also under scrutiny from an evolutionary perspective. As mentioned previously, the origin of viruses and cells has been under intense debate, especially after the discovery of large DNA viruses such as the EhV or the Mimivirus. One hypothesis proposes that these viruses represent ancient cellular forms that gained viral form by progressive loss of genes (Claverie, 2006; Suzan-Monti, La Scola, and Raoult, 2006). Along similar lines of thought hypotheses have been drawn that viruses may have appeared before the separation of the current cellular domains, and consequently influenced the entire evolution of life as we know it (Forterre, 2006a; Forterre, 2006b; Forterre and Gadelle, 2009; Hendrix, 1999). Other authors propose that large DNA viruses are the result of a tendency to indiscriminately acquire genes from all different “horizontal” sources (direct hosts or not) (Koonin, 2005; Moreira and Brochier-Armanet, 2008; Moreira and Lopez-Garcia, 2005). On the contrary according to Monier et al (2007) despite the fact that HGT events play a significant role in the dynamics of gene transfer between the different reservoirs of genetic diversity in the oceans, such events still account for only a minority of the gene composition found in most viruses. This observation suggests that the extremely large sizes of the genomes of some large viruses (for example the Mimivirus) are not due to recent accretion of foreign genes. By extrapolation, the capacity to capture foreign genes is unlikely to be the major factor that determines the tremendous variation in genome size for DNA viruses (Claverie et al., 2006).

Clearly viral HGT, its magnitude and impacts, remain a very hot topic in today’s virology. The recognition of HGT events is highly dependent on the capacity of recognizing homologies between potentially phylogeneticaly close DNA sequences. To that extent we

must not forget that the great majority of the genes present in NCLDV genomes, or in the viral metagenomic databases, remain of unknown function given their dissimilarity with the actual characterized genetic diversity (for example see Raoult et al., 2004; Wilson et al., 2005b). This situation can be the result of a very old origin and/or rapid parallel evolution of viral genes. Hence, even if for a few genes the probability of correctly identifying HGT events is high, the reality is that on the whole it remains difficult to determine the extent of HGT events in these large viral genomes.

2. Phycodnaviridae

The phycodnaviruses are a family of large dsDNA viruses that infect a very diverse group of aquatic eukaryotic organisms. The phycodnaviruses isolated and characterized so far infect different protist lineages comprising green algae, haptophytes, and stramenopiles, as well as multicellular organisms belonging to the brown algae group. They are generally very large viruses that contain also some of the largest viral genomes ever found. Among the phycodnaviruses we find Emiliania huxleyi Virus, which has been the central object of study throughout this thesis.

2.1. Taxonomy and distribution

The phycodnavirus group comprise a genetically diverse (Dunigan, Fitzgerald, and Van Etten, 2006; Iyer et al., 2006), yet morphologically similar, family of large icosahedral viruses that infect marine or freshwater eukaryotic organisms. Their big dsDNA genomes can range from 180 kb to 560 kb (Van Etten et al., 2002). To present date members of the Phycodnaviridae are grouped into six genera (named after the hosts groups they infect): *Chlorovirus*, *Coccolithovirus*, *Prasinovirus*, *Prymnesiovirus*, *Phaeovirus*, and *Raphidovirus* (Table 1). We should also mention here the mimivirus group. These are huge dsDNA viruses (genome reaching up to 1.2 Mb) that, so far, have been found to infect amoeba (Raoult et al., 2004). Even if their potential hosts are not algae, cumulating evidence indicates that they occupy a phylogenetic position within the phycodnaviridae (Larsen et al., 2008; Monier et al., 2008; Wu et al., 2009). From herein in this text, all mentions to phycodnaviruses should be regarded as that wider group that includes also the mimivirus.

Phycodnaviruses are widely distributed in nature. Viral isolates have been obtained from eutrophic and oligotrophic water masses, and even sediments (Castberg et al., 2002; Cottrell and Suttle, 1991; Jacobsen, Bratbak, and Heldal, 1996; Lawrence, Chan, and Suttle, 2001; Nagasaki and Yamaguchi, 1997; Sandaa et al., 2001; Suttle and Chan, 1995). These probably represent only a tiny fraction of the enormous diversity of the existing phycodnaviruses. Other culture independent techniques have allowed a glimpse into the magnitude of their variability and dispersion. Given their large size phycodnaviruses can be identified and quantified using flow cytometric techniques (Brussaard, 2004b). Marie and colleagues (Marie et al., 1999) recurred to such techniques to show that a clearly distinct group of phycodnaviruses was always present in sea water samples from mesotrophic through oligotrophic environments. Moreover, genetic fingerprints based on polymerase chain reaction (PCR), denaturing gradient gel electrophoresis (DGGE), and metagenomic sequencing reveal that phycodnaviruses are very diverse and a regular component of all aquatic environments (Chen and Suttle, 1995; Chen, Suttle, and Short, 1996; Larsen et al., 2008; Monier, Claverie, and Ogata, 2008; Short and Suttle, 2002; Short and Suttle, 2003).

Table 1. Taxonomy and characteristics of some phycodnaviruses[a]

Genus	**Type species**	**Host Species / group**	**Source**	**Particle diameter (nm)**	**Genome size (kbp) and conformation**	**Latent period (h)**	**Burst size**
Chlorovirus	*Paramecium bursaria* chlorella virus (PBCV-1)	*Chlorella* NC64A (Green algae)	FW	190	313-370 Closed linear dsDNA, hairpin termini	6-8	200-350
Prymnesiovirus	*Chrysochomulina brevifilum* virus PW1 (CbV-PW1) *Phaeocystis pouchetii virus*	*Chrysochomulina sp.* (Haptophytes)	MW	120-160	485-510	12-19	400-4100
Prasinovirus	*Micromonas pusilla* virus SP1 (MpV-SP1) *Ostreococcus tauri* virus 5 (OtV-5)	*Micromonas sp.* *Pyraminomonas sp.* *Ostreococcus tauri* (Green algae)	MW	115-200 113-131	200-560 186-192	7-14 8-12	200-1000 25
Phaeovirus	*Ectocarpus siliculosus* virus 1 (EsV-1)	*Ectocarpus siliculosus* (Brown algae)	MW	130-200	160-340 Open linear, single stranded regions	ND	>1.10^6
Coccolithovirus	*Emiliania huxleyi* virus 86 (EhV-86)	*Emiliania huxleyi* (Haptophytes)	MW	160-200	407-415 Circular	4-6	400-1000
Raphidovirus	*Heterosigma akashiwo* virus 01 (HaV-01)	*Heterosigma akashiwo* (Stramenopiles)	MW	202	294	30-33	770

FW, fresh water; MW, marine/coastal water; ND, not determined.

[a] Data from Dunigan *et al.* (2006), Derelle *et al.* (2008), Weynberg *et al.* (2009).

2.2. Phylogeny and evolution

To date several phycodnavirus genomes have been completely sequenced. They correspond to representatives of the genera *chlorovirus* (Zhang et al., 1994), *coccolithovirus* (Wilson et al., 2005b), *phaeovirus* (Delaroque et al., 2001; Schroeder et al., 2009), *prasinovirus* (Derelle et al., 2008; Weynberg et al., 2009), and *mimivirus* (Raoult et al., 2004). Evolutionary analysis of their genomes places them within a major, monophyletic assemblage of large eukaryotic dsDNA viruses termed the Nucleo-Cytoplasmic Large DNA Viruses (NCLDVs) (Fig. 2) (Allen et al., 2006c; Derelle et al., 2008; Iyer, Aravind, and Koonin, 2001; Iyer et al., 2006; Raoult et al., 2004). Five families are currently attributed to the NCLDVs clade, including Poxviridae, Iridoviridae, Asfarviridae, Phycodnaviridae. The inclusion of the phycodnaviruses within the NCLDVs is significant for, as the name suggests, it implies a likely propagation mechanism where replication would initiate in the nucleus, and be completed in the cytoplasm (Iyer et al., 2006; Raoult et al., 2004; Villarreal and DeFilippis, 2000). A total of nine gene products are present in all NCLDVs identified to date, and 33 more gene products are present in at least two of these five viral families (Iyer, Aravind, and Koonin, 2001; Raoult et al., 2004). Phylogeny of the NCLDVs constructed by cladistic analysis indicates that the major families may have diverged prior to the divergence of the major eukaryotic lineages 1-2 billion years ago (Iyer et al., 2006; Raoult et al., 2004). Regarding the Phycodnaviridae, the finding that only 14 genes (from a pool of approximately 1000 genes) are shared between three genomes from different genera (*chlorovirus*, *coccolithovirus* and *phaeovirus*) supports the idea that these groups also diverged a long time ago (Allen et al., 2006c).

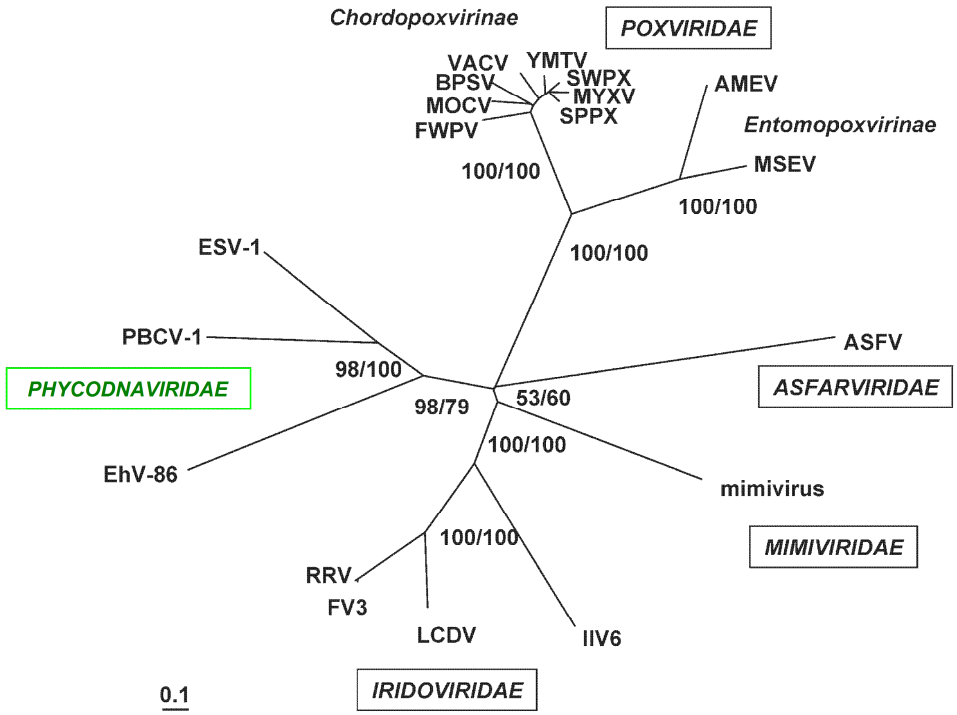

Figure 2. Phycodnavirus occupy a phylogenetic clade within the NCLDVs. Tree based on a distance matrix algorithm between the concatenated conserved domains from A18-like helicase, D6R-like helicase, A32-like ATPase, D5-like ATPase, DNA polymerase, thiol-oxidoreductase, and the two largest RNA polymerase subunits from members of the NCLDV group (Neighbor, in PHYLIP version 3.6b). Numbers at nodes indicate bootstrap values retrieved from 100 replicates for both the neighbour-joining and parsimony analyses. The bar depicts 1 base substitution per 10 amino acids. Viruses included are *(cont.)* African swine fever virus (AFSV), *Amsacta moorei* entomopoxvirus (AMEV), *Melanoplus sanguinipes* entomopoxvirus (MSEV), bovine papular stomatitis virus (BPSV), fowlpox virus (FWPV), sheeppox virus (SPPX), swinepox virus (SWPV), vaccinia virus (VACV), *Molluscum contagiosum* virus (MOCV), myxoma virus (MYXV), Yaba monkey tumor virus (YMTV), *Paramecium bursaria chlorella* virus 1 (PBCV-1), Ectocarpus siliculosus virus 1 (ESV-1), *Emiliania huxleyi* virus 86 (EhV-86), frog virus 3 (FV3), invertebrate iridescent virus 6 (IIV6), *Regina ranavirus* (RRV), lymphocystis disease virus 1 (LCDV) and mimivirus. *Adapted from Allen et al. (2006c).*

The phylogenetic relations within the Phycodnaviridae are far from being conclusive (Allen et al., 2006c; 2008; Larsen et al., 2008). The genera *chlorovirus*, *prymnesiovirus* and *phaeovirus* seem to correlate in accordance to their host's phylogeny. Yet, the genera *coccolithovirus* and *phaeovirus* seem to have a more ancient divergence (Fig. 3). The formation of the genus *coccolithovirus* has brought some confusion to the phycodnavirus taxonomy. *Coccolithovirus* infect *Emiliania huxleyi* (an alga species in the class Prymnesiophyceae), and hence it was expected that they occupy a phylogenetic position within the *prymnesiovirus* group. However, phylogenetic analysis of the Major capsid protein

gene from these viruses has indicated they belong to a distinct genus (Schroeder et al., 2002, see Fig. 3).

It should also be noted that, as mentioned above (section 1.2. Plankton Viruses – abundance and host mortality) not all viruses that infect eukaryotic phytoplankton belong to the phycodnavirus family. Indeed, other types of viruses that infect algae are being discovered and characterized (e.g., ssRNA, dsRNA, and ssDNA containing viruses) (Brussaard et al., 2004a; Tai et al., 2003; Tomaru et al., 2004).

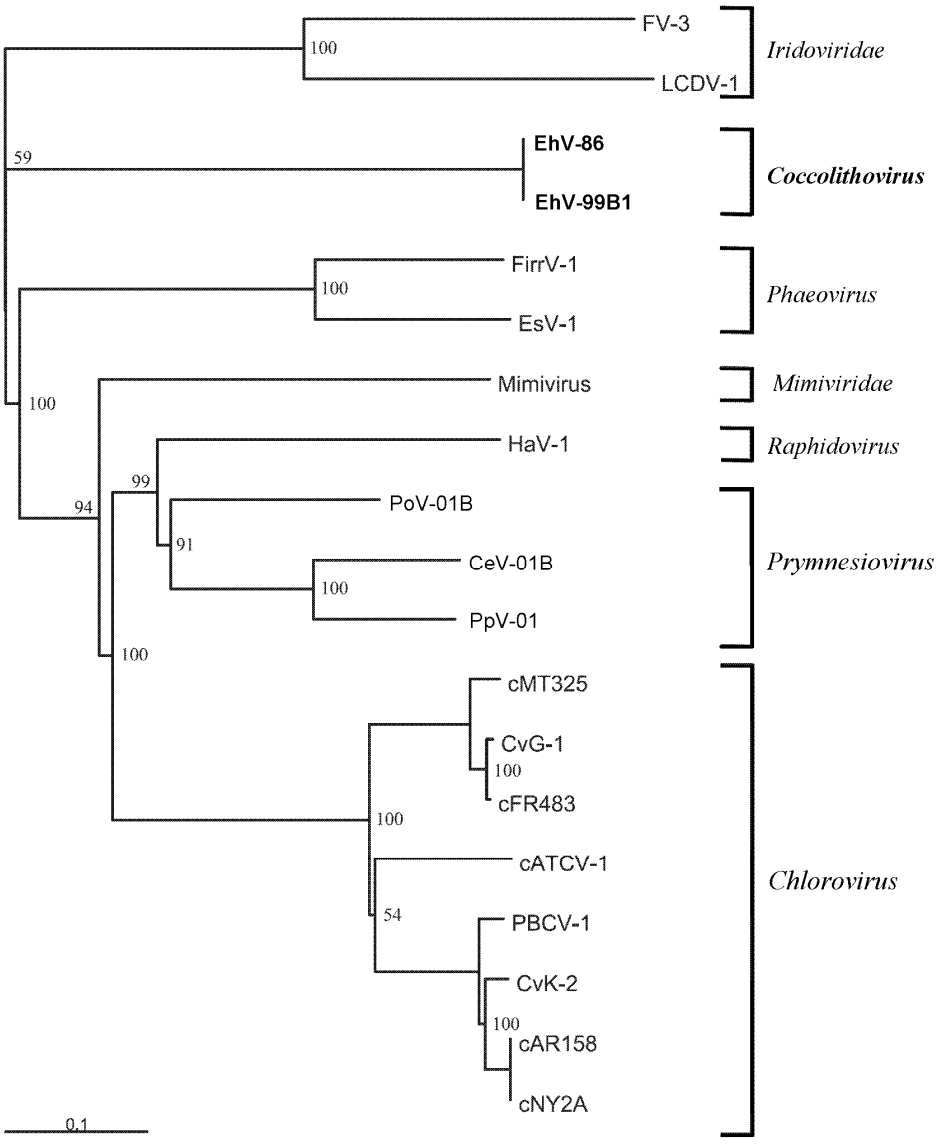

Figure 3. Phylogenetic inference using the complete protein sequence of the major capsid protein from eighteen members of the Phycodnaviridae family. Numbers at nodes indicate bootstrap values retrieved from neighbour-joining analysis using 1000 replicates. The tree was rooted using the two sequences of Frog Virus 3 and Lymphocystis Disease Virus 1 of the Iridoviridae family. The scale bar indicates number of amino acid substitutions per residue. *Adapted from Larsen et al. (2008).*

2.3. Propagation strategies

The study of phycodnaviruses is still in its infancy, thus hitherto we do not know much about their propagation strategies. These viruses possess very large genomes that encode for a wide panoply of genes. This potentially confers phycodnaviruses a relatively high degree of independence from their host's cellular machinery. The infection strategies used by these viruses appear to be quite diverse. We give here a brief description of what is known about the replication strategies of two phycodnaviruses, the Chlorovirus and the Phaeovirus. Data on the replication strategy of the coccolithovirus will be presented in detail in **Chapter 2**.

Phaeoviruses are lysogenic and only infect free-swimming, wall-less gametes of their filamentous brown algae hosts *Ectocarpus siliculosus* and *E. fasciculatus* (Muller, Kapp, and Knippers, 1998; Muller et al., 1996). Following virion attachment the genetic material immediately moves to the nucleus to be incorporated in the host's genome (Maier, Müller, and Katsaros, 2002). The viral genome is then replicated and transmitted from cell to cell during mitosis in the host vegetative cells (Delaroque et al., 1999). Once the host becomes sexually mature and produces reproductive cells, the transcription of the viral genome is integrated with subsequent production of viral capsids and release of newly formed infectious virions from the cells.

Chloroviruses infect freshwater unicellular green algae from the genus *Chlorella* (Wilson, Van Etten, and Allen, 2009). Virion attachment to the cell provokes the degradation of the cell wall, followed by release of the viral genetic material (DNA and virion-associated proteins) into the host, and a rapid depolarization of the cell membrane (Frohns et al., 2006). The viral genetic material is then within 5-10 mins. transferred into the cell nucleus where early transcription starts. The early mRNAs are then transported to the cytoplasm for translation, and the early proteins presumably return to the nucleus to initiate DNA replication, which begins 60–90 min post infection, followed by late gene transcription (Schuster et al., 1986). Late mRNAs are transported to the cytoplasm for translation, and many of these late proteins are targeted to the virus assembly centers, in the cytoplasm, where virus capsids are formed (Meints, Lee, and Van Etten, 1986). Six to eight hours after infection the algal cell membrane and wall lyse, allowing the release of around 300 viral particles, from which only 30% (approximately) are infectious (Van Etten et al., 1983).

3. Coccolithophores

Coccolithophores are unicellular chlorophyll a + c containing eukaryotes that belong to the Phylum Haptophyta (Fig. 4), and more particularly to the Class Prymnesiophyceae. They occur as solitary free-living motile cells possessing two smooth flagella. Although the Haptophyta are distinguished by the presence of a unique organelle called a *haptonema* (from the Greek *hapsis* - touch, this organ is superficially similar to a flagellum but differs in the arrangement of its microtubules and in its use for prey capture or attachment), in many coccolithophores it is reduced to a vestigial structure.

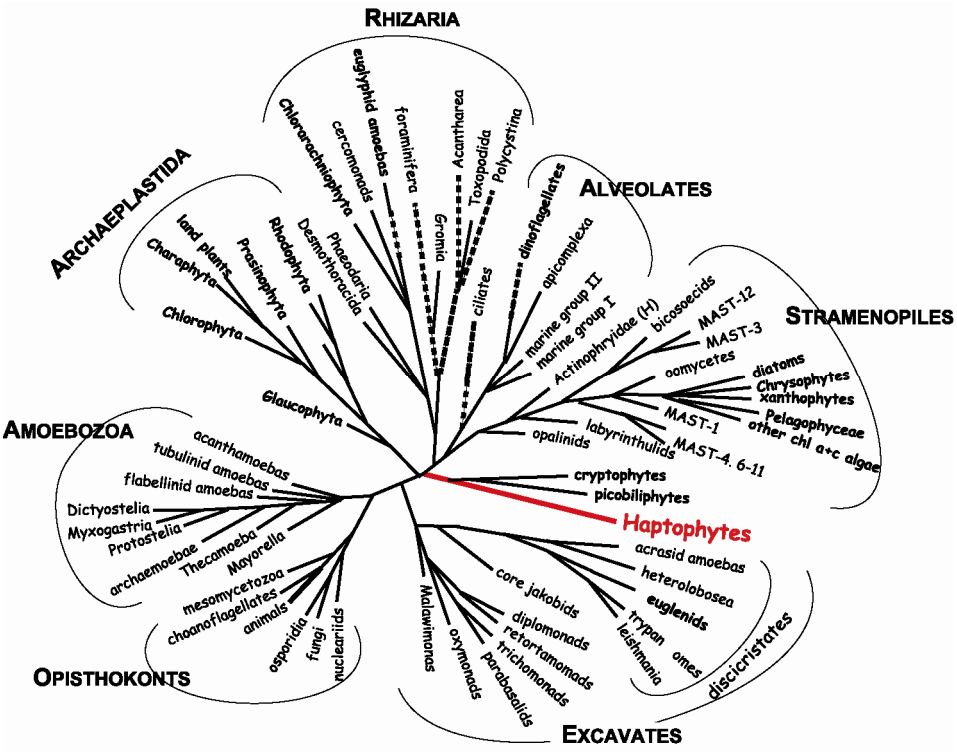

Figure 4. The haptophytes in the eukaryote tree of life. The tree shown is a consensus phylogeny of eukaryotes based on a combination of molecular phylogenetic and ultrastructural data. *Modified from Baldauf (2003).*

The name *coccolithophore* relates to the extraordinary capacity these organisms have to internally produce calcareous scales, the coccoliths, that they extrude and deposit around the cell. The term coccolith (literally meaning *round stones*) was coined by Huxley in 1858 (1858). Wallich (1877) described for the first time the association between coccoliths and the cells producing them: the coccolithophores. Coccoliths are believed to have their origin in pre-existing organic scales which are very common among Prymnesiophytes (de Vargas et al., 2007; Leadbeater, 1994). These organisms use a particularly large and highly polarized

dictyosome to produce those organic scales (Leadbeater, 1994; Pienaar, 1994). In coccolithophores this organelle is not only involved in the synthesis of organic scales as it has also acquired a particular function: coccolithogenesis (Billard and Inouye, 2004). Coccolithogenesis may vary slightly according to the morphological type of coccolith produced, but the basic principle is the same in all cells: the coccolith is produced intracellularly (in Golgi-derived compartments), and it is only extruded to the cell surface (generally close to the flagellar pole) when fully calcified.

3.1. **Evolution and life cycle**

According to Bown (1987) the first reliable coccolith fossil traces back to the Triassic (~220 My) (Fig. 5). Coccolithophore evolution seems to have started in coastal environments, followed by a clear expansion and colonization of the open oceans along the Jurassic (de Vargas et al., 2007). The transition of coccolithophores from coastal water environments to the deep ocean was a remarkable step in their evolution, and a crucial event that would forever transform the Earth's biogeochemical system (see section 3.2).

Molecular data (Saez et al., 2004) as well as studies on coccolith biomineralization homology (Young et al., 1999; Young et al., 1992; Young, Geisen, and Probert, 2005) support the idea that coccolithophores form a monophyletic clade within the Class Prymnesiophyceae. This means that a common prymnesiophyte ancestor has probably developed the capacity to control the intracellular precipitation of calcite onto pre-existing organic plate scales, and assembly of mature carbonate scales at the cell surface (de Vargas et al., 2007; Leadbeater, 1994). Some living descendents inside the coccolithophore clade (or sometimes stages of their complex life cycle) have posteriorly lost the capacity to produce calcareous scales (Billard and Inouye, 2004; de Vargas and Probert, 2004). Hence, the "presence or absence of coccoliths" cannot be used as unique feature to classify potential coccolithophore cells. More recently, and based on a wide range of haptophyte SSU and LSU rDNA sequence data (de Vargas et al., 2007; Saez et al., 2004), de Vargas and colleagues (2007) have proposed the creation of a new subclass Calcihaptophycidae. This group will comprise all potentially calcifying haptophytes, which by definition includes all coccolithophores (Fig. 5).

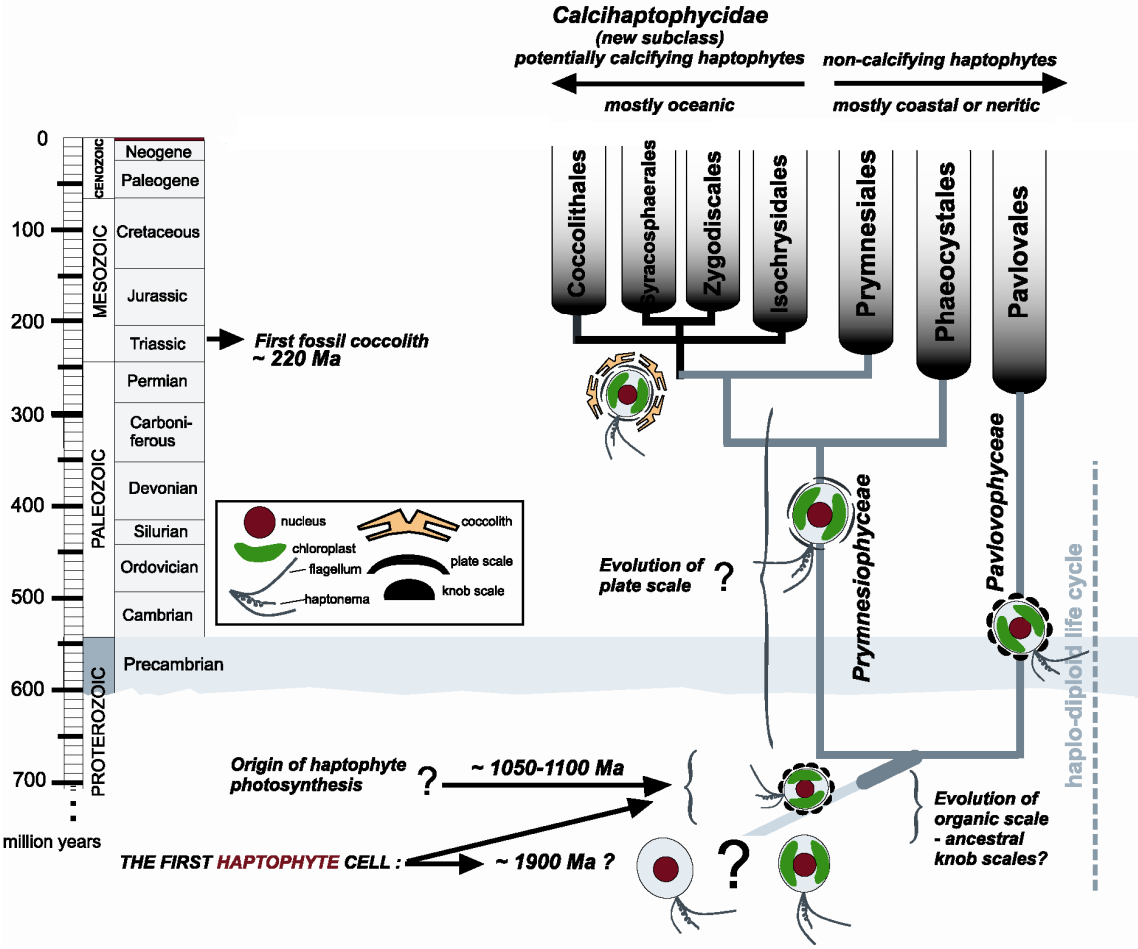

Figure 5. Evolutionary history of the coccolithophores within the haptophyte phylum. Major innovations are shown along a geological time scale on the left side of the figure, and a synthesis of recent molecular phylogenetic data using representative species of the seven extant haptophyte orders (de Vargas et al. 2007 and Saez et al. 2004) is depicted on the right side. Biological, phylogenetic, and paleontological data tend to support a scenario according to which the haptophytes have broadly evolved from coastal or neritic heterotrophs/mixotrophs to oceanic autotrophs since their origination in the Proterozoic. *Adapted from de Vargas et al. (2007).*

Fossil studies estimate that throughout their evolution coccolithophores diversified into >4000 morphological species (de Vargas et al., 2007) (see some examples in Fig. 6). This number is however probably significantly underestimated. A recent detailed study of sinking planktonic assemblages (Andruleit, Rogalla, and Stager, 2004) has shown that most of the morphological diversity is entirely dissolved in the upper water column. Among the ~280 types of coccosphere (morphospecies) known from the modern plankton, only 57 are present in Holocene sediments (Young, Geisen, and Probert, 2005). Thus, it can be estimated that up to 70% of the past coccolithophore diversity has been erased from the fossil record.

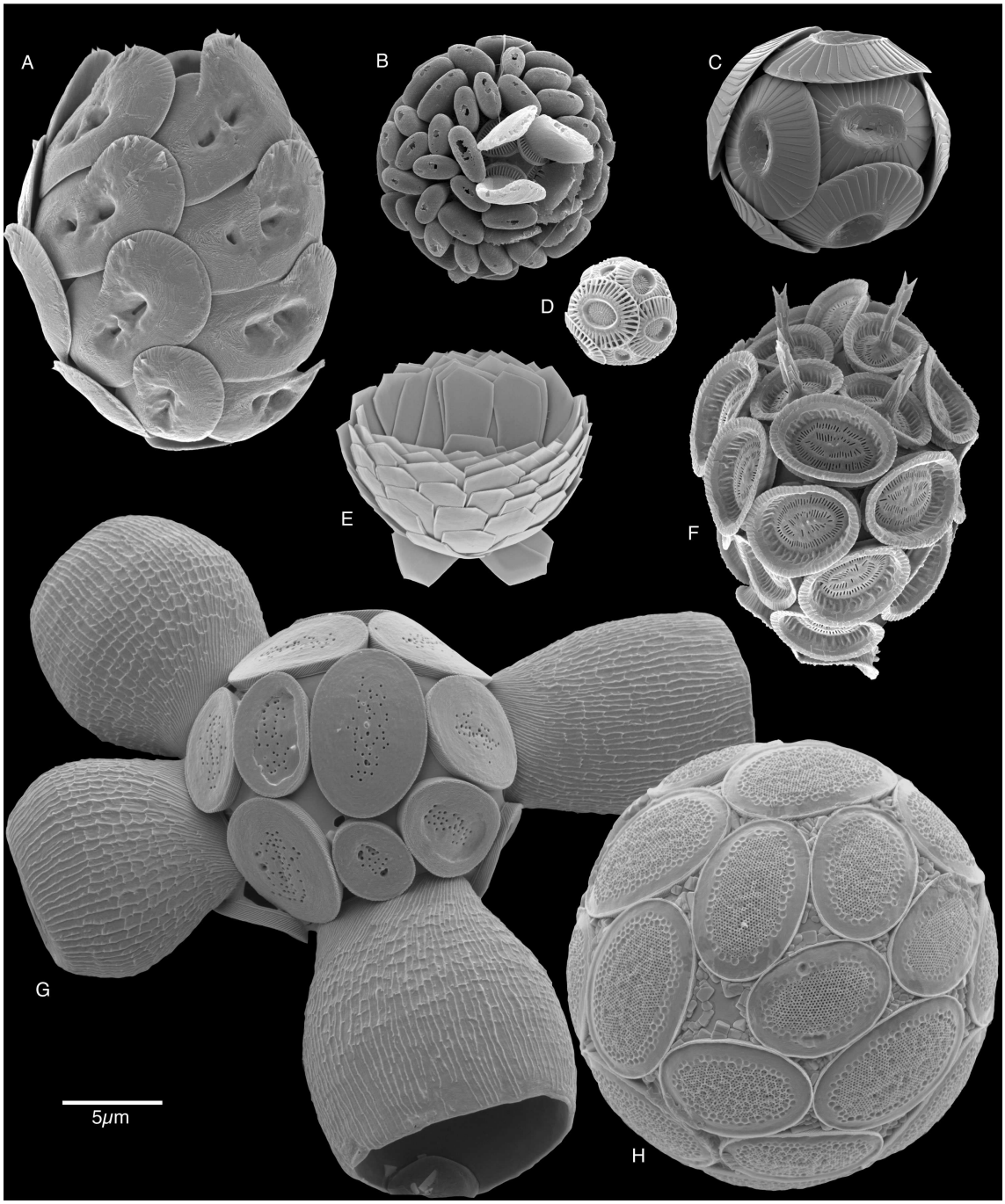

Figure 6. Morphostructural diversity in extant coccospheres and their coccoliths. This plate illustrates the astounding calcareous morphostructures observed in some modern coccolithophores. (A) *Helicosphaera carteri*, (B) *Algirosphaera robusta*, (C) *Coccolithus pelagicus*, (D) *Emiliania huxleyi*, (E) *Florisphaera profunda*, (F) *Syracosphaera pulchra*, (G) *Scyphosphaera apsteinii*, and (H) *Pontosphaera japonica*. *Adapted from de Vargas et al. (2007).*

Coccolithophores usually reproduce asexually by binary fission and, following mitotic division, the coccoliths are redistributed on the daughter cells (Billard and Inouye, 2004). Life cycle of coccolithophores is generally haplodiploid (Green, Course, and Tarran, 1996; Houdan et al., 2003; Larsen and Edvardsen, 1998; Vaulot et al., 1994), as occurs in the

majority of the prymnesiophytes (considered a synapomorphic trait among this group). In a haplodiploid life cycle both stages, haploid and diploid, are capable of independent asexual reproduction. This capacity to grow vegetatively under both haploid and diploid genomes, and expressing radically different phenotypes, is frequent among unicellular eukaryotes. The haplo-diploid strategy clearly involves fitness costs for each life stage, and hence it must be balanced by advantages of evolutionary and/or ecological significance. De Vargas and colleagues (2007) suggest that it is likely a strategy to rapidly escape negative selection pressures exerted on one stage, such as grazing, parasite or virus infection, or abrupt environmental changes. However, the factors triggering shifts from diploid to haploid stages (and vice-versa) in coccolithophores have never been clearly identified. In that regard a recent study (Frada et al., 2008) showed that the *Emiliania huxleyi* Virus (EhV) infects exclusively the diploid life stage of the species *Emiliania huxleyi*. According to the authors this specific negative selective pressure imposed on the diploid life stage could be a major evolutionary force behind the maintenance of a haplodiploid life cycle, and eventually be linked to the origins of sex in evolutionary biology (further developments in **Chapter 2**).

3.2. **Biogeochemical and ecological roles**

As photosynthesizers and calcifying organisms, coccolithophores assume a rather complex and extremely important role on the regulation of the Earth's system, mainly in what regards carbon flux between atmosphere/ocean/lithosphere (Fig. 7). Coccolithophores are unicellular photosynthetic organisms, and hence integral part of the oceanic phytoplankton. Phytoplankton uses light energy to sequester dissolved carbon dioxide (CO_2) and produce particulate organic carbon (POC), and oxygen (O_2). This so-called photosynthetic process [$CO_2+H_2O \rightarrow CH_2O$ (POC) + O_2] participates to maintain the atmospheric CO_2 concentration 150 to 220 ppmv below what it would be if phytoplankton did not exist (Falkowski et al., 2000). It is estimated that 25% of the carbon fixed by phytoplankton is exported to the deep oceans, in a total of 11 to 16 Gt of carbon per year (Falkowski, Barber, and Smetacek, 1998; Laws et al., 2000). The complex system of oceanic biological and physico-chemical processes that transport carbon from the epipelagic zone to the abyssal ocean floor is designated the "biological pump" (Volk and Hoffert, 1985).

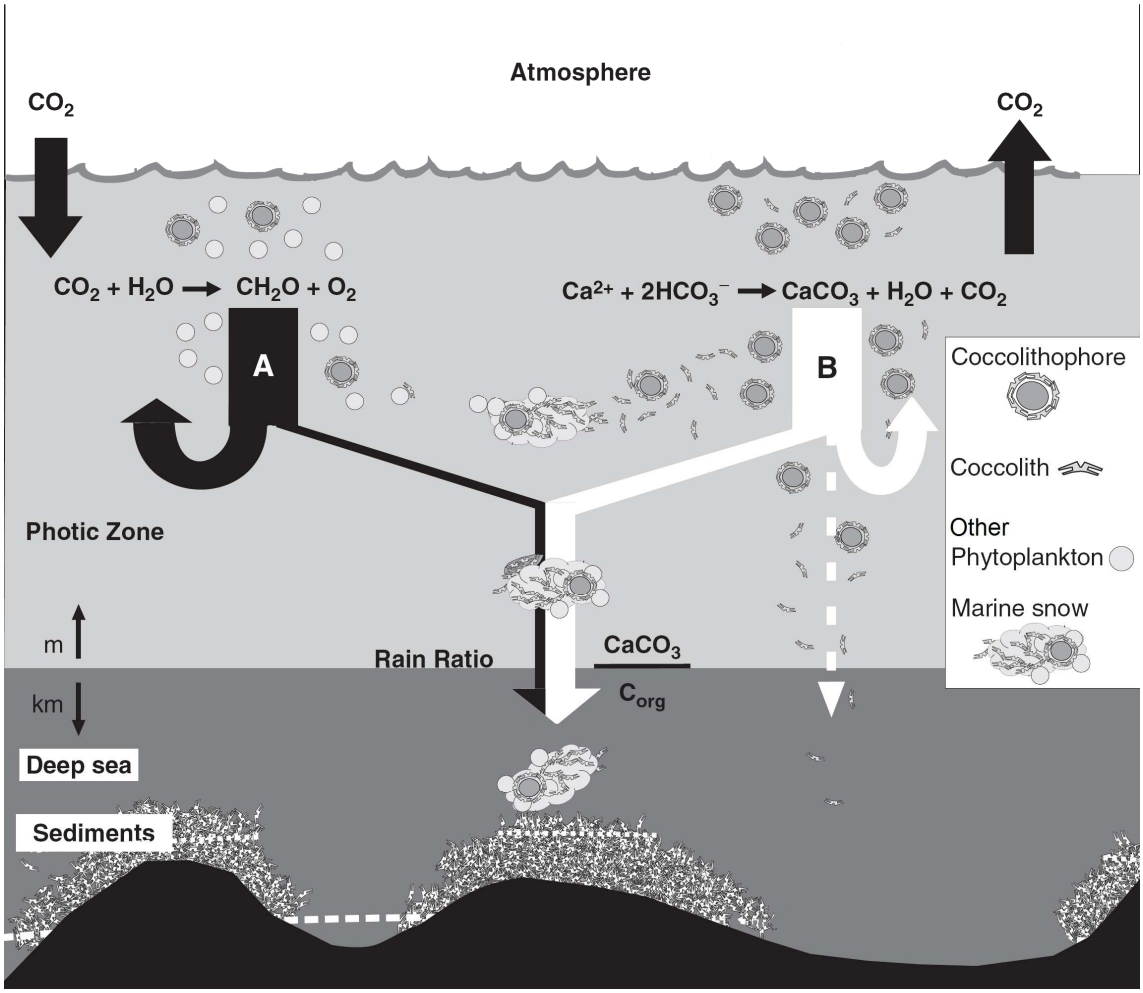

Figure 7. Role of coccolithophores in biogeochemical cycles. Through the production of their coccoliths, coccolithophores actively participate in gas exchange between seawater and the atmosphere and to the export of organic matter (C_{org}) and carbonate ($CaCO_3$) to deep oceanic layers and deep-sea sediments. Via the ballasting effect of their coccoliths on marine snow, coccolithophores are the main driver of the organic carbon pump (A), which removes CO_2 from the atmosphere. They are also the main actors of the carbonate counter-pump (B), which, through the calcification reaction, is a short-term source of atmospheric CO_2. Thus, organic and carbonate pumps are tightly coupled through coccolithophore biomineralization. Ultimately, certain types of coccoliths particularly resistant to dissolution are deposited at the seafloor, where they have built a remarkable fossil archives for the last 220 My. *Adapted from de Vargas et al. (2007).*

Coccolithophores also have also a second important particularity, they produce calcium carbonate structures. This function provides them with a more elaborate role in the carbon cycle. As they secrete their calcareous exoskeleton, they are ultimately sequestering carbon from the atmosphere and stocking it into dense inorganic structures (the coccoliths). These structures will then act as ballasts of the oceanic dead-matter aggregates (or marine snow). Being denser than sea water, these agglomerates sink to the deep ocean. It is estimated that this process is responsible for half of the total $CaCO_3$ deposition in modern oceans (Milliman,

1993). As a result, around 35 % of the ocean floor is covered by a calcareous layer (which may reach several kilometres deep), which acts as one of the most important stabilizing components of the Earth's carbonate compensation system (Broecker and Peng, 2005). A major role played by coccolithophores is hence to provoke a substantial flow of carbon from the atmosphere into the Earth's mantle (by subduction), where it will be prisoner of the lithosphere for millions of years.

Coccolithophore interference in the global carbon cycle started to have a serious impact as they started proliferating in the Cretaceous oceans (Fig. 2). With an outstanding coccolithophore expansion taking place (which led to the occupation of new deep ocean niches), these organisms provoked a clear change in the oceanic carbon deposition sites. A process that before occurred mostly in the coastal shallow water regions moved to the deep ocean for the first time in the Earth's history (Hay, 2004), leading to a revolution in ocean carbon chemistry regulation (Ridgwell and Zeebe, 2005).

4. *Emiliania huxleyi*

Emiliania huxleyi (Lohmann) Hay et Mohler is a very young coccolithophore morphospecies. Its first appearances in the fossil record date from 268,000 years ago (Thierstein, Geitzenauer, and Molfino, 1977). It diverged within the *Gephyrocapsa* lineage, and became the most abundant and ubiquitous coccolithophore in today's oceans (Brown and Yoder, 1994). It grows from warm and nutrient depleted shallow surface mixed layers, to freshly stratified waters from temperate and sub-arctic latitudes following termination of spring diatom blooms (Tyrrell and Merico, 2004).

E. huxleyi's haplo-diplontic life cycle (Fig. 8) comprises two forms: the diploid (2N), nonmotile, coccolith-bearing phase, and the haploid (N) flagellated phase that possesses nonmineralized organic scales overlying the cell membrane (Paasche, 2001). Recent data shows that both life forms can usually be found co-habiting in the sea; however the diploid stage always represents the great majority of the *E. huxleyi* cells found (Frada, 2009).

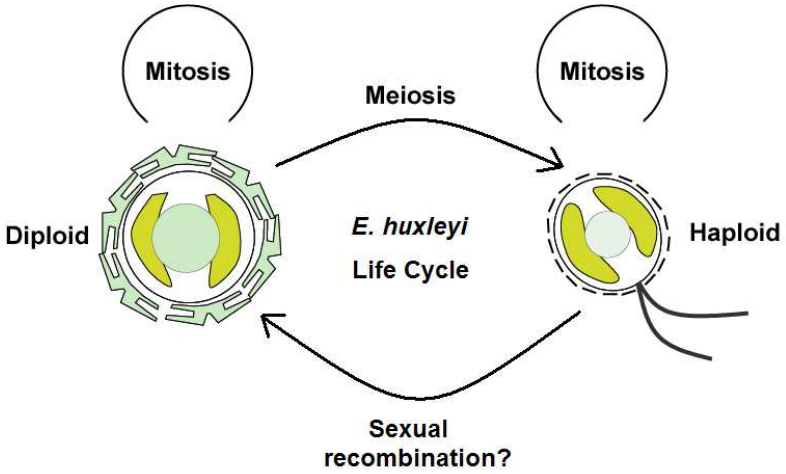

Figure 8. Schematic representation of the life cycle of *Emiliania huxleyi*. The recombination of haploid cells which would allow the cycle to continue onto the 2N stage has never been documented. *Adapted from Frada (2009).*

E. huxleyi presents an unusual capacity to form vast blooms that can reach up to 10^8 cells litre^{-1}. The capacity to form vast blooms appears to be exclusive to the diploid cells, clearly R-selected organisms, that present a very high maximum growth rate (up to 2,8 doublings per night (Brand and Guillard, 1981). *E. huxleyi* blooms can cover such large areas that they can easily be visualized from space (Fig. 9). One of the largest registered blooms happened in 1991 south of Iceland (Atlantic North), with an extension of 250,000 km^2 (Holligan et al., 1993). When these vast coastal and mid-oceanic *E. huxleyi* blooms disappear there is a substantial flux of calcite to the seabed (Ziveri et al., 2000), a situation that has clear impact on the biological pump (see explanation above). Moreover *E. huxleyi* blooms can also produce other significant regional environmental impacts, such as an increase in water albedo (reflectance) (Tyrrell, Holligan, and Mobley, 1999), and the release of cloud-forming dimethyl sulfide (DMS) to the atmosphere (Malin and Steinke, 2004).

Several studies have now demonstrated that viruses are the major cause of bloom termination (Bratbak, Egge, and Heldal, 1993; Bratbak et al., 1996; Castberg et al., 2002; Jacquet et al., 2002; Schroeder et al., 2003; Wilson et al., 2002). In a recent article Frada and co-workers have shown that those viruses (Coccolithovirus; *Emiliania huxleyi* Virus (EhV)) infect only the diploid cells (Frada et al., 2008). These workers hypothesize that EhV is a crucial factor inducing meiosis and the production of haploid cells. This leads to the idea that the haploid stage could be acting as an escape strategy from viral infection, in which case the evolutionary cost of maintaining two completely distinct forms (haploid and diploid) could be surmounted by the fitness of a form that is viral resistant. Viruses may thus assume a major selective force for the maintenance of sex.

Figure 9. Satellite image of an *Emiliania huxleyi* bloom in the English Channel, off the south-western English coast, covering a region of at least 8000 km^2.

5. Thesis Objectives

The first report of viral infection associated with an *E. huxleyi* bloom appeared 17 years ago (Bratbak, Egge, and Heldal, 1993). On a first stage, and accompanying what we consider to be the first major trend in phytoplankton virology, this discovery was followed by a number of studies that confirmed the major role played by viruses in the termination of *E. huxleyi* blooms (Bratbak, Egge, and Heldal, 1993; Bratbak et al., 1995; Brussaard et al., 1996b; Castberg et al., 2001; Wilson, Tarran, and Zubkov, 2002). Enhanced by the development of specific techniques, in particular, epifluorescence microscopy (Wen, Ortmann, and Suttle, 2004), flow cytometry (Brussaard, Marie, and Bratbak, 2000), and genomic fingerprinting techniques (Wommack et al., 1999), these studies confirmed that, at times, at least 50% of the *E. huxleyi* cells in a natural bloom can be simultaneously infected, resulting in an important release of organic carbon to the environment that is re-utilized by bacteria. Even at this early stage of research, the data obtained added to a new understanding

of the planktonic realm, one in which viruses are a significant source of phytoplankton mortality, influencing phytoplankton (bloom) dynamics and have fundamental impacts on the microbial food web (Wilhelm and Suttle, 1999).

The first isolation of an *Emiliania huxleyi* specific virus (EhV; coccolithovirus) (Bratbak, Wilson, and Heldal, 1996) opened the door to new research possibilities. Following the discovery that EhV is a "giant" DNA virus (both with regards to genome and capsid size), the first DNA sequences retrieved from the genome (from the major capsid protein and DNA polymerase genes) immediately placed EhV among the wide group of Nucleo Cytoplasmic Large DNA Viruses (NCLDVs), and with close affinity to other phycodnaviruses (group of large DNA viruses that infect algae) (Castberg et al., 2002). However, it was at this early stage of the characterization of EhV that the first signs of this virus' peculiar nature were identified, reflected in the very deep phylogenetic position it occupies within the Phycodnaviridae, and separation from the other current four genera of phycodnavirus (Chlorovirus, Prasinovirus, Prymesiovirus, and Phaeovirus). This led to the classification of EhV in a newly created genus, the Coccolithovirus (Schroeder et al., 2002).

Less than 5 years ago, Wilson and co-workers (2005b) released the full genome sequence of the coccolithovirus EhV-86. This giant viral genome contained 472 open reading frames (ORFs). The great majority of these (86%) corresponded to sequences without match in the existing databases, but their functionality was immediately perceived since at least 65% were expressed during lytic infection. As usually happens with viruses, several unexpected genes were found in the EhV genome, most notably six RNA polymerase subunits and a unique *de novo* sphingolipid biosynthesis pathway. The presence of RNA polymerase subunits in this viral genome raised the hypothesis that EhV actually encoded its own transcription machinery and hence, expression of some EhV-86 transcripts could occur in the cytoplasm rather than the nucleus unlike all other known phycodnaviruses. This unique feature provided further evidence of the individuality of coccolithovirus among the other known phycodnavirus. Yet, it was the discovery of the coccolithovirus *de novo* sphingolipid biosynthesis pathway that was responsible, by far, for the generation of most of the intrigue associated with the coccolithoviruses. Sphingolipids are membrane lipids often involved in cell signalling and stress responses (Hannun and Obeid, 2008; Merrill, 2002). Notably ceramide, usually the final product of this pathway, is often implicated in the control of programmed cell death and apoptosis (Pettus, Chalfant, and Hannun, 2002; Siskind, 2005). This was the first time a sphingolipid biosynthesis pathway was discovered in a viral genome,

which raised the hypothesis that this algal virus could encode a mechanism for manipulating its host's cell death (Wilson et al., 2005b).

This thesis was planned taking into account all these findings, but also, and most particularly, the availability of the newly sequenced *E. huxleyi* genome. All the works presented herein were hence made possible by the recently developed access to both host (*E. huxleyi*) and virus (EhV) genomes.

In **Chapter 2** is presented a bibliographic review of the current knowledge we have on this viruses.

In **Chapter 3** we seek explanations for the possible origin of the extraordinary and totally unexpected EhV *de novo* sphingolipid biosynthesis pathway. The possibility of these viral genes being the product of horizontal gene transfer (HGT) was put on the table. We report the discovery of clear close homology between that viral metabolic pathway and the one present in its host's genome, and discuss the most probable case of a unique horizontal gene "importation" from *Emiliania huxleyi* to its lytic virus. An extended search for HGT events across the EhV genome resulted in the discovery of 35 most probable cases. These analyses and data are presented in **Annexe 1**.

Coming back to the EhV sphingolipid pathway, and seeking answers to the functionality of these viral genes in the environment, transcription qPCR studies were performed from field RNA samples collected during mesocosm *E. huxleyi* blooms in North Atlantic waters. In **Chapter 4** we report the clear expression of this viral pathway during EhV infection in the ocean, and discuss possible implications of the observed host to virus transcription shift.

Afterwards, the scope of a transcription analysis during natural EhV infections was enlarged to wide representation of both host (first time attempt) and virus genomes through the use of microarray techniques. The resulting observation of consistent viral takeover and viral genome activation during infection in the wild are reported in **Chapter 5**. A panorama of metabolic requirements during EhV infection is presented, including interesting novel metabolic features previously not reported in the EhV infection process.

Attempts to isolate new coccolithophore viruses made a significant part of this PhD thesis. A brief report on these attempts is thus presented in **Chapter 6**.

Chapter 2.
Coccolithovirus – a review*

António Pagarete[1], Colomban de Vargas[1], Michael Allen[2] and William Wilson[3]

1. Equipe EPPO-Evolution du Plancton et PaléoOcéans, CNRS-UMR7144, Université Pierre et Marie Curie, Station Biologique, FR-29682 Roscoff, France.

2. Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, PL1 3DH, UK.

3. Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, 04575 ME, USA.

1. Biological properties

Coccolithoviruses infect numerous strains of the Prymnesiophyte alga *Emiliania huxleyi*, the most abundant and ubiquitous coccolithophore in today's oceans. Their type species is *Emiliania huxleyi* virus 86 (EhV-86). Their host's life cycle is haplo-diplontic, with the 2N phase being responsible for periodic extensive blooms in temperate ocean water. These viruses are only capable of infecting the diploid phase (Frada et al., 2008). Their lytic infection is very severe and is the main cause of the *E. huxleyi* bloom demise (Bratbak, Egge, and Heldal, 1993; Bratbak, Wilson, and Heldal, 1996). By the end of the blooms a clear distinct population of EhV like particles can be recognized using flow cytometry analysis (Brussaard et al., 1996b; Castberg et al., 2001; Jacquet et al., 2002). Densities can reach up to 10^7 virus particles ml^{-1} (Wilson, Tarran, and Zubkov, 2002). The early stages of bloom development are usually characterized by the presence of many different EhV genotypes, which seem to suffer strong selection, leaving only a few genotypes present by the end of the bloom (Martinez-Martinez et al., 2006; Schroeder et al., 2003).

Coccolithoviruses have proven relatively easy to isolate from water samples taken at the end of *E. huxleyi* bloom events (Castberg et al., 2002). Susceptible host strains usually lyze between 2 and 7 days after the addition of 0.45 µm filtered infectious seawater (Wilson et al., 2002). Clonal isolates can be obtained by plaque or dilution assays (Schroeder et al., 2002). To date 18 EhV strains have been isolated from *E. huxleyi* blooms occurring in

* *The information presented here will be published in the form of a review on Coccolithoviruses in the next edition of the book Big Encyclopedia of Viruses (Springer Editions).*

different locations of the North Atlantic ocean, including the English Channel, Norwegian fjords, and the coast of Maine (West Atlantic) (Table 1).

Table 1. List of the different EhV strains isolated to date.

Strain name	Geographic location	Isolation date
EhV-84	English Channel	1999
EhV-86 (type species)	English Channel	1999
EhV-88	English Channel	1999
EhV-163	Raunefjorden, Norway	2000
EhV-201	English Channel	2001
EhV-202	English Channel	2001
EhV-203	English Channel	2001
EhV-204	English Channel	2001
EhV-205	English Channel	2001
EhV-206	English Channel	2001
EhV-207	English Channel	2001
EhV-208	English Channel	2001
EhV-209	English Channel	2001
EhV-V1	Raunefjorden, Norway	2003
EhV-V2	Raunefjorden, Norway	2003
EhV-2KB1	Maine, North-West Atlantic	2008
EhV-2KB2	Maine, North-West Atlantic	2008
EhV-99B1	Maine, North-West Atlantic	2008

To date, little is known on the natural distribution/variation of the different EhV strains, as well as the rates of selection to which these viruses are subjected. Recent data, corroborated by the accordance between three complementary techniques (phylogeny based on DNA polymerase and major capsid protein gene sequences; host range infection assays; microarray-based wide genome approach), indicated that the EhV strains isolated, in different years, in the English Channel and in Norway cluster phylogeneticaly according to both temporal and geographical proximity (Allen et al., 2007).

2. The EhV virion

2.1. Morphological and structural properties

The virion size ranges from 170 to 190 nm. It has an electron dense core, surrounded by a clearly defined icosahedral capsid and an external lipid envelope (Fig. 1) (Mackinder et al., 2009; Schroeder et al., 2002). EhV virions have a density of approx. 1.2 g / ml after CsCl centrifugation (Schroeder et al., 2002).

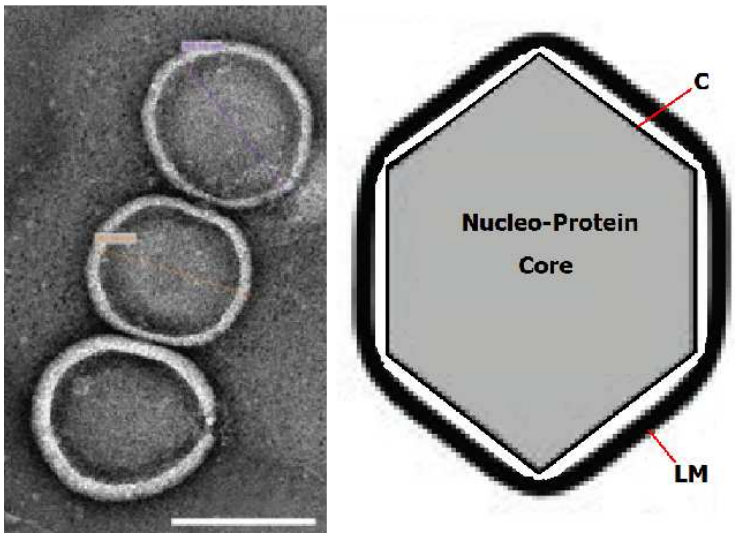

Figure 1. On the left a transmission electron micrograph depicting the presence of EhV particles in a lysed *E. huxleyi* culture (*courtesy of Dr. Declan Schroeder*). On the right a diagrammatic view of the EhV virion highlighting the central **Nucleo-protein core**, surrounded by an icosahedral capsid (**C**) and an exterior lipid membrane (**LM**). Bar 200 nm.

2.2. Virion proteome

The coccolithovirus virion is composed of at least 28 proteins, 23 of which are predicted to be membrane proteins (Allen et al., 2008). From the total 28 proteins, 10 have been assigned putative functions including the major capsid protein, two lectin proteins, a thioredoxin and a serine/threonine protein kinase (Table 2). According to Allen and colleagues (2008) the other proteins suggest potential roles involved with viral budding, caspase activation, signalling, antioxidation, virus adsorption and host range determination.

Table 2. Proteins identified in the EhV virion with an assigned putative function[a].

Gene number[b]	Top Blast hit	Blast Score
ehv035	similar to SMC2 protein, *Bos taurus*	0.058
ehv036	HlyD family secretion protein, *Agrobacterium tumefaciens*	0.004
ehv085	major capsid protein, Heterosigma akashiwo virus	$7e^{-39}$
ehv175	serine/threonine protein kinase, *Populus tomentosa*	0.66
ehv182	diaminopimelate decarboxylase, *Streptococcus pneumoniae*	0.48
ehv301	NB-ARC domain containing protein, *Oryza sativa*	0.31
ehv325	envelope glycoprotein, Simian immunodeficiency virus	1.1
ehv333	CRISPR-associated protein, Cse1 family, *Pseudomonas mendocina*	0.35
ehv340	fimbrial associated sortase-like protein, *Corynebacterium diphtheriae*	0.42
ehv461	Fatty acid synthesis protein, *Herminiimonas arsenicoxydans*	2.6

[a]Adapted from Allen et al. (2008).

[b]Gene number corresponds to the designation of each gene in the EhV genome (Genbank accession number AJ890364).

2.3. Lipids

Electron and confocal microscopy imagery have shown that coccolithovirus release occurs via budding at the host membrane (Mackinder et al., 2009). Hence, the EhV virion particles are coated in a lipid membrane as they are released from infected cells. This is corroborated by flow cytometry data suggesting that virus release occurs before cellular disintegration. Membrane proteins identified as components of the virion are potentially responsible for coordinating this viral budding through the formation of lipid rafts at the plasma membrane. The presence of a sphingolipid biosynthesis pathway in the virus genome (Han et al., 2006; Wilson et al., 2005b) further enhances this hypothesis, since sphingolipids have also been implicated with lipid raft formation.

2.4. Nucleic acids and genome organization

All coccolithoviruses studied so far have genomes with an estimated size of 410 kb. The genome of the type strain, EhV-86, has been sequenced in its entirety (Wilson et al., 2005b). Good sequence coverage (>80%) is available for a second Norwegian virus, EhV-163

(Allen et al., 2006b), and sequencing of a nine other strains is underway (Bratbak and Allen, pers. com.).

The genome consists of a single molecule of dsDNA. Originally believed to be linear in conformation, PCR amplification over the termini revealed a random A/T single nucleotide overhang (50% A, 50% T), suggesting that the virus genome has both linear and circular phases (Allen et al., 2006a; Wilson et al., 2005b). The identification of a putative origin of replication (similar in structure to that of the Epstein Barr virus) suggests a possible rolling circle mechanism involved in genome replication. Furthermore, the presence of a DNA ligase gene closely associated with the putative origin of replication (in tandem with four endonucleases at various locations on the genome) hints that a linear genome may be packaged into the virion, which later circularizes to allow DNA replication (Allen, Schroeder, and Wilson, 2006).

In the EhV-86 genome a total of 472 coding sequences (CDSs) are predicted, with an average gene length of 786 bp. Coding density is 91%. Of the total predicted CDSs, only 66 (14%) have been annotated with functional product predictions on the basis of sequence similarity or protein domain matches (Table 3) (Wilson et al., 2005b). Twenty-five of those genes belong to the common core of genes present in the Nucleo-cytoplasmic large DNA viruses (NCLDVs) (Allen et al., 2006c).

Table 3. Functions of encoded proteins in the EhV genome[a].

Gene number	Putative protein function	Top BLAST hit	BLAST score
Nucleotide metabolism, transcription, replication and repair			
ehv018	Endonuclease	*Homo sapiens*	E = 1.8e^{-55}
ehv026	Ribonucleoside-diphosphate reductase small chain	*Nicotiana tabacum*	E = 4.4e^{-81}
ehv030	DNA polymerase delta catalytic subunit	*Mus musculus*	E = 2.1e^{-79}
ehv041	Endonuclease	*Paramecium bursaria chlorella* virus 1	E = 4.5e^{-08}
ehv064	DNA-dependent RNA polymerase II largest subunit	*Glaucosphaera vacuolata*	E = 1.9e^{-89}
ehv072	DNA-binding protein	*Paramecium bursaria chlorella* virus 1	E = 4.9e^{-19}
ehv093	HNH endonuclease family protein	*Methanosarcina mazei*	E = 6.0e^{-05}
ehv104	Putative helicase	*Drosophila melanogaster*	E = 7.9e^{-12}
ehv105	Transcription factor S-II (TFIIS) family protein	*Aeropyrum pernix*	E = 3.4e-5
ehv108	DNA-directed RNA polymerase subunit I	*Encephalitozoon cuniculi*	E = 1.1e^{-5}
ehv110	RING finger protein	*Arabidopsis thaliana*	E = 3.7e^{-06}
ehv136	Nucleic acid–binding protein	*Caenorhabditis briggsae*	E = 6.3e^{-4}
ehv158	DNA ligase	*Arabidopsis thaliana*	E = 4.0e^{-}

			103
ehv166	RING finger protein	*Schistosoma japonicum*	E = 2.7e^{-3}
ehv167	DNA-directed RNA polymerase subunit	*Homo sapiens*	E = 3.2e^{-5}
ehv230	Endonuclease	Bacteriophage T4	E = 5.6e^{-15}
ehv393	DnaJ domain-containing protein	*Plasmodium yoelii*	E = 1.4e^{-3}
ehv397	Deoxyuridine 5'-triphosphate nucleotidohydrolase	Fowlpox virus	E = 3.6e^{-21}
ehv399	DNA-directed RNA polymerase subunit	*Methanococcus jannaschii*	E = 2.6e^{-08}
ehv401	Ribonuclease	*Chlamydia trachomatis*	E = 1.2e^{-23}
ehv428	Ribonucleoside-diphosphate reductase protein	*Homo sapiens*	E = 2.3e^{-200}
ehv430	Helicase	Bacteriophage KVP40 UvsW	E = 6.0e^{-05}
ehv434	DNA-directed RNA polymerase II subunit	*Homo sapiens*	E = 3.3e^{-148}
ehv444	DNA topoisomerase	*Schizosaccharomyces pombe*	E = 1.1e^{-95}
ehv453	mRNA capping enzyme	*Homo sapiens*	E = 9.5e^{-06}
ehv459	Nucleic acid–independent nucleoside triphosphatase	*Paramecium bursaria chlorella* Virus 1	E = 2.9e^{-21}
Structural proteins and proteins involved in morphogenesis			
ehv085	Major capsid protein	EhV-86	E = 2.4e^{-90}
Fatty acid metabolism			
ehv028	Lipase	*Photobacterium profundum*	E = 2.5e^{-5}
ehv031	Sterol desaturase	*Macaca fascicularis*	E = 1.2e^{-23}
ehv050	Serine palmitoyltransferase	*Homo sapiens*	E = 2.5e^{-55}
ehv061	Fatty acid desaturase	*Anopheles gambiae* str. PEST agcp14456	E = 1.4e^{-43}
ehv077	Transmembrane fatty acid elongation protein	*Homo sapiens*	E = 2.2e^{-23}
ehv079	Lipid Phosphate phosphatase	*Arabidopsis thaliana*	E = 5.2e^{-08}
ehv415	Putative fatty acid desaturase	*Trichoplusia* NI	E = 4.1e^{-09}
Proteases			
ehv021	Serine protease	*Homo sapiens*	E = 5.0e^{-12}
ehv109	OTU-like cysteine protease	*Oryza sativa*	E = 7.1e^{-4}
ehv133	ATP-dependent protease proteolytic subunit	*Deinococcus radiodurans*	E = 1.4e^{-06}
ehv151	Serine protease	*Bombyx mori*	E = 3.0e^{-15}
ehv160	Serine protease	*Meriones unguiculatus*	E = 6.6e^{-06}
ehv349	Protease	*Arabidopsis thaliana*	E = 9.2e^{-10}
ehv361	Serine protease	*Drosophila melanogaster*	E = 1.2e^{-05}
ehv447	Serine protease	*Penaeus vannamei*	E = 9.7e^{-17}
Other Proteins			
ehv020	Putative proliferating cell nuclear antigen	*Nicotiana tabacum*	E = 2.0e^{-33}
ehv023	Deoxycytidylate deaminase	*Homo sapiens*	E = 2.1e^{-35}
ehv060	Lectin protein	*Paramecium tetraurelia*	E = 2.0e^{-13}
ehv101	Hydrolase	*Mycoplasma genitalium*	E = 2.2e^{-5}
ehv103	Vesicle-associated membrane protein	*Homo sapiens*	E = 1.1e^{-05}
ehv113	Bifunctional dihydrofolate reductase–thymidylate synthase	*Paramecium tetraurelia*	E = 2.3e^{-78}
ehv117	Phosphate permease	*Neurospora crassa*	E = 2.4e^{-22}
ehv128	ERV1/ALR family protein	*Chilo iridescent* Virus	E = 9.8e^{-5}
ehv141	Hypothetical protein	*Brachydanio rerio*	E = 2.2e^{-49}

ehv179	Major facilitator superfamily protein	*Arabidopsis thaliana*	E = $9.7e^{-18}$
ehv358	Thioredoxin	*Triticum aestivum*	E = $6.7e^{-5}$
ehv363	Lipase esterase	*Mycoplasma mycoides*	E = $6.6e^{-05}$
ehv402	Protein kinase	*Schizosaccharomyces pombe*	E = 3.9 e^{-2}
ehv403	Hypothetical protein	*Ectocarpus siliculosus* Virus	E = $3.8e^{-16}$
ehv431	Thymidylate kinase	*Clostridium tetani*	E = $7.8e^{-12}$
ehv440	Proliferating cell nuclear antigen protein	*Schizosaccharomyces pombe*	E = $8.9e^{-2}$
ehv451	Protein kinase	*Homo sapiens*	E = $1.1e^{-05}$
ehv455	Sialidase	*Homo sapiens*	E = $2.1e^{-5}$
ehv465	Putative thioredoxin protein	*Arabidopsis thaliana*	E = $9.5e^{-3}$

[a] adapted from Wilson et al. (2005b).

Coccolithoviruses possess several unique features among the Phycodnaviruses. Their genome encodes six RNA polymerase genes, all of which are expressed during infection (Wilson et al., 2005b). Three families of distinctly different repeat sequences appear throughout the EhV genome, designated Family A, B and C (Allen, Schroeder, and Wilson, 2006). Family A repeats are non-coding, found immediately upstream of 86 predicted coding sequences (CDSs) and are likely to play a crucial role in controlling the expression of the associated CDSs. Family B repeats are GC rich, coding, and correspond to possible calcium binding sites in 22 proline-rich domains found in the protein products of eight predicted EhV-86 CDSs. Family C repeats are AT-rich, non-coding, and form part of the putative origin of replication. Being involved with transcriptional control (Family A), virus adsorption/release (Family B) and DNA replication (Family C), these repeat regions are potentially of fundamental importance for virus propagation.

The analysis of the EhV-86 genome also revealed the presence of genes involved in sphingolipid biosynthesis, and a further two genes encoding desaturases (Wilson et al., 2005b). Sphingolipids are membrane lipids present in all eukaryotes and some prokaryotes. They play a key role in several processes, particularly signal transduction (Futerman and Hannun, 2004). Sphingolipid biosynthesis usually leads to the formation of ceramide (Merrill, 2002), a known suppresser of cell growth and an intracellular signal for apoptosis (Hannun and Obeid, 1995; Obeid et al., 1993).

Another potentially important feature of the coccolithovirus genome is a 100-kbp "ORFanage" region. It is located in the middle of the genome between 105 kb and 205 kb, and its function is still unclear. It is commonly referred to as "ORFanage" due to lack of known function associated with most of its CDSs (Wilson et al., 2005b). It contains approximately 150 CDSs (Allen et al., 2006d). More than half of these genes (87) are associated with a unique promoter element that drives their expression during the earliest

stages of infection (see earlier, Family A repeats) (Allen et al., 2006a; Allen, Schroeder, and Wilson, 2006). These genes encode proteins that undoubtedly play a crucial and integral role during virus infection, yet, so far their function remains a mystery.

Genomic variation among different EhVs is substantial. In a microarray analysis of 425 of the 472 predicted genes of the EhV-86 genome, more than 70 were found to be absent or sufficiently variable to cause a negative hybridization in the genome of one or more coccolithovirus isolates (Allen et al., 2007). Direct comparison of EhV-86 and EhV-163 genomic sequences reveal that of the 202 CDSs for which there is full sequence in both genomes, only 20 are completely identical at the nucleotide level and an additional 17 at the protein level (Allen et al., 2006b). Nonetheless the EhV genome consistency is still clear. The overall genome size is similar between all coccolithoviruses, and all 25 "core" NCLDV genes are present in all EhV isolates analysed so far (Allen et al., 2007).

2.5. Replication strategy

After contact with the host cell membrane the adsorption of the virion happens in a matter of a few minutes. Unlike other characterized phycodnaviruses (for example the Chlorovirus described in Chapter 1) that directly inject their DNA content into the host's plasma (Van Etten et al., 2002), the coccolithovirus virion initially maintains its integrity following entry into the cell. After passing the host's exterior membrane the EhV capsids can be seen intact in the cytoplasm with their nucleoprotein core encapsulated by the major capsid protein. Then it takes only a few seconds for the nucleoprotein core to disassemble and release its DNA in the cell cytoplasm or directly in the nucleus. An eclipse period then takes place while the viral machinery takes over the cell metabolism and starts the assembly of new virions. The first newly produced viral capsids start to appear around 3h p.i. (Mackinder et al., 2009). According to Castberg et al. (2002) around 400 to 1000 assembled virions can be seen accumulating in the host cytoplasm before progressive release. Exit of the viral capsids occurs through a budding mechanism, in which the viruses gain a lipid envelope made of their host's membrane (Fig. 2). Ultimately, the process leads to the disintegration of the host cell (Mackinder et al., 2009).

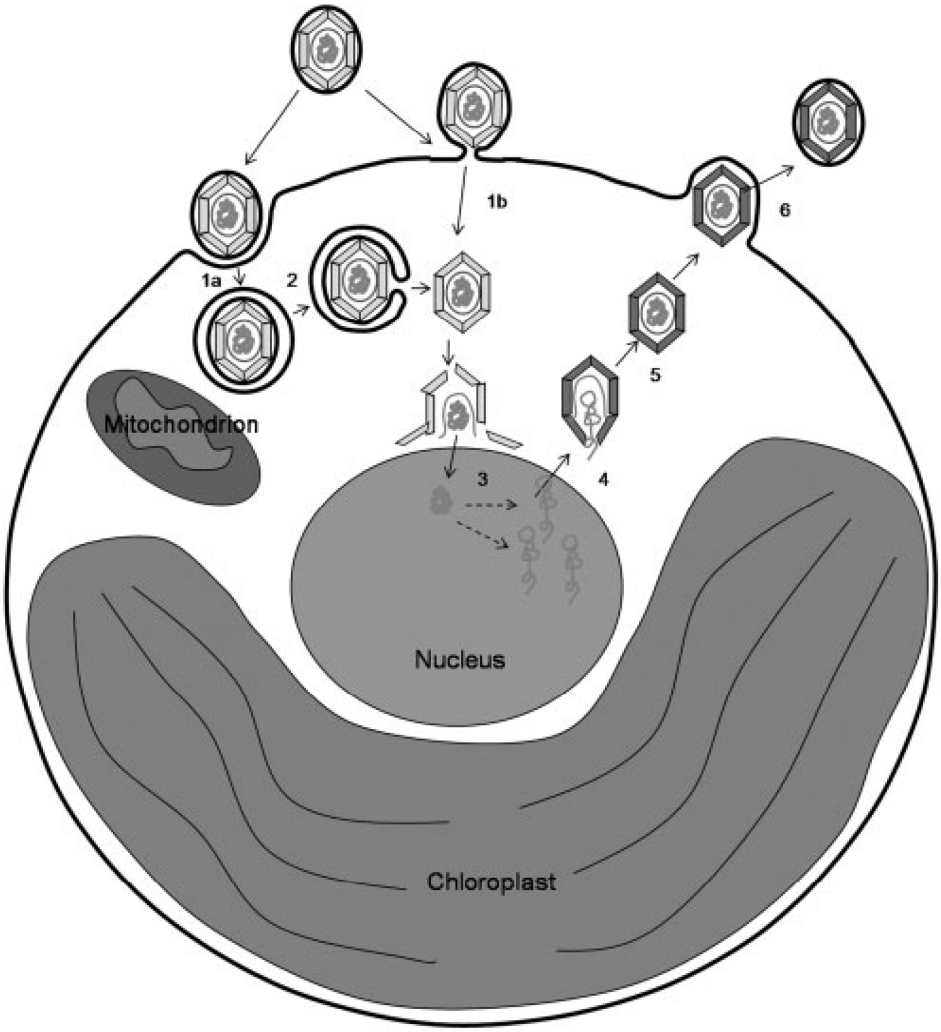

Figure 2. Schematic representation of the EhV replication cycle. The EhV virion enters the cell either via an endocytotic mechanism (step 1a) followed by fusion of its envelope with the vacuole membrane (step 2), or by fusion of its envelope with the host plasma membrane (step 1b). The capsid rapidly breaks down, releasing the viral genome (step 3). Early viral transcription starts occurring in the nucleus using host's RNA polymerase, followed by possible cytoplasm transcription using the viral encoded RNA polymerase present in EhV genome (step 4). Capsid assembly takes place in the cytoplasm (step 5), and the release of the newly formed virions occurs via a budding mechanism (step 6). (*adapted from Mackinder et al. 2009*).

The coccolithoviruses have a different propagation strategy in comparison with the latent Phaeovirus and the lytic Chlorovirus systems (presented in Chapter 1). The EhV genome possesses 6 RNA polymerase subunit genes, meaning that their replication strategy could be partially independent from the host nucleus (Wilson et al., 2005b). Viral transcription begins immediately after infection, and it is limited to a distinct 100 kb "ORFanage" region in the virus genome; this region contains a unique promoter element

(Allen et al., 2006a; Allen, Schroeder, and Wilson, 2006). The only genes transcribed during the first hour post infection are associated with this element. Proteomic analysis of EhV-86 virions has failed to detect any transcriptional machinery packaged in mature virions (Allen et al., 2008), therefore a host nuclear RNA polymerase(s) is presumably responsible for this early transcription. Between 1 and 2 hr post infection, a second transcriptional phase begins with gene expression occurring from the remainder of the genome (Allen et al., 2006a). Since viral RNA polymerase components are expressed in this second phase, viral replication may no longer be nuclear dependent at this stage and transcription may move to the cytoplasm (Fig. 2, step 4).

2.6. Phylogeny and evolution

The phylogenetic position of the coccolithoviruses is still in debate with high uncertainty regarding its evolutionary history. Several independent phylogenetic studies (Allen et al., 2006c; Larsen et al., 2008; Schroeder et al., 2002; Wilson et al., 2006) have always placed the EhV within the family *Phycodnaviridae*. However, the coccolithovirus do not cluster with the other Prymnesiovirus identified to date (whose hosts are phylogeneticaly close to *E. huxleyi*), but instead occupy a very deep position in the phycodnavirus clade (Fig. 3). This differentiation from the other members of the *Phycodnaviridae* led to the creation of the new genus *Coccolithovirus*.

The 6 RNA polymerase subunits present in the EhV genome (unique among the known phycodnaviruses) add to the singularity of the coccolithoviruses among other algal viruses. Since ancestral NCLDV contained the RNA polymerase function, it is likely that of all the phycodnaviruses sequenced so far, EhV-86 represents the virus with the lifestyle most similar to the ancestral virus (Allen et al., 2006d). The change in lifestyle represented by this loss of RNA polymerase function (i.e. from nuclear independence to nuclear dependent transcription) probably contributes to the high diversity among present day genera in the *Phycodnaviridae*.

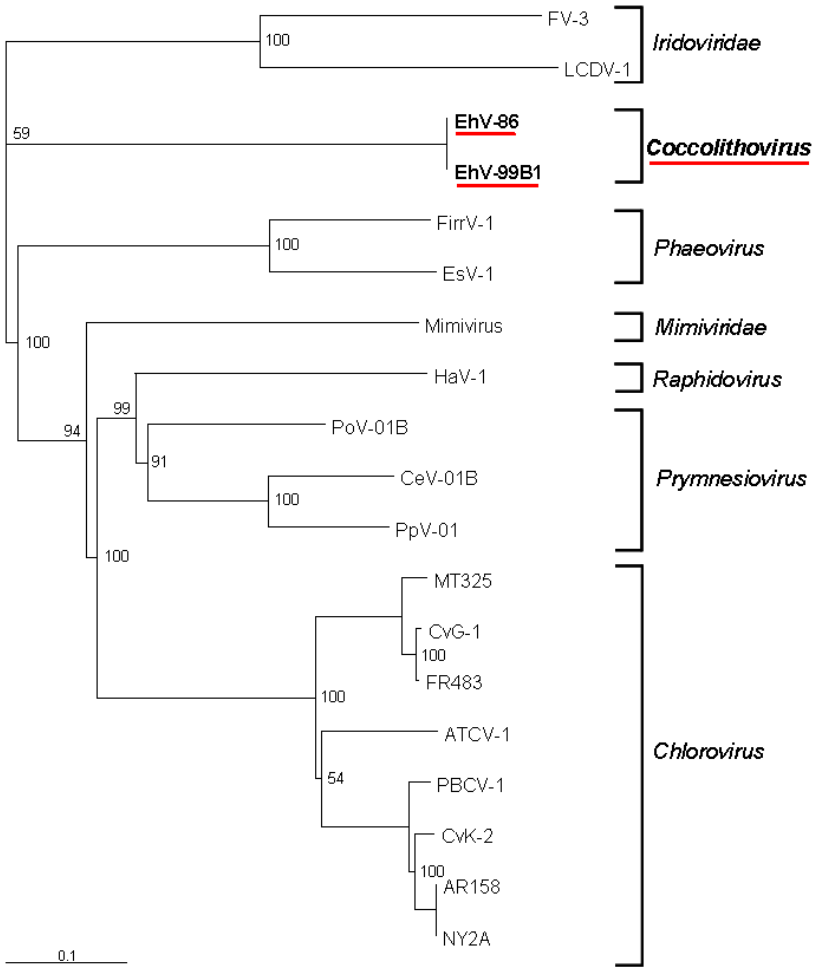

Figure 3. Phylogenetic inference using the complete protein sequences of the MCPs from 16 members of the *Phycodnaviridae* family. The tree was rooted using the sequences of FV-3 and LCDV-1 of the *Iridoviridae* family. The scale bar indicates the number of amino acid substitutions per residue. Viruses included are Frog virus (FV-1), Lymphocystis disease virus (LCDV-1), Emiliania huxleyi virus (EhV-86, EhV-99B1), Feldmannia irregularis virus (FirrV-1), Ectocarpus siliculosus virus (ESV-1), Mimivirus, Heterosigma akashiwo virus (HaV-1), Pyramimonas orientalis virus (PoV-01B), Chrysochromulina ericina virus (CeV-01B), Phaeocystis pouchetii virus (PpV-01), and Paramecium bursaria chlorella virus (MT325, CvG-1, FR483, ATCV-1, PBCV-1, CvK-2, AR158, NY2A). *Adapted from Larsen et al. (2008).*

Chapter 3.

Horizontal gene transfer of an entire metabolic pathway between a eukaryotic alga and its DNA virus*

Adam Monier[1], António Pagarete[2], Colomban de Vargas[2], Michael J. Allen[3], Betsy Read[4], Jean-Michel Claverie[1] and Hiroyuki Ogata[1]

1. Structural and Genomic Information Laboratory, CNRS-UPR2589, Mediterranean Institute of Microbiology (IFR-88), Université de la Méditerranée, Parc Scientifique de Luminy, 163 avenue de Luminy Case 934, FR-13288 Marseille, France.

2. Equipe EPPO-Evolution du Plancton et PaléoOcéans, CNRS-UMR7144, Université Pierre et Marie Curie, Station Biologique, FR-29682 Roscoff, France.

3. Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, PL1 3DH, UK.

4. Department of Biological Science, California State University San Marcos, CA, 92096-0001, USA.

1. Summary

Interactions between viruses and phytoplankton, the main primary producers in the oceans, affect global biogeochemical cycles and climate. Recent studies are increasingly revealing possible cases of gene transfers between cyanobacteria and phages, which might have played significant roles in the evolution of cyanobacteria/phage systems. However, little has been documented about the occurrence of horizontal gene transfer in eukaryotic phytoplankton/virus systems. Here we report phylogenetic evidence for the transfer of seven genes involved in the sphingolipid biosynthesis pathway between the cosmopolitan eukaryotic microalga *Emiliania huxleyi* and its large DNA virus EhV. PCR assays indicate that these genes are prevalent in *E. huxleyi* and EhV strains isolated from different geographic locations. Patterns of protein and gene sequence conservation support that these genes are functional in both *E. huxleyi* and EhV. This is the first clear case of horizontal gene transfer of multiple functionally-linked enzymes in a eukaryotic phytoplankton-virus system. We examine arguments for the possible direction of the gene transfer. The virus-to-host direction suggests the existence of ancient viruses that controlled the complex metabolic pathway in order to

** Published in Genome Research 19(8): 1441-1449, 2009*

infect primitive eukaryotic cells. In contrast, the host-to-virus direction suggests that the serial acquisition of genes involved in the same metabolic pathway might have been a strategy for the ancestor of EhVs to stay ahead of their closest relatives in the great evolutionary race for survival.

2. Introduction

Oceanic phytoplankton (cyanobacteria and eukaryotic microalgae) is responsible for over half of the Earth's primary production. The high level of biological production by microalgae is principally due to their fast turnover rate of 2 to 6 days on average, relative to 19 years for land plants (Field et al., 1998). During the last two decades, oceanographers discovered that viral infection is one of the major causes of death of marine microorganisms, including microalgae (Nagasaki, 2008; Suttle, 2007). It is estimated that 20-40% of microorganisms in surface sea waters are killed by viruses each day. Being basal components of marine microbial food webs, viruses thus significantly contribute to the cycling of energy and nutrient on a global scale.

Emiliania huxleyi is the most prominent modern coccolithophore, a group of photosynthetic marine unicellular eukaryotes that play a critical role in ocean biogeochemistry (de Vargas et al., 2007). *E. huxleyi* is a member of the Haptophyta, one of the deepest branching lineages of the eukaryotic tree of life. This microalgal species is known for its beautiful exoskeleton made of calcium carbonate scales ("coccoliths"), and its recurrent blooms turning extensive areas (>10,000 km^2) of oceanic surface waters milky-white. *E. huxleyi* actively participates to CO_2 exchange between the atmosphere, seawater, and the lithosphere, through the synthesis of coccoliths and by driving massive sinking of organic and inorganic carbon into the deep sea. It thus plays a critical role on global carbon cycling and climate change (Charlson et al., 1987; Westbroek et al., 1994). *E. huxleyi* blooms suddenly terminate with a sharp increase in the abundance of giant viruses ("coccolithoviruses") which infect and lytically kill the microalgae (Bratbak, Egge, and Heldal, 1993; Delille et al., 2005). Coccolithoviruses are large double stranded DNA viruses, and form a monophyletic group within the virus family *Phycodnaviridae* (Allen et al., 2006c; Schroeder et al., 2002). *E. huxleyi* virus 86 (EhV-86) is the type species of the genus *Coccolithovirus*, and was originally isolated from a seawater sample collected from a dying *E. huxleyi* bloom in the English Channel. The 407 kbp-genome of EhV-86 is predicted to encode 472 proteins (Wilson et al.,

2005b), and is the second largest among sequenced eukaryotic viral genomes (Raoult et al., 2004).

Along with the sequencing of the EhV-86 genome came the discovery, for the first time in a virus, of a series of enzyme-coding genes predicted to be involved in the biosynthesis of sphingolipids (Wilson et al., 2005b). Sphingolipids are membrane lipids present in all eukaryotes and some prokaryotes, that are involved in the regulation of various cellular processes (Futerman and Hannun, 2004). Sphingolipid metabolism has been mostly studied in mammalian and yeast cells, where it was found to play key roles in signal transductions (Hannun, Luberto, and Argraves, 2001). The *de novo* sphingolipid biosynthesis leads to the production of ceramide (Merrill, 2002), which serves as the backbone for all complex sphingolipids and has a fundamental role in coordinating eukaryotic cell stress responses including activation of apoptosis (Guenther et al., 2008; Hannun, 1996; Yang et al., 2004).

The genome of EhV-86 encodes at least seven enzymes predicted to be involved in the biosynthesis of sphingolipids (Wilson et al., 2005b). These include four enzymes central to this metabolic pathway (Fig. 1, (Merrill, 2002)): serine palmitoyltransferase (SPT), dihydroceramide synthase (longevity assurance factor 1, LAG1), dihydroceramide desaturase (Dsd1-like fatty acid desaturase, FAD) and sphingosine 1-phosphate phosphatase (lipid phosphate phosphatase, LPP). The remaining viral enzymes related to sphingolipid biosynthesis are a sterol desaturase, a transmembrane fatty acid elongation protein, and an Aco1-like FAD (Table 1). Together, these enzymes constitute an almost entire *de novo* sphingolipid biosynthesis pathway. The seven viral genes are dispersed in the EhV-86 genome, which does not encode any obvious homologue for 3-ketosphinganine reductase (3-KSR), the enzyme catalyzing the second step of the sphingolipid biosynthesis pathway.

Transcription of the viral genes involved in the sphingolipid pathway is coordinately regulated. It starts at 2 hours postinfection, corresponding to an early stage of the viral replication cycle, which lasts from > 4 hours up to 2 days (Allen et al., 2006a). This timing coincides with the first expression of the viral RNA polymerase (Allen et al., 2006a; Wilson et al., 2005b), suggesting that these viral genes may be transcribed by the virally encoded transcription machinery and thus expressed in the host cytoplasm (Allen et al., 2006a). The viral sphingolipid enzymes are not packaged in EhV-86 virions according to a proteomic survey (Allen et al., 2008). The sole EhV sphingolipid enzyme biochemically characterized to date is the SPT, which exhibits atypical domain fusion architecture. In most eukaryotes, SPTs are heterodimers comprised of two aminotransferase subunits, the long chain base 1 (LCB1)

and LCB2. Han *et al.* (Han et al., 2006) found that the EhV-86 gene ehv050 encodes a single polypeptide with N-terminal LCB2-like and C-terminal LCB1-like domains. They also found similar fusion proteins in EST libraries from *E. huxleyi* and *Entamoeba histolytica*. The viral SPT was further expressed in yeast, demonstrating its activity and unusual preference for myristoyl-CoA (C14) rather than palmitoyl-CoA (C16) (Han et al., 2006). Despite these recent efforts to characterize the EhV sphingolipid biosynthesis genes, their function in the virus replication cycle remains unknown. Based on the role of ceramide as an inducer of cell death in mammalian and yeast cells (Guenther et al., 2008; Siskind, 2005; Susin et al., 1997), several authors proposed that the viral sphingolipid/ceramide pathway may activate host cell death, thus helping disseminate newly generated virions in the host population (Bidle et al., 2007; Wilson et al., 2005b).

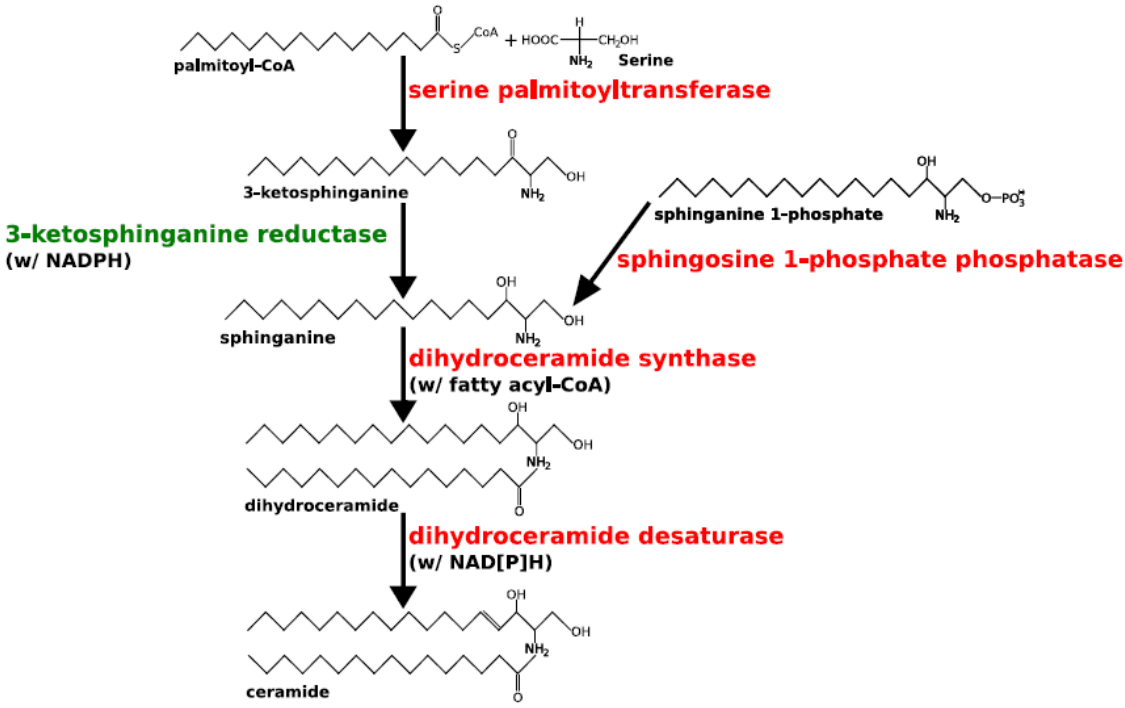

Figure 1. A model of *de novo* sphingolipid/ceramide biosynthesis pathway. The enzymes found in both EhV-86 and *E. huxleyi* are indicated in red. The enzyme present only in *E. huxleyi* is indicated in green.

The unique presence of sphingolipid enzyme genes in only EhV among all known viruses and their ubiquitous distribution in eukaryotes suggest the possibility of horizontal gene transfers (HGTs) of these functionally linked enzyme genes between ancestral virus and eukaryotic host lineages. Viruses are known to carry a variety of host genes. Recent genomics studies are increasingly revealing interesting cases of HGT between prokaryotic

phytoplankton (cyanobacteria) and their viruses (cyanophages) (Lindell et al., 2005; Lindell et al., 2004; Sullivan et al., 2006; Yoshida et al., 2008).

Table 1. Sphingolipid biosynthesis enzymes in the giant virus EhV-86 and its coccolithophore host *E. huxleyi.*

Enzymes	**EhV-86 CDS**	***E. huxleyi* CDS***	***E. huxleyi* scaffold ID / scaffold size / CDS position**
Serine palmitoyltransferase (SPT)	YP_293804 (*ehv050*)	432901	Scaff 7 / 1.4 Mb / 1020001-1016564
3-ketosphinganine reductase (3-KSR)	absent	437991	Scaff 68 / 604 kb / 377451-376298
Dihydroceramide synthase (longevity assurance factor 1, LAG1)	YP_293768 (*ehv014*)	200862	Scaff 13 / 1.1 Mb / 88654-89151
Fatty acid desaturase (Dsd1-like)	YP_293875 (*ehv061*)	54601	Scaff 675 / 28 kb / 8222-7338
Lipid phosphate phosphatase (LPP)	YP_293833 (*ehv079*)	193908	Scaff 1 / 3 Mb / 540559-541506
Transmembrane fatty acid elongation protein	YP_293831 (*ehv077*)	70214	Scaff 118 / 428 kb / 11947-12897
Sterol desaturase	YP_293785 (*ehv031*)	210457	Scaff 43 / 769 kb / 605740-606537
Fatty acid desaturase (Aco-1 like)	YP_294173 (*ehv415*)	236135	Scaff 16 / 1.1 Mb / 267821-266730

* CDS IDs from the JGI reduced protein set.

From the observation of "host-like" genes in bacteriophage genomes, Hendrix and other authors proposed a modular theory of phage evolution, in which phages evolve through the stepwise acquisition of genes from diverse sources (Brussow and Hendrix, 2002; Hendrix et al., 2000). Eukaryotic large DNA viruses also exhibit genes with homologs in cellular organisms, such as those related to immune system in poxviruses (Hughes and Friedman, 2005), and homologs of cellular genes found in the amoeba-infecting giant mimivirus; albeit with controversy on the timing, mechanism and frequency for possible gene transfers (Claverie, 2006; Filee, Siguier, and Chandler, 2007; Moreira and Brochier-Armanet, 2008; Ogata and Claverie, 2007; Raoult et al., 2004). Nevertheless, little has been documented about the occurrence of gene transfer in eukaryotic alga-virus systems due to the limited availability of genomic sequence data for such host-virus pairs. A recent comparative genomics study of the green alga *Ostreococcus tauri* and its virus OtV5 could reveal only one putative case of HGT for this eukaryotic alga-virus pair (Derelle et al., 2008). Here we test the hypothesis that HGT is at the origin of the EhV sphingolipid biosynthesis genes using the recently released draft genome sequences of *E. huxleyi* diploid strain CCMP1516 (7809 scaffolds, 168 Mb, 10X coverage) determined by the International *E. huxleyi* Genome Sequencing Consortium.

3. Materials and Methods

Sphingolipid biosynthesis gene sequences from E. huxleyi CCMP1516

The genome sequence data of *E. huxleyi* CCMP1516 strain were produced by the International *E. huxleyi* Genome Sequencing Consortium in collaboration with the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/). The genome sequence data are being analyzed by the consortium members and will be published elsewhere. The amino acid sequences corresponding to the seven EhV-86 sphingolipid biosynthesis genes (Table 1) were used to identify their homologs in the *E. huxleyi* genome sequences, using BLASTP searches (Altschul et al., 1997) against the host's ORFeome (the JGI reduced protein set as of April 4, 2008; E-value<10^{-20}). For the detection of *E. huxleyi* 3-KSR homolog, 3-KSR homologs from green plants (*Arabidopsis thaliana* and *Ostreococcus tauri*) were used as TBLASTN queries.

PCR-amplification and sequencing of sphingolipid biosynthesis genes from host and virus strains

Six *E. huxleyi* and eleven EhV strains were chosen by taking into account their distant geographical origins (Table S1) and distinct behavior regarding susceptibility to EhV infection (data not shown). To extract *E. huxleyi* DNA, 250 ml of late exponential growing cultures were harvested by centrifugation (14000 rpm for 2 mins). A 0.5ml pellet was recovered and initially treated with proteinase K (5 mg/ml) in a lysis buffer containing 20 mM EDTA, pH 8 and 0.5% SDS (w/v) at 65 ºC for 1 h. Major cell debris was removed by adding 600 µl of phenol to each sample and centrifuging at maximum speed for 10 min. The top layer was recovered and the DNA was extracted using an equal volume of chloroform:isoamyl alcohol (24:1). The DNA was precipitated with the addition of 0.5 × volume 7.5 M ammonium acetate, pH 7.5, and 2.5 × volume absolute ethanol. The pellet was washed 3 times in 300 µl of ice-cold 70% ethanol, after which it was dried and re-suspended in 30 µl of DNase free water. The virus isolates were directly used as DNA template for PCR without prior DNA purification. The on-line application Primer3 (Rozen and Skaletsky, 2000) was used to design primers that target homologous regions in both host and viral genes (Table

S2). The PCR reaction was set up as follows: 1 µl of DNA template (extracted DNA in case of the hosts, viral isolate in the case of the virus) was added to a 25 µl reaction mixture which contained: 1 U *Taq* DNA polymerase (Promega), 1 × PCR reaction buffer (Promega), BSA, 0.25 mM dNTPs, 2.5 mM $MgCl_2$, 10 pmol of each primer. The PCR was conducted in a PTC-100™ cycler (MJ Research) with an initial denaturing step of 95 ºC (5 min), followed by 35 cycles of denaturing at 95 ºC (60 s), annealing at 56 ºC (60 s), and extension at 74 ºC (60 s). A SequiTherm EXCEL II DNA Sequencing Kit-LC (EpicentreTechnologies) with a LI-COR Automated DNA Sequencer was used to sequence the PCR products. Obtained sequence data were deposited in GenBank (accession numbers from FJ531546 to FJ531633).

EhV-86/EhV-163 orthologs

We extracted open reading frames (≥ 60 aa) from the partial genome sequence data of EhV-163 (Allen et al., 2006b) using EMBOSS/GETORF software (Rice, Longden, and Bleasby, 2000). EhV-86/EhV-163 orthologous sequence pairs were determined using the reciprocal BLASTP best hit criterion.

Phylogenetic analysis

EhV-86 and *E. huxleyi* sphingolipid biosynthesis-related protein sequences were used as queries for BLASTP searches against the NCBI non-redundant database (Pruitt, Tatusova, and Maglott, 2007) (E-value < 10^{-3}) to identify and retrieve their homologs. We generated multiple sequence alignments using MUSCLE (Edgar, 2004). The SPT protein sequences from EhV-86, *E. huxleyi*, *O. tauri* and *E. histolytica* were split into sub-sequences according to their particular domain architecture. All gap containing sites were removed from the alignments for the following phylogenetic analyses. Maximum-likelihood phylogenetic analyses were performed using PHYML (Guindon and Gascuel, 2003) with the Jones-Taylor-Thornton substitution model (Jones, Taylor, and Thornton, 1994) and with 100 bootstrap replicates. Maximum parsimony phylogenetic analyses were performed for PCR-amplified sequences using Phylip/DNAPARS (Felsenstein, 2004).

Estimation of Ka and Ks

Each orthologous protein sequences were aligned by MUSCLE, and then back-translated into codon alignments. The maximum likelihood computation of synonymous (Ks) and non-synonymous (Ka) substitution rates and their ratio (ω) for each orthologous pairs of sequences was performed using CODEML from the PAML 4 package (Yang, 2007). For the comparison of EhV-86/EhV-163 orthologs, we discarded all the sequence pairs showing estimated evolutionary parameters (Ks, Ka) with >50% of standard errors and those having Ks>1.0.

4. Results

Emiliania huxleyi possesses a full set of sphingolipid biosynthesis genes

Homologs of the seven viral proteins predicted to be involved in the sphingolipid/ceramide biosynthesis were readily identified (BLASTP, E-value<10^{-10}) in the *E. huxleyi* genome (Table 1). Remarkably, these *E. huxleyi* proteins (except for SPT) were the most similar to their viral counterparts, with which they shared from 26% to 49% identical residues. In addition, a 3-KSR homolog, apparently missing in EhV-86 genome, was identified in the *E. huxleyi* genome using green plant and yeast 3-KSR sequences as queries. These host enzymes were found encoded in the middle of different scaffolds with various sizes from 28 Kb up to 3 Mb, thus do not cluster in a small region of a host chromosome. The host and viral protein sequences were aligned with a wide phylogenetic array of homologs, and were examined regarding the conservation of previously reported sequence features (Jiang et al., 1998; Lindqvist et al., 1996; Mitchell and Martin, 1997; Oh et al., 1997; Stukey and Carman, 1997; Winter and Ponting, 2002). We confirmed the presence of most of the sequence motifs and conserved catalytic residues (Fig. S1), suggesting that both viral and host enzymes are functional.

The viral LPP (ehv079) and its closest host homolog (JGI_193908) belong to the phosphatidic acid phosphatase type 2 (PAP2) superfamily (Pfam PF01569; E-value<$7x10^{-9}$). At least six PAP2 superfamily proteins were found encoded in the *E. huxleyi* genome. Notably, one of the PAP2 sequences was located in the C-terminus of JGI_432901 corresponding to the enzyme SPT (E-value=0.0096), indicating a fusion of three domains for

the protein (LCB2, LCB1 and PAP2; Fig. 2). We found the same tri-domain architecture in the homologous protein from the green alga *O. tauri*.

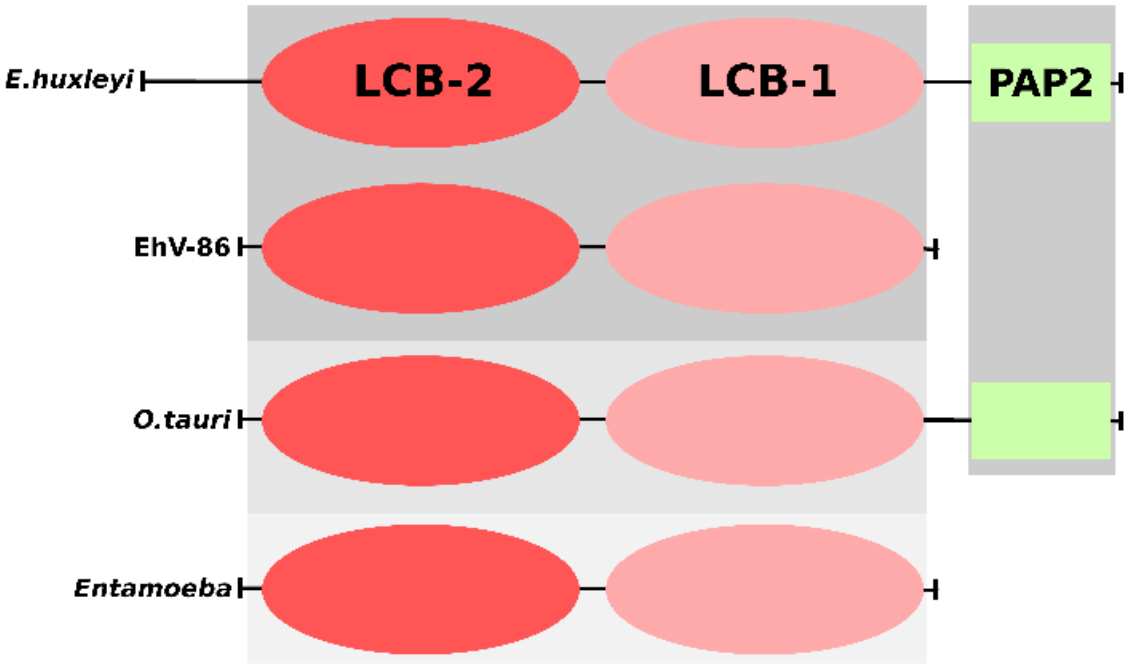

Figure 2. Domain architectures of serine palmitoyltransferases from *E. huxleyi*, EhV-86, *O. tauri* and *Entamoeba* spp. LCB2-like domains correspond to red ovals, LCB1-like domain pink ovals, and PAP2 domains green rectangles. Background gray scales correspond to the level of sequence similarity from the *E. huxleyi* SPT sequence (darker grey for higher BLAST scores).

E. huxleyi thus possesses a complete set of enzymes for the sphingolipid biosynthesis. We attempted to detect a distant 3-KSR homolog in EhV-86, using the newly identified host 3-KSR sequence and a Pfam profile for 3-KSRs (PF00106) as queries, but no such homologue could be found in the viral genome.

Evidence of horizontal gene transfers between giant DNA virus and its eukaryotic host

Maximum likelihood phylogenetic trees were reconstructed for the seven sphingolipid biosynthesis enzymes shared by EhV-86 and *E. huxleyi*, including the widest taxonomic range of homologs available in GenBank (Fig. 3, Fig. S2). In all cases, the trees displayed a monophyletic grouping of EhV-86 and *E. huxleyi* sequences, including their respective SPT's LCB1 and LCB2 domains (Fig. 3a). In five cases, >90% bootstrap values supported the clustering of EhV-86 and *E. huxleyi* protein sequences (96% for the C-terminal LCB1 domain, Fig. 3a; 100% for the LAG1, Fig. 3b; 100% for the Dsd1-like FAD, Fig. 3d; 100% for the sterol desaturase, Fig. S2A; and 91% for the transmembrane fatty acid elongation

protein, Fig. S2C). The branching positions of the EhV/*E. huxleyi* sequence groups for these genes are generally compatible with a deep phylogenetic origin of *E. huxleyi* within eukaryotes, except for the Aco1-like FADs (Fig. S2B).

These results strongly suggest that the seven sphingolipid biosynthesis genes were horizontally transferred between the eukaryotic and viral lineages leading to *E. huxleyi* and EhV-86, respectively. The EhV/*E. huxleyi* Aco1-like homologs were more similar to bacterial homologs than to eukaryotic homologs in terms of both domain organization and sequence; fungal homologs have two delta-9-desaturase domains and an additional chytochrome b_5-domain (Sperling et al., 2003), while a single delta-9-desaturase domain was identified for EhV, *E. huxleyi* and bacterial homologs.

Legend of Figure 3 (next page). Maximum likelihood phylogenetic trees based on the amino acid sequences of the four central enzymes in the sphingolipid biosynthesis pathway. (a) Serine palmitoyltransferase LCB1 and LCB2 domain sequences and their homologs. (b) Dihydroceramide synthases (LAG1). (c) Dsd1-like fatty acid desaturases. (d) Lipid phosphate phosphatase (LPP), the PAP2-domain sequence from the *E. huxleyi* SPT, and their homologs. These trees are unrooted *per se*, although we have arbitrarily chosen a root (mostly by mid-point rooting) for each tree only for visualization purpose. The number of substitutions per site is indicated under the scale bar. In (d), sequences best hitting to sphingosine 1-phosphate phosphatases (cd03388) after NCBI/CDD searches are marked by '#', those best hitting to phosphatidic acid phosphatases (cd03390) are marked by '*',

and those best hitting to the wunen subfamily sequences (a family of membrane associated phosphatidic acid phosphatases; cd03384) are marked by '§'.

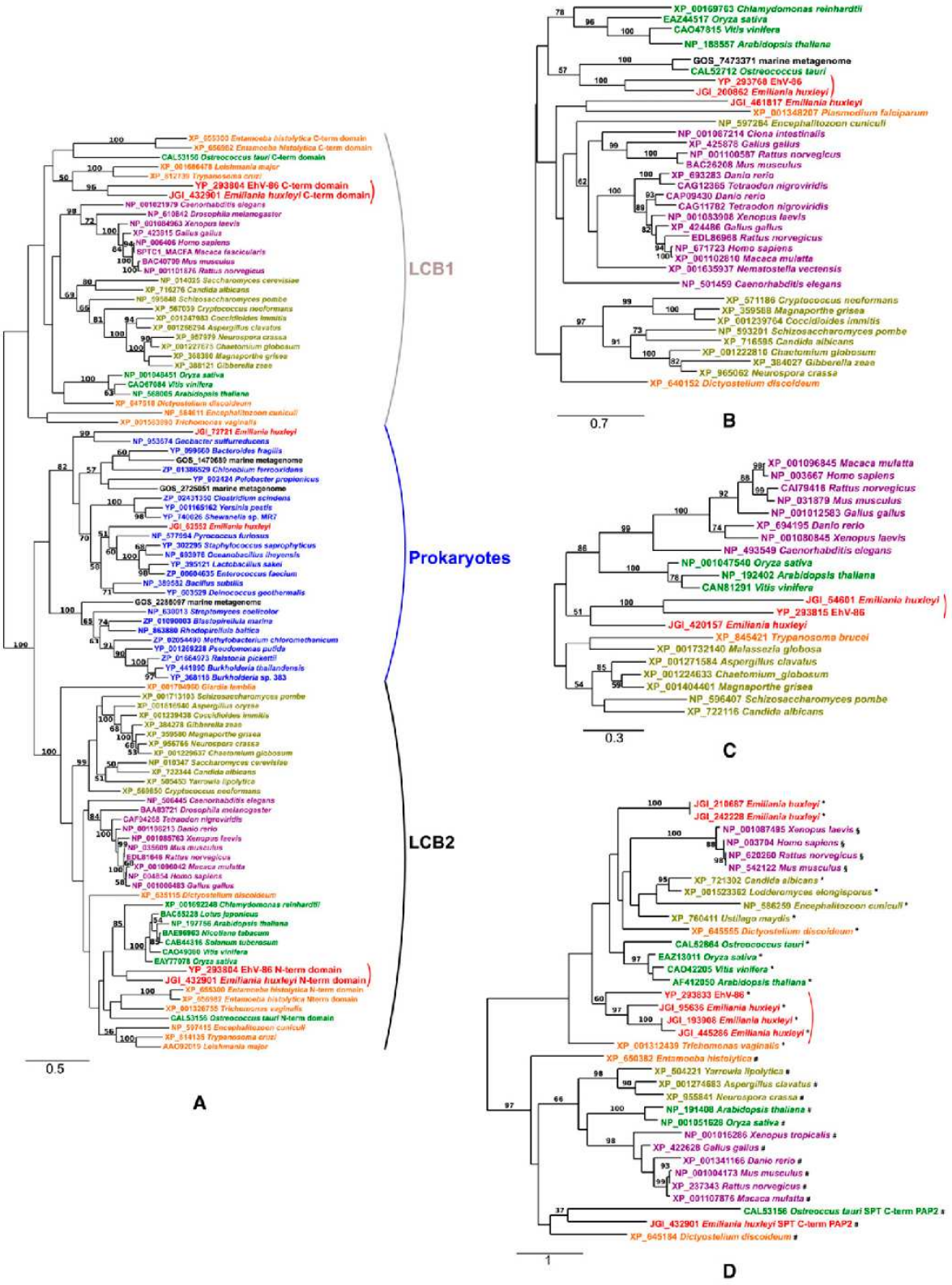

Figure 3. *(legend on previous page)*

Viral sphingolipid biosynthesis genes are widespread and functional

To further assess the presence and function of the sphingolipid/ceramide biosynthesis genes in the *E. huxleyi*/EhV system, we looked for 6 of them in various *E. huxleyi* and EhV strains covering a wide geographic range (Table S1). We could successfully PCR amplify and sequence most of the genes in the host strains, and all of them, except for 3-KSR, in the 11 virus strains. The sphingolipid biosynthesis genes thus appear to be prevalent in different *E. huxleyi* and EhV strains. The amplified gene sequences from the host strains were highly similar (Fig. S3). In contrast, the viral sequences exhibited substantial inter-strain variation, and were further used to investigate the functional status of the encoded enzymes by assessing evolutionary rates and selection pressure at the sequence level.

Maximum parsimony trees for the five genes from the eleven viral strains showed topologies compatible with each other and revealed a clustering of the strains into two groups, I and II (Fig. S4). This clustering is consistent with previous reports (Allen et al., 2007), and correlates with the isolation time points and/or geographical origins of the strains (Table S1). Group I is composed of three strains isolated from the English Channel in 1999 and three strains isolated during mesocosm experiments in Norway in 2000 and 2003, while group II comprises exclusively strains isolated from the English Channel in 2001. We computed synonymous (Ks) and non-synonymous (Ka) substitution rates, and their ratio (Ka/Ks=ω) between group I and group II sequences (Table 2). The average ω values were substantially below 1 ranging from 0.09 for Aco1-like FAD to 0.43 for LPP.

Table 2. Nucleotide substitution rates and their ratio of the viral sphingolipid biosynthesis genes measured by the comparisons between the group I and group II strains.

Genes	Ka †	Ks †	ω †
SPT	0.005 (0 – 0.021)	0.044 (0.034 – 0.069)	0.101 (0.001 – 0.336)
LAG1	0.012 0.011 – 0.014	0.074 (0.055 – 0.088)	0.176 (0.134 – 0.224)
Dsd1-like FAD	0.0232	0.1141	0.2036
Aco1-like FAD	0.008 (0.002 – 0.013)	0.084 (0.083 – 0.086)	0.093 (0.032 – 0.154)
LPP	0.014 (0.011 – 0.016)	0.038 (0.029 – 0.054)	0.383 (0.312 – 0.49)

† The average Ka, Ks and ω values followed by the range in parentheses.

Finally, we assessed the evolutionary rates of the viral sphingolipid biosynthesis genes relative to other EhV genes by computing Ka and Ks values for orthologous gene pairs between two viral strains, EhV-86 and EhV-163 (Allen et al., 2006b). The substitution rates

of the sphingolipid biosynthesis related genes were found comparable to those of other EhV genes (Fig. 4). Overall, our results suggest that a negative selection acted on these viral genes, again supporting that they are fully functional.

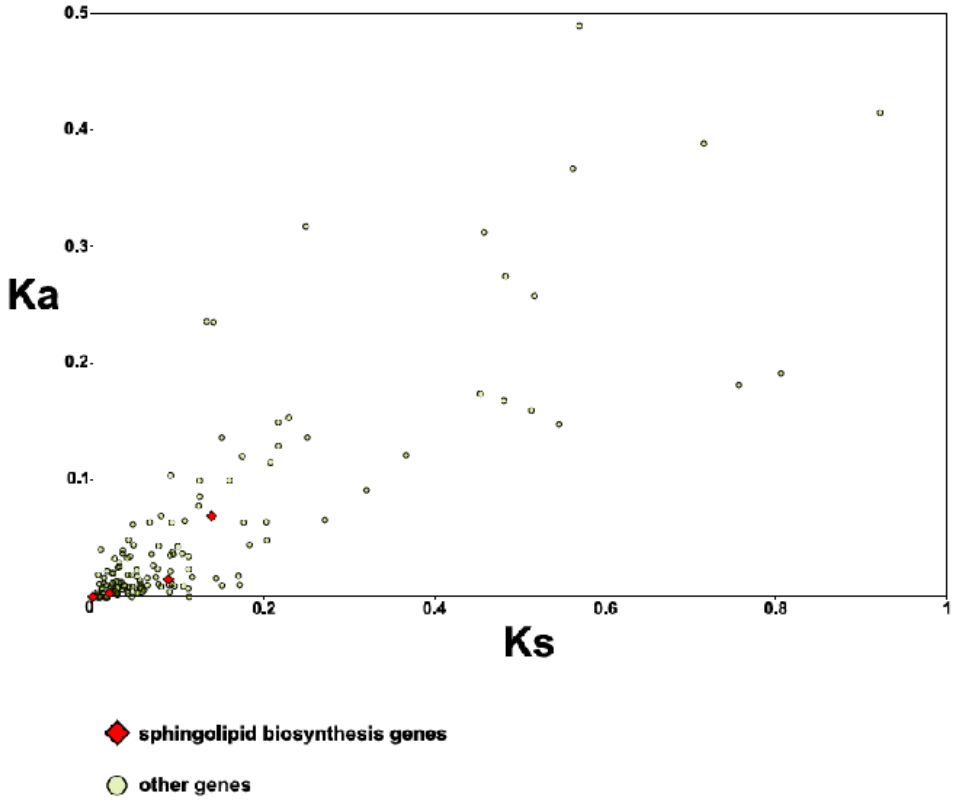

Figure 4. Synonymous (Ks) and non-synonymous (Ka) substitution rates for the orthologous sequences between EhV-86 and EhV-163. Of the seven EhV-86 sphingolipid enzyme genes, EhV-163 orthologs were found for four genes (SPT, LAG1, Dsd1-like FAD, LPP), which are indicated in red rectangles in this figure.

5. Discussion

We have provided here clear evidence supporting HGT(s) between *E. huxleyi* and EhV for seven genes probably involved in the biosynthesis of sphingolipids. Given the previously reported co-transcription of the viral genes (Allen et al., 2006a) and the validated enzymatic activity of the viral SPT (Han et al., 2006), all of these viral enzymes are likely to be functional. The prevalence of those genes in a wide range of host and virus strains and the pattern of their amino acid and DNA sequence conservation revealed by this study further support that these enzymes are functional in both *E. huxleyi* and EhVs.

The presence of conserved sequence motifs and catalytic residues in the viral and host enzyme sequences, and their phylogenetic positions within individual enzyme families are generally compatible with the current annotations of their enzymatic functions. Regarding the classification of these putative enzymes, the PAP2 superfamily sequences are worth mentioning. The enzyme LPP belongs to the PAP2 superfamily, which includes a variety of phosphatase subfamilies (Pfam PF01569). Our phylogenetic analysis (Fig. 3c) indicates that the viral LPP (ehv079) and its closest host homolog (JGI_193908) are more similar to phosphatidic acid phosphatases than to sphingosine 1-phosphate phosphatases. Thus the *bona fide* substrate of these LPPs may not be sphingosine 1-phosphate. By contrast, the PAP2 domain embedded in the C-terminal region of the *E. huxleyi* SPT sequence (JGI_432901) appears more similar to many sphingosine 1-phosphate phosphatases than other phosphatases. This algal protein may thus be involved in two distinct steps shown in the sphingolipid/ceramide biosynthesis pathway model in Fig. 1.

Wilson *et al.* (2005b) originally proposed that the viral enzymes would be part of a viral lysis strategy to kill the host cell. This hypothesis was based on the role of ceramide as an inducer of cell death in mammalian and yeast cells (Guenther et al., 2008). The presence of eight protease genes in the EhV-86 genome (Wilson et al., 2005b) parallels the known connection between protease activation and ceramide-induced programmed cell death in mammalian cells (Siskind, 2005; Susin et al., 1997). Bidle et al. (2007) recently demonstrated the activation of programmed cell death of *E. huxleyi* cells upon EhV infection and concomitant induction of host metacaspases. Notably, the EhV-86 genome encodes eight proteins with caspase cleavage recognition sequence motifs. Wilson *et al.* (2009) also suggested that the viral sphingolipid biosynthesis pathway might act to temporarily inhibit cell death process to prolong the length of infection. Given the wide variety of eukaryotic signaling pathways known to be induced by sphingolipids, however, possible biological roles of the viral sphingolipid pathway may not be necessarily linked to the control of lysis. For example, the viral sphingolipid biosynthesis may have a direct role in the highly specific cell/virus recognition processes involving membrane interactions (Allen et al., 2008). Sphingolipids have fundamental functions in host-pathogen membrane interactions (Riethmuller et al., 2006), through the organization of membrane domains (called “membrane rafts”) where different sphingolipids, cholesterol, receptors and signaling molecules are recruited to coordinate the dynamics of membrane structures. EhV virions are thought to contain a lipid membrane layer within their capsid, which in turn is enveloped by another lipid membrane upon their release from cells (Allen et al., 2008). The EhV sphingolipid

biosynthesis enzymes may thus modulate the lipid profile of the host and/or viral membranes (Han et al., 2006) to facilitate the release of newly formed virions by membrane budding (Allen et al., 2008) or to enhance the host-virus recognition process upon infection. It is also worth noting a recently revealed connection between EhV infectivity and host's life cycle. Frada *et al.* showed that EhVs are capable of infecting the diploid phase of *E. huxleyi* cells, but not the haploid phase of this alga (Frada et al., 2008). Intriguingly, sphingolipids are known to be involved in the regulation of meiotic division in higher animals (Strum et al., 1995; Yang et al., 2004). Given this new connection, it is tempting to speculate that EhVs may take advantage of their own sphingolipid pathway to control host's sexual life cycle.

Mixing of genetic pools between viruses and their hosts by horizontal transfer or through symbiotic association might have played significant roles in the evolution of viruses as well as of their hosts. The "eukaryogenesis" theory (Bell, 2001; Takemura, 2001) proposes that viruses might have been at the origin of the nucleus and that viruses have provided several "viral" features to primitive cells, which are now seen as properties of modern eukaryotic cells. Early and possibly bi-directional (Claverie, 2006) genetic exchanges between viruses and their hosts through such a process could lead to a situation where homologs of certain viral genes were found in all or most of modern eukaryotic species (Claverie, 2006; Roossinck, 2005; Villarreal, 2005), as previously proposed for instance for the eukaryotic DNA polymerase (Forterre, 2006c; Villarreal and DeFilippis, 2000). However, such early evolutionary processes prior to the divergence of major eukaryotic lineages do not readily explain the HGT(s) of the sphingolipid biosynthesis genes analyzed in this study, since the viral sphingolipid genes show a much higher sequence similarity to the *E. huxleyi* homologs than to the homologs in other eukaryotes; the HGT(s) are likely to have occurred after the separation of the lineage leading to *E. huxleyi* from other major lineages of eukaryotes.

Regarding the possible direction of the sphingolipid pathway gene transfer, the direction from viruses to the ancestors of *E. huxleyi* ("V2H direction") has an advantage in that it would easily minimize the number of required evolutionary events by invoking genome or *en bloc* gene transfer. The sphingolipid genes are dispersed in both the EhV-86 and *E. huxleyi* genomes. Transfer of all or part of a relatively small viral genome to a larger host genome (by unknown mechanism), probably followed by the elimination of the original eukaryotic homologs, would require fewer evolutionary steps than the transfer of a set of genes in the reverse direction. The V2H scenario predicts that the *E. huxleyi* homologs may retain viral homolog-like properties that are missing in other protist lineages. In this regard,

the atypical fusion of the LCB2/LCB1 domains in SPT (Fig. 2) would deserve future investigation. The domain fusion was observed not only in EhVs and *E. huxleyi* but also in other eukaryotes such as *O. tauri* and *Entamoeba* clearly placed outside the group of EhV/*E. huxleyi* sequences in the LCB2/LCB1 tree (Fig. 3a). Notably, an additional PAP2-domain was found in the SPTs of *E. huxleyi* and *O. tauri*. Whether these SPT domain organizations support the V2H scenario is unclear due to the lack of estimate for the relative frequency between this sort of domain fusion/deletion and gene transfer among viruses and eukaryotes. Sequencing of other protists including other members of the Haptophyta may provide important clues to better understand the origin of the SPT domain organizations and to assess the likeliness of the V2H hypothesis. The V2H hypothesis implies that sphingolipid biosynthesis pathway genes were present in the genomes of very ancestral viruses infecting primitive eukaryotic cells.

The direction from the ancestors of *E. huxleyi* or their relatives to the viral lineages leading to EhV ("H2V direction") is also possible and appears more parsimonious than the V2H direction in several aspects. First, these sphingolipid genes are ubiquitous in eukaryotes. The V2H scenario would require an additional and earlier HGT for a set of these genes between ancestral eukaryotes and ancestral viruses (for instance, an earlier HGT prior to the divergence of eukaryotic lineages). Furthermore, the branching positions of the EhV/*E. huxleyi* sequence groups for these genes are globally compatible with a deep phylogenetic origin of *E. huxleyi* within eukaryotes, with one exception for the viral and host Aco1-like FADs (Fig. S2B). The V2H scenario would require an additional evolutionary mechanism (or constraint) forcing the sequences of the EhV homologs to be placed near the basis of eukaryotic trees without long branches. It should be noted that branches longer for viruses than for their eukaryotic hosts are often obtained by phylogenetic tree reconstruction (Claverie, Abergel, and Ogata, 2009; Forterre and Gadelle, 2009; Moreira and Brochier-Armanet, 2008). In the H2V scenario, the viral acquisition of the sphingolipid enzyme genes was probably gradual, through multiple HGT events, rather than through a single *en bloc* transfer of multiple genes. However, it is difficult to reliably assess the relative timing of these HGT events from the current sequence data due to the large sequence divergence between the viral and host homologs. The initial acquisition of one of the genes of this metabolic pathway, for instance the enzyme SPT (the rate limiting step of this pathway), might have been sufficient for the virus to start modulating its host's life span or lipid profile, thus giving this altered virus a selective advantage on other viral strains. The later acquisitions of additional genes could have further enhanced the viral capacity to modulate the cellular

metabolism. This type of serial gene acquisition by a virus could be a possible way to increase its fitness, and might be a driving force in the Red Queen evolution of viral strains infecting the same host species.

6. Acknowledgements

We are grateful to the anonymous referees for their precious comments. The comments from one of the referees on an earlier version of the manuscript were helpful to improve the discussion about the direction of gene transfers. This study is part of the Ph.D. thesis work of AM (supported by the EuroPathoGenomics European network of excellence) and is part of the Ph.D. thesis work of AP (supported by a Marie-Curie Early Stage Training fellowship (EU-FP6) awarded to the Station Biologique Roscoff). AM has performed the reported bioinformatics analyses. AP has performed the reported experimental work. AM and AP wrote the initial draft of the manuscript. *E. huxleyi* genome sequence data were produced by the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/) in collaboration with the International *E. huxleyi* Genome Sequencing Consortium community. MJA/AP were supported by grants awarded to Dr. William H. Wilson from the Natural Environment Research Council (NERC) Environmental Genomics thematic program (ref. NE/A509332/1 and NE/D001455/1) and Oceans 2025. This work is part of the pluridisciplinary project *BOOM* (Biodiversity of Open Ocean Microcalcifiers, CdV) funded by the French *Agence Nationale de la Recherche*, grant ANR-05-BIODIV-004. This work was in part supported by the PACA-BioInfo Platform and Marseille-Nice Genopole.

7. Supplementary data - Tables

Table S1. Name and origin of the EhV and *Emiliania huxleyi* strains utilized for PCR/sequencing of 6 genes involved in the sphingolipid biosynthesis.

Strains	Year	Area*	PCR amplification†					
			SPT	LAG1	Dsd1	LPP	Aco1	3-KSR
Virus (group)								
EhV-84 (I)	1999	EC	+	+	+	+	+	NA
EhV-86 (I)	1999	EC	+	+	+	+	+	NA
EhV-88 (I)	1999	EC	+	+	+	+	+	NA
EhV-V1 (I)	2003	RN	+	+	+	+	+	NA
EhV-V2 (I)	2003	RN	+	+	+	+	+	NA
EhV-163 (I)	2000	RN	+	+	+	+	+	NA
EhV-201 (II)	2001	EC	+	+	+	+	+	NA
EhV-202 (II)	2001	EC	+	+	+	+	+	NA
EhV-205 (II)	2001	EC	+	+	+	+	+	NA
EhV-207 (II)	2001	EC	+	+	+	+	+	NA
EhV-208 (II)	2001	EC	+	+	+	+	+	NA
E. huxleyi								
RCC1242 (CCMP1516)	1991	Pacific Ocean / Offshore Peru	+	+	+	+	+	+
RCC1215 (TW1)	1998	MSS	+	+	+	+	+	+
RCC 1259 (CCMP374)	1989	Gulf of Maine	+	+	+	+	+	+
RCC1255 (CCMP370)	1959	Oslo Fjord	+	-	+	+	+	+
RCC1235 (VF20)	2006	Mediterranean Sea (France)	+	+	+	+	+	-
RCC1253 (OS-2)	2006	Sea of Japan	+	+	+	+	-	-

* EC: Western English Channel (off the coast of Plymouth, UK); RN: Raunefjorden (Western Norway during a mesocosm experiment); MSS: Mediterranean Sea (Spain). See Schroeder *et al.* (2002) for more detail on the EhV isolates.
† + for positive, and – for negative amplification.

Table S2. List of the different primers used for this study.

Target Gene	Oligo Name	Direction	Position	Oligo Sequence	Product Size
Emiliania huxleyi					
Serine palmitoyltranferase	CP_EX_1F	Forward	1023	CGACGAGGTCTTCAAGCAG	668
Serine palmitoyltranferase	CP_EX_1R	Reverse	1690	AGCCGACGTAGAGGTCAATC	
3 ketosphinganine reductase	CP_EX_2F	Forward	338	ACGTGCAGGTCGTCTTTCTC	574
3 ketosphinganine reductase	CP_EX_2R	Reverse	911	ATCTTGTAGTCGGGCGTGAG	
Dihydrocermaide synthase	CP_EX_3F	Forward	356	TGCATCAGCTGGTGTACCTC	508
Dihydrocermaide synthase	CP_EX_3R	Reverse	863	TCCAGCTCCTTGCCACTTAG	
Dihydrocermaide desaturase DSD1 like	CP_EX_4F	Forward	43	GAGGTGAAGCAGCTCTTTGG	681
Dihydrocermaide desaturase DSD1 like	CP_EX_4R	Reverse	723	CCAGTTGTACGAGCTTGCAG	
Fatty acid desaturase Aco1 like	CP_EX_5F	Forward	4	AATCTGCTGCTGACGATGG	508
Fatty acid desaturase Aco1 like	CP_EX_5R	Reverse	511	GGTTTCGGATGTTGAACCAC	
Sphingosine 1 phosphate phosphatase	CP_EX_6F	Forward	156	CATCATCAACATCGCAGGAG	517
Sphingosine 1 phosphate phosphatase	CP_EX_6R	Reverse	672	GAATGCGCCAGCCACTAC	
EhV					
Serine palmitoytransferase	CP_EhV_1F	Forward	1475	ACACCGTTTCGGTGAAAAAG	508
Serine palmitoytransferase	CP_EhV_1R	Reverse	1982	CGCAATGCGATAATACATGG	
Dihydroceramide synthase	CP_EhV_2F	Forward	337	GCCGGGTTTTATATTCACCA	510
Dihydroceramide synthase	CP_EhV_2R	Reverse	846	CGGATTTCCTGCAATGACTT	
Dihydroceramide destaurase DSD1-like	CP_EhV_3F	Forward	51	GCATGCTGAACGTAAGCAAA	571
Dihydroceramide destaurase DSD1-like	CP_EhV_3R	Reverse	621	AAATGGGGCGATACCATACA	
Fatty acid destaurase Aco1-like	CP_EhV_4F	Forward	170	CGCATTCCGCATATAAAACA	540
Fatty acid destaurase Aco1-like	CP_EhV_4R	Reverse	709	TACCAAGCGATGGCCTTACT	
Sphingosine 1 phosphate phosphatase	CP_EhV_5F	Forward	150	TGATCATCCGCTGATTGAAG	531
Sphingosine 1 phosphate phosphatase	CP_EhV_5R	Reverse	680	AACCCGCCAATTAAAAATCC	

9. Supplementary data – Figures

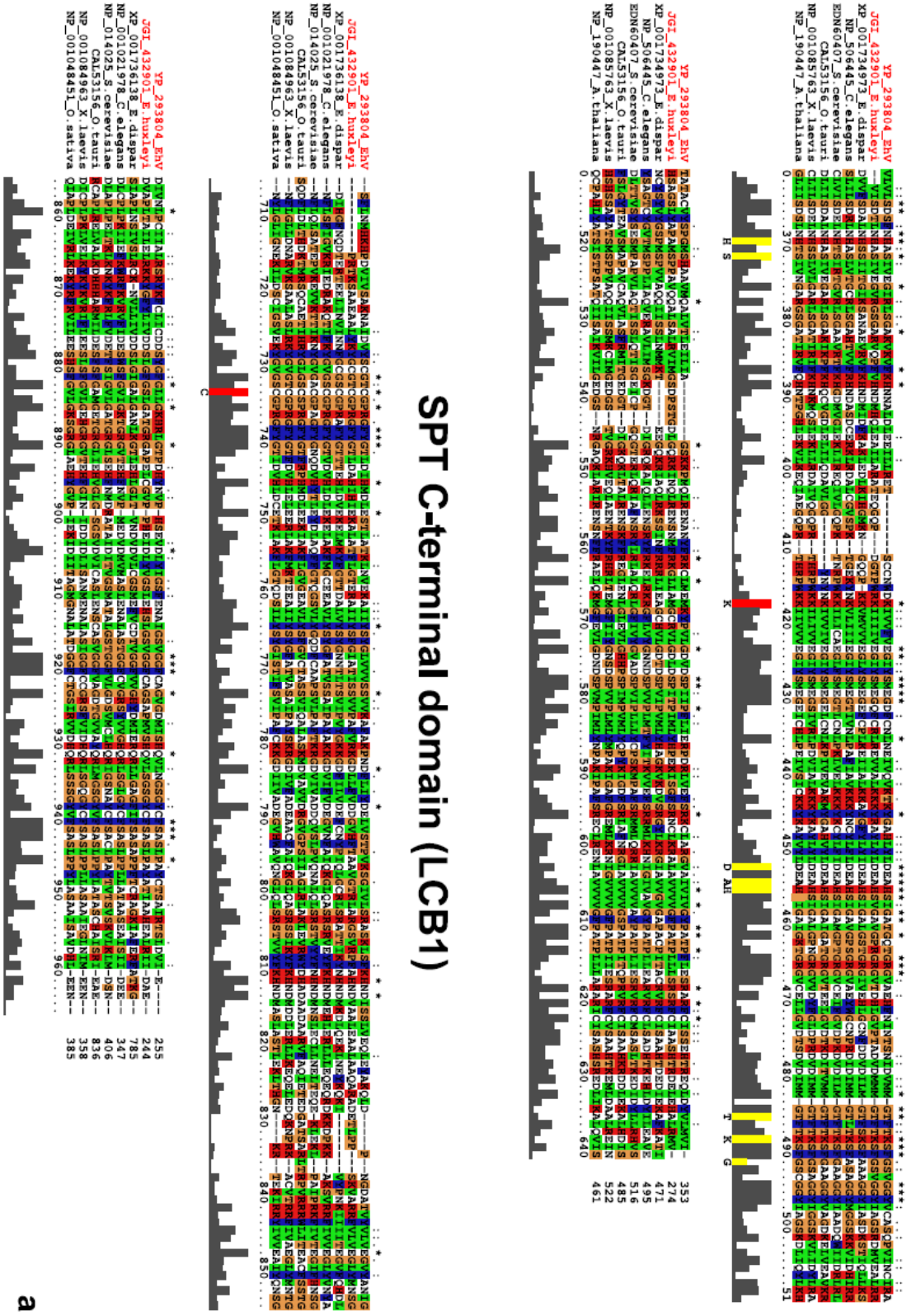

Figure S1. Multiple sequence alignments for the sphingolipid biosynthesis enzymes (explanation and continuation on next page).

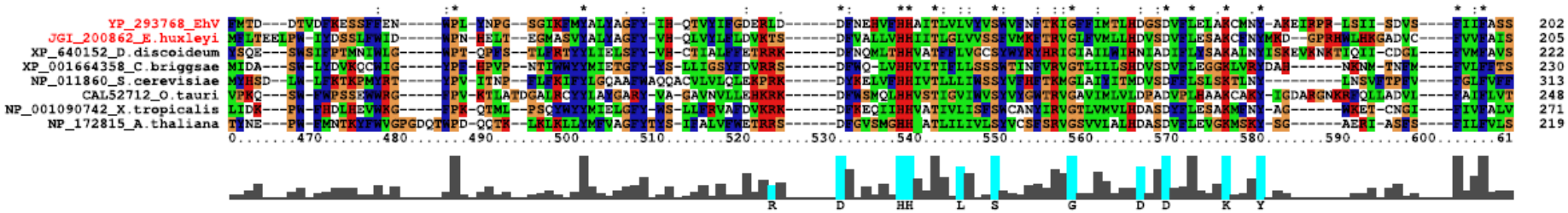

Figure S1. (cont.) The positions of previously reported catalytic residues, residues located in the active sites, and other conserved residues are respectively highlighted in red, yellow and cyan colored bars. Consensus sequences for those conserved residues are indicated under the bars.

b

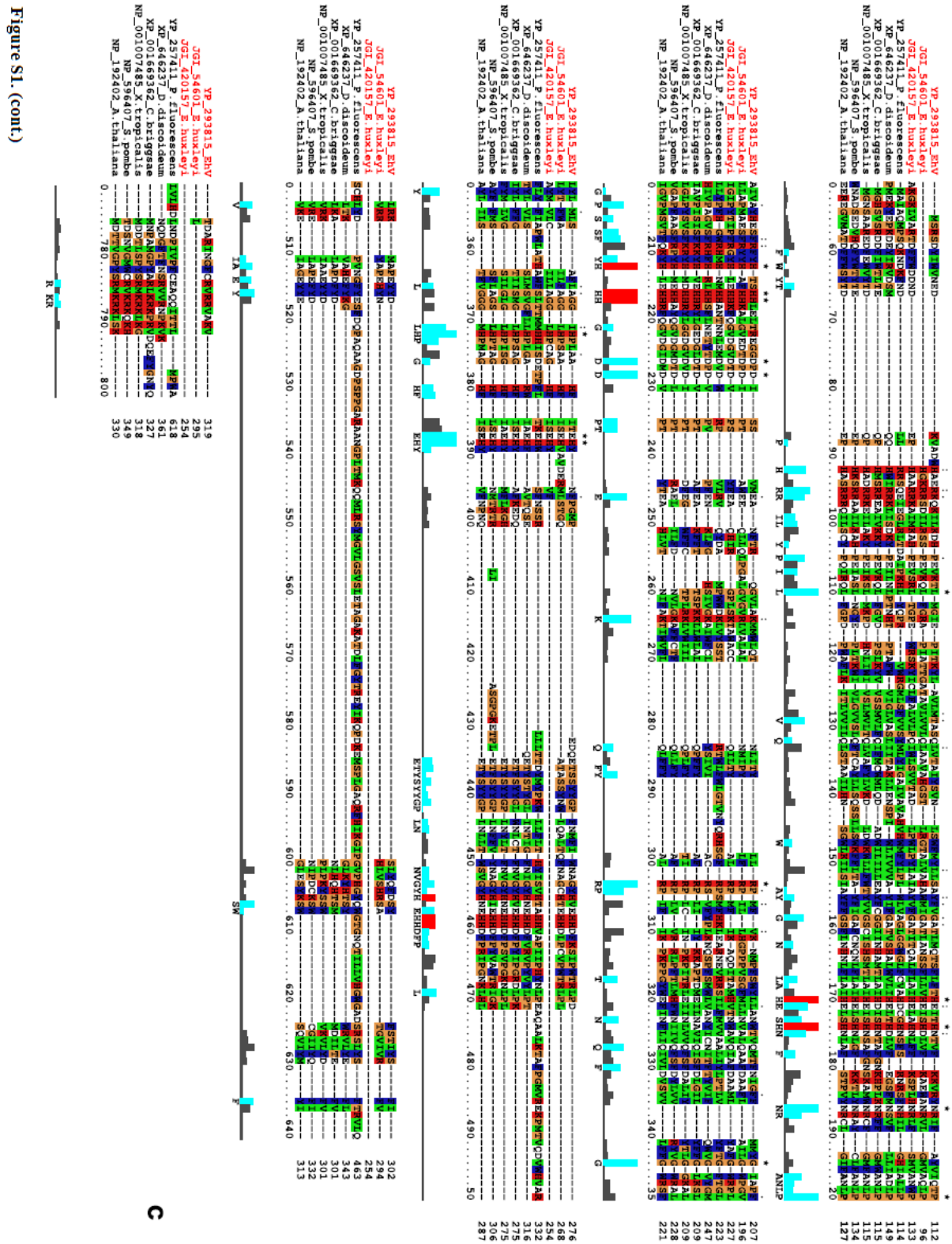

Figure S1. (cont.)

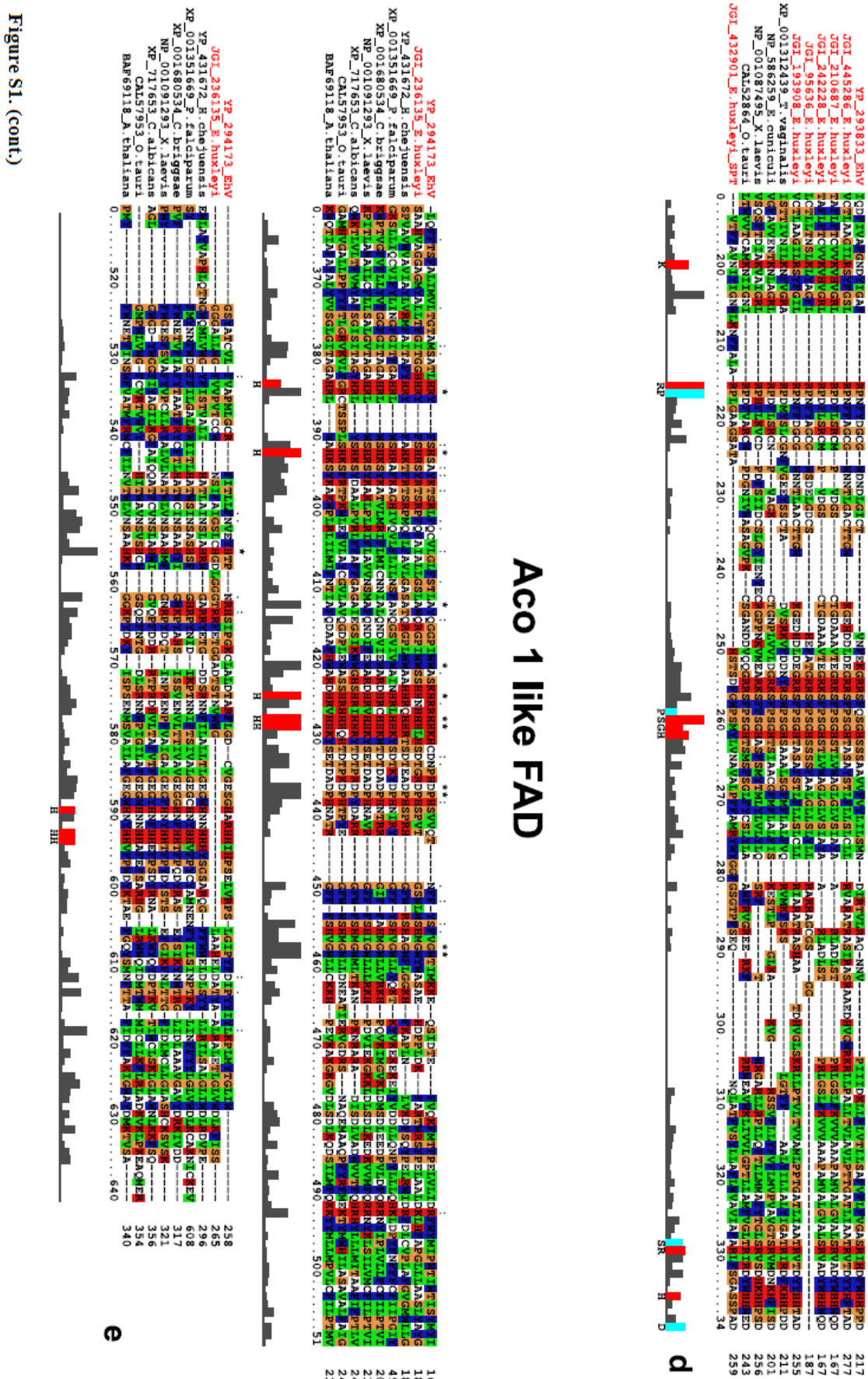

Figure S1. (cont.)

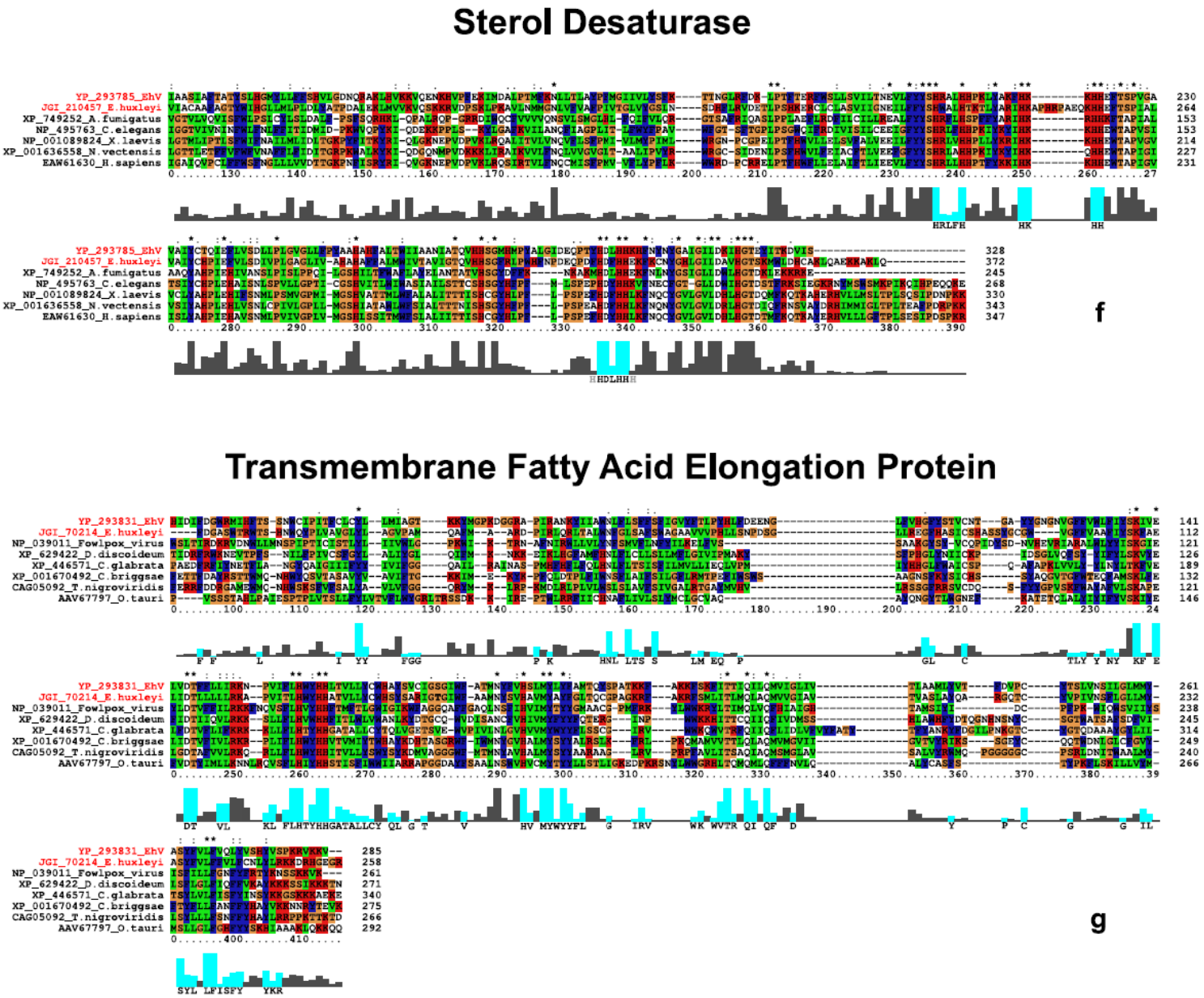

Figure S1. (cont.)

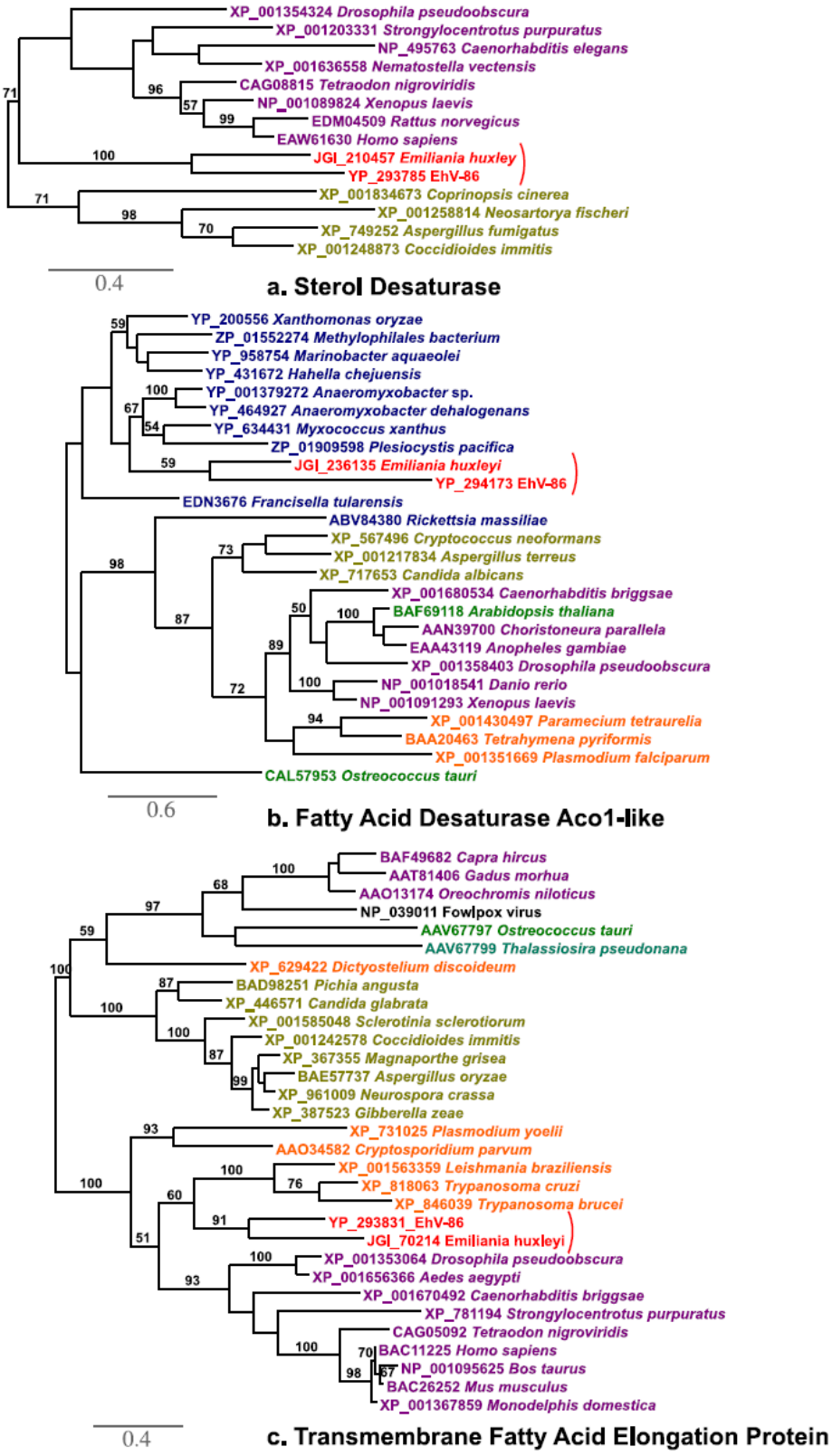

Figure S2. Maximum likelihood phylogenetic trees based on the amino acid sequences of three enzymes predicted to be involved in the sphingolipid biosynthesis pathway. (a) Sterol desaturase. (b) Aco1-like fatty acid desaturase. (c) Transmembrane fatty acid elongation protein. The number of substitutions per site is indicated under the scale bar. These trees are unrooted *per se*, although we have arbitrarily chosen a root (mostly by mid-point rooting) for each tree only for visualization purpose.

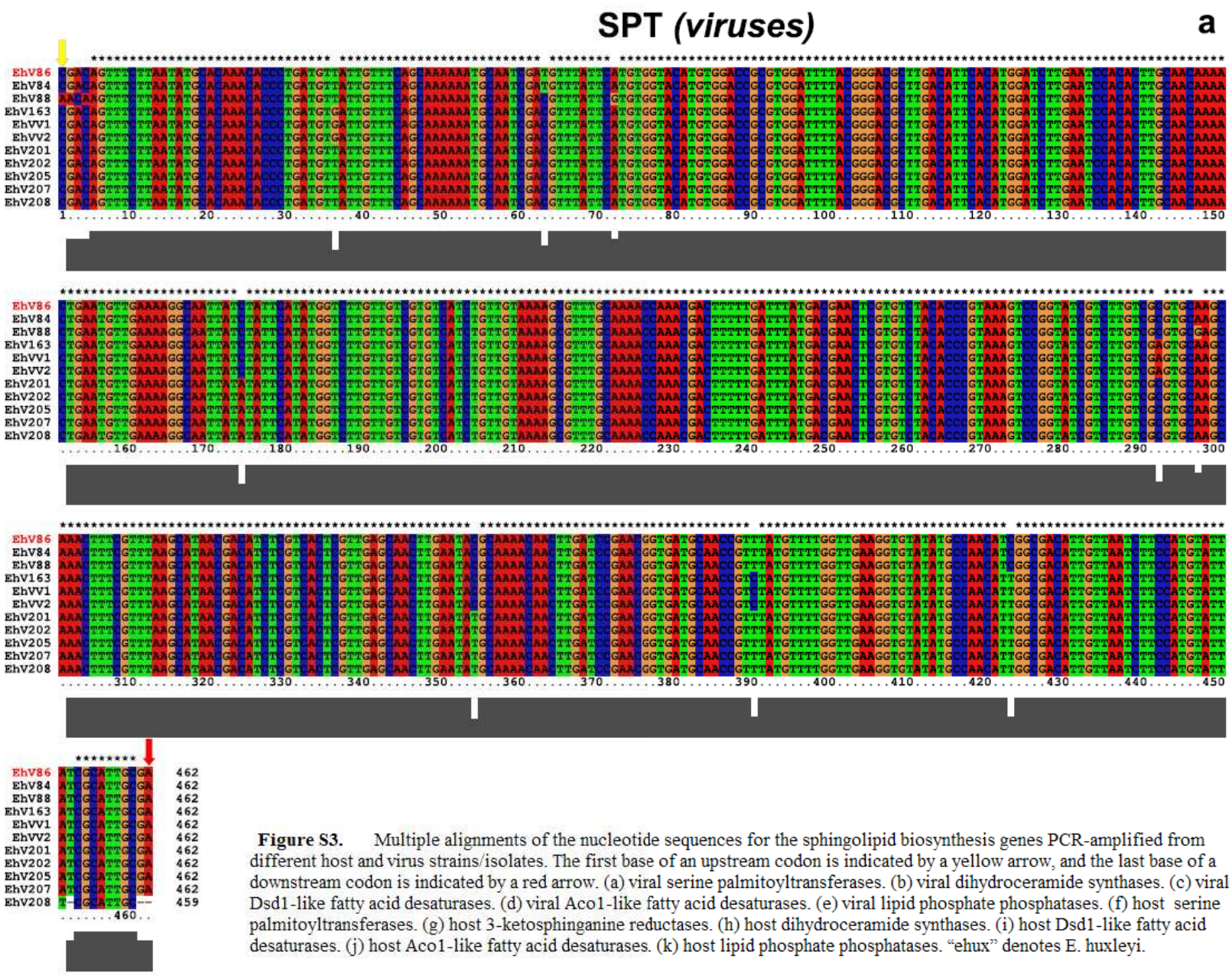

Figure S3. Multiple alignments of the nucleotide sequences for the sphingolipid biosynthesis genes PCR-amplified from different host and virus strains/isolates. The first base of an upstream codon is indicated by a yellow arrow, and the last base of a downstream codon is indicated by a red arrow. (a) viral serine palmitoyltransferases. (b) viral dihydroceramide synthases. (c) viral Dsd1-like fatty acid desaturases. (d) viral Aco1-like fatty acid desaturases. (e) viral lipid phosphate phosphatases. (f) host serine palmitoyltransferases. (g) host 3-ketosphinganine reductases. (h) host dihydroceramide synthases. (i) host Dsd1-like fatty acid desaturases. (j) host Aco1-like fatty acid desaturases. (k) host lipid phosphate phosphatases. "ehux" denotes E. huxleyi.

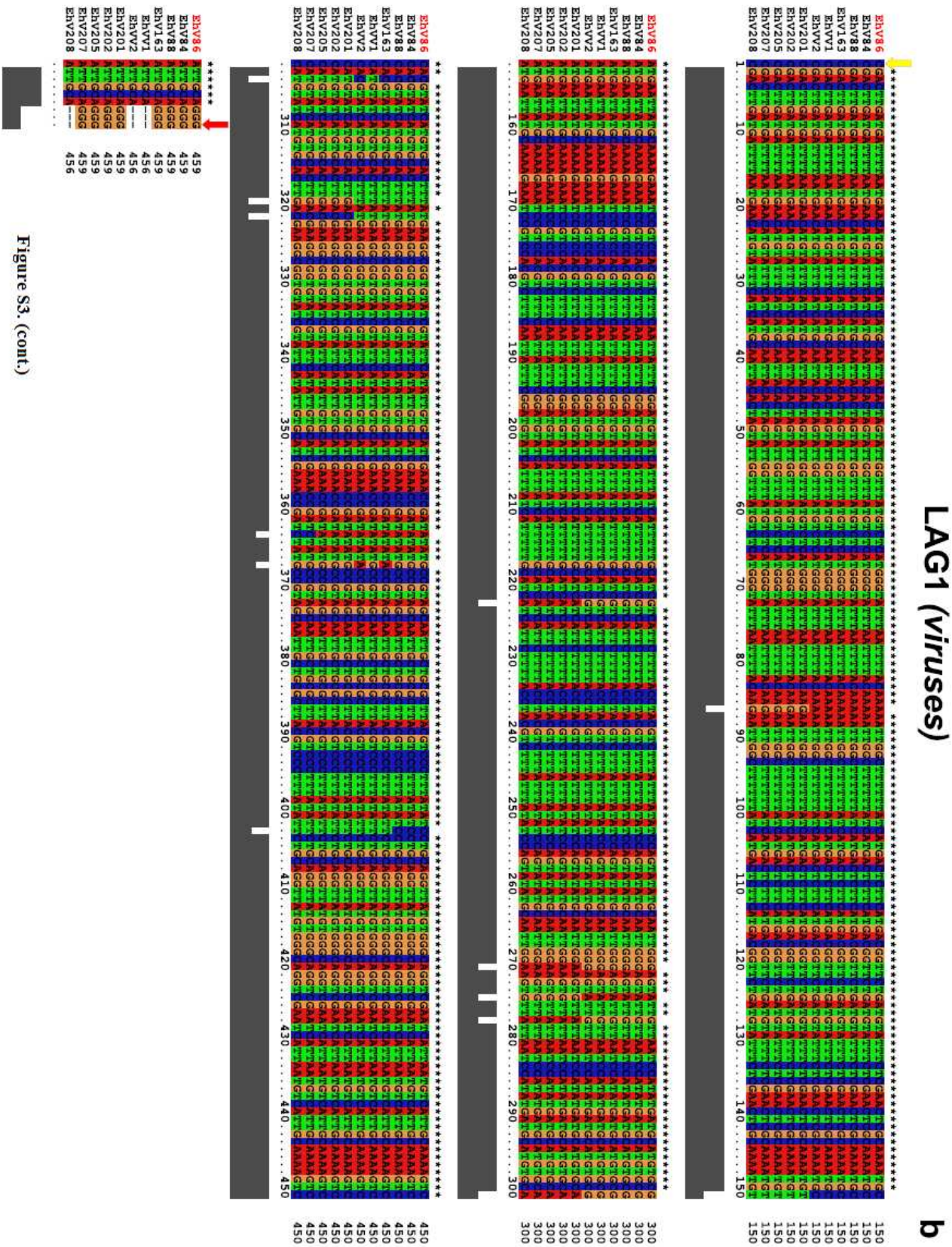

Figure S3. (cont.)

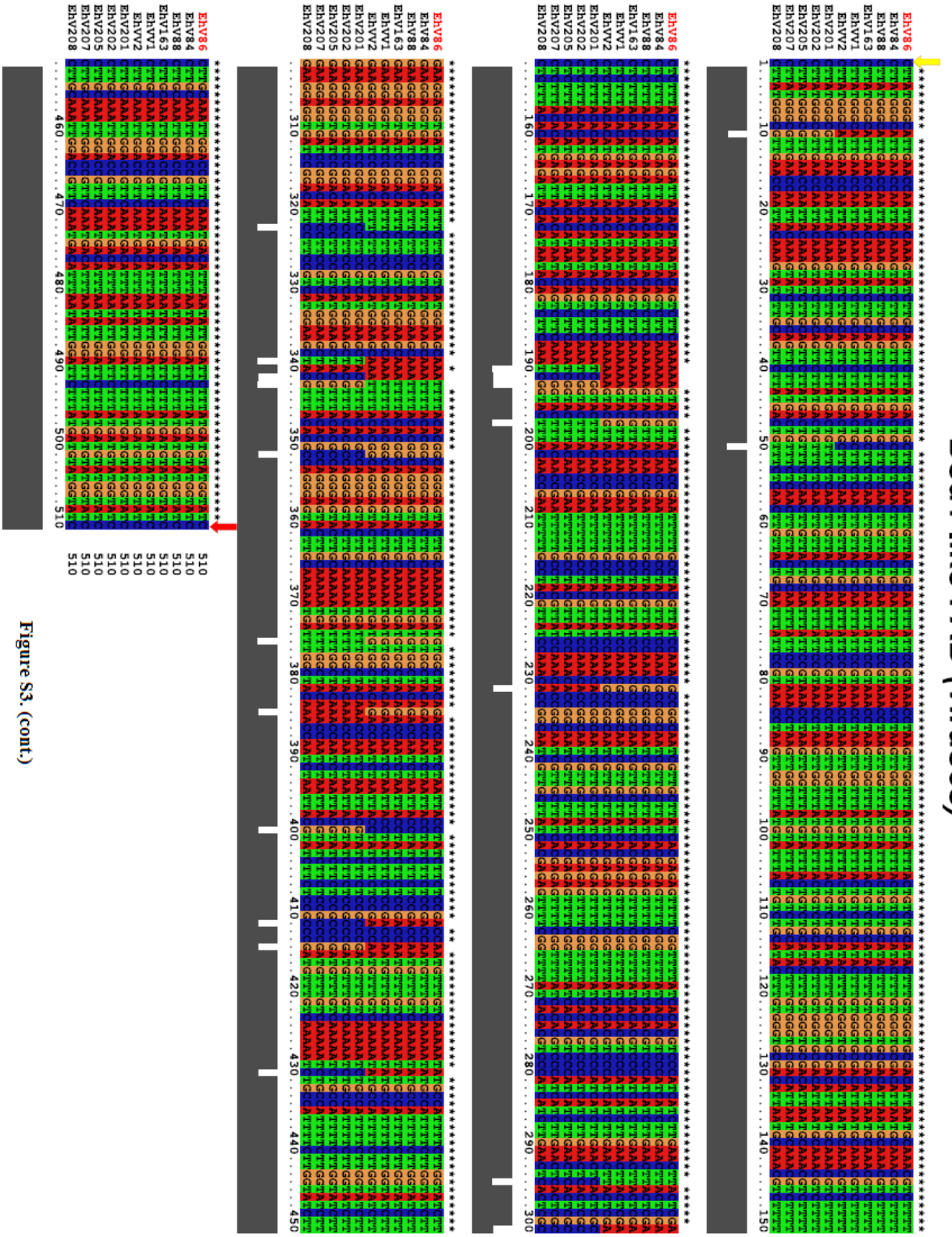

Figure S3. (cont.)

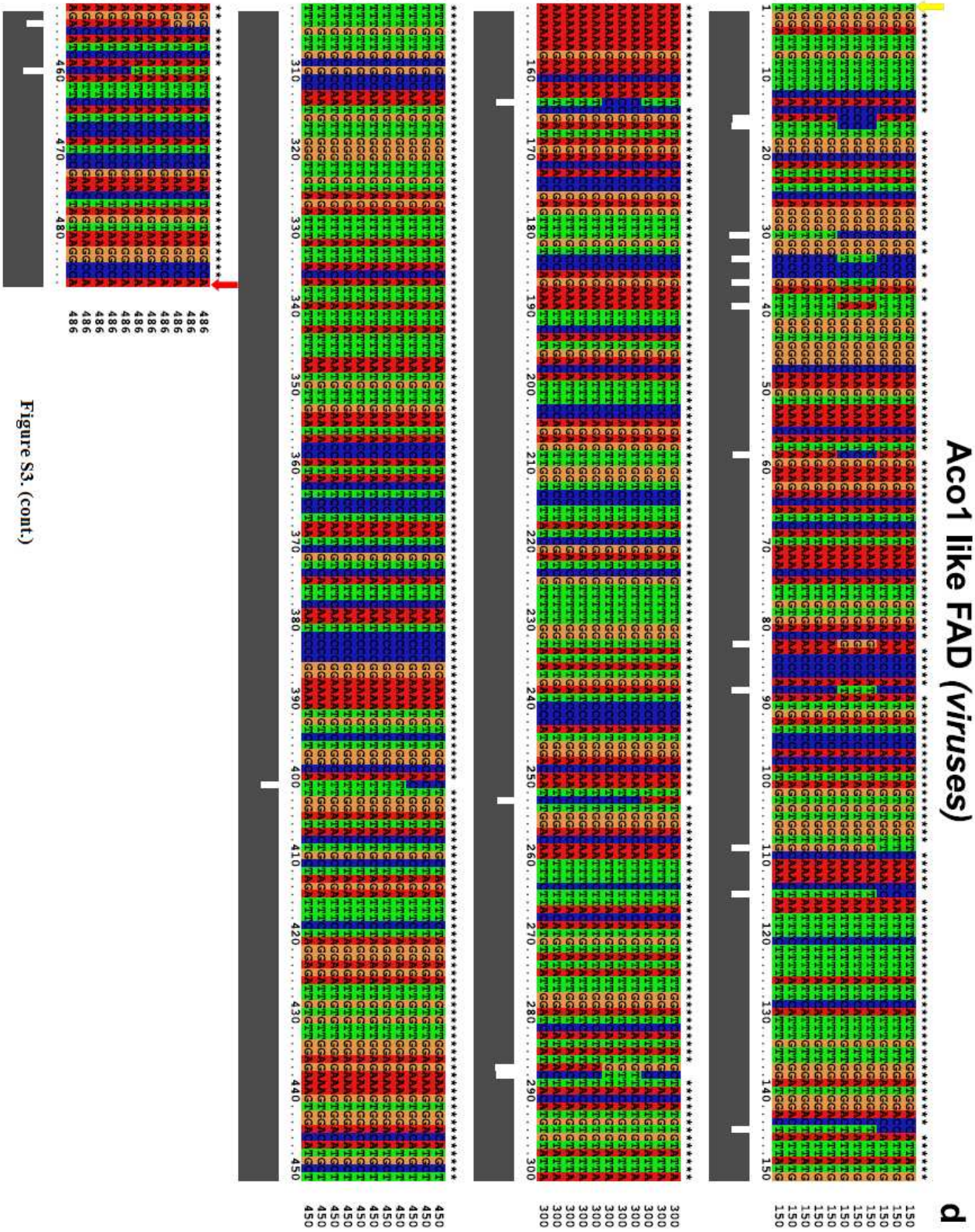

Figure S3. (cont.)

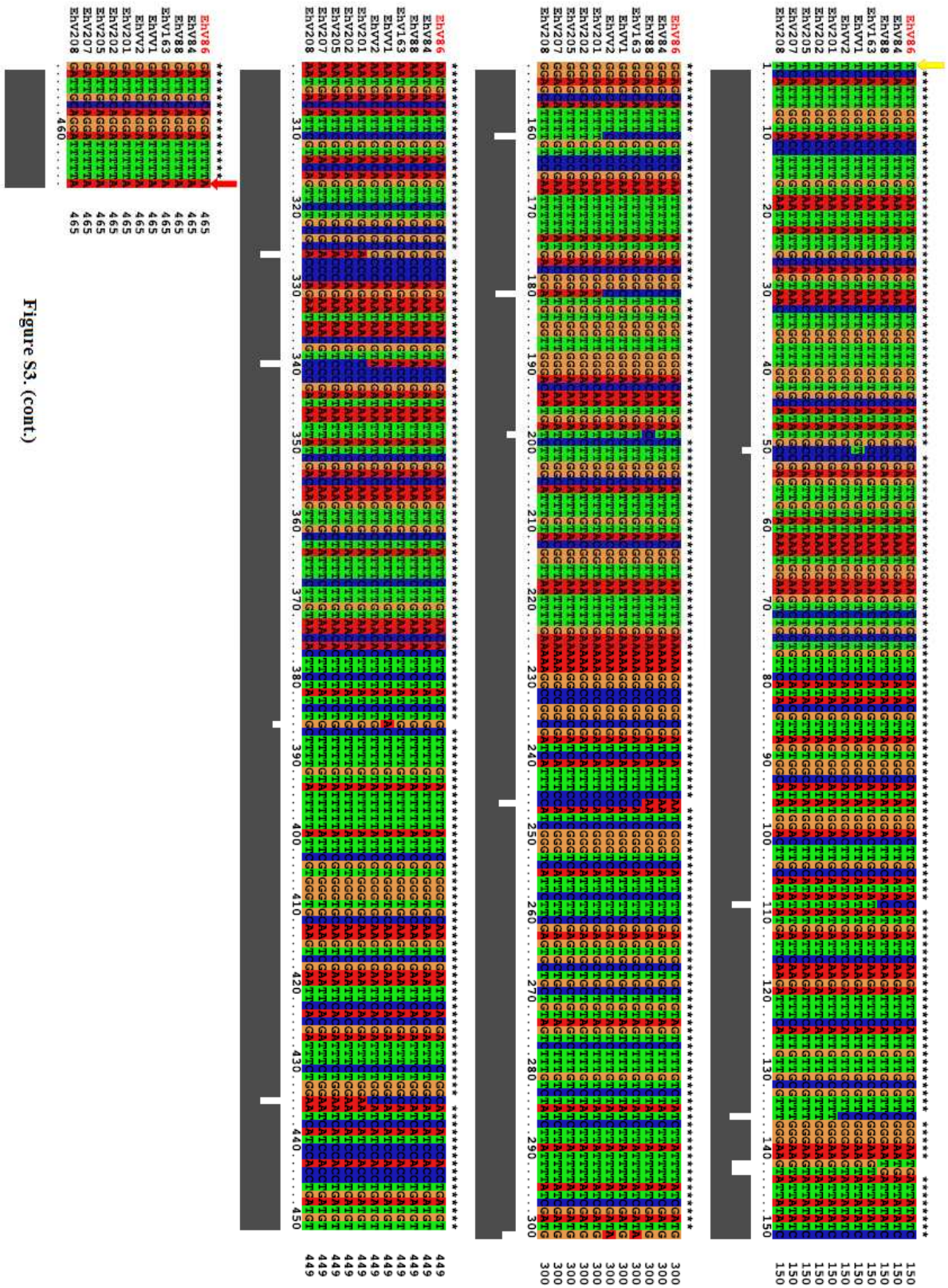

Figure S3. (cont.)

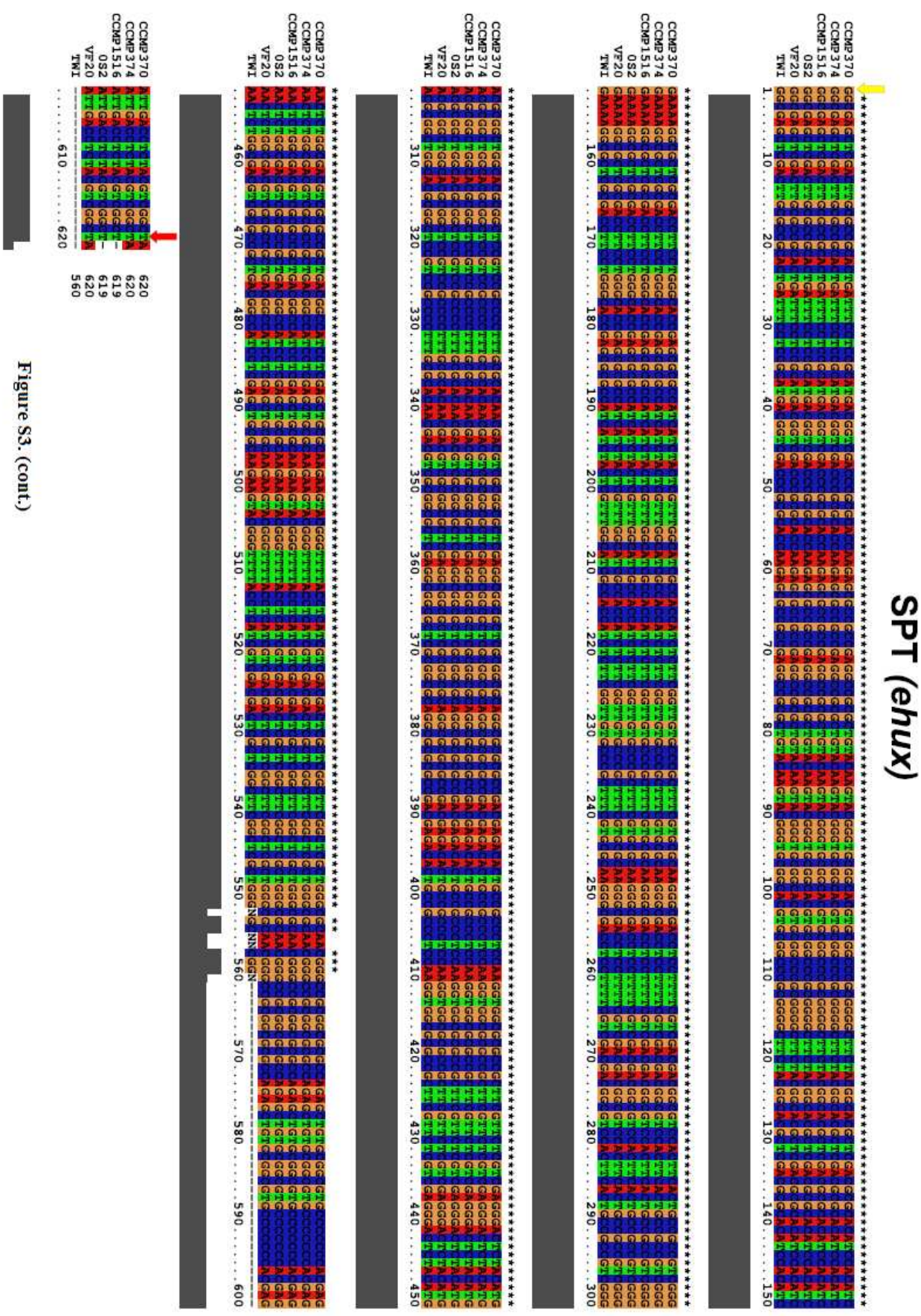

Figure S3. (cont.)

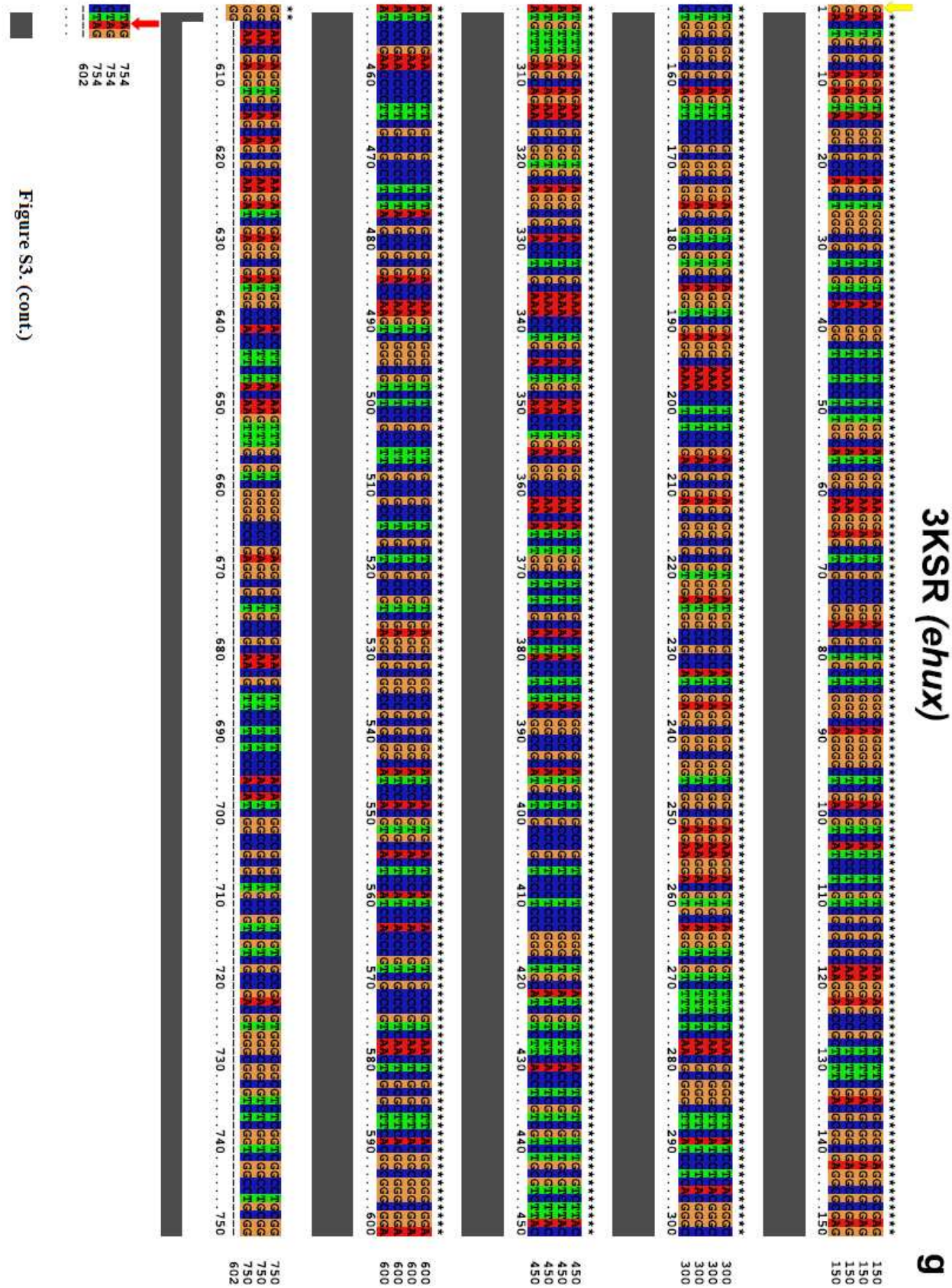

Figure S3. (cont.)

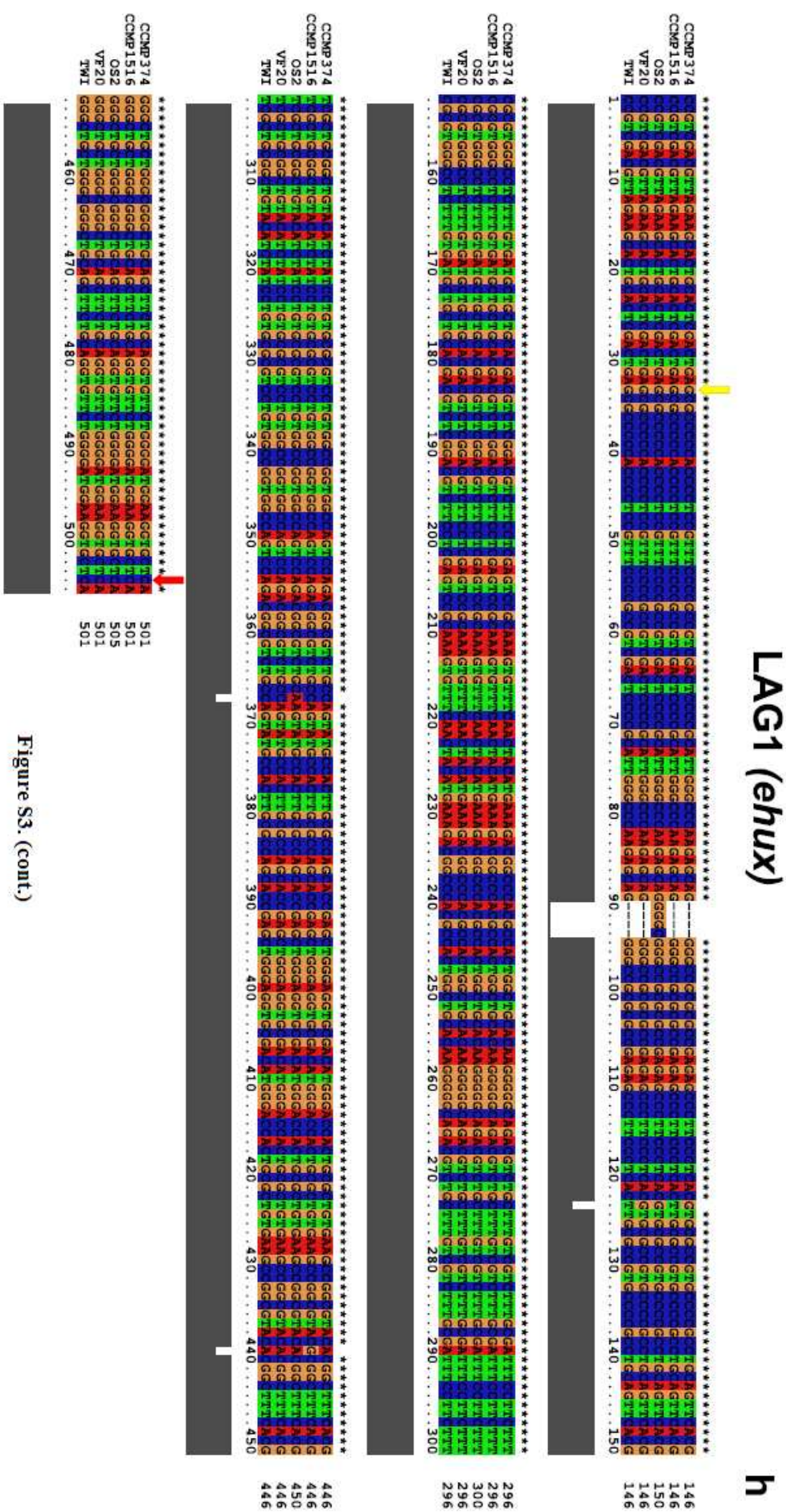

Figure S3. (cont.)

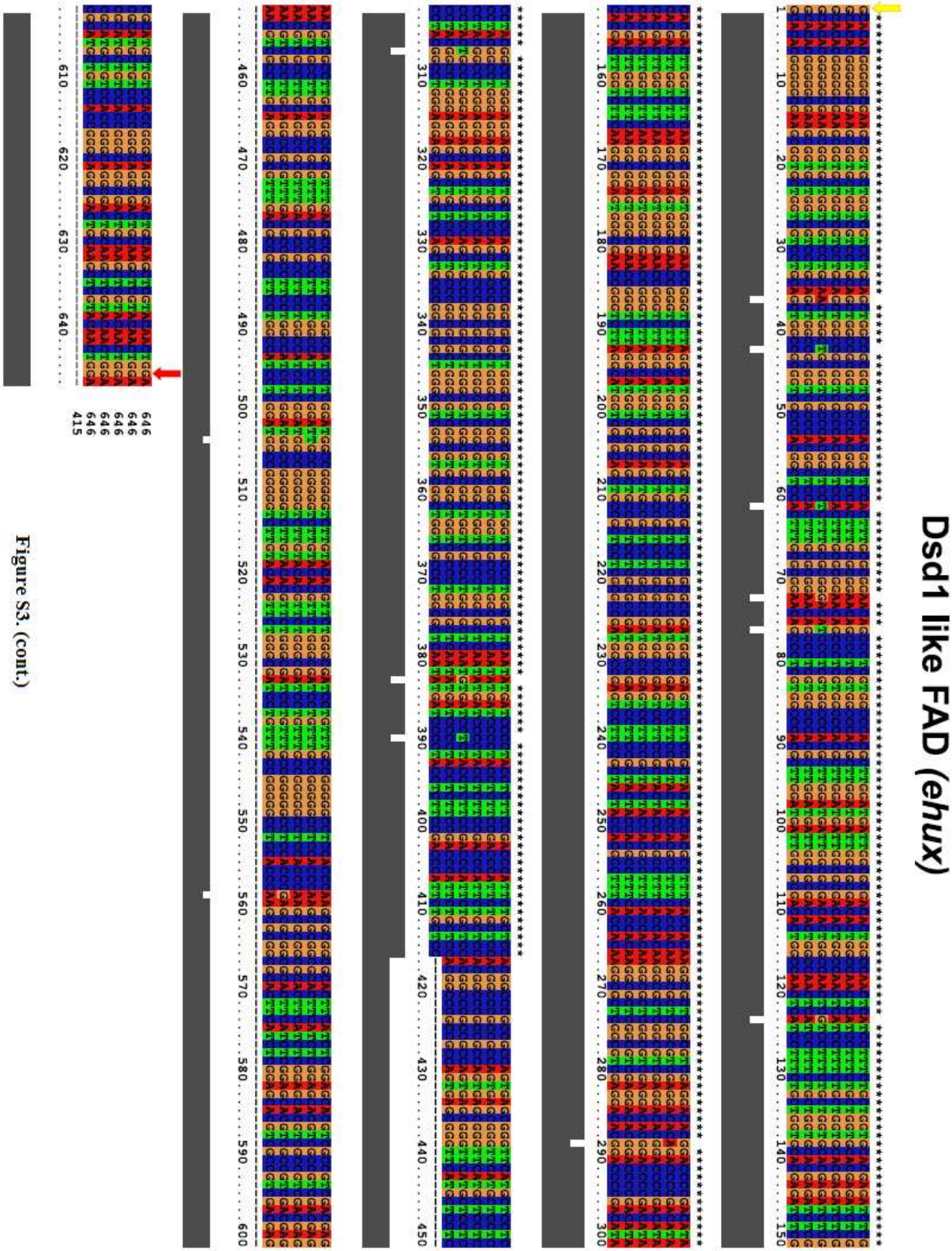

Figure S3. (cont.)

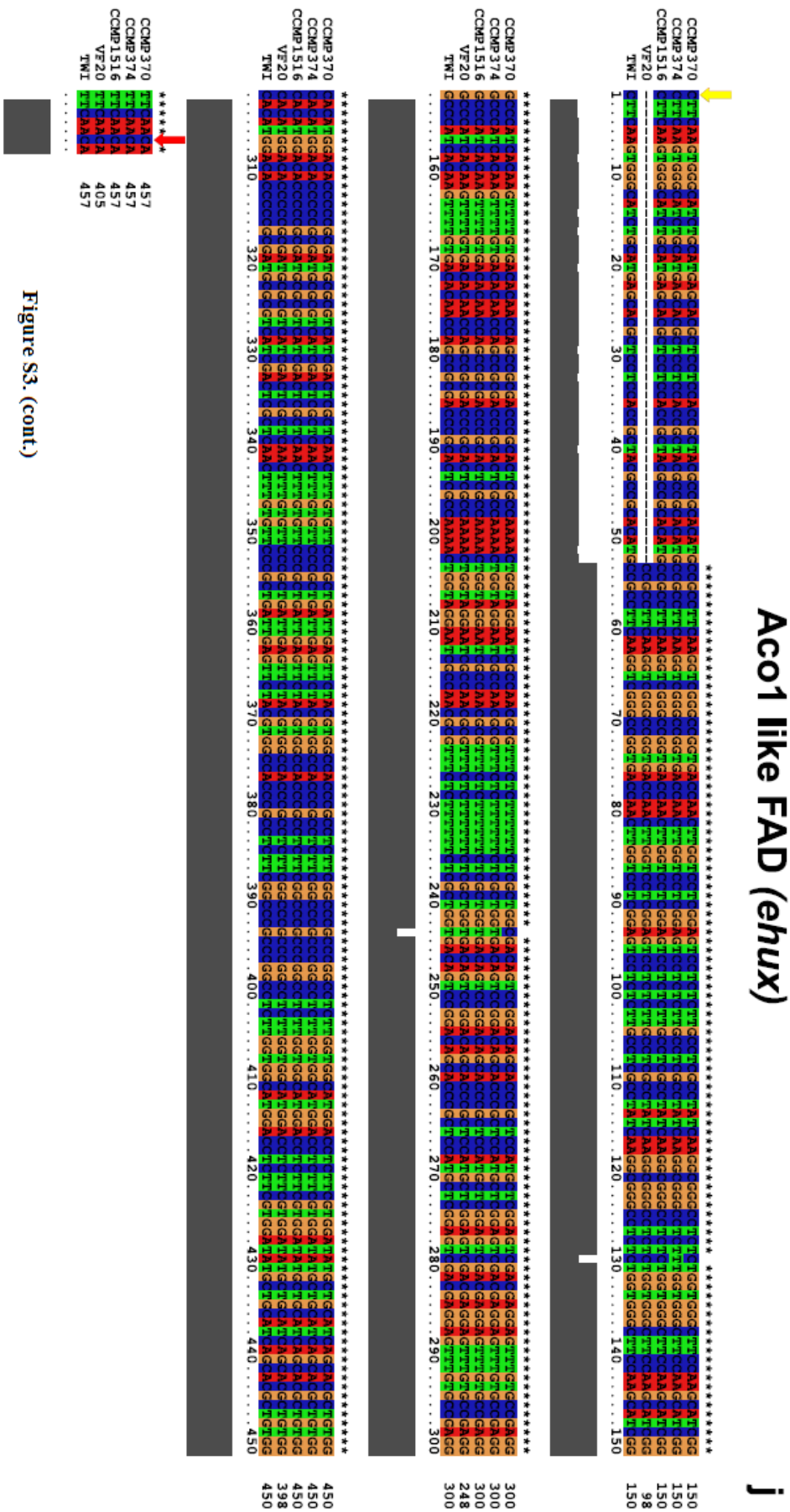

Figure S3. (cont.)

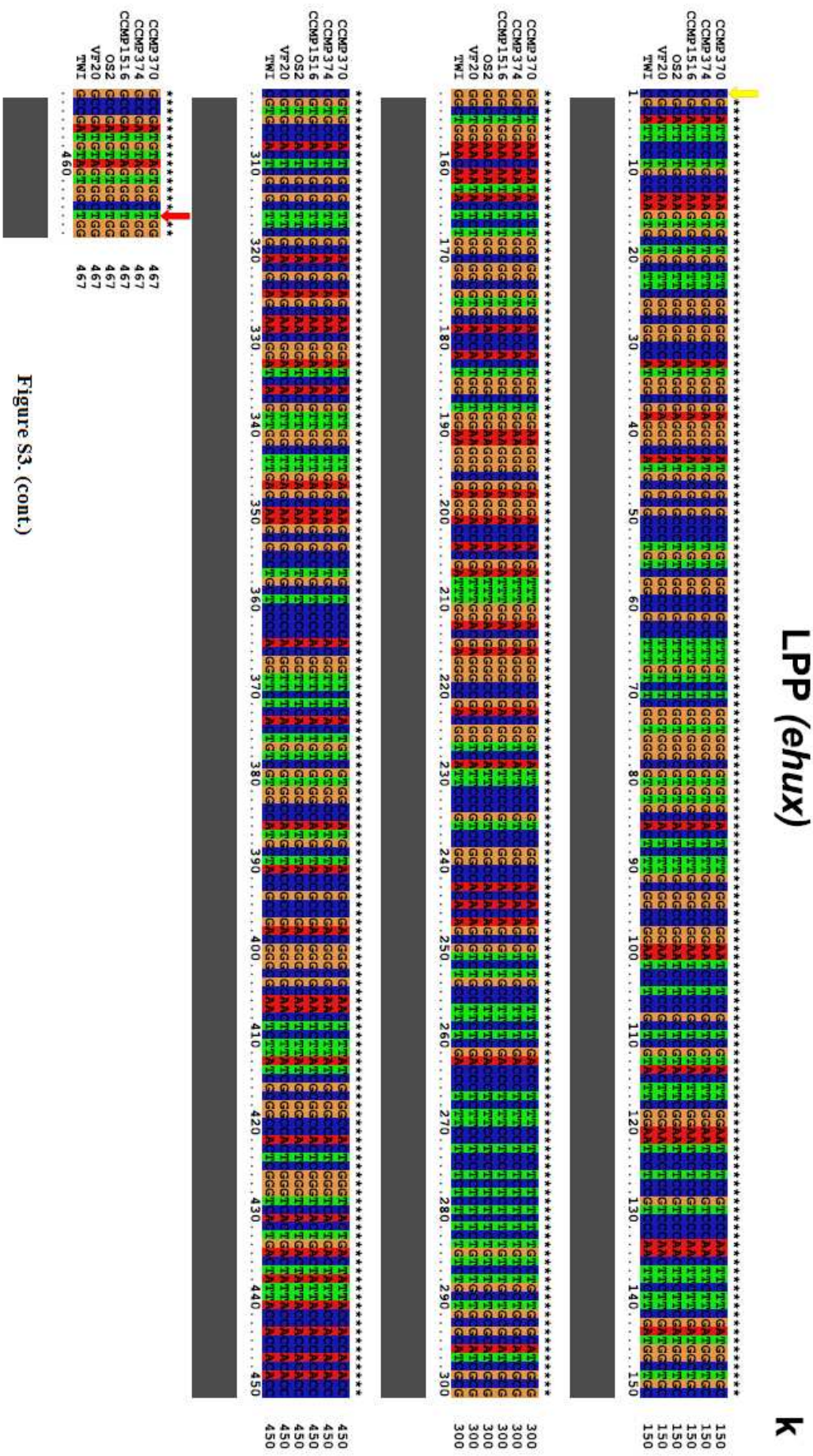

Figure S3. (cont.)

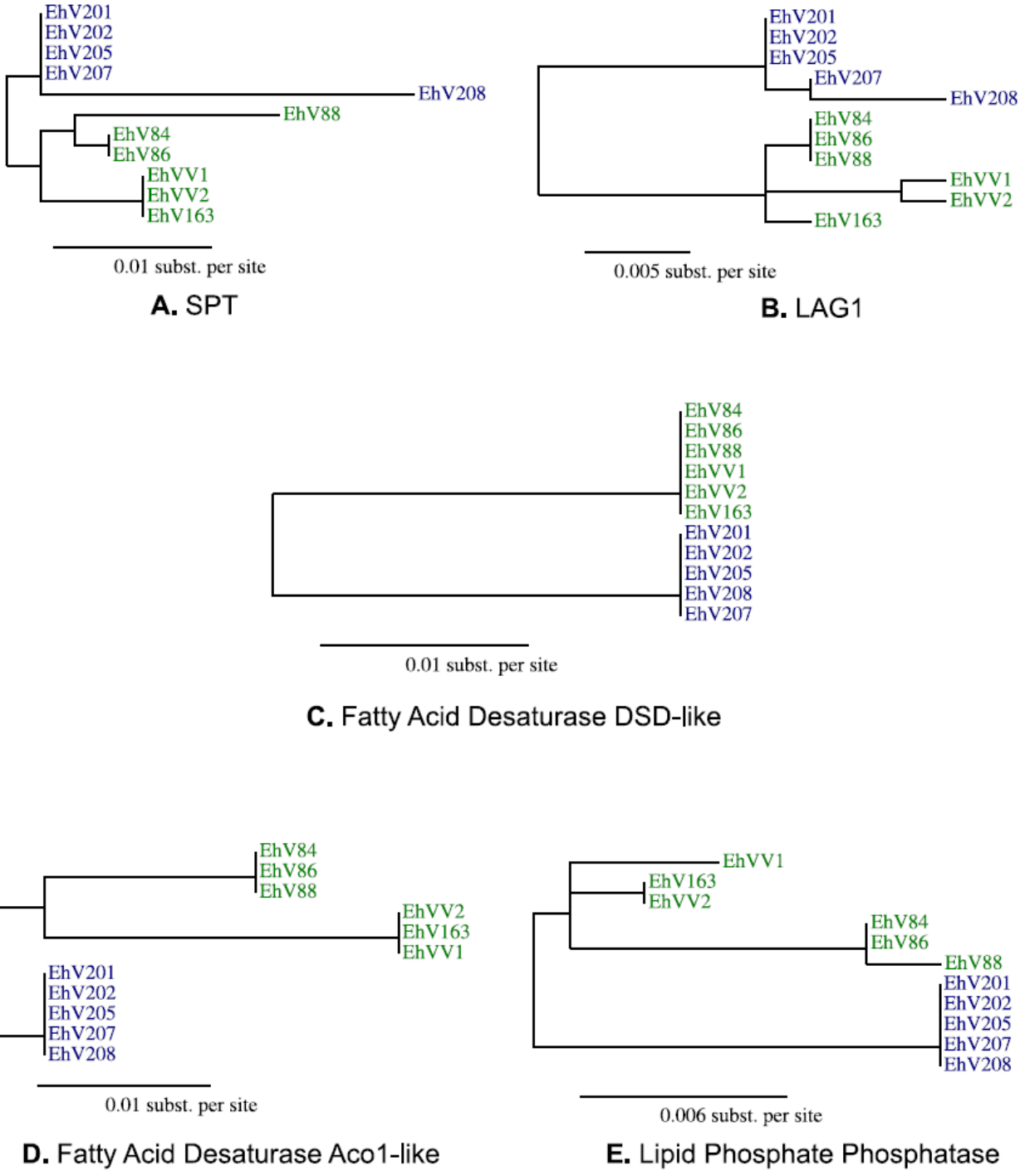

Figure S4. Maximum parsimony trees based on the nucleotide sequences of the sphingolipid biosynthesis genes amplified from eleven viral stain/isolates. Viral strains belonging to group I are indicated in green, and those belonging to group II are indicated in blue. (A) Serine palmitoyltransferase (SPT). (B) Dihydroceramide synthase (LAG1). (C) Dsd1-like fatty acid desaturase. (D) Aco1-like fatty acid desaturase. (E) Lipid phosphate phosphatase (LPP).

Chapter 4.

Host-virus shift of the sphingolipid pathway along an *Emiliania huxleyi* bloom: survival of the fattest*

António Pagarete[1], Michael J. Allen[2], William H. Wilson[2,3], Susan A. Kimmance[2], and Colomban de Vargas[1]

1. Equipe EPPO-Evolution du Plancton et PaléoOcéans, CNRS-UMR7144, Université Pierre et Marie Curie, Station Biologique, FR-29682 Roscoff, France.

2. Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, PL1 3DH, UK.

3. Department of Biological Science, California State University San Marcos, CA, 92096-0001, USA.

1. Summary

The interactions between viruses and phytoplankton play a key role in shaping the ecological and evolutionary dynamics of oceanic ecosystems. One of the most fascinating examples of horizontal gene transfer between a eukaryotic host and its virus is a *de novo* sphingolipid biosynthesis pathway (SBP) found in the genomes of both *Emiliania huxleyi* and its coccolithovirus EhV-86. Here, we focus on a natural *E. huxleyi*/coccolithovirus system off the coast of Norway and investigate the dynamics of host and virus homologous gene expression for two of the most important sphingolipid biosynthesis enzymes, serine palmitoyl transferase (SPT) and dihydroceramide desaturase (DCD). Transcriptional dynamics display three defined stages along *E. huxleyi* bloom formation and decline, with the coccolithovirus transcripts taking over and controlling the SBP in stages 2 and 3. The observed patterns fit the hypothesis according to which viral sphingolipids are involved in the timing and physical processes of virion release from the host cells. This study provides a unique insight into the transcriptional interplay of homologous metabolic pathways between virus and host during temporal progression of oceanic *E. huxleyi* blooms.

** Published in Environmental Microbiology 11(11): 2840-2848, 2009*

2. Introduction

Viruses that infect phytoplankton play a key role in shaping the evolution and dynamics of the oceanic micro-scale ecosystem (Fuhrman, 1999; Sandaa, 2008; Suttle, 2005b). Several studies have highlighted the role of viruses as major triggers for high phytoplankton turnover rates, a process termed the viral shunt (Wilhelm and Suttle, 1999). The interplay of viruses with their host communities is complex, and may assume different forms. Traditionally regarded as simple agents of mortality and catalysts for nutrient transformation (Suttle, 2005b; Weinbauer and Rassoulzadegan, 2004), viruses are now also believed to play a fundamental role in controlling the biodiversity and functioning of their host communities (Frada et al., 2008; Thingstad, 2000; Thingstad and Lignell, 1997). More recently, studies have revealed that host-virus gene transfers significantly contribute to the emergence of novel viral infection strategies in cyanobacteria/phage systems (Clokie et al., 2006; Lindell et al., 2005; Lindell et al., 2004; Millard et al., 2004; Yoshida et al., 2008).

One of the rare and most fascinating examples of horizontal gene transfer between a eukaryotic host and its virus is the almost complete *de novo* sphingolipid biosynthesis pathway (SBP) found in the genomes of *Emiliania huxleyi* (Haptophyta) and its virus EhV-86 (Coccolithovirus; Phycodnaviridae) (Monier et al., 2009; Wilson et al., 2005b). *Emiliania huxleyi* (Lohmann) Hay et Mohler, a single celled phytoplankton, is the most abundant and ubiquitous coccolithophore in today's oceans (Brown and Yoder, 1994). Despite its very young age (estimated at 260,000 years) (Thierstein, Geitzenauer, and Molfino, 1977), *E. huxleyi* is a key player in global ecology, in particular marine primary productivity and the biogeochemical cycles of carbon and sulphur (Burkill et al., 2002; Westbroek et al., 1993). Recently, it became clear that specific viruses are closely linked to the sudden crashes of the vast coastal and mid oceanic *E. huxleyi* blooms (Bratbak, Egge, and Heldal, 1993; Bratbak et al., 1996; Castberg et al., 2002; Jacquet et al., 2002; Schroeder et al., 2003; Wilson et al., 2002).

The eukaryotic *de novo* SBP leads to the production of ceramide (Merrill, 2002), a metabolite implicated in the regulation of critical cellular processes such as senescence, differentiation, apoptosis, and cell cycle arrest (for a review see Hannun and Obeid, 2008 and references therein). Allen et al. (2006) demonstrated that the viral SBP genes are expressed during infection under laboratory conditions, and the functionality of the viral serine palmitoyl transferase (SPT), the first and rate-limiting enzyme of sphingolipid biosynthesis, was confirmed using an over-expression system (Han et al., 2006).

On the other hand, numerous studies on eukaryotic organisms have demonstrated the involvement of viruses in the control of host Programmed Cell Death (PCD), either repressing or triggering its activation (see Hay and Kannourakis, 2002 for a clear and extended review). Although counter intuitive, PCD has the potential to be an advantageous mechanism to maximise virion spread. In such case, the apoptotic demise of a cell results in the formation of small membrane-bound entities known as apoptotic bodies. Apoptotic bodies are commonly consumed by the phagocytic action of neighbouring cells, thus providing a route for the dissemination of the virus (Teodoro and Branton, 1997). Recently, a study showing the involvement of coccolithovirus in metacaspase activation further speculated that this virus could be using its own SBP pathway to trigger the host's PCD (Bidle et al., 2007).

Here, we focus on a natural *E. huxleyi*/coccolithovirus system and investigate the dynamics of host and virus gene expression for two of the most important SBP enzymes, serine palmitoyl transferase (SPT) and dihydroceramide desaturase (DCD). SPT catalyses the first and rate-limiting step of the pathway, while DCD leads to direct production of ceramide (Merrill, 2002; Perry et al., 2000). A mesocosm environment was used to monitor, for the first time, the transcription levels of both host and virus homologous SBP genes in a marine microbial community during temporal progression and decline of a naturally occurring *E. huxleyi* bloom.

3. Experimental Procedures

Set-up of the mesocosm experiment

The *E. huxleyi*-induced blooms were conducted in the Raunefjorden, Western Norway coast, at the Marine Biological Field Station, for 17 days (5th to the 21st of June 2008). Six mesocosm bags (11 m^3 each) were filled with unfiltered Fjord water pumped from 10m depth adjacent to the raft. Homogeneous water masses within the enclosures were ensured by pumping water from the bottom of the bag to the surface. The 6 enclosures were divided in 2 treatment groups allowing triplication of each treatment: phosphate replete (enclosures #1, #3 and #5) and phosphate deplete (enclosures #2, #4 and #6). Nutrients were added approx 1500 h daily at an N:P ratio of 15:1 (1.5 µM $NaNO_3$ and 0.1 µM KH_2PO_4) to the phosphate replete enclosures and at a ratio of 75:1 (1.5 µM $NaNO_3$ and 0.02 µM KH_2PO_4) to the phosphate deplete enclosures. Only samples from the phosphate replete enclosures (#1, #3 and #5) are

used for the purpose of this current study. Four daily samples (0600 h, 1200 h, 1800 h, and 2400 h) were taken from the surface of each mesocosm with 20 L carboys. The carboys were immediately brought back to the lab for sample filtration.

Flow Cytometry

Emiliania huxleyi and Coccolithovirus concentrations in each bag were measured using flow cytometry (FCM). All FCM analyses were performed with a FACSCalibur flow cytometer (Becton Dickinson, Franklin Lakes, NJ, USA) equipped with an air-cooled laser providing 15 mW at 488 nm and with standard filter setup. Algal counts were taken from fresh samples, while for viruses the samples were fixed with glutaraldehyde (0.5% final concentration), stored at 4ºC in the dark for 30 min, frozen in liquid nitrogen and stored at -80ºC (Marie et al., 1999). The samples were stained with SYBR Green I (Molecular Probes, Eugene, OR, USA) and analysed according to Marie et al. (1999).

RNA extraction

Water samples (1.5 L) from each experimental bag were filtered onto 0.45-µm-pore-size, 47-mm-diameter Supor-450 filters (PALL Corp.). The cells were resuspended from the filter using 2 ml PBS buffer, and transferred into an Eppendorff tube. Tubes were centrifuged for 1 min at 15500 × g, and the supernatant discarded. Total RNA was extracted with RNeasy Midi *Kits* (Qiagen) according to the manufacturer's instructions. Briefly, the samples were resuspended in 2 ml RLT lysis buffer, vortexed for 1 min. (5 sec. bursts), and spun for 5 min at 15500 × g. The supernatant was transferred to a new tube, and the nucleic acids were precipitated using 2 ml of 70% ethanol. The samples were loaded onto an RNA-binding midiprep column (Qiagen). The column was washed with RW1 and RPE solutions, and RNA was eluted in 200 µl of RNase free water. To remove contaminating DNA, the samples were DNase treated twice for 25 min. at 37º C with 2 U of Turbo DNA-free kit (Ambion) according to the manufacturer's instructions. The quality of the RNA samples was determined with a Bioanalyzer 2100 (Agilent, Waldbronn, Germany). Samples were stored at -80º C until further use.

Reverse Transcription

The quantity of the isolated RNA samples was measured with a NanoDrop ND 1000 (NanoDrop Technologies, Wilmington, DE). RNA was reverse-transcribed into first-strand cDNA using Superscript II (Invitrogen). In each case, 200 ng of total RNA was combined with 250 ng of random hexamer primers (Promega), and 1 µl dNTP mix (10 mM each), heated to 65° C for 2 min and snap cooled on ice for 2 min. Reaction buffer (5 ×) and DTT were added to the respective concentrations of 1x and 10 mM, the contents were gently mixed and incubated at 25º C for 2 min. 200 U of Superscript II enzyme were added for a final volume of 20 µl. Reactions were incubated at 42° C for 50 min, after which SuperScript II was inactivated by incubation at 70° C for 15 min. cDNA samples were then diluted in 100 ul of DNase free water. For DNA contamination control RNA (200 ng) was diluted in DNase free water without being subjected to reverse transcription. All samples were stored at -20 ºC.

Primer design and real-time PCR procedures

Primers were designed to target two key homologous sphingolipid pathway genes from *E. huxleyi* and coccolithovirus: the serine palmitoyl transferase, and dihydroceramide desaturase. *E. huxleyi's* β-tubulin and coccolithovirus major capsid protein genes were also used to serve as biological controls of host fitness and infection stage, respectively (Table S1). All primers were built based on sequence alignments including a wide geographical range of either *E. huxleyi* or EhV strains (check Table S2 for details on strain origin and Genbank sequence accession numbers). They were designed in areas of the genes that were conserved among all the tested viral or *E. huxleyi* strains, respectively. Sequence alignments for the SPT and DCD genes were extracted from Monier et al. (2009). The significant genetic distance separating the viral and host homologues facilitated the design of *E. huxleyi* and EhV specific primers (Figures S1 and S2, for SPT and DCD, respectively). Regular PCR using monoclonal *E. huxleyi* cultures and EhV isolates was used to confirm primers specificity and range. MCP gene sequences from Genbank were aligned (Figure S3) and used to design EhV primers tested against multiple viral strains. The *E. huxleyi* β-tubulin primers were designed from conservative parts of the *E. huxleyi* CCMP1516 relevant sequence (genome project http://genome.jgi-psf.org/Emihu1/Emihu1.home.html. Protein ID: 451245). Regular PCR and sequencing were then used to confirm that this set of primers amplifies the targeted β-tubulin gene sequence among a wide range of *E. huxleyi* strains (Figure S4). These novel sequences

were submitted to GenBank (Table S2). All primer sets were created using the online application Primer3 (http://frodo.wi.mit.edu/primer3/input.htm) (Rozen and Skaletsky, 2000). Each gene was cloned using the pGEM-T cloning kit (Promega), in order to generate standard serial dilutions and test primer efficiency.

Quantitative real-time PCR assays were performed in a total volume of 20 µl containing 4 µl of the above-described cDNA, 0.04 µM each of the 3′ and 5′ primers, 10 µl of Absolute™ qPCR Sybr® Green Rox Mix (Thermo Fisher Scientific - Abgene, France), and 4.4 µl of DNase free water. The amplification protocol was 40 cycles of 95 °C for 15 sec followed by 60 °C for 1 min. All real-time PCRs were performed with a CFB-3240 Chromo4 Detection System (Biorad). For accuracy every sample measurement was performed in duplicate, and ran against a non-reverse-transcribed (-RT) control to evaluate percentage of DNA contamination.

Real-time qPCR data analysis

An assumption-free analysis of qPCR data was used to estimate relative gene expression values following the method described in Ruijter et al. (2009). Based on the raw SYBR Green I fluorescence data, the application LinRegPCR v11.1 (Ruijter et al., 2009) was first used to calculate individual PCR efficiencies for each sample, followed by an estimation of mean PCR efficiency (E) and adequate fluorescence threshold (F) for each gene. Ct values were then used to calculate relative expression values (N) for each qPCR measure according to the basic equation for PCR amplification [N=F/(E^Ct)]. The final expression value for each gene in each time point was obtained by the difference between the respective sample and –RT control values (the latter corresponding to DNA contamination that should not be accounted for in an estimation of gene expression).

Data normalization

For each gene the lowest measure of expression was taken as the minimum level of detection. Transcript abundance for each host and viral genes was then normalized to the respective minimum level of detection, and finally normalised to the abundance of *E. huxleyi* cells (previously enumerated by flow cytometry).

4. Results

General bloom/infection dynamics

E. huxleyi bloom and consequent coccolithovirus infection evolved in a similar manner in three replicate enclosures (encl.1, 3 & 5) (Figure 1). Coccolithophore bloom progression increased exponentially from day 6-8 onwards, with encl.3 entering exponential phase approximately 2 days prior to encl.1 and encl.5. Initial *E. huxleyi* abundance was $1\times10^3 \pm 7.3\times10^1$ cells ml^{-1} in all enclosures, reaching a maximum number of cells (1.5×10^5, 1.7×10^5, 1.6×10^5 cells ml^{-1} in encl.1, 3 and 5 respectively) 5-7 days after the onset of exponential phase (day 13, 12, and 15 for encl. 1, 3 and 5), followed by sharp decline. The decline in *E. huxleyi* numbers coincided with the appearance and exponential increase of coccolithoviruses from day 12 onwards. A maximum concentration of 3.1×10^7 coccolithoviruses ml^{-1} was registered in encl.3 on day 16. *E. huxleyi* decline to pre-bloom abundance following the appearance of coccolithoviruses went to completion in encl.3, but was not observed in encl.1 or encl.5, due to the termination of experimental sampling. Previous experiments under these conditions have shown that the *E. huxleyi* and associated coccolithovirus population are *E. huxleyi* and coccolithovirus groups (Bratbak, Egge, and Heldal, 1993; Jacquet et al., 2002; Martinez-Martinez et al., 2006b), this was confirmed in this work through quantitative real time PCR (see below).

Gene transcription analysis

Quantitative real-time PCR was used to determine the dynamics of sphingolipid pathway gene expression in both *E. huxleyi* and coccolithovirus within the natural community. Total RNA samples obtained during the *E. huxleyi* exponential growth phase and subsequent decline (days 6 – 16) were interrogated for transcript abundance. Coccolithophore and coccolithovirus-specific primers were designed based on alignments of sequences from multiple *Emiliania huxleyi* and EhV strains encompassing a wide geographic diversity. The β-tubulin gene of *E. huxleyi* was used to follow a host housekeeping gene transcriptional levels, while coccolithovirus major capsid protein (MCP) gene expression was monitored as a general measure of virus propagation (Table S1). Sphingolipid pathway transcript levels were monitored using primers for serine palmitoyl transferase (SPT) and dihydroceramide desaturase (DCD) (the first and last steps of the pathway, respectively) for both host *E.*

huxleyi- and Coccolithovirus-derived transcripts (Table S1). Host and viral SPT and DCD genes were sufficiently divergent (54 % and 57 % nucleotide sequence divergence, respectively) (Monier et al., 2009) to allow the design of specific primers for SPTh and SPTv, and DCDh and DCDv, respectively. Gene expression measures were normalized to the concentrations of *E. huxleyi* cells.

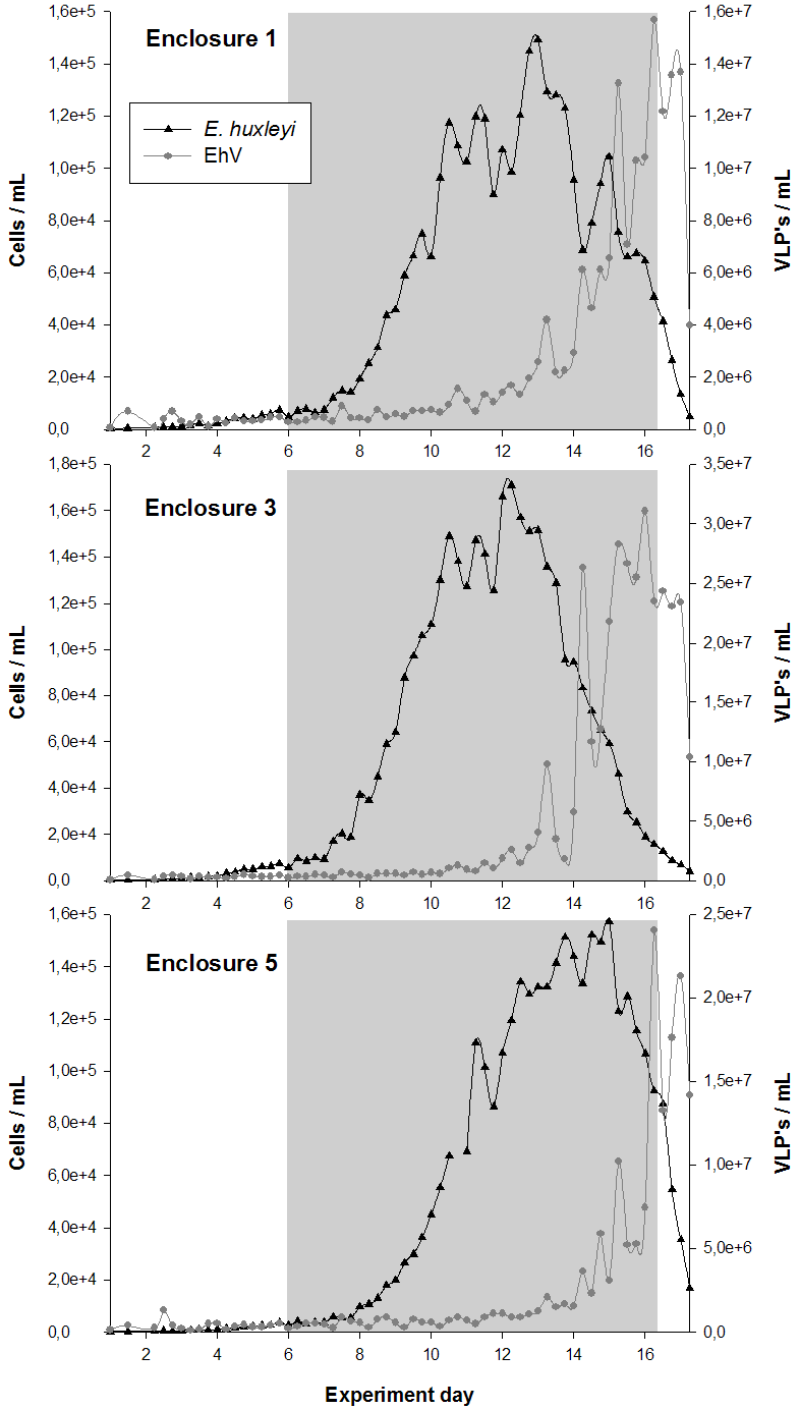

Figure 1. *E. huxleyi*/coccolithovirus abundance in Enclosures 1, 3 and 5 as determined by flow cytometry. Grey areas indicate the sections selected for posterior gene transcription analysis.

Tubulin expression remained consistently low throughout the duration of the sampling period. Initial tubulin transcript levels were approximately 10^2 relative expression units (REU), but decreased by an order of magnitude in all enclosures with the onset of bloom formation (Figure 2). Tubulin transcript abundance remained low (1-10 REU) in all enclosures for the remainder of the sampling period, except in encl.3 where a return to original pre-bloom levels was observed following the near total demise of the *E. huxleyi* population (Figure 2).

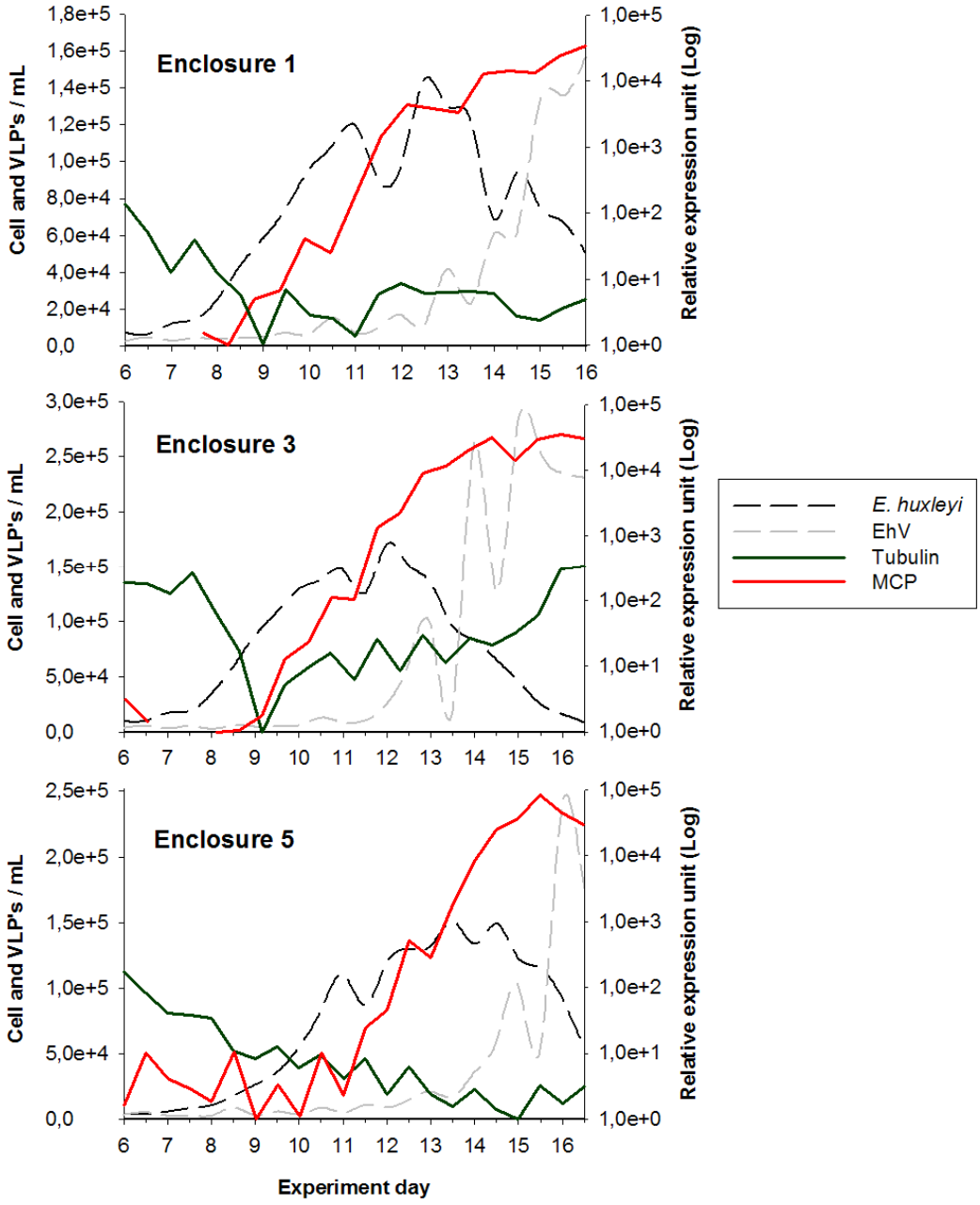

Figure 2. Relative gene expression (Log scale) of *E. huxleyi* β-tubulin and coccolithovirus major capsid protein genes (right axis). Cell and virus numbers are plotted for reference (black and grey dashed lines, respectively) (left axis). Viral capsid values (grey lines) were lowered of 2 orders of magnitude for better perception.

Viral MCP transcript abundance increased exponentially over a 4-5 day period from an initial 1-10 REU under pre-bloom conditions in all enclosures to maximum levels of over 10^4 REU (Figure 2). These high levels were maintained over a 3-4 day period despite near total bloom demise in encl.1 and encl.3; encl.5 only attained maximum MCP transcript abundance 1 day prior to the completion of sampling. The onset of major MCP expression followed *E. huxleyi* bloom formation by 1 day, but preceded major coccolithovirus release by approximately 4 days in each enclosure.

SPTh and DCDh followed near identical expression profiles, as did SPTv and DCDv (Figure 3). From initial levels between 10^1-10^2 REU, there was a general trend to decrease to 1-10 REU observed for SPTh and DCDh over the 10-day sampling period. Conversely, SPTv and DCDv varied between undetectable and 10^2 REU initially, decreased with a similar trend to their host counter parts during the first half of the sampling period, and then increased dramatically akin to MCP expression levels in the second half of the sampling period. The dramatic increase in SPTv and DCDv expression (to levels of 10^3-10^4 REU in all enclosures) followed *E. huxleyi* bloom formation by approximately 2 days, preceding major virion release by 3 days in all enclosures.

When considered as a whole, the transcriptional patterns observed for the SBP genes (SPTh, SPTv, DCDh and DCDv) can be split into 3 defined transcriptional stages along bloom formation and decline. These stages, intrinsically linked and dependent on host/virus population abundance, are transcriptionally defined and not simply based on population dynamics (i.e. boom and burst phases). The three-stage pattern was observed in all three enclosures. In Stage 1 (4-5 days long), *E. huxleyi* abundance increases in the exponential phase as both host and virus SBP relative expression levels decrease. This slow decrease continues until a balance point is reached (the stage 1/2 boundary) when virus SBP takes over almost entirely and host SBP is reduced to minimal transcriptional levels. This event occurred on day 10 in encl.3 and day 11 in encl.1 and encl.5 (Figure 3). *E. huxleyi* population growth begins to plateau. Viral SBP transcription increases exponentially through stage 2 which lasts 2-3 days. Virion abundance remains low throughout stage 2. The third transcriptional stage is defined by the onset of a plateau in virus SBP expression levels (Day 12 in encl.3, day 13 in encl.1 and encl.5). Virus SBP expression maintains its maximum level throughout Stage 3 as major coccolithovirus release occurs and *E. huxleyi* abundance rapidly declines.

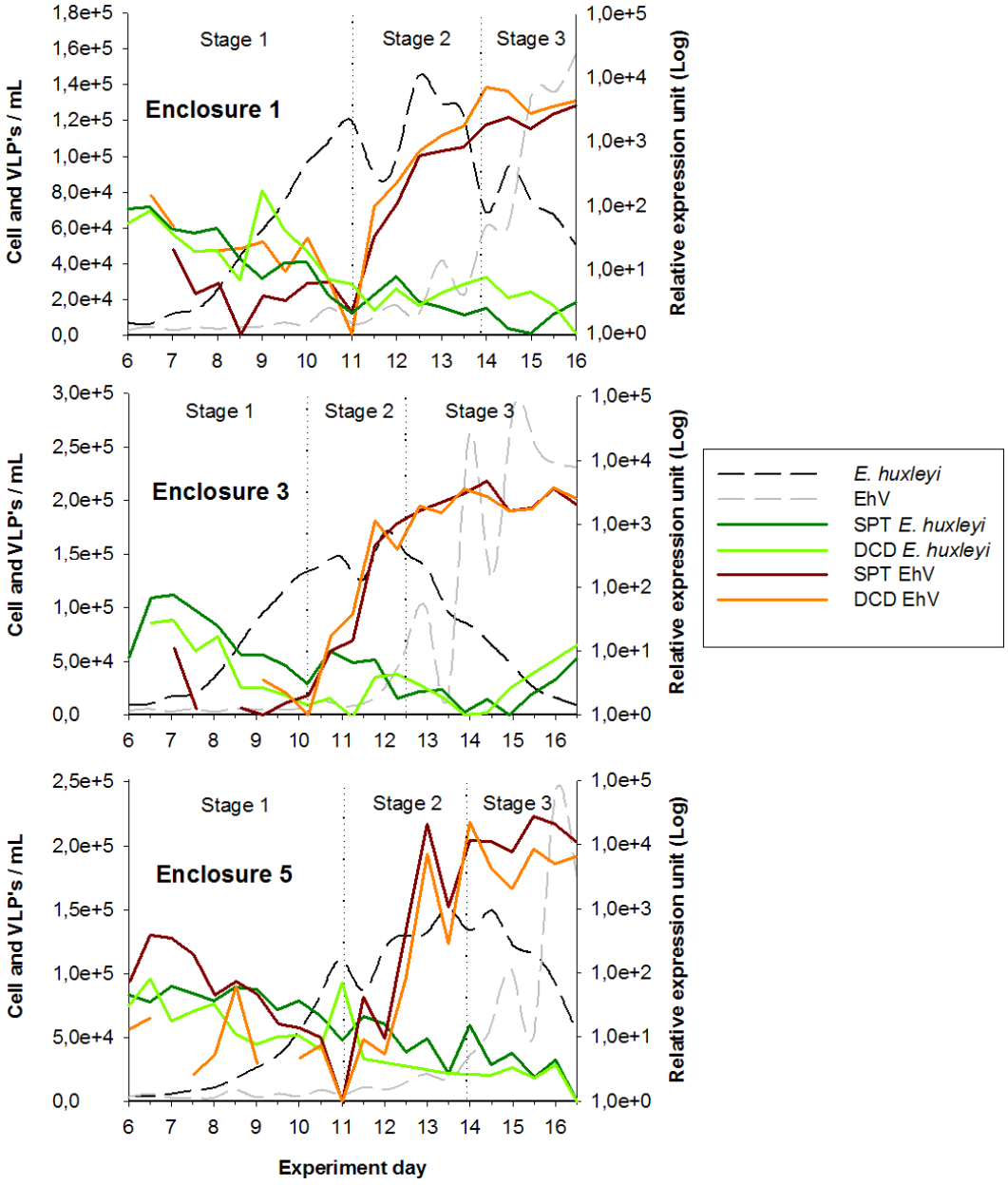

Figure 3. Relative gene expression (Log scale) of *E. huxleyi* and coccolithovirus serine palmitoyl transferase, dihydroceramide desaturase (right axis). Cell and virus numbers are plotted for reference (black and grey dashed lines, respectively) (left axis). Viral capsid values (grey lines) were lowered of 2 orders of magnitude for better perception.

5. Discussion

All mesocosm enclosures showed similar expression profile and population dynamics; the 1-2 day delay observed between encls.1/5 and encl.3 can be attributed to the typical variability inherent to semi-natural experiments (Martinez-Martinez et al., 2006a).

Consistently in all enclosures relative β-tubulin expression per cell decreased as the viral infection took over, eventually returning to pre-bloom levels by the end of the infection in encl.3. If encl.1 and encl.5 were sampled beyond day 16 they too may have shown a similar increase. The source of the β-tubulin in stage-3 could be either resistant diploid cells which have avoided infection, and/or a developing and virus-resistant haploid population as has been observed at bloom demise (Frada et al., 2008).

The coccolithovirus MCP gene displayed a major increase in expression that clearly corresponded to the demand for virion assembly as a consequence of the infection of the exponentially growing *E. huxleyi* population. However, the massive demand for virion assembly components was not reflected immediately in free virus abundance which only increased approximately 4 days after the heavy requirement for building components was induced. This could be due to immediate adsorption of newly released viruses onto uninfected cells or, alternatively, intracellular virus retention by infected hosts prior to total cellular breakdown after the prolonged 4-day infection period. The truth may lie somewhere in between the two scenarios: infected cells may release newly synthesised virus at a low rate initially (and these viruses are rapidly adsorbed onto exponentially growing uninfected cells) prior to total cellular disruption and major release of their contents.

The expression pattern obtained for the coccolithophore and coccolithovirus SBP genes tested here suggests that sphingolipid biosynthesis is a crucial factor for a successful infection. The reproducible expression profiles observed for all genes tested in three independent enclosures reinforce the robustness of the data which may well be a good representation of natural bloom dynamics. The synchronized decrease in host and viral SBP expression in stage-1 may hint at an underlying transcriptional arms race going on within freshly infected cells. Since the *E. huxleyi* and EhV SBP genes share a common ancestor, they may well have conserved ancestral regulatory mechanisms (Monier et al., 2009). The infected host may thus respond to viral infection by rapid gene silencing or transcriptional control which could account for the observed decrease in virus SBP transcript abundance. The following rapid and exponential increase in virus SBP message between days 10/11 and 12/13 (Fig.3) suggests a massive increase in sphingolipid requirement as the infection progresses.

The presence of a near complete sphingolipid pathway in the coccolithovirus genome, in tandem with its substantial expression during the infection, suggests a crucial role for sphingolipids in coccolithovirus life cycle. However, the physiological function(s) of the sphingolipids in either virus or host is still to be resolved. In numerous other systems, *de novo* ceramide metabolism has been identified as playing an important role in bioactive lipid

signalling, namely the activation of apoptosis and other crucial pathways (Hannun, 1996; Hannun, Luberto, and Argraves, 2001; Hannun and Obeid, 1995; Hannun and Obeid, 2002; Hannun and Obeid, 2008). Previously, the production of a sphingolipid such as ceramide in this host/virus system has been proposed as a trigger mechanism for *E. huxley*i programmed cell death (Allen et al., 2008; Frada et al., 2008; Han et al., 2006; Monier et al., 2009), potentially preventing higher levels of community contamination. Yet, the decrease in host SBP expression observed here suggests a reduction, not induction, of this pathway by the host in response to infection. The possibility still exists that the virus SBP could actually function to up regulate PCD in order to promote virion and/or apoptotic bodies release. The demonstrated activation of host metacaspase activity by the coccolithoviruses (Bidle et al., 2007) adds credence to theory of PCD manipulation, at some level, in this system.

Control of host PCD has been described in other viruses belonging to the large family of NCLDVs (Nucleo-Cytoplasmic Large DNA Viruses). In all cases viruses were reported to encode genomic machinery impeding host's PCD (Hay and Kannourakis, 2002; McLean et al., 2008). The potential virus driven activation of host PCD described herein remains so far an isolated and extraordinary feature among the NCLDVs, consistent with the unique life cycle and propagation strategies exhibited by the coccolithovirus and its host (Allen et al., 2006a; Allen et al., 2006c; Allen, Schroeder, and Wilson, 2006).

However, the assumption that homologous genes produce proteins with homologous function has clearly not been tested in this system. It may well be that the sphingolipids produced by the host using its own SBP are different structurally and functionally to those produced by the viral SBP. For example, it is tempting to speculate that a host specific sphingolipid could be involved in bioactive lipid signalling among host cells, eventually triggering meiosis in a part of the cells population to escape from viral infection (Frada et al., 2008). In such case, virus specific sphingolipids may well play the role of stereochemical mimics inhibiting this process.

Alternatively, the function of the virally produced sphingolipid may be related to its inherent physical properties in the cell membrane. The observation that exponential virus assembly (MCP transcription), but not release, occurs 1 day prior to the exponential induction of the virus SBP, suggests that sphingolipids could be intrinsically involved in virion release and not cellular signalling. One theory is that these induced sphingolipids promote the formation of lipid rafts in the membrane becoming focal points on the membrane for viral budding and release. Indeed, the accumulation of sphingolipids within infected cells may be the trigger for virion release; only once a critical concentration is reached within the

membrane do high numbers of virions get released. This could account for the >24 hour delay in release of virions observed in laboratory experiments (Unpublished data).

This study provides a unique insight into the transcriptional interplay between eukaryotic host and viral homologous metabolic pathways that occurs in an infected population of marine phytoplankton. The gene expression patterns observed clearly show the existence of strict constraints and stages along the bloom/infection process. Ultimately, the coccolithovirus take-over and control of SBP, possibly resulting in the distinct separation of viral capsid production from their release, may allow maximal virion outreach. The different hypotheses here raised demand further research not only at the transcriptomic level, but also at the proteomic and metabolomic levels.

The constant evolutionary battle between hosts and their viruses has been ongoing since life evolved, it is an arms race that will not end until either the host or its virus goes extinct. This year marks the 150th anniversary since Darwin published his seminal work "On the Origin of the Species" from which the phrase the survival of the fittest was born. It is worth pondering if, in the case of the *E. huxleyi*/coccolithovirus system described herein, the survival of the fittest may actually be dependent on the survival of the fattest.

6. Acknowledgements

The authors would like to thank Sarah Romac and Ian Probert for help provided in sample treatment, and Arnauld Tanguy, Laurence Garczarek, Bénédicte Charrier, and Nicolas Taris for technical advices concerning the qPCR procedures and analysis. We acknowledge Hiroyuki Ogata and Adam Monier for the help on concept framing. AP, CdV, MJA, and WHW conceived and designed research. AP, MJA, SAK, and WHW performed the reported mesocosm experiments. SAK and WHW performed the flow cytometry analysis. AP and MJA performed the gene transcriptional analysis. AP, MJA, WHW, and CdV contributed to the drafting of the manuscript. AP was supported by a Marie-Curie Early Stage Training fellowship (EU-FP6) awarded to the Station Biologique Roscoff, and by a grant awarded in the framework of ‘‘Marine Genomics Europe’’ European Network of excellence (2004–2008) (GOGE-CT-505403). MJA and SAK were supported by grants awarded to WHW from the Natural Environment Research Council (NERC) Environmental Genomics thematic program (ref. NE/A509332/1 and NE/D001455/1) and Oceans 2025. WHW was supported by a National Science Foundation (NSF) grant ref. EF0723730. This work is part of the

pluridisciplinary project *BOOM* (Biodiversity of Open Ocean Microcalcifiers) funded by the French Agence Nationale de la Recherche, Grant ANR-05-BIODIV-004.

8. Supplementary data – Tables

Table S1. *E. huxleyi* and coccolithovirus primer sequences used in this study.

Protein name	Oligo name	Direction	Position	Oligo Sequence	Product Size
E. huxleyi					
Serine palmitoyl transferase (SPTh)	EX_SPT_F	Forward	1093	ACTGATTTCCTCCGCATGAC	190
	EX_SPT_R	Reverse	1282	CGATGCCAAACGAGTAGATG	
Dihydroceramide desaturase (DCDh)	EX_DCD_F	Forward	244	AAGGCGGAGTGGGCGAAC	90
	EX_DCD_R	Reverse	333	GGCGTGGTAGTAGCGGAAGG	
β-tubulin	Tubβ_Fw	Forward	1037	CGCTGTACGACATCTGCTT	164
	Tubβ_Rv	Reverse	1201	GGAAGGGGATCATGTTGAC	
Coccolithovirus					
Serine palmitoyltranferase (SPTv)	EhV_SPT_F	Forward	1793	AGTCCGGTATCGTCTTGTCG	190
	EhV_SPT_R	Reverse	1982	CGCAATGCGATAATACATGG	
Dihydroceramide desaturase (DCDv)	EhV_DCD_F	Forward	417	GGACATTTCTTCCGTCATGG	153
	EhV_DCD_R	Reverse	569	ACGGTCCAATTTGCAAGAAG	
Major capsid protein (MCP)	MCP1Fw	Forward	255	ACGCACCCTCAATGTATGGAAGG	90
	MCP90Rv	Reverse	344	AGCCAACTCAGCAGTCGTTC	

Table S2. Details (year and geographical origin) of the monoclonal *E. huxleyi* strains and coccolithovirus isolates used as reference for primer building, and Genbank accession numbers of the respective gene sequences.

	Year	Geographical origin*	MCP	β-Tubulin	SPT**	DCD**
Coccolithovirus						
EhV-84	1999	EC	AF453849		FJ531546	FJ531569
EhV-86	1999	EC	AF453848		FJ531547	FJ531570
EhV-88	1999	EC	AF453850		FJ531548	FJ531571
EhV-V1	2003	RN			FJ531549	FJ531572
EhV-V2	2003	RN			FJ531550	FJ531573
EhV-163	2000	RN	AF453851		FJ531551	FJ531579
EhV-201	2001	EC	AF453857		FJ531552	FJ531574
EhV-202	2001	EC	AF453856		FJ531553	FJ531575
EhV-205	2001	EC	AF453854		FJ531554	FJ531576
EhV-207	2001	EC	AF453853		FJ531555	FJ531578
EhV-208	2001	EC	AF453852		FJ531556	FJ531577
E. huxleyi						
RCC1242 (CCMP1516)	1991	Pacific Ocean / Offshore Peru		GQ232275	FJ531558	FJ531565
RCC1215 (TW1)	1998	Mediterranean Sea (Spain)		GQ232274	FJ531559	FJ531567
RCC1259 (CCMP374)	1989	Gulf of Maine		GQ232273	FJ531557	FJ531564
RCC1255 (CCMP370)	1959	Oslo Fjord		GQ232272	FJ531562	FJ531563
RCC1235 (VF20)	2006	Mediterranean Sea (France)		GQ232278	FJ531560	FJ531568
RCC1237 (VF22)	2005	Mediterranean Sea (France)		GQ232276		
RCC1247 (ESP7410)	1999	Mediterranean Sea (Spain)		GQ232277		
RCC1253 (OS-2)	2006	Sea of Japan			FJ531561	FJ531566

* EC: Western English Channel (off the coast of Plymouth, UK); RN: Raunefjorden (Western Norway during a mesocosm experiment); See Schroeder et al. (2002) for more detail on the EhV isolates.

** From Monier et al. (in press).

9. Supplementary data – Figures

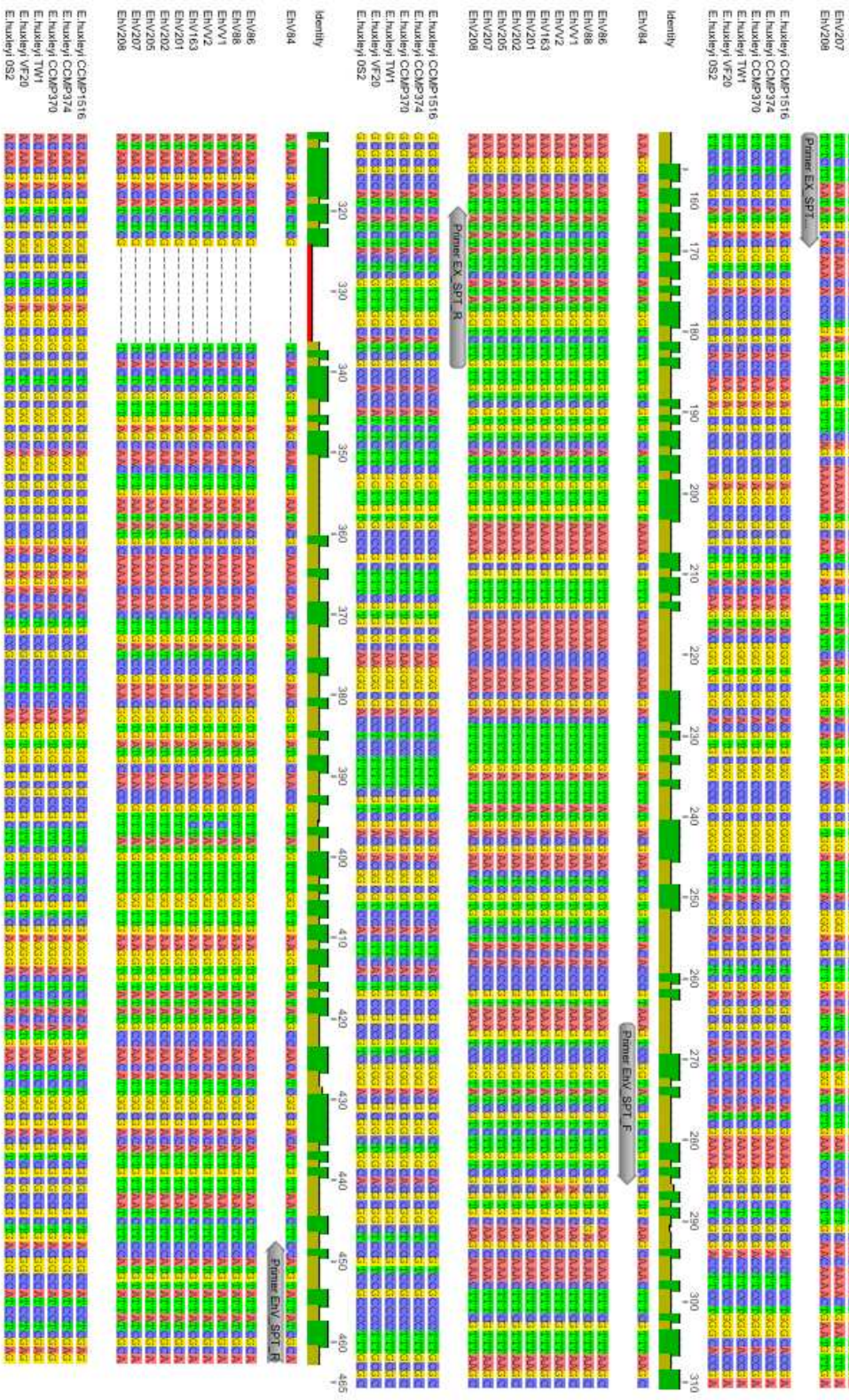

Figure S1. Serine palmitoyltransferase (SPT) sequence alignment, indicating chosen primer regions.

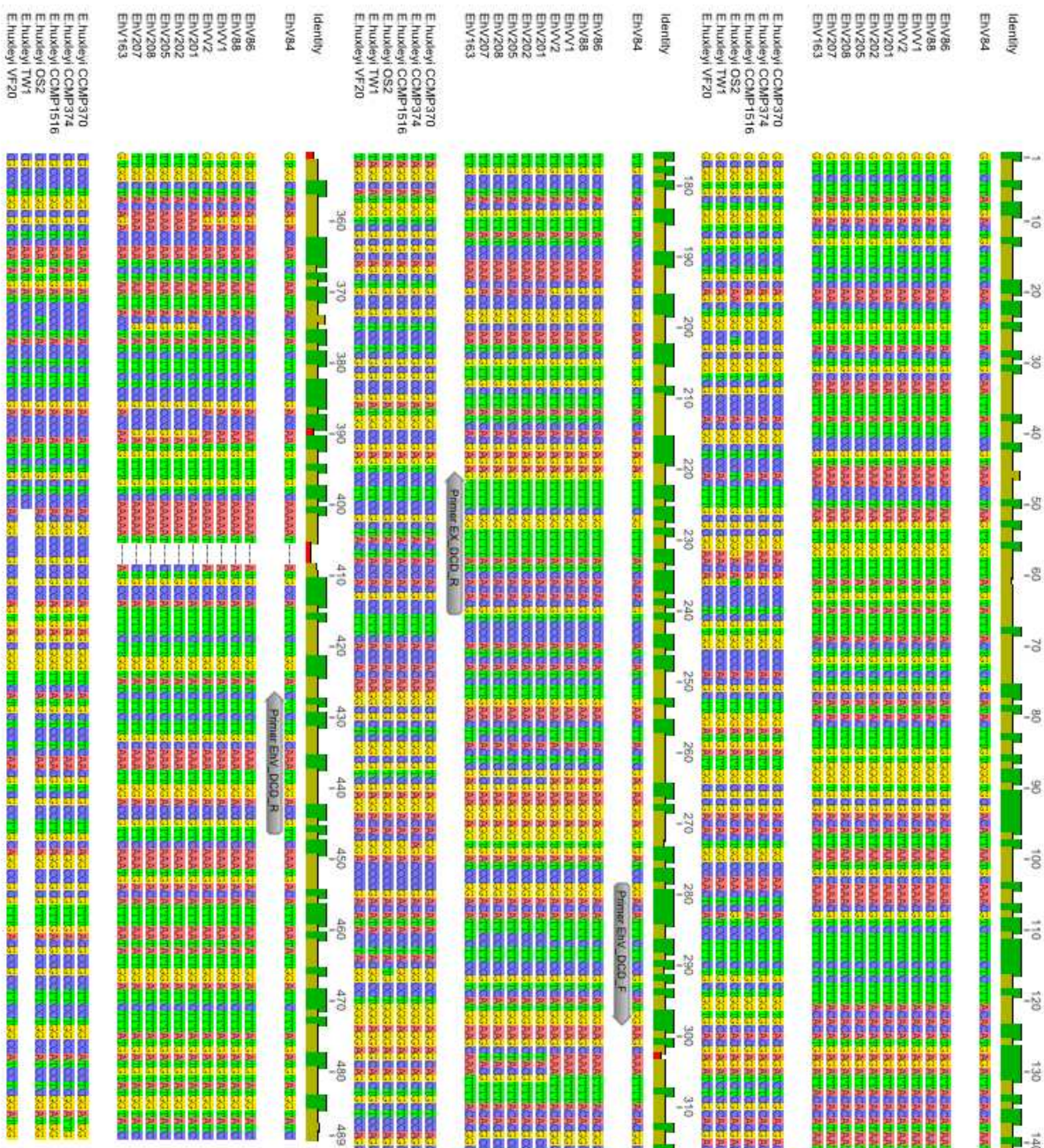

Figure S2. Dihydroceramide desaturase (DCD) sequence alignment, indicating chosen primer regions.

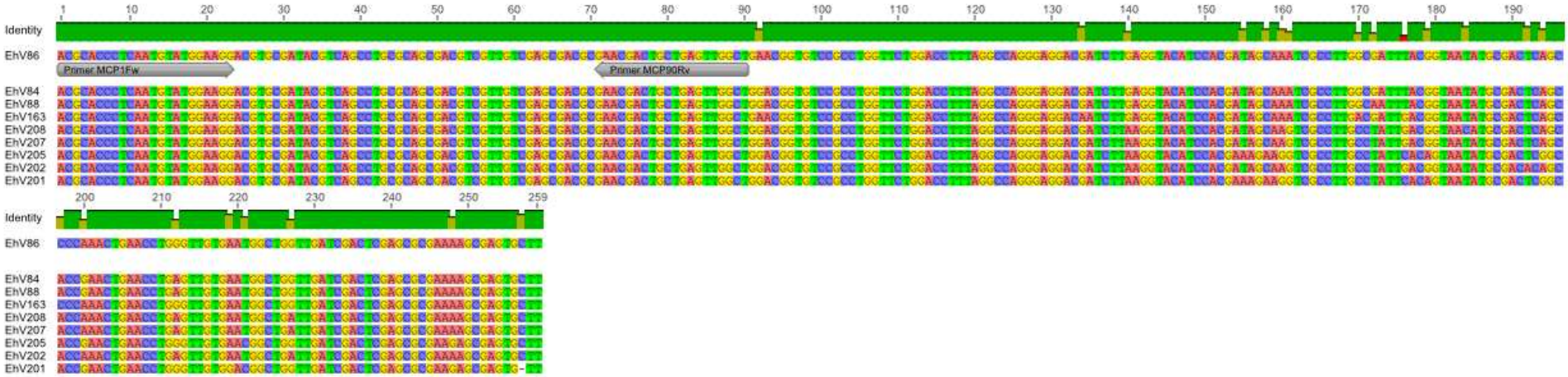

Figure S3. Major capsid protein (MCP) sequence alignment, indicating chosen primer regions.

Figure S4. β-tubulin sequence alignment, indicating chosen primer regions.

Chapter 5.

Novel transcription features unveiled during natural coccolithovirus infection

António Pagarete[1], Gildas Corguillé[2], Bela Tiwari[3], Hiroyuki Ogata[4], Lisa Olohan[5], Colomban de Vargas[1], William H. Wilson[6] and Michael J. Allen[7]

[1] Equipe EPPO-Evolution du Plancton et PaléoOcéans, CNRS-UMR7144, Université Pierre et Marie Curie, Station Biologique, FR-29682 Roscoff, France.

[2] CNRS/UMPC, FR2424, Service Informatique et Génomique, Station Biologique de Roscoff, Place Georges Teissier, BP74, 29682 Roscoff Cedex, France.

[3] NERC Environmental Bioinformatics Centre, Centre for Ecology and Hydrology, Maclean Bldg, Benson Lane Crowmarsh Gifford, Wallingford, UK.

[4] Structural and Genomic Information Laboratory, CNRS-UPR2589, Mediterranean Institute of Microbiology (IFR-88), Université de la Méditerranée, Parc Scientifique de Luminy, 163 avenue de Luminy Case 934, FR-13288 Marseille, France.

[5] School of Biological Sciences, University of Liverpool, Liverpool L69 3BX, UK

[6] Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, 04575 ME, USA.

[7] Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, PL1 3DH, UK.

1. Summary

Over the last two decades oceans' virology has progressively increased the distance to the original global community targeted approaches. While allowing more profound understanding of specific infection mechanisms employed by a few domesticated "lab-model" viruses, such approaches often are not capable of addressing the complex panoply of phenomena occurring in the natural environment. Here we employed a microarray approach to present the first attempt of studying host (*Emiliania huxleyi*; Haptophyta) and virus (EhV; coccolithovirus; phycodnaviridae) global transcriptomics occurring within a natural oceanic community enclosed in a mesocosm. Our results showed that during host blooms there is a

** Article under preparation. The data here presented is still in the form of a preliminary analysis, and should hence be understood with the corresponding caution. We decided to present the analysis done so far given that this project consisted of an important part of the PhD program. Further and more robust statistical analyses will be performed before final presentation of these data to the broad scientific community.*

synchronized moment of viral takeover clearly reflected in the global transcriptomic signal. Among the 279 genes that significantly increased in abundance from pre to post viral takeover the majority (52%) corresponded to viral sequences for which there is so far no match in the protein databases. *E. huxleyi* and EhV gene transcripts with putative functions (that significantly increased abundance from pre to post viral takeover) corresponded not only to expected functions such as genetic information processing, but also to some less expected genes probably involved in posttranslational control, intracellular trafficking mechanisms, or even control of programmed cell death. While far from being conclusive in what regards the real role played by these genes, our results (1) indicate that a complex infection strategy is used by the coccolithovirus, one that faces/demands the balanced use of intricate host machinery, and (2) provide precious lights for the progressive unveiling of the so far largely unknown coccolithovirus infection processes and genomic functions.

2. Introduction

Over the past two decades the study of oceanic virioplankton has gained progressive importance as the scientific community unveiled the overwhelming abundance, and extreme functional and genetic diversity presented by marine viruses (Suttle, 2005b). The study of virioplankton started by using global approaches to understand viral impact at the community level. Numerous works attempted to measure viral presence, development of the viral communities and its correlation to the development of host groups (for extended reviews see Brussaard et al., 2008; Weinbauer, 2004; Wommack and Colwell, 2000). Studies were able to clearly demonstrate that abundance and distribution of major viral groups change accordingly to the progression of major planktonic goups (Larsen et al., 2001; Rodriguez et al., 2000; Wommack et al., 1992). Viral control of host development becomes more evident in situations of algal blooming, with viruses clearly being responsible for extensive bloom termination (Brussaard, Kuipers, and Veldhuis, 2005; Castberg et al., 2001; Larsen et al., 2004; Maranger, Bird, and Juniper, 1994; Nagasaki et al., 1994). These advances were made possible by the development of specific epifluorescence microscopy (Wen, Ortmann, and Suttle, 2004), flow cytometry (Brussaard, Marie, and Bratbak, 2000), and genomic fingerprinting techniques (Wommack et al., 1999).

The first isolations and characterization of specific phytoplankton viruses brought with it the possibility of understanding in a more profound manner the nature of specific host/virus

interactions. Studies that started by classical host range approaches and infection kinetics by means of cell and virus numbers (Bergh et al., 1989; Borsheim, Bratbak, and Heldal, 1990; Bratbak et al., 1990) soon picked up on the development of new genomic and transcription characterization tools to push the field of phytoplankton virology one step further. The development of easily accessible culture independent genomic sequencing tools lead to the amazing discovery of a whole new range of prokaryotic and eukaryotic viruses (Breitbart et al., 2007; Suttle, 2005a). With genome sizes sometimes scaling in the same order of magnitude as the smallest cells (for examples please check Raoult et al., 2004; Wilson et al., 2005b) these viruses have reheated the debate on the definition of life, pushing further the concept of life itself (Claverie et al., 2006; Moreira and Lopez-Garcia, 2009). These discoveries have had a significant impact in a world that considered viruses as mere plasmidic parasites, rather than consistent evolutionary entities matching in complexity with the hosts they use to assure the continuity of their genes. Despite remaining obligate intracellular parasites with a requirement for suitable host cell environment to replicate, many of these viruses have been proven to carry in their genomes vast and complex gene "cellular" machinery. Some authors introduced even the concept of a "living virus factory" to express the manner in which some giant viruses (notably the extreme case of the mimivirus, whose genome encodes 1262 putative genes, Raoult et al., 2004) take the host cell environment and metabolic machinery to become themselves a new cell inside the cell (Claverie, 2006; Novoa et al., 2005). In such thinking, free virions would be considered to be merely "inactive spores" of the virus factory.

Within the genomes of the newly discovered phytoplankton viruses several novel and unexpected metabolic functions have been identified. For example cyanophages inherited from their hosts (cyanobacteria), through horizontal gene transfer (HGT), all the necessary genetic machinery to restore the cell's photosynthetic apparatus (Sharon et al., 2009). During the lytic cycle, most of the host's transcription is shut down by phage, which then replaces it by its own virus message. To keep proton motive force cyanophage-encoded photosynthesis proteins are expressed during the infection cycle, hence providing the virus with improved selective fitness (Clokie and Mann, 2006; Lindell et al., 2005). Another unexpected set of viral genes is found in the *Emiliania huxleyi* virus (EhV). This virus encodes in its genome a near complete de novo sphingolipid biosynthesis pathway also believed to be the product of direct HGT from its host (Monier et al., 2009, see Chapter 3). The sphingolipid pathway leads to the production of different sphingolipids, molecules often involved in cell signalling and stress response (Hannun and Obeid, 2008; Merrill, 2002). Notably ceramide, the final product

of this pathway, is often implicated in the control of programmed cell death and apoptosis (Pettus, Chalfant, and Hannun, 2002; Siskind, 2005; Vardi et al., 2009).

The major trend throughout these past 15 years has hence been to go from global ecological dynamics to specific genomic and functional interactions studied with the few host/virus models available for lab based experiments. The phytoplankton virology field is now faced with the quest of developing new tools to push our understanding of functional viral impact in a global ecological scale. Moreover, whilst single gene analysis can provide a particular functional insight into a process, a system wide approach is essential for broader understanding of physiological interdependence.

Here we present the first attempt of studying host and virus global transcriptomics occurring within a natural oceanic community. In order to do so, we focused on the progression of an *E. huxleyi* bloom followed by the inevitable crash provoked by the severe lytic *Emiliania huxleyi* virus (EhV). *E. huxleyi* is a calcifying eukaryotic microalgae that is also the most numerous and ubiquitous coccolithophore in today's oceans. EhVs have repeatedly been identified as the major cause of bloom demise. We designed a microarray set to target both *E. huxleyi* and EhV wide gene transcription patterns. Such an approach has so far been neglected given the high genomic variability levels commonly associated with natural phytoplankton and virioplankton communities. In such context any recognizable gene expression pattern can only stand out if (1) there is significant genetic resemblance between the target natural strains and the host and virus strains used in the microarray design, and (2) if, within the complex natural community there is an extense synchronization of transcription events. We consider this experiment in a mesocosm environment as a step along the path to bring the study of phytoplankton and virioplankton dynamics from the lab back to the natural oceanic communities.

3. Materials and methods

Set-up of the mesocosm experiment

The *E. huxleyi*-induced blooms were conducted in the Raunefjorden, Western Norway coast, at the Marine Biological Field Station, for 17 days (5th to the 21st of June 2008). Six transparent polyethylene enclosures (11 m^3; 90% penetration of photosynthetically active radiation) purchased from ANI-TEX (Notodden, Norway) were mounted on floating frames

moored along the south side of a raft (Egge and Heimdal, 1994) and filled with unfiltered fjord water pumped from 10 m depth adjacent to the raft. Homogeneous water masses within the enclosures were ensured by pumping water from the bottom of the bag to the surface. The 6 enclosures (encs.) were divided in 2 treatment groups allowing triplication of each treatment: phosphate replete (encs. #1, #3 and #5) and phosphate deplete (encs. #2, #4 and #6). Nutrients were added approx 15:00 h daily at an N:P ratio of 15:1 (1.5 µM $NaNO_3$ and 0.1 µM KH_2PO_4) to the phosphate replete enclosures and at a ratio of 75:1 (1.5 µM $NaNO_3$ and 0.02 µM KH_2PO_4) to the phosphate deplete enclosures. Four daily samples (06 h, 12 h, 18 h, and 24 h) were taken from the surface of each mesocosm with 20 l carboys. Samples were immediately brought to the lab where 1.5 l of each sample was filtered onto 0.45-µm-pore-size 47-mm-diameter Supor-450 filters (PALL Corp.).

In depth flow cytometry and diversity analysis was performed previously and reported in **Annexe B** (Kimmance et al., in press).

RNA extraction

The cells were resuspended from the filter using 2 ml PBS buffer, and transferred into an Eppendorff tube. Tubes were centrifuged for 1 min at 15500 × g, and the supernatant discarded. Total RNA was extracted with RNeasy Midi *Kits* (Qiagen) according to the manufacturer's instructions. Briefly, the samples were resuspended in 2 ml RLT lysis buffer, vortexed for 1 min. (5 sec. bursts), and spun for 5 min at 15500 × g. The supernatant was transferred to a new tube, and the nucleic acids were precipitated using 2 ml of 70% ethanol. The samples were loaded onto an RNA-binding midiprep column (Qiagen). The column was washed with RW1 and RPE solutions, and RNA was eluted in 200 µl of RNase free water. To remove contaminating DNA, the samples were DNase treated twice for 25 min. at 37º C with 2 U of Turbo DNA-free kit (Ambion) according to the manufacturer's instructions. The quality of the RNA samples was determined with a Bioanalyzer 2100 (Agilent, Waldbronn, Germany). All samples tested negative for genomic DNA contamination. Samples were stored at -80º C until further use.

Fluorescent labelling of cRNA

Random amplification of the entire mRNA population was achieved using the Microarray Target Amplification Kit (Roche). This required the use of a special primer,

containing a random sequence with no significant homology to any known sequences in public databases (Target Amplification Sequence, [TAS]) in addition to T7 promoter and oligo (dT) sequences, for first-strand cDNA synthesis. The TAS-region generates a 3´ anchor on the cDNA for subsequent PCR amplification with a TAS-PCR primer. First and second strand cDNA synthesis was performed from 250 ng of total RNA and purified using the Target Purification Kit (Roche). For expression profiling, purified cDNA was randomly amplified in a 100 µl volume by 21 cycles of PCR using TAS primers according to the manufactures instructions. PCR products (produced after the appropriate number of PCR cycles) were purified using the Microarray Target Purification kit (Roche). The resulting PCR-amplified cDNA was then transcribed into Cy5 (for each sample) and Cy3 (for a pool of random hexamers to be used in a control channel) fluorescently labelled cRNA using the T7 Microarray RNA Target Synthesis Kit (Roche). The appropriate labelled cRNAs were combined and then purified using the Microarray Target Purification kit.

Microarray design

The microarray is described extensively in the MIAME compliant database entry E-MAXD-23, available at www.ebi.ac.uk/arrayexpress. Briefly, 70mer oligonucleotides were designed and synthesised for an *Emiliania Huxleyi* 1516 EST sequence library (http://www.nematodes.org/neglectedgenomes/emiliania/) and for every EhV-86 gene by Operon GmbH, for a microarray hybridisation temperature of 68 °C. Hairpin-free, hybridisation probes were designed with a 3’ bias within selected stability limits. Oligonucleotide length was initially fixed to 70, Tm range was set between 73 and 81 °C, and GC content between 40 and 60%. Probe elements were deposited and immobilised onto amino silane treated glass slides (Corning GAP II) using a BioRobotics MicroGrid 2 printer. Each probe was printed in quintuplicate along with a collection of negative and positive control probes. Probes were printed in a 1 × 7 meta-grid, each sub-grid composed of 53 rows and 12 columns. Ten SpotReport Alien PCR products (Stratagene) and were used as a positive control and to confirm consistency between cyanine-3 (Cy3) and cyanine-5 (Cy5) scanning channels. Other control probes included were 3 × SSC buffer, human COT-1 DNA, poly(dA) (40-60 bases in length, single stranded).

Sample hybridization to microarray

Total cRNA present in each sample was quantified using a NanoDrop ND 1000 (NanoDrop Technologies, Wilmington, DE). For each sample 7.5 µl 20 × SSC (Sigma), 1.0 µl 10% SDS and 100 ng of labelled cRNA were combined in a total volume of 45 µl. 33.3 µl of this dye-labelled target were added to 16.7 µl of pre-heated (50 °C) 3 × Hybridisation buffer (consisting of 9 × SSC, 0.6% SDS), and incubated at 100 °C for 2 mins. of denaturation in the dark. Tubes were kept for a few minutes at 50 °C while waiting to be loaded onto microarray slide. 46 µl of target were loaded into an AO Maui mixer (BioMicro Systems, Inc.) and transferred to the preheated Maui hybridisation system (BioMicro Systems, Inc.). Mixing was started and arrays incubated overnight for 16 h at 50°C. Microarray slides were then given three post-hybridisation washes. The first wash was in 50 ml of 1 × SSC, 2% SDS for 5 mins with constant agitation; followed by a second wash for a further 5 mins (with constant agitation) in 50 ml of 0.5 × SSC, 2% SDS; then a third wash for 10 secs in 50 ml of 0.5 × SSC. Residual liquid was removed by centrifugation at 1000 rpm for 5 mins at room temperature, and the slides stored in the dark prior to scanning.

Scanning and image processing

Hybridised arrays were scanned in an Agilent DNA Microarray Scanner with Surescan High-Resolution Technology with 5 µm resolution. Each array was scan at 20%, 50%, and 100% gain settings in order to determine and select the optimal scan setting producing a high dynamic signal range without saturation (Forster, Roy, and Ghazal, 2003). Images from all scans were quantified and background corrected using BlueFuse for Microarrays v3.6 (7145). Raw microarray data (including microarray design; hybridisation and analysis) as well as curated data will be available at EnvBase, the NERC Environmental Genomics Data Catalogue, (http://envgen.nox.ac.uk/) and also at the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo).

Overall host vs virus transcript signal comparison

Trimmed means were calculated for the 5 replicate gene spots on each array. These were then normalised into 10 quantile categories, given ordinal factors from 0 to 10 (for example, 3 meaning the normalised and trimmed mean signal for that gene is in a percentile

range between 30 and 40% of the intensities on a given array). Overall percentile position was then calculated for both *E. Huxleyi* and EhV spots, respectively.

Pre- vs post-viral takeover transcript analysis

A second analysis was performed in order to identify genes that consistently changed expression from pre to post infection stages. Based on the overall host *vs* virus overall transcript variation (described above), two samples, corresponding to pre (T0) and post (T1) viral takeover, were chosen for each mesocosm enclosure (**T0**: day 8 / encs. 1 to 6; **T1**: day 14 / encs.1 and 4; day 12 / enc. 2; day 15 / encs. 2 and 5; day 16 / enc. 6). Overall fluorescence values between different microarray chips were normalized using a robust multi-array average analysis (RMA) (Irizarry et al., 2003). After, T0 and T1 median expression values were calculated for each probe. Up-regulated calls for each gene were generated on the basis of a Log_2 (T1/T0) detection threshold above 1.

Sequence analysis and annotation

E. huxleyi EST sequences represented in our microarray were searched against UniProt protein sequence database (The UniProt, 2009) using BLASTX (Altschul et al., 1997) with an E-value cutoff of $1e^{-3}$. We also extracted all possible stop-to-stop open reading frames (≥ 50 aa) from EST sequences. The amino acid sequences derived from these ORFs were used to search against NCBI/KOG database (Koonin et al., 2004) using RPS-BLAST (Altschul et al., 1997) with an E-value cut-off of $1e^{-5}$. Regarding coccolithovirus gene annotation the data was retrieved from NCBI genbank (http://www.ncbi.nlm.nih.gov/Genbank).

Comparison with qPCR data

The same RNA preparations used for microarray hybridization were analysed, using qPCR techniques, to confirm transcript abundance fluctuation of three *E. huxleyi* and EhV genes, respectively. Those works are described in detail in Pagarete et al. (2009, see Chapter 4). Briefly primers were designed to target *E. huxleyi's* β-tubulin and coccolithovirus major capsid protein genes, as well as two key homologous sphingolipid pathway genes present in both virus and host: the serine palmitoyl transferase, and dihydroceramide desaturase (Table

S1). Quantitative real-time PCR assays were performed in a total volume of 20 µl containing 4 µl of the above-described cDNA, 0.04 µM each of the 3′ and 5′ primers, 10 µl of Absolute™ qPCR Sybr® Green Rox Mix (Thermo Fisher Scientific - Abgene, France), and 4.4 µl of DNase free water. The amplification protocol was 40 cycles of 95 °C for 15 sec followed by 60 °C for 1 min. All real-time PCRs were performed with a CFB-3240 Chromo4 Detection System (Biorad). For accuracy every sample measurement was performed in duplicate, and ran against a non-reverse-transcribed (-RT) control to evaluate percentage of DNA contamination. Based on the raw SYBR Green I fluorescence data, the application LinRegPCR v11.1 (Ruijter et al., 2009) was first used to calculate individual PCR efficiencies for each sample, followed by an estimation of mean PCR efficiency (E) and adequate fluorescence threshold (F) for each gene. Ct values were then used to calculate relative expression values (N) for each qPCR measure according to the basic equation for PCR amplification [N=F/(E^Ct)]. The final expression value for each gene in each time point was obtained by the difference between the respective sample and –RT control values (the latter corresponding to DNA contamination that should not be accounted for in an estimation of gene expression). For each gene the lowest measure of expression was taken as the minimum level of detection. Transcript abundance for each host and viral genes was then normalized to the respective minimum level of detection, and finally normalised to the abundance of *E. huxleyi* cells (previously enumerated by flow cytometry).

4. Results

General bloom/infection dynamics

Details on *E. huxleyi* bloom and consequent coccolithovirus infection are described in (Kimmance et al., in press, see Annexe B). Briefly days 7 – 13 of the study were characterised by exponential growth of the *E. huxleyi* population (Fig. S1). A clear split was observed between P-replete (encs. 1,3,5) and P-deplete (encs. 2,4,6) treatments, with maximum cell concentrations reaching 170,944 ml^{-1} in the P-replete treatment. *E. huxleyi* exponential growth phase was followed by sharp decline. That decline coincided with the exponential increase of coccolithoviruses from day 12 onwards. A maximum concentration of 3.1×10^{7} coccolithoviruses ml^{-1} was registered in enc. 3 on day 16. *E. huxleyi* decline to pre-bloom abundance following the appearance of coccolithoviruses went to completion in encl3, but

was not observed in the other enclosures, due to the termination of experimental sampling. In the P-deplete treatment, *E. huxleyi* cells showed lower growth rates with maximum cell concentrations only reaching 7×10^4 ml^{-1}. Coccolithovirus concentrations also remained lower in the P-deplete enclosures, with a registered maximum of 1×10^6 capsids ml^{-1}.

Based in this bloom dynamics analysis, RNA extractions and transcription analyses were performed on samples taken from days 8 to 16.

Overall host versus virus transcript abundance

The microarray used herein contained a total of 3571 gene probes; 2271 (63.6%) matching *E. huxleyi* ESTs, and 1300 (36.4%) matching EhV-86 and EhV-163 genomic sequences. Presence of transcript message was detected for the large majority of the probes present in the array (empty spots were taken as reference for minimum detection threshold), independently of having host or viral origin. Their relative proportions changed throughout the experiment. A quartile rank based analysis of *E. huxleyi versus* EhV global gene expression (Fig. 1) showed the same similar pattern of viral transcript activation in all 6 replicate enclosures. Overall EhV transcript abundance made an increasing progression along the time lapse of the experiment, accompanying its host's global community development, from initial pre-bloom stages towards late bloom decline. Several minor peaks of viral signal were observed in some enclosures (day 10 / encs. 5 and 6; day 11 / enc. 2; day 12 / enc. 4) that preceded a moment, observed in all 6 enclosures, where major overall viral transcript production was triggered (day 11 / enc. 3; day 12 / enc. 1; day 13 / encs. 2, 4 and 5; day 15 / enc. 6). This moment of viral takeover comprised at least two distinct stages: (1) major viral transcript activation (enc. 3 / day 11; encs. 1 and 2 / day 12; encs. 4 and 5 / day 13; enc. 6 / day 14) which happened 24-48h before (2) host cell number reach maxima and start of bloom decline (enc. 3 / day 12; encs. 1 and 2 / day 13; enc 5 / day 14; encs. 4 and 6 / day 15). This second stage also coincided with major viral capsid release.

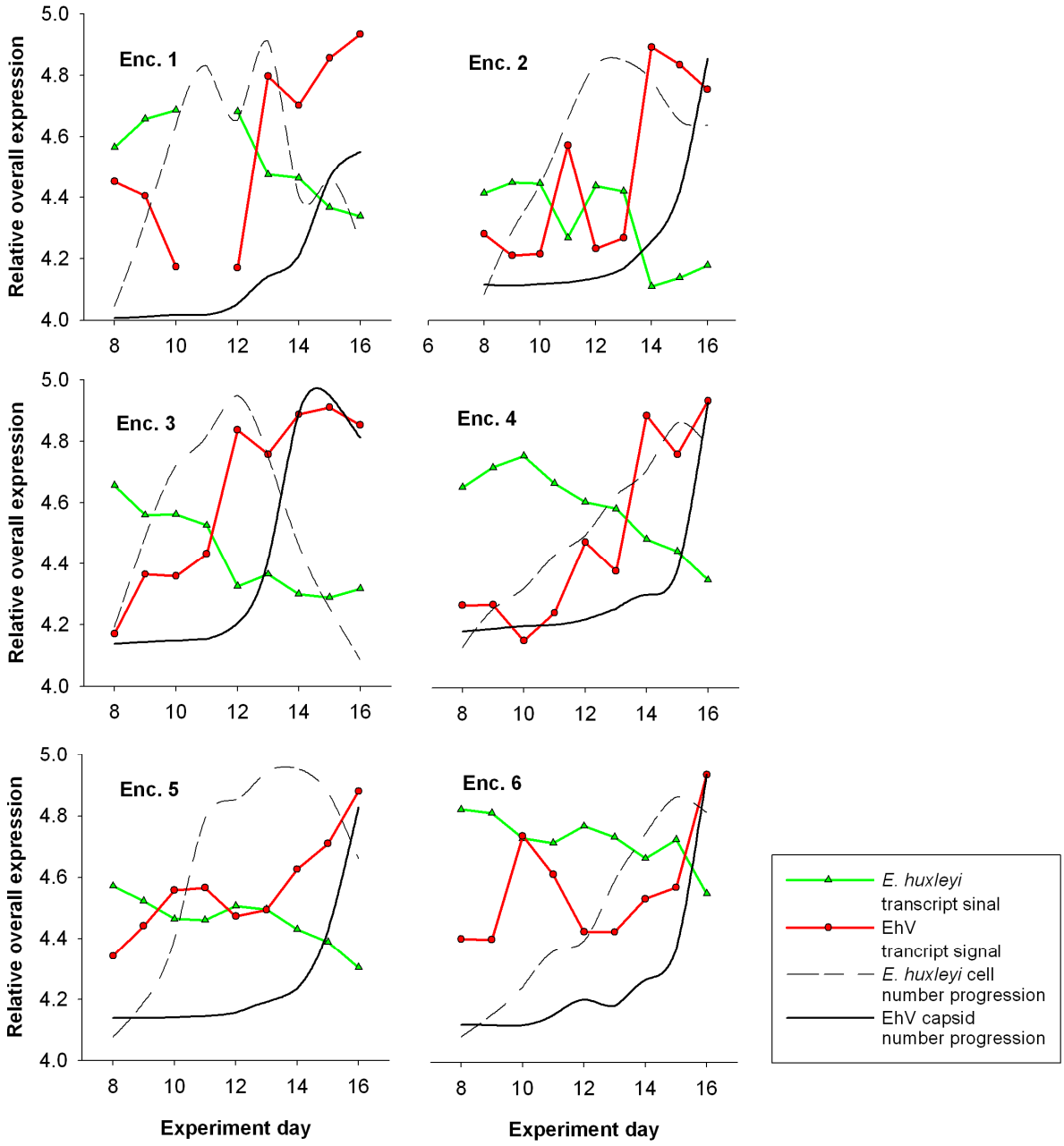

Figure 1. Relative overall progression of *E. huxleyi* and coccolithovirus transcript signal. Y axis scale corresponds to the average quartile position occupied by overall host and virus probe fluorescence signal, respectively (please refer to the Materials and methods section for further explanation). Relative cell and virus number progression is plotted for reference (detailed numbers presented in Fig. S1).

qPCR gene expression analysis specifically targeting three *E. huxleyi* genes (beta-tubulin, serine palmitoyltransferase, and dihydroceramide desaturase) and three EhV genes (major capsid protein, serine palmitoyltransferase, and dihydroceramide desaturase) corroborated both host and viral global transcription patterns (Fig. S2), and confirmed the moment in each enclosure when major viral takeover occurred.

Gene transcription analysis before vs after viral takeover

Based on the overall host *vs* virus transcript variation (described above), two samples, representing pre (T0) and post (T1) major viral takeover, were chosen for each mesocosm enclosure (**T0**: day 8 / encs. 1 to 6; **T1**: day 14 / encs.1 and 4; day 12 / enc. 2; day 15 / encs. 2 and 5; day 16 / enc. 6).

The great majority of probes used in this microarray corresponded to genes for which function is not yet identified (74%); 41% of these refer to *E. huxleyi* genes and the other 33% to EhV. Comparison between T0 and T1 revealed a total of 279 (7.81%) probes presenting at least two fold transcript abundance increase from pre to post major viral takeover (Fig. 2). This gene up-regulation did not occur randomly throughout the entire microarray (Fig. 3). The group of EhV genes without predicted function increased its presence in the pool of up-regulated genes (33 to 52%). At the same time *E. huxleyi* genes, both with and without predicted function, were less well represented in the final up-regulated pool (23 to 15% and 41 to 30%, respectively).

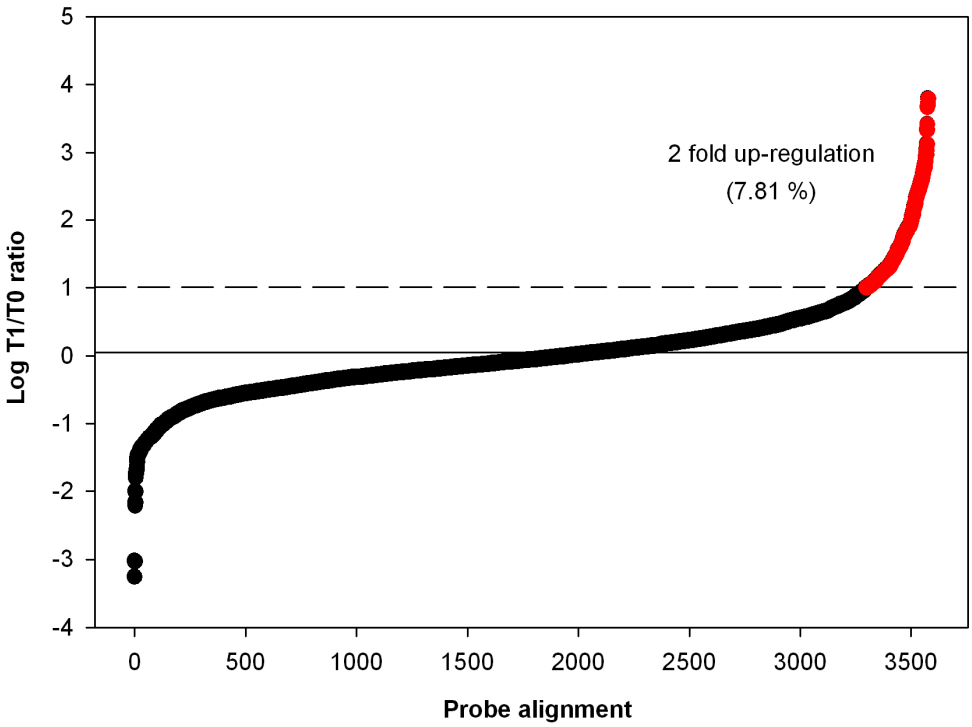

Figure 2. Gene expression ratio between post (T1) and pre (T0) major viral takeover. X scale refers to probe alignment according to Log_2 (T1/T0). Red marks correspond to genes with an up-regulation of at least 2 fold (7.81% of the total targeted probes).

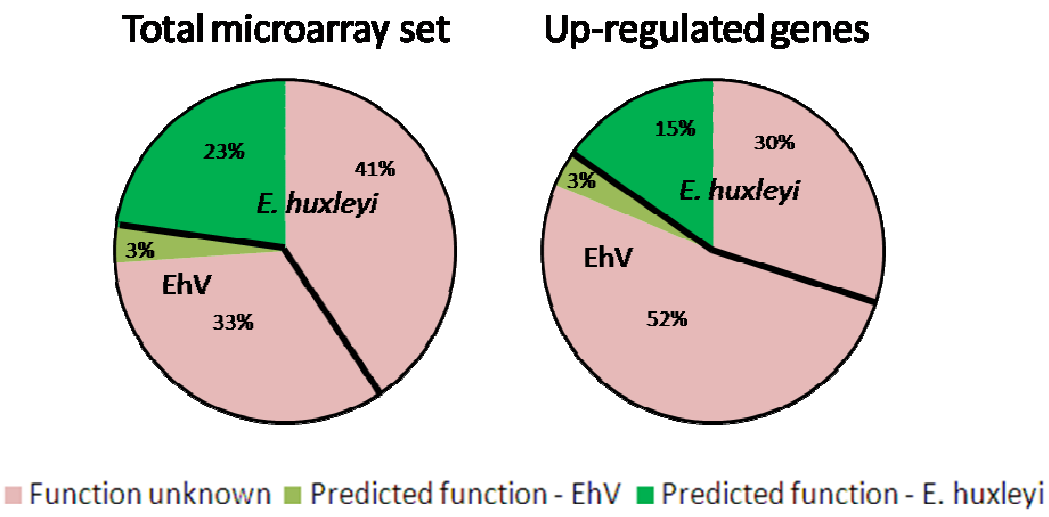

Figure 3. Percentages of total *E. huxleyi* and EhV probes present in the microarray (pie chart on the left) and pool of up-regulated genes during viral takeover (pie chart on the right).

Genes with a predicted function that were consistently utilized during EhV infection represented three major KOG groups: (1) cellular processes and signalling, (2) information storage and processing, and (3) metabolism (Table 1). The majority refers to *E. huxleyi* (15% of the total) while only 3% come from the EhV genome (Figs. 3 and 4). Among these, genes related to metabolism (nucleotide, amino acid, and lipid) were present in large scale (around 10% of the total probes present in the microarray). Regarding information storage and processing many different genes involved in nucleotide replication, RNA processing and translation were identified. Cellular processes and signalling were also represented, namely through metacaspases (involved in cell cycle control and apoptosis), vesicle related proteins (involved in intracellular trafficking), actin and related proteins (involved in cytoskeleton), a mitogene-activated kinase (presumably involved in signal transduction mechanisms), and several genes involved on posttranslational modification (Table 1).

Pre to post viral takeover comparison showed 154 viral genes for which relative abundance increased above the established 2 fold threshold. The huge majority of these viral sequences correspond to unannotated genes (94%). Among the minority that have a predicted function we observed genes related to: posttranslational modification, protein turnover and chaperones (including serine protease and thioredoxin); nucleotide metabolism, transcription, replication and repair (including DNA polymerase and topoisomerase); lipid metabolism (including serine palmitoyltransferase); inorganic ion transport represented by a phosphate permease gene. A viral lectin protein was also identified (Fig. 4, Table 1).

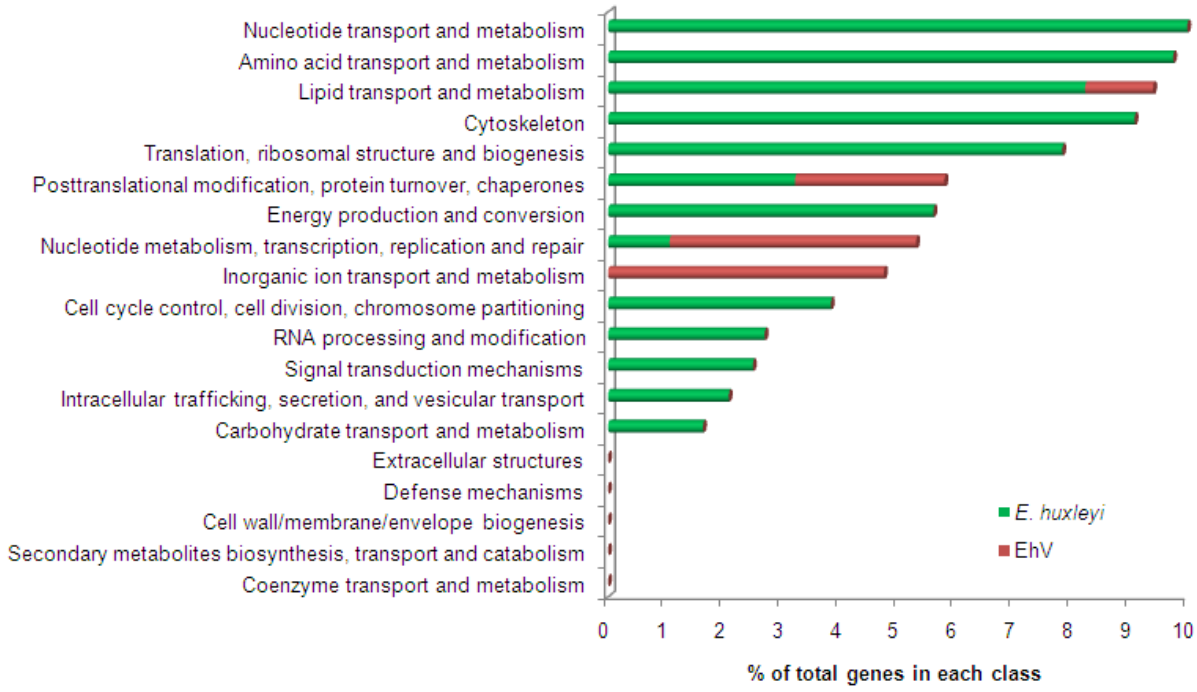

Figure 4. Percentage of *E. huxleyi* ane EhV gene up-regulation per KOG functional class.

Table 1. List of up-regulated genes for which a function can be predicted.

KOG group	KOG class	KOG definition	id[1]	Host/Virus
CELLULAR PROCESSES AND SIGNALING	Cell cycle control, cell division, chromosome partitioning	Metacaspase involved in regulation of apoptosis	KOG1546	*E. huxleyi*
	Cytoskeleton	Actin and related proteins	KOG0676	*E. huxleyi*
	Intracellular trafficking, secretion, and vesicular transport	Vesicle coat complex COPI, alpha subunit	KOG0292	*E. huxleyi*
	Posttranslational modification, protein turnover, chaperones	26S proteasome regulatory complex, ATPase RPT4	KOG0651	*E. huxleyi*
		Aspartyl protease	KOG1339	*E. huxleyi*
		ERV1/ALR family protein	NA	EhV
		Molecular chaperones GRP78/BiP/KAR2, HSP70 superfamily	KOG0100	*E. huxleyi*
		DnaJ domain-containing protein	K09528	EhV
		Serine protease	Not defined	EhV
		Thioredoxin	KOG2501	EhV
	Signal transduction mechanisms	Mitogen-activated protein kinase	KOG0660	*E. huxleyi*
INFORMATION STORAGE AND PROCESSING	Nucleotide metabolism, transcription, replication and repair	5'-3' exonuclease HKE1/RAT1	KOG2044	*E. huxleyi*
		DNA polymerase delta catalytic subunit	COG0417	EhV
		DNA topoisomerase	COG0550	EhV
	RNA processing and modification	mRNA splicing factor ATP-dependent RNA helicase	KOG0923	*E. huxleyi*
	Translation, ribosomal structure and biogenesis	40S ribosomal protein S11	KOG1728	*E. huxleyi*
		40S ribosomal protein S13	KOG0400	*E. huxleyi*
		40S ribosomal protein S25	KOG1767	*E. huxleyi*
		60s ribosomal protein L34	KOG1790	*E. huxleyi*

		60S ribosomal protein L5	KOG0875	*E. huxleyi*
		Glutaminyl-tRNA synthetase	KOG1148	*E. huxleyi*
		Mitochondrial/chloroplast ribosomal protein S2	KOG0832	*E. huxleyi*
		Nucleolar RNA-binding protein NIFK	KOG4208	*E. huxleyi*
		Predicted mitochondrial/chloroplast ribosomal protein S17	KOG1740	*E. huxleyi*
		Translation elongation factor EF-1 alpha/Tu	KOG0052	*E. huxleyi*
		Ubiquitin/40S ribosomal protein S27a fusion	KOG0004	*E. huxleyi*
METABOLISM	Amino acid transport and metabolism	ATPase component of ABC transporters	KOG0062	*E. huxleyi*
		Glutamine synthetase	KOG0683	*E. huxleyi*
		Glycine cleavage system H protein (lipoate-binding)	KOG3373	*E. huxleyi*
		Glycine/serine hydroxymethyltransferase	KOG2467	*E. huxleyi*
	Carbohydrate transport and metabolism	3-phosphoglycerate kinase	KOG1367	*E. huxleyi*
	Energy production and conversion	Aldehyde dehydrogenase	KOG2450	*E. huxleyi*
		ATP synthase F0 subunit 6 and related proteins	KOG4665	*E. huxleyi*
		NADH-dehydrogenase (ubiquinone)	KOG2495	*E. huxleyi*
		Vacuolar H+-ATPase V0 sector, subunit d	KOG2957	*E. huxleyi*
		Vacuolar H+-ATPase V1 sector, subunit A	KOG1352	*E. huxleyi*
	Inorganic ion transport and metabolism	Phosphate permease	COG0306	EhV
	Lipid transport and metabolism	17 beta-hydroxysteroid dehydrogenase type 3, HSD17B3	KOG1014	*E. huxleyi*
		3-oxoacyl-(acyl-carrier-protein) synthase (I and II)	KOG1394	*E. huxleyi*
		3-oxoacyl-(acyl-carrier-protein) synthase (I and II)	KOG1394	*E. huxleyi*
		Delta 6-fatty acid desaturase/delta-8 sphingolipid desaturase	KOG4232	*E. huxleyi*
		Hydroxyacyl-CoA dehydrogenase/enoyl-CoA hydratase	KOG1683	*E. huxleyi*
		Serine palmitoyltransferase	COG0156	EhV
		Squalene synthetase	KOG1459	*E. huxleyi*
	Nucleotide transport and metabolism	Phosphoribosylformylglycinamidine synthase	KOG1907	*E. huxleyi*
NA	Precise function unknown	Lectin protein	Not defined	EhV

[1] COG refers to NCBI's list of **C**lusters of **O**rthologous **G**roups of proteins; KOG refers to the list Eu**K**aryotic **O**rthologous **G**roups, which is a eukaryote-specific version of COG.

5. Discussion

The microarray set used in this study (based on known sequences from one *E. huxleyi* and two EhV strains) was capable of detecting a wide range of host and virus transcripts present in the natural environment. In our view this could only be achieved if there was high genomic identity among strains of *E. huxleyi* and EhV, respectively, present in the natural environment. Natural *E. huxleyi* communities have previously been reported to be genetically rich, but still highly conserved (Martinez-Martinez et al., 2006b). Moreover DGGE based studies performed during this same mesoscosm experiment demonstrated that at least five GPA genotypes could always be detected (Sorensen et al., 2009), which corresponded to *E. huxleyi* sequences previously known to occur in these fjords (Martinez-Martinez et al.,

2006b). If finding such consistency was not so surprising when considering a eukaryotic species like *E. huxleyi*, that was not the case regarding coccolithoviruses. An increasing number of studies have highlighted the huge diversity observed among environmental viruses (Breitbart et al., 2004; Breitbart et al., 2007; Suttle, 2005a). However, using a microarray set based on expressed sequences from only two EhV strains, we obtained extensive hybridization obtained for the great majority of the probes present. We hence think that our results, while not denying the existence of high genomic diversity levels among the EhV natural communities, at least reflect the existence of highly conserved regions in their genomes. This idea is corroborated in the coccolithovirus genomic diversity study produced by Allen and co-workers (Allen et al., 2007).

In the six mesocosm replicates enclosures under analysis we observed a clear moment of viral takeover. The 1-3 day delay observed between enclosures was already expected and can be attributed to the typical variability inherent to mesocosm experiments (Martinez-Martinez et al., 2006a). The clear transcriptomic response observed was far from expected when we consider the natural community approach used in this experiment. Specific *E. huxleyi* and EhV RNA transcripts were targeted among a huge and diversified pool of other transcripts being produced by the global planktonic community. This means that the precise moment of viral takeover consistently observed has at least two major implications. First, it reflects the outstanding community overtake produced by *E. huxleyi* cells during bloom stages (maximum numbers reaching 1.7 cells ml^{-1}). If this was not the case then any transcriptomic message being produced inside *E. huxleyi* cells would most probably be "diluted" in the pool of total transcripts present in the water. Second, it corroborates the idea that bloom/infection processes are regulated by strict constraints and stages, ultimately resulting in the separation of viral capsid production from their release (Pagarete et al., 2009). Hence the precise moment of viral takeover observed in every enclosure was probably the result of an extensive and synchronized viral activation throughout the global *E. huxleyi* "blooming" community.

This study also demonstrated that several gene functions are clearly being utilized on a large scale during coccolithovirus infection. EhV-86 possesses a very complex and "rich" genome with 472 estimated open reading frames (Wilson et al., 2005b). Given this genetic richness one could hypothesize that these viruses have an infection strategy that is mostly independent from their host's machinery. In that case a broad replacement of host cellular machinery by a viral one would be instigated imediately upon infection, with the host cell functioning as a closed "suited" environment for infection to take place, instead of a complex machine of cellular tools at viral disposal. That first scenario was not verified in our study. A

large panoply of host gene functions remained activated or were up-regulated in the course of infection. At this stage we cannot distinguish viral gene activation for infection purposes from host mediated responses to viral infection. Nonetheless the utilization of numerous host genes during infection indicates that a complex infection strategy is used by the coccolithovirus, one that faces/demands the balanced use of intricate host machinery.

One of the major "priorities" for a virus during infection is to replicate and translate its genetic information into newly synthesized viruses. In this study we observed the recruitment of numerous amino acid and nucleotide metabolism related genes, most probably in response to that need for genetic information processing. Several transcription related EhV encoded genes were highly transcribed during infection, such as viral DNA polymerase and topoisomerase. At the same time numerous *E. huxleyi* genes related to m-RNA splicing, ribosomal function and translation were also clearly up-regulated. Further down the line several posttranslational processes were also consistently required, with virus and host sharing potentially interlinked functions namely through the use of protein cleavage enzymes and protein chaperones. The recruitment of EhV ERV_1/ALR family proteins along with thioredoxin could indicate the use of a viral cytoplasmic pathway of disulfide bond formation, eventually similar to what has been verified in other nucleo-cytoplasmic large dsDNA viruses (NCLDVs), such as the Poxviridae (Senkevich et al., 2000).

Other transcription features were identified for the first time in the course of this natural EhV infection. For example, the manipulation of cell cytoskeleton machinery pathways (through actin transcription regulation). To avoid molecular crowding caused by high organelle and protein concentration several viruses require active mechanisms for directed transport inside the cell (Radtke, Dohner, and Sodeik, 2006). Among the large NCLDV group, herpesviruses and poxviruses, for example, have been shown to use the host actin and microtubule transport systems for several steps during their life cycle (Forest, Barnard, and Baines, 2005; Schramm and Locker, 2005). If the need for active intracellular transport mechanisms is related to the size of the particles in question, then this would probably be of crucial importance for a "giant" virus such as EhV (capsid diameter between 160-180 nm). Moreover, active transport processes demand energy production and conversion. This could explain the up-regulation verified for genes like ATP synthase, NADH-dehydrogenase or vacuolar H+ ATPase. Viruses are also known to use cellular "highways" for more distant displacements, either as hitchhikers inside cellular vesicles or organelles, or as cytosolic viral complexes that hijack microtubule motors directly (Smith and Enquist, 2002). In that regard up regulation of *E. huxleyi*'s vesicle coat complex (COPI) could

probably be related to coccolithovirus use of vesicle based intracellular trafficking mechanism for capsid displacement. This would correlate to the recent observations of EhV-86 capsid migration inside the cell (Mackinder et al., 2009). Even more surprisingly this same study argues that coccolithoviruses have an infection mechanism different from that employed by other algal NCLDVs, with entry and exit strategies showing a greater analogy to animal-like NCLDVs. This is based on observations of EhV-86 entering its host via either an endocytotic or an envelope fusion mechanism, while presenting also a budding mediated mechanism for virus progeny release. In our study several lipid transport and metabolism genes were identified, which could also be involved in membrane and vesicle related capsid transport. Although lipid function during EhV infection is still poorly understood, it is becoming more and more evident that lipids play a fundamental role (Vardi et al., 2009).

With regards to lipid metabolism there is a particular group of EhV genes that has received particular attention. The EhV genome encodes a unique de novo sphingolipid pathway (Wilson et al., 2005b). These genes are predicted to be fully functional and expressed during infection (Han et al., 2006; Pagarete et al., 2009). Sphingolipids are membrane lipids whose function has often been implicated in signal transduction mechanisms (Hannun and Obeid, 2008). Moreover, ceramide, the final product of this pathway (Merrill, 2002), is recognized as a major actor in the control of apoptosis (Pettus, Chalfant, and Hannun, 2002). This has lead to hypotheses regarding a possible viral involvement in the control of *E. huxleyi*'s programmed cell death, a situation that has been reported in other host/virus systems (McLean et al., 2008). In this study we observed the clear recruitment of the first and rate limiting enzyme of this pathway, serine palmitoyltransferase. Along with it we also observed the clear up-regulation of *E. huxleyi*'s metacaspases, which are also key precursors in activation of apoptosis (Thornberry and Lazebnik, 1998). This data is in accordance with recent studies that demonstrated not only caspase activation upon EhV infection and presence of caspase cleavage recognition sequences within virally encoded proteins (Bidle et al., 2007), as well as a clear relation between glycosphingolipid production and induction of the lytic infection in oceanic water masses (Vardi et al., 2009).

It was known from the start that a microarray approach like the one presented here does not allow to conclude on the definitive functions associated with these genes, but it retrieves important indications for future studies. We based our interpretation of possible gene function on (1) the current knowledge we have of the *E. huxleyi* / EhV model and (2) the known examples from other host/virus systems. The scenario here presented remains hence our interpretation (even if educated) of the most probable functions these genes might be

playing during coccolithovirus infection. Moreover, we are aware that a substantial fraction of the total RNA transcript present in the environment at all stages of the experiment does not correspond to genes being directly activated by the viral machinery. Those would mainly correspond to *E. huxleyi* genes related to the "normal" functioning of the cell being utilized by uninfected cells, as well as host genes being utilized in infected cells as a means of host response to viral infection. Nevertheless, the fact that the genes here highlighted substantially increase transcript abundance from pre to post viral takeover makes us more confident to believe that these genes are in some way regulated during coccolithovirus infection.

As a final note we would like to remark the fact that the list of genes here highlighted for which relative abundance significantly increased during EhV infection corresponds still to the tip of an iceberg. Many other viral and host genes are surely in use during that process, although having more subtle changes in transcript levels which could not be diagnosed using the methods here described. For example using qPCR techniques one could clearly identify the activation of other coccolithovirus specific genes, namely the expected structural major capsid protein gene, and two other genes involved in sphingolipid production (Pagarete et al., 2009). That said one should also keep in mind that the genes presenting very high transcriptional changes throughout EhV infection remain very useful as potential markers for ecological monitoring of coccolithovirus communities. More importantly, they could serve as precious lights in the quest to better understand the processes and genomic demands that are inherent to coccolithovirus infection.

We launched this study with the aim of bringing phytoplankton virology from specific lab/strain approaches back to a wide natural community targeted approach. We did this by creating a microarray set that we expected would still be capable of distinguishing specific synchronized transcriptomic responses happening within *E. huxleyi* natural community during coccolithovirus infection. The data here presented clearly demonstrates the validity of this approach. With a microarray design aimed at both host and virus transcripts we could identify a complex network of metabolic functions that are consistently evocated along a natural coccolithovirus infections. From genetic information processing, passing by posttranslational control, intracellular trafficking mechanisms, and even probable implications in the control of programmed host cell death, the results here showed add to the idea of "viral factory" (Novoa et al., 2005) acting as a spinning wheel of metabolic functions that interact in perfect balance with each other with the ultimate goal of producing the highest number possible of new viral capsids.

8. Supplementary data - Tables

Table S1. *E. huxleyi* and coccolithovirus qPCR primer sequences used in this study.

Protein name	Oligo name	Direction	Position	Oligo Sequence	Product Size
E. huxleyi					
Serine palmitoyl transferase (SPTh)	EX_SPT_F	Forward	1093	ACTGATTTCCTCCGCATGAC	190
	EX_SPT_R	Reverse	1282	CGATGCCAAACGAGTAGATG	
Dihydroceramide desaturase (DCDh)	EX_DCD_F	Forward	244	AAGGCGGAGTGGGCGAAC	90
	EX_DCD_R	Reverse	333	GGCGTGGTAGTAGCGGAAGG	
β-tubulin	Tubβ_Fw	Forward	1037	CGCTGTACGACATCTGCTT	164
	Tubβ_Rv	Reverse	1201	GGAAGGGGATCATGTTGAC	
Coccolithovirus					
Serine palmitoyltranferase (SPTv)	EhV_SPT_F	Forward	1793	AGTCCGGTATCGTCTTGTCG	190
	EhV_SPT_R	Reverse	1982	CGCAATGCGATAATACATGG	
Dihydroceramide desaturase (DCDv)	EhV_DCD_F	Forward	417	GGACATTTCTTCCGTCATGG	153
	EhV_DCD_R	Reverse	569	ACGGTCCAATTTGCAAGAAG	
Major capsid protein (MCP)	MCP1Fw	Forward	255	ACGCACCCTCAATGTATGGAAGG	90
	MCP90Rv	Reverse	344	AGCCAACTCAGCAGTCGTTC	

9. Supplementary data – Figures

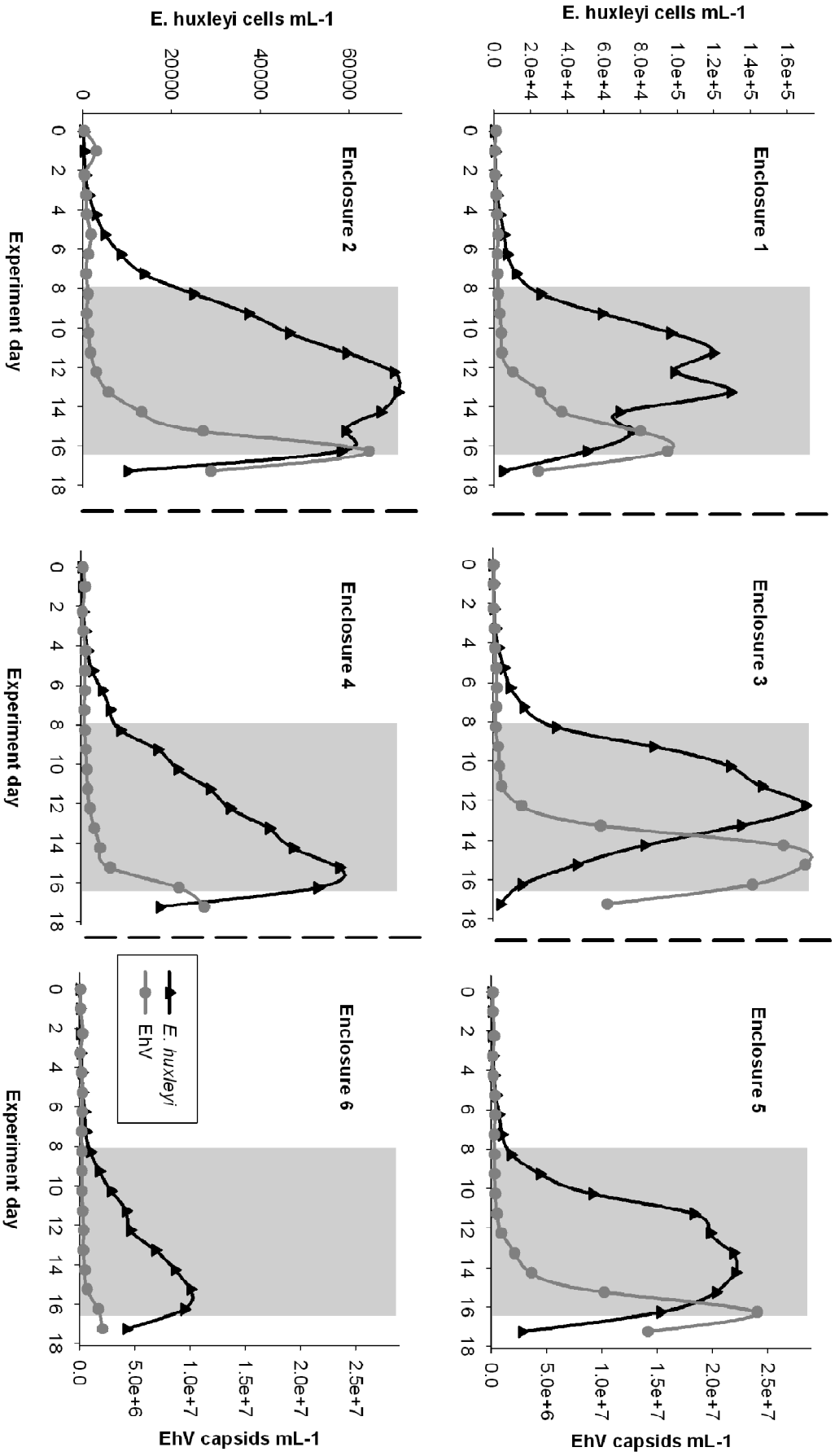

Figure S1. *E. huxleyi* and coccolithovirus abundances in Enclosures 1 to 6 as determined by flow cytometry. Grey sections indicate the samples selected for gene transcription analysis.

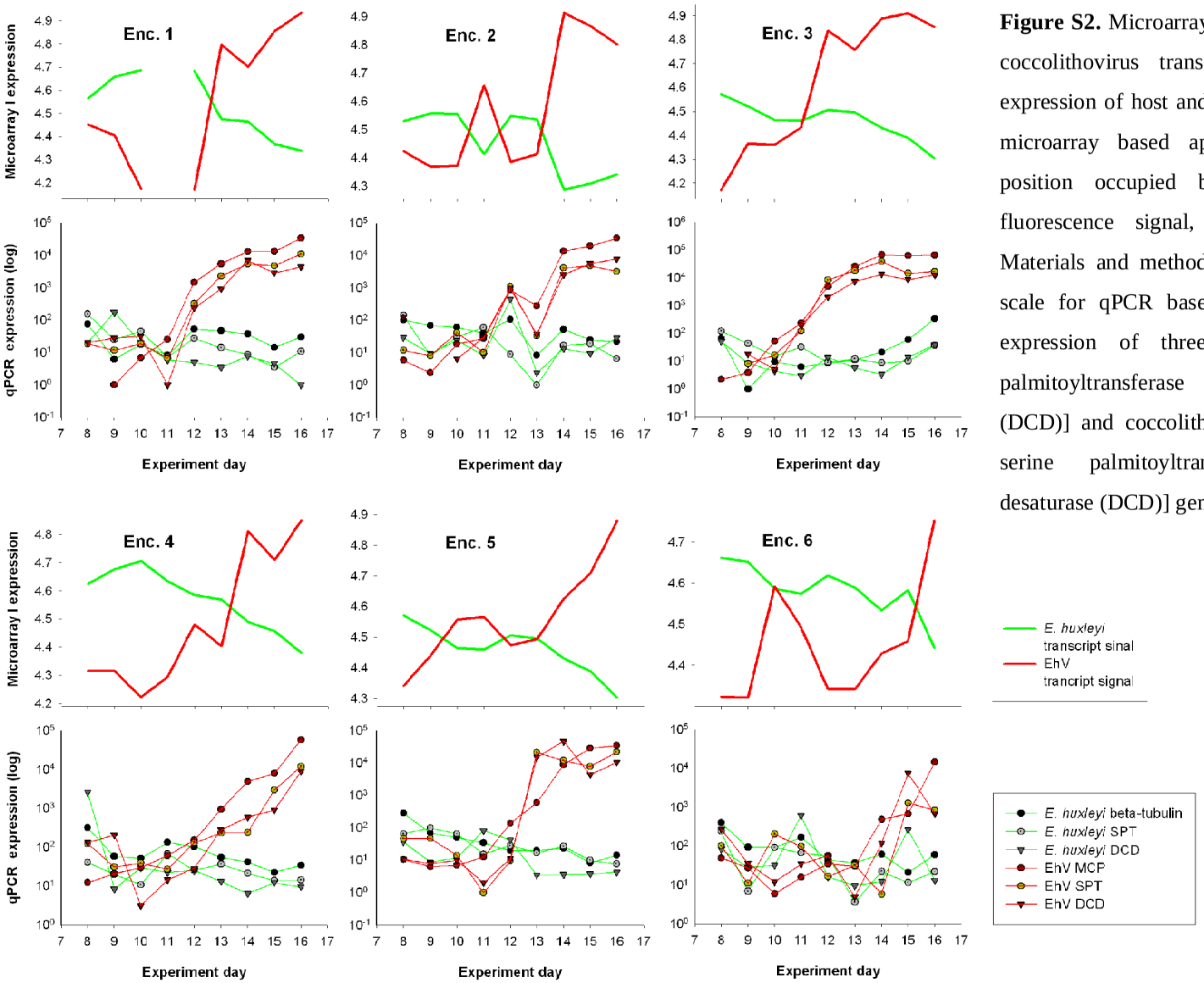

Figure S2. Microarray overall progression of *E. huxleyi* and coccolithovirus transcript signal *versus* qPCR relative expression of host and virus genes, respectively. Y scale for microarray based approach represents average quartile position occupied by overall host and virus probe fluorescence signal, respectively (please refer to the Materials and methods section for further explanation). Y scale for qPCR based approach represents relative gene expression of three *E. huxleyi* [beta-tubulin; serine palmitoyltransferase (SPT); dihydroceramide desaturase (DCD)] and coccolithovirus [major capsid protein (MCP); serine palmitoyltransferase (SPT); dihydroceramide desaturase (DCD)] genes, respectively.

Chapter 6.
Short report on attempts to isolate new coccolithophore viruses

1. Introduction

The discovery phase of marine viruses is still in its infancy as the discovery of new viruses continues to progress at an elevated pace. Along with those discoveries comes the acknowledgement of the major roles played by viruses as regulators of their host's populations (Brussaard, 2004a) and vectors of genetic information (Rohwer and Thurber, 2009; Willner, Thurber, and Rohwer, 2009). Nonetheless the biological processes underlying the evolution and maintenance of oceanic viral diversity are still poorly understood. High host-specificity (i.e. in transmissibility and/or virulence) has been reported, particularly in viruses infecting marine eukaryotic phytoplankton (Brussaard, 2004a). Thus, the continuous arms race between oceanic viruses and eukaryotic microbes may actually be highly specific and localized co-evolutionary processes may explain the rapid diversification and speciation observed in both hosts and viruses.

In recent years, coccolithophores (calcifying haptophyte algae that belong to the prymnesiophyte clade) have become an important model to study the evolutionary interactions of eukaryotic phytoplankton viruses with their hosts. This is based both on the evolutionary traits presented by the coccolithophores, and on the achievement to isolate and characterize a family of viruses that infects one of these species.

Among the marine phytoplankton dominating the modern-ocean, the coccolithophores play critical roles in primary production, gas exchange between the oceans and atmosphere, and carbon export to the deep oceanic layers and deep-sea floor (see Chapter 1 for further details). The sedimentation of the their $CaCO_3$ scales (coccoliths) over the last 200 My has produced one of the best fossil records, making them key biomarkers for paleo-climate and stratigraphic reconstructions, and for the study of marine biodiversity and evolution (Bown et al., 2004). The excellent fossil framework of coccolithophores, the presence of sister-species that can be recognized morphologically, and the alternation between haploid and diploid population of cells that can be isolated in culture, combined with the growing knowledge about coccolithophore viruses (Allen et al., 2006c; Schroeder et al., 2003; Schroeder et al.,

2002; Wilson et al., 2005b), make this group an ideal model to test host/virus evolution hypotheses.

Several large dsDNA viruses that infect algal hosts have already been isolated and all classified into a distinct family, the Phycodnaviridae. Among the phycodnaviruses we find the coccolithoviruses (type species Emiliania huxleyi virus 86). To date around 16 strains of coccolithoviruses (EhV's) have been isolated and characterized, all infecting the most ubiquitous and abundant coccolithophore species *Emiliania huxleyi* (Fig. 1). In nature the coccolithoviruses have been shown to play a crucial role in the dynamics of *E. huxleyi* blooms, community structure, and nutrient cycling (Allen et al., 2006c; Jacquet et al., 2002; Martinez et al., 2007; Schroeder et al., 2002) . The coccolithovirus genome EhV-86 has been completely sequenced, revealing 472 predicted genes (Wilson et al., 2005b).

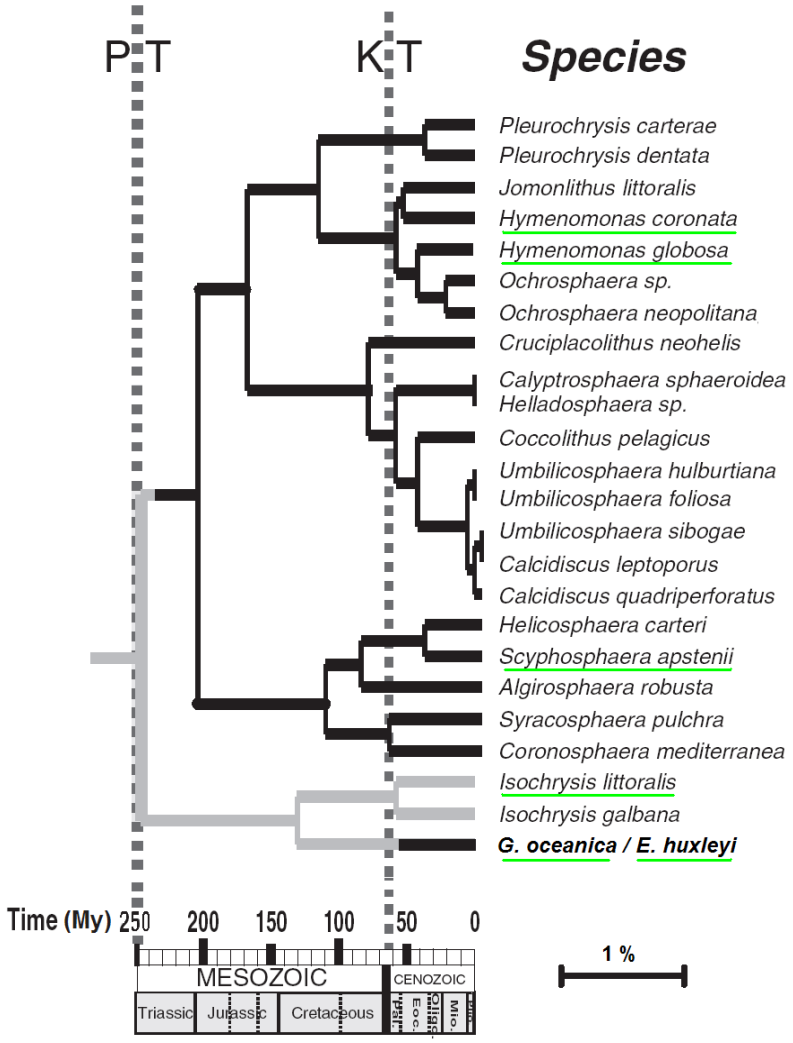

Figure 1. *Emiliania huxleyi* in the coccolithophore phylogenetic tree. Underlined in green are some of the other species that have been used in our attempts to isolate new viruses. Black branches indicate groups where calcification occurs. Tree is based in LSU rDNA sequence data. In this representation *E. huxleyi* occupies the same phylogenetic branch as its sister group *Gephyrocapsa oceanica* given that their divergence is younger than 300.000 years. *Adapted from de Vargas et al. (2007).*

EhV is the only coccolithovirus isolated to date. It clearly occupies a phylogenetic position within the large family Phycodnaviridae, however (and unexpectedly) these viruses are phylogeneticaly distant from the other prymnesiophyte viruses (Genus *Prymnesiovirus*) (Allen et al., 2006c; Larsen et al., 2008). This situation led to the creation of a separate Genus the *Coccolithovirus* (check Fig. 3 in Chapter 1). Two central hypotheses have been raised on the possible origin of EhV: (1) descending from an older lineage of coccolithophore viruses that is isolated from the other known prymnesioviruses, or (2) being the result of an "evolutionary jump" between phylogenetic isolated clades. Whichever was the case mechanisms of co-evolution between hosts and their viruses, where both are fighting to respectively decrease and increase virulence, are certainly playing critical roles in both virus and phytoplankton diversification. What is the role viruses may have played in the diversification of coccolithophores?, have viruses accompanied the diversification pattern of their hosts, and can that be inferred from their phylogenetic trees?

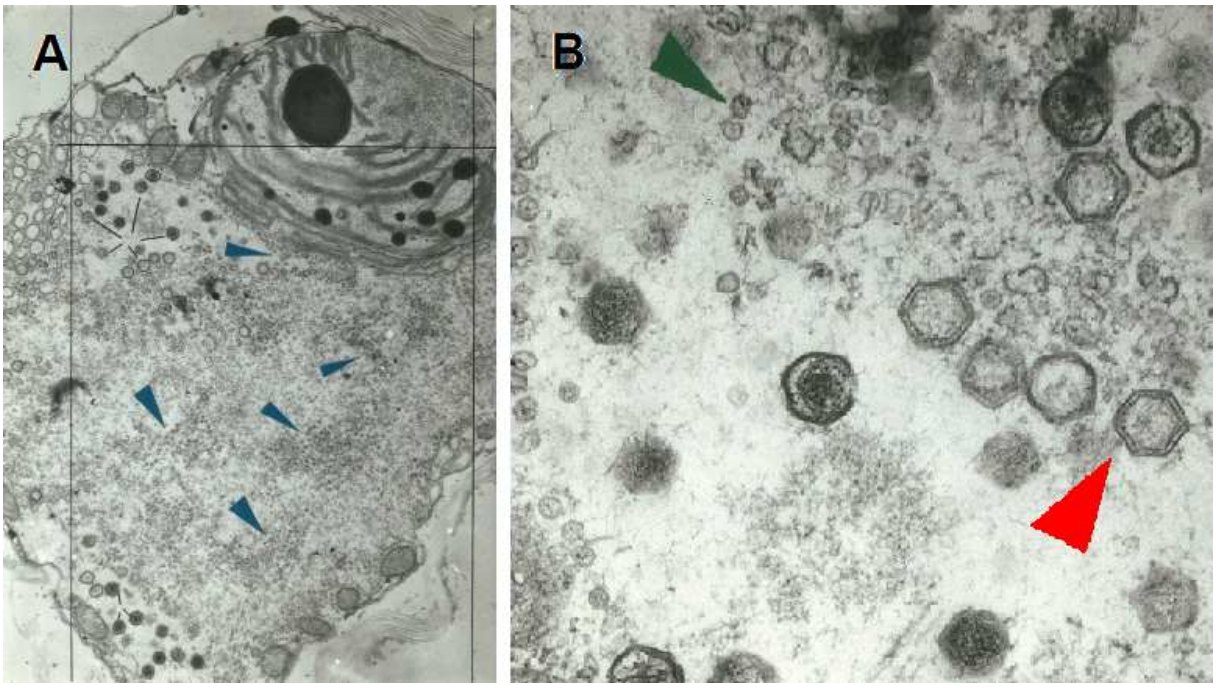

Figure 2. Viral like particles inside the coccolithophore *Hymenomonas globosa* (A) and detail (B). Two types of particles (at least) can be seen in what seems to be a case of co-infection. One group of large particles (red arrow) around 160 nm diameter (clearly resembling other known phycodnaviruses), and a second (more numerous) group of smaller particles (green arrow) around 60 nm in diameter. Electron microscopy images (from 1978) kindly provided by Chantal Billard (University of Caen).

To start answering these questions it will be extremely useful to have access to newly isolated viruses that infect coccolithophore species other than *E. huxleyi*. It has become

almost common knowledge that for every cellular organism probably exists at least one viral infective agent. Finding those new coccolithophore viruses should be just a matter of time. Moreover, in the 70's, Chantal Billard and colleagues (University of Caen, France) reported the presence of large viral particles inside other coccolithophore cells, namely belonging to the genus *Hymenomonas* (Fig. 2).

During the first year of this PhD thesis we performed attempts to isolate new coccolithophore viruses. Here we present the procedures used and the most significant results we obtained during those attempts.

2. Materials and Methods

Host cell culturing

Different coccolithophore species/strains from the Roscoff Culture Collection (http://www.sb-roscoff.fr/Phyto/RCC) were used as potential hosts for the isolation of new viruses. Besides *E. huxleyi*, these included *E. huxleyi*'s sister species *Gephyrocapsa oceanica*, *Hymenomonas globosa* (where viral like particles had previously been identified), among other coccolithophore species (Table 1). The different strains were grown in f/2 (Guillard, 1975), based on aged and filtered (0.22 µm) seawater. The temperature of the cultures was maintained at 16.0 ± 2.0°C, and light was supplied as a light-dark cycle of 16 and 8 h at a photon flux density of 100 ± 20 µmol m^{-2} s^{-1}. Cultures were maintained in exponential growth phase in order to maximise the chances of generating a viral infection upon inoculation.

Water collection and concentration

Attempts were made to infect the coccolithophore cultures with viruses potentially present in new sea water samples. The water samples were collected from the SOMLIT station (English Channel, offshore Roscoff) every two weeks, for 10 months. We also episodically used water samples from Roscoff's shore, the Black Sea , and from the North Pacific, Canadian and Japanese coasts. Each water sample was first filtrated through a 0,45 µm filter in order to separate the viral fraction from all other bigger components. The filtrate was then concentrated using a VivaFlow tangential flow filtration system, 100,000 MWCO PES (Sartorius Stedim Biotech) according to the manufacturer's protocol. The concentrated viral fraction was preserved at 4 ºC until further use.

Table 1. Host coccolithophore species/strains used to attempt new viral infections.

Genus / Species	Strain
Chrysochromulina sp.	A17
Chrysochromulina stipitata	AC4
Dicrateria inornata	AC49
Emiliania huxleyi	MM4-3
Emiliania huxleyi	CCMP1516
Emiliania huxleyi	NS6-2
Gephyrocapsa oceanica	JS1
Gephyrocapsa oceanica	ARC1 2N
Gephyrocapsa oceanica	ARC1 N
Gephyrocapsa oceanica	PC71
Hymenomonas coronata	AC58
Hymenomonas coronata	AC115
Hymenomonas globosa	AC30
Isochrysis litoralis	AC18
Isochrysis sp.	AC80
Isochrysis sp.	AC66
Scyphosphaera sp.	AC504

Flow cytometry was used to characterize the viral fraction present in the filtrate. The methods used followed the protocol described in Marie et al. (1999). Briefly samples were fixed with glutaraldehyde 0.25% (final concentration) for 30 min. followed by fast freezing in liquid nitrogen. The samples were then tawed at room temperature, and incubated with TE buffer (10 mM Tris, 1 mM EDTA [pH 7.5]) and SYBR-green (0.5×10^{-4} final concentration) in the dark for 15 min. Viral counts were done with a FacsCalibur flow cytometer (Becton Dickinson, San Jose, Calif.). Each sample was analyzed for 2 min. at a delivery rate of 50 ml min^{-1}. The discriminator was set to green fluorescence, which is proportional to the nucleic acids–SYBR-I complex. Standard laser settings were: FL1 treshold at 100, SSC 600, FL2 500, FL3 500. Parameters were collected on logarithmic scales and plots were generated with the custom-designed software WinMDI version 2.9 (http://facs.scripps.edu/software.html).

Inoculation and infection check

Infection experiments were run in sterile 24 well plates. For each test 1 ml of exponentially growing culture was inoculated with 50 µl of viral fraction concentrate. This was done in triplicate, and with three negatives controls. The plates were then left to incubate

in different phytoplankton growing chambers with temperatures varying from 15º to 20º C. Possible signs of infection were checked once a day though the use of light microscopy for at least 15 days. Whenever a culture presented signs of infection (lysed cells, or inoculated sample with clear fewer cells than negative control), both positive and negative samples were taken and filtered through 0.45 µm filters in order to recover the viral fraction. The samples were then analysed through flow cytometry (following the same procedure has described above) to check for the presence of distinct viral populations.

Finally any filtrate from such potential infections was used then used to re-infect the same host strain, in an attempt to demonstrate re-infection by the lytic agent.

3. Results

Among all the situations where a potential viral infection was observed, a few presented very promising results (view list in Table 2). The different water samples used to provoke such infections all clearly presented one or several distinct populations of potential large dsDNA viruses (Fig. 3). The sample from the Black Sea was extremely rich, at least 4 distinct groups of large viruses could be recognized (Fig. 3c).

Table 2. List of the different potential viral lysis situations encountered.

Species	Strain	Water sample	Days p.i.	ISOLATE	Room Temp.
Emiliania huxleyi	ARC1 2N	Roscoff shore	6	EHARROSP1	20º C
Emiliania huxleyi	ARC1 2N	Roscoff shore	4	EHARROST1	20º C
Emiliania huxleyi	MT0610B	English Channel	13	EHMTSOM1	20º C
Gephyrocapsa oceanica	NS6-2	Roscoff shore	6	GONSROSP1	20º C
Gephyrocapsa oceanica	NS6-2	Roscoff shore	6	GONSROST1	20º C
Isochrysis sp.	AC80	English Channel	15	ISACSOM2	20º C
Isochrysis littoralis	AC18	Black sea	10	ILACVB1	20º C
Dicrateria inornata	AC49	Black sea	5	DIACVB1	20º C
Hymenomonas globosa	AC30	English Channel	13	HGACSOM1	20º C
Hymenomonas globosa	AC30	Black sea	10	HGACVB1	20º C

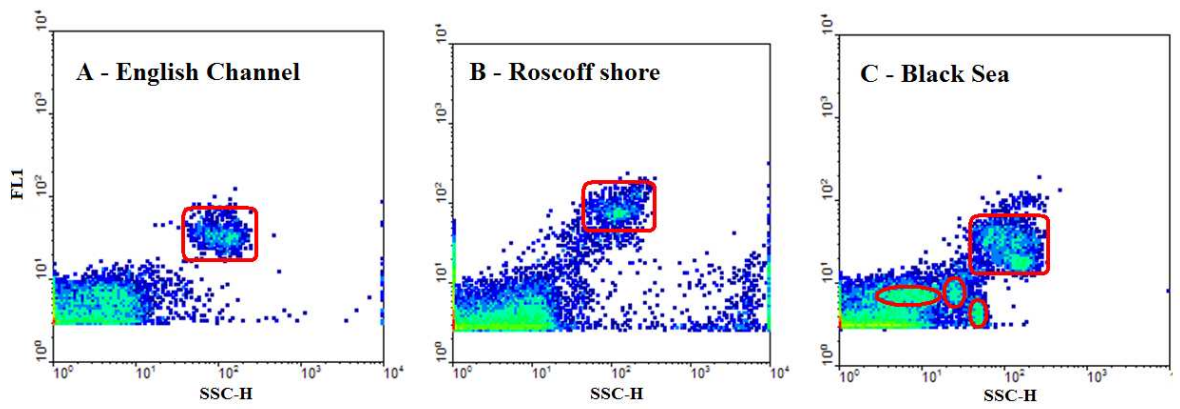

Figure 3. Flow cytometry plots of the viral fractions (< 0.45 um) present in the three water samples that present the most promising infection results. Distinct groups of potential large dsDNA viruses are marked with red circles/boxes.

All the potential infections reported here presented cell lysis (diagnosed using light microscopy and by comparison with a negative control). Posterior analysis using flow cytometry in each case clearly denounced distinct populations of viral like particles present in the < 0.45 µm fraction (Fig. 4). The signatures of the viral like particles obtained tended to appear in the region where large dsDNA viruses usually occur (compare with EhV, Fig. 4a).

Several viral populations could be observed after inoculation with the viral fraction present in water samples from the English Channel (offshore and Roscoff's shore. Three potentially new viral isolates were obtained with *E. huxleyi* (Fig. 4b, c, and d), two with *G. oceanica*, the sister species of *E. huxleyi* (Fig. 4e and f), and also isolates taken from lysed cultures of *Isochrysis* and *Hymenomonas* (Fig. 4g and j). The water samples also proved fruitful in terms of the different viral populations present (Fig. 3). Clear distinct viral signatures were obtained for *Isochrysis*, *Dicrateria*, and *Hymenonomonas* (Fig. 4h, i, k). For *Isochrysis* and *Hymenonomonas* the viral like particles observed had the same flow cytometry signature in both the English Channel and the Black Sea samples.

Upon potential viral lysis, the culture was filtered through 0,2 µm, and the isolate was used to attempt re-infection of the original host strain. None of the isolates here reported was able to maintain the same levels of (potential) infectivity verified in the initial inoculations, since no clear signs of infection could be distinguish after the second or third attempts of re-inoculation.

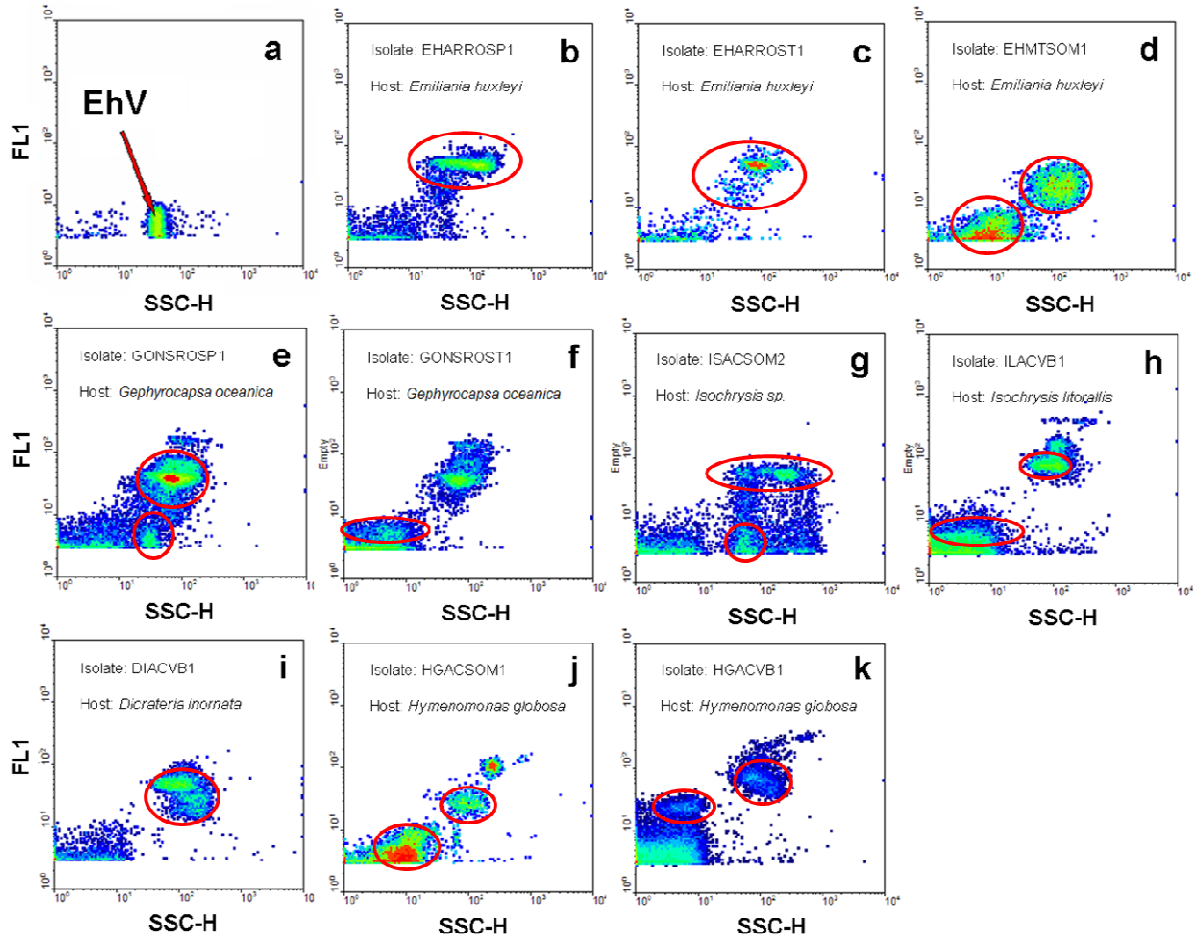

Figure 4. Flow cytometry plots of the isolates that were most similar to viral infections. "a" plot obtained for pure EhV stock shown here for the purpose of comparison. Potential viral groups are marked with red circles/boxes.

4. Discussion

The major objective of this project was to isolate new coccolithophore viruses. Such goal was not achieved given the failure to demonstrate re-infection capacity with the potential new viral isolates. Despite that draw back, these works are far from being uninformative. The location of viral particles on a flow cytometry plot is the result off different factors such as the form or the size of the particles (Marie et al., 1999). In the case of large dsDNA viruses, given their elevated fluorescence (after nucleic acid staining with SybrGreen), they tend to occupy a particularly elevated FL1 (forward laser 1) position that clearly separates them from smaller viruses, namely the (usually) very numerous phages (Brussaard, 2004b). During the course of the infection attempts here reported we came across with lysates that had a flow cytometry signature very similar to that of known large dsDNA viruses (Fig. 4), which lead us to believe

that those lysates most likely to contained large DNA viruses. This comes to corroborate the previous information provided by Billard and co-workers (University of Caen) who had clearly identified (using electron microscopy) large viral particles in coccolithophore species other than *E. huxleyi* (see Fig. 2). To date we do not know yet what is the nature of other coccolithophore viruses (besides EhV), but this report, and theirs, leaves very few doubts that coccolithophores are commonly infected by large DNA viruses.

Moreover, the water samples originating from the Black Sea seemed to be extremely rich in large DNA viruses, namely when compared to the regular samples taken in the English Channel, offshore Roscoff. Not only were the large virus like particles very numerous, they could clearly be identified into several distinct groups. Interestingly those water samples were among the ones providing the most promising results in terms of possible viral isolations. The basic premise to be able to isolate a virus from an environmental sample consists in enhancing the contact between the right host and the right virus. The Black Sea is a rather constrained marine environment, probably leading to some sort of segregated evolution among their phytoplankton communities. It has already been demonstrated that viral impact in the phytoplankton communities is usually very severe (Brussaard, 2004a), leading to a fast selective pressure that clearly affects intra-species succession (Martinez-Martinez et al., 2006a; Martinez-Martinez et al., 2006b). Ultimately this would implicate a tight evolutionary connection between phytoplankton and viruses that infect it. In that sense one could predict that it would be hard to find the right virus/host correspondence when trying to infect phytoplankton cells collected in the North Sea with viruses collected in the Black Sea. Therefore it was with some surprise that we saw that the Black Sea viral community contains viruses capable of infecting coccolithophores in the North Sea.

In what regards EhV, we already knew that strains from the North-East Atlantic region are able to infect *E. huxleyi* cells originating from the most distant locations as New Zealand or the Japanese Sea (data not published). Although being strictly associated with that same one species, it seems clear that the largest marine distances possible in our planet or the possible isolation between oceanic basins are not enough to create an evolutionary barrier that would lead to a localized specialization of EhV. The potential infections obtained between the viral community present in the samples from the Black Sea and the English Channel coccolithophore strains also corroborate this idea. Either phytoplankton host/virus evolutionary pressures are not strong enough to make these viruses specialize to their hosts according to their co-localization, or the flux of viruses and plankton cells between such

(seemingly) isolated marine basins is enough to erase possible traces of localized co-evolution, or both.

We must not forget though that all the potential viral infections observed during these works were not successful in terms of producing a viral isolate capable of re-infecting the same strains. We could clearly observe (at least sometimes) the production of viral capsids on first-time inoculations with complex viral fractions existing in natural water samples. These same viruses however were not capable of maintaining their infectivity after 2 or 3 rounds of re-inoculation. It seems that something is occurring in between inoculation rounds leading to decrease in infectivity. For example, successful infection may rely on a community effect, which is progressively lost as newly formed viruses go through successive infection bottlenecks. A natural sample of filtered water contains a complex community of numerous viruses and bacteria. Regarding our results we should not exclude the possibility that (at least some) successful infections may rely on interactions among different viruses, or maybe even including bacteria. Such enhancing community effect could potentially decrease with progressive bottleneck selection that occurs from one inoculation to the next.

Numerous hypotheses could be drawn to explain the observed incapacity to retain infectivity reported in these works. Although for now this question will stay without response, future studies on the subject might help getting a closer understanding of the real infectious interactions between coccolithophores and their viruses in natural oceanic situations.

5. Acknowledgements

The execution of the works presented here was only possible with the help from numerous sources. I would like to acknowledge Dominique Marie (for his time and patience to teach me the reported flow cytometry techniques, and also Fabienne Jalabert (for her crucial help with the sampling and guidance in the concentration of the water samples). A word also to Dr. Daniela Petrova and Dimitar Borisov (Institute of Fishing Resources Varna, Bulgaria), Dr. Chris Brown (National Oceanic and Atmospheric Administration, USA), Dr. Lisa Eisner (Ted Stevens Marine Research Institute, USA), and Yoshiyuki Ishitani (Department of Earth and Planetary Sciences from Kyushu University, Japan) for their kindness in collecting and sending me viral isolates from the different report locations.

Chapter 7.
Final discussion and perspectives

With new DNA sequencing techniques being developed at an astonishing pace, full genomic sequencing "invaded" the oceanic realm, opening new possibilities for comparative genomics, namely at the level of host-virus interactions. Such techniques allowed Lindell and co-workers (Lindell et al., 2004) to demonstrate for the first time a clear case of horizontal gene transfer (HGT) of photosynthetic genes between *Prochlorococcus* (a cyanobacteria and one of the most abundant members of the phytoplankton communities) and the genomes of three phages from two viral families (Myoviridae and Podoviridae). The possibility of viral control of host photosynthetic apparatus had important implications for viral and host fitness. By acquiring genes from their hosts (through HGT), and eventually transferring these genes back to their hosts after a period of accelerated evolution in the virus, mechanisms of HGT could be influencing the present fitness landscape of hosts and phages in the surface oceans, and at the same time playing a major role in the host-virus evolutionary dynamics.

Having the potential for HGT in mind, we started this thesis by focusing on the coccolithovirus *de novo* sphingolipid biosynthesis pathway (from herein referred to as sphingolipid pathway), and in particular trying to find the answer for two defined questions: (1) what is the origin of these viral genes? and (2) what role do they have during infection?

Using database mining and blast searches we confirmed the existence of the *E. huxleyi* sphingolipid pathway, and discovered a clear homology between host and virus pathways. This was the first clear case of horizontal gene transfer of multiple functionally linked enzymes in a eukaryotic phytoplankton–virus system (see Chapter 3). PCR assays and sequence comparison also indicated that these genes were prevalent in *E. huxleyi* and EhV strains isolated from different geographic locations. Patterns of protein and gene sequence conservation supported the functionality of both host and virus pathways, a hypothesis that came corroborate three studies recently published (Allen et al., 2006; Han et al., 2006; Vardi et al., 2009).

At this stage it was impossible to confirm the direction of the HGT, however the relevance of that discussion is much wider than this host-virus model; indeed, it is directly linked

to the debate on the origin of virus, and viral role in the evolution of cellular life. Growing evidence indicates that all life domains (bacteria, archaea, eukarya, and viruses) are linked by ancestral genetic traits which would report to some form of pre-cellular world (Forterre, 2006), in extreme a last universal common ancestor, LUCA (Forterre, Gribaldo, and Brochier, 2005). Viruses share homologous features across the three domains of cellular life, suggesting that many genes were created in an ancestral virosphere, being later on transferred (by HGT) to the different cellular domains (Forterre and Gadelle, 2009).

Regarding the *E. huxleyi* / coccolithovirus sphingolipid pathway the virus-to-host direction would suggest the existence of ancient viruses that controlled this complex metabolic pathway in order to infect primitive eukaryotic cells. This idea would add to the above mentioned ancestrality of viral genes and their major impact in the creation and development of cellular life (Villarreal, 2005). In this specific case though, we are more inclined to support the host-to-virus direction. First, these sphingolipid genes are ubiquitous in eukaryotes. The virus-to-host scenario would require an additional and earlier HGT for a set of these genes between ancestral eukaryotes and ancestral viruses (e.g., an earlier HGT prior to the divergence of eukaryotic lineages). Furthermore, the branching positions of the EhV / *E. huxleyi* sequences for these genes are globally compatible with a deep phylogenetic origin of *E. huxleyi* within eukaryotes. The virus-to-host scenario would again require an additional (not parsimonious) evolutionary mechanism forcing the sequences of the EhV homologs to be placed near the basis of eukaryotic trees without long branches, a situation often not found when reconstructing phylogenetic trees with viral sequences (Forterre and Gadelle, 2009; Moreira and Brochier-Armanet, 2008).

In our interpretation we considered that, in a host-to-virus gene transfer scenario, the viral acquisition of the sphingolipid enzyme genes was probably gradual, through multiple HGT events, rather than through a single *en bloc* transfer of multiple genes (even if it is difficult to reliably assess the relative timing of these HGT events). The initial acquisition of one of the genes of this metabolic pathway, for instance, the enzyme serine palmitoyl transferase (the first and rate limiting step of this pathway), could have been sufficient for the virus to start modulating its host's life span or lipid profile, thus giving this altered virus a selective advantage on other viral strains. The later acquisitions of additional genes could have further enhanced the viral capacity to modulate the cellular metabolism. This type of serial gene acquisition by a virus could be a possible way to increase its fitness, and hence a driving force in the Red Queen evolutionary race between virus and hosts.

In the recent years a growing number of studies have argued that HGT mechanisms might be playing a central role in the evolution of viruses and their hosts (Hendrix et al., 2000; Koonin, 2009). It is worth mentioning here that in the case of *E. huxleyi* and coccolithovirus the example of the sphingolipid pathway is most likely not alone in what regards HGT. Further similarity searches and phylogenetic reconstruction revealed 35 cases of possible HGTs between *E. huxleyi* and viruses, the vast majority of which (29 of the 35 cases) happening with coccolithovirus (see Annexe A). Besides the sphingolipid pathway, the other *E. huxleyi* / EhV HGTs were associated with DNA/RNA processing, transporting function and other metabolisms. The direction of these HGTs could be in both ways between virus and host, although a host-to-virus scenario seems to be the most parsimonious for many of these genes. Nevertheless, there are also several cases for which close homologs of *E. huxleyi* genes were found only in viruses (and more remotely in bacteria). These latter add to the idea that (also in the case of EhV) viruses might be representing an accessible vast genetic pool capable of influencing cellular evolution through mechanisms of gene transfer.

These results do not contradict in any way the probable ancestrality of viral genes and the existence of a very ancient virosphere that preceded the creation of cellular life. They reflect the complexity of life, and the impossibility we have to establish rules that verify in all cases. It is reasonable to think that many existing genes are reminiscent from an ancient viral world, while many other genes could have a posterior cellular origin. In our opinion, it is of the utmost importance to link this new information on viral-mediated HGT with the fact that that viruses represent the largest genetic pool in our planet (Angly et al., 2006; Suttle, 2005). The accelerated mutation rates usually associated with viral genes and the inherent genetically promiscuous relation they endure with their cellular hosts (which as we are seeing potentially enhances HGT events), makes viruses crucial vectors of genetic information between life domains and an “endless” reservoir of diverse genetic tools at the disposal of cells.

After the initial quest to unveil possible traces of HGT events present in the *E. huxleyi* and EhV genomes, we re-centred our attention on the functionality of the sphingolipid pathway, and more precisely its use in a complex community situation of a natural *E. huxleyi* bloom and EhV infection. A mesocosm environment was hence used to monitor, for the first time, the transcription levels of both host and virus homologous sphingolipid pathway genes in a marine microbial community during temporal progression and decline of an *E. huxleyi* bloom (Pagarete

et al., 2009, see Chapter 4). More specifically we were able to investigate the dynamics of host and virus gene expression for two of the most important enzymes in the sphingolipid pathway, serine palmitoyl transferase (SPT) and dihydroceramide desaturase (DCD). SPT catalyses the first and rate-limiting step of the pathway, while DCD leads to direct production of ceramide (Merrill, 2002; Perry et al., 2000).

The use of real time quantitative PCR proved to be a very useful and powerful approach to investigate abundance dynamics of specific genes in natural oceanic samples. Our study reported a clear transcriptional pattern displayed along *E. huxleyi* bloom formation and decline, with the coccolithovirus transcripts taking over and progressively controlling the sphingolipid pathway. This experiment was realized in different naturally occurring *E. huxleyi* blooms. The reproducible expression profiles observed for all genes tested in those independent enclosures reinforced the robustness of the data, and lead us to think that the patterns observed may well be a good representation of natural bloom dynamics.

Crucially, the observed patterns were in accordance with the hypothesis of possible involvement of viral sphingolipids in the timing and physical processes of virion release from the host cells. Moreover the EhV sphingolipid pathway genes showed a strikingly similar pattern of transcript abundance as the structural major capsid protein gene. This denotes to what extent these "newly imported" pathway (through the above demonstrated HGT events) has been integrated into the EhV genome, acquiring a functional use that is now seemingly interlinked with the complex gene machinery used during coccolithovirus infection. The presence of a near complete sphingolipid pathway in the coccolithovirus genome, in tandem with its substantial expression during the infection, indicate that sphingolipid biosynthesis is a crucial factor for a successful coccolithovirus infection.

The physiological function(s) of sphingolipids in either virus or host remains to be resolved. In numerous other systems, the de novo ceramide metabolism has been identified as playing an important role in bioactive lipid signalling, activation of apoptosis and other crucial pathways (Hannun and Obeid, 1995; Hannun and Obeid, 2008). The demonstrated activation of host metacaspase activity by the coccolithoviruses (Bidle et al., 2007) and production of viral glycosphingolipids during natural *E. huxleyi* blooms (Vardi et al., 2009) adds credence to theory of PCD manipulation in this system. Control of host PCD has been described in other viruses belonging to the large family of NCLDVs. In all those cases viruses were reported to encode genomic machinery impeding host's PCD (Hay and Kannourakis, 2002; McLean et al., 2008).

The potential coccolithovirus driven activation of host PCD would hence be (for the time being) another extraordinary feature among the NCLDVs.

Other hypotheses exist though for the function of the coccolithovirus sphingolipid pathway. It could be that the sphingolipids produced by the host using its own pathway are different structurally and functionally to those produced by the virus. For example, the host SBP could be involved in bioactive lipid signalling among cells, while the viral homologous pathway could eventually be playing a counter role of stereochemical mimic to inhibit that process. Another theory is that the function of the virally produced sphingolipid could be related to its inherent physical properties in the cell membrane. These induced sphingolipids could be involved in vesicle mediated transport inside the cells and/or promote the formation of lipid rafts viral budding and release. This idea correlates to the recent observations of EhV-86 capsid budding (to host membrane) and migration inside the cell (Mackinder et al., 2009). Independently of the outcome of future researches on this subject, this study will stand as the first time one quantified the transcriptional interplay of homologous metabolic pathways between virus and host during temporal progression of oceanic *E. huxleyi* blooms.

Throughout the past 15 years plankton virology has followed a general tendency to go from the initial global ecological dynamics to specific genomic and functional interactions, whose works rely on the few available host/virus models available. The previously discussed works on HGTs between *E. huxleyi* and EhV are a good demonstration of this. We are convinced though, that in order to understand the plankton host / virus interactions in their complexity we should try to complement the lab/strain based approaches, commonly used nowadays, with wider ecological studies that target real and natural planktonic communities. In that regard the development of new genomic and transcriptomic tools will be essential to scrutinize through global metabolic processes and identify major patterns of utilization of genomic information taking place in the oceans. We had this idea in mind when we decided to perform RNA transcript measurements on natural samples using qPCR. The clear viral takeover pattern observed was very promising, and proved the strength of using that approach. However, whilst single gene analysis can provide a particular functional insight into a particular metabolic process, a genome wide approach would be essential for broader understanding of physiological interdependence between genes and metabolical pathways.

That led us to attempt for the first time a study on both host and virus global transcriptomics occurring within a natural oceanic community (see Chapter 5). In order to do so, we focused on the progression of an *E. huxleyi* bloom followed by the inevitable crash provoked by the severely lytic EhV. We designed a microarray set to target both *E. huxleyi* and EhV wide gene transcription patterns. Before conducting this microarray study the ideas existing on the EhV infection strategy, and metabolic requirements during infection, were based solely on coccolithovirus transcription data. Having novel access to host probes we could push our comprehension a step further with a study that combined simultaneous host and virus transcript analysis.

We think that when measuring global host/virus transcript levels using natural community samples, the significance of the results depends largely on two major factors: (1) that there is significant genetic resemblance between the targeted natural strains and the host/virus strains used for microarray design; and (2) that transcription patterns are extensively synchronized within the community. In our opinion both these conditions were present in our study. We were able to observe specifically *E. huxleyi* and EhV transcription profile evolution among all the other transcripts being produced by the global planktonic community. This reflected the outstanding community overtake produced by *E. huxleyi* cells during bloom stages. If this was not the case then any transcriptomic message being produced inside *E. huxleyi* cells would most probably be "diluted" in the pool of total transcripts present in the water. Moreover, our results also supported the idea that bloom/infection processes involve synchronized interactions between cells, which seem regulated by strict constraints and stages. Through the use of a microarray approach, we could confirm that there is an extensive and synchronized viral activation throughout the global *E. huxleyi* "blooming" community as infection develops. That viral activation was synchronized, allowing the recognition of the same viral takeover pattern that had previously been observed with real time PCR data (Pagarete et al., 2009, see Chapter 4).

A big difficulty when trying to perform this type of study is to identify distinct moments where gene expression corresponds to different biological stages. In the case of microalgae blooms and consequent viral infections, that task may be simplified by the existence of clearly recognizable pre and post viral takeover stages. We already knew that the great majority of the EhV genes are expressed during infection (Allen et al., 2006; Wilson et al., 2005). Moreover, qPCR data had also showed that EhV genes involved in very different functions (such as the

major capsid protein, or the sphingolipid pathway) are clearly activated during infection, with transcript abundance values raising of several orders of magnitude (Pagarete et al., 2009, see Chapter 4). We chose, hence, to focus our analysis on the genes that significantly increased in transcript abundance from pre to post viral takeover. The fact that the abundance of these transcripts oscillates around a moment of viral takeover made us more confident to believe that their regulation is linked to the process of EhV infection.

The first and clearest result was to discover that, from the bulk of *E. huxleyi* and EhV genes, the great majority of those that significantly increased in transcript abundance from pre to post viral takeover were of viral origin. Although being an expected situation, given that previous reports had already showed that there is an extensive activation of the coccolithovirus genome upon infection (Allen et al., 2006; Wilson et al., 2005), this was the first time we confirmed this phenomena among a complex community of wild coccolithovirus.

Moreover, our pre to post viral takeover analysis revealed not only the requirement of EhV genes, but also the increase in transcript abundance of numerous *E. huxleyi* genes putatively related to many different metabolic pathways. It should be noted that, although we could predict the function of some of those genes (through sequence comparison and phylogenetic inference), the large bulk of them remain to be functionally annotated. This is a recurrent situation in the characterization of large viral genomes, for which most of the genes have no homologues in the existing genomic and protein databases.

An analysis of the annotated genes allowed us to build an interpretation of possible metabolic mechanisms being required during coccolithovirus infection. Logically, that interpretation was based on (1) the current knowledge we have of the *E. huxleyi* / EhV model and (2) the examples of characterized gene functions from other host / virus systems. Hence, among a very large list of annotated *E. huxleyi* and EhV genes that significantly increased transcript abundance from pre to post viral takeover, we pointed out a few genes that are probably involved in (expected) functions such as genetic information processing, and also in some (less expected) functions such as posttranslational control, intracellular trafficking mechanisms, or even control of programmed cell death.

One of the major “priorities” for a virus during infection is to replicate and translate its genetic information into newly synthesized viruses. It was hence logical that our results indicated the recruitment of numerous amino acid and nucleotide metabolism related genes, most probably in response to that need for genetic information processing. Several transcription related EhV

encoded genes were highly transcribed during infection, such as viral DNA polymerase and topoisomerase. Moreover several *E. huxleyi* genes related to mRNA splicing, ribosomal function and translation were also clearly up-regulated. Further down the line post-translational processes were consistently required, with virus and host sharing potentially interlinked functions namely through the use of protein cleavage enzymes and protein chaperones.

Another striking example of transcriptional features identified for the first time in the course of this experiment regards the manipulation of cell cytoskeleton machinery pathways (through actin transcription regulation). Similar cases had been identified in numerous other viruses as a mean to avoid molecular crowding inside the cell (Radtke, Dohner, and Sodeik, 2006). Namely among the large NCLDV group, herpesviruses and poxviruses, for example, it has been shown the use of host actin and microtubule transport systems for several steps during their life cycle (Forest, Barnard, and Baines, 2005; Schramm and Locker, 2005). We propose that if the need for active intracellular transport mechanisms is related to the size of the particles in question, then this would probably be of crucial importance for a "giant" virus such as EhV (capsid diameter between 160-180 nm). Moreover this potential EhV active transport mechanism would demand energy production and conversion, which could explain the up-regulation verified for genes like ATP synthase, NADH-dehydrogenase or vacuolar ATPase. This data was in accordance with the recent observation of EhV capsid displacement inside the host cells during infection (Mackinder et al., 2009).

Even more surprisingly this same study (Mackinder et al., 2009) argued that coccolithoviruses have an infection mechanism different from that employed by other algal NCLDVs, with entry and exit strategies showing a greater analogy to animal-like NCLDVs. This was based on observations of EhV-86 entering its host via either an endocytotic or an envelope fusion mechanism, and also on a budding mediated mechanism for virus progeny release. In our study several lipid transport and metabolism genes were identified, which could also be involved in membrane and vesicle related capsid transport. Although lipid function during EhV infection is still poorly understood, our study added to the idea that lipids play a fundamental role, recently demonstrated in the field (Vardi et al., 2009).

Overall, the gene utilization features observed with our microarray study seem to relate to the conception of virus factories commonly known for other large DNA viruses. When a virus enters a cell, the subsequent steps in its replication cycle involve interactions between different types of viral and host components. These interactions may be more or less complex depending

on the infections strategy used by each virus. In the case of large DNA viruses, which are known for carrying particularly rich and complex genomes, it is common to observe large areas of the infected cell modified for the creation of functional dynamic structures where efficient viral morphogenesis occurs. That functional viral structure, known as "viral factory" (Novoa et al., 2005), has been reported for several complex NCLDVs, namely the *Poxviridae, Iridoviridae, Asfarviridae, Herpesviridae,* and *Mimiviridae*. According to our results EhV probably employs a similar "viral factory" strategy. This supports the idea of an evolutionary continuum that would link distant groups of large DNA viruses, such as the coccolithovirus, to other NCLDVs that infect animals and other life forms.

We must not forget that the knowledge we have of the mechanisms of infection used by EhV is still very incipient, which is why all the hints of possible gene function requirement here presented are breakthroughs on the path to better understand this virus. At this stage it would be too early to present a detailed picture of the metabolic interactions happening inside *E. huxleyi* cells during coccolithovirus infection. Any putative gene functions being utilized by either virus or host metabolisms need to be confirmed through functionality tests. Nonetheless, the genes presenting very high transcriptional changes throughout EhV infection may reveal very useful as potential markers for ecological monitoring of coccolithovirus communities.

Perspectives for future research

In the works here described we presented the discovery of outstanding HGT events that corroborate the very close interactions occurring between the cellular and viral life, and confirm the extraordinary role that viruses play in evolution and ecology. These examples though may well be the tip of an iceberg. In the immediate future the list of 35 reported cases of HGT demands further characterization, both with phylogenetic analysis that will allow understanding better the origin of these genes, as well as with functionality assays to discover their potential importance and role in the coccolithovirus genome.

We also discovered the utilization of specific host and virus genes during EhV infection, which not only allowed us to foresee possible cellular mechanisms that EhV uses during infection, but also contributed to the newly forming idea of an evolutionary continuum between the different NCLDVs, with microalgae viruses hypothetically representing ancestral forms of this large group of DNA viruses. These insights into the metabolic processes required during

EhV infection are still very incipient though. Further and more detailed characterization of gene use is demanded. For instance, it is essential to develop tools that will allow disentangling the viral activated metabolic mechanisms from the host responses to the viral attack. Such *demarche* should be accompanied of functionality tests using proteomic tools to clearly unveil the role played by the different proteins that we now know are required during infection. Here we should not forget that in what regards the great majority of host and virus genes we are severely conditioned by the impossibility (so far) of even predicting their function. Future hybridization tests could be envisaged that will take advantage of the knowledge we have so far, both on probable gene functions and utilization of certain genes, to open the door to many of the unknown genes that make the complex EhV genome black box.

On an ecological perspective, we were able to demonstrate the utilization of the outstanding coccolithovirus sphingolipid pathway during natural *E. huxleyi* blooms, and its close relation to the dynamics of the bloom and infection development. These genes remain a source of discussion and of high scientific interest. Our studies, with others, confirmed the importance that this pathway has in the course of EhV infection. However, some of the most interesting and elementary questions regarding these genes remain to be answered: what is the real role of the coccolithovirus sphingolipid pathway during infection? is it related to a form of controlling host cellular death, or is it implicated in the vesicle trafficking and budding mechanisms that are probably occurring? The answer to these questions will once again demand entering the proteomic world, and the utilization of functionality tests. Moreover, it would be extremely useful the development of gene manipulation tools (so far inexistent in this system), to allow the possibility of introducing artificial mutations and performing gene knockouts. We are positive that the possible discovery that this coccolithophore virus has acquired mechanisms to, and is capable of, controlling the death of its host would be a remarkable scientific breakthrough. Certainly it would significantly condition the way we conceive phytoplankton virology and the role played by protist viruses in the oceans. The observed scenario of increased lipid production during coccolithovirus infection is also, and on its own, a very interesting feature, especially in a world that is turning into microalgae culturing as a source of renewable energy. Viruses may well represent excellent tools to efficiently breakdown cells for lipid harvesting, with the possible advantage of increasing the amounts of produced lipids in the process.

As a final note we would like to say that the isolated microalgae viruses that are available in today's laboratories remain very scarce examples of the colossal diversity of viruses existing in

the oceans. This clearly reflects the embryonic state of phytoplankton virology. We consider that the most significant achievements to be made in near future in this field pass by the isolation and characterization of new viruses. Indeed, a significant amount of this thesis's period was spent trying to isolate new coccolithophore viruses (see Chapter 6). Our attempts proved that what seems to be a somewhat straightforward task for some viruses and some host cells, may reveal a very hard objective in others. This means that the conditions to isolate viruses are not standardized and that specific methodologies should be adapted for each virus and host interaction under study. This is clearly a daunting task. Hopefully in the future we will be able to develop protocols to allow the large scale assessment for many different microalgae, allowing the discovery of a whole new range of viruses.

Plankton virology is implicated in very important and diversified areas of scientific research. These go from the most fundamental debate on the origin of cellular life, have crucial implications for the comprehension of major biological interactions in the oceans and regulation of Earth's chemical cycles, and even include a vast panoply of potential energetic and medical applications with high impact for mankind. Working at the same time on the vastest reservoir of genetic diversity existing on our planet, and on the limits of life as we know it, plankton virology has developed into one of the most exciting and promising fields of modern science, one that surely will not stop enlightening us with significant and unexpected discoveries in the decades to come.

Annexe A.

Horizontal gene transfer between *Emiliania huxleyi* and viruses*

Hiroyuki Ogata[1],Adam Monier[1], António Pagarete[2], Jean-Michel Claverie[1].

1. Structural and Genomic Information Laboratory, CNRS-UPR2589, Mediterranean Institute of Microbiology (IFR-88), Université de la Méditerranée, Parc Scientifique de Luminy, 163 avenue de Luminy Case 934, FR-13288 Marseille, France.

2. Equipe EPPO-Evolution du Plancton et PaléoOcéans, CNRS-UMR7144, Université Pierre et Marie Curie, Station Biologique, FR-29682 Roscoff, France.

1. Brief report

Horizontal gene transfers (HGTs) between phytoplankton and their viruses could be a driving force in the co-evolutionary processes of their genomes (Hendrix et al., 2000). Recent reports on HGTs between cyanobacteria and cyanophages supports this view for prokaryotic systems (Lindell et al., 2007; Yoshida et al., 2008). However, the occurrence of HGTs is much less documented for eukaryotic systems due to the lack of genomic sequence data. Our analysis combining similarity searches and phylogenetic reconstruction revealed 35 cases of possible HGTs between *E. huxleyi* and viruses (**Table 1, Table S1**).

Table 1. Putative lateral gene transfer between *E. huxleyi* and viruses.

Ehux protein ID	Viral species	Best viral homolog (%-identity, E-value)	Putative function	Putative biological process
54601	*Emiliania huxleyi* virus 86	ehv061 (34%, 3x10-36)	Dihydroceramide desaturase	Sphingolipid pathway
200862	*Emiliania huxleyi* virus 86	ehv014 (37%, 1x10-44)	Dihydroceramide synthase (Lag1)	Sphingolipid pathway
43654	*Emiliania huxleyi* virus 86	ehv077 (55%, 1x10-63)	Long chain fatty acid elongation	Sphingolipid pathway
196284	*Emiliania huxleyi* virus 86	ehv415 (31%, 3x10-22)	Fatty acid desaturase (Aco1, delta-9)	Sphingolipid pathway
432901	*Emiliania huxleyi* virus 86	ehv050 (44%, 1x10-145)	Serine palmitoyltransferase (tri-domain architecture of LCB2/LCB1/PAP2)	Sphingolipid pathway
193908	*Emiliania huxleyi* virus 86	ehv079 (26%, 4x10-15)	Lipid phosphate phosphatase (PAP2 superfamily)	Sphingolipid pathway
210457	*Emiliania huxleyi* virus 86	ehv031 (42%, 1x10-54)	ERG3, Sterol desaturase	Sphingolipid pathway

**António Pagarete contributed to the annotation of the genes presented in this report.*

102590	*Emiliania huxleyi* virus 86	ehv116 (30%, 7x10-34)	ERG3, Sterol desaturase	Sphingolipid pathway
432191	*Emiliania huxleyi* virus 86	ehv395 (25%, 1x10-12)	Sec14p-like lipid-binding domain	Membrane dynamics
97888	*Emiliania huxleyi* virus 86	ehv400 (36%, 5x10-19)	Lipocalin-like protein	Membrane dynamics (possibly transporter of lipids)
420219	*Emiliania huxleyi* virus 86	ehv158 (57%, 3x10-98)	ATP-dependent DNA ligase	DNA/RNA processing
208320	*Emiliania huxleyi* virus 86	ehv141 (35%, 2x10-39)	DNA repair and recombination protein pif1-like with HRDC (Helicase and RNase D C-terminal) domain	DNA/RNA processing
215136	*Emiliania huxleyi* virus 86	ehv431 (37%, 1x10-44)	Tmk, Thymidylate kinase	DNA/RNA processing
61414	*Emiliania huxleyi* virus 86	ehv117 (54%, 4x10-54)	Phosphate permease	Transporter
212478	*Emiliania huxleyi* virus 86	ehv179 (41%, 3x10-95)	MFS_1, Major Facilitator Superfamily	Transporter
446612	*Emiliania huxleyi* virus 86	ehv056 (36%, 2x10-22)	Methyltransferase	Metabolism
235604	*Emiliania huxleyi* virus 86	ehv421 (27%, 3x10-17)	Glycosyl transferase family 8-like protein	Metabolism
434519	*Emiliania huxleyi* virus 86	ehv032 (40%, 2x10-32)	Hypothetical protein	-
197639	*Emiliania huxleyi* virus 86	ehv092 (44%, 7x10-9)	Hypothetical protein	-
432978	*Emiliania huxleyi* virus 86	ehv095 (49%, 3x10-60)	Hypothetical protein	-
432205	*Emiliania huxleyi* virus 86	ehv155 (58%, 1x10-54)	Hypothetical protein	-
443105	*Emiliania huxleyi* virus 86	ehv161 (43%, 1x10-78)	Hypothetical protein	-
111551	*Emiliania huxleyi* virus 86	ehv176 (50%, 1x10-11)	Hypothetical protein	-
440222	*Emiliania huxleyi* virus 86	ehv186 (29%, 2x10-7)	Hypothetical protein	-
461715	*Emiliania huxleyi* virus 86	ehv192 (41%, 2x10-21)	Hypothetical protein	-
193896	*Emiliania huxleyi* virus 86	ehv222 (25%, 8x10-10)	Hypothetical protein	-
200323	*Emiliania huxleyi* virus 86	ehv225 (38%, 5x10-11)	Hypothetical protein	-
205088	*Emiliania huxleyi* virus 86	ehv408 (24%, 5x10-8)	Hypothetical protein	-
260349	*Emiliania huxleyi* virus 86	ehv424 (40%, 3x10-27)	Hypothetical protein	-
508420	Mimivirus	MIMI_L315 (26%, 4x10-24)	Formamidopyrimidine-DNA glycosylase	DNA/RNA processing
454190	NCLDVs (Mimivirus, EtV1, OtV5, PBCVs, ASCV1)	OsV5_067f (25%, 2x10-23)	SSL2, DNA or RNA helicases of superfamily II	DNA/RNA processing
242737	PBCV-1, OtV5	OsV5_146f (37%, 7x10-10)	YqaJ viral recombinase family	DNA/RNA processing
243604	NCLDVs (Mimivirus, PBCVs, ASCV1)	Q98540_PBCV1 (34%, 4x10-11)	Hypothetical protein	-
122629	Enterobacteria phages	A5PJ32_9CAUD (36%, 5x10-26)	DNA cytosine methylase	DNA/RNA processing
439872	*Synechococcus* phage Syn9	Q0QZH8_9CAUD (39%, 6x10-29)	2OG-FeII_Oxy domain-containing protein	-

Remarkably, a vast majority of the HGTs (29 of the 35 cases) were between *E. huxleyi* and *E. huxleyi* viruses (EhV's), large lytic DNA viruses that regulate the yearly "boom and bust" successions of *E. huxleyi* blooms (Martinez et al., 2007; Wilson et al., 2005b). Ten of these genes were predicted to be involved in lipid metabolism and/or membrane trafficking, including a series of enzyme genes belonging to the same biosynthetic pathway of ceramide (Monier et al., 2009). Other *E. huxleyi*/EhV HGTs were associated with DNA/RNA processing (4 cases), transporting function (2 cases), other metabolisms (2 cases) and 11 hypothetical proteins. These 29 HGTs correspond to only 6% of the genes encoded in the 407 kb EhV genome (Wilson et al., 2005b), albeit being substantially higher than the previously reported HGTs (<1%) between the green alga *Ostreococcus tauri* (12.6 Mbp) and its virus OtV5 (186 kbp) (Derelle et al., 2008). The remaining six putative HGTs concerned other large eukaryotic viruses (4 cases) and phages (2 cases). We also found several types of relatively short interspersed DNA repeats uniquely shared between *E. huxleyi* and EhV (**Table 2**, **Fig. 1**).

Table 2. EhV-86 genomic sequence regions showing significant similarities in the *E. huxleyi* genome sequences.

EhV-86 genomic coordinate			BLASTN hits (E-value<1e-10) in the *E. huxleyi* genome		Sequence feature
Start	End	Length (bp)	# of HSPs	# *E. huxleyi* scaffold	
59758	59861	104	1	1	GGN (or NCC) rich sequence
60397	60466	70	1	1	GGN (or NCC) rich sequence
127859	128101	243	66	34	GGN (or NCC) rich sequence
128234	128353	120	8	4	GGN (or NCC) rich sequence
137808	137892	85	21	17	A rearrangement between *E. huxleyi* and EhV
181582	181764	183	149	7	GGN (or NCC) rich sequence
184489	184601	113	1	1	GGN (or NCC) rich sequence
187728	187936	209	86	53	GGN (or NCC) rich sequence
188151	188192	42	2	2	GGN (or NCC) rich sequence
188559	188731	173	238	21	GGN (or NCC) rich sequence
188970	189333	364	251	41	GGN (or NCC) rich sequence
197822	199069	1248	794	79	GGN (or NCC) rich sequence
272296	272338	43	4	4	GGN (or NCC) rich sequence
307015	307463	449	10	9	GGN (or NCC) rich sequence
307474	307724	251	3	3	GGN (or NCC) rich sequence
307820	307997	178	2	2	GGN (or NCC) rich sequence
308008	308258	251	3	3	GGN (or NCC) rich sequence
308281	308590	310	3	3	GGN (or NCC) rich sequence
308621	308818	198	47	5	GGN (or NCC) rich sequence
309583	309656	74	13	12	GGN (or NCC) rich sequence
309726	309799	74	2	2	GGN (or NCC) rich sequence
309856	310016	161	37	12	GGN (or NCC) rich sequence
310278	310545	268	39	14	GGN (or NCC) rich sequence
311553	311664	112	1	1	GGN (or NCC) rich sequence
349695	349772	78	35	33	GGN (or NCC) rich sequence

These results suggest that lateral gene transfers have taken place during the evolution of *E. huxleyi* and its viruses. The direction of these HGTs could be in both ways between virus and host. On one hand, the direction from *E. huxleyi* to its virus is likely for many of these HGTs, where the host and viral genes exhibit close homologs in other eukaryotes (**Fig. S1**). On the other hand, in several cases (for instance ehux454190 in **Fig. S1**), close homologs of *E. huxleyi* genes were found only in viruses (and more remotely in bacteria). This latter suggests that *E. huxleyi* might have an access to the genetic pool of viruses and that of bacteria through viruses in an evolutionary time scale.

(a)

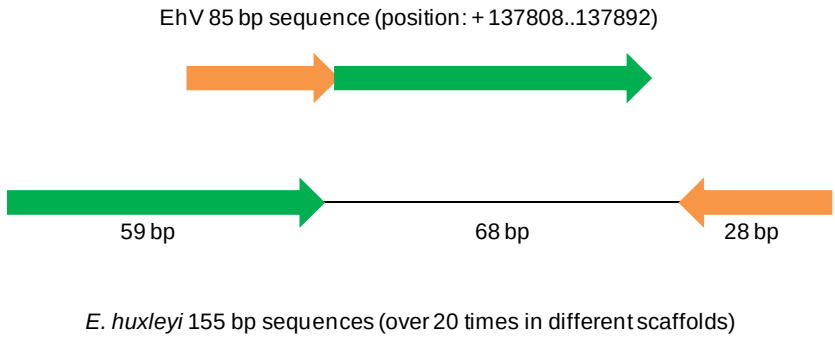

Figure 1. Interspersed repeats uniquely present in *E. huxleyi* and EhV. (**a**, **b**) Repeated sequences found over 20 times in different *E. huxleyi* scaffolds and once in the EhV genome. (**c**) An example of the alignment for the GGN-rich interspersed repeats found in both E. huxleyi and EhV.

(b)

```
                 (EhV-86: + 137834..137892)
EhV-86           CACACGCACACAC-CACTCCATACGCGAGCACGCTCTCGCTCCGCCGCTCGTGCGCCC--
Scaffold_68      CACACACACACACACACTCCATGGGCGAGCCCGCTCTCGCTCCGCCGCTCGTGCGCCC--
Scaffold_138.2   CACACACACGCACACACTCCATACGCGATCCTGCTCTCGCTCCGCCGCTCGTGCGCCC--
Scaffold_147     CACACACACACACACACACCATACGCGAGCCTGCTCTCGCTCCGCCGCTCGTGCGCCC--
Scaffold_170     CACACACACACACACACACCATACGCGAGCCCGCTCTCGCTCCGCCGCTCGTGCGCCC--
Scaffold_3.1     CACACACACACACACACTCCATACGCGAGCCCGCTCTCGCTCCGCCGCTCGTGCGCCC--
Scaffold_470     CACACACACACACACACTCCATACGCGAGCCCGCTCTCGCTCCGCCGCTCGTGCGCCC--
Scaffold_19.1    CACACACACACACACACTCCATACGCGAGCCCGCTCTCGCTCCGCCGCTCGTGCGCCC--
Scaffold_416     CACACACACACACACACTCCATACGCGAGCCCGCTCTCGCTCCGCCGCTCGTGCGCCC--
Scaffold_120     CACACACACACACACACTCCATACGCGAGCCCGCTCTCGCTCCGCCGCTCGTGCGCCC--
Scaffold_99      CACACACACACAC-CACTCCATACGCGAGCCCGCTCTCGCTCCGCCGCTCGTGCGCCCCG
Scaffold_61      CACACACACACAC-CACTCCATACGCGAGCCCGCTCTCGCTCCGCCGCTCGTGCGCCCCG
Scaffold_55      CACACACACACAC-CACTCCATACGCGAGCCCGCTCTCGCTCCGCCGCTCGTGCGCCCCG
Scaffold_34      CACACACACACAC-CACTCCATACGCGAGCCCGCTCTCGCTCCGCCGCTCGTGCGCCCCG
Scaffold_27      CACACACACACAC-CACTCCATACGCGAGCCCGCTCTCGCTCCGCCGCTCGTGCGCCCCG
Scaffold_198     CACACACACACAC-CACTCCATACGCGAGCCCGCTCTCGCTCCGCCGCTCGTGCGCCCCG
Scaffold_334     CACACACACACACACACTCCATACGCGAGCCCGCTCTCGCTCCGCCGCTCGTGCGCCC--
Scaffold_138.1   CACACACACACACACACTCCATACGCGAGCCCGCTCTCGCTCCGCCGCTCGTGCGCCC--
Scaffold_3.3     CACACACATACACACACTCCATACGCGAGCCCGCTCTCGCTCCGCCGCTCGTGCGCCC--
Scaffold_3.2     CACACACACACACACACTCCATACGCGAGCCCGCTCTCGCTCCGCCGCTCGTGCGCCC--

EhV-86           CG
Scaffold_68      CGCACATGTCGGCTTTCTCATGCCTCCGGCCTCCGAGCGCCCACCTCGGGCGCTCGGGCG
Scaffold_138.2   CGCGCATGTCGGCTTTCTCATGCCTCCGGCCTCCGAGCGCCCTCCTCGGGCGCTCGGGCG
Scaffold_147     CGCGCATGTCGGCTTTCTCATGCCTCCGGCCTCCGAGCGCCCACCTCGGGCGCTCGGGCA
Scaffold_170     CGCGCATGTCGGCTTTCTCATGCCTCCGGCCTCCGAGCGCCCACCTCCGGCGCTCGGGCG
Scaffold_3.1     CGCGCATGTCGGCTTTCTCATGCCTCCAGCCTCCGAGCGCCCACCTCGGGCGCTCGGGCG
Scaffold_470     CGCGCACGTCGGCTTTCTCATGCCTCCGGCCTCCGAGCGCCCACCTCGGGCGCTCGGGCG
Scaffold_19.1    CGCGCACGTCGGCTTTCTCATGCCTCCGGCCTCCGAGCGCCCACCTCGGGCGCTCGGGCG
Scaffold_416     CGCGCATGTCGGCTTTCTCATGCCTCCGGCCTCCGAGCGCCCACCTTGGGCGCTCGGGCG
Scaffold_120     CGCGCATGTCGGCCTTCTCATGCCTCCGGCCTCCGAGCGCCCACCTCGGGCGCTCGGGCG
Scaffold_99      CGCGCATGTCGGCCTTCTCATGCCTCCGGCCTCCGAGCGCCCACCTCGGGCGCTCGGGCG
Scaffold_61      CGCGCATGTCGGCCTTCTCATGCCTCCGGCCTCCGAGCGCCCACCTCGGGCGCTCGGGCG
Scaffold_55      CGCGCATGTCGGCCTTCTCATGCCTCCGGCCTCCGAGCGCCCACCTCGGGCGCTCGGGCG
Scaffold_34      CGCGCATGTCGGCCTTCTCATGCCTCCGGCCTCCGAGCGCCCACCTCGGGCGCTCGGGCG
Scaffold_27      CGCGCATGTCGGCCTTCTCATGCCTCCGGCCTCCGAGCGCCCACCTCGGGCGCTCGGGCG
Scaffold_198     CGCGCATGTCGGCCTTCTCATGCCTCCGGCCTCCGAGCGCCCACCTCGGGCGCTCGGGCG
Scaffold_334     CGCGCATGTCAGCTTTCTCATGCCTCCGGCCTCCGAGCGCCCACCTCGGGCGCTCGGGCG
Scaffold_138.1   CGCGCATGTCAGCTTTCTCATGCCTCCGGCCTCCGAGCGCCCACCTCGGGCGCTCGGGCG
Scaffold_3.3     CGCGCATGTCAGCCTTCTCATGCCTCCGGCCTCCGAGCGCCCACCTCGGGCGCTCGGGCG
Scaffold_3.2     CGCGCATGTCAGCCTTCTCATGCCTCCGGCCTCCGAGCGCCCACCTCGGGCGCTCGGGCG
```

```
                          (EhV-86: - 137808..137835)
EhV-86                    TGGCATATCGGTTCCCTGGGCGTTCACT
Scaffold_68         GTGGGGAATGTCATATCGGTTCCCTGGGCGTTCACT
Scaffold_138.2      GTTGGGGATGGCATATTGGTTCCCTGGGCGTTCACT
Scaffold_147        GTTGGGGATGACATATCGGTTCCCTGGGCGTTCACT
Scaffold_170        GTTGGGGATGGCATATCGGTTCCCTGGGCGTTCACT
Scaffold_3.1        GTTGGGGATTGCATATCGGTTCCCTGGGCGTTCACT
Scaffold_470        GTTGGGG-TGGCATATCGGTTCCCTGGGCGTTCAC-
Scaffold_19.1       GTTGGGG-TGGCATATCGGTTCCCTGGGCGTTCAC-
Scaffold_416        GTTGGGGATGGCATATCGGTTCCCTGGGCGTTCACT
Scaffold_120        GTTGGGGATGGCATATCGGTTCCCTGGGCGTTCACT
Scaffold_99         GTTGGGGATGGCATATCGGTTCCCTGGGCGTTCACT
Scaffold_61         GTTGGGGATGGCATATCGGTTCCCTGGGCGTTCACT
Scaffold_55         GTTGGGGATGGCATATCGGTTCCCTGGGCGTTCACT
Scaffold_34         GTTGGGGATGGCATATCGGTTCCCTGGGCGTTCACT
Scaffold_27         GTTGGGGATGGCATATCGGTTCCCTGGGCGTTCACT
Scaffold_198        GTTGGGGATGGCATATCGGTTCCCTGGGCGTTCACT
Scaffold_334        GTTGGGGATGGCATATCGGTTCCCTGGGCGTTCACT
Scaffold_138.1      GTTGGGGATGGCATATCGGTTCCCTGGGCGTTCACT
Scaffold_3.3        GTTGGGGATGGCATATCGGTTCCCTGG---------
Scaffold_3.2        GTTGGGGATGGCATATCGGTTCCCTGG---------
```

(c)

```
Score =  529 bits (267), Expect = e-149
Identities = 776/945 (82%), Gaps = 3/945 (0%)
Strand = Plus / Plus


Query: 226   gtggtggagatggtggcggtggagatggtggcggtggagatggcggtggaggcgatggtg 285
             ||||||| |||||||| ||||| | |||||| ||||| | ||| ||||| || | |||||
Sbjct: 43746 gtggtggtgatggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 43805


Query: 286   gcgatggtggcggtggagatggcggtggtggagatggtggcggtggagatggcggtggtg 345
             ||| ||||||||||||   || ||||||||| |||||||| ||||| | ||| |||||||
Sbjct: 43806 gcggtggtggcggtgg---tgacggtggtggtgatggtggtggtggtggtggtggtggtg 43862


Query: 346   gtggagatggtggtggtggagatggcggtggtggtgatggtggtggtggtgatggtggtg 405
             |||| | ||||||||| || | |||||||||||| | ||||||||||||||| ||||||||
Sbjct: 43863 gtggtggtggtggtggcggtggtggcggtggtggcggtggtggtggtggtggtggtggtg 43922


Query: 406   gcggagatggtggtggtggagatggcggtggtggagatggtggcggtggtgatggtggcg 465
             | || | |||||||||||| | ||| |||||||| | ||||||| ||||||| ||||||| |
Sbjct: 43923 gtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 43982


Query: 466   gtgatggtggtggtggcggtggagatggcggtggaggtgatggtggtgggggcgatggtg 525
             ||| |||||||||||| ||||| | ||| ||||| |||| ||||||||| || | |||||
Sbjct: 43983 gtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 44042


Query: 526   gtgggggtgatggtggcggtggtgatggtggtggtggtgatggtggtggcggagatggtg 585
             |||| |||| |||||| ||||||| ||||||||||||||| ||||||||| || | |||||
Sbjct: 44043 gtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 44102


Query: 586   gtggtggagatggcggtggtggagatggtggcggtggtgatggtggcggtgatggtggtg 645
             ||||||| | ||| |||||||| | ||||||| ||||||| ||||||| |||| ||||||||
Sbjct: 44103 gtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 44162


Query: 646   gtggcggtggagatggcggtggaggtgatggtggtgggggcgatggtggtgggggtgatg 705
             |||| ||||| | | | ||||| |||| ||||||||| || | ||||||||| |||| ||
Sbjct: 44163 gtggtggtggtggtagtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 44222


Query: 706   gtggcggtggagatggcggtggtggagatggtggtggcggtgatggtggtggaggcgatg 765
             |||| ||||| | ||| |||||||| | ||||||||| |||| ||||||||| || | ||
Sbjct: 44223 gtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 44282


Query: 766   gtggtgggggtgatggtggtggaggcgatggcggtggtggtggtggtggcggagatggcg 825
             ||||||| |||| ||||||||| || | ||| |||||||||||||||| || | ||| |
Sbjct: 44283 gtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 44342


Query: 826   gtggtggagatggcggtggtggagatggtggtggtggagatggcggtggaggtgatggtg 885
             ||||||| | ||| |||||||| | |||||||||||| | ||| ||||| |||| |||||
Sbjct: 44343 gtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 44402


Query: 886   gtgggggcgatggtggtgggggtgatggtggtgggggtgatggcggtggtggagatggcg 945
             |||| || | ||||||||| |||| ||||||||| |||| ||| |||||||| | ||||
```

```
Sbjct: 44403 gtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggcc 44462


Query: 946   gtggtggagatggtggtggcggtgatggtggtggcggtgatggtggtggtggagatggtg 1005
              ||||||| | ||||||||| ||   ||||  | | |||| |||||||||||| | |||||
Sbjct: 44463 gtggtggtggtggtggtggtggaagtggtcctcgtggtggtggtggtggtggtggtggtg 44522


Query: 1006  gcgggggtgatggtggtgggggtgatggtggtgggggtgatggtggcgggggtgatggtg 1065
              | ||      ||||||||| |||| ||||||||| |||| |||||| || |||| |||||
Sbjct: 44523 gtggccaccgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 44582


Query: 1066  gcgggggtgatggtggtgggggtgatggtggtggaggcgatggtggtgggggtggtggtg 1125
              | || |||| |||||||||| ||| ||||||||| |  | ||||||||  |||| |||||
Sbjct: 44583 gtggtggtggtggtggtgggtgtggtggtggtggtgttggtggtggtgttggtgttggtg 44642


Query: 1126  gcggagatggtggtggtggtggagatggtggcgggggtgatggtg 1170
              | |  | ||||||||||||||| | |||||| ||| ||| |||||
Sbjct: 44643 gtgttggtggtggtggtggtggtggtggtggtgggtgtggtggtg 44687
```

2. Methods

Homologous sequences of *E. huxleyi* protein models (the JGI reduced set) were gathered from UniProt (UniProtConsortium, 2008) using BLASTP, PSI-BLAST and BLASTCLUST (Altschul et al., 1997). Multiple sequence alignments were generated using MUSCLE (Edgar, 2004). All the gap-containing sites in the alignment were excluded in the phylogenetic analysis. Phylogenetic analyses were performed using the neighbor-joining (NJ) method implemented in ClustalW (Chenna et al., 2003) and the maximum likelihood (ML) method implemented in PhyML (Dereeper et al., 2008; Guindon and Gascuel, 2003). NJ analysis was performed based on the distances with Kimura's correction. ML analysis was performed with WAG substitution model and a gamma low (four rate categories). We considered *E. huxleyi* and viral proteins as potentially originating from horizontal gene transfer if their closest homologs were only found in *E. huxleyi* and viruses, or if the *E. huxleyi* and viral protein sequences formed a monophyletic group in both NJ- and ML-analyses. Statistical supports for the phylogenetic reconstructions are provided in **Fig. 1** and **Table S1**.

3. Acknowledgements

The works here presented were conducted under the frame of the *E. huxleyi* genome project consortium (http://genome.jgi-psf.org/Emihu1/Emihu1.home.html), and will hence be published as an integral part of an article discussing the various aspects ligated with the

sequencing of the *E. huxleyi* genome. HO and JMC were the main heads behind the project, AP contributed to the annotation of the genes analysed in this report.

5. Supplementary data - Tables

Table S1. Putative lateral gene transfer between *E. huxleyi* and viruses.

JGI Ehux protein ID	Annotation	Scaf-fold ID	Best hit species (BLASTP/ UniProt)*	Best hit E-value	Number of BLAST hits (E-value < 0.001)	Viral hits	Euka-ryote hits	Bacte-rial hits	Archa-eal hits	Other hits
43654	ELO, GNS1/SUR4 family (long chain fatty acid elongation systems that produce the 26-carbon precursors for ceramide and sphingolipid synthesis)	39	EhV-86	1E-63	232	1	231	0	0	0
54601	Dihydroceramide desaturase (Dsd1, delta-4)	675	EhV-86	3E-36	95	1	92	2	0	0
61414	(Nutrient) Phosphate permease	3	*Tetraselmis chuii.*	3E-57	250	1	137	112	0	0
97888	Lipocalin-like (Lipocalins are transporters for small hydrophobic molecules, such as lipids, steroid hormones, bilins, and retinoids.)	12	EhV-86	5E-19	7	1	4	2	0	0
102590	ERG3, Sterol desaturase	41	EhV-86	7E-34	5	1	4	0	0	0
111551	Hypothetical protein	153	EhV-86	1E-11	1	1	0	0	0	0
122629	Putative DNA cytosine methylase	1385	Enterobacteria phage TLS	5E-26	5	4	0	1	0	0
193896	Hypothetical protein	1	EhV-86	8E-10	1	1	0	0	0	0
193908	Lipid phosphate phosphatase (PAP2 superfamily)	1	EhV-86	4E-15	2	1	1	0	0	0
196284	Fatty acid desaturase (Aco1, delta-9)	4	EhV-86	3E-22	233	1	86	146	0	0
197639	Hypothetical protein	5	EhV-86	7E-09	1	1	0	0	0	0
200323	Hypothetical protein	11	EhV-86	5E-11	1	1	0	0	0	0
200862	Dihydroceramide synthase (longevity-assurance LAG1 family)	13	EhV-86	1E-44	198	1	197	0	0	0
205088	Hypothetical protein	23	EhV-86	5E-08	2	2	0	0	0	0
208320	(DNA processing) DNA repair and recombination protein pif1-like with HRDC (Helicase and RNase D C-terminal) domain	35	EhV-86	2E-39	137	1	86	47	2	1
210457	ERG3, Sterol desaturase	43	EhV-86	1E-54	194	1	190	3	0	0
212478	MFS_1, Major Facilitator Superfamily	52	EhV-86	3E-95	45	1	43	1	0	0
215136	(DNA processing) Tmk, Thymidylate kinase	64	*Trypanosoma brucei.*	5E-48	80	3	41	31	5	0
235604	(Sugar metabolism) Glycosyl transferase family 8-like protein	223	EhV-86	3E-17	1	1	0	0	0	0
242737	(DNA processing) YqaJ viral recombinase family	367	*Arabidopsis thaliana*	5E-11	22	2	15	5	0	0
243604	Hypothetical protein	393	Paramecium bursaria Chlorella virus 1	4E-11	12	12	0	0	0	0
260349	Hypothetical protein	10861	EhV-86	3E-27	1	1	0	0	0	0

Table S1. *(cont.).*

420219	(DNA processing) ATP-dependent DNA ligase	1	*Dictyostelium discoideum*	1E-104	250	44	150	4	52	0
432191	Sec14p-like lipid-binding domain	3	EhV-86	1E-12	20	1	19	0	0	0
432205	Hypothetical protein	3	EhV-86	1E-54	2	1	1	0	0	0
432901	Serine palmitoyltransferase (tri-domain architecture of LCB2/LCB1/PAP2)	7	EhV-86	1E-145	250	1	191	54	4	0
432978	Hypothetical protein	8	EhV-86	3E-60	1	1	0	0	0	0
434519	Hypothetical protein	19	EhV-86	2E-32	2	2	0	0	0	0
439872	2OG-FeII_Oxy domain-containing protein (Prolyl 4-hydroxylase alpha subunit-like)	110	Synechococcus phage syn9.	6E-29	64	6	0	58	0	0
440222	Hypothetical protein	120	EhV-86	2E-07	1	1	0	0	0	0
443105	Hypothetical protein	218	EhV-86	1E-78	2	1	0	1	0	0
446612	Methyltransferase	913	EhV-86	2E-22	7	1	0	6	0	0
454190	(DNA processing) SSL2, DNA or RNA helicases of superfamily II	103	Paramecium bursaria Chlorella virus AR158.	2E-23	61	9	0	48	4	0
461715	Hypothetical protein	7	EhV-86	2E-21	1	1	0	0	0	0
508420	(DNA processing) Formamidopyrimidine-DNA glycosylase	652	Acanthamoeba polyphaga mimivirus	4E-24	14	1	1	12	0	0

* EhV-86 stands for Emiliania Huxleyi Virus 86

6. Supplementary data - Figures

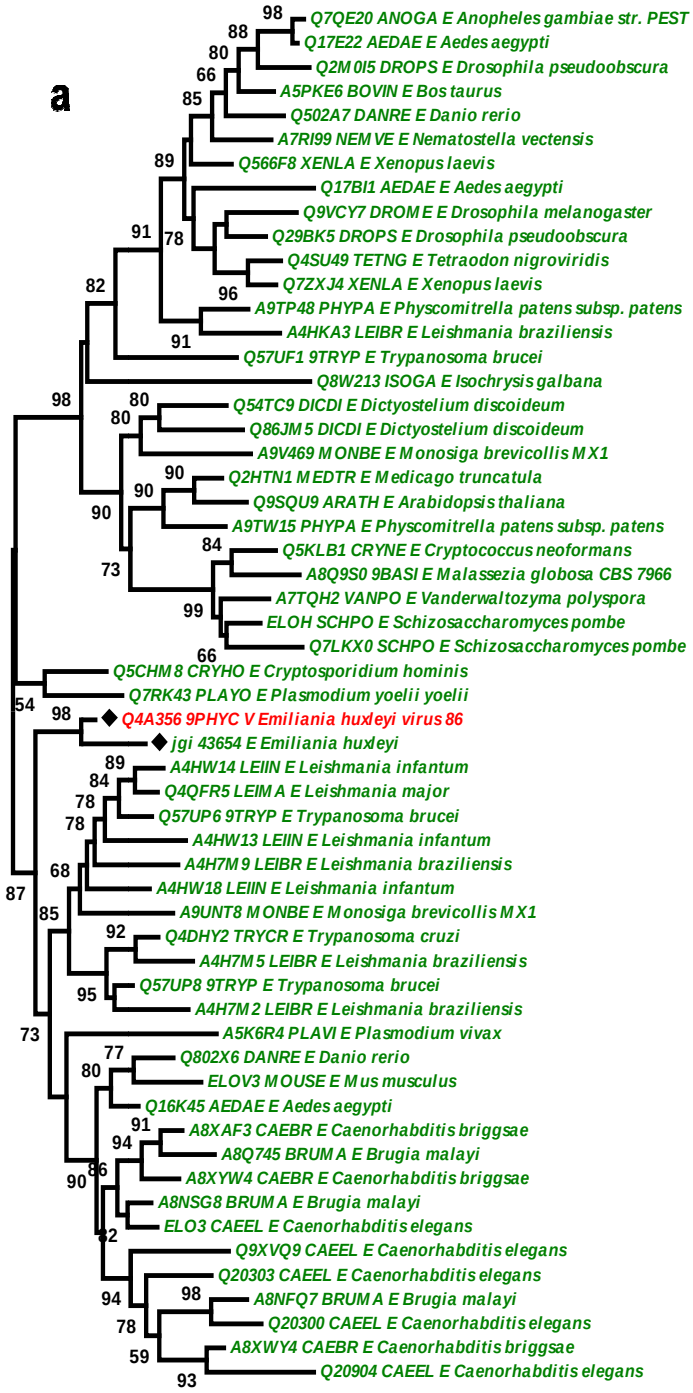

Figure S1. Maximum likelihood phylogenetic trees for putative lateral gene transfers between *E. huxleyi* and viruses. Trees are essentially unrooted. Mid-point rooting was used only for presentation purpose. Eukaryotes, Viruses, Bacteria, and Archaea are colored in green, red, black, and blue, respectively. The branch labels indicate the minimum value of the Chi2-based parametric branch support (i.e. approximate likelihood ratio test) and the Shimodaira-Hasegawa-like non-parametric branch support. **S1a** corresponds to *E. huxleyi* gene 43654: ELO, GNS1/SUR4 family (long chain fatty acid elongation systems that produce the 26-carbon precursors for ceramide and sphingolipid synthesis). *Continued in the next pages.*

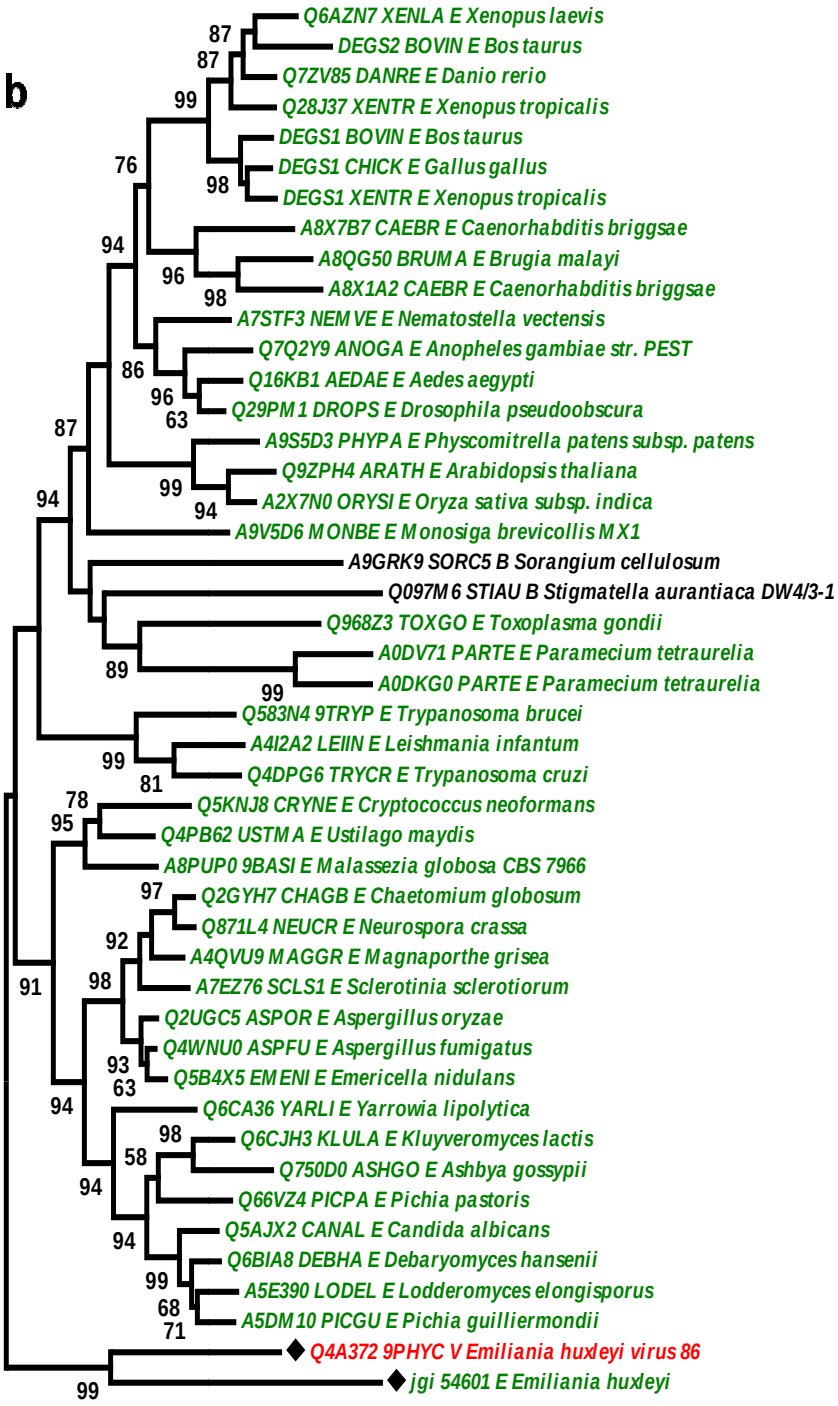

Figure S1b. *(cont.) E. huxleyi* gene 54601: Dihydroceramide desaturase (Dsd1, delta-4).

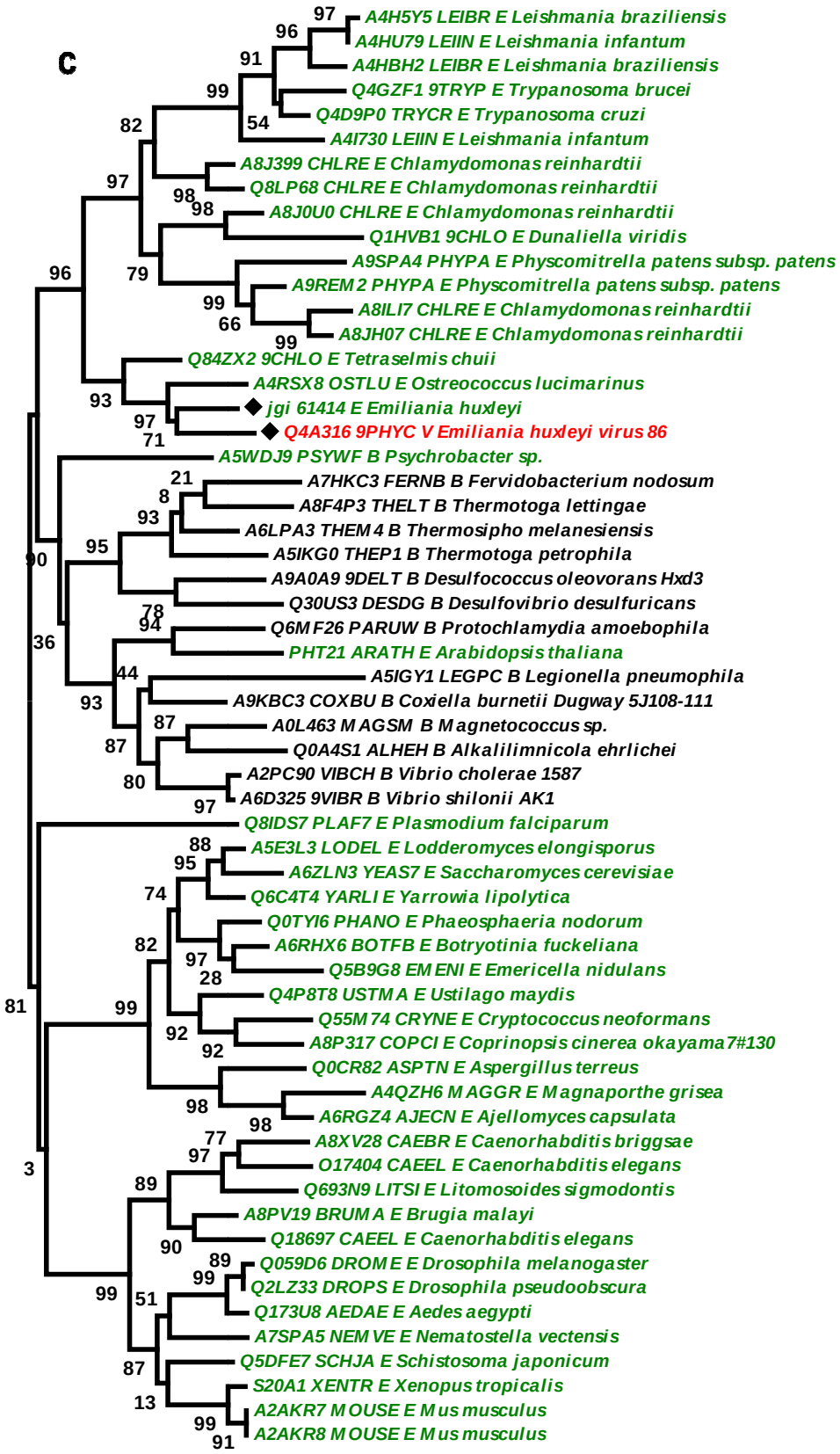

Figure S1c. *(cont.) E. huxleyi* gene 61414: Phosphate permease.

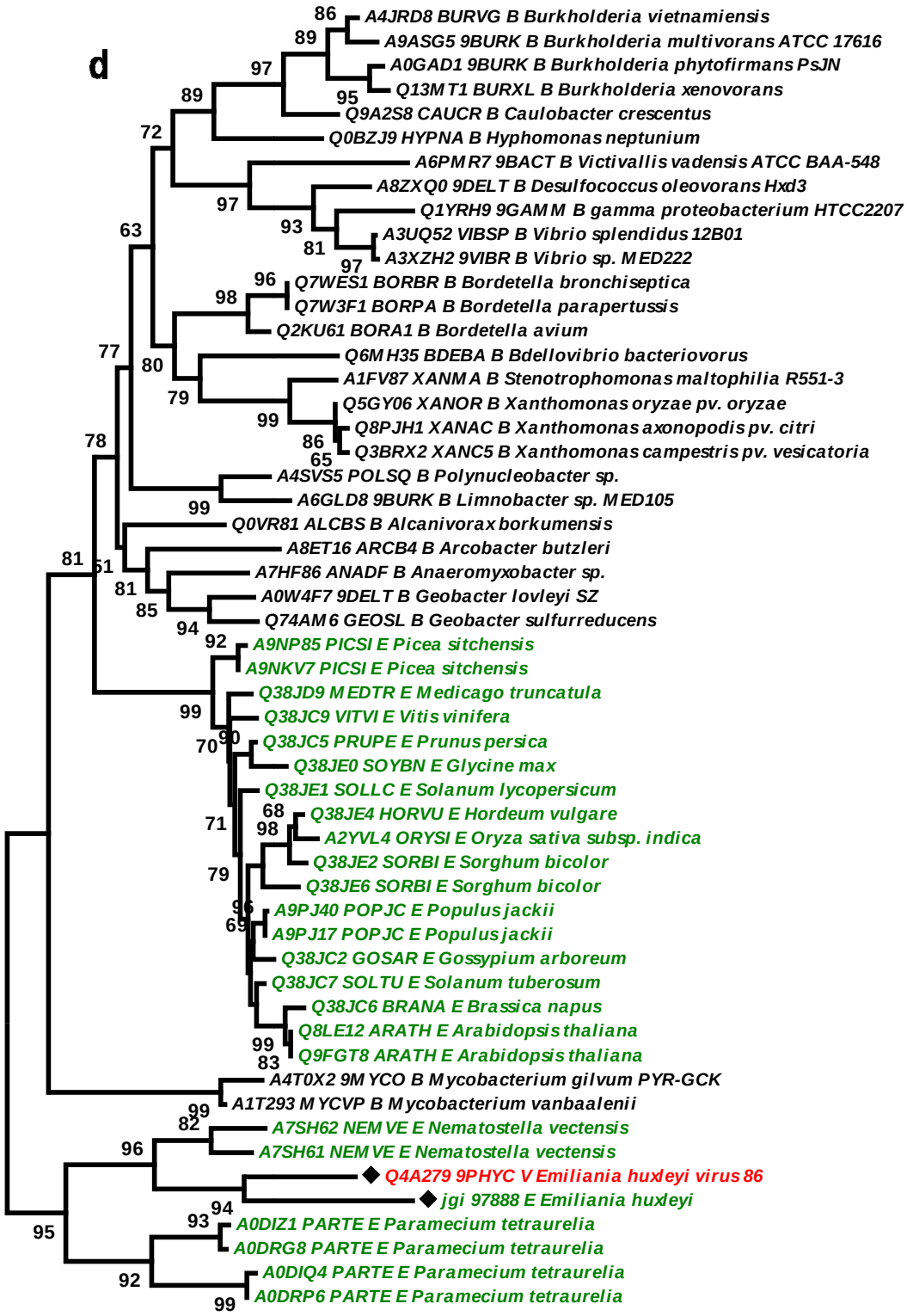

Figure S1d. *(cont.) E. huxleyi* gene 97888: Lipocalin-like (Lipocalins are transporters for small hydrophobic molecules, such as lipids, steroid hormones, bilins, and retinoids.).

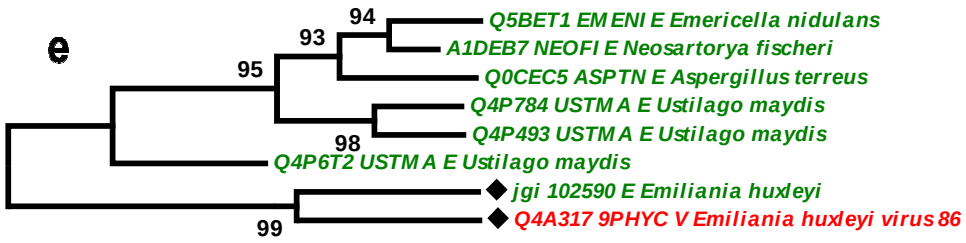

Figure S1e . *(cont.) E. huxleyi* gene 102590: ERG3, Sterol desaturase.

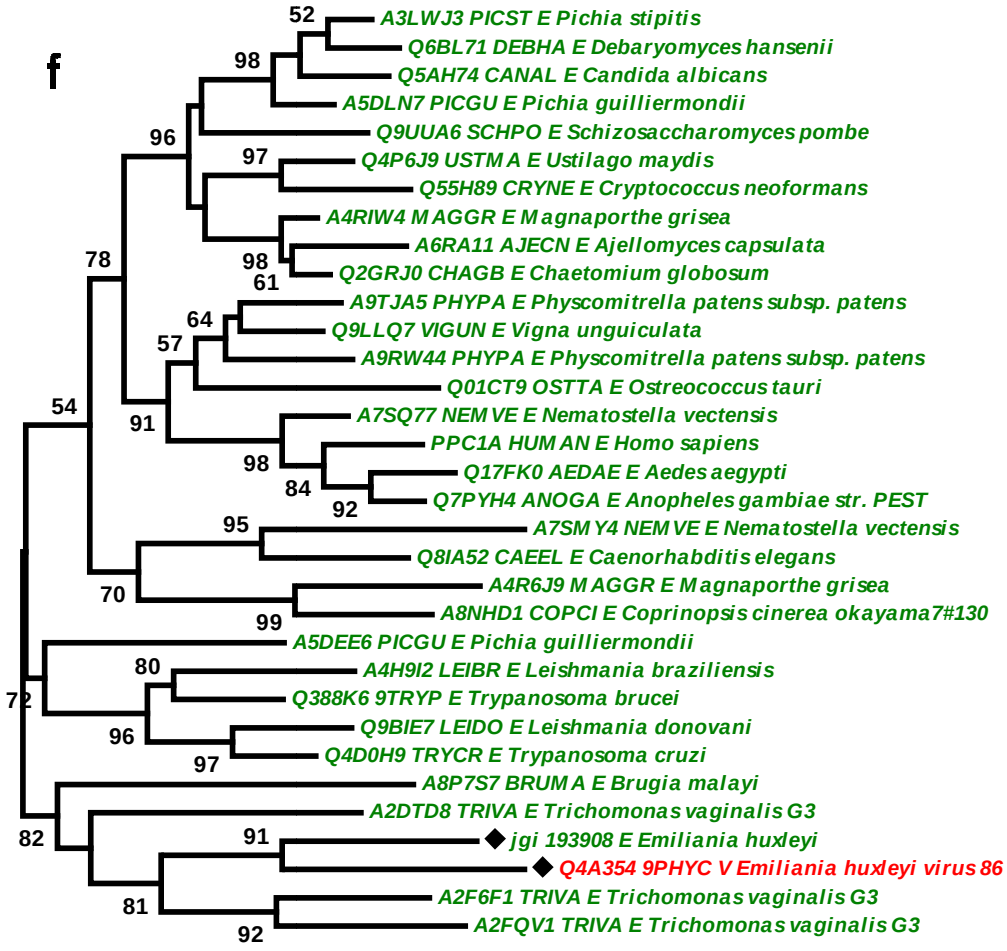

Figure S1f. *(cont.) E. huxleyi* gene 193908: Lipid phosphate phosphatase (PAP2 superfamily).

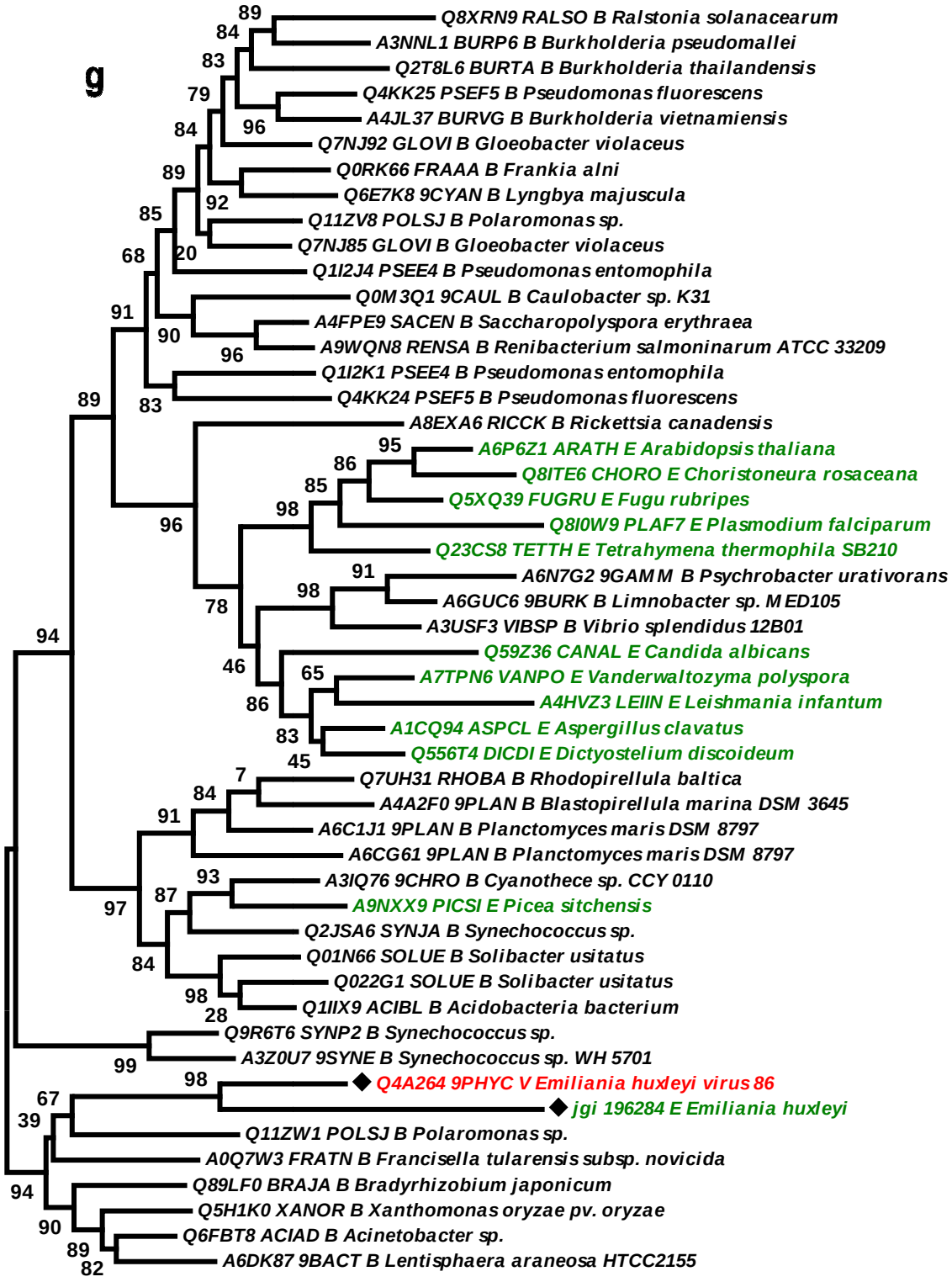

Figure S1g. *(cont.) E. huxleyi* gene 196284: Fatty acid desaturase (Aco1, delta-9).

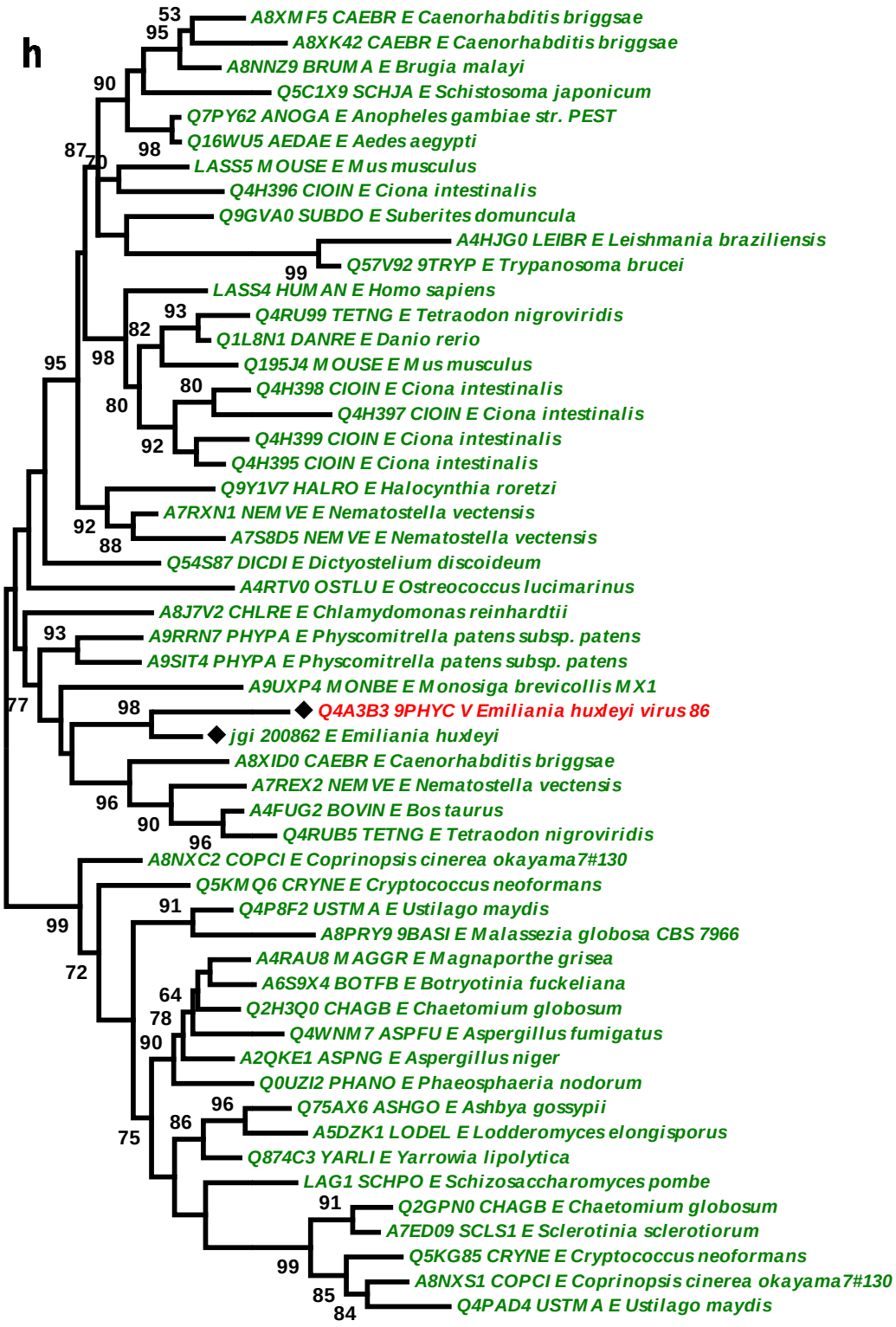

Figure S1h. *(cont.) E. huxleyi* gene 200862: Dihydroceramide synthase (longevity-assurance LAG1 family).

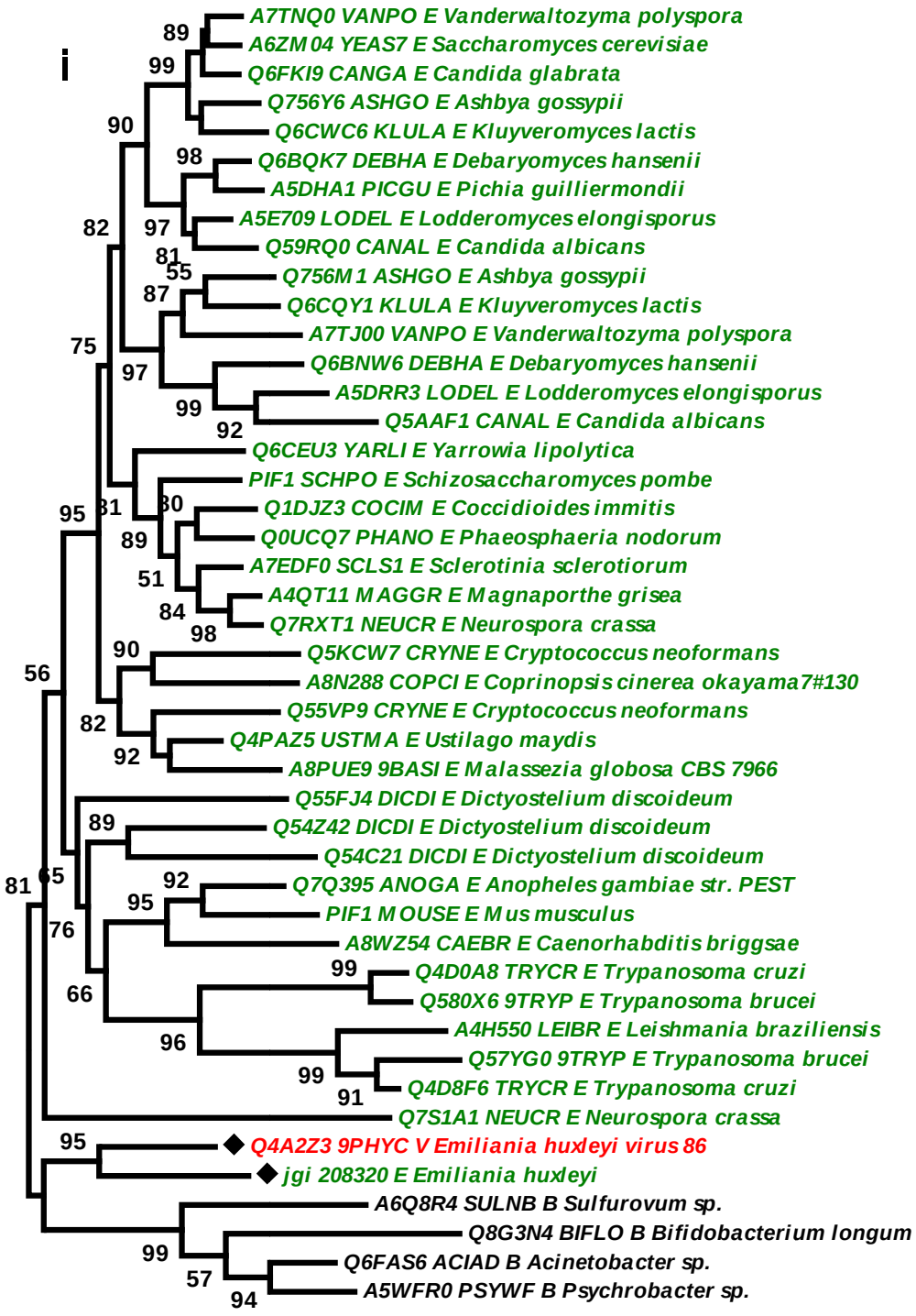

Figure S1i. *(cont.) E. huxleyi* gene 208320: DNA repair and recombination protein pif1-like with HRDC (Helicase and RNase D C-terminal) domain.

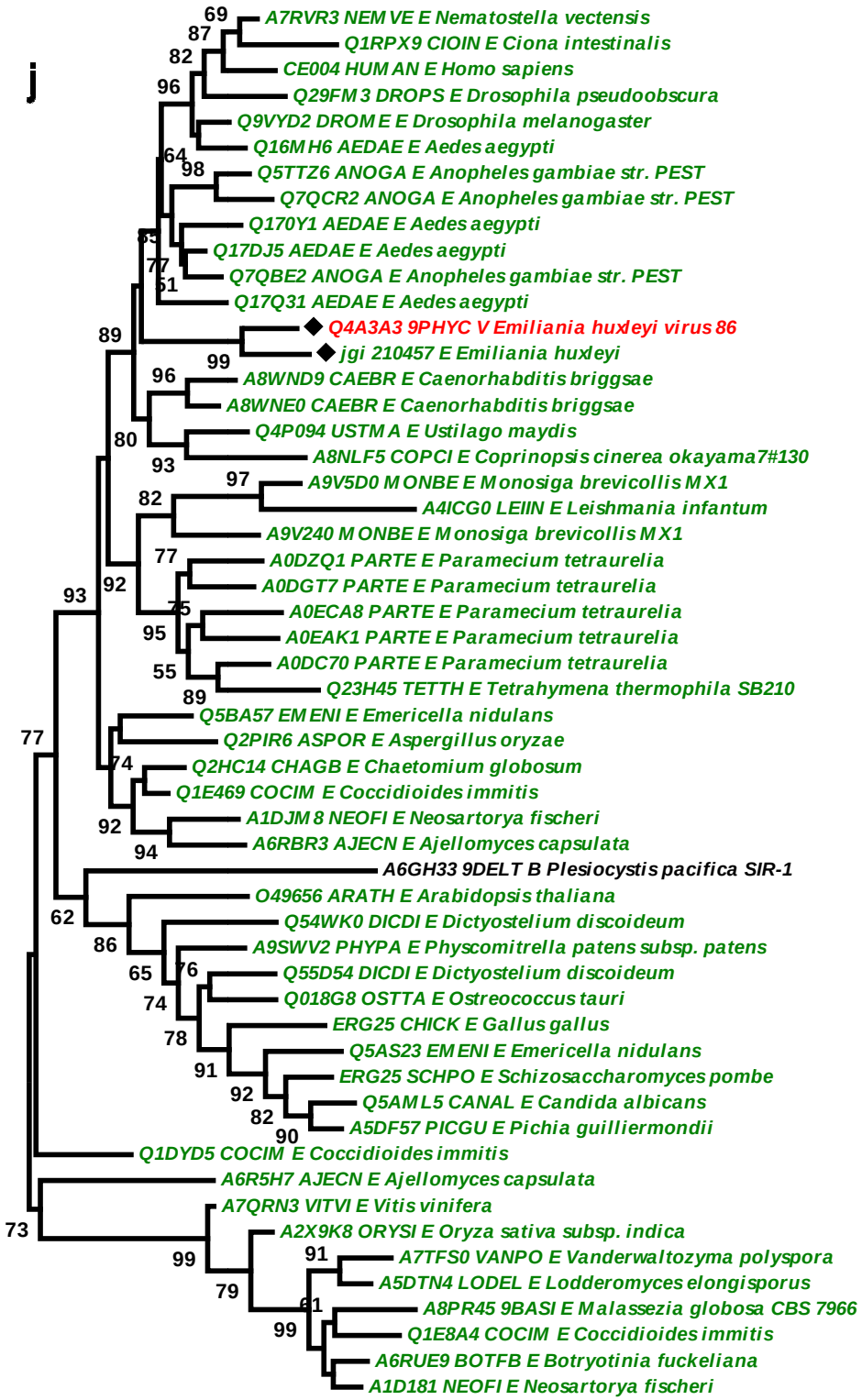

Figure S1j. *(cont.) E. huxleyi* gene 210457: ERG3, Sterol desaturase.

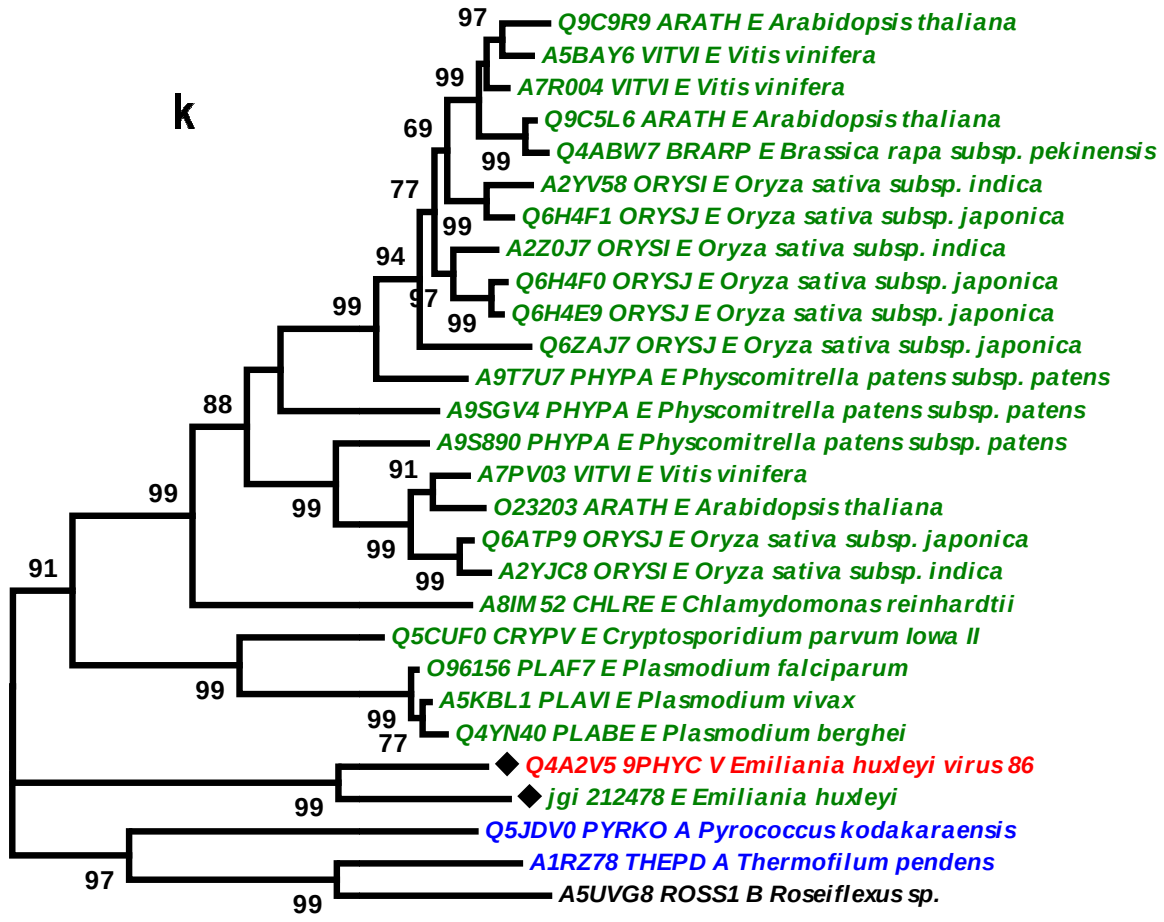

Figure S1k. *(cont.) E. huxleyi* gene 212478: MFS-1, Major Facilitator Superfamily.

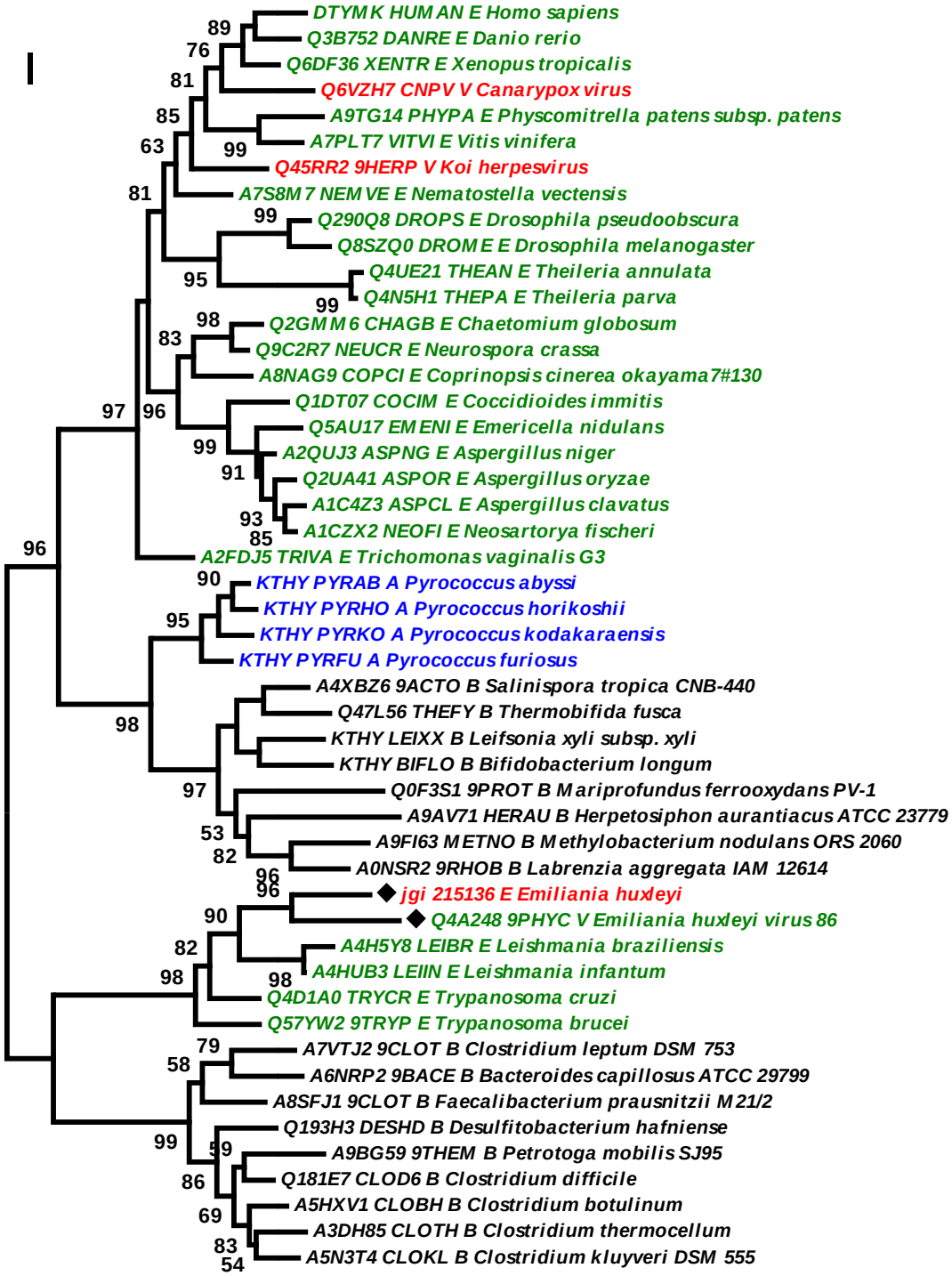

Figure S1l. *(cont.) E. huxleyi* gene 215136: Tmk, Thymidylate kinase.

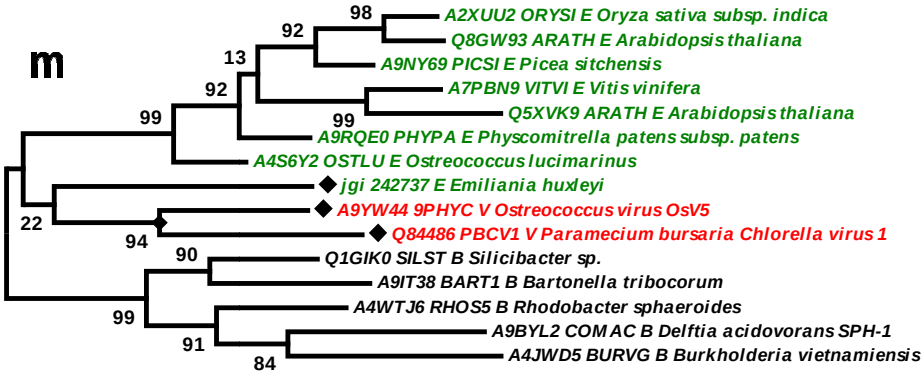

Figure S1m. *(cont.) E. huxleyi* gene 242737: YqaJ viral recombinase family.

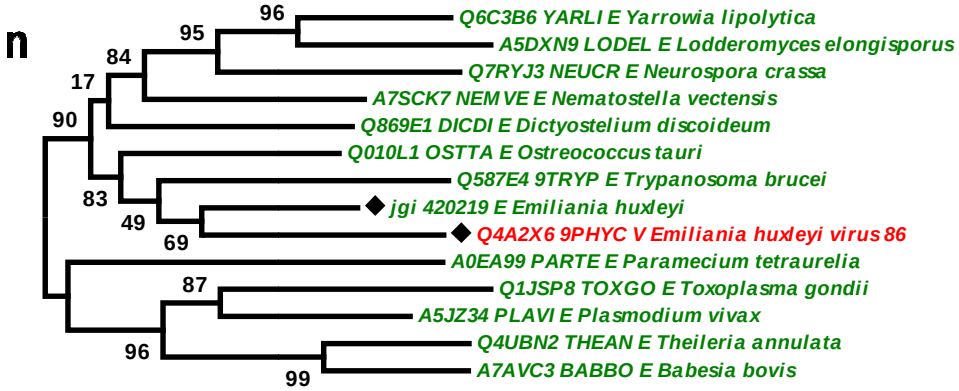

Figure S1n. *(cont.) E. huxleyi* gene 420219: ATP-dependent DNA ligase.

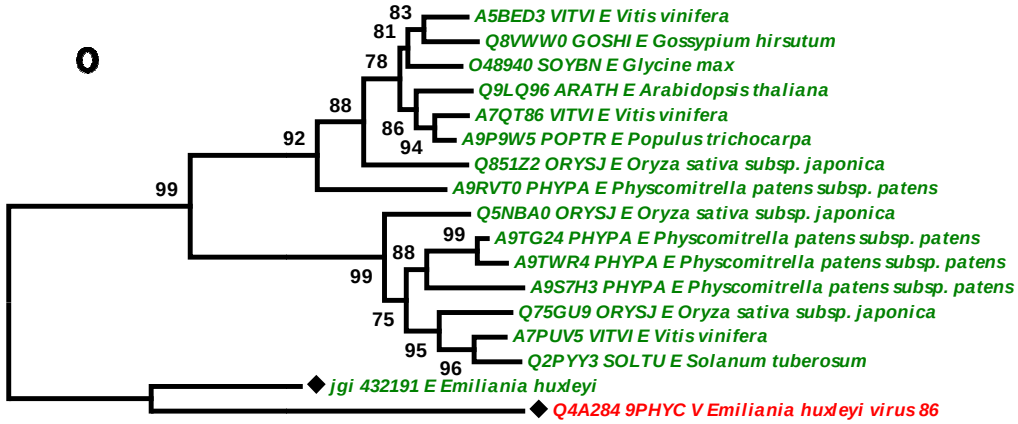

Figure S1o. *(cont.) E. huxleyi* gene 432191: Sec14p-like lipid-binding domain.

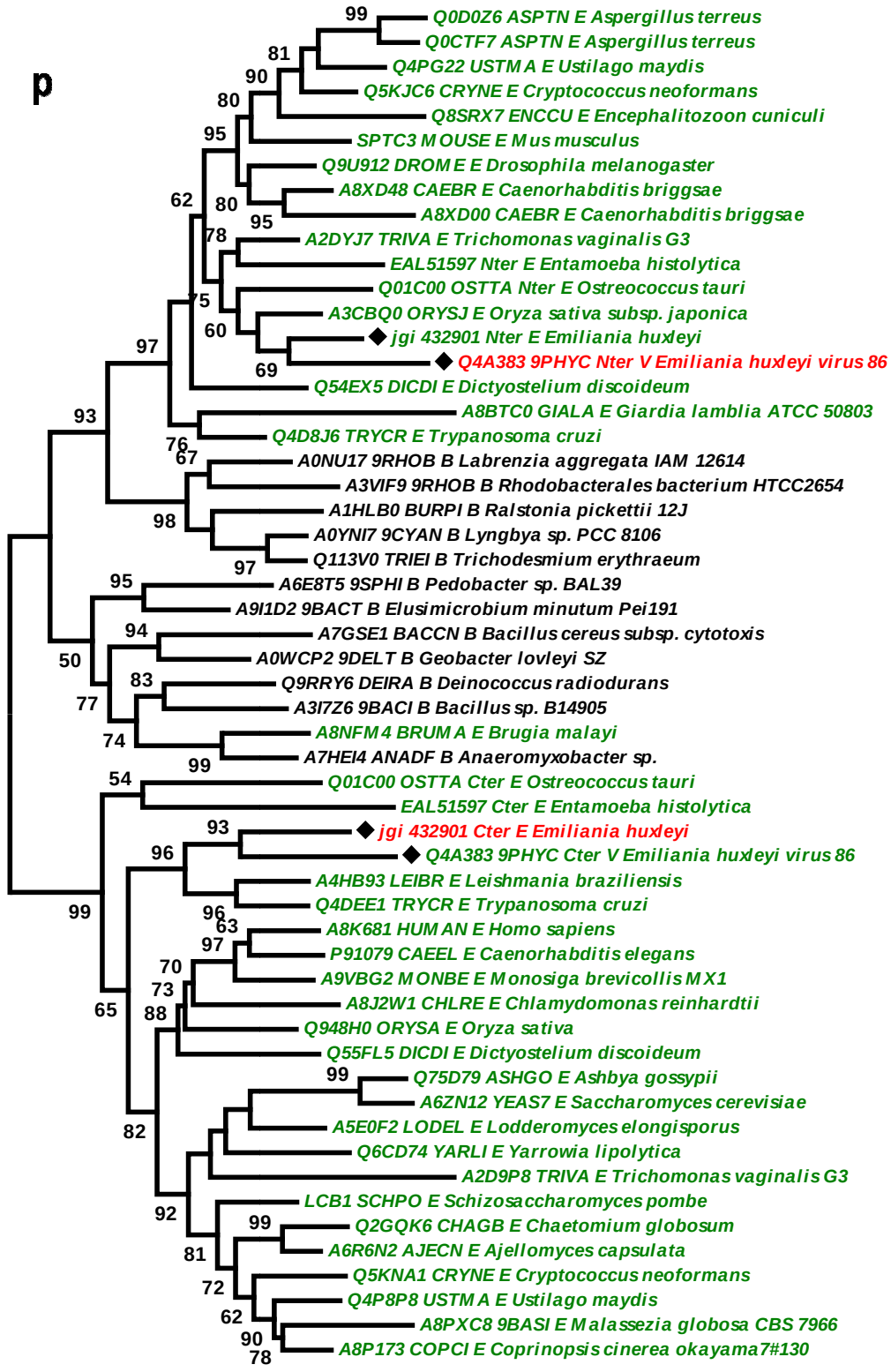

Figure S1p. *(cont.) E. huxleyi* gene 432901: Serine palmitoyltransferase (tri-domain architecture of LCB2/LCB1/PAP2): this tree shows the sequence relationships for the N-terminal LCB2 and C-terminal LCB1 domains.

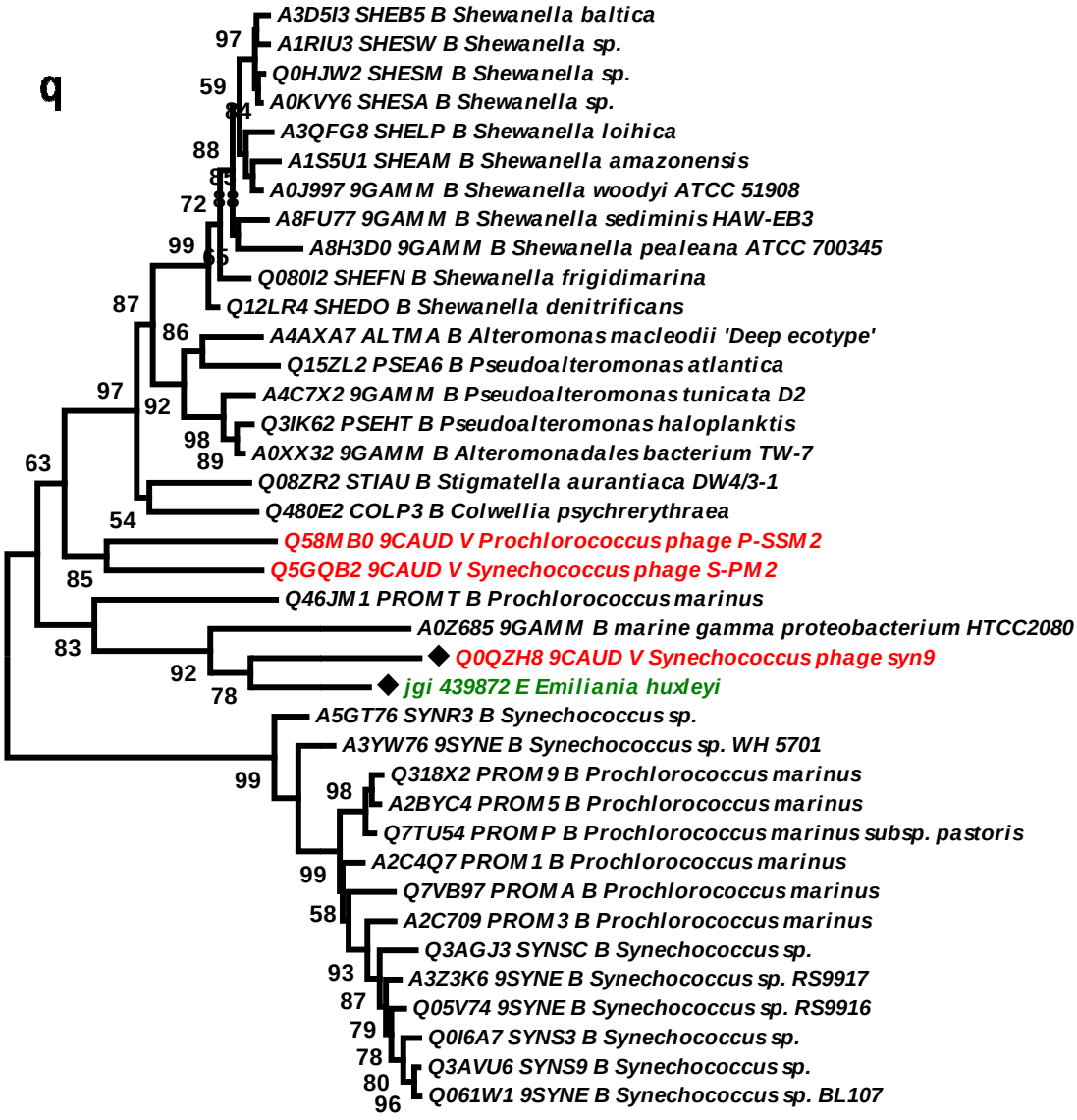

Figure S1q. *(cont.) E. huxleyi* gene 439872: 2OG-FeII_Oxy domain-containing protein (Prolyl 4-hydroxylase alpha subunit-like).

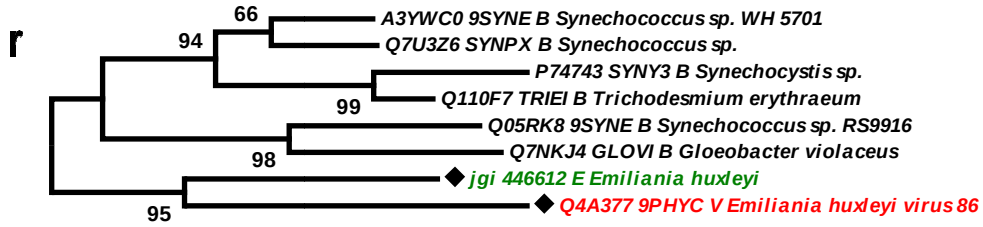

Figure S1r. *(cont.) E. huxleyi* gene 446612: Methyltransferase.

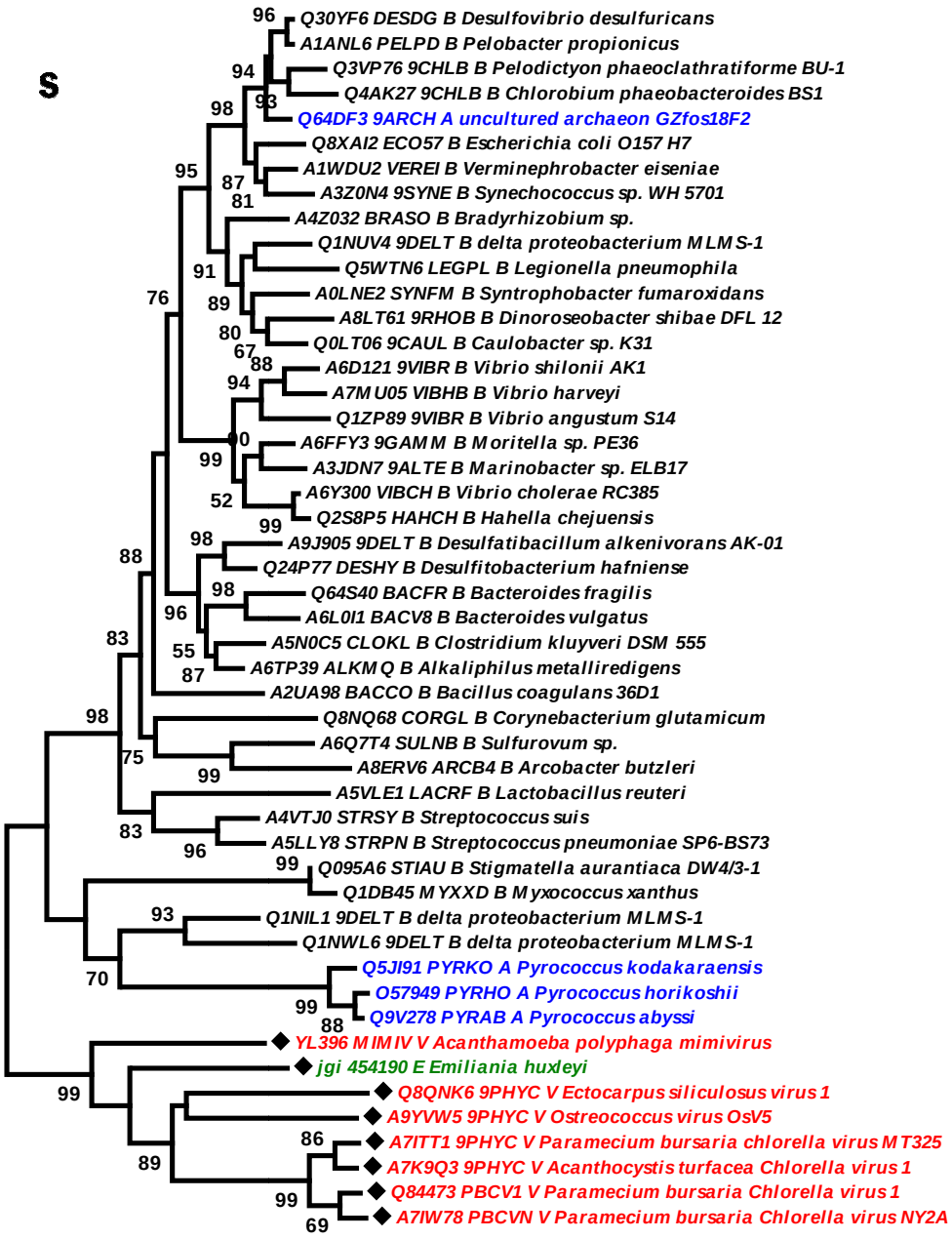

Figure S1s. *(cont.) E. huxleyi* gene 454190: SSL2, DNA or RNA helicases of superfamily II.

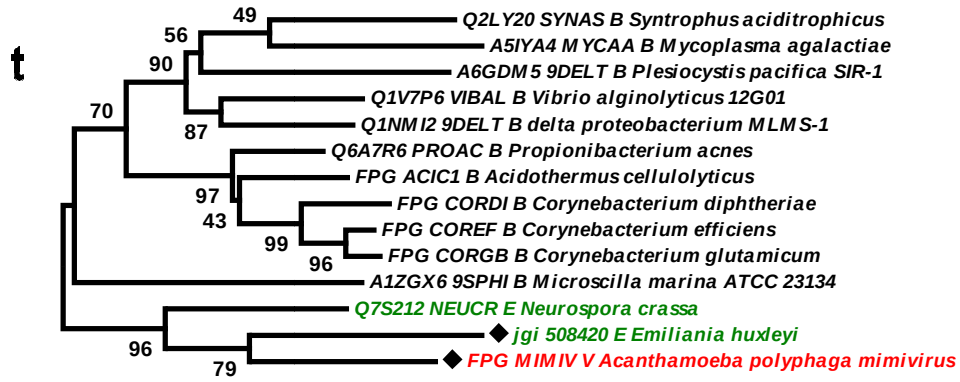

Figure S1t. *(cont.) E. huxleyi* gene 508420: Formamidopyrimidine-DNA glycosylase.

Annexe B.

Uncoupling of *Emiliania huxleyi* photosynthesis: virus infection versus nutrient stress

Susan A. Kimmance[1], Michael J. Allen[1], Joaquin Martínez- Martínez[3], António Pagarete[2] and William H. Wilson[1,3]

[1] *Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, PL1 3DH, UK.*

[2] *Equipe EPPO-Evolution du Plancton et PaléoOcéans, CNRS-UMR7144, Université Pierre et Marie Curie, Station Biologique, FR-29682 Roscoff, France.*

[3] *Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, 04575 ME, USA.*

1. Summary

Viruses can significantly reduce primary production mediated carbon cycling through infection and lysis of a wide range of autotrophs. The cellular mechanisms involved in lytic viral infection and the impacts on host physiology are not well understood, but infection inevitably results in physiological consequences to host metabolism. Here we assessed infection dynamics in a natural *Emiliania huxleyi*/coccolithovirus system to determine the implications for host photophysiology and efficiency of PSII photochemistry. A nutrient-enriched mesocosm environment off the coast of Norway was used to identify the photophysiological mechanisms that distinguish the response of natural *E. huxleyi* populations to abiotic (nutrient limitation) versus biotic (viral) stress. The magnitude of the *E. huxleyi* blooms and their fate appeared to be dependent on phosphate availability. Prior to bloom collapse there was an uncoupling of the diel pattern of photosynthesis which appeared to be linked to viral stress. Interaction between nutrient and viral stress increased the magnitude of photophysiological suppression. This is the first demonstration of a direct viral impact on photosystem II (PSII) photochemistry in natural coccolithophore populations, and as *E. huxleyi* plays such a vital role in the global carbon cycle, it may have important implications for carbon and nutrient flux.

** Article under preparation. António Pagarete was involved in this work, contributing to the mesocosm experiment set-up, development, and sampling stages.*

2. Introduction

Not only are viruses the most abundant entities in our oceans (Bergh et al., 1989; Suttle, 2007) but they are capable of infecting the main primary producers, and play a crucial role in both nutrient (Wilhelm and Suttle, 1999) and biogeochemical cycling (Fuhrman, 1999). Yet the precise physiological and, in turn, ecological effects of viral infection are poorly understood due to the high complexity of natural systems. Laboratory based experimentation, usually based on limited numbers of strains of interest can offer a glimpse at the workings of a host/viral system, yet it lacks the realism of the natural environment where whole communities interact at all trophic levels. Mesocosm experiments have gone some way to address the need to study whole communities in as natural an environment as possible. Through the enclosure of natural bodies of water in transparent bags, mesocosms offer the opportunity to study natural systems under semi-controlled conditions. Their realism, reliability and reproducibility (Egge and Heimdal, 1994; Martinez-Martinez et al., 2006) make them well suited for studying the effects of chemical, physical and biological manipulations on natural plankton communities.

The bloom-forming *Emiliania huxleyi* (Lohmann) Hay and Mohler (Prymnesiophyceae) is the most abundant species of coccolithophore in the world's oceans. *E. huxleyi* greatly impacts on marine ecosystems and in particular, on the global carbon and sulphur cycles (Burkill et al., 2002; Westbroek et al., 1993). Blooms of this ubiquitous microalga are known to affect the oceanic carbon pump (Elderfield, 2002) and climate (Charlson et al., 1987). As such, it is one of the most intensively studied phytoplankton species. Until recently, the mechanisms of *E. huxleyi* bloom disintegration were poorly understood but it is now accepted that viruses (termed coccolithoviruses) are intrinsically linked to these sudden crashes (Bratbak, Egge, and Heldal, 1993; Jacquet et al., 2002; Schroeder et al., 2003; Wilson et al., 2002), and a diverse range of coccolithoviruses that infect *E. huxleyi* have now been isolated and characterized (Allen et al., 2007; Schroeder et al., 2002; Wilson et al., 2005b; Wilson et al., 2002).

Viral infection of a phytoplankton cell inevitably results in physiological consequences to the host metabolism leading to the activation of stress and defence mechanisms (Evans et al., 2006) which in turn can alter growth dynamics and lifecycle (Frada et al., 2008). Changes in photosynthetic activity can have major implications for the physiological status of a cell, which may have consequences for its fate. Recently, changes in photophysiology using chlorophyll fluorescence measurements have become an important

and easily measurable parameter of the photophysiological state of phytoplankton such as *E. huxleyi*. Variations in parameters such as the maximum quantum yield of PSII photochemistry (F_v/F_m) and the absorption cross-section of PSII ($_{\sigma PSII}$) are frequently used as indicators of physiological stress in phytoplankton (Geider and Laroche, 1994; Moore et al., 2006; Timmermans et al., 2001).

However, there are relatively few studies on the photophysiological response of phytoplankton to viral stress and none previously has investigated this phenomenon in natural populations. A previous *E. huxleyi* culture study suggests that viral infection elevates oxidative stress (Evans et al., 2006). It is likely that this is linked to interruption of electron transport between photosystems I and II (Balachandran et al., 1997), which results in photochemical quenching of excess energy through fluorescence to prevent photoinhibition (Seaton, Lee, and Rohozinski, 1995). Photoinhibition is characterized by damage to the proteins, lipids, and pigments of the photosynthetic membrane, and in particular, the vulnerable D1 protein (Melis, 1999) and oxygen-evolving complex associated with PSII (Nishiyama, Allakhverdiev, and Murata, 2006). Abiotic and biotic stresses dictate the degree of photoinhibition by inhibiting the rate of repair of PSII, thus exacerbating photodamage. Ultimately this may cause reduced rates of carbon fixation resulting in an overall reduction in the photosynthetic rate (Hewson, O'Neil, and Dennison, 2001). Thus, photosynthetic efficiency is a good indicator of the metabolic and physiological status of chloroplast-containing cells (Seaton, Lee, and Rohozinski, 1995).

Disruption of cellular processing or photophysiological state caused by stress factors such as virus infection or nutrient limitation should be measurable by observing changes in PSII photochemistry. To test this hypothesis, we took advantage of a field mesocosm experiment that was conducted at the large scale facilities at the Marine Biological Field Station, University of Bergen, Norway in June 2008. During this experiment large volume (>10 m3) mesocosm enclosures were nutrient-manipulated to induce *E. huxleyi* blooms under both P-replete and P-deplete conditions to investigate the role of P-availability on coccolithovirus - *E. huxleyi* dynamics (as part of a separate study, Pagarete et al., 2009). More specifically, this field study allowed us to take opportunistic samples to investigate temporal photosynthetic community photophysiology profiles and interpret them in the context of nutrient availability and microbial community succession dynamics. Using a fluorescence-based method to assess changes in PSII photochemical efficiency, PSII antenna size and PSII photoinhibition, the overall aim was to help identify the photophysiological mechanisms that distinguish the response of natural *E. huxleyi* populations to viral versus nutrient stress.

3. Materials and methods

Study site and experimental design

The mesocosm experiment was carried out in Raunefjorden, western Norway, at the Marine Biological Field Station, Espeland 20 km south of Bergen, from 2nd until June 25th 2008. Six enclosures of 11 m^3 (4 m deep and 2 m wide) made of 0.15 mm thick polyethylene (90% light penetration of the photosynthetic active radiation) were mounted on floating frames moored along the south side of a raft in the middle of the bay (for details see Egge and Aksnes, 1992) and numbered 1 to 6 in the east-west direction (Fig. 1). The enclosures were filled simultaneously by piecemeal accretion on June 3rd with unfiltered, unscreened seawater from 6 m depth using a submersible centrifugal pump. The seawater in the enclosures was kept homogeneous by means of airlifts.

Nutrient enrichment

The 6 mesocosms were divided into 2 treatment groups allowing triplication of each treatment: phosphate-replete (enclosures 1, 3 and 5) and phosphate-deplete (enclosures 2, 4 and 6). To induce a bloom of *Emiliania huxleyi* nutrients were added daily as concentrated stock solutions (between 13:00 and 14:00 h; after the daily sampling) in a N:P ratio of 15:1 (1.5 µmol L^{-1} $NaNO_3$ and 0.1 µmol L^{-1} KH_2PO_4) to the P-replete enclosures and at a ratio of 75:1 (1.5 µmol L^{-1} $NaNO_3$ and 0.02 µmol L^{-1} KH_2PO_4) to the P-deplete mesocosms from 5 to 25 June. Nutrient concentrations were analysed once daily using standard methods (Strickland and Parsons, 1968) adapted to an autoanalyzer equipped with autosampling, detection and computing methods from SANplus segmented Flow Analyzer (Skalar Analytic). Briefly, a 100 mL sample was taken daily from each enclosure, preserved in chloroform (0.8 % final concentration) and stored in the dark at 4°C prior to analysis.

Physical and environmental parameters

Temperature, salinity and oxygen concentration were measured daily in all enclosures using a multi-parameter water quality monitor OTS, Isi Model 85 (data not shown).

Meteorological data was obtained from Space Monitoring Information Support laboratory (SMIS IKI RAN).

Phytoplankton abundance and composition

Phytoplankton composition and abundance estimates were determined four times daily (06:00, 12:00, 18:00 and 00:00 h) from all enclosures and directly from Raunefjorden surface waters (adjacent to enclosures 3 and 4) by analysis of fresh samples on a FACScan flow cytometer (Becton Dickinson, Oxford, UK) equipped with a 15 mW laser exciting at 488 nm and with a standard filter set up. Samples were analysed at high flow rate (~70 µL min^{-1}) and specific phytoplankton groups were discriminated by differences in their forward or right angle light scatter (FALS, RALS) and chlorophyll (and phycoerythrin for *Synechococcus* populations) fluorescence. Files were analysed using WinMDI 2.8 software (Joseph Trotter, [http://facs.scripps.edu]).

Virus abundance.

Virus abundance was determined four times daily (see Phytoplankton abundance above) using the flow cytometric protocol of (Brussaard, 2004b). Samples for viral analysis were fixed with glutaraldehyde (0.5 % final concentration) for 30 min at 4°C, snap frozen in liquid nitrogen and stored at –80°C. Samples were subsequently defrosted at room temperature and diluted 500 fold with TE buffer (10 mmol L^{-1} Tris-HCL pH8, 1 mmol L^{-1} EDTA), stained with SYBR Green 1 (Molecular Probes, Marie et al., 1999) at a final dilution of 5 x 10^{-5} the commercial stock, incubated at 80°C for 10 min in the dark, then allowed to cool for 5 min before analysis using a FACSort flow cytometer (Becton Dickinson, Oxford, UK). Samples were analysed for 2 min at a flow rate of ~ 70 µL min^{-1} and virus groups were discriminated on the basis of their RALS versus green fluorescence. Data files were analysed using WinMDI 2.8 software (see above).

Photophysiology measurements using FIRe fluorometry

To assess phytoplankton photophysiological potential for processing light with respect to both nutrient limitation and viral infection, discrete variable chlorophyll fluorescence measurements were acquired using a fluorescence induction and relaxation (FIRe)

fluorometer (Satlantic Inc., Halifax, Nova Scotia, Canada). Prior to each fluorescent measurement samples were dark-adapted for 30-45 min at a controlled-temperature to match in-situ conditions.

Dark-adapted samples (3 mL) were analysed three times daily (12:00, 18:00 and 00:00 h) within cylindrical 1 cm path length cuvettes placed into the FIRe fluorometer cuvette holder. Excitation was provided by a high luminosity blue and green LED array (450 and 500 nm peak heights, each with 30 nm bandwidth). The FIRe was employed with a four-step measurement protocol: (1) single turnover (ST) excitation from a 100 µs pulse, (2) ST relaxation from a weak modulated light over 500 ms, (3) multiple turnover (MT) excitation from a 100 ms pulse, and (4) MT relaxation from a weak modulated light over 1 s. Forty sequential acquisitions of each four-step sequence were cumulatively averaged for each fluorescence profile to increase the signal-to-noise ratio. Fluorescence profiles were fitted with the biophysical (KPF) model of Kolber et al. (1998) using the software FIREPRO (v.1.3, Satlantic Inc.). Filtered (0.2 µm) sample blanks were analysed at the gain chosen for the measurement on the sample and subtracted from the sample fluorescence sequence at the time of fitting the KPF model. The retrieved PSII photochemistry parameters utilized in this study are the minimum (F_o) and maximum (F_m) fluorescence yields, the maximum photochemical efficiency of PSII (F_v/F_m) and the relative functional absorption cross-section of PSII, σ_{PSII}, (the product of the light-harvesting capability of the light-harvesting pigments and the efficiency of excitation transfer to the reaction centre (Kolber and Falkowski, 1993).

4. **Results**

Nutrient availability

Changes in nutrient concentrations in the mesocosms reflected the daily additions made during the 17-day experimental period. Phosphate (P) concentrations in enclosures 2, 4, and 6 (P-deplete) were almost identical and remained at close to zero for most of the experiment (Fig. 1a); with the exception of day 12 where there was a pulse of P in all mesocosms to <0.1 µmol L^{-1}. In contrast, mesocosms 1, 3, and 5 (P-replete) were more variable, with P concentrations ranging from 0 to 0.23 µmol L^{-1} (Fig. 1a). Although P was higher in P-replete mesocosms during phase 1 and into phase 2 (see below for description of

bloom phases), P-concentrations then decreased during phase 2 and were more variable again during phase 3.

The difference between P-replete and P-deplete mesocosms was most noticeable for nitrate (N) concentrations, and a deviation between treatments was visible from day 4 onwards (Fig. 1b). N concentrations fluctuated but remained consistently low, < 2 µmol L^{-1} in P-replete mesocosms with a marked decline in phase 2 of the bloom after day 9 to beyond the limit of detection; this was followed by an observed increase to 1 µmol L^{-1} in phase 3. In contrast, P-deplete mesocosms revealed a steady rise in N concentrations from the start of the study well into phase 2, from undetectable to between 8 - 9 µmol L^{-1}. This was followed by a decrease in N starting between days 11 – 13 in phase 2 which dropped to < 2 µmol L^{-1} in phase 3 (Fig. 1b). N:P ratios largely reflected the N concentrations in each of the treatments since P concentrations were so low (Fig. 1c). N:P ratios suggest that P-replete mesocosms were N-limited for most of the experimental period and P-deplete mesocosms were P-limited.

Mesocosm bloom dynamics

The photosynthetic community in the mesocosms was composed of 4 major discernable groups, measurable by flow cytometry: *Synechococcus*, picoeukaryotes, nanoeukaryotes and *Emiliania huxleyi* (see Fig 2 for description of groups). Temporal progression and succession of the photosynthetic community was split into 3 distinct phases over the 17-day study (Fig. 3).

Phase 1 was between days 2 – 7 of the study and was characterised by rapid increase then decline of picoeukaryotes (Fig. 3b) and nanoeukaryotes (Fig. 3c). Cell concentrations increased from between 500 – 1000 mL^{-1} to a maximum of 17,046 cells mL^{-1} for picoeukaryotes, and over 37,336 cells mL^{-1} for nanoeukaryotes in the space of 3 days, then immediately crashed back to initial concentrations over the next 3 days. *Synechococcus* concentrations remained between 5,000 – 10,000 cells mL^{-1} (Fig. 3a), gradually decreasing in each mesocosm during the phase 1 period. *E. huxleyi* concentrations gradually increased during this phase from only approx. 200 cells mL^{-1} to a maximum of 10,000 cells mL^{-1} (Fig. 3d). There was no discernable difference in photosynthetic community abundance between P-deplete and P-replete treatments in phase 1 (Figs. 3).

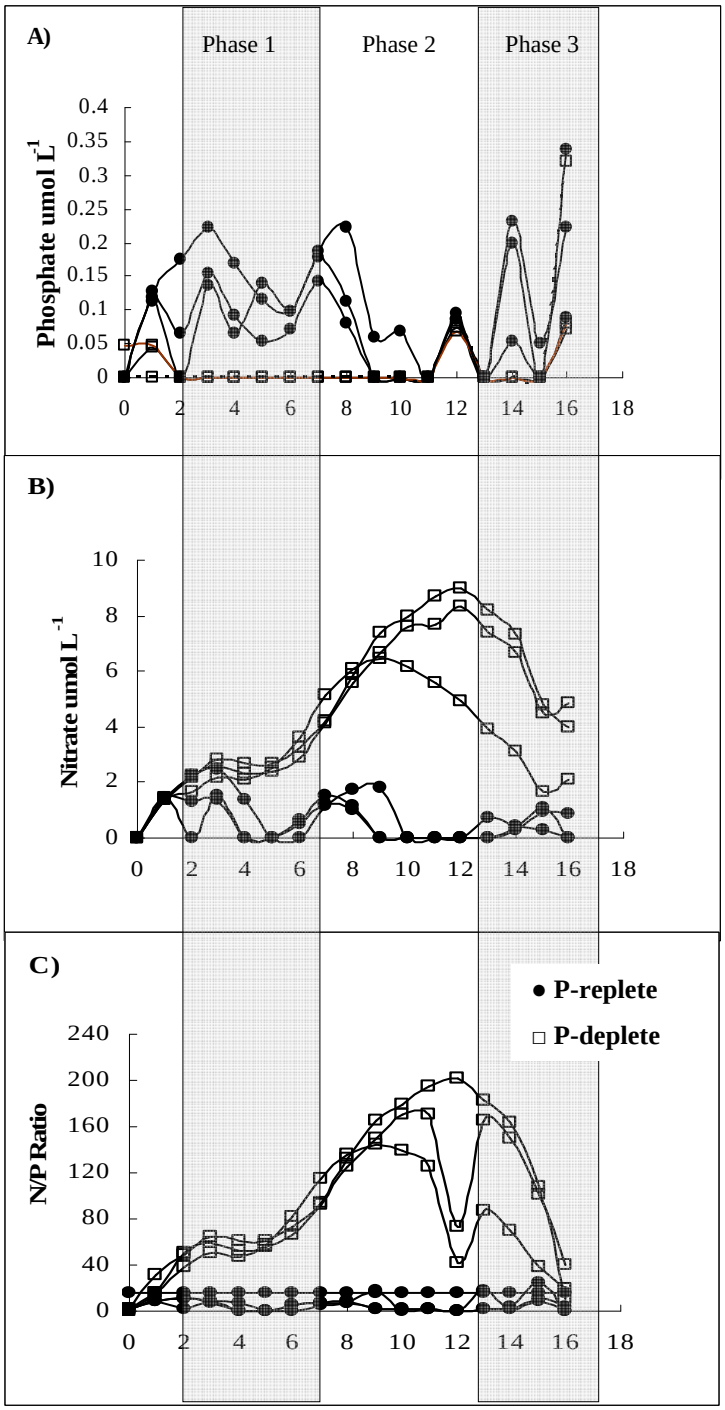

Figure 1. Dynamics of (A) PO_4 and (B) NO_3 concentrations measured once daily in P-replete mesocosms 1, 3 and 5; (1.5 µmols, L^{-1} $NaNO_3$ and 0.1µM KH_2PO_4) and P-deplete mesocosms 2, 4 and 6; (1.5 µmols, L^{-1} $NaNO_3$ and 0.02 µmols, L^{-1} KH_2PO_4) during the 17 day experiment in June 2008. Fig. 1C shows the progression of the ratio between N and P.

Phase 2 was between days 7 – 13 of the study and was primarily characterised by exponential growth of the *E. huxleyi* population (Fig. 3d). A clear split was observed between P-replete and P-deplete treatments, with maximum cell concentrations reaching 170,944 mL^{-1} in the P-replete treatment towards the end of phase 2, where they reached apparent stationary phase. In addition, *E. huxleyi*-specific virus (EhV) concentrations started to increase exponentially by the end of phase 2 with concentrations reaching 4 x 10^6 mL^{-1} from a low of

4 x 10^5 mL^{-1} at the start of the phase in the P-replete treatments (Fig. 3e). In the P-deplete treatment, *E. huxleyi* cells showed lower growth rates and were still in exponential phase towards the end of phase 2 with maximum cell concentrations only reaching 70,125 mL^{-1}. EhV concentrations remained low and variable in the P-deplete phase 2, ranging from undetectable up to a maximum of 1 x 10^6 mL^{-1}.

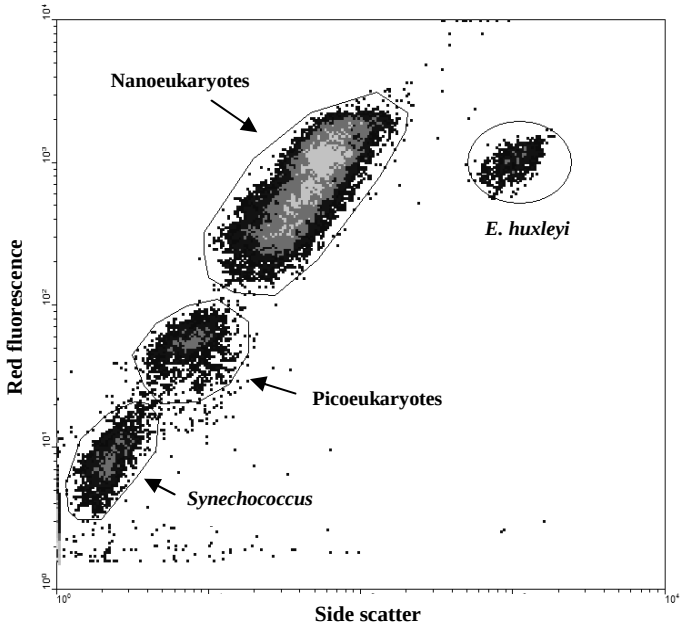

Figure 2. Typical flow cytometry scatter plot distinguishing the 4 groups identified in the photosynthetic community (*Emiliania huxleyi*; nanoeukaryotes; picoeukaryotes and *Synechococcus*). Populations were discriminated based on their differences in forward or right angle light scatter (FALS, RALS) and chl *a* red fluorescence (FL3) and phycoerythrin for *Synechococcus* populations (FL2).

Other components of the photosynthetic community exhibited mixed responses in phase 2 (Figs. 3) although, in general there was an observed bifurcation between P treatments. *Synechococcus* concentrations started increasing exponentially in the P-deplete treatment reaching a maximum of 43,510 cells mL^{-1}. Interestingly, the P-replete populations exhibited a variable response, reflecting differences observed between mesocosms 1, 3 and 5: an increase of different magnitudes followed by a crash in 1 and 3, versus a drop then increase in mesocosm 5. However all three mesocosms ended up at similar concentrations (approx. 20,000 cells mL^{-1}) by the end of phase 2; (Fig. 3a). Picoeukaryote and nanoflagellate populations both exhibited slow but variable increases in abundance, with differences starting to emerge between P treatments by the end of phase 2.

Phase 3 was between days 13 – 17 of the study and was primarily characterised by crash of the *E. huxleyi* populations and a concurrent exponential increase in EhV concentrations (Figs. 3d; 3e). *E. huxleyi* in the P-deplete treatments did not start to crash until

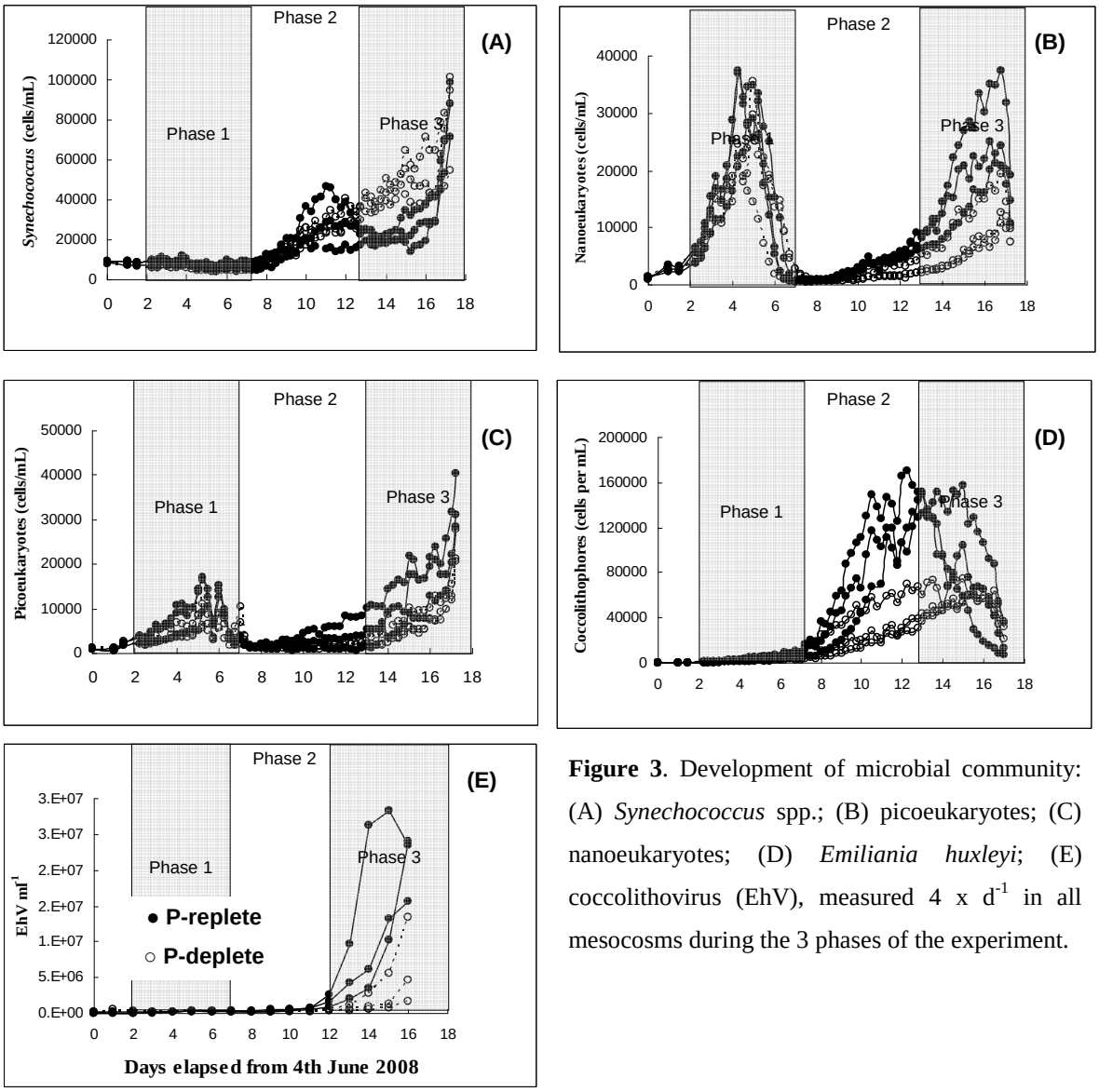

Figure 3. Development of microbial community: (A) *Synechococcus* spp.; (B) picoeukaryotes; (C) nanoeukaryotes; (D) *Emiliania huxleyi*; (E) coccolithovirus (EhV), measured 4 x d^{-1} in all mesocosms during the 3 phases of the experiment.

half way through phase 3, around day 15, after a maximum cell concentration of 76,222 cell mL^{-1}. The magnitude of the crash was lower in the P-deplete treatment, with cell concentrations dropping by approximately 50% of maximum, compared to a 95% - 99% drop in the P-replete treatment. Succession of the other photosynthetic components occurred concurrently, with exponential increases in *Synechococcus*, picoeukaryote and nanoeukaryote populations. Each population exhibited discernable differences in cell concentrations between P-replete and P-deplete treatments. *Synechococcus* showed a clear preference for P-deplete conditions (Fig. 3a) whereas; picoeukaryote and nanoeukaryote populations showed a preference for P-replete conditions (Figs. 3b & 3c respectively).

Photophysiological response

At the start of the experiment (day 0) dark-adapted variable fluorescence measurements showed that there was no difference between the mesocosms regarding photosynthetic community PSII photochemistry. At this point, the ratio of variable to maximum fluorescence (F_v/F_m) was approx. 0.51 ± 0.01 (average for all mesocosms). As described above, succession of the photosynthetic community was split into 3 distinct phases over the 17-day study. The photophysiological response of this community will now be discussed in relation to these 3 phases.

Phase 1. Daily addition of nutrients during this initial phase gradually increased F_v/F_m by approx. 16 % until stabilising at a maximum of approx. 0.59 ± 0.01 around day 5 (Fig. 4a). During Phase 1 the value of F_v/F_m was similar in both the P-replete and P-deplete treatments indicating that the different nutrient treatments did not significantly affect the maximal PSII photochemical efficiency during this time. Similarly, the value of σ_{PSII} was also comparable in both treatments during phase 1 indicating that P-availability did not significantly affect the PSII cross-section during this period. However, in contrast to F_v/F_m, σ_{PSII} decreased by approx. 30 % during the first couple of days after nutrient addition until stabilising around day 3 (Fig. 4b).

Phase 2. After day 8 a clear diel pattern emerged in all mesocosms corresponding to the daily light cycle, with lowest F_v/F_m values recorded at midday. Phase 2 was the point in the experiment at which *E. huxleyi* populations started to increase and become the dominant phytoplankton biomass. During phase 2 there was a separation between the two P treatments with a general trend of lower F_v/F_m in the P-deplete enclosures (Fig. 4a). In particular, towards the end of phase 2, between days 11 and 13, F_v/F_m in the P-deplete treatment was up to 11 % lower at midday than in the P-replete treatment (Fig. 4a). After day 8 the PSII cross-section also followed a diel pattern, with highest σ_{PSII} at noon with a subsequent decrease during the later part of the day. This corresponds with trends in both F_v/F_m and the diel cycle of chlorophyll fluorescence which was typically higher at the earliest part of the day (Fig. 4c). The diel chlorophyll fluorescence pattern was more prominent in the P-deplete treatment and was not visible in the P-replete treatment once the *E. huxleyi* population increased and became the dominant phytoplankton group (Fig. 4c).

From the start of bloom phase 2 there appeared to be a P-treatment effect with higher σ_{PSII} in P-deplete compared to P-replete mesocosms (Fig. 4b). Under P-stress the PSII cross-section increased by up to 40% during phase 2. This also corresponds with a marked

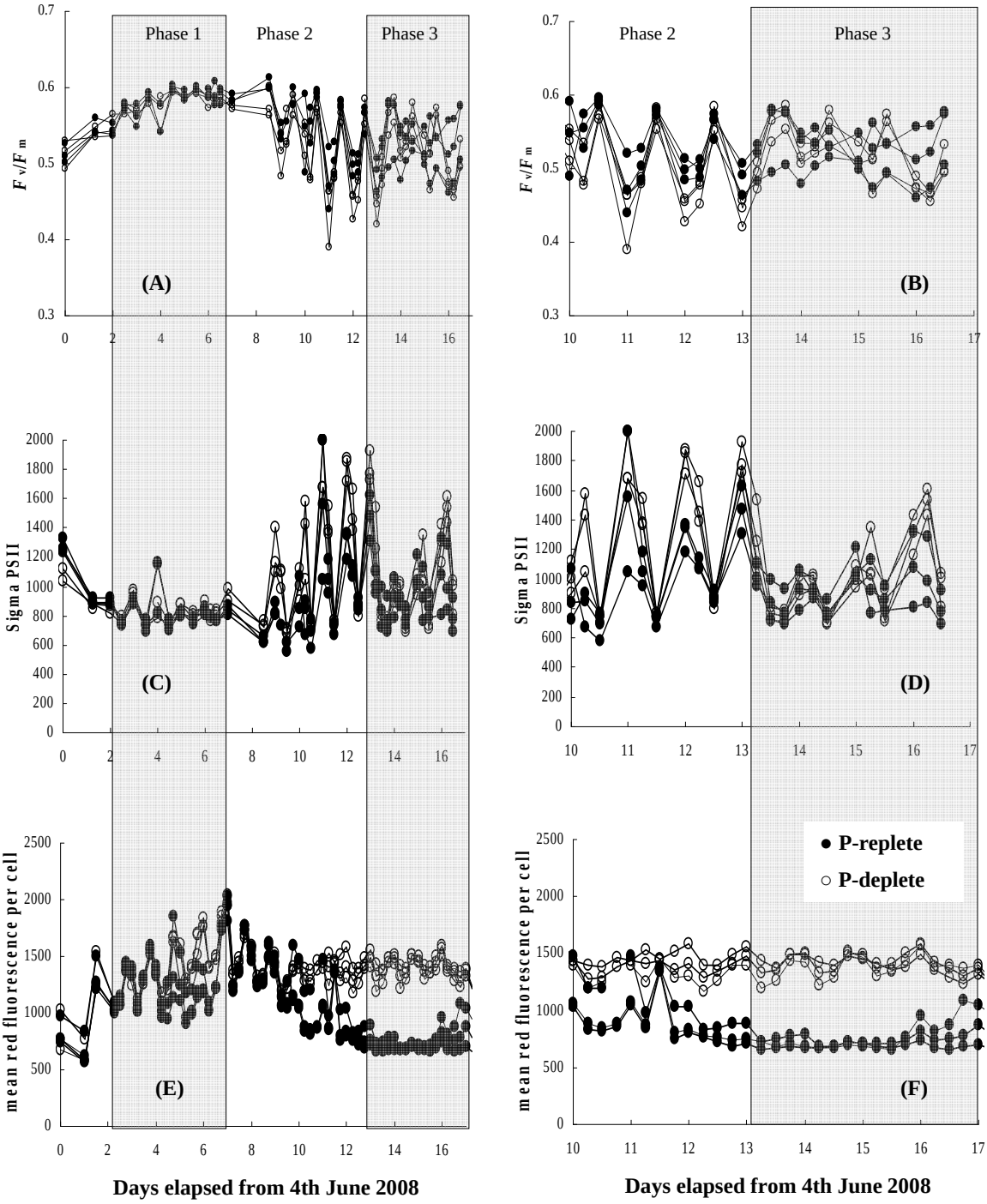

Figure 4. Changes in photosynthetic community fluorescence parameters (A, B) F_v/F_m, (C, D) σ_{PSII}, and (E, F), mean *Emiliania huxleyi* chlorophyll fluorescence per cell (arbitrary units; derived using flow cytometry) measured 3 x d^{-1} in all mesocosms during the 3 phases of the experiment.

separation of mean chlorophyll fluorescence that was observed in both P-treatments after day 9 (with the exception of mesocosm 5), which remained for the duration of the experiment; P-deplete populations had higher mean chlorophyll fluorescence than the P-replete populations (Fig. 4c). Also noticeable was a sharp increase in chlorophyll fluorescence in P-replete enclosures 1 and 3 which occurred towards the end of phase 2 (Fig. 4c) just before these *E. huxleyi* populations entered negative growth (Fig. 5a). For both P-treatments there was a correlation between declining growth rate and increasing σ_{PSII} around the point of negative growth (Fig. 5b).

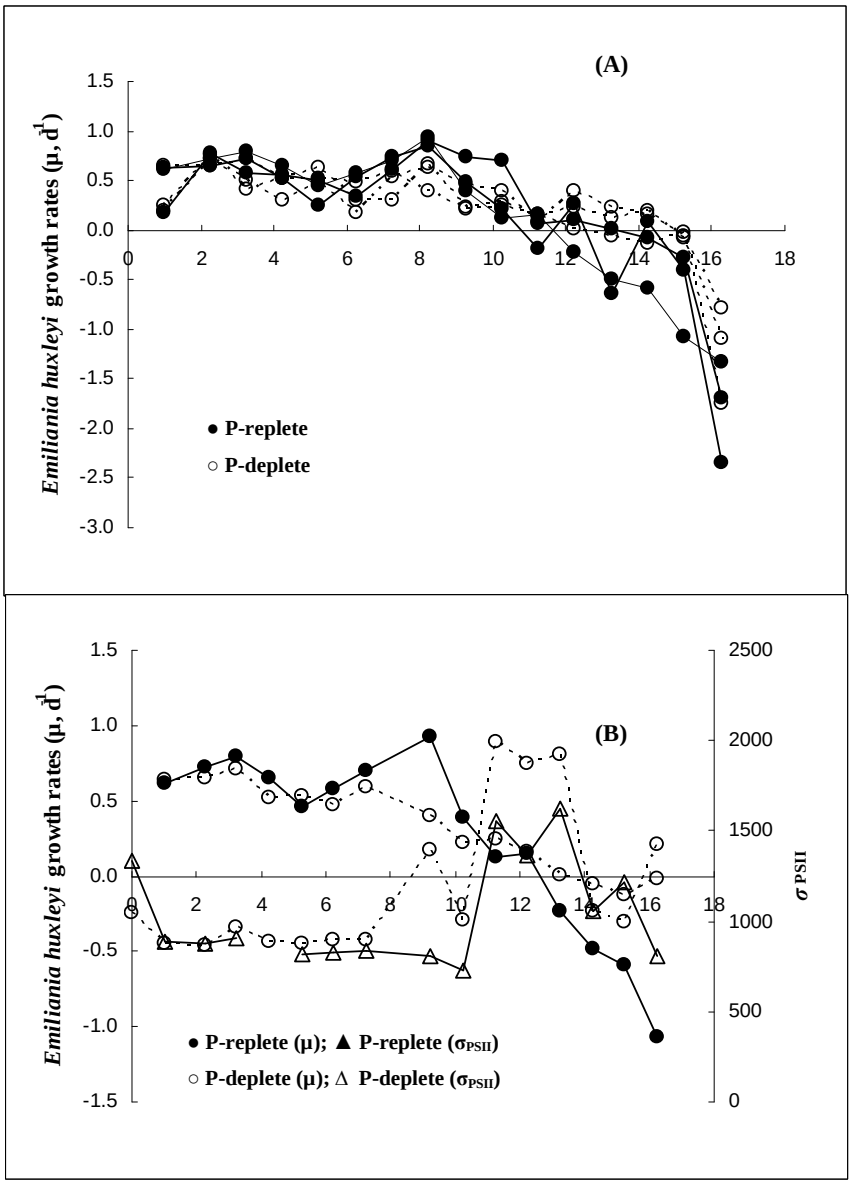

Figure 5. (A) *Emiliania huxleyi* net growth rates (μ, d^{-1}) in all mesocosms during the 17 day experiment in June 2008 (rates determined from 0600 data); (B) Plot showing the correlation between declining *Emiliania huxleyi* growth rates (μ, d^{-1}) and increasing σ_{PSII} during phase 3 of the experiment.

Phase 3. As the photosynthetic community entered phase 3 a marked uncoupling of photophysiology occurred in all mesocosms (Figs. 4a, b), interrupting the diel cycle in both F_v/F_m and σ_{PSII}. The change was first noticeable and most dramatic in the P-replete treatment (specifically mesocosm 3; Fig. 4b). F_v/F_m reached a low in this mesocosm of 0.48 during days 13-15 and displayed an overall decrease in F_v/F_m of 24 % compared to bloom phase 2. The decrease in F_v/F_m was most evident at midday and correlated with a crash of the *E. huxleyi* population shortly afterwards. For all mesocosms there was a corresponding trend between decreasing *E. huxleyi* net growth rate (Fig. 5a) and decreasing F_v/F_m (data not shown). On day 14, PSII cross-section was much reduced in range in all mesocosms compared to previous days and there was no difference between P-treatments; a slight recovery was visible by day 15. During the time of *E. huxleyi* dominance (days 8-16) an inverse relationship was determined between F_v/F_m and σ_{PSII} in all mesocosms, suggesting a linear correlation between the extent of photoinhibition and the increase in PSII antenna size (Fig. 6), (Ragni et al. 2008). The slopes of this relationship were greater in the P-deplete (y = 1881.5x -2563.7, r = 0.95) compared to P-replete populations (y = 1542.7x -1955.6, r = 0.86).

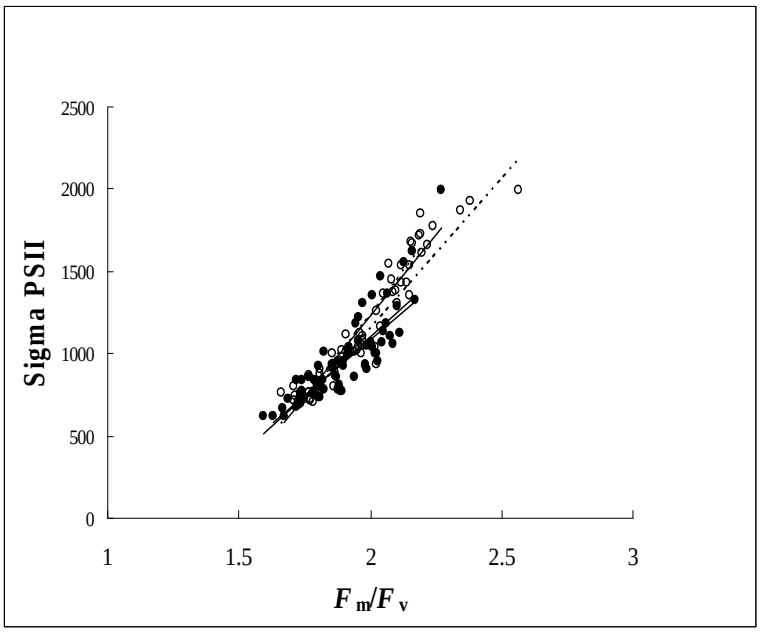

Figure 6. Relationship between σ_{PSII} and F_m/F_v; the plot contains all the data during days 9 to 16 when *E. huxleyi* populations dominated the photosynthetic community biomass. Superimposed linear fit (R = 0.95 and 0.86, for P-deplete and P-replete populations respectively).

5. Discussion

Addition of nutrients to 6 fjord seawater enclosures stimulated microbial succession and induced coccolithophore blooms, which in 3 enclosures subsequently collapsed with corresponding increases in large, viral-like particles (LVLPs). Flow cytometric analysis (FCM) allowed us to identify the major microbial groups that developed and to examine in fine-detail the temporal changes in microbial dynamics that occurred during this experiment. Based on the flow cytometric profiles and verified by molecular analysis (Pagarete et al., 2009; Sorensen et al., 2009) the coccolithophore and LVLP populations were identified as *Emiliania huxleyi* and EhV genotypes respectively. The initiation and dominance of the *E. huxleyi* populations probably resulted from a combination of several natural and enhanced factors. However the magnitude of these blooms and their fate appeared to be dependent on nutrient status, specifically P availability. Based on the rapid increase in EhV corresponding with the decline in *E. huxleyi* we assume that the viruses caused the major collapse of the *E. huxleyi* populations. This view is further supported by previous experimental data that corresponds with our findings (Bratbak, Egge, and Heldal, 1993; Jacquet et al., 2002; Martinez-Martinez et al., 2006; Schroeder et al., 2003). At the point of bloom collapse there was an uncoupling of photosynthetic community photosynthesis, which appeared to be linked to viral stress. Results from this experiment are comparable with those of previous studies that induced monospecific blooms of *E. huxleyi* under similar conditions (Bratbak, Egge, and Heldal, 1993; Castberg et al., 2001; Egge and Heimdal, 1994; Jacquet et al., 2002; Martinez-Martinez et al., 2006). However this study is the first demonstration of viral-driven interruption of PSII photochemistry in natural *E. huxleyi* populations.

Photophysiological response during bloom progression; impact of P-limitation

Very few studies have considered the impact of P-limitation in relation to PSII photochemistry (Beardall, Young, and Roberts, 2001; Graziano et al., 1996; Kromkamp and Peene, 1999; Lippemeier et al., 2003; Wood and Oliver, 1995); the role of N or Fe is more typical. However, examining the role of phosphorous limitation is imperative because P can constrain photosynthesis through both its requirement in nucleic acid and protein synthesis relating to the photosynthetic apparatus, and its function in the production and regeneration of substrates for carbon fixation (Falkowski and Raven, 2007). The impact of availability of phosphate on *E. huxleyi* growth was clearly visible in this study and had a huge impact on the

ultimate size and demise of the *E. huxleyi* blooms. The high abundance of *E. huxleyi* in the P-replete enclosures suggests that they were not limited by any other nutrients during the experimental period and as they were dividing more than once per day during bloom increase (max. μ = 0.93, d^{-1}) with corresponding F_v/F_m of 0.6, were growing at their maximum. In contrast, the development of the P-deplete *E. huxleyi* populations was slower and growth rates lower (max μ = 0.67) than for the P-replete cells. The peak in these populations also occurred 2 days later and delayed the onset of the collapse of these populations. This was also reflected by a lower maximum F_v/F_m in these populations of 0.5.

During the second and third phases of the bloom the *E. huxleyi* populations in all mesocosms displayed a series of sequential photophysiological responses as a result of the interactive effects of light, nutrient and viral stress. Individually these factors have previously been shown to affect the maximum quantum yield of PSII photochemistry (F_v/F_m) in *E. huxleyi* (e.g. Evans et al., 2006; Kruskopf and Flynn, 2006; Ragni et al., 2008); however examining the interaction between them is novel. In the natural environment phytoplankton typically display a diel photosynthetic pattern in response to the daily light cycle of elevated photosynthetic rate at the start of the day, followed by an afternoon depression (Marra, 1980; Sournia, 1974). Photoinhibition is typically greatest around midday when high light leads to a reduction in photosynthetic rate. In this study, once *E. huxleyi* populations were established, diel patterns emerged for various physiological parameters: cell abundance, Chl fluorescence, F_v/F_m and the relative functional absorption cross-section of PSII (σ_{PSII}). These initial changes in *E. huxleyi* physiology were clearly linked to the daily light cycle and have been observed previously for natural *E. huxleyi* populations (Jacquet et al., 2002). Over the 24 h period lowest F_v/F_m values were obtained at midday regardless of nutrient status in the mesocosms. However the impact on P-limited cells was greater in that the interaction between nutrient limitation and light stress resulted in lower F_v/F_m and elevated σ_{PSII} in these populations compared to P-replete cells.

Clear differences in Chl fluorescence were also observed between the P-limited and P-replete enclosures. As is typical with P-limited cells, mean Chl fluorescence was higher than in the P-replete cells as a reflection of their increased size (Muller, Antia, and LaRoche, 2008), but unlike Jacquet et al. (2002) the differences in chlorophyll fluorescence between the P treatments remained visible until the end of the experiment. These observations agree with previous studies showing that PSII photochemistry is responsive to nutrient limitation (e.g. Graziano et al., 1996; Moore et al., 2006; Sylvan et al., 2007) and that σ_{PSII} is typically higher in nutrient deplete cells relative to healthy cells (Kolber, Zehr, and Falkowski, 1988).

PSII photochemistry is depressed during nutrient stress because the ability of a cell to repair photodamage is decreased (Falkowski and Raven, 2007; Kolber, Zehr, and Falkowski, 1988; Marshall, Geider, and Flynn, 2000) and a larger fraction of the absorbed excitation energy is dissipated as heat and fluorescence (non-photochemical quenching, NPQ), rather than being used for photochemistry (Falkowski and Raven, 2007; Kolber, Zehr, and Falkowski, 1988). Unfortunately there are no measurements of NPQ or photorepair directly in this study; however the drop in F_v/F_m we observed at midday when photosynthetic populations were exposed to supraoptimal light levels, may reflect the net photoinhibition resulting from the imbalance between damage and repair of PSII (Adir et al., 2003; Critchley, 2000). Intriguingly, in previous culture studies *E. huxleyi* has been shown to be high-light tolerant and thus not prone to photoinhibition because of efficient photorepair mechanisms (Nanninga and Tyrrell, 1996; Ragni et al., 2008). These mechanisms are thought to be the strategy that makes this phytoplankton so competitive in high-light environments. However the stress tolerances of natural populations may be more limited in range compared to cultured species (Sylvan et al., 2007), and exacerbated when combined with changes in other environmental parameters such as temperature or nutrient shifts and biotic factors such as viral infection.

As well as playing a crucial role in the development of the *Emiliania huxleyi* blooms, nutrient availability also appeared to be important in determining their fate. Collapse of the *E. huxleyi* populations was rapid as found previously (Bratbak, Egge, and Heldal, 1993; Castberg et al., 2001; Jacquet et al., 2002) and the greatest decline was in P-replete enclosure 3 where *E. huxleyi* abundance fell by 95 % in 4 days from 1.7 x 10^6 cells mL^{-1} on day 12, to ~ 7 x 10^3 cells mL^{-1} by day 16. Though *Eh*V abundance increased gradually in the P-replete enclosures from day 8 onwards it was not until day 12 that numbers increased exponentially suggesting that as observed previously by Jacquet et al. (2002), cell infection and lysis was occurring before the host population decrease. This appeared to be reflected in the photophysiological response; just before bloom collapse there was an uncoupling of photosynthesis that was first visible in mesocosm 3 and resulted in a breakdown of the diel pattern of F_v/F_m and σ_{PSII}. This interruption of *E. huxleyi* photosynthesis is most likely a stress response to viral infection that was not clearly visible until a high enough proportion of the population were infected. During viral infection of plants a rapid initial reduction in exciton trapping efficiency by PSII results in initiation of photoprotective mechanisms to alleviate excess excitation (Balachandran et al., 1997). When host photochemistry begins to be affected, non-photochemical energy dissipation (through fluorescence or heat) is increased to

protect the photosynthetic apparatus from further damage by excess light, preventing photoinhibition (Horton, Ruban, and Walters, 1996). This is indicated by a decrease in dark-adapted F_v/F_m and usually associated with an increase in xanthophyll cycle pigments (Demmig-Adams and Adams, 1992). A similar impairment of photophysiology following virus infection has also been observed in phytoplankton (Hewson et al., 2001; Juneau et al., 2003; Seaton, Lee, and Rohozinski, 1995; Suttle and Chan, 1993; Suttle, Chan, and Cottrell, 1990; Waters and Chan, 1982), but the degree of PSII inhibition and the timing with regard to the lytic cycle is variable between species. In *Heterosigma akashiwo*, non-photochemical energy dissipation increased drastically as a result of viral infection, suggesting that heat dissipation was the main energy dissipation process in infected cells. Viral-induced non-photochemical quenching has also been observed in *Chlorella* (Seaton, Lee, and Rohozinski, 1995) and in filamentous cyanobacteria (Hewson, O'Neil, and Dennison, 2001). In *Micromonas pusilla* photoassimilation of CO_2 was significantly reduced soon after infection (~ 2 h) but cells continued to fix carbon photosynthetically until near the end of the lytic cycle (Waters and Chan, 1982). However in contrast, Lindell et al. (2005) showed that viral impact on PSII performance in *Prochlorococcus* was minimal, and suggested that this was because sustained photosynthesis was necessary in this system for maximum phage production. Likewise in *Synechococcus* (Mackenzie and Haselkorn, 1972) and *Phaeocystis pouchetti* (Bratbak, Jacobsen, and Heldal, 1998) photosynthesis continued unabated until the point of lysis. Thus, the impact of viral infection on phytoplankton photophysiology appears to be variable and may depend on whether virus replication is dependent on host energy derived from photosynthesis. What implications this has for primary production and biogeochemical cycling is still unclear, as very few studies have investigated direct viral-induced effects on photochemistry. In terms of phytoplankton biomass we know that viruses can significantly reduce carbon fixation and primary production (e.g. Hewson et al., 2001; Suttle, 1992; Suttle, Chan, and Cottrell, 1990). However previous studies have only considered this factor in terms of carbon loss via cell lysis. What has not been considered previously is the impact of viral infection on photosynthesis and carbon fixation during the lytic cycle before the conversion of POC to DOC. This could be an important carbon sink currently not considered in carbon cycling budgets. Furthermore, it has implications for higher trophic levels if cells that are grazed are less carbon-rich as a result of viral infection. Particularly if infected cells are preferentially and more rapidly grazed than non-infected prey (Evans and Wilson, 2008).

Although F_v/F_m and PSII antenna size was affected in all mesocosms during the third phase of the bloom, the impact was greatest in the P-stressed populations. This may be a

result of high viral nucleic acid requirements during infection (Bratbak, Egge, and Heldal, 1993) that may have exacerbated P constraints on *E. huxleyi* photosynthesis through P competition between host and virus. Previously it has been suggested that this high viral phosphate-to-protein requirement may make viral production sensitive to P-limitation (Bratbak, Egge, and Heldal, 1993; Wilson, Carr, and Mann, 1996). Indeed, the fully sequenced coccolithovirus isolate, EhV-86 was found to contain a gene encoding a putative phosphate permease. A second strain, EhV-163 (isolated from Bergen), has been shown to have a gene replacement at this locus, swapping the permease for an endonuclease. Thus, essentially scavenging P from within the cell instead of from external sources, like EhV-86. In the present experiment there was a distinct difference in EhV abundance from mesocosms that were P-limited compared to the P-replete enclosures, which was not only a result of the lower host density, viral production rates were also reduced with P-depletion. This correlates with previous studies where increased lytic cycle length and decreased burst size occurred when host cell growth was phosphate limited (Wilson, Carr, and Mann, 1996). If during EhV infection of *E. huxleyi* photosynthesis continues through to host lysis, as has been demonstrated for other alga:virus systems (e.g. Brown, Campbell, and Lawrence, 2007; Lindell et al., 2005), viral demands would decrease host P and energy reserves further. Support for this hypothesis comes from the increase in Chl fluorescence that was observed in P-replete mesocosms towards the end of phase 3 which may be an indication of the increased P-demand due to viral proliferation.

The trend of decreasing F_v/F_m and a corresponding increase in σ_{PSII} was visible over the three days before *E. huxleyi* population crash. This pattern has been observed previously during photoinhibition of *E. huxleyi*, but as a result of light stress (Ragni et al., 2008). During photoinhibition, inactivated/damaged PSII reaction centres still transfer energy to the functional PSII reaction centres, with the effect of increasing the effective size of the antenna serving the latter (Ragni et al., 2008). Increases in Sigma PSII cross-section are thought to enhance the susceptibility of PSII reaction centres to damage (Vassiliev et al., 1994) and this stress also results in a decline in F_v/F_m. Although a decrease in F_v/F_m has been observed previously during laboratory *E. huxleyi* viral infection experiments (Bidle et al., 2007; Evans et al., 2006), to our knowledge this is the first study to demonstrate this phenomenon in natural *E. huxleyi* populations. Unlike most natural phytoplankton communities which contain a mixture of photosynthetic algal groups, establishment of a dominant *E. huxleyi* bloom means that the majority of the fluorescence signal obtained from the analysis is derived from these populations. Evidence supporting this is the clear diel signal that was only observed

after day 8 when *E. huxleyi* starting growing rapidly. Interestingly, the uncoupling of a diel pattern first occurred in mesocosm 3 which contained populations that collapsed first, had the highest virus:host ratio, and maximum viral abundance. At the end of the experiment there appeared to be a recovery in F_v/F_m in all mesocosms but occurred initially in the P-replete treatments. This is probably associated with the rising abundance of other eukaryote populations after the collapse of the coccolithophores. Previously there has been contention regarding the use of F_v/F_m as an indicator of nutrient stress (e.g. Kruskopf and Flynn, 2006; Parkhill, Maillet, and Cullen, 2001; Young and Beardall, 2003), however this study shows the potential of this parameter and provides the first demonstration of its utility for assessing the physiological status of natural *E. huxleyi* populations.

6. Acknowledgements

This research was supported by a grant awarded to WHW from the Natural Environment Research Council (NERC) (ref NE/D001455/1). The research also forms part of the NERC Oceans 2025 programme which SAK and MJA are funded through at PML. WHW was supported by a National Science Foundation (NSF) grant ref. EF0723730. We would like to thank all the participants of the mesocosm, the staff of Espeland Station and the University of Bergen for making our stay both highly enjoyable and productive.

REFERENCES

Adir, N., Zer, H., Shochat, S., and Ohad, I. (2003). Photoinhibition - a historical perspective. *Photosynthesis Research* **76**(1-3)**,** 343-370.

Agusti, S., and Duarte, C. M. (2000). Strong seasonality in phytoplankton cell lysis in the NW Mediterranean littoral. *Limnology and Oceanography* **45**(4)**,** 940-947.

Agusti, S., and Sanchez, M. C. (2002). Cell viability in natural phytoplankton communities quantified by a membrane permeability probe. *Limnology and Oceanography* **47**(3)**,** 818-828.

Allen, M. J., Forster, T., Schroeder, D. C., Hall, M., Roy, D., Ghazal, P., and Wilson, W. H. (2006a). Locus-specific gene expression pattern suggests a unique propagation strategy for a giant algal virus. *J Virol* **80**(15)**,** 7699-7705.

Allen, M. J., Howard, J. A., Lilley, K. S., and Wilson, W. H. (2008). Proteomic analysis of the EhV-86 virion. *Proteome Sci* **6,** 11.

Allen, M. J., Martinez-Martinez, J., Schroeder, D. C., Somerfield, P. J., and Wilson, W. H. (2007). Use of microarrays to assess viral diversity: from genotype to phenotype. *Environmental Microbiology* **9**(4)**,** 971-982.

Allen, M. J., Schroeder, D. C., Donkin, A., Crawfurd, K. J., and Wilson, W. H. (2006b). Genome comparison of two Coccolithoviruses. *Virology J* **3,** 15.

Allen, M. J., Schroeder, D. C., Holden, M. T., and Wilson, W. H. (2006c). Evolutionary history of the Coccolithoviridae. *Mol Biol Evol* **23**(1)**,** 86-92.

Allen, M. J., Schroeder, D. C., and Wilson, W. H. (2006). Preliminary characterisation of repeat families in the genome of EhV-86, a giant algal virus that infects the marine microalga *Emiliania huxleyi*. *Arch Virol* **151**(3)**,** 525-535.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**(17)**,** 3389-402.

Andruleit, H., Rogalla, U., and Stager, S. (2004). From living communities to fossil assemblages: origin and fate of coccolithophores in the northern Arabian Sea. *Micropaleontology* **50,** 5-21.

Angly, F. E., Felts, B., Breitbart, M., Salamon, P., Edwards, R. A., Carlson, C., Chan, A. M., Haynes, M., Kelley, S., Liu, H., Mahaffy, J. M., Mueller, J. E., Nulton, J., Olson, R., Parsons, R., Rayhawk, S., Suttle, C. A., and Rohwer, F. (2006). The Marine Viromes of Four Oceanic Regions. *PLoS Biology* **4**(11)**,** e368.

Azam, F., Smith, D. C., Steward, G. F., and Hagström, Å. (1994). Bacteria-organic matter coupling and its significance for oceanic carbon cycling. *Microbial Ecology* **28**(2)**,** 167-179.

Balachandran, S., Hurry, V. M., Kelley, S. E., Osmond, C. B., Robinson, S. A., Rohozinski, J., Seaton, G. G. R., and Sims, D. A. (1997). Concepts of plant biotic stress. Some insights into the stress physiology of virus-infected plants, from the perspective of photosynthesis. *Physiologia Plantarum* **100**(2)**,** 203-213.

Baldauf, S. L. (2003). The deep roots of eukaryotes. *Science* **300**(5626)**,** 1703-1706.

Bandea, C. I. (1983). A new theory on the origin and the nature of viruses. *Journal of Theoretical Biology* **105**(4)**,** 591-602.

Beardall, J., Young, E., and Roberts, S. (2001). Approaches for determining phytoplankton nutrient limitation. *Aquatic Sciences* **63**(1)**,** 44-69.

Bell, P. J. (2001). Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? *J Mol Evol* **53**(3)**,** 251-6.

Bergh, O., Borsheim, K. Y., Bratbak, G., and Heldal, M. (1989). High abundance of viruses found in aquatic environments. *Nature* **340**(6233)**,** 467-468.

Bidle, K. D., Haramaty, L., Barcelos, E. R. J., and Falkowski, P. (2007). Viral activation and recruitment of metacaspases in the unicellular coccolithophore, Emiliania huxleyi. *Proc Natl Acad Sci U S A* **104**(14)**,** 6049-54.

Billard, C., and Inouye, I. (2004). What's new in coccolithophore biology. *In* "Coccolithophores: from molecular processes to global impact" (H. Thierstein, and J. R. Young, Eds.), pp. 1-29. Springer Verlag.

Boehme, J., Frischer, M. E., Jiang, S. C., Kellogg, C. A., Pichard, S., Rose, J. B., Steinway, C., and Paul, J. H. (1993). Viruses, bacterioplankton, and phytoplankton in the southeastern Gulf of Mexico - distribution and contribution to oceanic DNA pools. *Marine Ecology-Progress Series* **97**(1)**,** 1-10.

Borsheim, K. Y., Bratbak, G., and Heldal, M. (1990). Enumeration And Biomass Estimation Of Planktonic Bacteria And Viruses By Transmission Electron-Microscopy. *Applied and Environmental Microbiology* **56**(2)**,** 352-356.

Bown, P. R. (1987). Taxonomy, biostratigraphy, and evolution of late Triassic-early Jurassic calcareous nannofossils. *Special Papers in Palaeontology* **38,** 1-118.

Brand, L. E., and Guillard, R. R. L. (1981). The effects of continuous light and light-intensity on the reproduction rates of 22 species of marine phytoplankton. *Journal of Experimental Marine Biology and Ecology* **50**(2-3)**,** 119-132.

Bratbak, G., Egge, J. K., and Heldal, M. (1993). Viral mortality of the marine alga *Emiliania huxleyi* (Haptophyceae) and termination of algal blooms. *Marine Ecology-Progress Series* **93**(1-2)**,** 39-48.

Bratbak, G., Heldal, M., Norland, S., and Thingstad, T. F. (1990). Viruses as partners in Spring bloom microbial trophodynamics. *Applied and Environmental Microbiology* **56**(5)**,** 1400-1405.

Bratbak, G., Heldal, M., Thingstad, T. F., and Tuomi, P. (1996). Dynamics of virus abundance in coastal seawater. *Fems Microbiology Ecology* **19**(4)**,** 263-269.

Bratbak, G., Jacobsen, A., and Heldal, M. (1998). Viral lysis of Phaeocystis pouchetii and bacterial secondary production. *Aquatic Microbial Ecology* **16**(1)**,** 11-16.

Bratbak, G., Levasseur, M., Michaud, S., Cantin, G., Fernandez, E., Heimdal, B. R., and Heldal, M. (1995). Viral activity in relation to *Emiliania huxleyi* blooms: A mechanism of DMSP release? *Marine Ecology-Progress Series* **128**(1-3)**,** 133-142.

Bratbak, G., Wilson, W., and Heldal, M. (1996). Viral control of *Emiliania huxleyi* blooms? *Journal Of Marine Systems* **9**(1-2)**,** 75-81.

Breitbart, M., Felts, B., Kelley, S., Mahaffy, J. M., Nulton, J., Salamon, P., and Rohwer, F. (2004). Diversity and population structure of a near-shore marine-sediment viral community. *Proc Biol Sci* **271**(1539)**,** 565-74.

Breitbart, M., Middelboe, M., and Rohwer, F. (2008). Marine viruses: community dynamics, diversity and impact on microbial processes. *In* "Microbial Ecology of the Oceans" (D. L. Kirchman, Ed.), pp. 443-479. John Wiley & Sons, Inc.

Breitbart, M., Thompson, L. R., Suttle, C. A., and Sullivan, M. B. (2007). Exploring the Vast Diversity of Marine Viruses. *Oceanography* **20**(2)**,** 135-139.

Broecker, W. S., and Peng, T. (2005). The role of $CaCO_3$ compensation in the glacial to interglacial atmospheric CO_2 change. *Global Biogeochem Cycles* **1**(1)**,** 15-29.

Brown, C. M., Campbell, D. A., and Lawrence, J. E. (2007). Resource dynamics during infection of Micromonas pusilla by virus MpV-Sp1. *Environmental Microbiology* **9**(11)**,** 2720-2727.

Brown, C. W., and Yoder, J. A. (1994). Coccolithophorid Blooms In The Global Ocean. *Journal Of Geophysical Research-Oceans* **99**(C4)**,** 7467-7482.

Brussaard, C. P. (2004a). Viral control of phytoplankton populations--a review. *J Eukaryot Microbiol* **51**(2)**,** 125-38.

Brussaard, C. P. D. (2004b). Optimization of procedures for counting viruses by flow cytometry. *Applied and Environmental Microbiology* **70**(3)**,** 1506-1513.

Brussaard, C. P. D., Gast, G. J., vanDuyl, F. C., and Riegman, R. (1996a). Impact of phytoplankton bloom magnitude on a pelagic microbial food web. *Marine Ecology-Progress Series* **144**(1-3)**,** 211-221.

Brussaard, C. P. D., Kempers, R. S., Kop, A. J., Riegman, R., and Heldal, M. (1996b). Virus-like particles in a summer bloom of Emiliania huxleyi in the North Sea. *Aquatic Microbial Ecology* **10**(2)**,** 105-113.

Brussaard, C. P. D., Kuipers, B., and Veldhuis, M. J. W. (2005). A mesocosm study of Phaeocystis globosa population dynamics - 1. Regulatory role of viruses in bloom. *Harmful Algae* **4**(5)**,** 859-874.

Brussaard, C. P. D., Marie, D., and Bratbak, G. (2000). Flow cytometric detection of viruses. *Journal Of Virological Methods* **85**(1-2)**,** 175-182.

Brussaard, C. P. D., Noordeloos, A. A. M., Sandaa, R. A., Heldal, M., and Bratbak, G. (2004a). Discovery of a dsRNA virus infecting the marine photosynthetic protist Micromonas pusilla. *Virology* **319**(2)**,** 280-291.

Brussaard, C. P. D., Riegman, R., Noordeloos, A. A. M., Cadee, G. C., Witte, H., Kop, A. J., Nieuwland, G., Vanduyl, F. C., and Bak, R. P. M. (1995). Effects of grazing, sedimentation and phytoplankton cell-lysis on the structure of a coastal pelagic food-web. *Marine Ecology-Progress Series* **123**(1-3)**,** 259-271.

Brussaard, C. P. D., Short, S. M., Frederickson, C. M., and Suttle, C. A. (2004b). Isolation and phylogenetic analysis of novel viruses infecting the phytoplankton *Phaeocystis globosa* (Prymnesiophyceae). *Applied and Environmental Microbiology* **70**(6)**,** 3700-3705.

Brussaard, C. P. D., Wilhelm, S. W., Thingstad, F., Weinbauer, M. G., Bratbak, G., Heldal, M., Kimmance, S. A., Middelboe, M., Nagasaki, K., Paul, J. H., Schroeder, D. C., Suttle, C. A., Vaque, D., and Wommack, K. E. (2008). Global-scale processes with a nanoscale drive: the role of marine viruses. *Isme Journal* **2**(6)**,** 575-578.

Brussow, H., and Hendrix, R. W. (2002). Phage genomics: small is beautiful. *Cell* **108**(1)**,** 13-6.

Burkill, P. H., Archer, S. D., Robinson, C., Nightingale, P. D., Groom, S. B., Tarran, G. A., and Zubkov, M. V. (2002). Dimethyl sulphide biogeochemistry within a coccolithophore bloom (DISCO): an overview. *Deep-Sea Research Part Ii-Topical Studies In Oceanography* **49**(15)**,** 2863-2885.

Castberg, T., Larsen, A., Sandaa, R. A., Brussaard, C. P. D., Egge, J. K., Heldal, M., Thyrhaug, R., van Hannen, E. J., and Bratbak, G. (2001). Microbial population dynamics and diversity during a bloom of the marine coccolithophorid *Emiliania huxleyi* (Haptophyta). *Marine Ecology-Progress Series* **221,** 39-46.

Castberg, T., Thyrhaug, R., Larsen, A., Sandaa, R. A., Heldal, M., Van Etten, J. L., and Bratbak, G. (2002). Isolation and characterization of a virus that infects Emiliania huxleyi (Haptophyta). *Journal Of Phycology* **38**(4)**,** 767-774.

Charlson, R. J., Lovelock, J. E., Andreae, M. O., and Warren, S. G. (1987). Oceanic Phytoplankton, Atmospheric Sulfur, Cloud Albedo and Climate. *Nature* **326**(6114)**,** 655-661.

Chen, F., and Suttle, C. A. (1995). Amplification of DNA-Polymerase Gene Fragments from Viruses Infecting Microalgae. *Applied and Environmental Microbiology* **61**(4)**,** 1274-1278.

Chen, F., Suttle, C. A., and Short, S. M. (1996). Genetic diversity in marine algal virus communities as revealed by sequence analysis of DNA polymerase genes. *Applied and Environmental Microbiology* **62**(8)**,** 2869-2874.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., and Thompson, J. D. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31**(13)**,** 3497-500.

Clasen, J. L., and Suttle, C. A. (2009). Identification of freshwater Phycodnaviridae and their potential phytoplankton hosts, using DNA pol sequence fragments and a genetic-distance analysis. *Appl Environ Microbiol* **75**(4)**,** 991-7.

Claverie, J. M. (2006). Viruses take center stage in cellular evolution. *Genome Biol* **7**(6)**,** 110.

Claverie, J. M., Abergel, C., and Ogata, H. (2009). Mimivirus. *Curr Top Microbiol Immunol* **328,** 89-121.

Claverie, J. M., Ogata, H., Audic, S., Abergel, C., Suhre, K., and Fournier, P. E. (2006). Mimivirus and the emerging concept of "giant" virus. *Virus Res* **117**(1)**,** 133-44.

Clokie, M. R. J., and Mann, N. H. (2006). Marine cyanophages and light Marine cyanophages and light. *Environmental Microbiology* **8,** 2074-2082.

Clokie, M. R. J., Shan, J. Y., Bailey, S., Jia, Y., Krisch, H. M., West, S., and Mann, N. H. (2006). Transcription of a 'photosynthetic' T4-type phage during infection of a marine cyanobacterium. *Environmental Microbiology* **8**(5)**,** 827-835.

Cottrell, M. T., and Suttle, C. A. (1991). Wide-Spread Occurrence and Clonal Variation in Viruses Which Cause Lysis of a Cosmopolitan, Eukaryotic Marine Phytoplankter, Micromonas-Pusilla. *Marine Ecology-Progress Series* **78**(1)**,** 1-9.

Critchley, C. (2000). Photoinhibition. *In* "Photosynthesis: A comprehensive Treatise" (S. Raghavendra, Ed.), pp. 264–272. Cambridge University Press, Cambridge, UK.

de Vargas, C., Aubry, M.-P., Probert, I., Young, J., Falkowski, P., and Knoll, A. (2007). Origin and evolution of Coccolithophores: from coastal hunters to oceanic farmers. *In* "Evolution of primary producers in the sea" (P. Falkowski, and A. Knoll, Eds.), pp. 251-285. Elsevier, Burlington.

de Vargas, C., and Probert, I. (2004). New keys to the past: Current and future DNA studies in coccolithophores. *Micropaleontology* **50,** 45-54.

Delaroque, N., Maier, I., Knippers, R., and Muller, D. G. (1999). Persistent virus integration into the genome of its algal host, *Ectocarpus siliculosus* (Phaeophyceae). *J Gen Virol* **80,** 1367-70.

Delaroque, N., Muller, D. G., Bothe, G., Pohl, T., Knippers, R., and Boland, W. (2001). The complete DNA sequence of the Ectocarpus siliculosus virus EsV-1 genome. *Virology* **287**(1)**,** 112-132.

Delille, B., Harlay, J., Zondervan, I., Jacquet, S., Chou, L., Wollast, R., Bellerby, R., Frankignoulle, M., Vieira Borges, A., Riebesell, U., and Gattuso, J.-P. (2005). Response of primary production and calcification to changes of pCO2 during experimental blooms of the coccolithophorid Emiliania huxleyi. *Global Biogeochem Cy* **19**(doi:10.1029/2004GB002318)**,** GB2023.

Demmig-Adams, B., and Adams, W. W. (1992). Photoprotection and Other Responses of Plants to High Light Stress. *Annual Review of Plant Physiology and Plant Molecular Biology* **43,** 599-626.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., Claverie, J. M., and Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **36**(Web Server issue)**,** W465-9.

Derelle, E., Ferraz, C., Escande, M.-L., Eychenie, S., Cooke, R., Piganeau, G., Desdevises, Y., Bellec, L., Moreau, H., and Grimsley, N. (2008). Life-cycle and genome of OtV5, a large DNA virus of the pelagic marine unicellular green alga *Ostreococcus tauri*. *PLoS ONE* **3**(5)**,** e2250, 1-13.

Dimmock, N., Easton, A., and Leppard, K. (2007). "Introduction to Modern Virology." 5th edition ed. Blackwell Publishing.

Dinsdale, E. A., Edwards, R. A., Hall, D., Angly, F., Breitbart, M., Brulc, J. M., Furlan, M., Desnues, C., Haynes, M., Li, L. L., McDaniel, L., Moran, M. A., Nelson, K. E., Nilsson, C., Olson, R., Paul, J., Brito, B. R., Ruan, Y. J., Swan, B. K., Stevens, R., Valentine, D. L., Thurber, R. V., Wegley, L., White, B. A., and Rohwer, F. (2008). Functional metagenomic profiling of nine biomes (vol 452, pg 629, 2008). *Nature* **455**(7214)**,** 830-830.

Dunigan, D. D., Fitzgerald, L. A., and Van Etten, J. L. (2006). Phycodnaviruses: A peek at genetic diversity. *Virus Research* **117**(1)**,** 119-132.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**(1)**,** 113.

Egge, J. K., and Aksnes, D. L. (1992). Silicate as Regulating Nutrient in Phytoplankton Competition. *Marine Ecology-Progress Series* **83**(2-3)**,** 281-289.

Egge, J. K., and Heimdal, B. R. (1994). Blooms of Phytoplankton Including *Emiliania huxleyi* (Haptophyta) - Effects of Nutrient Supply in Different N-P Ratios. *Sarsia* **79**(4)**,** 333-348.

Elderfield, H. (2002). Climate change: Carbonate mysteries. *Science* **296**(5573)**,** 1618-+.

Essbauer, S., and Ahne, W. (2001). Viruses of lower vertebrates. *Journal of Veterinary Medicine Series B-Infectious Diseases and Veterinary Public Health* **48**(6)**,** 403-475.

Evans, C., Malin, G., Mills, G. P., and Wilson, W. H. (2006). Viral infection of Emiliania huxleyi (Prymnesiophyceae) leads to elevated production of reactive oxygen species. *Journal of Phycology* **42**(5)**,** 1040-1047.

Evans, C., and Wilson, W. H. (2008). Preferential grazing of *Oxyrrhis marina* on virus-infected *Emiliania huxleyi*. *Limnology and Oceanography* **53**(5)**,** 2035-U12.

Falkowski, P., and Raven, A. (2007). "Aquatic photosynthesis." Princeton University Press.

Falkowski, P., Scholes, R. J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Hogberg, P., Linder, S., Mackenzie, F. T., Moore, B., Pedersen, T., Rosenthal, Y., Seitzinger, S., Smetacek, V.,

and Steffen, W. (2000). The global carbon cycle: A test of our knowledge of earth as a system. *Science* **290**(5490)**,** 291-296.

Falkowski, P. G., Barber, R. T., and Smetacek, V. (1998). Biogeochemical controls and feedbacks on ocean primary production. *Science* **281**(5374)**,** 200-206.

Felsenstein, J. (2004). PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**(5374)**,** 237-40.

Filee, J., Pouget, N., and Chandler, M. (2008). Phylogenetic evidence for extensive lateral acquisition of cellular genes by Nucleocytoplasmic large DNA viruses. *Bmc Evolutionary Biology* **8**.

Filee, J., Siguier, P., and Chandler, M. (2007). I am what I eat and I eat what I am: acquisition of bacterial genes by giant viruses. *Trends Genet* **23**(1)**,** 10-5.

Forest, T., Barnard, S., and Baines, J. D. (2005). Active intranuclear movement of herpesvirus capsids. *Nature Cell Biology* **7**(4)**,** 429-U54.

Forster, T., Roy, D., and Ghazal, P. (2003). Experiments using microarray technology: limitations and standard operating procedures. *Journal of Endocrinology* **178**(2)**,** 195-204.

Forterre, P. (2003). The great virus comeback - from an evolutionary perspective. *Research in Microbiology* **154**(4)**,** 223-225.

Forterre, P. (2006a). The origin of viruses and their possible roles in major evolutionary transitions. *Virus Research* **117**(1)**,** 5-16.

Forterre, P. (2006b). The role of virus in the origin of DNA genome and early cellular evolution. *Origins Of Life And Evolution Of The Biosphere* **36**(3)**,** 219-221.

Forterre, P. (2006c). Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. *Proc Natl Acad Sci U S A* **103**(10)**,** 3669-74.

Forterre, P., and Gadelle, D. (2009). Phylogenomics of DNA topoisomerases: their origin and putative roles in the emergence of modern organisms. *Nucleic Acids Res* **37**(3)**,** 679-92.

Forterre, P., Gribaldo, S., and Brochier, W. (2005). Luca : the last universal common ancestor. *M S-Medecine Sciences* **21**(10)**,** 860-865.

Frada, M. (2009). "The haplo-diplontic life cycles of the haptophytes, with emphasis on the eco-physiology of *Emiliania huxleyi*." Université Pierre et Marie Curie (Paris VI), Paris.

Frada, M., Probert, I., Allen, M. J., Wilson, W. H., and de Vargas, C. (2008a). The "Cheshire Cat" escape strategy of the coccolithophore *Emiliania huxleyi* in response to viral infection. *Proceedings of the National Academy of Sciences of the United States of America* **105**(41)**,** 15944-15949.

Frohns, F., Kasmann, A., Kramer, D., Schafer, B., Mehmel, M., Kang, M., Van Etten, J. L., Gazzarrini, S., Moroni, A., and Thiel, G. (2006). Potassium ion channels of Chlorella viruses cause rapid depolarization of host cells during infection. *J Virol* **80**(5)**,** 2437-44.

Fuhrman, J. A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature* **399**(6736)**,** 541-548.

Fuhrman, J. A., and Noble, R. T. (1995). Viruses and protists cause similar bacterial mortality in coastal seawater. *Limnology and Oceanography* **40**(7)**,** 1236-1242.

Fuhrman, J. A., and Suttle, C. (1993). Viruses in marine planktonic systems. *Oceanography* **6,** 51-63.

Futerman, A. H., and Hannun, Y. A. (2004). The complex life of simple sphingolipids. *Embo Reports* **5**(8)**,** 777-782.

Geider, R. J., and Laroche, J. (1994). The Role of Iron in Phytoplankton Photosynthesis, and the Potential for Iron-Limitation of Primary Productivity in the Sea. *Photosynthesis Research* **39**(3)**,** 275-301.

Gobler, C. J., Hutchins, D. A., Fisher, N. S., Cosper, E. M., and Sanudo-Wilhelmy, S. A. (1997). Release and bioavailability of C, N, P, Se, and Fe following viral lysis of a marine chrysophyte. *Limnology and Oceanography* **42**(7)**,** 1492-1504.

Gorbalenya, A. E., Koonin, E. V., and Wolf, Y. I. (1990). A new superfamily of putative NTP-binding domans encoded by genomes of small DNA and RNA viruses. *Febs Letters* **262**(1)**,** 145-148.

Graziano, L. M., Geider, R. J., Li, W. K. W., and Olaizola, M. (1996). Nitrogen limitation of North Atlantic phytoplankton: Analysis of physiological condition in nutrient enrichment experiments. *Aquatic Microbial Ecology* **11**(1)**,** 53-64.

Green, J. C., Course, P. A., and Tarran, G. A. (1996). The life-cycle of Emiliania huxleyi: A brief review and a study of relative ploidy levels analysed by flow cytometry. *Journal of Marine Systems* **9**(1-2)**,** 33-44.

Greenblatt, R. J., Quackenbush, S. L., Casey, R. N., Rovnak, J., Balazs, G. H., Work, T. M., Casey, J. W., and Sutton, C. A. (2005). Genomic variation of the fibropapilloma-associated marine turtle herpesvirus across seven geographic areas and three host species. *Journal of Virology* **79**(2)**,** 1125-1132.

Guenther, G. G., Peralta, E. R., Rosales, K. R., Wong, S. Y., Siskind, L. J., and Edinger, A. L. (2008). Ceramide starves cells to death by downregulating nutrient transporter proteins. *Proc Natl Acad Sci U S A.*

Guillard, R. R. L. (1975). Culture of phytoplankton for feeding marine invertebrates. *In* "Culture of marine invertebrate animals." (W. L. Smith, and M. H. Chanley, Eds.), pp. 29-60. Plenum Press, New York, N.Y.

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**(5)**,** 696-704.

Han, G., Gable, K., Yan, L., Allen, M. J., Wilson, W. H., Moitra, P., Harmon, J. M., and Dunn, T. M. (2006). Expression of a novel marine viral single-chain serine palmitoyltransferase and construction of yeast and mammalian single-chain chimera. *J Biol Chem* **281**(52)**,** 39935-42.

Hannun, Y. A. (1996). Functions of ceramide in coordinating cellular responses to stress. *Science* **274**(5294)**,** 1855-1859.

Hannun, Y. A., Luberto, C., and Argraves, K. M. (2001). Enzymes of sphingolipid metabolism: From modular to integrative signaling. *Biochemistry* **40**(16)**,** 4893-4903.

Hannun, Y. A., and Obeid, L. M. (1995). Ceramide - an Intracellular Signal for Apoptosis. *Trends in Biochemical Sciences* **20**(2)**,** 73-77.

Hannun, Y. A., and Obeid, L. M. (2002). The ceramide-centric universe of lipid-mediated cell regulation: Stress encounters of the lipid kind. *Journal of Biological Chemistry* **277**(29)**,** 25847-25850.

Hannun, Y. A., and Obeid, L. M. (2008). Principles of bioactive lipid signalling: lessons from sphingolipids. *Nature Reviews Molecular Cell Biology* **9**(2)**,** 139-150.

Hara, S., Koike, I., Terauchi, K., Kamiya, H., and Tanoue, E. (1996). Abundance of viruses in deep oceanic waters. *Marine Ecology Progress Series* **145,** 269-277.

Hay, S., and Kannourakis, G. (2002). A time to kill: viral manipulation of the cell death program. *Journal of General Virology* **83,** 1547-1564.

Hay, W. W. (2004). Carbonate fluxes and calcareous nannoplankton. *In* "Coccolithophores: from molecular processes to global impact" (H. Thierstein, and J. R. Young, Eds.), pp. 509–527. Springer.

Hendrix, R. W. (1999). Evolution: The long evolutionary reach of viruses. *Current Biology* **9**(24)**,** R914-R917.

Hendrix, R. W., Lawrence, J. G., Hatfull, G. F., and Casjens, S. (2000). The origins and ongoing evolution of viruses. *Trends in Microbiology* **8**(11)**,** 504-508.

Hennes, K. P., Suttle, C. A., and Chan, A. M. (1995). Fluorescently labeled virus probes show that natural virus populations can control the structure of marine microbial communities. *Applied and Environmental Microbiology* **61**(10)**,** 3623-3627.

Hewson, I., O'Neil, J. M., and Dennison, W. C. (2001). Virus-like particles associated with *Lyngbya majuscula* (Cyanophyta; Oscillatoriacea) bloom decline in Moreton Bay, Australia. *Aquatic Microbial Ecology* **25**(3)**,** 207-213.

Hewson, I., O'Neil, J. M., Heil, C. A., Bratbak, G., and Dennison, W. C. (2001). Effects of concentrated viral communities on photosynthesis and community composition of co-occurring benthic microalgae and phytoplankton. *Aquatic Microbial Ecology* **25**(1)**,** 1-10.

Holligan, P. M., Fernandez, E., Aiken, J., Balch, W. M., Boyd, P., Burkill, P. H., Finch, M., Groom, S. B., Malin, G., Muller, K., Purdie, D. A., Robinson, C., Trees, C. C., Turner, S. M., and Vanderwal, P. (1993). A biogeochemical study of the coccolithophore, *Emiliania huxleyi*, in the North-Atlantic. *Global Biogeochemical Cycles* **7**(4)**,** 879-900.

Horton, P., Ruban, A. V., and Walters, R. G. (1996). Regulation of light harvesting in green plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **47,** 655-684.

Houdan, A., Billard, C., Marie, D., Not, F., aacute, ez, A. G., Young, J. R., and Probert, I. (2003). Holococcolithophore-heterococcolithophore (Haptophyta) life cycles: flow cytometric analysis of relative ploidy levels. *Systematics and Biodiversity* **1**(04)**,** 453-465.

Hughes, A. L., and Friedman, R. (2005). Poxvirus genome evolution by gene gain and loss. *Mol Phylogenet Evol* **35**(1)**,** 186-95.

Hutchinson, G. E. (1961). The paradox of plankton. *American Naturalist* **95**(882)**,** 137-145.

Huxley, T. H. (1858). Appendix A. *In* "Deep Sea Soundings in the North Atlantic Ocean between Irelandd and Newfoundland" (J. Dayman, Ed.), pp. 63-68. H. M. Stationary Office, London.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., and Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**(2)**,** 249-264.

Iyer, L. M., Aravind, L., and Koonin, E. V. (2001). Common origin of four diverse families of large eukaryotic DNA viruses. *Journal of Virology* **75**(23)**,** 11720-11734.

Iyer, L. M., Balaji, S., Koonin, E. V., and Aravind, L. (2006). Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Research* **117**(1)**,** 156-184.

Jacobsen, A., Bratbak, G., and Heldal, M. (1996). Isolation and characterization of a virus infecting Phaeocystis pouchetii (Prymnesiophyceae). *Journal Of Phycology* **32**(6)**,** 923-927.

Jacquet, S., Heldal, M., Iglesias-Rodriguez, D., Larsen, A., Wilson, W., and Bratbak, G. (2002). Flow cytometric analysis of an Emiliana huxleyi bloom terminated by viral infection. *Aquatic Microbial Ecology* **27**(2)**,** 111-124.

Jiang, J. C., Kirchman, P. A., Zagulski, M., Hunt, J., and Jazwinski, S. M. (1998). Homologs of the yeast longevity gene LAG1 in Caenorhabditis elegans and human. *Genome Res* **8**(12)**,** 1259-72.

Jiang, S. C., and Paul, J. H. (1996). Occurrence of lysogenic bacteria in marine microbial communities as determined by prophage induction. *Marine Ecology-Progress Series* **142**(1-3)**,** 27-38.

Jiang, S. C., and Paul, J. H. (1998a). Gene transfer by transduction in the marine environment. *Applied and Environmental Microbiology* **64**(8)**,** 2780-2787.

Jiang, S. C., and Paul, J. H. (1998b). Significance of lysogeny in the marine environment: Studies with isolates and a model of lysogenic phage production. *Microbial Ecology* **35**(3)**,** 235-243.

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1994). A mutation data matrix for transmembrane proteins. *FEBS Lett* **339**(3)**,** 269-75.

Juneau, P., Lawrence, J. E., Suttle, C. A., and Harrison, P. J. (2003). Effects of viral infection on photosynthetic processes in the bloom-forming alga *Heterosigma akashiwo*. *Aquatic Microbial Ecology* **31**(1)**,** 9-17.

Kimmance, S., Allen, M. J., Martinez-Martinez, J., Pagarete, A., and Wilson, W. H. (in press). Uncoupling of *Emiliania huxleyi* photosynthesis: virus infection versus nutrient stress. *Environmental Microbiology*.

Kolber, Z., and Falkowski, P. G. (1993). Use of Active Fluorescence to Estimate Phytoplankton Photosynthesis in-Situ. *Limnology and Oceanography* **38**(8)**,** 1646-1665.

Kolber, Z., Zehr, J., and Falkowski, P. (1988). Effects of Growth Irradiance and Nitrogen Limitation on Photosynthetic Energy-Conversion in Photosystem-Ii. *Plant Physiology* **88**(3)**,** 923-929.

Koonin, E., Fedorova, N., Jackson, J., Jacobs, A., Krylov, D., Makarova, K., Mazumder, R., Mekhedov, S., Nikolskaya, A., Rao, B., Rogozin, I., Smirnov, S., Sorokin, A., Sverdlov, A., Vasudevan, S., Wolf, Y., Yin, J., and Natale, D. (2004). A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biology* **5**(2)**,** R7.

Koonin, E., Senkevich, T., and Dolja, V. (2006). The ancient virus world and evolution of cells. *Biology Direct* **1**(1)**,** 29.

Koonin, E. V. (2005). Virology: Gulliver among the Lilliputians. *Curr Biol* **15**(5)**,** R167-9.

Koonin, E. V. (2009). Darwinian evolution in the light of genomics. *Nucleic Acids Research* **37**(4)**,** 1011-1034.

Kromkamp, J., and Peene, J. (1999). Estimation of phytoplankton photosynthesis and nutrient limitation in the Eastern Scheldt estuary using variable fluorescence. *Aquatic Ecology* **33**(1)**,** 101-104.

Kruskopf, M., and Flynn, K. J. (2006). Chlorophyll content and fluorescence responses cannot be used to gauge reliably phytoplankton biomass, nutrient status or growth rate. *New Phytologist* **169**(3)**,** 525-536.

Larsen, A., Castberg, T., Sandaa, R. A., Brussaard, C. P. D., Egge, J., Heldal, M., Paulino, A., Thyrhaug, R., van Hannen, E. J., and Bratbak, G. (2001). Population dynamics and diversity of phytoplankton, bacteria and viruses in a seawater enclosure. *Marine Ecology-Progress Series* **221,** 47-57.

Larsen, A., and Edvardsen, B. (1998). Relative ploidy levels in Prymnesium parvum and P-patelliferum (Haptophyta) analyzed by flow cytometry. *Phycologia* **37**(6)**,** 412-424.

Larsen, A., Flaten, G. A. F., Sandaa, R. A., Castberg, T., Thyrhaug, R., Erga, S. R., Jacquet, S., and Bratbak, G. (2004). Spring phytoplankton bloom dynamics in Norwegian coastal waters: Microbial community succession and diversity. *Limnology And Oceanography* **49**(1)**,** 180-190.

Larsen, J. B., Larsen, A., Bratbak, G., and 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**(10)**,** 3048-3057.

Lawrence, J. E., Chan, A. M., and Suttle, C. A. (2001). A novel virus (HaNIV) causes lysis of the toxic bloom-forming alga Heterosigma akashiwo (Raphidophyceae). *Journal Of Phycology* **37**(2)**,** 216-222.

Laws, E. A., Falkowski, P. G., Smith, W. O., Ducklow, H., and McCarthy, J. J. (2000). Temperature effects on export production in the open ocean. *Global Biogeochemical Cycles* **14**(4)**,** 1231-1246.

Leadbeater, B. S. C. (1994). Cell coverings. *In* "The Haptophyte Algae" (J. C. Green, and B. S. C. Leadbeater, Eds.), pp. 23-46. Systematics Association Special Vol 51, Clarendon Press, Oxford.

Lindell, D., Jaffe, J. D., Coleman, M. L., Futschik, M. E., Axmann, I. M., Rector, T., Kettler, G., Sullivan, M. B., Steen, R., Hess, W. R., Church, G. M., and Chisholm, S. W. (2007). Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449**(7158)**,** 83-6.

Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M., and Chisholm, S. W. (2005). Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**(7064)**,** 86-89.

Lindell, D., Sullivan, M. B., Johnson, Z. I., Tolonen, A. C., Rohwer, F., and Chisholm, S. W. (2004). Transfer of photosynthesis genes to and from Prochlorococcus viruses. *Proc Natl Acad Sci U S A* **101**(30)**,** 11013-8.

Lindqvist, Y., Huang, W., Schneider, G., and Shanklin, J. (1996). Crystal structure of delta9 stearoyl-acyl carrier protein desaturase from castor seed and its relationship to other di-iron proteins. *Embo J* **15**(16)**,** 4081-92.

Lippemeier, S., Frampton, D. M. F., Blackburn, S. I., Geier, S. C., and Negri, A. P. (2003). Influence of phosphorus limitation on toxicity and photosynthesis of Alexandrium minutum (Dinophyceae) monitored by in-line detection of variable chlorophyll fluorescence. *Journal of Phycology* **39**(2)**,** 320-331.

Mackenzie, J. J., and Haselkorn, R. (1972). Photosynthesis and the development of blue-green algal virus SM-1. *Virology* **49**(2)**,** 517-21.

Mackinder, L. C. M., Worthy, C. A., Biggi, G., Hall, M., Ryan, K. P., Varsani, A., Harper, G., Wilson, W. H., Brownlee, C., and Schroeder, D. C. (2009). A unicellular algal virus, *Emiliania huxleyi* virus 86, exploits an animal-like infection strategy. *J Gen Virol* **90,** 2306–2316.

Maier, I., Müller, D. G., and Katsaros, C. (2002). Entry of the DNA virus, *Ectocarpus fasciculatus* virus type 1 (Phycodnaviridae), into host cell cytosol and nucleus. *Phycological Research* **50**(3)**,** 227-231.

Makeyev, E. V., and Grimes, J. M. (2004). RNA-dependent RNA polymerases of dsRNA bacteriophages. *Virus Research* **101**(1)**,** 45-55.

Malin, G., and Steinke, M. (2004). Dimethyl sulfide production: what is the contribution of the coccolithophores? *In* "Coccolithophores - From Molecular Processes to Global Impact" (H. R. Thierstein, and J. R. Young, Eds.), pp. 127-164. Springer, Berlin, Germany.

Mann, N. H., Clokie, M. R. J., Millard, A., Cook, A., Wilson, W. H., Wheatley, P. J., Letarov, A., and Krisch, H. M. (2005). The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine Synechococcus strains. *Journal of Bacteriology* **187**(9)**,** 3188-3200.

Maranger, R., Bird, D. F., and Juniper, S. K. (1994). Viral and bacterial dynamics in Artic sea-ice during the spring algal bloom near Resolute, NWT, Canada. *Marine Ecology-Progress Series* **111**(1-2)**,** 121-127.

Marie, D., Brussaard, C. P. D., Thyrhaug, R., Bratbak, G., and Vaulot, D. (1999). Enumeration of marine viruses in culture and natural samples by flow cytometry. *Applied and Environmental Microbiology* **65**(1)**,** 45-52.

Marra, J. (1980). Time Course of Light-Intensity Adaptation in a Marine Diatom. *Marine Biology Letters* **1**(3)**,** 175-183.

Marshall, H. L., Geider, R. J., and Flynn, K. J. (2000). A mechanistic model of photoinhibition. *New Phytologist* **145**(2)**,** 347-359.

Martinez-Martinez, J., Norland, S., Thingstad, T. F., Schroeder, D. C., Bratbak, G., Wilsow, W. H., and Larsen, A. (2006). Variability in microbial population dynamics between similarly perturbed mesocosms. *Journal Of Plankton Research* **28**(8)**,** 783-791.

Martinez, J. M., Schroeder, D. C., Larsen, A., Bratbak, G., and Wilson, W. H. (2007). Molecular dynamics of *Emiliania huxleyi* and co-occurring viruses during two separate mesocosm studies. *Appl Environ Microbiol* **73**(2)**,** 554-62.

McDaniel, L. D., delaRosa, M., and Paul, J. H. (2006). Temperate and lytic cyanophages from the Gulf of Mexico. *Journal of the Marine Biological Association of the United Kingdom* **86**(3)**,** 517-527.

McLean, J. E., Ruck, A., Shirazian, A., Pooyaei-Mehr, F., and Zakeri, Z. F. (2008). Viral manipulation of cell death. *Current Pharmaceutical Design* **14**(3)**,** 198-220.

Meints, R. H., Lee, K., and Van Etten, J. L. (1986). Assembly site of the virus PBCV-1 in a Chlorella-like green alga: Ultrastructural studies. *Virology* **154**(1)**,** 240-245.

Melis, A. (1999). Photosystem-II damage and repair cycle in chloroplasts: what modulates the rate of photodamage in vivo? *Trends in Plant Science* **4**(4)**,** 130-135.

Merrill, A. H. (2002). De novo sphingolipid biosynthesis: A necessary, but dangerous, pathway. *Journal of Biological Chemistry* **277**(29)**,** 25843-25846.

Middelboe, M., Riemann, L., Steward, G. F., Hansen, V., and Nybroe, O. (2003). Virus-induced transfer of organic carbon between marine bacteria in a model community. *Aquatic Microbial Ecology* **33**(1)**,** 1-10.

Millard, A., Clokie, M. R. J., Shub, D. A., and Mann, N. H. (2004). Genetic organization of the psbAD region in phages infecting marine *Synechococcus* strains. *Proceedings of the National Academy of Sciences of the United States of America* **101**(30)**,** 11007-11012.

Milliman, J. D. (1993). Production and accumulation of calcium-carbonate in the ocean - budget of a nonsteady state. *Global Biogeochemical Cycles* **7**(4)**,** 927-957.

Mindell, D. P., and Villarreal, L. P. (2003). Don't forget about viruses. *Science* **302**(5651)**,** 1677-1677.

Mitchell, A. G., and Martin, C. E. (1997). Fah1p, a Saccharomyces cerevisiae cytochrome b5 fusion protein, and its Arabidopsis thaliana homolog that lacks the cytochrome b5 domain both function in the alpha-hydroxylation of sphingolipid-associated very long chain fatty acids. *J Biol Chem* **272**(45)**,** 28281-8.

Monier, A., Claverie, J. M., and Ogata, H. (2007). Horizontal gene transfer and nucleotide compositional anomaly in large DNA viruses. *Bmc Genomics* **8**.

Monier, A., Claverie, J. M., and Ogata, H. (2008). Taxonomic distribution of large DNA viruses in the sea. *Genome Biology* **9**(7).

Monier, A., Larsen, J. B., Sandaa, R. A., Bratbak, G., Claverie, J. M., and Ogata, H. (2008). Marine mimivirus relatives are probably large algal viruses. *Virology Journal* **5**.

Monier, A., Pagarete, A., de Vargas, C., Allen, M. J., Read, B., Claverie, J.-M., and Ogata, H. (2009). Horizontal gene transfer of an entire metabolic pathway between a eukaryotic alga and its DNA virus. *Genome Research* **19**(8)**,** 1441-1449.

Moore, C. M., Suggett, D. J., Hickman, A. E., Kim, Y. N., Tweddle, J. F., Sharples, J., Geider, R. J., and Holligan, P. M. (2006). Phytoplankton photoacclimation and photoadaptation in response to environmental gradients in a shelf sea. *Limnology and Oceanography* **51**(2)**,** 936-949.

Moreira, D., and Brochier-Armanet, C. (2008). Giant viruses, giant chimeras: the multiple evolutionary histories of mimivirus genes. *BMC Evolutionary Biology* **8**(12)**.**

Moreira, D., and Lopez-Garcia, P. (2005). Comment on "The 1.2-megabase genome sequence of Mimivirus". *Science* **308**(5725)**,** 1114; author reply 1114.

Moreira, D., and Lopez-Garcia, P. (2009). Ten reasons to exclude viruses from the tree of life. *Nature Reviews Microbiology* **7**(4)**,** 306-311.

Muhling, M., Fuller, N. J., Millard, A., Somerfield, P. J., Marie, D., Wilson, W. H., Scanlan, D. J., Post, A. F., Joint, I., and Mann, N. H. (2005). Genetic diversity of marine Synechococcus and co-occurring cyanophage communities: evidence for viral control of phytoplankton. *Environmental Microbiology* **7**(4)**,** 499-508.

Muller, D. G., Kapp, M., and Knippers, R. (1998). Viruses in marine brown algae, Vol. 50, pp. 49 - 67.

Muller, D. G., Sengco, M., Wolf, S., Brautigam, M., Schmid, C. E., Kapp, M., and Knippers, R. (1996). Comparison of two DNA viruses infecting the marine brown algae *Ectocarpus siliculosus* and *E. fasciculatus*. *J Gen Virol* **77,** 2329-33.

Muller, M. N., Antia, A. N., and LaRoche, J. (2008). Influence of cell cycle phase on calcification in the coccolithophore Emiliania huxleyi. *Limnology and Oceanography* **53**(2)**,** 506-512.

Nagasaki, K. (2008). Dinoflagellates, diatoms, and their viruses. *J Microbiol* **46**(3)**,** 235-43.

Nagasaki, K., Ando, M., Itakura, S., Imai, I., and Ishida, Y. (1994). Viral mortality in the final stage of *Heterosigma akashiwo* (Raphidophyceae) red tide. *Journal of Plankton Research* **16**(11)**,** 1595-1599.

Nagasaki, K., and Yamaguchi, M. (1997). Isolation of a virus infectious to the harmful bloom causing microalga Heterosigma akashiwo (Raphidophyceae). *Aquatic Microbial Ecology* **13**(2)**,** 135-140.

Nanninga, H. J., and Tyrrell, T. (1996). Importance of light for the formation of algal blooms by Emiliania huxleyi. *Marine Ecology-Progress Series* **136**(1-3)**,** 195-203.

Nishiyama, Y., Allakhverdiev, S. I., and Murata, N. (2006). A new paradigm for the action of reactive oxygen species in the photoinhibition of photosystem II. *Biochimica Et Biophysica Acta-Bioenergetics* **1757**(7)**,** 742-749.

Noble, R. T., and Fuhrman, J. A. (2000). Rapid virus production and removal as measured with fluorescently labeled viruses as tracers. *Applied and Environmental Microbiology* **66**(9)**,** 3790-3797.

Novoa, R. R., Calderita, G., Arranz, R., Fontana, J., Granzow, H., and Risco, C. (2005). Virus factories: associations of cell organelles for viral replication and morphogenesis. *Biology of the Cell* **97**(2)**,** 147-172.

Obeid, L. M., Linardic, C. M., Karolak, L. A., and Hannun, Y. A. (1993). Programmed Cell-Death Induced by Ceramide. *Science* **259**(5102)**,** 1769-1771.

Ogata, H., and Claverie, J. M. (2007). Unique genes in giant viruses: regular substitution pattern and anomalously short size. *Genome Res* **17**(9)**,** 1353-61.

Oh, C. S., Toke, D. A., Mandala, S., and Martin, C. E. (1997). ELO2 and ELO3, homologues of the Saccharomyces cerevisiae ELO1 gene, function in fatty acid elongation and are required for sphingolipid formation. *J Biol Chem* **272**(28)**,** 17376-84.

Paasche, E. (2001). A review of the coccolithophorid Emiliania huxleyi (Prymnesiophyceae), with particular reference to growth, coccolith formation, and calcification-photosynthesis interactions. *Phycologia* **40**(6)**,** 503-529.

Padan, E., Shilo, M., and Oppenhei.Ab (1972). Lysogeny of blue-green-alga *Plectonema boryanum* by LPP2-SPI cyanophage. *Virology* **47**(2)**,** 525-526.

Pagarete, A., Allen, M. J., Wilson, W., Kimmance, S., and de Vargas, C. (2009). Host-virus shift of the sphingolipid pathway along an *Emiliania huxleyi* bloom: survival of the fattest. *Environmental Microbiology* **11**(11)**,** 2840-2848.

Parkhill, J. P., Maillet, G., and Cullen, J. J. (2001). Fluorescence-based maximal quantum yield for PSII as a diagnostic of nutrient stress. *Journal of Phycology* **37**(4)**,** 517-529.

Paul, J. H. (2008). Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *Isme Journal* **2**(6)**,** 579-589.

Perry, D. K., Carton, J., Shah, A. K., Meredith, F., Uhlinger, D. J., and Hannun, Y. A. (2000). Serine palmitoyltransferase regulates de novo ceramide generation during etoposide-induced apoptosis. *Journal of Biological Chemistry* **275**(12)**,** 9078-9084.

Pettus, B. J., Chalfant, C. E., and Hannun, Y. A. (2002). Ceramide in apoptosis: an overview and current perspectives. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1585**(2-3)**,** 114-125.

Pienaar, R. N. (1994). Ultrastructure and calcification of coccolithophores. *In* "Coccolithophores" (A. Winter, and W. G. Siesser, Eds.), pp. 13-37. Cambridge University Press, Cambridge.

Poorvin, L., Rinta-Kanto, J. M., Hutchins, D. A., and Wilhelm, S. W. (2004). Viral release of iron and its bioavailability to marine plankton. *Limnology and Oceanography* **49**(5)**,** 1734-1741.

Prangishvili, D., Stedman, K., and Zillig, W. (2001). Viruses of the extremely thermophilic archaeon Sulfolobus. *Trends in Microbiology* **9**(1)**,** 39-43.

Proctor, L. M., and Fuhrman, J. A. (1990). Viral mortality of marine Bacteria and Cyanobacteria. *Nature* **343**(6253)**,** 60-62.

Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35**(Database issue)**,** D61-5.

Radtke, K., Dohner, K., and Sodeik, B. (2006). Viral interactions with the cytoskeleton: a hitchhiker's guide to the cell. *Cellular Microbiology* **8**(3)**,** 387-400.

Ragni, M., Airs, R. L., Leonardos, N., and Geider, R. J. (2008). Photoinhibition of PSII in *Emiliania huxleyi* (Haptophyta) under high light stress: The roles of photoacclimation, photoprotection, and photorepair. *Journal of Phycology* **44**(3)**,** 670-683.

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., and Claverie, J. M. (2004). The 1.2-megabase genome sequence of Mimivirus. *Science* **306**(5700)**,** 1344-50.

Renault, T., and Novoa, B. (2004). Viruses infecting bivalve molluscs. *Aquatic Living Resources* **17**(4)**,** 397-409.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**(6)**,** 276-7.

Ridgwell, A., and Zeebe, R. E. (2005). The role of the global carbonate cycle in the regulation and evolution of the Earth system. *Earth and Planetary Science Letters* **234**(3-4)**,** 299-315.

Riethmuller, J., Riehle, A., Grassme, H., and Gulbins, E. (2006). Membrane rafts in host-pathogen interactions. *Biochim Biophys Acta* **1758**(12)**,** 2139-47.

Rodriguez, F., Fernandez, E., Head, R. N., Harbour, D. S., Bratbak, G., Heldal, M., and Harris, R. P. (2000). Temporal variability of viruses, bacteria, phytoplankton and zooplankton in the western English Channel off Plymouth. *Journal Of The Marine Biological Association Of The United Kingdom* **80**(4)**,** 575-586.

Rohwer, F., and Thurber, R. V. (2009). Viruses manipulate the marine environment. *Nature* **459**(7244)**,** 207-212.

Roossinck, M. J. (2005). Symbiosis versus competition in plant virus evolution. *Nat Rev Microbiol* **3**(12)**,** 917-24.

Roulston, A., Marcellus, R. C., and Branton, P. E. (1999). Viruses and apoptosis. *Annual Review of Microbiology* **53,** 577-628.

Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132,** 365-86.

Ruijter, J. M., Ramakers, C., Hoogaars, W. M. H., Karlen, Y., Bakker, O., van den Hoff, M. J. B., and Moorman, A. F. M. (2009). Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data, pp. gkp045.

Saez, A. G., Probert, I., Young, J. R., Edvardsen, B., Eikrem, W., and Medlin, L. K. (2004). A review of the phylogeny of the Haptophyta. *In* "Coccolithophores: from molecular processes to global impact" (H. Thierstein, and J. R. Young, Eds.), pp. 251-270. Springer Verlag.

Sandaa, R. A. (2008). Burden or benefit? Virus-host interactions in the marine environment. *Research in Microbiology* **159**(5)**,** 374-381.

Sandaa, R. A., Heldal, M., Castberg, T., Thyrhaug, R., and Bratbak, G. (2001). Isolation and characterization of two viruses with large genome size infecting *Chrysochromulina ericina* (Prymnesiophyceae) and *Pyramimonas orientalis* (Prasinophyceae). *Virology* **290**(2)**,** 272-280.

Schramm, B., and Locker, J. K. (2005). Cytoplasmic organization of POXvirus DNA replication. *Traffic* **6**(10)**,** 839-846.

Schroeder, D. C., Oke, J., Hall, M., Malin, G., and Wilson, W. H. (2003). Virus succession observed during an Emiliania huxleyi bloom. *Applied and Environmental Microbiology* **69**(5)**,** 2484-2490.

Schroeder, D. C., Oke, J., Malin, G., and Wilson, W. H. (2002). Coccolithovirus (Phycodnaviridae): characterisation of a new large dsDNA algal virus that infects *Emiliana huxleyi*. *Arch Virol* **147**(9)**,** 1685-98.

Schroeder, D. C., Park, Y., Yoon, H. M., Lee, Y. S., Kang, W., Meints, R. H., Ivey, R. G., and Choi, T. J. (2009). Genomic analysis of the smallest giant virus - Feldmannia sp virus 158. *Virology* **384**(1)**,** 223-232.

Schuster, A. M., Girton, L., Burbank, D. E., and Van Etten, J. L. (1986). Infection of a Chlorella-like alga with the virus PBCV-1: Transcriptional studies. *Virology* **148**(1)**,** 181-189.

Seaton, G., Lee, K., and Rohozinski, J. (1995). Photosynthetic Shutdown in Chlorella NC64A Associated with the Infection Cycle of Paramecium bursaria Chlorella Virus-1. *Plant Physiol* **108**(4)**,** 1431-1438.

Senkevich, T. G., White, C. L., Koonin, E. V., and Moss, B. (2000). A viral member of the ERV1/ALR protein family participates in a cytoplasmic pathway of disulfide bond formation, Vol. 97, pp. 12068-12073.

Sharon, I., Alperovitch, A., Rohwer, F., Haynes, M., Glaser, F., Atamna-Ismaeel, N., Pinter, R. Y., Partensky, F., Koonin, E. V., Wolf, Y. I., Nelson, N., and Beja, O. (2009). Photosystem I gene cassettes are present in marine virus genomes. *Nature* **461**(7261)**,** 258-262.

Sharon, I., Tzahor, S., Williamson, S., Shmoish, M., Man-Aharonovich, D., Rusch, D. B., Yooseph, S., Zeidner, G., Golden, S. S., Mackey, S. R., Adir, N., Weingart, U., Horn, D., Venter, J. C., Mandel-Gutfreund, Y., and Beja, O. (2007). Viral photosynthetic reaction center genes and transcripts in the marine environment. *Isme Journal* **1**(6)**,** 492-501.

Shibata, A., Kogure, K., Koike, I., and Ohwada, K. (1997). Formation of submicron colloidal particles from marine bacteria by viral infection. *Marine Ecology Progress Series* **155,** 303-307.

Short, S. M., and Suttle, C. A. (1999). Use of the polymerase chain reaction and denaturing gradient gel electrophoresis to study diversity in natural virus communities. *Hydrobiologia* **401,** 19-32.

Short, S. M., and Suttle, C. A. (2002). Sequence analysis of marine virus communities reveals that groups of related algal viruses are widely distributed in nature. *Appl Environ Microbiol* **68**(3)**,** 1290-6.

Short, S. M., and Suttle, C. A. (2003). Temporal dynamics of natural communities of marine algal viruses and eukaryotes. *Aquatic Microbial Ecology* **32**(2)**,** 107-119.

Siskind, L. J. (2005). Mitochondrial ceramide and the induction of apoptosis. *Journal of Bioenergetics and Biomembranes* **37**(3)**,** 143-153.

Smith, G. A., and Enquist, L. W. (2002). Break ins and break outs: Viral interactions with the cytoskeleton of mammalian cells. *Annual Review of Cell and Developmental Biology* **18,** 135-161.

Sorensen, G., Baker, A. C., Hall, M. J., Munn, C. B., and Schroeder, D. C. (2009). Novel virus dynamics in an Emiliania huxleyi bloom. *Journal of Plankton Research* **31**(7)**,** 787-791.

Sournia, A. (1974). Circadian Periodicities in Natural-Populations of Marine Phytoplankton. *Advances in Marine Biology* **12,** 325-389.

Sperling, P., Ternes, P., Zank, T. K., and Heinz, E. (2003). The evolution of desaturases. *Prostaglandins Leukot Essent Fatty Acids* **68**(2)**,** 73-95.

Steward, G. F., Smith, D. C., and Azam, F. (1996). Abundance and production of bacteria and viruses in the Bering and Chukchi Seas. *Marine Ecology Progress Series* **131,** 287-300.

Strickland, J. D. H., and Parsons, T. R. (1968). A practical handbook of seawater analysis. *Bull. Fish. Res. Board Can.*

Strum, J. C., Swenson, K. I., Turner, J. E., and Bell, R. M. (1995). Ceramide triggers meiotic cell cycle progression in Xenopus oocytes. A potential mediator of progesterone-induced maturation. *J Biol Chem* **270**(22)**,** 13541-7.

Stukey, J., and Carman, G. M. (1997). Identification of a novel phosphatase sequence motif. *Protein Sci* **6**(2)**,** 469-72.

Sukhumsirichart, W., Kiatpathomchai, W., Wongteerasupaya, C., Withyachumnarnkul, B., Flegel, T. W., Boonseang, V., and Panyim, S. (2002). Detection of hepatopancreatic parvovirus (HPV) infection in Penaeus monodon using PCR-ELISA. *Molecular and Cellular Probes* **16**(6)**,** 409-413.

Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F., and Chisholm, S. W. (2005). Three Prochlorococcus cyanophage genomes: Signature features and ecological interpretations. *Plos Biology* **3**(5)**,** 790-806.

Sullivan, M. B., Lindell, D., Lee, J. A., Thompson, L. R., Bielawski, J. P., and Chisholm, S. W. (2006). Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *Plos Biology* **4**(8)**,** 1344-1357.

Susin, S. A., Zamzami, N., Larochette, N., Dallaporta, B., Marzo, I., Brenner, C., Hirsch, T., Petit, P. X., Geuskens, M., and Kroemer, G. (1997). A cytofluorometric assay of nuclear apoptosis induced in a cell-free system: application to ceramide-induced apoptosis. *Exp Cell Res* **236**(2)**,** 397-403.

Suttle, C. A. (1992). Inhibition of photosynthesis in phytoplankton by the submicron size fraction concentrated from seawater. *Marine Ecology-Progress Series* **87**(1-2)**,** 105-112.

Suttle, C. A. (1994). The significance of viruses to mortality in aquatic microbial communities. *Microbial Ecology* **28**(2)**,** 237-243.

Suttle, C. A. (2005a). The viriosphere: the greatest biological diversity on Earth and driver of global processes. *Environmental Microbiology* **7**(4)**,** 481-482.

Suttle, C. A. (2005b). Viruses in the sea. *Nature* **437**(7057)**,** 356-61.

Suttle, C. A. (2007). Marine viruses - major players in the global ecosystem. *Nature Reviews Microbiology* **5**(10)**,** 801-812.

Suttle, C. A., and Chan, A. M. (1993). Marine Cyanophages Infecting Oceanic and Coastal Strains of *Synechococcus* - Abundance, Morphology, Cross-Infectivity and Growth-Characteristics. *Marine Ecology-Progress Series* **92**(1-2)**,** 99-109.

Suttle, C. A., and Chan, A. M. (1994). Dynamics and Distribution of Cyanophages and Their Effect on Marine Synechococcus Spp. *Applied and Environmental Microbiology* **60**(9)**,** 3167-3174.

Suttle, C. A., and Chan, A. M. (1995). Viruses Infecting The Marine Prymnesiophyte Chrysochromulina Spp - Isolation, Preliminary Characterization And Natural-Abundance. *Marine Ecology-Progress Series* **118**(1-3)**,** 275-282.

Suttle, C. A., Chan, A. M., and Cottrell, M. T. (1990). Infection of phytoplankton by viruses and reduction of primary productivity. *Nature* **347**(6292)**,** 467-469.

Suzan-Monti, M., La Scola, B., and Raoult, D. (2006). Genomic and evolutionary aspects of Mimivirus. *Virus Research* **117**(1)**,** 145-155.

Sylvan, J. B., Quigg, A., Tozzi, S., and Ammerman, J. W. (2007). Eutrophication-induced phosphorus limitation in the Mississippi River plume: Evidence from fast repetition rate fluorometry. *Limnology and Oceanography* **52**(6)**,** 2679-2685.

Tai, V., Lawrence, J. E., Lang, A. S., Chan, A. M., Culley, A. I., and Suttle, C. A. (2003). Characterization of HaRNAV, a single-stranded RNA virus causing lysis of Heterosigma akashiwo (Raphidophyceae). *Journal Of Phycology* **39**(2)**,** 343-352.

Takemura, M. (2001). Poxviruses and the origin of the eukaryotic nucleus. *J Mol Evol* **52**(5)**,** 419-25.

Tarutani, K., Nagasaki, K., and Yamaguchi, M. (2000). Viral impacts on total abundance and clonal composition of the harmful bloom-forming phytoplankton *Heterosigma akashiwo*. *Appl Environ Microbiol* **66**(11)**,** 4916-20.

Teodoro, J. G., and Branton, P. E. (1997). Regulation of apoptosis by viral gene products. *Journal of Virology* **71**(3)**,** 1739-1746.

The UniProt, C. (2009). The Universal Protein Resource (UniProt) in 2010, pp. gkp846.

Thierstein, H. R., Geitzenauer, K. R., and Molfino, B. (1977). Global synchroneity of late quaternary coccolith datum levels - validation by oxigen isotopes. *Geology* **5**(7)**,** 400-404.

Thingstad, T. F. (2000). Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography* **45**(6)**,** 1320-1328.

Thingstad, T. F., and Lignell, R. (1997). Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquatic Microbial Ecology* **13**(1)**,** 19-27.

Thornberry, N. A., and Lazebnik, Y. (1998). Caspases: Enemies within. *Science* **281**(5381)**,** 1312-1316.

Timmermans, K. R., Gerringa, L. J. A., de Baar, H. J. W., van der Wagt, B., Veldhuis, M. J. W., de Jong, J. T. M., Croot, P. L., and Boye, M. (2001). Growth rates of large and small Southern Ocean diatoms in relation to availability of iron in natural seawater. *Limnology and Oceanography* **46**(2)**,** 260-266.

Tomaru, Y., Hata, N., Masuda, T., Tsuji, M., Igata, K., Masuda, Y., Yamatogi, T., Sakaguchi, M., and Nagasaki, K. (2007). Ecological dynamics of the bivalve-killing dinoflagellate Heterocapsa circularisquama and its infectious viruses in different locations of western Japan, Vol. 9, pp. 1376-1383.

Tomaru, Y., Katanozaka, N., Nishida, K., Shirai, Y., Tarutani, K., Yamaguchi, M., and Nagasaki, K. (2004). Isolation and characterization of two distinct types of HcRNAV, a single-stranded RNA virus infecting the bivalve-killing microalga Heterocapsa circularisquama. *Aquatic Microbial Ecology* **34**(3)**,** 207-218.

Tyrrell, T., Holligan, P. M., and Mobley, C. D. (1999). Optical impacts of oceanic coccolithophore blooms. *Journal Of Geophysical Research-Oceans* **104**(C2)**,** 3223-3241.

Tyrrell, T., and Merico, A. (2004). *Emiliania huxleyi*: bloom observations and the conditions that induce them. *In* "Coccolithophores: from molecular processes to global impact" (H. Thierstein, and J. R. Young, Eds.), pp. 75-97. Springer Verlag, Berlin, Germany.

UniProt Consortium (2008). The universal protein resource (UniProt). *Nucleic Acids Res* **36**(Database issue)**,** D190-5.

Van Bressem, M. F., Van Waerebeek, K., and Raga, J. A. (1999). A review of virus infections of cetaceans and the potential impact of morbilliviruses, poxviruses and papillomaviruses on host population dynamics. *Diseases of Aquatic Organisms* **38**(1)**,** 53-65.

Van Etten, J. L., Burbank, D. E., Xia, Y., and Meints, R. H. (1983). Growth cycle of a virus, PBCV-1, that infects Chlorella-like algae. *Virology* **126**(1)**,** 117-125.

Van Etten, J. L., Graves, M. V., Muller, D. G., Boland, W., and Delaroque, N. (2002). Phycodnaviridae - large DNA algal viruses. *Archives Of Virology* **147**(8)**,** 1479-1516.

Vardi, A., Van Mooy, B. A. S., Fredricks, H. F., Popendorf, K. J., Ossolinski, J. E., Haramaty, L., and Bidle, K. D. (2009). Viral Glycosphingolipids Induce Lytic Infection and Cell Death in Marine Phytoplankton. *Science* **326**(5954)**,** 861-865.

Vassiliev, I. R., Prasil, O., Wyman, K. D., Kolber, Z., Hanson, A. K., Prentice, J. E., and Falkowski, P. G. (1994). Inhibition of PS II photochemistry by PAR and UV radiation in natural phytoplankton communities. *Photosynthesis Research* **42**(1)**,** 51-64.

Vaulot, D., Birrien, J. L., Marie, D., Casotti, R., Veldhuis, M. J. W., Kraay, G. W., and Chretiennotdinet, M. J. (1994). Morphology, ploidy, pigment composition, and genome size of cultured strains of Phaeocystis (Prymnesiophyceae). *Journal of Phycology* **30**(6)**,** 1022-1035.

Villarreal, L. P. (2005). "Viruses and the evoltuion of life." American Society for Microbiology Press, Washington, DC.

Villarreal, L. P., and DeFilippis, V. R. (2000). A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. *J Virol* **74**(15)**,** 7079-84.

Volk, T., and Hoffert, M. I. (1985). Ocean carbon pumps: analysis of relative strength and efficiencies of in ocean-driven circulation atmospheric CO^2 changes. *In* "The carbon cycle and atmospheric CO^2: natural variation Archean to Present: AGU Monograph 32." (E. T. Sundquist, and W. S. Broecker, Eds.), pp. 99-110. American Geophysical Union, Washington, DC.

Wallich, G. C. (1877). Observations on the coccosphere. *Ann Magazine Natural History* **4,** 342-350.

Waterbury, J. B., and Valois, F. W. (1993). Resistance to cooccurring phages enables marine Synechococcus communities to coexist with Cyanophages abundant in seawater. *Applied and Environmental Microbiology* **59**(10)**,** 3393-3399.

Waters, R. E., and Chan, A. T. (1982). Micromonas pusilla Virus: the virus growth-cycle and associated physiological events within the host-cells; host range mutation. *Journal of General Virology* **63,** 199-206.

Weinbauer, M. G. (2004). Ecology of prokaryotic viruses. *Fems Microbiology Reviews* **28**(2)**,** 127-181.

Weinbauer, M. G., Fuks, D., Puskaric, S., and Peduzzi, P. (1995). Diel, seasonal, and depth-related variability of viruses and dissolved DNA in the northern Adriatic Sea. *Microbial Ecology* **30**(1)**,** 25-41.

Weinbauer, M. G., and Rassoulzadegan, F. (2004). Are viruses driving microbial diversification and diversity? *Environmental Microbiology* **6**(1)**,** 1-11.

Weinbauer, M. G., and Suttle, C. A. (1996). Potential significance of lysogeny to bacteriophage production and bacterial mortality in coastal waters of the Gulf of Mexico. *Applied and Environmental Microbiology* **62**(12)**,** 4374-4380.

Weinbauer, M. G., and Suttle, C. A. (1999). Lysogeny and prophage induction in coastal and offshore bacterial communities. *Aquatic Microbial Ecology* **18**(3)**,** 217-225.

Weiner, A. M., and Maizels, N. (1994). Molecular evolution - unlocking the secrets of retroviral evolution. *Current Biology* **4**(6)**,** 560-563.

Wen, K., Ortmann, A. C., and Suttle, C. A. (2004). Accurate estimation of viral abundance by epifluorescence microscopy. *Applied and Environmental Microbiology* **70**(7)**,** 3862-3867.

Westbroek, P., Brown, C. W., Vanbleijswijk, J., Brownlee, C., Brummer, G. J., Conte, M., Egge, J., Fernandez, E., Jordan, R., Knappertsbusch, M., Stefels, J., Veldhuis, M., Vanderwal, P., and Young, J. (1993). A Model System Approach to Biological Climate Forcing - the Example of Emiliania-Huxleyi. *Global and Planetary Change* **8**(1-2)**,** 27-46.

Westbroek, P., van Hinte, J. E., Brummer, G. J., Veldhuis, M., Brownlee, C., Green, J. C., Harris, R., and Heimdal, B. R. (1994). Emiliania huxleyi as a key to biosphere-geosphere interactions. *In* "The haptophyte algae" (J. Green, and B. Leadbeater, Eds.), pp. 321-334. Oxford University Press, Oxford.

Weynberg, K., Allen, M. J., Ashelford, K., Scanlan, D. J., and Wilson, W. H. (2009). From small hosts come big viruses: the complete genome of a second *Ostreococcus tauri* virus, OtV-1. *Environmental Microbiology* **11**(11), 2821-2839.

Wilhelm, S. W., and Suttle, C. A. (1999). Viruses and Nutrient Cycles in the Sea - Viruses play critical roles in the structure and function of aquatic food webs. *Bioscience* **49**(10)**,** 781-788.

Willner, D., Thurber, R. V., and Rohwer, F. (2009). Metagenomic signatures of 86 microbial and viral metagenomes. *Environmental Microbiology* **11**(7)**,** 1752-1766.

Wilson, W. H., Carr, N. G., and Mann, N. H. (1996). The effect of phosphate status on the kinetics of cyanophage infection in the oceanic cyanobacterium *Synechococcus* sp WH7803. *Journal of Phycology* **32**(4)**,** 506-516.

Wilson, W. H., Dale, A. L., Davy, J. E., and Davy, S. K. (2005a). An enemy within? Observations of virus-like particles in reef corals. *Coral Reefs* **24**(1)**,** 145-148.

Wilson, W. H., Schroeder, D. C., Allen, M. J., Holden, M. T. G., Parkhill, J., Barrell, B. G., Churcher, C., Harnlin, N., Mungall, K., Norbertczak, H., Quail, M. A., Price, C., Rabbinowitsch, E., Walker, D., Craigon, M., Roy, D., and Ghazal, P. (2005b). Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. *Science* **309**(5737)**,** 1090-1092.

Wilson, W. H., Schroeder, D. C., Ho, J., and Canty, M. (2006). Phylogenetic analysis of PgV-102P, a new virus from the English Channel that infects Phaeocystis globosa. *Journal Of The Marine Biological Association Of The United Kingdom* **86**(3)**,** 485-490.

Wilson, W. H., Tarran, G., and Zubkov, M. V. (2002). Virus dynamics in a coccolithophore-dominated bloom in the North Sea. *Deep-Sea Research Part Ii-Topical Studies In Oceanography* **49**(15)**,** 2951-2963.

Wilson, W. H., Tarran, G. A., Schroeder, D., Cox, M., Oke, J., and Malin, G. (2002). Isolation of viruses responsible for the demise of an Emiliania huxleyi bloom in the English Channel. *J Mar Biol Assoc U.K.* **82**(3)**,** 369-377.

Wilson, W. H., Van Etten, J. L., and Allen, M. J. (2009). The Phycodnaviridae: the story of how tiny giants rule the world. *Curr Top Microbiol Immunol* **328,** 1-42.

Winter, E., and Ponting, C. P. (2002). TRAM, LAG1 and CLN8: members of a novel family of lipid-sensing domains? *Trends Biochem Sci* **27**(8)**,** 381-3.

Wommack, K. E., and Colwell, R. R. (2000). Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* **64**(1)**,** 69-114.

Wommack, K. E., Hill, R. T., Kessel, M., Russekcohen, E., and Colwell, R. R. (1992). Distribution of viruses in the Chesapeake Bay. *Applied and Environmental Microbiology* **58**(9)**,** 2965-2970.

Wommack, K. E., Ravel, J., Hill, R. T., Chun, J. S., and Colwell, R. R. (1999). Population dynamics of Chesapeake bay virioplankton: Total-community analysis by pulsed-field gel electrophoresis. *Applied and Environmental Microbiology* **65**(1)**,** 231-240.

Wood, M. D., and Oliver, R. L. (1995). Fluorescence transients in response to nutrient enrichment of nitrogen-limited and phosphorus-limited *Microcystis aeruginosa* cultures and natural phytoplankton populations - a measure of nutrient limitation. *Australian Journal of Plant Physiology* **22**(2)**,** 331-340.

Wu, G. A., Jun, S.-R., Sims, G. E., and Kim, S.-H. (2009). Whole-proteome phylogeny of large dsDNA virus families by an alignment-free method. *Proceedings of the National Academy of Sciences* **106**(31)**,** 12826-12831.

Yang, J., Yu, Y., Sun, S., and Duerksen-Hughes, P. J. (2004). Ceramide and other sphingolipids in cellular responses. *Cell Biochem Biophys* **40**(3)**,** 323-50.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**(8)**,** 1586-91.

Yoshida, T., Nagasaki, K., Takashima, Y., Shirai, Y., Tomaru, Y., Takao, Y., Sakamoto, S., Hiroishi, S., and Ogata, H. (2008). Ma-LMM01 infecting toxic *Microcystis aeruginosa* illuminates diverse cyanophage genome strategies. *Journal of Bacteriology* **190**(5)**,** 1762-1772.

Young, E. B., and Beardall, J. (2003). Rapid ammonium and nitrate-induced perturbations to Chl *a* flourescence in nitrogen-stressed *Dunaliella tertiolecta* (Chlorophyta) *J. Phycol.* **39**(2)**,** 332-342.

Young, J. R., Davis, S. A., Bown, P. R., and Mann, S. (1999). Coccolith ultrastructure and biomineralisation. *Journal of Structural Biology* **126**(3)**,** 195-215.

Young, J. R., Didymus, J. M., Bown, P. R., Prins, B., and Mann, S. (1992). Crystal assembly and phylogenetic evolution in Heterococcoliths. *Nature* **356**(6369)**,** 516-518.

Young, J. R., Geisen, M., and Probert, I. (2005). Review of selected aspects of coccolithophore biology with implications for paleobiodiversity estimation. *Micropaleontology* **51**(4)**,** 267-288.

Zhang, Y., Strasser, P., Grabherr, R., and Van Etten, J. L. (1994). Hairpin loop structure at the termini of the chlorella virus PBCV-1 genome. *Virology* **202**(2)**,** 1079-82.

Ziveri, P., Broerse, A. T. C., van Hinte, J. E., Westbroek, P., and Honjo, S. (2000). The fate of coccoliths at 48 degrees N 21 degrees W, northeastern Atlantic. *Deep-Sea Research Part Ii-Topical Studies In Oceanography* **47**(9-11)**,** 1853-1875.

DETAILED INDEX

Abstract

Emiliania huxleyi Virus (EhV) is a giant nucleo-cytoplasmic double stranded DNA virus that belongs to the Phycodnavirus family. It has the capacity to infect *Emiliania huxleyi*, the most abundant coccolithophore in today's oceans. Population dynamics of these eukaryotic microalgae is clearly controlled by the severe lytic action of EhV. After an extended bibliographic review on the current knowledge existing on these viruses, we present a series of approaches conducted with the major aim of unveiling functional genomic features of the EhV. Evidence for the transfer of genes between *E. huxleyi*'s and the EhV genomes is presented. In particular, we investigate the origin of seven genes involved in the unique viral sphingolipid biosynthesis pathway encoded in EhV genome. This is the first clear case of horizontal gene transfer of multiple functionally-linked enzymes in a eukaryotic phytoplankton-virus system. We then focus on a natural *E. huxleyi*/EhV system and investigate the dynamics of host and virus homologous gene expression for two of the most important genes of this pathway, serine palmitoyl transferase and dihydroceramide desaturase. Three defined transcriptional stages are reported during the bloom, with the coccolithovirus transcripts taking over and controlling the SBP. Further on, host and virus global transcript abundance occurring within a natural oceanic community was investigated. The majority of the genes that significantly increased in abundance from pre to post viral takeover corresponded to viral sequences for which there is so far no match in the protein databases. Nonetheless, novel transcription features associated with EhV infection were discovered, namely the utilization of genes potentially related to genetic information processing, posttranslational control, intracellular trafficking mechanisms, and control of programmed cell death. On a final note, the collection of the works is discussed, followed by the potential implications of these findings and future research perspectives in the plankton virology field.

Key words

Virioplankton, coccolithovirus, coccolithophore, *Emiliania huxleyi*, horizontal gene transfer (HGT).