

Unveiling the transcriptional features associated with coccolithovirus infection of natural *Emiliania huxleyi* blooms

António Pagarete^{1,2}, Gildas Le Corguillé³, Bela Tiwari⁴, Hiroyuki Ogata⁵, Colomban de Vargas¹, William H. Wilson⁶ & Michael J. Allen²

¹Equipe EPPO-Evolution du Plancton et PaléoOcéans, CNRS-UMR7144, Université Pierre et Marie Curie, Station Biologique, Roscoff, France; ²Plymouth Marine Laboratory, The Hoe, Plymouth, UK; ³CNRS/UMPC, FR2424, Service Informatique et Génomique, Station Biologique, Roscoff, France; ⁴NERC Environmental Bioinformatics Centre, Centre for Ecology and Hydrology, Wallingford, UK; ⁵Structural and Genomic Information Laboratory, CNRS-UPR2589, Mediterranean Institute of Microbiology (IFR-88), Aix-Marseille University, Marseille, France; and ⁶Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, ME, USA

Correspondence: Michael J. Allen, Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth PL1 3DH, UK. Tel.: +44 1752 633472; fax: +44 1752 633101; e-mail: mija@pml.ac.uk

Present addresses: António Pagarete, Department of Biology, University of Bergen, Thormøhlensgate 53 A & B, N-5006, Bergen, Norway; Bela Tiwari, CLC bio, Finlandsgade 10-12, Katrinebjerg, 8200 Aarhus N, Denmark.

Received 9 May 2011; revised 1 August 2011; accepted 12 August 2011. Final version published online 22 September 2011.

DOI: 10.1111/j.1574-6941.2011.01191.x

Editor: Gary King

Keywords

host-virus interactions; functional genomics; *Phycodnaviridae*; marine microbial ecology; marine transcriptomics.

Introduction

Ever since the overwhelming abundance and extreme functional and genetic diversity represented by marine viruses was revealed, the study of oceanic virioplankton has gained increasing attention (reviewed by Suttle, 2005; Jacquet *et al.*, 2010). Viral control of host population development becomes particularly evident in situations of algal blooming, with viruses now unambiguously identified as being responsible for bloom termination in many natural environmental systems (Maranger *et al.*, 1994; Nagasaki *et al.*, 1994; Castberg *et al.*, 2001; Larsen *et al.*, 2004; Brussaard *et al.*, 2005). Accordingly, the study of

Abstract

Lytic viruses have been implicated in the massive cellular lysis observed during algal blooms, through which they assume a prominent role in oceanic carbon and nutrient flows. Despite their impact on biogeochemical cycling, the transcriptional dynamics of these important oceanic events is still poorly understood. Here, we employ an oligonucleotide microarray to monitor host (Emiliania huxlevi) and virus (coccolithovirus) transcriptomic features during the course of E. huxleyi blooms induced in seawater-based mesocosm enclosures. Host bloom development and subsequent coccolithovirus infection was associated with a major shift in transcriptional profile. In addition to the expected metabolic requirements typically associated with viral infection (amino acid and nucleotide metabolism, as well as transcription- and replication-associated functions), the results strongly suggest that the manipulation of lipid metabolism plays a fundamental role during host-virus interaction. The results herein reveal the scale, so far massively underestimated, of the transcriptional domination that occurs during coccolithovirus infection in the natural environment.

> viral ecology is now firmly placed at the forefront of marine research (Brussaard, 2004).

> The extensive application of metagenomic and metatranscriptomic sequencing techniques has provided an abundance of novel genetic information relating to marine viruses (Breitbart *et al.*, 2007). Yet, determining the functional impact and role of viruses on a global and ecological scale remains a formidable challenge. *Emiliania huxleyi* is the most numerous and ubiquitous coccolithophore (calcifying eukaryotic microalgae) in today's oceans and forms enormous mesoscale seasonal blooms (Brown & Yoder, 1994). Specific viruses increase in abundance during these *E. huxleyi* blooms and are closely linked to their

sudden crashes (Bratbak et al., 1993, 1996; Jacquet et al., 2002). In 1999, a lytic DNA virus, referred to as E. huxleyi virus (EhV) from the genus coccolithovirus (family Phycodnaviridae), was isolated at the terminal stages of one such bloom in the English Channel (Wilson et al., 2002). This virus is phylogenetically related to other Phycodnaviruses, which are in their turn members of an extensive group of nucleocytoplasmic large DNA viruses (NCLDVs) (Iyer et al., 2006). A recent study demonstrated a clear correlation between the dynamics of natural E. huxleyi blooms and the transcriptional activation of several EhV genes (Pagarete et al., 2009). Whilst a restricted and targeted gene analysis provides in depth functional insight into particular processes of interest (in this case sphingolipid biosynthesis), a system-wide approach is essential for broader understanding of the physiological and biochemical interactions that occur during viral infection. Furthermore, laboratory-based studies, whilst offering an invaluable opportunity to study the infection process under highly controllable conditions, often do not reflect the transcriptional dynamics of what actually occurs under natural and variable environmental conditions.

Thus, here, we present the transcriptional profiles of *E. huxleyi* blooms within natural oceanic communities using a microarray-based approach. We simultaneously monitor the progression of host and virus transcript abundance in a coccolithophore-induced bloom from a mesocosm experiment which focussed on the role of nutrient availability on coccolithophores–coccolithovirus dynamics. The transcriptional profiles obtained were analysed to assess diel cycling, phosphate availability and temporal bloom development in the natural environment.

Materials and methods

Set-up of the mesocosm experiment

The E. huxleyi-induced blooms were conducted in the Raunefjorden, Western Norway coast, at the Espeland Marine Biological Field Station, for 17 days (5-21 of June 2008). Six transparent polyethylene enclosures (11 m³; 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 & Heimdal, 1994) and filled with unfiltered fiord water collected 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 six enclosures (enc.) were divided into two treatment groups allowing triplication of each treatment: phosphate replete (enc. 1, 3 and 5) and phosphate deplete (enc. 2, 4 and 6). Nutrients were added at 15:00 h daily at an N/P ratio of 15 : 1 (1.5 µM NaNO₃ and 0.1 μ M KH₂PO₄) to the phosphate replete enclosures and at a ratio of 75 : 1 (1.5 μ M NaNO₃ and 0.02 μ M KH₂PO₄) to the phosphate deplete enclosures. Four daily samples (06, 12, 18 and 24 h) were taken from the surface of each mesocosm with 20 L carboys. Samples were immediately brought to the laboratory where 1.5 L of each sample was filtered onto 0.45- μ m pore size and 47-mm-diameter Supor-450 filters (PALL Corp.).

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) equipped with an air-cooled laser providing 15 mW at 488 nm and with standard filter set-up. Algal counts were taken from fresh samples, with the addition of 1-µm fluorescent beads (Molecular Probes, Eugene, OR). Autotrophic groups were discriminated on the basis of their forward light scatter and chlorophyll fluorescence (for details see Jacquet *et al.*, 2002). For viruses, the samples were fixed with glutaraldehyde (0.5% final concentration), stored at 4 °C in the dark for 30 min, frozen in liquid nitrogen and stored at -80 °C. The samples were stained with SYBR Green I (Molecular Probes) and analysed according to Marie *et al.* (1999).

RNA extraction, message amplification, labelling and microarrays

All molecular biology and bioinformatic protocols are described extensively in the Supplementary Methods. RNA was extracted as described previously (Pagarete et al., 2009). Random amplification of the entire mRNA population was achieved using the Microarray Target Amplification kit (Roche) and T7 Microarray RNA Target Synthesis kit (Roche) and purified using the Microarray Target Purification kit (Roche) according to the manufacturer's instructions. The microarray is described extensively in the NCBI's Gene Expression Omnibus (accession number GSE24341) (http://www.ncbi.nlm.nih.gov/geo/ query/acc.cgi?acc=GSE24341). Briefly, 70mer oligonucleotides were designed and synthesized for an E. huxleyi 1516 EST sequence library (http://www.nematodes.org/Neglected Genomes/EMILIANIA/) and for every EhV gene by Operon GmbH, making 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. Noncoding 70 base primers were also included which corresponded to the sequence directly upstream from the start codon to the EhV-86 CoDing Sequences (CDSs) found in a 104-kbp section of the genome that has previously been identified as containing unique putative promoter elements known as family A repeats (Wilson et al., 2005; Allen et al., 2006a). The data discussed in this publication have also been deposited in NCBI's Gene Expression

Omnibus (Edgar *et al.*, 2002) and are accessible through GEO Series accession number GSE24341. Overall fluorescence values between different microarray chips were normalized using the quantile method using R (Team RDC, 2009) and the limma package (Smyth, 2005).

Overall host vs. virus transcript signal comparison

Evolution of global host and virus transcript signals was assessed using a quantile distribution analysis. Medians for each gene probe were assigned into 10 quantile categories, given ordinal factors from 0 to 9 (3, for example, meaning the median fluorescent signal for a certain probe was in a percentile range between 30% and 40% of the overall intensities on a given array). Average percentile position was then calculated for the total of the *E. huxleyi* and EhV probes.

Hierarchical cluster analysis

The microarray data from all the enclosures and from all time points were combined to perform hierarchical clustering analyses using TIGR Multi Experiment Viewer (Saeed *et al.*, 2006). Nonnegative Matrix Factorization (NMF) was used, a technique that makes use of an algorithm based on decomposition by parts of an extensive data matrix into a small number of relevant metasamples (Brunet *et al.*, 2004). A battery of hierarchical clustering algorithms based on different distance metrics (Euclidean distance, Manhattan distance, Pearson Correlation, Pearson Uncentered) was also performed to test the NMF results.

Pre- vs. postviral takeover transcript analysis

Based on the clusters obtained with the hierarchical analyses, a two-unpaired Significance Analysis of Microarrays (SAM, Tusher *et al.*, 2001) was performed to identify genes that consistently changed expression from pre- to postinfection stages (Cluster 1 and Cluster 2, respectively). Upregulated calls for each gene were generated on the basis of a Delta value of 2.793 (FDR median = 0%), combined with a Cluster2/Cluster1 detection threshold above 2 and 4 for EhV and *E. huxleyi* probes, respectively.

Sequence analysis and annotation

Emiliania huxleyi EST sequences represented on the microarray were searched against UNIPROT protein sequence database (UniProt, 2010) using BLASTX (Altschul *et al.*, 1997) with an *E*-value cut-off of $1e^{-3}$. All possible stop-to-stop open reading frames (\geq 50 aa) were extracted 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 $1e^{-5}$. Coccolithovirus gene annotation data were retrieved from NCBI genbank (http://www.ncbi.nlm.nih.gov/Genbank), accession number AJ890364. The majority of probes used in the microarray targeted genes for which function could not be predicted (74%); 41% of these refer to *E. huxleyi* genes and the other 33% to EhV.

Comparison with qPCR data

The RNA preparations used for microarray hybridization were analysed, using qPCR techniques, to confirm transcript abundance fluctuation of two E. huxleyi and three EhV genes, respectively. qRT-PCR data referring to P-replete enclosures (enc. 1, 3 and 5) have been published previously in the study by Pagarete et al. (2009). gRT-PCR data referring to P-deplete enclosures (enc. 2, 4 and 6) are presented here for the first time. qRT-PCR procedures are described in detail in the Supplementary Methods. Briefly, primers were designed to target E. huxleyi's β-tubulin and coccolithovirus major capsid protein genes, as well as two key homologous sphingolipid pathwayencoding genes present in both virus and host: serine palmitoyltransferase and dihydroceramide desaturase (Supporting information Table S1). To eliminate any possible bias because of the presence of DNA traces in the RNA isolates, the final expression value for each gene in each sample was adjusted by subtracting the amplification signal of the corresponding '-Reverse Transcription' control (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 normalized to the abundance of E. huxleyi cells (previously enumerated by FCM).

Results

General bloom/infection dynamics

Initial *E. huxleyi* concentration in the fjord was 1.7×10^2 cells mL⁻¹. A consistent pattern of *E. huxleyi* bloom development followed by coccolithovirus infection was observed in all six enclosures. Days 7–13 of the study were characterized by exponential growth of the *E. huxleyi* population in all enclosures (Fig. S1). Phosphate deplete enclosures displayed reduced growth rates in comparison with the phosphate replete enclosures. Maximum *E. huxleyi* concentrations of 1.3×10^5 cells mL⁻¹ (day 13), 1.7×10^5 cells mL⁻¹ (day 12) and 1.2×10^5

cells mL⁻¹ (day 15) were observed in the phosphate replete enc. 1, 3 and 5, respectively. Maximum E. huxleyi concentrations of 7.6 \times 10⁴ cells mL⁻¹ (day 14), 6.1 \times 10^4 cells mL⁻¹ (day 15) and 6.1 \times 10⁴ cells mL⁻¹ (day 15) were observed in the phosphate deplete enc. 2, 4 and 6, respectively. Emiliania huxleyi exponential growth phase was followed by sharp decline in all enclosures. This decline coincided with the exponential increase of coccolithovirus particles in the water from around day 12 onwards in all enclosures except enc. 6. In phosphate replete enclosures (enc. 1, 3 and 5), following the rapid increase in coccolithovirus, there followed a rapid decline in coccolithovirus abundance, a decline which was observed in only one of the phosphate deplete enclosures (enc. 2). The slower E. huxlevi and subsequent coccolithovirus population development observed in enc. 4 and 6 prevented a similar observation in those enclosures owing to the termination of the experimental sampling period. Maximum concentrations of 1.6×10^7 (day 15), 1.3×10^7 (day 15), 3.1×10^7 (day 14) and 2.4×10^7 (day 15) coccolithoviruses per mL were observed in enc. 1, 2, 3 and 5, respectively.

Global transcriptional analysis: overview

Based on the observed bloom dynamics, RNA extractions were utilized for transcriptomic analyses using microarrays on samples taken between days 8 to 16 from all six enclosures. To analyse the transcriptional signature of bloom progression, samples taken at 6 AM from all enclosures were analysed. As part of this analysis strategy, the impact of phosphate availability could also be assessed by treating enc. 1, 3 and 5, and enc 2, 4 and 6 as biological triplicates for phosphate replete and deplete scenarios, respectively. To assess the transcriptional signature associated with daily cycling, as well as enabling the finer mapping of bloom progression, samples were also analysed from enc. 2 (nutrient deplete) and 3 (nutrient replete) at 12 PM, 6 PM and 12 AM between days 8 and 16.

Global transcript abundance: bloom progression

To simplify the process of analysing a complex community transcriptional data set, transcriptional profiles were initially assessed using a quantile distribution analysis. Each gene on each microarray was assigned into one of the ten quantile categories depending on the intensity of its fluorescent signal (0 representing the lowest, 9 the highest). Average quantile position was then calculated and studied over the course of the sampling period for individual genes, as well as for groups of genes such as



Fig. 1. Relative progression of *E. huxleyi* and coccolithovirus global transcript signals in the six replicate mesocosm enclosures (enc. 1–6). A similar pattern of viral transcript increase was observed in the six enclosures, accompanying the development of the host bloom. *Y*-axis scale represents the average quantile position occupied by host and virus probe fluorescence signal. Relative cell and virus abundances are plotted for reference (black dashed and solid lines, respectively; for absolute concentrations, please refer to Fig. S1). Background white/grey colour code corresponds to the two sample groups retrieved after the hierarchical clustering analyses: white, Cluster 1 (samples from early *E. huxleyi* bloom development); grey, Cluster 2 (samples from the late bloom where viral expression had taken over and was widespread throughout the *E. huxleyi* community).

'host' or 'virus'. In the six replicate enclosures, an initial period of host transcript dominance (reflected by occupying quantile scores > 4.5) was followed by a progressive increase in viral transcripts (Fig. 1). This trend accompanied the development of the coccolithovirus infection and correlated with the appearance of newly formed virions in the environment (Fig. 1). Furthermore,

hierarchical clustering analysis (using NMF) consistently separated microarray hybridizations into two clusters displaying two significantly different transcription profiles (cophenetic correlation = 0.99). Cluster 1 contained samples taken only from the initial stages of E. huxleyi bloom development, while Cluster 2 contained samples from the later stages of the bloom where viral expression had become more dominant within the E. huxlevi community (Fig. 1 and Fig. S2, Table S2). These results were corroborated by other hierarchical clustering analyses based on different distance metrics (Euclidean distance, Manhattan distance, Pearson Correlation, Pearson Uncentered).

The NMF clustering analysis failed to retrieve any significant diel cycle–related pattern from the intensive profiling of enc. 2 or 3, nor did clustering relate to either of the two different phosphate treatments (Fig. S3). As verified by the global quantile analysis, the only significant transcriptomic distinction observed separated earlier bloom stage samples against later (post-infection) stages.

Gene transcript variation along infection: Cluster 1 (earlier bloom) vs. Cluster 2 (late bloom/viral takeover)

Based on the major clusters obtained in the previous hierarchical clustering analyses, SAM was used to access significant changes in specific transcript abundances between Cluster 1 and Cluster 2. Significant transcriptional changes were detected, almost always in cases of transcript increase. Significant transcript decrease was detected for three EhV-86 probes (ehv226, ehv326_451_520 and mija_ehv218). For ehv226, an alternative probe, ehv226_1659_1728, actually increased transcription levels, but only 1.4-fold. Significant decrease in fluorescence could not be detected for any of the *E. huxleyi* probes.

In total, 218 EhV-associated probes significantly increased transcription signal towards the later bloom stages. From these, 101 corresponded to predicted EhV genes: 96 unique genes (see details in Table S4), approximately 21% of the previously predicted EhV CDSs



Fig. 2. Upregulation of EhV genomic elements from early bloom stages (Cluster 1) to later bloom stages (Cluster 2), plotted on the circular representation of the EhV genome. The outside scale is numbered clockwise in bp. Circle 1 (from outside in) represents the CDSs from the complete EhV-86 genome, starting with CDS ehv001 at position 276 bp. CDSs are colour coded: red, significant increase in transcriptomic signal; grey, no significant increase. Circle 2 represents the CDSs from the EhV-163 genome used in this microarray set. Colour codes correspond to: orange, significant increase in transcriptomic signal; grey, no significant increase. In circles 1 and 2, green colour represents CDS that was upregulated simultaneously for EhV-86 and EhV-163 probes. Circle 3 represents location of the family A promoter regions (Wilson *et al.*, 2005; Allen et al., 2006a, b, c, d). Colour code indicates: blue, significant increase in transcriptomic signal; grey, no significant increase.

(Wilson et al., 2005). Regarding their position, and using the EhV-86 genome as a reference, these CDSs are scattered throughout the genome (Fig. 2). The majority of these probes (79) corresponded to EhV-86 specific sequences, but also 22 of the probes corresponded to genes designed using information from the Norwegian isolate EhV-163 (Allen et al., 2006a, b, c, d) (Fig. 2, circles 1 and 2, respectively). Indeed, six genes (ehv060, ehv142, ehv173, ehv206, ehv235 and ehv374) displayed significant increases in transcript abundance for both EhV-86 and EhV-163 specifically designed probes (Fig. 2, genes marked in green). The majority of these EhV CDSs (87%) do not have a predicted function owing to the lack of homology to the existing protein databases. The upregulated EhV probes associated with a putative function belonged to three main KOG categories: amino acid transport and metabolism; lipid transport and metabolism; nucleotide metabolism, transcription, replication and repair (Table S4).

A section of the EhV-86 genome (approximately 104 kb) has previously been identified as containing unique putative promoter elements known as family A repeats (Allen et al., 2006a, b, c, d). From the 154 probes designed to target the 70 bases directly upstream of the starting methionine of CDSs in this region, 74 (48%) displayed significantly increased signal as EhV infection took over (Fig. 2, circle 3, marked in blue), suggesting that although noncoding, significant portions of the upstream regions may well be transcribed during the process of expressing the genes in this region. Moreover, 29 probes targeting unannotated but potential coding sequences in EhV-86 also significantly increased transcript levels.

Significant transcript increase as the bloom progressed from host to viral dominance (displaying at least fourfold variation) was also observed for 81 of the *E. huxleyi* probes (4% of the total host probes on the array). The corresponding genes related to different cellular functions (grouped into 14 KOG classes, Fig. 3, Table S4). The three classes with highest number of genes were the following: translation, ribosomal structure and biogenesis (eight genes); energy production and conversion (seven genes); lipid transport and metabolism (six genes). For 42 (52%) of the upregulated *E. huxleyi* probes, the respective genes could not be annotated because of the absence of homology in the existing protein databases.

Discussion

We launched this study with the aim of bringing phytoplankton virology from specific laboratory/strain approaches back to a wide natural community-targeted approach. We did this by creating a microarray data set that we expected would be capable of distinguishing specific synchronized transcriptomic responses that occur within the E. huxleyi natural community during coccolithovirus infection. The data presented here clearly demonstrate the validity of this approach. A clear transcriptional shift during natural E. huxleyi blooms was consistently reflected in the separation of the samples into two major groupings, those from earlier bloom stages and those of later stages. The community takeover by E. huxleyi cells during bloom stages (maximum numbers reaching 1×10^7 cells mL⁻¹) and its following repression by an extremely lytic viral assault is remarkable. The magnitude of this phenomenon was extensive enough to prevent the respective transcriptomic production from being too 'diluted' in the pool of total transcript noise naturally present in the environment.

Characteristic of any approach to identify environmental patterns, our study was not exempt from unexpected caveats. Despite our success at identifying two broad transcriptional stages during coccolithophore bloom progression, we were unable to identify the finer and more structured transcriptional nuances associated with host



Fig. 3. Upregulation of *E. huxleyi* genes, from early bloom (Cluster 1) to late bloom (Cluster 2), grouped by KOG functional class. KOG refers to the NCBI's list of EuKaryotic Orthologous Groups of proteins. Scale on the *x*-axis indicates number of genes per class. Please refer to Fig. S4 for further details.

function. Indeed, no significant diel cycle patterns or any nutrient starvation pathways were identified in the course of this work. These may be too 'quiet' to study successfully using this type of approach. It is also likely that natural communities do not display the high levels of synchronicity shown in laboratory systems. A more targeted approach (using qRT-PCR for example) has the potential to yield more insight into these transcriptional processes. The crucial factor here, though, is having *a priori* knowledge of the appropriate targets, knowledge you can acquire using the global and broad microarray approach.

The validity of the microarray transcription data was also assessed through comparison with qRT-PCR gene expression measurements obtained for five of the genes present in the microarray (published previously in Pagarete et al., 2009), specifically two E. huxleyi genes (betatubulin and serine palmitoyltransferase) and three EhV genes (major capsid protein, serine palmitoyltransferase, and dihydroceramide desaturase). For three of these genes (E. huxleyi's beta-tubulin, and EhV's serine palmitoyltransferase and dihydroceramide desaturase), the transcription profiles obtained with the two techniques (microarray and qRT-PCR) were clearly correlated (Fig. S4a-c, respectively). However, the microarray probe for EhV major capsid protein consistently presented high and saturated fluorescence levels, impeding the recognition of any transcription regulation dynamics for this gene using the microarray approach employed (Fig. S4d). Regarding E. huxleyi's serine palmitoyltransferase, the respective microarray probe presented a transcription pattern that clearly related to the gRT-PCR and microarray transcriptional data obtained for the viral homologue of this gene (Fig. S4e). Close analysis of the probe design showed that the E. huxleyi serine palmitoyltransferase probe shares 48% sequence identity with the respective EhV serine palmitoyltransferase probe, fundamentally owing to direct result of the shared evolutionary history because of horizontal gene transfer of the genes (Fig. S5) (Monier et al., 2009). From past experience, this low similarity between host and virus probes would not be considered likely to create a significant cross-hybridization event. Rigorous design parameters were employed during microarray design and fabrication to ensure that no significant cross-hybridization would occur among the known genomic material being targeted. However, and to ensure that false results did not mislead us as a consequence of previously unconsidered cross-hybridization between distinct species (i.e. the host and virus), we further blasted all 70mer oligonucleotides against the existing genomic information for E. huxleyi and coccolithoviruses. The retrieved 26 probes (host and virus serine palmitoyltransferase included) with potential,

albeit weak, for cross-hybridization were removed from further analysis (Table S3).

Beyond the mentioned caveats, the microarray set used in this study was capable of detecting a wide range of host and virus transcripts present in the natural environment, reflecting sufficient genomic identity between the model and the environmental strains. In what concerns the host, natural E. huxlevi communities have been reported to be genetically rich but still highly conserved. Namely, DGGE-based studies carried out during this same mesocosm experiment demonstrated that at least five genotypes of E. huxleyi's calcium binding protein could always be detected, corresponding to E. huxleyi sequences known to occur in these fjords (Martinez-Martinez et al., 2006; Sorensen et al., 2009). Regarding the virus, an unexpected high level of genomic compatibility between the EhV communities existing in that Norwegian fjord and the EhV-86-based probes which dominated the microarray (a virus that was isolated 9 years before in the English Channel, around 1000 km away) was observed. This was a surprising result, considering the extreme strain richness and diversity recently reported for EhV natural communities (Rowe et al., 2011), and a clear indication of global genomic conservation among this extensively spread oceanic virus (Allen et al., 2007). Notably, high levels of genomic conservation were also observed for the unique EhV promoter sequences found immediately upstream of 86 CDSs [all characterized by the presence of the nonamer sequence GTTCCC(T/C)AA] (Allen et al., 2006a, b, c, d), a result that provides more credence to the functional importance of these genomic elements during viral infection.

Coccolithoviruses possess a very complex and 'rich' genome of approximately 470 estimated coding sequences (Wilson *et al.*, 2005; Allen *et al.*, 2006a, b, c, d). Using this microarray, we were able to detect the requirement and consequent transcriptomic upregulation of a high percentage (at least 21%) of the currently predicted EhV genes. Numerous amino acid and nucleotide metabolism–related genes (both host and virus encoded) were highly transcribed during infection, most probably in response to the heavy viral requirement for genetic information processing. Translation and protein production functions also looked assured as there was a clear increase in transcriptomic signal from *E. huxleyi* genes related to mRNA splicing and ribosomes.

This study revealed a high demand for a surprisingly large number of virus and host genes involved in lipid transport and metabolism during infection. Eukaryotic cells display complex membrane systems, and although the role of lipids during EhV infection is still poorly understood, it is becoming more and more evident that they are fundamental to the infection process (Han *et al.*, 2006; Pagarete et al., 2009). In particular, the EhV genome encodes a unique de novo sphingolipid pathway which it acquired from its host through sequential horizontal gene transfer (Monier et al., 2009). These genes, often linked to the formation of lipid rafts and signal transduction mechanisms (Hannun & Obeid, 2008), are predicted to be fully functional and highly expressed during infection (Pagarete et al., 2009). Control of lipid production could be related to the need of controlling host membrane/virus interactions during infection, eventually leading to vesicle-mediated capsid transport inside a crowded intracellular eukaryotic environment. This hypothesis is plausible under the lights of other studies demonstrating that several viruses require active mechanisms for directed transport inside the cell (Novoa et al., 2005; Radtke et al., 2006). Indeed, a recent microscopybased study has clearly identified controlled EhV-86 capsid migration inside the cell and a membrane budding -mediated mechanism for virion release (Mackinder et al., 2009). Our transcriptomic results (high lipid metabolism demand during infection plus upregulation of different genes normally related to intracellular movement) add considerable credit to the theory of lipid raft formation and vesicle-mediated transport of viral proteins during EhV infection.

It was known from the outset that the microarray approach undertaken here would never allow definitive conclusions on the functions associated with the plethora of genes under study; however, such an approach has retrieved important indications for future studies. Hitherto, little information was available on the global transcriptional events that occur during coccolithovirus infection and even less under natural environmental conditions. Here, we have shown a wide range of host gene functions that remained active or, alternatively, were virally induced during the course of infection. At this stage, we cannot distinguish between viral-induced gene expression for infection purposes or the host-mediated defence response to viral infection. However, the results here confirm coccolithovirus infection to be an all-consuming and overwhelming phenomenon during E. huxleyi bloom formation, a process reflected in the community-wide transcriptional profile and indicative of the crucial role that viruses play in community population dynamics.

Acknowledgements

The authors leave here a special thank you to Laurence Garczarek for the discussions on microarray analysis. We acknowledge Margaret Hughes, Daryl Williams and Lisa Olohan (from the School of Biological Sciences, University of Liverpool, UK) for printing the arrays and for guidance with the microarray protocol. A.P. 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 (2004-08) (GOGE-CT-505403). A.P. and C.dV. were also supported by the EU-funded projects BioMarKs (EraNet Biodiversa) and EPOCA (FP7 grant agreement 211384). M.J.A. was supported by grants awarded to W.H.W. from the Natural Environment Research Council (NERC) Environmental Genomics thematic program (ref. NE/A509332/1 and NE/D001455/1) and Oceans 2025. W.H.W. was supported by a National Science Foundation (NSF) Grant ref. EF0723730. H.O. was in part supported by the French National Research Agency (Grant no. ANR-09-PCS-GENM-218).

References

- Allen MJ, Schroeder DC & Wilson WH (2006a) 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**: 525–535.
- Allen MJ, Schroeder DC, Holden MTG & Wilson WH (2006b) Evolutionary history of the Coccolithoviridae. *Mol Biol Evol* **23**: 86–92.
- Allen MJ, Schroeder DC, Donkin A, Crawfurd KJ & Wilson WH (2006c) Genome comparison of two Coccolithoviruses. *Virol J* 3: 15.
- Allen MJ, Forster T, Schroeder DC, Hall M, Roy D, Ghazal P & Wilson WH (2006d) Locus-specific gene expression pattern suggests a unique propagation strategy for a giant algal virus. *J Virol* **80**: 7699–7705.
- Allen MJ, Martinez-Martinez J, Schroeder DC, Somerfield PJ & Wilson WH (2007) Use of microarrays to assess viral diversity: from genotype to phenotype. *Environ Microbiol* **9**: 971–982.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.
- Bratbak G, Egge JK & Heldal M (1993) Viral mortality of the marine alga *Emiliania huxleyi* (Haptophyceae) and termination of algal blooms. *Mar Ecol Prog Ser* **93**: 39–48.
- Bratbak G, Heldal M, Thingstad TF & Tuomi P (1996) Dynamics of virus abundance in coastal seawater. *FEMS Microbiol Ecol* **19**: 263–269.
- Breitbart M, Thompson LR, Suttle CA & Sullivan MB (2007) Exploring the vast diversity of marine viruses. *Oceanography* **20**: 135–139.
- Brown CW & Yoder JA (1994) Coccolithophorid blooms in the global ocean. *J Geophys Res* **99**: 7467–7482.
- Brunet JP, Tamayo P, Golub TR & Mesirov JP (2004) Metagenes and molecular pattern discovery using matrix factorization. *PNAS* **101**: 4164–4169.

Brussaard CP (2004) Viral control of phytoplankton

populations – a review. *J Eukaryot Microbiol* **51**: 125–138. Brussaard CPD, Kuipers B & Veldhuis MJW (2005) A

mesocosm study of *Phaeocystis globosa* population dynamics – 1. Regulatory role of viruses in bloom. *Harmful Algae* 4: 859–874.

Castberg T, Larsen A, Sandaa RA *et al.* (2001) Microbial population dynamics and diversity during a bloom of the marine coccolithophorid *Emiliania huxleyi* (Haptophyta). *Mar Ecol Prog Ser* 221: 39–46.

Edgar R, Domrachev M & Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**: 207–210.

Egge JK & Heimdal BR (1994) Blooms of phytoplankton including *Emiliania huxleyi* (Haptophyta) – effects of nutrient supply in different N-P ratios. *Sarsia* 79: 333–348.

Han GS, Gable K, Yan LY *et al.* (2006) Expression of a novel marine viral single-chain serine palmitoyltransferase and construction of yeast and mammalian single-chain chimera. *J Biol Chem* 281: 39935–39942.

Hannun YA & Obeid LM (2008) Principles of bioactive lipid signalling: lessons from sphingolipids. *Nat Rev Mol Cell Bio* 9: 139–150.

Iyer LM, Balaji S, Koonin EV & Aravind L (2006) Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Res* 117: 156–184.

Jacquet S, Heldal M, Iglesias-Rodriguez D, Larsen A, Wilson W & Bratbak G (2002) Flow cytometric analysis of an *Emiliana huxleyi* bloom terminated by viral infection. Aquat Microb Ecol 27: 111–124.

Jacquet S, Miki T, Noble R, Peduzzi P & Wilhelm S (2010) Viruses in aquatic ecosystems: important advancements of the last 20 years and prospects for the future in the field of microbial oceanography and limnology. *Adv Oceanogr Limnol* 1: 71–101.

Koonin E, Fedorova N, Jackson J et al. (2004) A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol* 5: R7.

Larsen A, Flaten GAF, Sandaa RA *et al.* (2004) Spring phytoplankton bloom dynamics in Norwegian coastal waters: microbial community succession and diversity. *Limnol Oceanogr* **49**: 180–190.

Mackinder LCM, Worthy CA, Biggi G et al. (2009) A unicellular algal virus, *Emiliania huxleyi* virus 86, exploits an animal-like infection strategy. J Gen Virol **90**: 2306–2316.

Maranger R, Bird DF & Juniper SK (1994) Viral and bacterial dynamics in Artic sea-ice during the spring algal bloom near Resolute, NWT, Canada. *Mar Ecol Prog Ser* 111: 121–127.

Marie D, Brussaard CPD, Thyrhaug R, Bratbak G & Vaulot D (1999) Enumeration of marine viruses in culture and natural samples by flow cytometry. *Appl Environ Microbiol* **65**: 45–52.

Martinez-Martinez J, Schroeder DC, Larsen A, Bratbak G & Wilson WH (2007) Molecular dynamics of *Emiliania huxleyi* and their co-occurring viruses during two separate mesocosm studies. *Appl Environ Microbiol* **73**: 554–562.

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

Nagasaki K, Ando M, Itakura S, Imai I & Ishida Y (1994) Viral mortality in the final stage of *Heterosigma akashiwo* (Raphidophyceae) red tide. *J Plankton Res* 16: 1595–1599.

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

Pagarete A, Allen MJ, Wilson W, Kimmance S & de Vargas C (2009) Host–virus shift of the sphingolipid pathway along an *Emiliania huxleyi* bloom: survival of the fattest. *Environ Microbiol* 11: 2840–2848.

Radtke K, Dohner K & Sodeik B (2006) Viral interactions with the cytoskeleton: a Hitchhiker's guide to the cell. *Cell Microbiol* 8: 387–400.

Rowe JM, Fabre M-F, Gobena D, Wilson WH & Wilhelm SW (2011) Application of the major capsid protein as a marker of the phylogenetic diversity of *Emiliania huxleyi* viruses. *FEMS Microbiol Ecol* **76**: 373–380.

Saeed AI, Bhagabati NK, Braisted JC et al. (2006) TM4 Microarray Software Suite. Methods Enzymol 411: 134–193.

Smyth GK. (2005) Limma: linear models for microarray data.
Bioinformatics and Computational Biology Solutions using R and Bioconductor. (Gentleman R, Carey V, Dudoit S, Irizarry R & Huber H, eds), pp. 397–420. Springer, New York.

Sorensen G, Baker AC, Hall MJ, Munn CB & Schroeder DC (2009) Novel virus dynamics in an *Emiliania huxleyi* bloom. *J Plankton Res* 31: 787–791.

Suttle CA (2005) Viruses in the sea. Nature 437: 356-361.

Team RDC. (2009) R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Team RDC, Vienna, Austria. http://www. R-project.org (last accessed 2010).

Tusher VG, Tibshirani R & Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *PNAS* **98**: 5116–5121.

UniProt (2010) The Universal Protein Resource. The Uniprot Consorcium. http://www.uniprot.org (last accessed 2010).

Wilson WH, Tarran GA, Schroeder D, Cox M, Oke J & Malin G (2002) Isolation of viruses responsible for the demise of an *Emiliania huxleyi* bloom in the English Channel. *Journal of the Marine Biological Association of the UK* 82: 369–377.

Wilson WH, Schroeder DC, Allen MJ et al. (2005) Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. Science 309: 1090–1092.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. *E. huxleyi* and coccolithovirus concentrations in Enclosures 1 to 6 as determined by flow cytometry.

Fig. S2. Hierarchical clustering of all transcription profiles through the use of Non-negative Matrix Factorization (Brunet *et al.*, 2004).

Fig. S3. Non-negative Matrix Factorization (Brunet *et al.*, 2004) of transcription profiles from Enclosures 2 and 3 including four daily time points (6, 12, 18, and 24 h).

Fig. S4. Comparison microarray fluorescence signal vs. qPCR relative expression of host and virus genes, respectively.

Fig. S5. Sequence alignment showing the homologous regions (dots) between the coccolithovirus serine palmitoyltransferase gene sequence and the *E. huxleyi* microarray probe 'B.Read.TopHits.Contig12834_4576_46'. **Table S1.** *E. huxleyi* and coccolithovirus qPCR primer sequences used in this study

Table S2. Sample cluster obtained after different distance

 metrics (Euclidean distance, Manhattan distance, Pearson

Correlation, Pearson Uncentered) and Non-negative Matrix Factorization

Table S3. List of *E. huxleyi* probes that presented too high nucleotide alignment similarity to EhV genomic sequences, and hence were removed from analysis

Table S4. List of EhV and *E. huxleyi* probes that presented significant up-regulation from Cluster 1 to Cluster 2. KOG refers to the NCBI's list of EuKaryotic Orthologous Groups of proteins

Data S1 Supplementary Methods

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.