

Elucidating Genetic Variation and Mechanism of Virus Infection of *Emiliana huxleyi* via Genomic Approaches

Dissertation

zur Erlangung des Akademischen Grades eines

Doktors der Naturwissenschaften

- Dr. rer. Nat. -

im Fachbereich 2 (Biologie/Chemie)

der Universität Bremen

vorgelegt von

Jessica U. Kegel

Bremen, März 2009

Erster Gutachter: Prof. Dr. Dieter Wolf-Gladrow

Zweiter Gutachter: Prof. Dr. Allan Cembella

Tag des öffentlichen Kolloquiums: Universität Bremen, 15. April 2009

Eidesstattliche Erklärung

Hiermit erkläre ich nach § 6 Abs. 5 der Promotionsordnung der Uni Bremen (vom 14. März 2007), dass ich die vorliegende Dissertation

- (1) ohne unerlaubte Hilfe angefertigt habe,
- (2) keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe
und
- (3) die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Jessica Kegel

Dedicated to Mom and Dad

TABLE OF CONTENTS

1	GENERAL INTRODUCTION.....	1
1.1	IMPORTANCE OF MARINE PHYTOPLANKTON.....	1
1.2	THE COCCOLITHOPHORE <i>EMILIANA HUXLEYI</i>	3
1.3	MARINE VIRUSES.....	5
1.4	GENOMICS.....	7
1.4.1	EXPRESSED SEQUENCE TAGS.....	7
1.4.2	MICROARRAYS.....	9
1.4.2.1	GENE EXPRESSION PROFILING.....	10
1.4.2.2	COMPARATIVE GENOMIC.....	10
1.6	OUTLINE OF THE THESIS.....	11
2	PUBLICATIONS.....	12
2.1	LIST OF PUBLICATIONS.....	12
2.2	DECLARATION ON THE CONTRIBUTION OF EACH PUBLICATION	13
	PUBLICATION I.....	14
	PUBLICATION II.....	36
	PUBLICATION III.....	65
3	GENERAL DISCUSSION.....	94
3.1	EFFECTS OF VIRAL INFECTION ON <i>EMILIANA HUXLEYI</i>	94
3.2	BIODIVERSITY IN SEVERAL STRAINS OF <i>E. HUXLEYI</i>	98
3.2.1	VIRUS SUSCEPTIBILITY OF <i>EMILIANA HUXLEYI</i>	102
3.3	PERSPECTIVES OF FUTURE RESEARCH.....	104
4	REFERENCES.....	106
5	SUMMARY.....	121
6	ZUSAMMENFASSUNG.....	124
7	DANKSAGUNG.....	127

1 GENERAL INTRODUCTION

1.1 IMPORTANCE OF MARINE PHYTOPLANKTON

Marine phytoplankton is represented by more than 20,000 microscopic unicellular species of marine photoautotrophs (Falkowski *et al.*, 2003) and is ubiquitous in the world's oceans which cover around 70% of the planet's surface. Its contribution to the global primary production is often disregarded because they account for less than 1% of the global primary producer biomass (Falkowski *et al.*, 2004). However, it is responsible for more than 45% of the Earth's annual net primary production, which is roughly equal to the contribution of terrestrial plants (Field *et al.*, 1998). Grazing, viral attack, programmed cell death, and sinking into the deep ocean balance the phytoplankton production (Falkowski *et al.*, 1998). Consequently, the system is characterized by a high turnover rate and a small standing stock. Phytoplankton forms the base of the marine food chain and its growth is primarily limited by light, nutrients and temperature (Falkowski & Raven, 2007). Winter and autumn storms increase the availability of nutrients and thereby enhancing the growth in particular of bloom formers including diatoms, dinoflagellates and coccolithophores. These blooms can be observed near the coast and/or in upwelling ecosystems (Smetacek, 1999, Smayda, 2000). Diatom-dominated blooms occur mainly in turbulent, low-stratified waters during springtime (Smayda, 1997). In contrast coccolithophore-dominated blooms are found in nitrate-rich but

phosphate-poor, well stratified waters during late spring and early summer (Haidar & Thierstein, 2001). From about 250 coccolithophore species (Winter & Siesser, 1994), the two species *Gephyrocapsa oceanica* and *Emiliana huxleyi* are the only bloom-forming coccolithophores.

The importance of phytoplankton is due to its effect on global climate change through its key role in regulating geochemical cycles such as the global carbon and sulphur cycle. Hereby, marine phytoplankton is responsible for most of the transport of organic matter to the deep ocean and the sediment (Falkowski *et al.*, 2004) thus impacting on atmospheric carbon dioxide (CO₂) (Westbroek *et al.*, 1993). In this context the phytoplankton functional groups including coccolithophores also as well as dinoflagellates, diatoms and cyanobacteria are of major importance (Falkowski *et al.*, 2004). In the process of photosynthesis carbon dioxide is incorporated into particulate organic carbon (POC). Around 45 gigatons of POC are produced annually. More than a third is exported to the ocean interior (Falkowski *et al.*, 1998). A combination of two fundamental processes, the physical and the biological carbon pump, is responsible for the partitioning of CO₂ between atmosphere and ocean. The physical or so-called solubility pump describes the vertical carbon flux due to differences in CO₂ solubility of warm and cold water (Ito & Follows, 2003). The biological pump can be sub-divided into the organic carbon pump and the carbonate pump. The term “organic carbon pump” refers to the photosynthetic production of POC in the surface ocean and its sinking to depth (Volk & Hoffert, 1985). The carbonate pump includes the production of calcium carbonate (termed calcification) by marine organisms (mainly coccolithophores and foraminifera) and its subsequent transport to depth (Rost & Riebesell, 2004). Although both biological carbon pumps remove carbon from the surface ocean, they have, on the production level, opposite effects on the CO₂ concentration of surface waters as explained in the following. Photosynthesis consumes carbon in the form of CO₂, thus reducing the dissolved inorganic carbon (DIC) of the water without affecting total alkalinity (TA). This shifts the carbonate system towards lower CO₂ concentrations and higher pH. Calcification consumes carbon in the form of CO₃²⁻, thus reducing both DIC and TA in a 1:2 ratio. This shifts the carbonate system towards higher CO₂ concentrations and lower pH. Therefore the overall ratio of photosynthesis to calcification determines whether a plankton community increases or decreases CO₂ concentration of sea

surface water. Another important difference of the two biological carbon pumps is the preservation of the exported calcium carbonate that is buried in the sediments and eventually subducted (Van Capellen, 2003).

Coccolithophores also play an important role in other element cycles, e.g. the calcium cycle (De La Rocha & DePaolo, 2000) and the sulphur cycle (Malin *et al.*, 1994). When subject to grazing or during viral infection, *E. huxleyi*, a prolific coccolithophore, produces high amounts of dimethylsulfoniopropionate (DMSP), an important component in the sulphur cycle (Keller, 1989, Malin *et al.*, 1992). DMSP is the precursor of the trace gas dimethyl sulfide (DMS), its emission may contribute to marine cloud formation and climate regulation (Andreae, 1990, Malin *et al.*, 1992, Liss *et al.*, 1997, Stefels *et al.*, 2007).

Besides their importance in biogeochemical and nutrient cycles, marine phytoplankton is also intensively studied due to its contribution to biodiversity, value as a gene pool in times of global biodiversity loss (Pimm *et al.*, 1995), and as a potential source of natural products (Shimizu, 1996).

1.2 THE COCCOLITHOPHORE *EMILIANIA HUXLEYI*

Coccolithophores are unicellular, marine algae belonging to the division of Haptophyta and the class Prymnesiophyceae (Edwardsen *et al.*, 2000). One prominent feature of the coccolithophores is the ability to produce an exoskeleton formed of minute calcite plates, the coccoliths. The life cycle of coccolithophores consists of a diploid stage characterized by the production of so called heterococcoliths and a haploid stage, in which usually so called holococcoliths are produced (Billard, 1994). Heterococcoliths and holococcoliths have very different morphologies, which makes it easy to tell the two life cycle stages apart. On rare occasions combination cells are found, i.e. cells displaying both types of coccoliths (Geisen *et al.*, 2002). The first fossil record of coccolithophores can be traced back to the Late Triassic (~225 Ma) (Bown *et al.*, 2004). They first became abundant in the Jurassic (~150 Ma) (Morse & Mackenzie, 1990) and reached their greatest abundance in the Late Cretaceous (~80 Ma), becoming a major factor in the global carbonate cycle (Hay, 2004). Nowadays,

they are considered to be, besides foraminifera, the most productive calcifying organism on earth (Baumann *et al.*, 2004).

Emiliania huxleyi ranks among the ten most important coccolithophores in terms of calcite export (Baumann *et al.*, 2004). *E. huxleyi* has evolved from the older genus *Gephyrocapsa* 268.000 years ago (Thierstein *et al.*, 1977) and became dominant around 70.000 years ago. It is now the most abundant coccolithophore in the marine system except in polar waters (Brand, 1994, Winter *et al.*, 1994, Paasche, 2002, Marsh, 2003). *E. huxleyi* has spherical cells of 3-10 μm in diameter and is therefore one of the smaller coccolithophores. *E. huxleyi* is an atypical coccolithophore. Firstly, it does not produce holococcoliths. Secondly, its complex life cycle includes the coccolith-bearing non-motile ('C-cell') stage alternating with naked non-motile ('N-cell') and scale-bearing flagellated ('S-cell') stages (Klaveness, 1972). The C-cell and N-cell stages are typically diploid whereas the motile S-cell stage is haploid (Green *et al.*, 1996). Both diploid and haploid phases are capable of independent asexual reproduction. A third feature that makes *E. huxleyi* an atypical coccolithophore is the fact that it forms immense coastal and open ocean blooms. The blooms occur from sub-polar to tropical latitudes (Balch *et al.*, 1992, Brown & Yoder, 1993) and can cover more than 50.000 km^2 (Holligan *et al.*, 1993, Winter *et al.*, 1994, Sukhanova & Flint, 1998). These blooms can be detected via satellite imagery due to the reflection properties of the coccoliths (Holligan *et al.*, 1983, Balch *et al.*, 1991).

The size and intensity of these blooms makes *E. huxleyi* important for nutrient and CO_2 cycling and biogenic sulphur production (in the form of DMS) in the marine environment. Consequently it is a key species for current studies on global biogeochemical cycles and climate modelling (Westbroek *et al.*, 1994). Since viral infection is an important termination factor of the vast blooms of *E. huxleyi* (Bratbak *et al.*, 1993, Jacquet *et al.*, 2002), it is of particular interest to understand this host-virus interaction.

1.3 MARINE VIRUSES

Viruses are small, non-cellular particles composed of either DNA or RNA (double- or single-stranded) embedded in a protein coat known as capsid that may be surrounded by an envelope. They are metabolically inert and do not respire, move or grow. Outside their host cells, viruses exist as virus particles also named virions. The virion has the function to protect the genome of a virus and to deliver it into a host cell for replication and packaging into new virions. Since the viral genome is typically small, the question arises how viruses can encode all the information needed for their reproduction. They utilize host cell proteins, overlapping viral genes, and multifunctional viral proteins. Once introduced in a host cell, viruses utilize the host machinery in order to enhance the efficiency of the replication process. Therefore, the intracellular environment of their host is modified, which might include production of a new membranous structure, reduced expression of cell genes or enhancement of a cell process (e.g. transcription and translation).

In the oceans, viruses are the most abundant biological entities (Fuhrman, 1999, Suttle, 2000, Wommack & Colwell, 2000) and infect all organisms from bacteria to whales (Suttle, 2005). It is estimated that the marine environment contains 10^{30} viruses (Suttle, 2007). Most of the viruses described to date are species-specific: they infect a single host species and sometimes even a single strain within a species. Due to their immobility, viruses depend on passive movement to contact a suitable host (Brussaard, 2004, Weinbauer, 2004). Consequently the encounter rate between a virus and a host is directly affected by their relative abundances, respectively.

Several studies have shown the infection of a wide range of aquatic algae (Van Etten *et al.*, 1991, Van Etten & Meints, 2003) including bloom-forming marine phytoplankton (Jacobsen *et al.*, 1996, Sandaa *et al.*, 2001) like *Phaeocystis globosa* (Brussaard *et al.*, 2005), *Heterosigma akashiwo* (Nagasaki *et al.*, 1994a, Nagasaki *et al.*, 1994b, Nagasaki & Yamaguchi, 1997) and *Emiliania huxleyi* (Bratbak *et al.*, 1993). Through their various infection potential viruses are playing important roles in nutrient and biogeochemical cycling (Fuhrman, 1999, Wilhelm & Suttle, 1999), and influence structure and diversity of microbial and phytoplankton communities (Fuhrman, 1999, Wommack & Colwell, 2000). During the last two decades it became evident that viruses affect the biogeochemical cycles through the cell lysis of the

hosts. Viral lysis affects the efficiency of the biological pump by increasing or decreasing the relative amount of carbon in exported production (Suttle, 2007). This so called “viral shunt” moves material from heterotrophic and phototrophic microorganisms into particulate organic matter (POM) and dissolved organic matter (DOM) (Middelboe *et al.*, 1996, Gobler *et al.*, 1997, Middelboe & Lyck, 2002, Middelboe & Jorgensen, 2006), which is mostly converted to CO₂ by respiration and photodegradation (Fuhrman, 1999, Wilhelm & Suttle, 1999, Weinbauer, 2004, Suttle, 2005). Furthermore, the accelerated sinking rates of virus-infected cells increase the transport of organic molecules from the photic zone to the deep ocean (Lawrence *et al.*, 2002, Lawrence & Suttle, 2004). In addition, viral lysis of phytoplankton may also be an important source of DMSP and therefore influencing the global climate (Charlson *et al.*, 1987). Laboratory studies demonstrated the increase of DMSP in the media during viral lysis of *Phaeocystis pouchetii*, *Micromonas pusilla*, and *Emiliana huxleyi* (Hill *et al.*, 1998, Malin *et al.*, 1998, Wilson *et al.*, 2002).

Because of its importance for the global biogeochemical cycles, the bloom-former *E. huxleyi* is the most studied eukaryotic phytoplankton host-virus system to date (Bidle *et al.*, 2007). A range of different viruses specific for *E. huxleyi* (EhV) were first isolated from blooms in the English Channel and off Bergen, Norway (Castberg *et al.*, 2002, Wilson *et al.*, 2002). These viruses were further analyzed for their phylogeny (Schroeder *et al.*, 2002, Allen *et al.*, 2006c), ecological succession in mesocosm experiment (Schroeder *et al.*, 2003, Martínéz *et al.*, 2007), and genome structure of *Emiliana huxleyi* virus 86 (EhV-86) (Wilson *et al.*, 2005, Allen *et al.*, 2006b, Allen *et al.*, 2007). Characterization of their sequences revealed that *E. huxleyi* specific viruses are double-stranded DNA-containing lytic viruses with large genomes, approximately 410 kb in size (Wilson *et al.*, 2005) which belong to the *Coccolithoviruses* (Schroeder *et al.*, 2002) a genus within the family *Phycodnaviridae* (Van Etten *et al.*, 2002).

A recent study shows, that *E. huxleyi* can escape viral attack by switching its life cycle from a diploid to haploid (Frada *et al.*, 2008). This motile, noncalcifying haploid stage is impervious to viruses and therefore resistant to EhVs that infect and lyse the diploid calcifying phase. Besides this, *E. huxleyi* strains which are virus resistant show higher DMSP-lyase activity than strains that are susceptible to virus infection (Schroeder *et al.*, 2002). So far, nothing is known about the genes being

expressed in *E. huxleyi* during the viral infection. It was, therefore, one of the objectives of this study to elucidate genes involved in the host-virus interaction, to speculate on the infection mechanism.

1.4 GENOMICS

Considering the vital role of coccolithophores in the marine carbon cycle it is of great interest to get a deeper insight into their genetic variability, population biology and ecophysiological properties in order to be able to evaluate the influence of global environmental change.

Over the last decade genome-based technologies have contributed significantly to the understanding of algal ecology and evolution (Grossman, 2005). Marine ecological genomics is the study of the genomes of organisms, combining molecular biology with computing sciences, statistics and management, with the goal to understand the relationship between ecosystem processes and biodiversity (Lawton, 1994, van Straalen & Roelofs, 2006, Dupont *et al.*, 2007). The most popular genome technologies in this area include (1) whole genome sequencing of key organisms such as the red alga *Cyanidioschizon merolae* (Matsuzaki *et al.*, 2004, Nozaki *et al.*, 2007), the green alga *Chlamydomonas reinhardtii* (Merchant *et al.*, 2007), and the diatom *Phaeodactylum tricornutum* (Bowler *et al.*, 2008), (2) barcoding, (3) expressed sequence tag (EST) collections and (4) microarrays. The last two approaches are the most common methods used to date and will be discussed in detail.

1.4.1 EXPRESSED SEQUENCE TAGS

Expressed sequence tags (ESTs) are short sub-sequences produced from complementary DNA (cDNA) libraries with 200-800 bp length. cDNA libraries are constructed from mRNA isolated under specific conditions at a particular time. They are cost-effective and provide a robust sequence resource that can be exploited for

gene discovery, expression profiling, evolutionary and taxonomy studies, microarray design, genome annotation and comparative genomics (Rafalski *et al.*, 1998, Schmitt *et al.*, 1999, Rudd, 2003, Dupont *et al.*, 2007). Normalization procedures have been used to reduce the abundance of highly expressed genes thereby enriching the sampling of rarer transcripts (Soares *et al.*, 1994). More recently, subtraction techniques have been used to construct libraries depleted of clones already subjected to EST sampling (Bonaldo *et al.*, 1996).

Up to now, ESTs had helped in the discovering of genes in organisms for which genomic data are unavailable (Hackett *et al.*, 2005, Lidie *et al.*, 2005). Furthermore, ESTs identified novel genes involved in e.g. salinity or temperature stress response (Kore-eda *et al.*, 2004, Reusch *et al.*, 2008).

Several studies focusing on fungal- and viral-infected plants using ESTs were reported (Hsiang & Goodwin, 2003, Ventelon-Debout *et al.*, 2003, Goodwin *et al.*, 2004, Jantasuriyarat *et al.*, 2005) indicating the usefulness of ESTs for the discovery of genes involved in host-pathogen interaction. Once a virus has attached to an *E. huxleyi* cell, there follows a complex propagation strategy that is controlled largely by the virus, however, it is a life cycle which is still unknown. The question arises what kind of genes are involved in the viral lysis of *E. huxleyi* blooms. Which genes in *E. huxleyi* are expressed during the host-virus interaction related to the response to infection and possible resistance? Furthermore, what kind of viral genes are expressed during infection and how is this related to virulence and the ability to grow and reproduce in the host? Different scenarios are possible, from complete shutdown of the host on infection through to a predominantly host controlled process. These extremes are improbable and it is likely the truth lies somewhere in-between. Determination of the complete host response to infection is clearly beyond the scope of a single project. It was therefore one of the objectives of this thesis to gain more information about the response of *E. huxleyi* to viral infection and the interaction with EhV-86 during viral infection by taking advantage of EST libraries. The construction of ESTs from *E. huxleyi* at different stages of viral infection could thus be an effective means for expression analysis of virus infected cultures for which the viral genome is known. With the draft genome of *E. huxleyi* CCMP1516 and the complete genome of EhV-86 available, it is possible to determine the precise number of ESTs from both, the host and virus, in all EST libraries. Results of this study (**publication I**

and **II**) provide insights into the infection mechanisms of the virus EhV-86 in *E. huxleyi*.

1.4.2 MICROARRAYS

First applied in the mid 1990s (Schena *et al.*, 1995), microarray technology has become a routine and essential tool for gene expression profiling (Leung & Cavalieri, 2003). The advantage of microarray technology is the ability to study thousands of genes in a single experiment (Li *et al.*, 2002). Therefore DNA microarrays have a wide range of applications including gene expression profiling, gene discovery, detection of single nucleotide polymorphism (SNPs), comparative genomic hybridization (CGH), disease diagnostic, pharmacogenomics, and toxicology research (DeRisi *et al.*, 1997, Ye *et al.*, 2001, Li *et al.*, 2002).

DNA microarrays are based on a minimized, but high throughput form of a dot blot, and consist of an arrayed series of thousands of DNA fragments, immobilized onto a surface, such as coated glass slide or membrane (Ye *et al.*, 2001, Gentry *et al.*, 2006). They can be made either by the mechanical spotting of presynthesized DNA products like cDNAs of up to several hundred base pairs (DeRisi *et al.*, 1998, Eisen & Brown, 1999) or by the *in situ* synthesis of 60-mer oligonucleotides (Lipshutz *et al.*, 1999, Ye *et al.*, 2001, Li *et al.*, 2002). Following the production of a DNA microarray, microarray experiments are performed by sample isolation and preparation, hybridization and data analysis. Depending on the application either DNA from e.g. two different strains or RNA from e.g. an infected and uninfected sample is used as starting material. Prepared samples are labelled with two different fluorescent markers and co-hybridized to a microarray under high-stringency conditions. After hybridization the signal intensities are detected via fluorescent excitation by a microarray scanner.

1.4.2.1 GENE EXPRESSION PROFILING

Comparing ESTs and microarray analysis revealed that the combination of both methods is advantageous in estimating the expression level of gene transcripts (Munoz *et al.*, 2004). It has been shown that important functions in an organism are indicated by highly expressed genes (Dupont *et al.*, 2007). Transcripts of low abundance may not occur at all in an EST library but the absence is not necessarily evidence for not being expressed under a different condition (Bouck & Vision, 2007). Hence, the EST approach for simultaneous discovery and identification of host and viral genes involved in viral infection were complemented with microarray analysis to enable the detection of even more subtle changes in gene expression (**publication II**).

1.4.2.2 COMPARATIVE GENOMIC

Comparative genomic hybridization (CGH) is currently one of the most powerful microarray techniques to compare DNA copy numbers between the genomes of e.g. closely related taxa, such as sub-species and strains. CGH is used to compare the genes present, absent or divergent in the genomes of interest. Therefore, two different fluorescently labelled genomic DNA samples are compared by co-hybridization. Polymorphisms and insertions can be detected as a reduction or elevation of a hybridization signal (Gibson, 2002). Whole genome comparisons of different strains of various microbes indicate that polymorphism for gene content is not uncommon (Riley & Serres, 2000, Pearson *et al.*, 2003, Watanabe *et al.*, 2004), suggesting genetic adaptations to different ecological niches. Previous studies have reported different genome sizes among different morphotypes of *E. huxleyi* from different geographical regions (Medlin *et al.*, 1996, Iglesias-Rodriguez *et al.*, 2002). Results indicate the presence of different ecotypes of *E. huxleyi* potentially with differences in genome organization in response to environmental conditions or to potential threats, such as viral infections. For that reason, CGH were applied to estimate genetic variation at the genomic level of 16 *E. huxleyi* strains from different

geographic origin with the aim to identify genes correlated to virus susceptibility and morphology (**publication III**).

1.6 OUTLINE OF THE THESIS

In this thesis, molecular techniques were applied to improve our understanding of mechanisms and interactions of *Emiliania huxleyi* that take place during viral infection. Furthermore, genomic differences in several strains of *E. huxleyi* from different geographic origin were investigated to determine key genes in respect to viral susceptibility and morphology. Identified genes will be a starting point for further investigations using molecular approaches. The results of this thesis will improve our understanding of *E. huxleyi* and Coccolithoviruses as vital components of the global carbon cycle.

Publication I reports the first construction of EST libraries of *Emiliania huxleyi* throughout a viral infection process and shows the possibility to determine differentially expressed genes using cDNA libraries within this approach.

Publication II examines the effect of viral infection on *E. huxleyi* through ESTs in a larger way and made it possible to speculate on mechanism of the host-virus interaction that occur in the host cell during viral infection in both partners. Furthermore, a comparison of two different methods to determine differentially expressed genes is provided.

Publication III investigates the biodiversity of 16 *E. huxleyi* strains from different geographic origin with regard to virus susceptibility and morphology using genomic DNA for comparative genomic hybridization on oligoarrays.

In a concluding discussion main results of this thesis are summarized and discussed with respect to bloom dynamics, virus-host interactions, and genetic diversity. Finally, perspectives are given for future research.

2 PUBLICATIONS

2.1 LIST OF PUBLICATIONS

This doctoral thesis is based on the following publications:

- I. Jessica Kegel, Michael J. Allen, Katja Metfies, William H. Wilson, Dieter Wolf-Gladrow and Klaus Valentin. 2007. Pilot study of an EST approach of the coccolithophorid *Emiliana huxleyi* during a virus infection. *Gene* 406: 209-216.
- II. Jessica U. Kegel, Mark Blaxter, Michael J. Allen, Katja Metfies, William H. Wilson and Klaus Valentin. Transcriptional host-virus interaction of *Emiliana huxleyi* (Haptophyte) and EhV-86 deduced from combined analysis of expressed sequence tag and microarrays. Submitted to *European Journal of Phycology*.
- III. Jessica U. Kegel, Uwe John, Stephan Frickenhaus, Katja Metfies and Klaus Valentin. Comparative functional genomics of virus-susceptible and virus-resistant *E. huxleyi* strains. To be submitted to *BMC Genomics*.

2.2 DECLARATION ON THE CONTRIBUTION OF EACH PUBLICATION

Publication I

The experiments were planned together with Klaus Valentin and performed by the candidate. The data were interpreted and the manuscript written by the candidate in discussion with the co-authors.

Publication II

The experiments were planned together with Klaus Valentin and performed by the candidate. The data were interpreted and the manuscript written by the candidate in discussion with the co-authors.

Publication III

The experiments were planned together with Uwe John and performed by the candidate. S. Frickenhaus provided help in bioinformatics analysis and produced the R-figures. The data were interpreted and the manuscript written by the candidate in discussion with the co-authors.

PUBLICATION I

**Pilot study of an EST approach of the
coccolithophorid
Emiliana huxleyi during a virus infection**

JESSICA KEGEL^{1*}, MICHAEL J. ALLEN², KATJA METFIES¹, WILLIAM H.
WILSON³, DIETER WOLF-GLADROW¹ AND KLAUS VALENTIN¹

¹*Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12,
D-27570 Bremerhaven, Germany*

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

⁵*Bigelow Laboratory for Ocean Sciences, 180 McKown Point Road, POB 475, Maine
04575, USA*

Key words: *Emiliana huxleyi*, coccolithophore, EhV86, EST sequencing

Abbreviations: bp, base pair; cDNA, complementary to RNA; cfu, colony forming units; EST, expressed sequence tag; *e*-value, expectation value; fcp, fucoxanthin/chlorophyll binding protein; HSP70, heat shock protein 70; mRNA, messenger RNA

Abstract

Blooms of the coccolithophorid *Emiliana huxleyi* can be infected by viruses, which can lead to bloom-termination. This pilot study used an expressed sequence tag (EST) approach to get a first view of gene expression changes that occur during viral infection of *E. huxleyi*. cDNA libraries were constructed from uninfected cultures and 6, 12, and 24 h after infection with *E. huxleyi*-specific virus 86 (EhV-86). From each library 60 – 90 ESTs were randomly selected and annotated manually with PhyloGena. Viral genes were identified using BLAST-Search of the known viral genome. The data of this study show, that 6 h after viral infection the algal transcriptome changed significantly although few viral transcripts were present. At this point, changes mainly concerned transcripts related to photosynthesis and protein metabolism. However, after 24 h viral transcripts were most abundant. Viral transcripts found at this stage of viral infection encode proteins involved in protein degradation, nucleic acid degradation, transcription and replication.

1. Introduction

Emiliania huxleyi (Lohmann, 1902, Hay et al., 1967) is the most abundant coccolithophore and an important member of the marine phytoplankton. It is well known for its immense coastal and open ocean blooms ranging from sub-polar to tropical latitudes (Balch et al., 1992, Brown and Yoder 1994) that can cover 10,000 km² or more (Holligan et al. 1993, Winter et al. 1994). *E. huxleyi* is also regarded as a major sink for calcium carbonate carbonate in the ocean (Eide 1990, Samtleben and Bickert 1990, Baumann et al. 2004). Due to the reflection of their coccoliths blooms can be observed by satellites (Holligan et al. 1983, Balch et al. 1991).

The abundance and wide distribution of *E. huxleyi* and its production of calcium carbonate coccoliths and dimethylsulfide (DMS) make it an important species with respect to sediment formation and to ocean climate and natural acid rain (Charlson et al. 1987, Westbroek et al. 1993, Malin et al. 1994). Furthermore it is a key species for current studies on global biogeochemical cycles (Westbroek et al. 1994).

Viral lysis is thought to be one of the main causes for the termination of *E. huxleyi* blooms. Several studies have investigated the role of viruses in controlling the bloom-forming of *E. huxleyi* (Bratbak et al. 1993, 1995, 1996, Brussaard et al. 1996, Castberg et al. 2001, Jaquet et al. 2002, Wilson et al. 1998, 2002a, 2002b). It became evident from these investigations that viruses are intrinsically linked to the decline of *E. huxleyi* blooms.

Viruses are the most abundant biological agents in marine aquatic environments (Bergh et al. 1989, Suttle 2000, Wommack and Colwell 2000) and it is likely that most microbial organisms can be infected with a particular viruses. Therefore they play important roles in nutrient (Wilhelm and Suttle 1999) and biogeochemical (Fuhrmann 1999) cycling, and influence structure and diversity of microbial and phytoplankton communities (Fuhrmann 1999, Wommack and Colwell 2000). Viruses have also been observed to infect a wide range of aquatic algae (van Etten et al. 1991, van Etten and Meints 1999), including bloom-forming marine phytoplankton (Nagasaki et al. 1994a, 1994b, Jacobsen et al. 1996, Nagasaki and Yamaguchi 1997, Sandaa et al. 2001).

A range of different viruses that infect *E. huxleyi* (EhV) was isolated from the English Channel and off the coast of Bergen, Norway (Castberg et al. 2002, Wilson et al. 2002b) and were analysed for their phylogeny (Schroeder et al. 2002), ecological successions in mesocosm experiment (Schroeder et al. 2003) and genome structure (EhV-86) (Allen et al. 2006, 2007, Wilson et al. 2005). Characterization of their sequences revealed that the *E. huxleyi* viruses are large double-stranded DNA viruses with genomes approximately 410 kbp in size and that they belong to a new virus genus termed Coccolithovirus based on the phylogeny of their DNA polymerase gene (Schroeder et al., 2002). Coccolithoviruses belong to the *Phycodnaviridae* (Wilson et al. 2005), a diverse family of large icosahedral viruses that infect marine or freshwater eukaryotic algae, they all contain dsDNA genomes ranging from 180 – 560 kb (Van Etten et al. 2002).

Expressed sequence tag (EST) analysis is a useful tool to study gene expression and to discover novel genes. ESTs are small pieces of DNA sequences that are generated by sequencing and based on the creation of a cDNA library. By statistical evaluation of the frequency of the sequences for specific genes it is possible to develop an expression profile at different environmental conditions for genes of different cDNA libraries. Thereby it is possible to investigate the up- and down regulation of genes (Schmitt et al. 1999) or to compare the gene expression under different conditions (Rafalski et al. 1998). The establishment of ESTs from *E. huxleyi* at different stages of viral infection could thus be an effective means for expression analysis of virus infected cultures for which the viral genome is known. ESTs specify the type and rate of viral and host transcripts at a particular time. As a result of that, it is possible to hypothesise on mechanisms of host-virus interaction that occur in the host cell during viral infection in both partners.

The aim of this work was to provide a functionally annotated preliminary set of ESTs from *E. huxleyi* expressed before and during a virus infection in order to determine differentially expressed genes. The results of this study made it possible to estimate the proportional abundance of viral transcripts in relation to the whole transcriptome of the host cell during progression of the infection.

2. Material & Methods

2.1 Strains and Growth Conditions

Cultures of *E. huxleyi* CCMP1516 were grown in f/2 medium (Guillard 1975) at 15°C with a 16:8 light-dark illumination (150 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$). Because of the availability of the complete genome sequence of the virus EhV-86, this species was used for the infection of *E. huxleyi*. Exponentially growing cultures (approx. 1.2×10^6 cells/ml) were inoculated with EhV-86 lysate (2 ml per litre of culture, approx. 1×10^6 pfu/ml) in the middle of the dark phase.

2.2 RNA extraction from uninfected cells

Cultures (50 ml) were harvested on 1.2 μm filters (Millipore), transferred into a cryogenic vial (Nalgene), immediately frozen in liquid nitrogen and stored at -80°C until use for analysis.

RNA of uninfected cultures was isolated at five different time points in series. This approach was chosen because RNA from infected cells (see below) was taken at different times after infection and as such at different phases of the cell cycle. Starting time was at the late exponential phase (approx. 1×10^6 cells/ml) and the last point was at the beginning of the stationary phase (approx. 3.3×10^6 cells/ml) (Fig. 1). Total RNA was isolated with the RNeasy Plant Mini Kit (Qiagen) according to the manual including one more washing step with buffer RW1 and buffer RPE and a DNase digestion subsequent to the original protocol. Afterwards mRNA was isolated with the Oligotex mRNA Mini Kit (Qiagen). Before library construction the mRNA from five different time points were pooled (Fig. 1) and precipitated with 0.5 volumes of LiCl overnight at -20°C. Following centrifugation (14,000 rpm, 60 min, 4°C), the supernatant was discarded and the pellet was washed three times with 100 μl ice-cold 70% ethanol. The pellet was air dried, resuspended in 10 μl DEPC-treated water and used for library construction.

2.3 RNA Extraction from Infected Cells

After 6, 12 and 24 hours of virus-infection cultures (250 ml) were filtered through 0.45 μm filters (Millipore). The filtrate was discarded and the filters transferred to clean petri dishes. Cells from each filter were resuspended in 2 ml of 1 x Phosphate buffered saline (PBS), centrifuged (20,000 g, 5 min), resuspended (by

vortexing) in 2 ml RNAlater (Qiagen) and stored at -20°C until ready for processing. RNA extraction was performed using an RNeasy Midi Kit (Qiagen). Samples were centrifuged (20,000 g, 2 min) and the pellet resuspended in 2 ml RLT buffer (+ 20 µl 2-mercaptoethanol). Following vigorous vortexing (1 min, in 5 second bursts), the samples were spun (20,000 g, 5 min) and the supernatant transferred to a 15 ml Falcon tube containing 2 ml 70% ethanol. Following vigorous mixing the samples were applied to a Qiagen MidiPrep column, centrifuged (3,200 g, 5 min) and the flow-through discarded. Columns were washed twice with 2.5 ml RPE buffer (3,200 g, 5 min) and transferred to a new Falcon tube. RNase free water (250 µl) was added, the samples incubated (room temperature, 1 min) and the RNA eluted by centrifugation (3,200 g, 5 min).

To precipitate RNA solutions, 0.5 volumes of 7.5 M NH₄Ac and 2 volumes of 100% ethanol was added and the samples incubated at -80°C overnight. Following centrifugation (20,000 g, 30 min), the supernatant was discarded and the pellet washed twice with 0.5 ml 70% ethanol (20,000 g, 30 min). The pellet was air dried, resuspended in 50 µl RNase free water and stored at -80°C.

2.4 Library construction

The uninfected library stems from pooled RNA collected throughout the growth curve and from 2 independent cultures (Fig. 1). Less than 1 µg of mRNA were used to establish a cDNA Library with the CloneMiner™ cDNA Library Construction Kit (Invitrogen) according to the manual. First strand synthesis was performed using a Biotin-*attB2*-Oligo(dT) primer with the following sequence: Biotin-GGCGGCCGCACAACTTTGTACAAGAAAGTTGGGT(T)₁₉ and SuperScript™ II Reverse Transcriptase. Subsequent to the second-strand synthesis using *Escherichia coli* DNA polymerase, blunt end products were ligated with an *attB1* adapter through T4 DNA Ligase. After size fractionation (> 500 bp) cDNA-fragments were cloned into the cloning vector pDONR 222 with the BP Clonase enzyme. Plasmids were used to transform ElectroMax DH10B competent cells via electroporation, and random clones were picked for quality control analysis.

2.5 Construction of cDNA libraries after virus-infection

Total RNA (4 µg each) was used for the construction of cDNA libraries prepared by vertis Biotechnologie AG (Munich, Germany). From each total RNA poly A+ RNA was prepared. With the poly A+ RNA first-strand cDNA synthesis was performed using an oligo(dT)-linker primer and M-MLV-RNase H⁻ reverse transcriptase. Synthesis of the second strand was carried out with a random linker primer and Klenow exo- DNA-polymerase. The resulting cDNAs were then amplified with 17 (6 and 12 hours p.i.) and 16 cycles (24h) of LA-PCR (Barnes 1994).

For cloning, the cDNAs were subjected to a limited exonuclease treatment to generate 5' overhangs at both ends of the cDNAs. After size fractionation on an agarose gel and elution of cDNAs > 0,5 kb the cDNAs were directionally ligated into the *Eco* RI and *Bam* HI sites of the plasmid vector pBS II sk+. Ligations were electroporated into T1 Phage resistant TransforMaxTM EC100TM-T1^R (Epicentre) electro-competent cells. After transformation, glycerol was added to a final concentration of 15% (v/v).

2.6 EST sequencing

Plasmid DNA was isolated using a standard alkaline lysis procedure, and unidirectional sequencing was accomplished using the M13 HEDGE forward primer (TGA GCG GAT AAC AAT TTC ACA CAG) for the uninfected library and the M13 forward primer (TGT AAA ACG ACG GCC AGT) for the infected libraries, providing sequence from the 5' end of cDNA clones. Sequencing was performed according to the principle of Sanger (1977) using BigDye terminator chemistry from Applied Biosystems.

2.7 Data analysis

For identifying the function of ESTs, sequences were analysed by the program PhyloGena (Hahnekamp *et al.* 2007) on the basis of the SwissProt database. This is a system for an automated phylogenetic annotation of ESTs, genes and genomes. It automatically constructs phylogenetic trees on a per ORF basis and allows annotation on the basis of the function of the neighbouring sequences in the tree. This method is more reliable than simply assuming the function of the

“best hit” in a BLAST search. As a threshold for a significant similarity we used an e-value of 10^{-7} .

Viral transcripts were identified by BLAST searches of all ESTs against the EhV-86 genome. A corresponding analysis verified the origin of all other transcripts from the alga by similar searches against the *E. huxleyi* draft genome, i.e. the trace files of the sequencing runs (<http://www.jgi.doe.gov/>).

3. Results

The aim of this work was to create an initial dataset of the gene-expression that occurs during infection of the *E. huxleyi* strain CCMP1516 with the virus EhV-86. Therefore we constructed an EST-library from non infected cells and compared it with three EST-libraries after virus-infection (6, 12 and 24 hours).

3.1 Identification and annotation of the ESTs

After elimination of vector- and other problematic sequences, high-quality ESTs with an average length of 520 nucleotides were used for the identification and annotation of the sequences. All ESTs were tested for their origin from the alga and from the virus by BLAST searches against the EhV-86 genome and the *E. huxleyi* draft genome. There was no EST present which did not match either database. All genes were manually annotated. The annotated ESTs had a BLAST e-value smaller than or equal to 10^{-7} . Identification was based on phylogenetic analysis of all ORFs using PhyloGena. In all libraries the largest fraction of ESTs, both host and virus, were those of unknown function or those not producing any significant hit in BLAST searches.

It was possible to annotate 78 sequences from the uninfected library. A proportion of 17% (13 ESTs) of the sequences encoded for fcp or fcp-like proteins involved in light harvesting and 50% (39 ESTs) were proteins with unknown function. Furthermore, proteins were identified involved in photosynthesis, the cell cycle, transcription and protein metabolism (Tab. 1).

From the library 6 hrs after virus infection 67 sequences were annotated, from which 64% (43 ESTs) were encoded for unknown proteins. Most of the identified proteins are ribosomal proteins and elongation factors, which are responsible for protein synthesis (Tab. 2). In this context, expression of HSP70 and s-adenosylhomocysteinase are conspicuous because they are an indication for

stress. In contrast to the uninfected library, no fcp-like proteins were identified in the EST library generated 6 hrs post viral infection. However, three genes of unknown function from the virus EhV-86 were identified.

Twelve hours post infection 82 sequences could be annotated. In this EST-library, 58.5% (48 ESTs) of the ESTs had no significant match against the SwissProt database in BLAST searches and were classified as proteins of unknown function. However, 12% of the 82 ESTs were ribosomal proteins which are involved in protein metabolism (Tab. 3). We could identify 5% (4 ESTs) of viral genes of unknown function. The expression of stress proteins like rotamase, RAS-like protein and HSP70 is also conspicuous. Fcp-like proteins were also missing.

After 24 hours the host viral assemblage transcriptome is dominated by the virus. 80 ESTs of 91 annotated sequences were viral genes. Two of the 11 host genes could be identified as 60S ribosomal protein L8 and GDP-D-mannose 4, 6-dehydratase. Only 10% of the viral genes had significant matches in SwissProt (Tab. 4). All of the viral proteins of unknown function had a length between 800 and 1100 bp.

By comparing the ESTs divided into different functional categories, we found that before virus infection photosynthesis-related genes dominate in the host, but after 6 hrs post infection their abundance decreases rapidly in the libraries (Fig. 2). Furthermore, after 6 hrs virus infection stress-induced host genes were identified. 24 hrs post infection the viral genes clearly dominate the library. In all libraries a high number of genes of unknown function were found.

4. Discussion

Our data show that it is possible to determine expression profiles throughout a viral infection process using EST libraries. In our case we were lucky to have available genome sequences for both partners, i.e. the virus and the host. Thus we were able to determine the exact number of ESTs from both partners in the libraries. Our data show, that a large proportion of genes active in both partners are of unknown function. These genes would have escaped the analysis given we had focussed on known genes and using classical approaches. It is one of the strengths of the EST approach that also those genes are found that are unknown. .

4.1 Possible infection mechanism of the virus EhV-86 by *E. huxleyi*

These conclusions are based on approx. 320 annotated EST-sequences, but there is still a trend recognizable. Less than 6 hrs post infection the virus seems to change the expression pattern of *E. huxleyi* significantly. Our data indicate down-regulation of photosynthesis genes, which is also known in the infection cycle of *Paramecium bursaria Chlorella Virus-1* (Seaton *et al.* 1995). Viral infection took place in the middle of the dark phase and the first sample was taken after 6 hr, i.e. early in the light phase. During this phase one would expect photosynthesis genes to be upregulated. In *E. huxleyi* downregulation of photosynthesis genes took place on the benefit of up-regulation of genes related to gene expression and protein synthesis possibly to enhance the expression of viral proteins. The up-regulation of transcription and translation genes of the host could be induced by the virus to facilitate transcription and translation of its genes. Only a few viral transcripts seem to be required for that, because in the 6h EST library only a minority of 4.5% of the sequences were found to have a viral origin.

It is remarkable that after 6 and 12 hrs virus infections only a few viral transcripts appear, but that, nevertheless, drastic changes in host gene expression patterns were induced. After 24 hrs the transcription of *E. huxleyi* seemingly comes to a standstill. Now the share of viral transcripts in the EST library has reached 90%.

Viral RNA polymerase genes are activated, which may preferably transcribe viral genes or host genes required by the virus. Furthermore, endonuclease and clp-protease appear which could be responsible for the degradation of host DNA and host proteins. The virus has taken over transcription at the latest 24 hrs after infection, approximately the doubling time of the alga in culture; this could mean that the infection is correlated to the cell cycle of *E. huxleyi*. The reason could be the need of particular host proteins, e.g. for transcription, replication or translation, all of which occur in an organised form during host cell cycle.

4.2 Outlook

More sequences are required to be able to draw more reliable conclusions about changes of the host expression pattern during viral infection. In the ongoing

project we have established EST libraries comprising several thousand sequences for *E. huxleyi*. The next step currently under way is to use the sequence information to establish genome arrays for alga and virus and to screen genomes of various *E. huxleyi* and virus strains for genomic differences, and subsequently for transcriptome differences.

Twelve hours after infection, the virus affects the host expression pattern of *E. huxleyi* significantly, but after 24 hrs the effect was drastic. During this twelve hour period there seems to occur a lot of change in the expression of the host and the virus. To get further insight into the infection cycle, it would be of particular interest to construct and analyze a cDNA-library after 14 -18 hours of virus-infection, because the virus, described by Castberg *et al.* (2002) has a latency of 12-14 hours.

Acknowledgements

This work was supported by Marine Genomics Europe (EU Contract n° 505403). We would like to thank Matthew Hall for technical assistance and Linda Medlin for support.

References

- Allen MJ, Schroeder DC, Holden MTG & WH Wilson 2006 Evolutionary History of the Coccolithoviridae. *Mol. Biol. Evol.* 23: 86-92
- Allen MJ, Martinez-Martinez J, Schroeder DC, Somerfield PJ & WH Wilson 2007. Use of microarrays to assess viral diversity: from genotype to phenotype. *Environ. Microbiol.* 9: 971-982
- Balch WM, Holligan PM, Ackleson SG & KJ Voss 1991 Biological and optical properties of mesoscale coccolithophore blooms in the Gulf of Maine. *Limnol. Oceanogr.* 36: 629-643
- Balch WM, Holligan PM & AK Kilpatrick 1992 Calcification, photosynthesis and growth of the bloom-forming coccolithophore, *Emiliania huxleyi*. *Cont. Shelf. Res.* 12:1353-1374
- Barnes WM 1994 PCR amplification of up to 35-kb DNA with high fidelity and high yield from λ bacteriophage templates. *Proc. Natl. Acad. Sci. USA* 91: 2216-2220

- Baumann K-H, Böckel B & M Frenz 2004 Coccolith contribution to South Atlantic carbonate sedimentation. *In: Thierstein HR & JR Young [Eds.] Coccolithophores*, Springer-Verlag, Berlin Heidelberg, pp 367-402
- Bergh O, Borsheim KY, Bratbak G & M Heldal 1989 High abundance of viruses found in aquatic environments. *Nature* 340: 467-468
- Bratbak G, Egge JK & M Heldal 1993 Viral mortality of the marine alga *Emiliana huxleyi* (Haptophyceae) and termination of algal blooms. *Mar. Ecol. Prog. Ser.* 93: 39-48
- Bratbak G, Levasseur M, Michaud S, Cantin G, Fernández E, Heimdal BR & M Heldal 1995 Viral activity in relation to *Emiliana huxleyi* blooms: a possible mechanism of DMSP release? *Mar. Ecol. Prog. Ser.* 128: 133-142
- Bratbak G, Wilson W & M Heldal 1996 Viral control of *Emiliana huxleyi* blooms? *J. Mar. Syst.* 9: 75-81
- Brown CW & JA Yoder 1994 Blooms of *Emiliana huxleyi* (Prymnesiophyceae) in surface waters of the Nova Scotian shelf and the Grand Bank. *J. Plankton. Res.* 15: 1429-1438
- Brussaard CPD, Kempers RS, Kop AJ, Riegman R & M Heldal 1996 Virus like particles in a summer bloom of *Emiliana huxleyi* in the North Sea. *Aquat. Microb. Ecol.* 10: 105-113
- Castberg T, Larsen A, Sandaa RA, Brussaard CPD, Egge JK, Heldal M, Thyrraug R, van Hannen EJ & G Bratbak 2001 Microbial population dynamics and diversity during a bloom of the marine coccolithophorid *Emiliana huxleyi* (Haptophyta). *Mar. Ecol. Prog. Ser.* 221: 39-46
- Castberg T, Thyrraug R, Larsen A, Sandaa RA, Heldal M, Van Etten JL & G Bratbak 2002 Isolation and characterization of a virus that infects *Emiliana huxleyi* (Haptophyta). *J. Phycol.* 38: 767-774
- Charlson RJ, Lovelock JE, Andreae Mo & SG Warren 1987 Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature (Lond.)* 326: 655-661
- Eide LD 1990 Distribution of coccolithophorids in surface sediments in the Norwegian-Greenland Sea. *Mar. Micropaleontol.* 16: 65-75

- Fuhrmann JA 1999 Marine viruses and their biogeochemical and ecological effects. *Nature (Lond.)* 399: 541-548
- Guillard RRL 1975 Culture of phytoplankton for feeding marine invertebrates. In Smith WL & MH Chanley [Eds.] *Culture of Marine Invertebrate Animals*. Plenum Press, New York, pp 29-60
- Hanekamp K, Bohnebeck U, Beszteri B & K Valentin 2007 PhyloGena – a system for automated phylogenetic annotation of (algal) expressed sequence tags (EST), genes and genomes. *Bioinformatics (in press)*
- Hay WW, Mohler HP, Roth PH, Schmidt RR & JE Boudreaux 1967 Calcareous nanoplankton zonation of the Gulf Coast and Caribbean-Antillean area and transatlantic correlation. *Trans. Gulf Coastal Assoc. Geol. Soc.* 17: 428-480
- Hedge P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Hughes JE, Snesrud E, Lee N & J Quackenbush 2000 A Concise Guid to cDNA Microarray Analysis. *BioTechniques* 29: 548-562
- Holligan PM, Viollier M, Harbour DS, Camus P & M Champagne-Philippe 1983 Satellite and ship studies of coccolithophore production along a continental shelf edge. *Nature (Lond.)* 304: 339-342
- Holligan PM, Fernández E, Aiken J, Balch WM, Burkill PH, Finch M, Groom SB, Malin G, Muller K, Purdie DA, Robinson C, Trees CC, Turner SM & P van der Wal 1993 A biogeochemical study of the coccolithophore *Emiliana huxleyi* in the north Atlantic. *Glob. Biogeochem. Cycles* 7: 879-900
- Jacobsen A, Bratbak G & M Heldal 1996 Isolation and characterization of a virus infecting *Phaeocystis pouchetii* (Prymnesiophyceae). *J. Phycol.* 32: 923-927
- Jaquet S, Heldal M, Iglesias-Rodriguez D, Larsen A, Wilson W & G Bratbak 2002 Flow cytometric analysis of an *Emiliana huxleyi* bloom terminated by viral infection. *Aquat. Microb. Ecol.* 27: 111-124
- Lohmann H 1902 Die Coccolithophoridae, eine Monographie der Coccolithen bildenden Flagellaten. Zugleich ein Beitrag zur Kenntnis des Mittelmeerauftriebs. *Arch. Protistenk.* 1: 89-165

- Malin G, Liss PS & SM Turner 1994 Dimethyl sulphide: production and atmospheric consequences. In Green JC& BSC Leadbeater [Eds.] *The haptophyte algae*. Systematics Association Special Volume No. 51. Clarendon Press, Oxford, pp 303-320
- Nagasaki K & M Yamaguchi 1997 Isolation of a virus infectious to the harmful bloom causing microalga *Heterosigma akashiwo* (Raphidophyceae). *Aquat. Microb. Ecol.* 13: 135-140
- Nagasaki K, Ando M, Imai I, Itakura S & Y Ishida 1994a Virus-like particles in *Heterosigma akashiwo* (Raphidophyceae): a possible red tide disintegration mechanism. *Mar. Biol.* 119: 307-312
- Nagasaki K, Ando M, Itakura S, Imai I & Y Ishida 1994b Viral mortality in the final stage of *Heterosigma akashiwo* (Raphidophyceae) red tide. *J. Plankton Res.* 16:1595-1599
- Rafalski JA, Hanafey M, Miao GH, Ching A, Lee JM, Dolan M & S Tingey 1998 New experimental and computational approaches to the analysis of gene expression. *Acta Biochim. Pol.* 45: 929-934
- Samtleben C & T Bickert 1990 Coccoliths in sediment traps from the Norwegian Sea. *Mar. Micropaleontol* 16: 39-64.
- Sandaa RA, Heldal M, Castber T, Thyraug R & G Bratbak 2001 Isolation and characterization of two viruses with large genome size infecting *Chrysochromulina ericina* (Prymnesiophyceae) and *Pyramimonas orientalis* (Prasinophyceae). *Virology* 290: 272-280
- Sanger F, Nickeln S & AR Coulson 1977 DNA-sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74: 5463-5467
- Schmitt AO, Specht T, Beckmann G, Dahl E, Pilarsky CP, Hinzmann B & A Rosenthal 1999 Exhaustive mining of EST libraries for genes differentially expressed in normal and tumour tissues. *Nucl. Acids Res.* 27: 4251-4260
- Schroeder DC, Oke J, Malin G & WH Wilson 2002 Coccolithovirus (*Phycodnaviridae*): characterization of a new large dsDNA algal virus that infects *Emiliania huxleyi*. *Arch. Virol.* 147: 1685-1698

- Schroeder DC, Oke J, Hall M, Malin G and WH Wilson 2003 Virus Succession Observed during an *Emiliana huxleyi* Bloom. *Appl. Environ. Microbiol.* 69: 2484-2490
- Seaton GGR, Lee K & J Rohozinski 1995 Photosynthetic Shutdown in *Chlorella* NC64A Associated with the Infection Cycle of *Paramecium bursaria* *Chlorella* Virus-1. *Plant Physiol.* 108: 1431-1438
- Suttle CA 2000 Ecological, evolutionary, and geochemical consequences of viral infection of cyanobacteria and eukaryotic algae. *In: Hurst CJ [Ed.] Viral Ecology.* Academic Press, New York, pp 247-296
- Van Etten JL, Lane LC & RH Meints 1991 Viruses and virus-like particles of eukaryotic algae. *Microbiol. Rev.* 55: 586-620
- Van Etten JL & RH Meints 1999 Giant viruses infecting algae. *Annu. Rev. Microbiol.* 53: 447-494
- Van Etten JL, Graves MV, Müller DG, Boland W & N Delaroque 2002 *Phycodnaviridae* – large DNA algal viruses. *Arch. Virol.* 147: 1479-1516
- Westbroek P, Brown CW, Van Bleijswijk J, Brownlee C, Jan Brummer G, Conte M, Egge J, Fernandez E, Jordan R, Knappertsbusch M, Stefels J, Veldhuis M, Van der Wal P & J Young 1993 A model system approach to climate forcing. The example of *Emiliana huxleyi*. *Global Planet Change* 8: 27-46
- Westbroek P, Van Hinte J, Brummer G-R, Veldhuis M, Brownlee C, Green Jc, Harris R & B Heimdal 1994 *Emiliana huxleyi* as a key to biosphere-geosphere interaction. *In: Green JC & B Leadbeater [Eds.] The Haptophyte Algae.* Clarendon Press, Oxford, pp 321-334
- Wilhelm SW & CA Suttle 1999 Viruses and nutrients cycles in the sea – viruses play critical roles in the structure and function of aquatic food webs. *Bioscience* 49: 781-788
- Wilson WH, Turner S & NH Mann 1998 Population Dynamics of Phytoplankton and Viruses in a Phosphate-limited Mesocosm and their Effect on DMSP and DMS Production. *Estuarine, Coastal and Shelf Science* 46: 49-59
- Wilson WH, Tarran GA & MV Zubkov 2002a Virus dynamics in a coccolithophore dominated bloom in the north Sea. *Deep Sea Res.* 49: 2951-2963

- Wilson WH, Tarran GA, Schroeder DC, Cox M, Oke J & G Malin 2002b Isolation of viruses responsible for the demise of an *Emiliana huxleyi* bloom in the English Channel. *J. Mar. Biol. Ass. U.K.* 82: 369-377
- Wilson WH, Schroeder DC, Allen M, Holden MTG, Parkhill J, Barrell BG, Churcher C, Hamlin N, Mungall K, Norbertczak H, Quail MA, Price C, Rabinowitsch E, Walker D, Craigon M, Roy D & P Ghazal 2005 Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. *Science* 309: 1090-1092
- Winter A, Jordan RW, & PH Roth 1994 Biogeography of living coccolithophores in ocean waters. In Winter A & WG Siesser [Eds.] *Coccolithophores*. Cambridge University Press, Cambridge, UK, pp 161-177
- Wommack KE & RR Colwell 2000 Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* 64: 69-114

Figure legends

Figure 1. Growth curve of an uninfected *Emiliana huxleyi* culture and the time points for the pooled mRNA (demonstrated by black arrows). The grey and dashed arrow shows the equivalent position where the virus was added to the culture (approx. 1.2×10^6 cells/ml) in the separate infection experiment.

Figure 2. EST expression profile of *Emiliana huxleyi* before and during virus infection divided into different functional categories based on their putative function in per cent.

Tables**Table 1.** EST assembly results of *E. huxleyi* before virus infection

Sequence name	Frequency	<i>e</i> value	Putative function
Unknown protein	39	-	-
Fcp-like	10	10 ⁻²⁰	Photosynthesis
Fucoxanthin Chlorophyll a/c binding protein (fcp)	3	10 ⁻¹⁰	Photosynthesis
Light harvesting complex (LHC)	3	10 ⁻¹⁰	Photosynthesis
Cyclin dependent kinase regulatory subunit, putative	2	10 ⁻³⁰	Signal transduction, cell communication, cell cycle
Alpha-glucosidase	2	10 ⁻⁵¹	Galactose metabolism
Beta-hydrocyacyl-ACP dehydratase	2	10 ⁻⁵⁴	Fatty acid synthesis
Hypothetical conserved protein,	2	10 ⁻¹⁸	--
Chromosome condensation regulator-like protein	1	10 ⁻⁸	Regulation of the condensation of the chromosomes
Eukaryotic translation initiation factor-like protein	2	10 ⁻²⁴	Initiation of translation
Trehalose-6-phosphate-synthase	2	10 ⁻²⁹	Starch and saccharose metabolism
Ankyrin related protein	1	10 ⁻¹⁰	Involved in binding of spectrin at the plasma membrane
Ferredoxin-NADP ⁺ -reductase	1	10 ⁻⁷⁸	Photosynthesis
Glutarredoxin, mitochondrial	1	10 ⁻³⁰	Glutathione metabolism
Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	1	10 ⁻¹¹⁷	Glycolysis, gluconeogenesis
N5'-nucleotidase, cytosolic	1	10 ⁻²⁰	Transcription
NADH-dehydrogenase	1	10 ⁻¹²	Respiration, fatty acid oxidation
Protein kinase	1	10 ⁻¹⁸	Signal transduction, phosphorylation
Putative membrane protein	1	10 ⁻¹²	--
Ribosomal protein L3	1	10 ⁻³⁷	Protein metabolism
RNA helicase	1	10 ⁻⁴⁸	Transcription

Table 2. EST assembly results of *E. huxleyi* after 6 hours virus infection.

Protein	Frequency	e value	Putative function
40S ribosomal protein S11	2	2.0e-42	Protein metabolism
40S ribosomal protein S13	2	9.0e-51	Protein metabolism
60S ribosomal protein L7	2	5.0e-31	Protein metabolism
60S ribosomal protein L25	1	2.0e-43	Protein metabolism
60S ribosomal protein L27	1	2.0e-42	Protein metabolism
Adenylate kinase	1	1.0e-42	Glycolysis; phosphorylation
Elongation factor 1-alpha-like protein (EF1 α)	1	1.0e-66	Protein metabolism
EF-Tu like protein	1	8.0e-67	Protein metabolism
Caltractin-like protein	1	1.0e-21	Calcium binding protein
Eukaryotic translation initiation factor 3 subunit 7	1	1.0e-47	Protein metabolism
Chloroplast ferredoxin-NADP ⁺ -reductase	1	9.0e-93	Photosynthesis
GAPDH cytosolic	1	1.0e-34	Glycolysis, gluconeogenese
weakly similar to ubiquitin	1	7.0e-23	Amino acid catabolism
weakly similar to phosphoribulokinase	1	4.0e-30	Calcin cycle
HSP70-like protein	1	8.0e-38	Stress induced protein
Phosphoribosylaminoimidazole- succinocarboxamide synthase	1	7.0e-32	Transcription
S-adenosylhomocysteinase	2	8.0e-51	Amino acid catabolism
unknown protein	43	--	--

Table 3. EST assembly results of *E. huxleyi* after 12 hours virus infection.

Protein	Frequency	<i>e</i> value	Putative function
40S ribosomal protein S19	2	4.0e-21	Protein metabolism
40S ribosomal protein S19-like	1	6.0e-18	Protein metabolism
60S ribosomal protein L10	2	1.0e-51	Protein metabolism
60S ribosomal protein L8	1	2.0e-66	Protein metabolism
78kd glucose regulating protein	1	1.0e-38	Stress-induced protein
Adenine phosphoribosyltransferase-like protein	1	1.0e-15	Transcription
ADP-ribosylation factor (Arf)	1	1.0e-40	Stress-induced protein
Elongation factor 1-alpha-like protein (EF1 α)	1	6.0e-64	Protein metabolism
Calmodulin	2	1.0e-74	Signaltransduction; calcium sensor
Calmodulin-like protein	2	8.0e-41	Signaltransduction; calcium sensor
Similar to cytosolic GAPDH	1	4.0e-20	Glycolysis, gluconeogenesis
Similar to nuclear GAPDH	1	2.0e-42	Glycolysis, gluconeogenesis
Glycin cleavage protein, mitochondrial	1	2.0e-37	Amino acid catabolism
HSP70-like protein	1	4.0e-29	Stress-induced protein
nuclear protein with unknown function, transcription regulator?	1	3.0e-32	--
Methionin-aminopeptidase	1	3.0e-44	Protein metabolism
Peptidyl-prolyl cis-trans isomerise (rotamase)	1	1.0e-70	Stress-induced protein
Similar to phosphoribulokinase	1	1.0e-49	Calvin cycle
Mitochondrial processing protease α subunit	1	6.0e-19	Protein metabolism; protein catabolism
Ras-like protein	1	2.0e-21	Stress-induced protein
Ras-related protein, GTPase	1	3.0e-21	Stress-induced protein
S-adenosyl-L homocysteine hydrolase-like protein	1	4.0e-51	Amino acid catabolism
weakly conserved protein with ATP binding site, similar to elongation factor	1	2.0e-26	Protein metabolism
Stress-activated protein kinase	1	1.0e-47	Stress-induced protein
Ubiquitin	1	2.0e-71	Amino acid catabolism
UDP-N-acetylglucosamine pyrophosphatase	1	5.0e-13	UDP-N-acetylgalactosamine biosynthetic process
unknown protein	48	--	--

Table 4. Identified viral genes after 24 hours infection and their putative function.

Protein	<i>e</i> value	Putative function
Clp protease (Casein lytic protein)	1.0e-12	Degradation of host proteins
Clp-like protein	4.0e-11	Degradation of host proteins
DNA topoisomerase II	2.0e-43	Packaging of viral DNA; expression of viral genes
DNA dependent RNA Polymerase I	3.0e-06	Priority of expression of viral genes
DNA dependent RNA Polymerase II	2.0e-62	Priority of expression of viral genes
Non histone chromosomal protein	7.0e-09	Packaging of viral DNA
Flap endonuclease	1.0e-48	Degradation of host nucleic acid
Deoxyuridin 5'-triphosphate nucleotidhydrolase	6.4e-42	Nucleic acid metabolism
Protein kinase	1.0e-12	Regulation

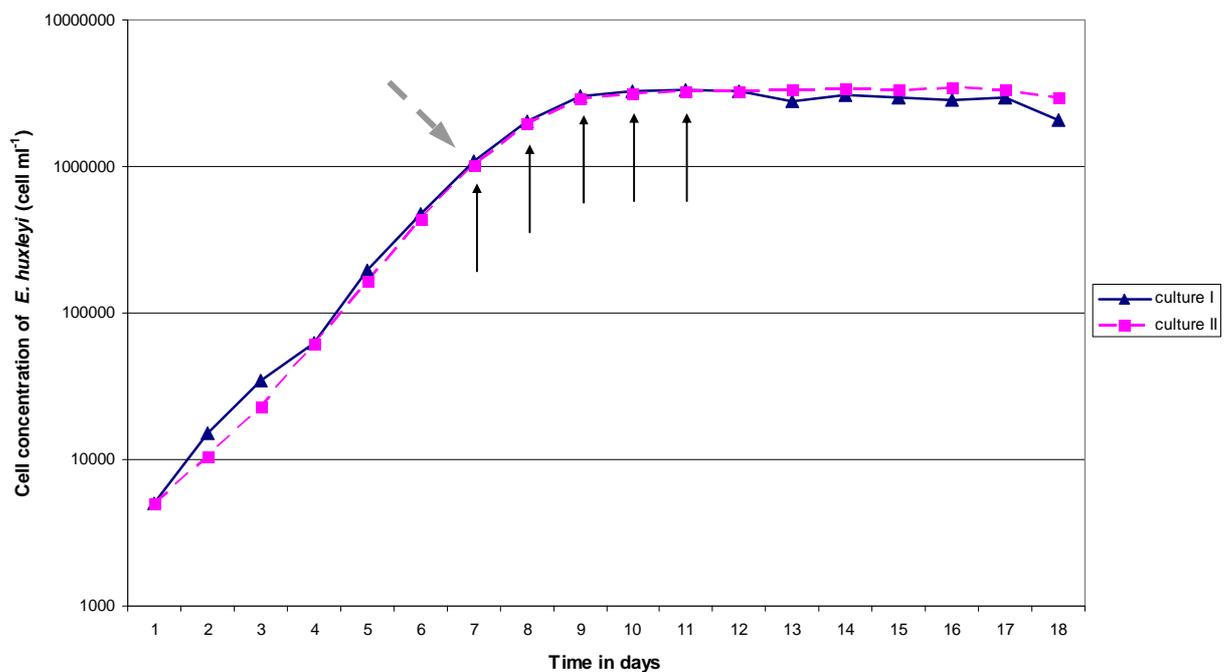
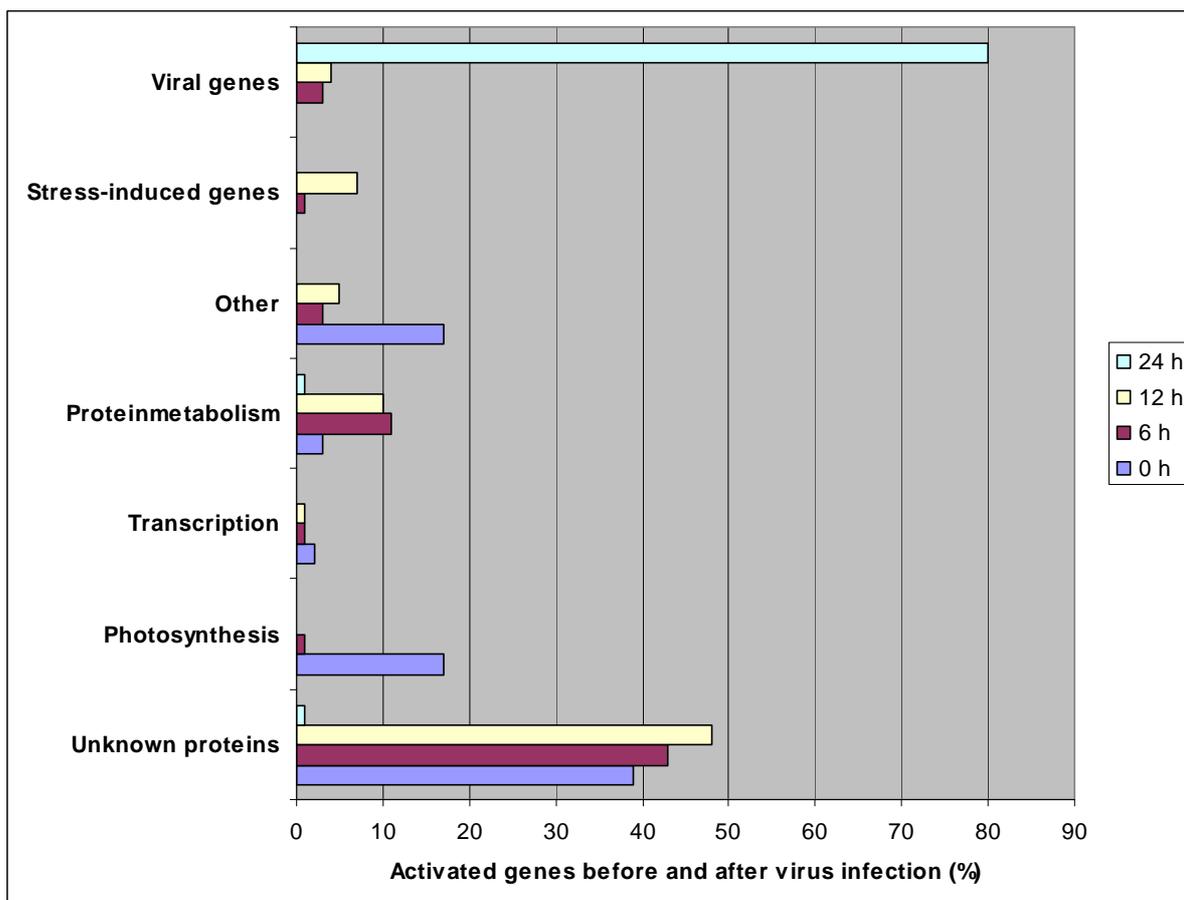
Figure 1

Figure 2

PUBLICATION II

Transcriptional host-virus interaction of *Emiliana huxleyi* (Haptophyceae) and EhV-86 deduced from combined analysis of expressed sequence tags and microarrays

Jessica U. Kegel^{1*}, Mark Blaxter², Michael J. Allen³, Katja Metfies⁴, William H. Wilson⁵ and Klaus Valentin¹

¹ Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany

² School of Biological Sciences, Ashworth Laboratories, King's Buildings, West Mains Rd, University of Edinburgh, Edinburgh EH9 3JT, UK

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

⁴ GKSS Research Center, Max Planck Strasse 1, 21502 Geesthacht, Germany

⁵ Bigelow Laboratory for Ocean Sciences, 180 McKown Point Road, POB 475, Maine 04575, USA

* Corresponding author

Running title: Gene expression of virus infected *E. huxleyi*

Abbreviations: bp, base pairs; EST, expressed sequence tag; e value, expectation value

Abstract

The cosmopolitan coccolithophore *Emiliana huxleyi* forms frequent massive blooms and thus is important for the global climate and carbon cycle. Lytic viral infection of this alga leads to termination of blooms and therefore influences the global climate. To understand the host-virus interaction of *E. huxleyi*, an expressed sequence tag (EST) approach was used to determine changes in gene expression during viral infection. Three cDNA libraries generated 6, 12 and 24 h post viral infection were compared to a library from an uninfected culture by sequencing, clustering and manual annotation of 1100-1500 ESTs per library. To verify the gene expression results of the ESTs we used two-colour oligonucleotide microarrays. A total of 4480 ESTs were assembled into 1871 clusters of which 223 clusters are of viral origin. Microarray expression analysis indicated that 231 out of 565 oligonucleotides of *E. huxleyi* changed their expression level in at least one time point in response to viral infection. Results suggest that viral infection affects the following processes: photosynthesis, transcription and translation, carbohydrate and lipid metabolism (particularly glycolysis), metabolism, and signal transduction. Results of this study provide insights into the infection mechanisms of the virus EhV-86 in *E. huxleyi*.

Key words: *Emiliana huxleyi*, host-virus interaction, EhV-86, EST, microarray, photosynthesis

Introduction

The marine coccolithophore *Emiliana huxleyi* is highly abundant and widely distributed in all marine systems except in polar waters (Paasche, 2001, Marsh, 2003). *E. huxleyi* is capable of forming immense coastal and open ocean blooms. The blooms occur from sub-polar to tropical latitudes (Balch *et al.*, 1992, Brown & Yoder, 1993) and can cover more than 50.000 km² (Balch *et al.*, 1991, Holligan *et al.*, 1993, Sukhanova & Flint, 1998). These blooms can be detected via satellite imagery due to reflection of its calcium carbonate coccoliths (Holligan *et al.*, 1983). Because of this bloom formation activity and its distribution and high abundance beyond blooms, *E. huxleyi* influences global climate by affecting the inorganic carbon system of seawater (Buitenhuis *et al.*, 1996, Buitenhuis *et al.*, 2001) and by organic carbon pumping. Due to the massive calcifying activity, *E. huxleyi* is considered to be the world's major producer of calcite and one of the largest single carbonate sinks in oceanic carbonate cycling (Eide, 1990, Samtleben & Bickert, 1990, Baumann *et al.*, 2004). It also plays an important role in the global sulphur cycle (Malin *et al.*, 1994). When subject to grazing or during viral infection, *E. huxleyi* produces dimethylsulfoniopropionate (DMSP), the precursor of the trace gas dimethyl sulphide (DMS), which is linked to marine cloud formation and climate regulation (Liss *et al.*, 1997, Stefels *et al.*, 2007).

Viral infection is an important termination factor of the blooms of *E. huxleyi* (Bratbak *et al.*, 1993, Jacquet *et al.*, 2002). As major bloom terminators double-stranded DNA-containing, lytic viruses (Schroeder *et al.*, 2002) have been isolated (Castberg *et al.*, 2002, Wilson *et al.*, 2002) and described as Coccolithoviruses (Wilson *et al.*, 2005). The interaction of *E. huxleyi* and these viruses specific to *E. huxleyi* are one of the best studied eukaryotic phytoplankton host-virus systems to date (Bidle *et al.*, 2007). To understand the molecular basis of viral lysis of *E. huxleyi* blooms, we need to learn which genes in *E. huxleyi* are expressed during the host-virus interaction and how this could be related to the response to infection and possible resistance. Furthermore, we need to know the virus genes which are expressed during infection in order to identify genes related to virulence and the ability to grow and reproduce in the host. Expressed Sequence Tags (ESTs) are one way to analyse the genes being expressed under specific conditions. They are cost-effective and provide a robust sequence resource that can be exploited for gene discovery, microarray design, genome annotation and comparative genomics (Rudd,

2003). ESTs can be successfully combined with information coming from completely sequenced genomes, in which case it is easy to identify which genome an EST belongs to. Several studies focussing on fungal- and viral-infected plants using ESTs were reported (Hsiang & Goodwin, 2003, Ventelon-Debout *et al.*, 2003, Goodwin *et al.*, 2004, Jantasuriyarat *et al.*, 2005). To our knowledge, this technique has only been applied to marine alga-virus interaction in our pilot study (Kegel *et al.*, 2007). With the draft genome of *E. huxleyi* CCMP1516 and the complete genome of *Emiliana huxleyi* virus 86 (EhV-86) available, we took advantage of an EST approach combined with genome sequence information to gain detailed insights into the host-virus interaction of this important microalgae.

In this study we investigated EST-libraries from *E. huxleyi* during virus infection and compared these to a library from a healthy culture. Furthermore, changes in gene expression levels, assessed through two-colour oligonucleotide microarrays, were compared with the gene redundancy in the EST libraries.

Material & Methods

Source of host and virus libraries

Four cDNA libraries from *Emiliana huxleyi* CCMP1516 constructed by Kegel *et al.* (2007) were compared: one pooled library (T0) from an uninfected culture and three libraries from a culture 6, 12 and 24 hours post infection with EhV-86 (in this publication referred to as T6, T12 and T24). For this study, between 1100 and 1500 clones were randomly selected from each library.

EST sequencing

Plasmid DNA was prepared from recombinant clones using a standard alkaline lysis procedure. Unidirectional sequencing was done using the M13 HEDGE (Hegde *et al.*, 2000) forward primer (TGA GCG GAT AAC AAT TTC ACA CAG) for the uninfected library and the M13 forward primer (TGT AAA ACG ACG GCC AGT) for the infected libraries, providing sequence from the 5' end of cDNA clones.

Sanger sequencing was performed by Max-Planck-Institute for Molecular Genetics of Berlin, Germany.

Sequence analysis and functional annotation

All ESTs were compared against the whole genome of EhV-86 (Wilson *et al.*, 2005) and the draft genome of *E. huxleyi* (JGI, <http://www.jgi.doe.gov/>) using BLASTn (Altschul *et al.*, 1990). The analysis was done using PartiGene (Parkinson *et al.*, 2004). Every EST had more than 86% identity to either *E. huxleyi* or EhV-86 genomic sequence.

ESTs were automatically clustered and annotated using the PartiGene software. Sequences were pre-processed concerning vector contamination, low complexity and repeat regions like the poly(A) tail. Low quality and very short ESTs (< 100 bp) were discarded. Subsequently, ESTs were clustered on the basis of sequence similarity into groups that putatively derive from the same gene and assembled into consensus sequences (Parkinson *et al.*, 2004). Following the clustering and assembling, these consensus sequences were annotated by BLAST searches against the UniProt database (<http://www.uniprot.org>), the whole genome of EhV-86, and the publicly accessible ESTs of *E. huxleyi*. Finally, an HTML summary table of all consensus sequences was produced, providing the number and list of ESTs for each cluster along with associated BLAST annotation. These are available at

http://www.nematodes.org/NeglectedGenomes/EMILIANIA/Emiliana_huxleyi.html.

For interpretation of the consensus sequences of the virus we used the best hit of the automated annotation against the whole genome of Eh-V86.

The consensus host sequences were manually analysed with PhyloGena (Hanekamp *et al.*, 2007) using the SwissProt database (<http://www.expasy.org/sprot/>) and were compared with Pfam (Bateman *et al.*, 2002) and the NCBI non-redundant protein database (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>, NCBI-nr). As a threshold for significant similarity we used an *e* value of 10^{-5} . Consensus sequences with BLAST results above that threshold were assigned as hypothetical proteins. Consensus sequences which showed significant similarity only to proteins of unknown function were annotated as conserved hypothetical proteins. Sequences which did not produce any significant hit in BLAST searches were classified as no significant hit.

Annotated EST clusters of the host were classified into the functional categories of KEGG using the application tool KAAS (KEGG Automatic Annotation Server: <http://www.genome.jp/kegg/kaas>, (Moriya *et al.*, 2007) with the SBH (single-

directional best hit) method, whereas EST clusters without any KEGG hit were manually assigned.

DNA Microarray Design and Validation

Based on the consensus sequence information generated in this project, microarrays were designed with Agilent's eArray online application tool version 5.0. Only EST clusters of *E. huxleyi* containing a significant BLAST hit were used to design the oligonucleotides for the microarray. The resulting 60-mer oligonucleotide probes and standard controls from Agilent were printed on glass slides by Agilent using an Ink Jet-based printing method (Agilent's SurePrint technology). Four arrays were printed on each 1 x 3-inch glass slide.

For the labelling of total RNA an Agilent Low RNA Input Fluorescent Linear Amplification kit (Agilent Technologies) was used according to the manufacturer's recommendations. Total RNA (500 ng) from the uninfected culture was labelled by fluorescent complimentary RNA (cRNA) synthesis with Cy3-CTP. The RNA of the infected culture (500 ng) was labelled with Cy5-CTP. A superscript III reverse transcriptase (0.5 μ L, Invitrogen) was added to the labelling reaction to increase the yield of longer transcripts. Therefore an additional incubation step (30 min, 50°C) was inserted prior to heat-inactivation of the enzymes. Amplified cRNAs were purified with RNeasy mini spin columns (Qiagen) and quantified by UV spectroscopy. Hybridizations were performed with 825 ng of each labelled cRNA at 65°C for 17 h within a hybridization chamber (Agilent Technologies). After hybridization, the microarrays were washed according to the manufacturer's instructions (Wash with Stabilization and Drying Solution, Agilent Technologies). Microarrays were imaged using an Agilent microarray scanner in eXtended Dynamic Range function mode and a scan resolution of 5 μ m. Signal intensities were detected by Feature Extraction software version 9.5 (Agilent Technologies). Differential expression was analysed using the MeV version 4.0.01 software package from TIGR (Saeed *et al.*, 2003). Significance Analysis of Microarray (SAM) analysis described by Tusher *et al.* (2001) was performed using following parameters: (i) One-Class test; (ii) 1000 permutations; (iii) Tusher *et al.* method with calculated *q*-values; (iv) K-nearest neighbor, 10 neighbors. Values were not restricted by setting any delta value and the resulting significant regulated genes with a *q*-value < 1% and a fold-change

of ≥ 2 were analysed. These data were then clustered by a hierarchical clustering algorithm using average linkage for the heuristic criterion and Euclidean distance as similarity metric (Eisen *et al.*, 1998). Original data files for all arrays were uploaded in MIAME format for expression arrays at GEO (<http://www.ncbi.nlm.nih.gov/geo/>; accession numbers XXXX-XXXX).

Results

Identification and annotation of the ESTs

A total of 4480 EST sequences (GenBank accession No. XXX-XXX) with good quality (phred 20, min. 100 bp length, no ribosomal RNA) and an average length of 644 bp were obtained from the uninfected library (T0) and 6, 12 and 24 hours post viral infection (in this publication referred to as T6, T12 and T24), clustered and assembled into 1871 different clusters. Of those, 1319 clusters comprised of unique clones and 552 clusters consisted of up to 135 ESTs. Out of the total 1871 clusters, a subset of 223 clusters (12 %) were significantly similar to the EhV-86 genome, of which only 44 clusters (20%) showed similarity to known proteins (Tab. S1). In addition, 569 (35%) of the remaining 1648 host clusters had significant similarity to known proteins or hypothetical proteins.

The uninfected library (T0) consisted of 1056 ESTs of good quality and could be divided into 554 different clusters. Less than half of these clusters (257) derived from 416 ESTs showed significant similarity to known proteins or hypothetical proteins. At T6, only 38 out of 985 ESTs were of viral origin and clustered into 26 viral and 435 host clusters. Twelve hours after viral infection, 1378 ESTs of good quality grouped into 626 clusters out of those, 19 clusters derived from 32 ESTs were of viral origin. At T24, a total of 946 out of 1061 ESTs showed significant similarity to the EhV-86 genome and clustered into 218 different clusters. Only 115 ESTs showed significant similarity to the *E. huxleyi* genome and resulted in 84 clusters.

ESTs were grouped into 11 gene functional categories derived from KEGG: (1) cellular processes; (2) signal transduction; (3) transport; (4) hypothetical; (5) transcription and translation; (6) carbohydrate and lipid metabolism; (7) metabolism; (8) folding, sorting and degradation; (9) chaperones and folding catalysts; (10) genetic information processing; and (11) photosynthesis. ESTs which did not fit in any of the categories were assigned to an additional category “other”.

In all libraries, the majority of ESTs (both host and viral in origin) showed no significant similarity to sequences in the public databases.

The E. huxleyi transcriptome

The most abundant transcripts prior to infection were related to photosynthesis (17.2%), metabolism (4.8%) and transcription and translation (3.4%). After 6, 12 and 24 hours the percentage of photosynthesis-related ESTs decreased to 2.5, 3.3, and 0.5%, respectively. In contrast, ESTs related to transcription and translation increased to 25.0% and 12.4% at T6 and T12 (Fig.1). In the carbohydrate and lipid metabolism category, the percentage of genes increased from 2.1% in the uninfected library to 3.4% and 6.4% in the infected libraries at 6 and 12 h after viral infection. Furthermore, the percentage of genes related to signal transduction increased from 0.5% in the uninfected library to 4.1% and 5.0% in the infected libraries at 6 and 12 h after viral infection, respectively. For example, at T6 and T12 the number of ESTs with similarity to ubiquitin was eight times higher than in the uninfected library. Furthermore, the number of ESTs with similarity to calmodulin, eukaryotic translation and elongation factors increased ten to twenty fold, whereas chloroplast light harvesting proteins and fucoxanthin chlorophyll a/c-binding protein, were 16 to 38 times decreased at T6 and T12 compared to T0 (Tab. 1). At T24, the host viral assemblage transcriptome was dominated by the viral transcripts (89.2%). Hence it was not possible to detect any obvious changes in the host transcripts.

The EhV-86 transcriptome

During viral infection 38 different viral genes were identified deriving from 44 clusters containing 164 ESTs with putative functions (Tab. 2). At T6 two copies of a transcript for a DNA-dependent RNA polymerase II largest subunit (EHV064), and one for a sialidase (EHV455) and a protease (EHV349) were found. Furthermore, at T12 a single transcript for a fatty acid desaturase (EHV061), a HNH endonuclease family protein (EHV093), and a proliferating cell nuclear antigen (EHV440) were detected. At T24 37 of the identified transcripts were for genes including ehv455, ehv349, ehv061, ehv093 and ehv440. The most frequent transcripts (19 ESTs) were for an ATP-dependent protease proteolytic subunit (EHV133), a fatty acid desaturase (15 ESTs, EHV415), and a deoxyuridine 5'-triphosphate nucleotidhydrolase (12

ESTs, EHV397). In addition, we could identify transcripts for three different DNA-directed RNA polymerase subunits (EHV108, EHV434 and EHV399), two endonucleases (EHV041 and EHV018), and two protein kinases (EHV451 and EHV402).

Identification of changes in gene expression using microarrays

A 60-mer oligonucleotide microarray based on the consensus sequences was used to investigate gene expression patterns of *E. huxleyi* during viral infection. To study the expression profile at each time point, technical triplicates (duplicate for 24 h) were performed comparing the RNA after 6, 12, and 24 hours with the uninfected culture as reference. Only features with absolute differential expression ≥ 2 -fold were used for trending. The ESTs of *E. huxleyi* from the four libraries were assigned to 569 clusters showing significant similarity to UniProt. The application tool eArray designed 565 oligonucleotides out of the 569 clusters. Due to technical reasons it was not possible to design oligonucleotides for the four remaining clusters.

A total of 231 (41%) out of the 565 gene transcripts were differentially expressed at least once for the tested time points. To characterize their expression levels, the 231 genes were clustered according to their expression patterns by the hierarchical clustering method using the correlation coefficient of average linkage of the \log_2 ratio (Fig. 2). We could identify 204 differently expressed gene transcripts at T6, 139 at T12, and 27 at T24. T6 and T12 had 112 transcripts in common including all 27 transcripts of T24 and three oppositional expressed transcripts at T6 and T12. The biggest change in transcript expression was found for a hypothetical protein (EVC00294) with a fold-change of 242.18, which was highly expressed in all viral stages. At T6 109 (53%) genes were up-regulated, whereas 95 (47%) were down-regulated. The number of up-regulated genes at T12 decreased to 79 (71%) and the number of down-regulated genes to 60 (29%). All of the 27 expressed transcripts at T24 were up-regulated (Tab. S3).

The category of metabolism showed similar up- and down-regulation pattern at T6 and T12 (Fig. 3A). Most genes involved in carbohydrate and lipid metabolism were up-regulated at T6 (25) and T12 (14). However, only four genes of this category were down-regulated at T6 and six at T12 (Fig. 3B). On the other hand, genes involved in photosynthesis were more down-regulated at T6 (21) than at T12 (4)

whereas the numbers of up-regulated genes were nearly equal for T6 (9) and T12 (8) (Fig. 3C). Genes involved in transcription and translation showed 15 up-regulated compared to two down-regulated genes at T6, and two up-regulated compared to five down-regulated at T12 (Tab. S2).

Among the hypothetical proteins, five genes were consistently up-regulated after 6, 12 and 24 hours of viral infection. In the category carbohydrate and lipid metabolism, we identified an up-regulation of enolase, glyceraldehyde-3-phosphate dehydrogenase and N-acetylneuraminic acid phosphate synthase at all three time points of viral infection. Moreover, among the genes involved in photosynthesis, only ferredoxin I was consistently up-regulated during viral infection with a fold-change between 10 and 12 (Tab. S3). Most of the consistently up-regulated genes at T6, T12 and T24 belong to the category metabolism and show similarity to acetyltransferase, beta-ketoacyl synthase, methyltransferase, farnesyl pyrophosphate synthetase, and NAD (P) H quinone oxidoreductase (Fig. 3A, Tab. S3).

Comparison of ESTs and Microarrays

Clusters of *E. huxleyi* with a total abundance of 15 ESTs or more were compared to the changes in gene expression identified with the microarrays (Fig. 4). A total of 25 clusters were used for trending. It was not possible to detect any changes at T24 because only 10% of the mRNA was of host origin. Therefore, only ESTs of T0, T6 and T12 were compared to the microarray analysis. At T6 and T12 21 genes were found (present) and 4 were not (absent). At T0 only 7 genes were present and 18 were absent. Microarray results displayed no differential expression for one of the absent genes at T6 and for four of the absent genes at T12. Moreover, 11 genes from the T6 library and 9 genes from the T12 library were differentially expressed.

At T6, four genes which were present in the library showed conflicting results. Three gene transcripts encoding for calmodulin were up-regulated according to the EST analysis. They were found 18 (EVC00667), 11 (EVC0450), and 3 (EVC00375) times in the library and were absent in the library generated from the healthy culture (Tab. 1). However, microarray results indicated that they were down-regulated at T6 with a fold-change of 3.56, 2.93 and 3.56, which means that calmodulin was more expressed in the healthy culture than at T6. The fourth gene transcript encoded for ubiquitin (EVC00050) and showed at both T6 and T12 contradictory results. The EST

analysis indicated an up-regulation of the gene, whereas the microarray analysis indicated a down-regulation. In the libraries it was found 5 times at T6 and 11 times at T12, but not in the healthy culture. Results of the microarray revealed that ubiquitin was down-regulated at both time points with a fold change of 4.9 (T6) and 2.6 (T12). In three cases where the gene was absent in the library at T6, microarray results could show that chloroplast light harvesting protein (EVC01904), fucoxanthin chlorophyll a/c-binding protein (EVC01939) and calmodulin (EVC00069) were down-regulated with a fold-change of 3.29, 2.11 and 2.96. These ESTs were found 38, 37 and 2 times in the healthy culture (Tab. 1 and S1).

Discussion

This study aimed at the identification of genes involved in the host-virus interaction of *E. huxleyi* and EhV-86 by taking advantage of EST libraries and DNA microarrays. Both methods contributed to the identification of many genes involved in the host's response to viral infection. We have demonstrated the power of EST libraries and DNA microarrays to obtain data on gene expression and regulation during viral infection. The differential expression of certain genes during viral infection suggests their involvement in the interaction between the host and the virus. This makes them suitable targets for further investigation. ESTs with a total abundance of 15 and above were selected for the comparison between EST redundancy and microarray analysis.

Since the genome sequences of both the virus and the host were available, it was possible to determine the precise number of ESTs from both in all four libraries. Consequently we are able to speculate on both, the host and virus mechanisms that occur in the host cell during viral infection.

The host transcriptome

The most prominent effect of the viral infection on the host transcriptome is the change in the expression of genes involved in photosynthesis, transcription and translation, glycolysis, fatty acid metabolism, and protein degradation.

The break-down of photosynthesis is of particular importance to photosynthetic organisms. During viral infection of *E. huxleyi*, EST and microarray results revealed significant reduction of genes involved in photosynthesis. Our

findings are in accordance with previous studies that observed the reduction of photosynthesis in aquatic communities following viral enrichment (Suttle *et al.*, 1990, Suttle, 1992, Hewson & Fuhrman, 2006). Similar results were also observed in *Chlorella* NC64A (Seaton *et al.*, 1995), *E. huxleyi* (Evans *et al.*, 2006, Llewellyn *et al.*, 2007), *Micromonas pusilla* (Waters & Chan, 1982), and *Heterosigma akashiwo* (Juneau *et al.*, 2003).

For viral replication the biosynthetic machinery of the host cell must be modified for the benefit of viral proteins. Hence, the up-regulation of genes related to protein and ATP synthesis in the mitochondria, genes involved in glycolysis and fatty acid synthesis, and genes involved in RNA-synthesis and translation of the host could be an indication for a boost in the expression of viral proteins which require large amounts of energy. Another indication for enhanced viral protein expression is the down-regulation of host genes involved in photosynthesis, protein degradation, signalling, pigment synthesis, RNA processing, the citric acid cycle, and protein import to the mitochondria (Tab.S2). In general, both organelles (chloroplast and mitochondria) of *E. huxleyi* seem to be down-regulated for the benefit of an increased transcription and translation used for viral replication. It seems that just a few viral proteins are required to alter the biosynthetic machinery of the host, as after 6 hours of viral infection only 38 (3.9%) ESTs were of viral origin.

A previous study has shown that host cells release virus particles between 4 and 48 h after infection while remaining intact (Allen & Wilson, 2006). Furthermore, it has been demonstrated that coexistence of host and virus is possible and that both can replicate during infection (Thyrhaug *et al.*, 2003). Supporting these findings, microarray results showed an up-regulation of several genes involved in photosynthesis during viral infection, which indicates, that at least some cells were intact and perform photosynthesis. In addition, the low abundance of viral ESTs at T12 (2.3%) indicates that perhaps many host cells remained intact and only a few infections took place. However, the change to 89.2% viral ESTs at T24 suggests that between 12 and 24 hours post infection the virus took over the biosynthetic machinery of the entire *E. huxleyi* population.

The virus transcriptome and putative infection mechanism

The virus drastically changes host transcription already after 6 hours, when only 38 (3.9%) viral ESTs are present. A good candidate gene to cause this effect is ehv064, encoding a putative DNA-dependent RNA polymerase II subunit which could be used for transcription of viral genes. Two ESTs for this gene were only found at T6 and not in any other library. The up-regulation of host genes involved in the protein synthesis at T6 and T12 suggest a favoured synthesis of viral proteins. In addition, at T6 a putative protease (EHV349) was identified which might be responsible for the degradation of host proteins. The down-regulation of host genes involved in protein degradation at T6 and T12 such as ubiquitin indicate the takeover of the host machinery. A HNH endonuclease family protein (EHV093) was found at T12 which could lead to the degradation of host DNA. Furthermore, a putative proliferating cell nuclear antigen (EHV440) was identified and which could help in DNA binding during transcription of viral genes. The occurrence of a putative fatty acid desaturase (EHV061) could be involved in membrane lysis, i.e. the lysis of host cells.

At 24 hours post infection 89.2% of mRNA is of viral origin. A putative major capsid protein (EHV085) whose function is well known and defined in viral systems (Allen *et al.*, 2008), and a putative DNA ligase (EHV158) which could be used for DNA replication were identified. These two genes could be involved in the packaging and the following release of new viruses. One copy of a Longevity-assurance (LAG1) family protein (EHV014) was identified which is involved in the ceramide synthesis. Ceramide can act in regulating apoptosis suggesting that LAG1 could be involved in the lysis of host cells. Another indication for the releasing of new viruses could be the up-regulation of host genes involved in the exchange with vacuoles such as Vacuolar ATP synthase catalytic subunit A and V-ATPase subunit d at T6 and T12 (Tab. S2). These suggestions are in accordance with previous studies (Wilson *et al.*, 2005, Allen *et al.*, 2006, Allen & Wilson, 2006). It was shown that at 4 h post infection viruses just started being released. However, at 33 h post infection the host cell is still in a steady stage of virus releasing and re-infection until the host finally lyses after around 48 h.

Gene expression - one goal two approaches

Combining and comparing ESTs and microarray analysis revealed that the combination of both methods is advantageous in estimating the expression level of gene transcripts (Munoz *et al.*, 2004). It has been shown that important functions in an organism are indicated by highly expressed genes (Dupont *et al.*, 2007). Transcripts of low abundance may not occur at all in an EST library but the absence is not necessarily evidence for not being expressed under a different condition (Bouck & Vision, 2007). In most of these cases, our microarray analysis proved helpful by revealing the presence of these genes. Several genes overrepresented within a library and not present at T0, showed a down-regulation in the microarray analysis at T6 and T12. This could be due to the fact that at T0 these genes were underrepresented in the EST library because of the high abundance of e.g. photosynthesis genes. This could be overcome by sequencing more clones to obtain weakly expressed genes.

In summary, we have demonstrated the advantages of an EST approach for simultaneous discovery and identification of host and viral genes involved in viral infection. We have also shown that complementing this approach with microarray analysis enables the detection of even more subtle changes in gene expression. The expression of *E. huxleyi* and EhV-86 genes changed significantly between 12 and 24 hours after infection. Further functional investigations of this infection period are required.

Acknowledgements

We would like to thank Matthew Hall, Susie Wharam, Erika Allhusen, Helga Mehl, and Sabine Strieben for technical assistance. We also thank Dr. Uwe John for consulting in microarray application. This project was funded by Marine Genomics Europe (EU Contract n° 505403).

References

- ALLEN, M.J., FORSTER, T., SCHROEDER, D.C., HALL, M., ROY, D., GHAZAL, P. & WILSON, W.H. (2006). Locus-specific gene expression pattern suggests a unique propagation strategy for a giant algal virus. *Journal of Virology* **80**(15): 7699-7705.
- ALLEN, M.J., HOWARD, J.A., LILLEY, K.S. & WILSON, W.H. (2008). Proteomic analysis of the EhV-86 virion. *Proteome Science* **6**: 11.
- ALLEN, M.J. & WILSON, W.H. (2006). The coccolithovirus microarray: an array of uses. *Brief Funct Genomic Proteomic* **5**(4): 273-279.
- ALTSCHUL, S.F., GISH, W., MILLER, W., MYERS, E.W. & LIPMAN, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**(3): 403-410.
- BALCH, W.M., HOLLIGAN, P.M., ACKLESON, S.G. & VOSS, K.J. (1991). Biological and optical properties of mesoscale coccolithophore blooms in the Gulf of Maine. *Limnology and Oceanography* **36**(4): 629-643.
- BALCH, W.M., HOLLIGAN, P.M. & KILPATRICK, K.A. (1992). Calcification, photosynthesis and growth of the bloom-forming coccolithophore, *Emiliana huxleyi*. *Continental Shelf Research* **12**(12): 1353-1374.
- BATEMAN, A., BIRNEY, E., CERRUTI, L., DURBIN, R., ETWILLER, L., EDDY, S.R., GRIFFITHS-JONES, S., HOWE, K.L., MARSHALL, M. & SONNHAMMER, E.L.L. (2002). The Pfam Protein Families Database. *Nucleic Acids Research* **30**(1): 276-280.
- BAUMANN, K.-H., BÖCKEL, B. & FRENZ, M. (2004). Coccolith contribution to South Atlantic carbonate sedimentation. In *Coccolithophores: From Molecular Processes to Global Impact* (THIERSTEIN, H.R. & YOUNG, J.R. editors), pp. 367-402. Springer, Berlin Heidelberg New York.
- BIDLE, K.D., HARAMATY, L., BARCELOS E RAMOS, J. & FALKOWSKI, P. (2007). Viral activation and recruitment of metacaspases in the unicellular coccolithophore, *Emiliana huxleyi*. *Proceedings of the National Academy of Sciences* **104**(14): 6049-6054.
- BOUCK, A. & VISION, T. (2007). The molecular ecologist's guide to expressed sequence tags. *Molecular Ecology* **16**: 907-924.

- BRATBAK, G., EGGE, J.K. & HELDAL, M. (1993). Viral mortality of the marine alga *Emiliana huxleyi* (Haptophyceae) and termination of algal blooms. *Marine Ecology Progress Series* **93**: 39-48.
- BROWN, C.W. & YODER, J.A. (1993). Blooms of *Emiliana huxleyi* (Prymnesiophyceae) in surface waters of the Nova Scotian Shelf and the Great Bank. *Journal of Plankton Research* **15**(12): 1429-1438.
- BUITENHUIS, E., VAN BLEIJSWIJK, J., BAKKER, D. & VELDHUIS, M. (1996). Trends in inorganic and organic carbon in a bloom of *Emiliana huxleyi* in the North Sea. *Marine Ecology Progress Series* **143**: 271-282.
- BUITENHUIS, E.T., VAN DER WAL, P. & DE BAAR, H.J.W. (2001). Blooms of *Emiliana huxleyi* are sinks of atmospheric carbon dioxide: A field and mesocosm study derived simulation. *Global Biogeochemical Cycles* **15**(3): 577-587.
- CASTBERG, T., THYRHAUG, R., LARSEN, A., SANDAA, R.-A., HELDAL, M., VAN ETTEN, J.L. & BRATBAK, G. (2002). Isolation and characterization of a virus that infects *Emiliana huxleyi* (Haptophyta). *Journal of Phycology* **38**(4): 767-774.
- DUPONT, S., WILSON, K., OBST, M., SKÖLD, H., NAKANO, H. & THORNDYKE, M.C. (2007). Marine ecological genomics: when genomics meets marine ecology. *Marine Ecology Progress Series* **332**: 257-273.
- EIDE, L.K. (1990). Distribution of coccoliths in surface sediments in the Norwegian-Greenland Sea. *Marine Micropaleontology* **16**: 65-75.
- EISEN, M.B., SPELLMAN, P.T., BROWN, P.O. & BOTSTEIN, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America* **95**(25): 14863-14868.
- EVANS, C., MALIN, G., MILLS, G.P. & WILSON, W.H. (2006). Viral infection of *Emiliana huxleyi* (Prymnesiophyceae) leads to elevated production of reactive oxygen species. *Journal of Phycology* **42**(5): 1040-1047.
- GOODWIN, P.H., OLIVER, R.P. & HSIANG, T. (2004). Comparative analysis of expressed sequence tags from *Malva pusilla*, *Sorghum bicolor*, and *Medicago truncatula* infected with *Colletotrichum* species. *Plant Science* **167**(3): 481-489.

- HANEKAMP, K., BOHNEBECK, U., BESZTERI, B. & VALENTIN, K. (2007). PhyloGena a user-friendly system for automated phylogenetic annotation of unknown sequences. *Bioinformatics* **23**(7): 793-801.
- HEGDE, P., QI, R., ABERNATHY, K., GAY, C., DHARAP, S., GASPARD, R., HUGHES, J.E., SNESRUD, E., LEE, N. & QUACKENBUSH, J. (2000). A concise guide to cDNA microarray analysis. *BioTechniques* **29**(3): 548-562.
- HEWSON, I. & FUHRMAN, J.A. (2006). Viral impacts upon marine bacterioplankton assemblage structure. *Journal of the Marine Biological Association of the UK* **86**(03): 577-589.
- HOLLIGAN, P.M., FERNÁNDEZ, 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. & WAL, P.V.D. (1993). A biogeochemical study of the coccolithophore, *Emiliana huxleyi*, in the North Atlantic. *Global Biogeochemical Cycles* **7**(4): 879-900.
- HOLLIGAN, P.M., VIOLLIER, M., HARBOUR, D.S., CAMUS, P. & CHAMPAGNE-PHILIPPE, M. (1983). Satellite and ship studies of coccolithophore production along a continental shelf edge. *Nature* **304**(5924): 339-342.
- HSIANG, T. & GOODWIN, P.H. (2003). Distinguishing plant and fungal sequences in ESTs from infected plant tissues. *Journal of Microbiological Methods* **54**(3): 339-351.
- 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. *Aquatic Microbial Ecology* **27**(2): 111-124.
- JANTASURIYARAT, C., GOWDA, M., HALLER, K., HATFIELD, J., LU, G., STAHLBERG, E., ZHOU, B., LI, H., KIM, H., YU, Y., DEAN, R.A., WING, R.A., SODERLUND, C. & WANG, G.-L. (2005). Large-Scale Identification of Expressed Sequence Tags Involved in Rice and Rice Blast Fungus Interaction. *Plant Physiol.* **138**(1): 105-115.
- JUNEAU, P., LAWRENCE, J.E., SUTTLE, C.A. & HARRISON, P.J. (2003). Effects of viral infection on photosynthetic processes in the bloom-forming alga *Heterosigma akashiwo*. *Aquat. Microb. Ecol.* **31**: 9-17.

- KEGEL, J., ALLEN, M.J., METFIES, K., WILSON, W.H., WOLF-GLADROW, D. & VALENTIN, K. (2007). Pilot study of an EST approach of the coccolithophorid *Emiliana huxleyi* during a virus infection. *Gene* **406**(1-2): 209-216.
- LISS, P.S., HATTON, A.D., MALIN, G., NIGHTINGALE, P.D. & TURNER, S.M. (1997). Marine sulphur emissions. *Philosophical Transactions of the Royal Society B: Biological Sciences* **352**(1350): 159-169
- LLEWELLYN, C.A., EVANS, C., AIRS, R.L., COOK, I., BALE, N. & WILSON, W.H. (2007). The response of carotenoids and chlorophylls during virus infection of *Emiliana huxleyi* (Prymnesiophyceae). *Journal of Experimental Marine Biology and Ecology* **344**(1): 101-112.
- MALIN, G., LISS, P.S. & TURNER, S.M. (1994). Dimethyl sulphide: production and atmospheric consequences. In *The Haptophyte Algae* (GREEN, J.C. & LEADBEATER, B.S.C. editors), pp. 303–320. Clarendon Press, Oxford.
- MARSH, M.E. (2003). Biochemistry and Molecular Biology : Regulation of CaCO₃ formation in coccolithophores. *Comparative Biochemistry and Physiology Part B* **136**(4): 743-754.
- MORIYA, Y., ITOH, M., OKUDA, S., YOSHIKAWA, A.C. & KANEHISA, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucl. Acids Res.* **35**(suppl_2): 182-185.
- MUNOZ, E., BOGARAD, L. & DEEM, M. (2004). Microarray and EST database estimates of mRNA expression levels differ: The protein length versus expression curve for *C. elegans*. *BMC Genomics* **5**(1): 30.
- PAASCHE, E. (2001). A review of the coccolithophorid *Emiliana huxleyi* (Prymnesiophyceae), with particular reference to growth, coccolith formation, and calcification-photosynthesis interactions. *Phycologia* **40**(6): 503-529.
- PARKINSON, J., ANTHONY, A., WASMUTH, J., SCHMID, R., HEDLEY, A. & BLAXTER, M. (2004). PartiGene--constructing partial genomes. *Bioinformatics* **20**(9): 1398-1404.
- RUDD, S. (2003). Expressed sequence tags: alternative or complement to whole genome sequences? *Trends in Plant Science* **8**(7): 321-329.
- SAEED, A.I., SHAROV, V., WHITE, J., LI, J., LIANG, W., BHAGABATI, N., BRAISTED, J., KLAPA, M., CURRIER, T., THIAGARAJAN, M., STURN,

- A., SNUFFIN, M., REZANTSEV, A., POPOV, D., RYLTSOV, A., KOSTUKOVICH, E., BORISOVSKY, I., LIU, Z., VINSAVICH, A., TRUSH, V. & QUACKENBUSH, J. (2003). TM4: a free, open-source system for microarray data management and analysis. *BioTechniques* **34**(2): 374-378.
- SAMTLEBEN, C. & BICKERT, T. (1990). Coccoliths in sediment traps from the Norwegian Sea. *Marine Micropaleontology* **16**: 39-64.
- SCHROEDER, D.C., OKE, J., MALIN, G. & WILSON, W.H. (2002). Coccolithovirus (Phycodnaviridae): Characterisation of a new large dsDNA algal virus that infects *Emiliana huxleyi*. *Archives of Virology* **147**(9): 1685-1698.
- SEATON, G.G.R., LEE, K. & ROHOZINSKI, J. (1995). Photosynthetic shutdown in *Chlorella* NC64A associated with infection cycle of *Paramecium busaria* *Chlorella* virus-1. *Plant Physiology* **108**: 1431-1438.
- STEFELS, J., STEINKE, M., TURNER, S., MALIN, G. & BELVISO, S. (2007). Environmental constraints on the production and removal of the climatically active gas dimethylsulphide (DMS) and implications for ecosystem modelling. *Biogeochemistry* **83**(1-3): 245-275.
- SUKHANOVA, I.N. & FLINT, M.V. (1998). Anomalous blooming of coccolithophorids over the Eastern Bering Sea Shelf. *Oceanology* **38**(4): 502-505.
- SUTTLE, C.A. (1992). Inhibition of photosynthesis in phytoplankton by the submicron size fraction concentrated from seawater. *Marine Ecology Progress Series* **87**: 105-112.
- SUTTLE, C.A., CHAN, A.M. & COTTRELL, M.T. (1990). Infection of phytoplankton by viruses and reduction of primary productivity. *Nature* **347**: 467-469.
- THYRHAUG, R., LARSEN, A., THINGSTAD, T.F. & BRATBAK, G. (2003). Stable coexistence in marine algal host-virus systems. *Marine Ecology Progress Series* **254**: 27-35.
- VENTELON-DEBOUT, M., NGUYEN, T.T.H., WISSOCQ, A., BERGER, C., LAUDIE, M., PIÉGU, B., COOKE, R., GHESQUIÈRE, A., DELSENY, M. & BRUGIDOU, C. (2003). Analysis of the transcriptional response to Rice

- Yellow Mottle Virus infection in *Oryza sativa indica* and *japonica* cultivars. *Molecular Genetics and Genomics* **270**(3): 253-262.
- WATERS, R.E. & CHAN, A.T. (1982). *Micromonas pusilla* Virus: the virus growth cycle and associated physiological events within the host cells; Host range and mutation. *J. Gen. Virol.* **63**: 199-206.
- WILSON, W.H., SCHROEDER, D.C., ALLEN, M.J., HOLDEN, M.T.G., PARKHILL, J., BARRELL, B.G., CHURCHER, C., HAMLIN, N., MUNGALL, K., NORBERTCZAK, H., QUAIL, M.A., PRICE, C., RABBINOWITSCH, E., WALKER, D., CRAIGON, M., ROY, D. & GHAZAL, P. (2005). Complete Genome Sequence and Lytic Phase Transcription Profile of a Coccolithovirus. *Science* **309**(5737): 1090-1092.
- WILSON, W.H., TARRAN, G.A., SCHROEDER, D., COX, M., OKE, J. & MALIN, G. (2002). Isolation of viruses responsible for the demise of an *Emiliana huxleyi* bloom in the English Channel. *J. Mar. Biolog. Assoc. UK* **82**(03): 369-377.

Figure legends

Figure 1. Distribution of functional categories derived from KEGG and percentage of *E. huxleyi* ESTs: (A) healthy culture; (B) 6 hours, (C) 12 hours and (D) 24 hours post infection.

Figure 2. General overview on the differentially expression of genes in *E. huxleyi* during viral infection with a fold change ≥ 2 in at least one timepoint. (A) Heat map generated by hierarchical clustering identifies overall up-regulated (red) and down-regulated (green) patterns of gene expression. The graphs in B and D show the number of the genes up- (B) and down-regulated (D) responding to viral infection after 6 (T1) and 12 h (T2) or both. (C and E) Distribution of up- (C) and down-regulated (E) genes at T1 (black bars), T2 (grey bars) and T3 (white bars) based on their fold change.

Figure 3. Differential expression of genes involved in (A) metabolism, (B) carbohydrate and lipid metabolism, and (C) photosynthesis during viral infection. (I) Heat maps generated by hierarchical clustering identifies overall up-regulated (red) and down-regulated (green) patterns of gene expression. (II) Positive values of barplots indicates the number of significantly up-regulated genes, whereas negative values shows the down-regulated genes at 6 (black bars), 12 (grey bars) and 24 hours (white bars) post infection. (III) Examples for up- (\uparrow) and down-regulated (\downarrow) gene transcripts during viral infection.

Figure 4. Individual ESTs (black bars) compared to microarray results (grey bars) of *E. huxleyi* after 6 h (T1) and 12 h (T2) viral infection. ESTs showing significant similarity to UniProt with a total abundance of 15 and above were selected for the comparison between EST abundance and microarray analysis. For the ESTs, a 1 indicates that at least one EST is present in the library and a 0 indicates absence from the library. For the microarray, a 1 indicates up-regulation of the gene; a -1 indicates down-regulation and a 0 indicates no differential expression.

Tables**Table 1.** The 20 most abundant ESTs in the *E. huxleyi* EST collection deduced from manual annotation.

Cluster ID	Abundance					Description	Process	e-value
	Total	0	6	12	24			
EVC00004	47	0	2	44	1	N-acetylneuraminic acid phosphate synthase	Carbohydrate metabolism	1.4e-71
EVC01904	38	38	0	0	0	Chloroplast light harvesting protein	Photosynthesis	4.5e-21
EVC01939	37	37	0	0	0	Fucoxanthin chlorophyll a/c-binding protein, chloroplast precursor	Photosynthesis	1.2e-23
EVC00115	44	1	26	15	2	60S ribosomal protein L7a	Translation	4.8e-76
EVC00069	25	0	0	25	0	Calmodulin (CaM)	Signal transduction	3.1e-75
EVC00071	36	0	24	10	2	40S ribosomal protein S11	Translation	9.9e-45
EVC00067	25	0	4	21	0	alpha-1,4-N-acetylglucosaminyltransferase EXTL3	Glycan biosynthesis and metabolism	2.7e-18
EVC00210	38	0	19	19	0	Elongation factor 1-alpha (EF-1-alpha)	Translation	0.0
EVC00024	23	0	3	19	1	60S ribosomal protein L3	Translation	6.9e-99
EVC00667	23	0	18	1	4	Calmodulin (CaM)	Signal transduction	1.2e-74
EVC00014	38	2	17	15	4	Ubiquitin	Folding, sorting and degradation	3.2e-35
EVC01892	16	16	0	0	0	Chloroplast light harvesting protein	Photosynthesis	1.5e-21
EVC00025	24	1	15	8	0	40S ribosomal protein S13	Translation	4.6e-58
EVC00375	19	0	3	15	1	Calmodulin (CaM)	Signal transduction	1.2e-74
EVC00045	22	2	14	6	0	Adenosylhomocysteinase	Amino acid metabolism	0.0
EVC01249	15	0	14	0	1	60S ribosomal protein L27	Translation	9.2e-42
EVC00200	19	0	13	5	1	Eukaryotic translation initiation factor 1A (EIF-1A)	Translation	6.2e-39
EVC00172	21	0	7	13	1	Hypothetical protein	Hypothetical	
EVC00457	19	0	12	5	2	Acetyltransferase-like	Metabolism	2.5e-11
EVC00054	17	0	4	12	1	NmrA family protein	Other	1.2e-17

0 = uninfected library; 6, 12 and 24 = hours after viral infection.

Table 2. EhV-86 genes in the virus infected EST libraries with known function and their putative process in viral infection.

Cluster ID	Abundance			Description	Putative process
	6	12	24		
EVC00969	0	0	19	ehv133 putative ATP-dependent protease proteolytic subunit	A
EVC01295	0	0	1	ehv361 putative serine protease	A
EVC00898	1	0	3	ehv349 putative protease	A
EVC00833	0	0	15	ehv15 putative fatty acid desaturase	B
EVC00358	0	1	5	ehv061 putative fatty acid desaturase	B
EVC01280	0	0	1	ehv028 putative lipase	B
EVC00896	0	0	7	ehv108 putative DNA-directed RNA polymerase subunit	C
EVC00899	0	0	6	ehv434 putative DNA-directed RNA polymerase II subunit	C
EVC01004	0	0	1		
EVC01069	0	0	2	ehv399 putative DNA-directed RNA polymerase subunit	C
EVC01347	2	0	0	ehv064 putative DNA-dependent RNA polymerase II largest subunit	C
EVC01224	0	0	1	ehv105 transcription factor S-II (TFIIS) family protein	C
EVC01054	0	0	2	ehv041 putative endonuclease	D
EVC01011	0	0	1		
EVC00938	0	0	5	ehv018 putative endonuclease	D
EVC00177	0	1	8	ehv093 HNH endonuclease family protein	D
EVC00936	0	0	4	ehv430 putative helicase	D
EVC00835	0	0	12	ehv397 putative deoxyuridine 5'-triphosphate nucleotidhydrolase	E
EVC01104	0	0	2		
EVC00838	0	0	7	ehv459 putative nucleic acid independent nucleoside triphosphatase	E
EVC00818	0	0	4	ehv026 ribonucleoside-diphosphate reductase small chain	E
EVC01059	0	0	2	ehv030 putative DNA polymerase delta catalytic subunit	E
EVC00999	0	0	1		
EVC00830	0	0	2	ehv428 putative ribonucleoside-diphosphate reductase protein	E
EVC00871	0	0	2	ehv136 putative nucleic acid-binding protein	F
EVC00891	0	0	2	ehv020 putative proliferating cell nuclear antigen	F
EVC01089	0	0	1		
EVC00506	0	1	4	ehv440 putative proliferating cell nuclear antigen	F
EVC00911	0	0	5	ehv152 putative DNA-binding protein	G
EVC00926	0	0	1	ehv184 putative DNA-binding protein	G
EVC01199	0	0	1	ehv431 putative thymidylate kinase	G
EVC00963	0	0	1	ehv451 putative protein kinase	H
EVC01201	0	0	1	ehv402 putative protein kinase	H
EVC00852	0	0	3	ehv179 Major Facilitator Superfamily protein	I
EVC00812	0	0	4	ehv444 putative DNA topoisomerase	J
EVC01114	0	0	2	ehv085 mcp putative major capsid protein	K
EVC00836	0	0	4	ehv158 putative DNA ligase	L
EVC00977	0	0	2	ehv014 Longevity-assurance (LAG1) family protein	M
EVC01106	0	0	2	ehv166 putative RING finger protein	N
EVC01230	0	0	1	ehv110 putative RING finger protein	

EVC00829	0	0	1	ehv023 putative deoxycytidylate deaminase	O
EVC00831	0	0	4	ehv103 putative vesicle-associated membrane protein	P
EVC00907	0	0	5	ehv455 putative sialidase	-
EVC01694	1	0	0		

A: degradation of host proteins; B: membrane lysis; C: virus gene transcription; D: degradation of host DNA; E: DNA replication; F: DNA binding; DNA polymerase processivity factor activity; G: viral DNA replication; H: signal transduction; I: membrane transport; J: insertion of viral DNA; K: capsid synthesis; L: DNA replication and host lysis; M: cell lysis; N: mediating protein-protein interactions; O: hydrolase activity, zinc ion binding; P: virus release, transport of capsid proteins.

Supplemental

Table S1. Excel-file of all ESTs used in this study separated into three different sheets. The first sheet contains all ESTs of *E. huxleyi* with a BLAST hit and its description and involved process, including their array and cluster ID, their oligo sequence for the microarrays, their abundance, their e-value and length, KEGG and EC numbers, and the species of their closest BLAST hit and its swissprot ID. The second sheet contains all host ESTs with no significant hit in BLAST searches. The third sheet contains all ESTs of EhV86 identified in this study, including their cluster ID, e-value, length, and description.

Table S2. Excel-file of all host gene transcripts which were differentially expressed after 6 or 12 hours viral infection. The array and cluster ID, the redundancy of the ESTs for each library, the description of the gene resulting from the EST analysis and its involvement in a process, and the fold-change at T6 and T12 are included.

Table S3. Excel-file of all consistently up-regulated genes at T6, T12 and T24 including their array and cluster ID, gene description and process involved, and their fold-change.

Figure 1

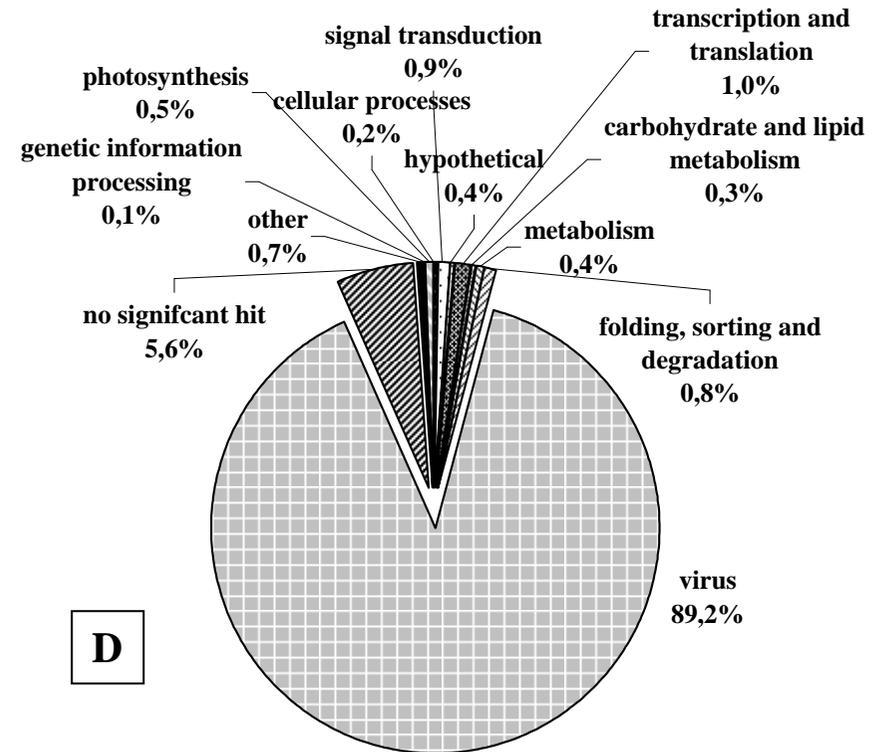
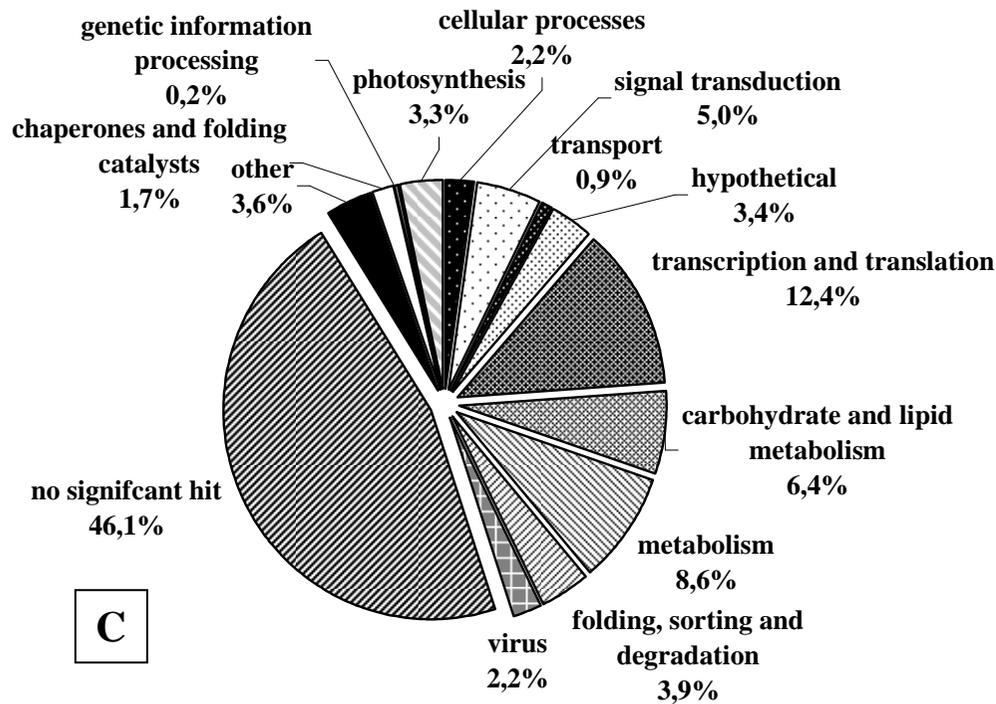
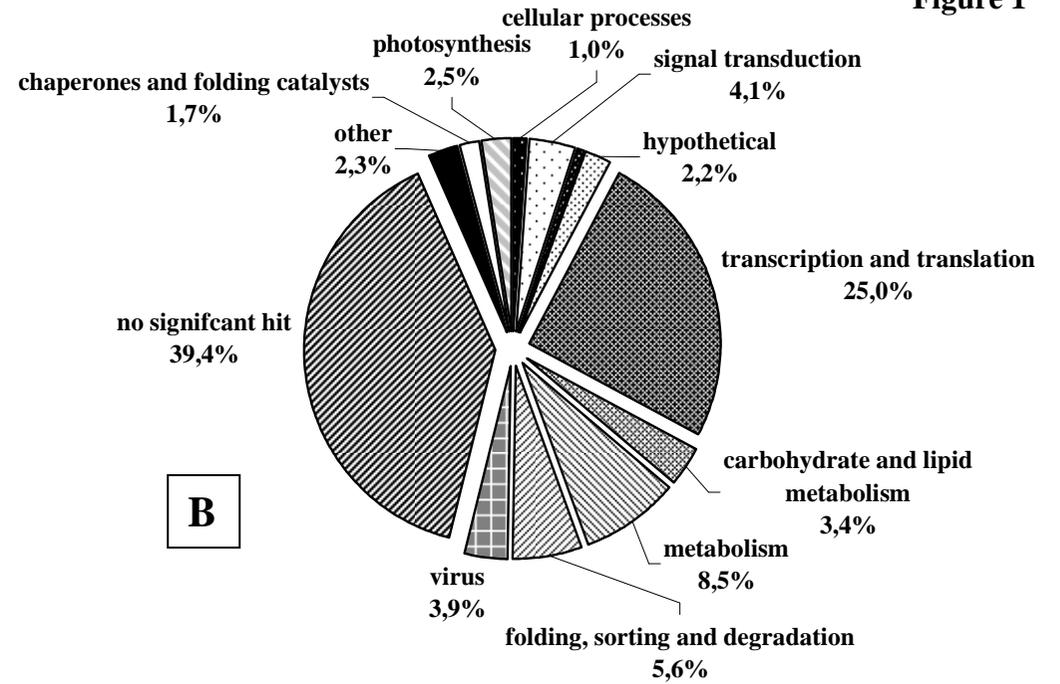
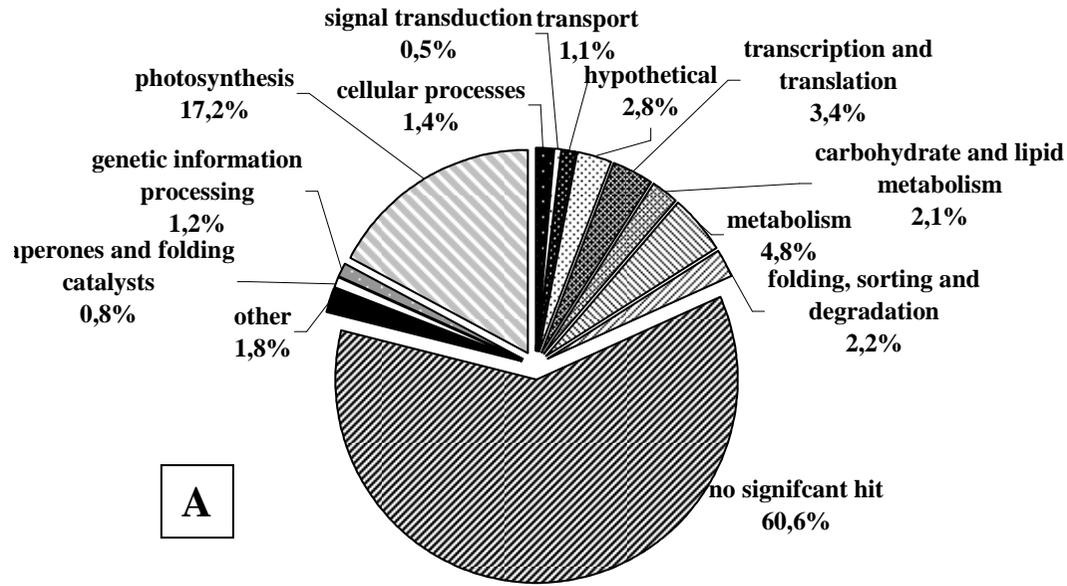


Figure 2

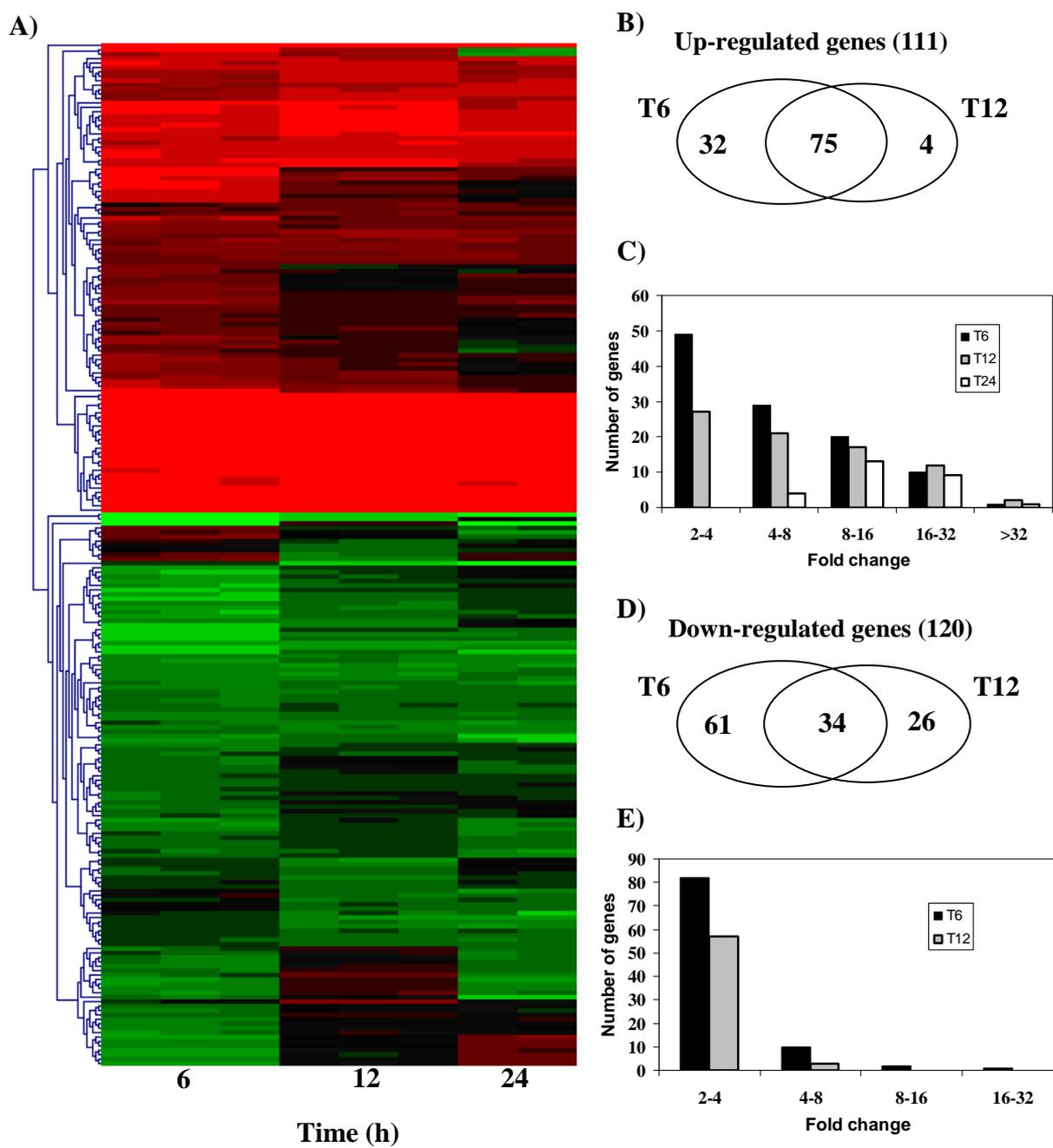
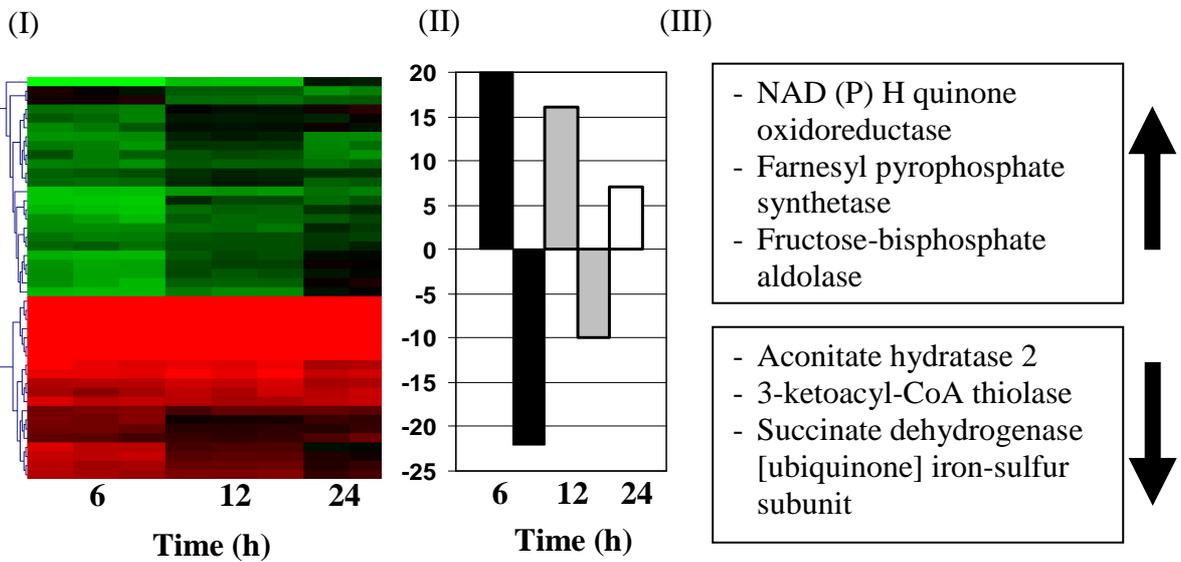
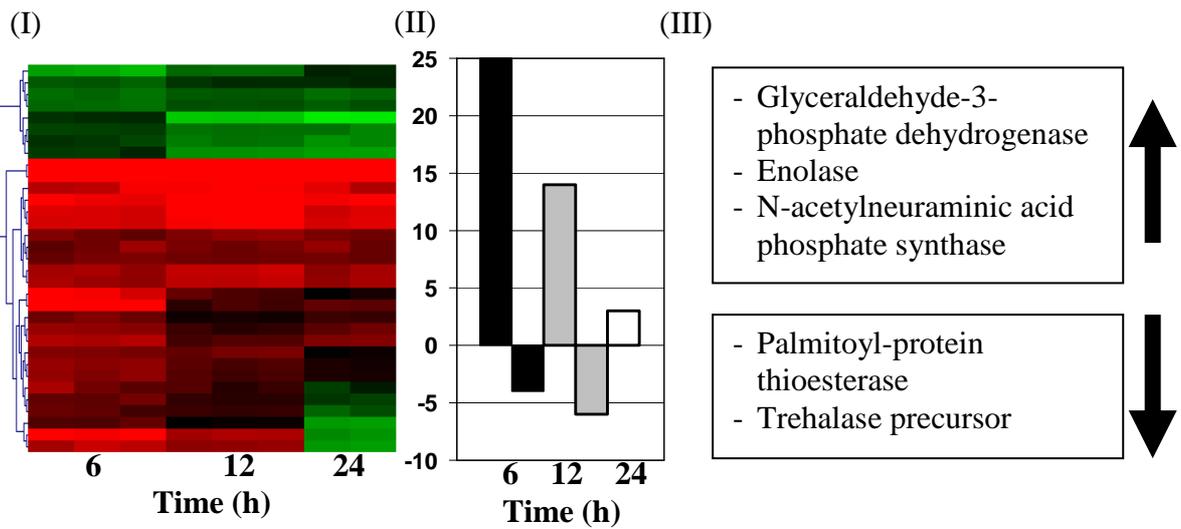


Figure 3

A



B



C

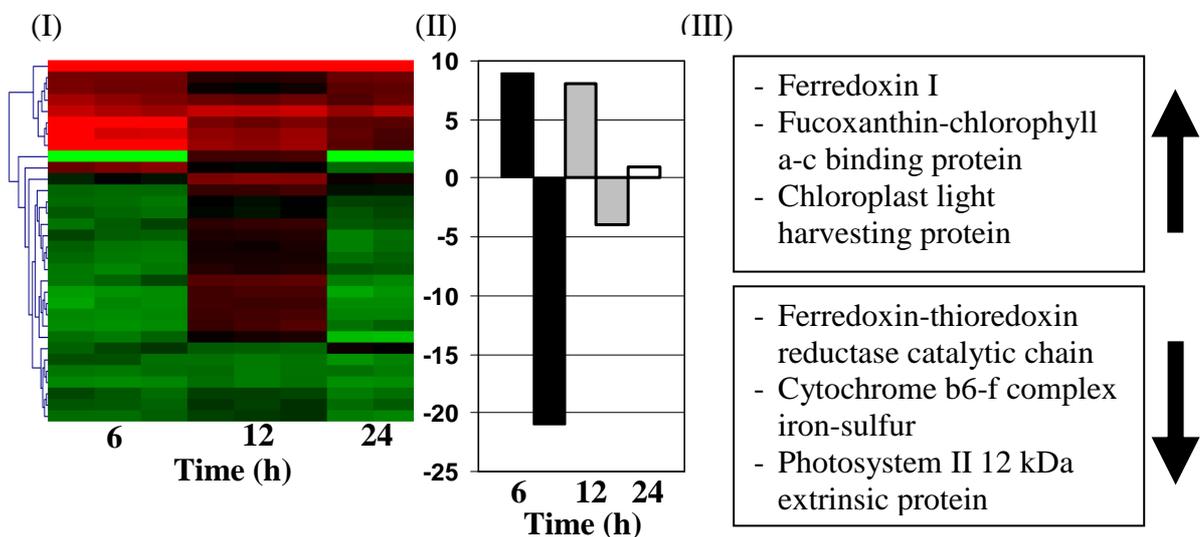
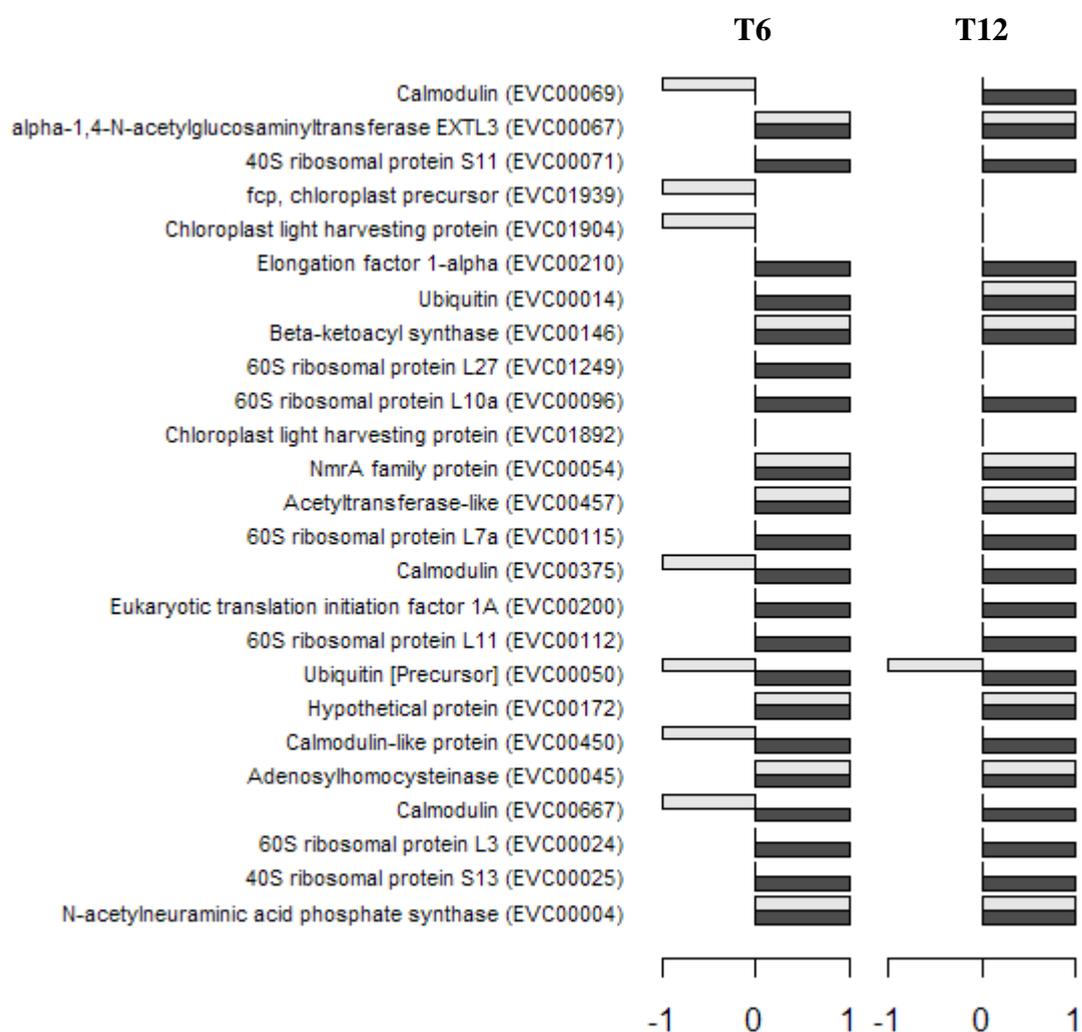


Figure 4



PUBLICATION III

Comparative functional genomics of virus-susceptible and virus-resistant strains of *Emiliana huxleyi*

Jessica U. Kegel^{1*}, Uwe John¹, Klaus Valentin¹ Katja Metfies² and Stephan
Frickenhaus¹

¹ Alfred Wegener Institute for Polar- and Marine Research Bremerhaven, Am
Handelshafen 12, 27570 Bremerhaven, Germany

² GKSS Research Center, Max Planck Strasse 1, 21501 Geesthacht, Germany

*Corresponding author

E-MAIL Jessica.Kegel@awi.de; TEL (0049) 471-4831-1996

Email addresses:

JK: Jessica.Kegel@awi.de

UW: Uwe.John@awi.de

SF: Stephan.Frickenhaus@awi.de

KV: Klaus.Valentin@awi.de

KM: Katja.Metfies@awi.de

Abstract

Background

Emiliania huxleyi is the most studied coccolithophore with respect to biogeochemical cycles, climatology, and host-virus interactions. Our aim was to investigate genomic diversity and plasticity of 16 *E. huxleyi* strains of different geographic origin and furthermore we emphasized to identify genes related to virus susceptibility and morphology, such as coccolith production and coverage. A microarray-based comparative genomic hybridization (CGH) was set up using 565 genes derived from an EST study of *E. huxleyi* strain CCMP1516 and 37880 genes from the ongoing genome project of the same strain. *Gephyrocapsa oceanica* and *Isochrysis galbana* were taken as out-groups.

Results

After normalisation and ANOVA test a total of 32395 genes had significant hybridization patterns. Comparisons with the sequenced *E. huxleyi* strain CCMP1516 revealed that 27% (8740 genes) to 57% (18581 genes) of the genes showed a pattern of hybridization concordant with deletion, nucleotide divergence or gene duplication within the species and up to 83% (26881 genes) between the genera. The largest variation was observed among the species for *E. huxleyi* strain 92F. Regarding variation with respect to virus susceptibility and morphology the most abundant genes with known function were associated with metabolism, transport, and transcription and translation. In addition, we identified two membrane receptors and two proteins related to ubiquitin which show significant differences between virus susceptible and resistant strains.

Conclusions

The results we have obtained by using CGH demonstrate that this method is appropriate to compare genome plasticity and the gene content of different *E. huxleyi* strains. We have successfully applied this method to identify genes related to virus susceptibility and morphology. Among others, the membrane receptors and the ubiquitin-related proteins that possibly play a role in virus infection deserve further attention.

Background

The prolific coccolithophore *Emiliana huxleyi* is distributed worldwide and has the ability to form immense coastal and open ocean blooms ranging from sub-polar to tropical latitudes [1,2] that can cover more than 50.000 km² [3-5]. These blooms can be detected via satellite imagery due to the reflection of the coccoliths [6,7]. This makes *E. huxleyi* an important factor influencing the global biogeochemical cycles of carbon and sulphur and one of the most important species on earth with respect to sediment formation and ocean climate [8,9]. Therefore, it is a key species for current studies on global biogeochemical cycles [10]. It is also of interest to scientists from fields as diverse as geology, biogeography, paleoclimatology, ecophysiology, material science, and medicine [11]. Whereas the bloom formation is stimulated from abiotic environmental factors, the bloom control and termination is highly influence by viral infection [12,13]. Consequential, a range of viruses specific to *E. huxleyi* (EhVs) have been isolated [14,15] and were further analyzed for their phylogeny [16,17], ecological succession in mesocosm experiment [18,19], and genome structure of *Emiliana huxleyi* virus 86 (EhV-86) [20-22]. Hence, it is one of the most studied eukaryotic phytoplankton host-virus systems to date [23].

Microarray-based comparative genomic hybridization (CGH) is well established in microbial and human (cancer) research, to determine DNA copy number variants in healthy subjects [24,25], genomic aberrations associated with various diseases and syndromes [26], and between the genomes of e.g. closely related taxa, such as species and strains [27-29]. CGH is used to compare the genes present, absent or divergent in the genomes of interest. Polymorphisms and insertions can be detected as a reduction or elevation of a hybridization signal [30]. Whole genome comparisons of different strains of various microbes indicate that polymorphism for gene content is not uncommon [27,31,32], suggesting genetic adaptations to different ecological niches. Previous studies have reported different genome sizes among different morphotypes of *E. huxleyi* from different geographical regions via DNA microsatellites and restriction fragment length polymorphism (RFLP) analysis [33,34]. Results indicate the presence of different ecotypes of *E. huxleyi* potentially with differences in genome organization in response to environmental conditions or to potential threats, such as viral infections. Furthermore, an example for a connection between genetic variation and virus susceptibility has been demonstrated [16]. It was

found that virus resistant strains of *E. huxleyi* display a higher dimethylsulfoniopropionate lyase (DMSP-lyase) activity than strains that are susceptible to virus infection. One reason for the different enzyme activities could be variations in the expression of the gene coding for the enzyme due to either a change in transcriptional regulation or a change of the copy number in the genomes.

In this study, we constructed and used a whole genome microarray comprising unique probes for each gene of the sequenced strain *E. huxleyi* CCMP1516 (reference strain) to examine the genetic diversity among 16 different strains of *E. huxleyi* of different geographic origin. We aimed not only at assessing the genetic diversity of *E. huxleyi* but also at the elucidation of genes responsible for the differential virus susceptibility and morphology of the different *E. huxleyi* strains.

Results

In order to generate a comprehensive DNA-microarray of *E. huxleyi*, we designed 37880 oligonucleotides by using the application tool eArray (Agilent technologies). The design was based on 39125 gene transcripts, mostly based on automated predicted gene models, of the whole genome of *E. huxleyi* strain CCMP1516. Due to technical reasons it was not possible to design oligonucleotides for all gene transcripts of the whole genome. In addition to the newly designed probes, the microarray contained 565 oligonucleotides based on sequence information previously generated in an expressed sequence tag (EST) study (Kegel *et al.* manuscript submitted). To investigate genetic diversity of *E. huxleyi*, genomic DNA from 15 different strains were compared with genomic DNA of the sequenced *E. huxleyi* strain CCMP1516 by co-hybridization to the microarray. *Gephyrocapsa oceanica* and *Isochrysis galbana* as phylogenetic closely related taxa were used as out-groups. By comparing the log₂-ratios (LR) of the hybridization signals from the different strains, it can be deduced whether a gene is present or absent in the genomes compared to the reference strain due to the variation in copy numbers of the genes.

As a first step an ANOVA test was performed (see methods) to assure the quality of our analysis. It resulted in a total of 32395 significant mean signals with p-values < 0.01. A threshold of the log₂-ratio (LR = sample/reference) for no significant difference between sample and reference (*E. huxleyi* CCMP1516) was

determined by a self-versus-self hybridization of the reference. The density of the reference LR showed an approximate normal distribution (data not shown) around 0. The boundaries for the threshold for the identification of genetic divergence were taken by computing a 99% interval of the reference LR, resulting in a positive cut-off of 0.48696 and a negative cut-off of -0.8270 for the LR. The number of genes possessing a LR ranging from -0.8270 to 0.48696 in each *E. huxleyi* strain showed between 42.6% and 70.7% homology to the reference strain (Fig. 1). The majority of differences in genome structure (19.6-36.9%) among these strains were in the $LR \geq 0.48696$ category, indicating increased copy number. The $LR \leq -0.8270$ indicating reduced copy numbers, deletion, or mismatches between probe and gene sequence were in a range between 10.3% and 20.5%. *E. huxleyi* strain 92F showed the highest degree of variability to the reference (57.4%), whereas *E. huxleyi* strain 12-1 and strain EH2 had the highest similarity to the reference strain (73.2% and 70.5%). *G. oceanica* showed 31.5% and *I. galbana* only 17.0% homology to the reference.

A neighbor-joining consensus dendrogram of the sixteen *E. huxleyi* strains and the two out-groups *G. oceanica* and *I. galbana* based on the CGH data is depicted in the dendrogram with bootstrap values in Figure 2. As expected, the two out-groups clustered perfectly outside of all *E. huxleyi* strains with 100% bootstrap support. The *E. huxleyi* strains grouped into two main clusters. The strains Van556, 92D, 92F and 373 showed the highest degree of divergence to all other *E. huxleyi* strains and clustered into one of the main groups. The second main cluster can be subdivided into two sub-groups. The strains EH2 and 12-1 were most similar to the reference strain and clustered into one of the two sub-groups. It is interesting to note that EH2 clustered directly with the reference whereas 12-1, which showed a higher similarity to the reference, clustered with the group of EH2 and the reference. The second sub-group consists of two well supported sub-clusters. The bootstrap values within the main clusters of the 16 *E. huxleyi* strains were higher (92-100%) than the support for the clustering into the two main clusters (49%).

To identify genes involved in virus susceptibility and morphology a non-parametric Wilcoxon exact rank-sum test was applied. Prior to analysis, all mean signals that failed to produce a positive result above the threshold of the reference ($LR > 0.48696$)

in at least one of the 16 different *E. huxleyi* strains (excluding *G. oceanica* and *I. galbana*) were removed from the data set. This resulted in 21371 signal combinations. Significance was defined as p-value < 0.01. The resulting genes were manually analysed by BLAST searches against the NCBI non-redundant protein database (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>, NCBI-nr) and the SwissProt database (<http://www.expasy.org/sprot/>) and were compared with Pfam [35].

The analysis of the genes concerning virus susceptibility or resistance yielded in 141 candidate gene transcripts (Fig. 3). Almost half of these gene transcripts (69) showed no significant similarity to any sequences in the public database or were of unknown function (Tab. S1). We found proteins involved in metabolism (25), transport (13), carbohydrate and lipid metabolism (6) (especially glycolysis, 3x), transcription and translation (11), protein interaction (2), cellular processes/cytoskeleton proteins (3), replication and repair (3), signal transduction (4), and chaperone and folding catalysts (1). Furthermore, we identified two ubiquitin-related proteins involved in folding, sorting and degradation and two membrane receptor proteins; one scavenger receptor protein and one receptor L domain-containing protein.

The identification of genes responsible for morphology resulted in 125 candidate gene transcripts (Fig. 4). A total of 72 out of the 125 gene transcripts showed no significant similarity to any sequences in the public database or were of unknown function. We could identify genes involved in metabolism (15), transcription and translation (7), transport (6), carbohydrate and lipid metabolism (6), amino acid metabolism (4), signal transduction (4), photosynthesis (3), proteolysis (2), replication and repair (2), cellular processes (1), folding, sorting and degradation (1), protein interaction (1), and RNA processing and modification (1). A glycosyltransferase could be identified within the group of metabolism. In addition, an elongation of very long chain fatty acids like-protein was identified within the group of lipid metabolism and a CASP-like protein involved in transport.

Discussion

The transcriptome of 16 *E. huxleyi* strains of different geographic origin were compared in order to identify genomic differences in terms of plasticity and possible relation to virus-susceptibility and morphology. Comparative genomic hybridization was used to characterize the 15 strains with respect to gene content similarities with the sequenced reference strain *E. huxleyi* CCMP1516. Hybridization intensities were compared to determine the relative copy number of each gene transcript.

The results of the DNA microarray-based CGH method revealed phylogenetic relationships between *E. huxleyi* strains and the two out-groups *G. oceanica* and *I. galbana* based on cluster analysis of the log₂-ratios. The analysis of the hybridization patterns showed the genetic distance between strains and ranged from 27% (8740 genes) to 57% (18581 genes). These genes were concordant with deletion, nucleotide divergence or gene duplication. When comparisons were made with the reference and the two out-groups, the genetic distance increased up to 83% (26881 genes).

The genetic distances between strains of *E. huxleyi* are in accordance with previous reports that demonstrated different genome sizes among different morphotypes of *E. huxleyi* from different geographical regions based on DNA microsatellites and restriction fragment length polymorphism (RFLP) analysis [33,34].

CGH has been extensively utilized to elucidate genetic diversity mainly in bacterial systems like *Helicobacter pylori*, *Campylobacter jejuni*, *Entamoeba histolytica*, *Francisella tularensis*, *Mycobacterium tuberculosis*, [32,36-39] but also in the eukaryotic systems of yeast [27]. Microarray analysis had indicated a limited genetic variation within the species and strains. Strain comparisons showed differential hybridization between 0.17 and 16.7% of the gene transcripts. The genetic diversity increased at most up to 90% within the subspecies. In contrast, our results revealed between 27% and 57% genetic variation within the species and up to 69% to the older genus *G. oceanica*. As *E. huxleyi* has evolved from *G. oceanica* only 268.000 years ago [40] and became dominant around 70.000 years ago, this high genetic diversity could indicate that *E. huxleyi* is still in its evolutionary radiation.

The highest variability amongst the strains was observed in the case of strain 92F (57.4%). This strain is virus susceptible and possesses the ability to produce

coccoliths, both suggesting a higher similarity to the reference strain (CCMP1516). An explanation for the high genomic deviation from the reference could be their geographic origin. The reference strain was isolated near the coast of Ecuador whereas 92F was obtained from the English Channel (Tab. 2). Another strain collected in the English Channel, 92E, showed 70.3% similarity to the reference strain. Both English Channel strains possess coccoliths but show different virus susceptibility. Our genomic comparisons support earlier findings of blooms being dominated by a succession of different populations [41]. During a bloom the community composition of *E. huxleyi* is affected by viruses in the role of a population controlling factor. One specific population is decimated by viruses making blooms of succeeding populations possible [41]. Affected populations induce a life-cycle transition to a haploid (1n) stage to escape viral infection as the haploid phase has been demonstrated to be resistant to viral infection [42]. At the time of isolation in 1992 during a bloom in the English Channel, the strain 92F was in a haploid stage while strain 92E was in a diploid (2n) stage. This suggests that 92F was escaping viral infection, as it was shown that virus attack could be one reason to induce life-cycle changes from 2n to 1n [42]. Consequently in this scenario 92E (2n) was blooming due to resistance to specific viruses. One reason of the genetic distance between these two strains could be still differences in the morphotypes and genome size.

Differences in the ecological strategies of *E. huxleyi* strains (e.g. bloom dynamics) could also cause these genetic differences [43]. Genome size differences can have different reasons and hence several hypothesis have been formulated [44,45]. Species with smaller genomes are streamlined to survive in stable environments. On the other hand, larger genomes provide organisms with a broader range of metabolic capabilities allowing them to take advantage of more complex and variable environments. However, more strains/ecotypes of this rather young species [46] should be analyzed before drawing definite conclusions on their genetic variation.

The Wilcoxon rank-sum test was used to elucidate whether a lack of certain genes, copy number changes or sequence divergence between reference and tester strain may explain the different biological properties of virus susceptibility or morphology.

We identified two receptor proteins; one scavenger receptor protein and one receptor L domain-containing protein which showed a decreased log₂-ratio in the virus susceptible group than in the resistant group. The first step of virus infection involves attachment of virus particles to host-specific cell surface receptors [47,48]. This prepares the way for the viruses to enter the host cell. Once inside the host cell, viruses utilize the host machinery in order to enhance the efficiency of its replication process. Consequently, the expression of the receptor on the outer surface of the host is a major determinant of the route of entry of the virus into the host and of the patterns of virus spread and pathogenesis in the host [47]. Viruses have evolved to exploit these receptors to gain entry into cells. As each virus is looking for a specific receptor that fits its attachment protein, the host receptor will, in part, determine the susceptibility of different hosts to the same virus. Previous studies have demonstrated that the lack of receptor expression restrict virus entry [49-51]. The identified scavenger receptor is a transmembrane glycoprotein and reminiscent of members of the immunoglobulin (Ig) superfamily [52]. Members of this family are known to be involved in entry of more than one virus into cells ([48] and references herein). The L domain of the second identified membrane receptor is also found in insulin receptor (IR) which is closely related to members of the tyrosine-kinase receptor superfamily. Members of this family play a role in different cellular processes, including division, proliferation, apoptosis, and differentiation [53]. Moreover, it has been shown that protein kinases are activated by viral infection [54], suggesting that the L domain-containing membrane receptor could be involved in virus susceptibility or infection. Different virus susceptibility could be due to differences in copy numbers or modifications of these two receptors. Therefore, the identified receptors are suitable targets for further investigations regarding virus susceptibility. Quantitative PCR (qPCR) can be used to determine the absence, presence and the real copy number of the target genes in each genome. Furthermore, sequence analysis of these receptors will be conducted among the studied strains and the two out-groups in order if sequence variation (deletion, insertion and base substitutions) can be determined and linked to virus susceptibility.

The occurrence of an ubiquitin and an ubiquitin-conjugation enzyme E1 might indicate its involvement in virus infection. Previous results of an EST study combined

with gene expression analysis by using microarrays had indicated the down-regulation of three ESTs related to the ubiquitin protein family and the up-regulation of two of them during viral infection (Kegel *et al.* manuscript submitted). Ubiquitin and its relatives regulate processes in eukaryotic cells by covalent attachment to other cellular proteins, thereby changing the stability, localization, or activity of the target protein [55]. The most prominent function of ubiquitin is the mediated proteolysis of labelled target proteins. Moreover, ubiquitin modifications are also involved in virus budding [56] indicating the importance of ubiquitin and its relatives for virus susceptibility or infection.

It would be of great importance to show that there are significant differences in the degree of variation in the genes associated with ubiquitin. However, more strains should be analyzed and other quantitative methods like qPCR should be applied before drawing definite conclusions on its involvement in virus infection.

The identification of genes related to morphology, i.e. the formation of coccoliths, revealed potential genes for further applications. We identified a protein for elongation of very long chain fatty acids which is in accordance with the discovery of a similar gene in a previous study associated with biomineralization [57]. As coccolith precursors are synthesized in the Golgi-derived structures [58,59], the identified CASP-like protein would be of interest due to its possible role in intra-Golgi transport. In addition, it has been shown that in the species *Pleurochrysis*, the coccolith formation is mediated by acidic polysaccharides [58,60]. These polysaccharides show a significant level of homology to glycosyltransferases. Therefore, our discovery of a glycosyltransferase in *E. huxleyi* could be possibly linked to coccolith formation. Several hypotheses about coccolith formation exist (for overview see [61]). Furthermore, novel genes possibly involved in calcification and coccolithogenesis were identified by EST approaches, microarrays for gene expression analysis, and suppressive subtractive hybridization [11,57,62]. Nevertheless, the process of coccolithogenesis and the exact genes involved in remain to be elucidated.

Conclusions

We successfully applied microarray-based CGH to compare the genomic content of different *Emiliania huxleyi* strains regarding virus susceptibility and morphology. Among others, the membrane receptors and the ubiquitin-related proteins that possibly play a role in virus infection deserve further attention. Future work will include the use of microarrays transcript profiling experiments and knock-out mutants to focus on the expression of the identified key genes. This will extend our understanding of virus susceptibility and viral infection of *E. huxleyi*.

Methods

Strains and culture conditions

Emiliania huxleyi strains (Tab. 1) and *Gephyrocapsa oceanica* were cultured in f/2 medium and *Isochrysis galbana* in K media at 15°C with a 16:8 light-dark cycle and $150 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$. EH2 and NZEH were treated with 1000 $\mu\text{g}/\text{mL}$ Kanamycin because they were too sensitive against the antibiotic mixture. All other cultures were treated with a mixture of Ampicillin, Gentamycin, Streptomycin, Chloramphenicol and Ciprofloxacin (Tab. 2). Antibiotic treatment took place over 10-12 days. After 5-6 days cultures grown in 200 mL treated with antibiotics were transferred to 800 mL antibiotic treated f/2 media. Five to six days later cells were harvested on 1.2 μm RTPP ISOPORE filters Millipore. Cultures were checked against bacteria with acridine-orange staining. Only samples with no observed bacteria were used for analysis, although we cannot reduce a highly reduced bacterial background.

Genomic DNA labelling

All steps were performed in technical triplicates in order to avoid methodological errors in the hybridisation patterns interpretation. Genomic DNA was isolated from the samples using Qiagen DNeasy Plant Mini Kit (Qiagen, Hilden, D) and were then subjected to amplification according to Agilent's protocol for oligonucleotide array-based CGH for genomic DNA (version 5.0, June 2007). Restriction digestion was performed with 200 ng of genomic DNA for 8 h at 37°C. Digested DNA from each test strain and species was labelled with Cy5-dUTP whereas *E. huxleyi* strain CCMP1516 was labelled with Cy3 as reference. Labelled DNA sample yields and dye incorporation efficiencies were assessed photometrical (Nanodrop ND-1000,

PecLab). Specific activity (pmol dyes per μg genomic DNA) were calculated as [pmol per μL dye/ μg per μL genomic DNA] from the results of photometry.

Microarray hybridizations

Labelled samples were then co-hybridized with the reference *E. huxleyi* strain CCMP1516 in triplicates to Agilent oligonucleotide-based 44k custom-made microarrays. One Array contained 37880 different transcripts derived from the *E. huxleyi* CCMP1516 genome project conducted by the U.S. department of Energy Joint Genome Institute (<http://www.jgi.doe.gov/>) using the best gene model for each locus and 565 transcripts from our EST study (Kegel *et al.* manuscript submitted). Microarrays were designed with Agilent's eArray online application tool version 5.0.

A self-versus-self hybridization was performed in triplicates for determining probe specificity, array reproducibility, and microarray feature uniformity. Hybridizations were done for 24 h with 20 rpm using a hybridization chamber (Agilent technologies). After hybridization, the microarrays were washed according to the manufacturer's instructions (Wash with Stabilization and Drying Solution, Agilent Technologies).

Data acquisition and analysis

Microarrays were scanned using a G25005B Agilent microarray scanner with 100% photomultiplier tube (PMT) settings for both channels and 5 μm scan resolution. Signal intensities were detected and normalized by Feature Extraction software version 9.5 (Agilent Technologies) using the GE protocol and matrix. Spots which were not well above background in the self-self hybridization were removed before further analysis. Results were first analyzed using the MeV software package from TIGR [63]. An ANOVA test was performed for all groups with a p-value < 0.01 and a standard Bonferroni correction. The average intensity from the significant genes of the triplicates was used for further analysis. Neighbor-joining trees of microarray data using Euclidean distance metrics (n = 1.000 bootstrap iterations) was performed in R with the ape-package (<http://www.r-project.org>, <http://ape-mol.ird.fr/>).

The exact Wilcoxon rank-sum test from the R-package exactRankTests [64] was used to compare log₂-ratios (LR) between groups of samples to identify genes regarding virus susceptibility and morphology. The reference strain (CCMP1516) and

the two out-groups *G. oceanica* and *I. galbana* were excluded in this analysis. Strains were grouped according to their virus susceptibility (Tab. 1) and morphology (Tab. 3), i.e. formation of coccoliths. Significance was assumed for as p-value < 0.01. For visualization, LR of significant gene transcripts were used in a heatmap clustering [65] which reorders rows (signals) and columns (strains) according to dissimilarity by a Manhattan metric and a hierarchical clustering by Ward's method [66]. The resulting heatmaps were drawn with a colour-scale from red (minimum) to white (maximum).

Original data files for all arrays were uploaded in MIAME format for expression arrays at GEO (<http://www.ncbi.nlm.nih.gov/geo/>; accession numbers XXXX-XXX).

Authors' contributions

JK performed all lab-work including DNA isolation and microarray hybridizations, carried out analysis and interpretation of microarray data, annotated the resulting genes manually, and drafted the manuscript. UJ participated in the design and interpretation of the study, and contributed to writing the manuscript. KV co-led the conception of and supervised this study, and contributed to the manuscript. KM conceived the study and contributed to writing the manuscript. SF provided the analysis, gene-lists and visualizations from data obtained after the ANOVA test and contributed to writing the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

We would like to thank Betsy Read and the JGI for giving us the opportunity to work with sequences of *E. huxleyi* CCMP1516 before they were submitted. This project was partly funded by Marine Genomics Europe (EU Contract n° 505403) and by the Alfred Wegener Institute for Polar and Marine Research.

References

1. **Balch WM, Holligan PM, Kilpatrick KA:** Calcification, photosynthesis and growth of the bloom-forming coccolithophore, *Emiliana huxleyi*. *Continental Shelf Research* 1992, 12(12):1353-1374.
2. **Brown CW, Yoder JA:** Blooms of *Emiliana huxleyi* (Prymnesiophyceae) in surface waters of the Nova Scotian Shelf and the Great Bank. *Journal of Plankton Research* 1993, 15(12):1429-1438.
3. **Winter A, Jordan RW, Roth PH:** Biogeography of living coccolithophores in ocean waters. In: *Coccolithophores*. Edited by Winter A, Siesser WG. Cambridge: Cambridge University Press; 1994: 161-177.
4. **Holligan PM, Fernández E, Aiken J, Balch WM, Boyd P, Burkill PH, Finch M, Groom SB, Malin G, Muller K et al:** A biogeochemical study of the coccolithophore, *Emiliana huxleyi*, in the North Atlantic. *Global Biogeochemical Cycles* 1993, 7(4):879-900.
5. **Sukhanova IN, Flint MV:** Anomalous blooming of coccolithophorids over the Eastern Bering Sea Shelf. *Oceanology* 1998, 38(4):502-505.
6. **Holligan PM, Viollier M, Harbour DS, Camus P, Champagne-Philippe M:** Satellite and ship studies of coccolithophore production along a continental shelf edge. *Nature* 1983, 304(5924):339-342.
7. **Balch WM, Holligan PM, Ackleson SG, Voss KJ:** Biological and optical properties of mesoscale coccolithophore blooms in the Gulf of Maine. *Limnology and Oceanography* 1991, 36(4):629-643.
8. **Malin G, Liss PS, Turner SM:** Dimethyl sulphide: production and atmospheric consequences. In: *The Haptophyte Algae*. Edited by Green JC, Leadbeater BSC. Oxford: Clarendon Press; 1994: 303–320.
9. **Westbroek P, Brown CW, Bleijswijk Jv, Brownlee C, Brummer GJ, Conte M, Egge J, Fernandez E, Jordan R, Knappertsbusch M et al:** A model system approach to biological climate forcing. The example of *Emiliana huxleyi*. *Global and Planetary Change* 1993, 8(1-2):27-46.
10. **Westbroek P, Hinte JEv, Brummer G-J, Veldhuis M, Brownlee C, Green JC, Harris R, Heimdal BR:** *Emiliana huxleyi* as a key to biosphere-geosphere interactions. In: *The Haptophyte Algae*. Edited by Green JC,

- Leadbeater BSC, vol. Systematics Association Special Volume No. 51. Oxford: Clarendon Press; 1994: 321-334.**
11. **Nguyen B, Bowers RM, Wahlund TM, Read BA:** Suppressive subtractive hybridization of and differences in gene expression content of calcifying and noncalcifying cultures of *Emiliana huxleyi* strain 1516. *Applied and Environmental Microbiology* 2005, 71(5):2564-2575.
 12. **Jacquet S, Heldal M, Iglesias-Rodriguez D, Larsen A, Wilson W, Bratbak G:** Flow cytometric analysis of an *Emiliana huxleyi* bloom terminated by viral infection. *Aquatic Microbial Ecology* 2002, 27(2):111-124.
 13. **Bratbak G, Egge JK, Heldal M:** Viral mortality of the marine alga *Emiliana huxleyi* (Haptophyceae) and termination of algal blooms. *Marine Ecology Progress Series* 1993, 93:39-48.
 14. **Castberg T, Thyrhaug R, Larsen A, Sandaa R-A, Heldal M, Van Etten JL, Bratbak G:** Isolation and characterization of a virus that infects *Emiliana huxleyi* (Haptophyta). *Journal of Phycology* 2002, 38(4):767-774.
 15. **Wilson WH, Tarran GA, Schroeder D, Cox M, Oke J, Malin G:** Isolation of viruses responsible for the demise of an *Emiliana huxleyi* bloom in the English Channel. *J Mar Biolog Assoc UK* 2002, 82(03):369-377.
 16. **Schroeder DC, Oke J, Malin G, Wilson WH:** Coccolithovirus (Phycodnaviridae): Characterisation of a new large dsDNA algal virus that infects *Emiliana huxleyi*. *Archives of Virology* 2002, 147(9):1685-1698.
 17. **Allen MJ, Schroeder DC, Wilson WH:** Preliminary characterisation of repeat families in the genome of EhV-86, a giant algal virus that infects the marine microalga *Emiliana huxleyi*. *Archives of Virology* 2006, 151(3):525-535.
 18. **Schroeder DC, Oke J, Hall M, Malin G, Wilson WH:** Virus succession observed during an *Emiliana huxleyi* bloom. *Applied and Environmental Microbiology* 2003, 69(5):2484-2490.
 19. **Martínez JM, Schroeder DC, Larsen A, Bratbak G, Wilson WH:** Molecular Dynamics of *Emiliana huxleyi* and Cooccurring Viruses during Two Separate Mesocosm Studies. *Applied and Environmental Microbiology* 2007, 73(2):554-562.

20. **Allen MJ, Martinez-Martinez J, Schroeder DC, Somerfield PJ, Wilson WH:** Use of microarrays to assess viral diversity: from genotype to phenotype. *Environmental Microbiology* 2007, 9(4):971-982.
21. **Allen MJ, Schroeder DC, Holden MTG, Wilson WH:** Evolutionary History of the Coccolithoviridae. *Molecular Biology and Evolution* 2006, 23(1):86-92.
22. **Wilson WH, Schroeder DC, Allen MJ, Holden MTG, Parkhill J, Barrell BG, Churcher C, Hamlin N, Mungall K, Norbertczak H et al:** Complete Genome Sequence and Lytic Phase Transcription Profile of a Coccolithovirus. *Science* 2005, 309(5737):1090-1092.
23. **Bidle KD, Haramaty L, Barcelos e Ramos J, Falkowski P:** Viral activation and recruitment of metacaspases in the unicellular coccolithophore, *Emiliana huxleyi*. *Proceedings of the National Academy of Sciences* 2007, 104(14):6049-6054.
24. **Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME et al:** Copy number variation: new insights in genome diversity. *Genome Research* 2006, 16(8):949-961.
25. **Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W et al:** Global variation in copy number in the human genome. *Nature* 2006, 444(7118):444-454.
26. **Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J et al:** Strong association of de novo copy number mutations with autism. *Science* 2007, 316(5823):445-449
27. **Watanabe T, Murata Y, Oka S, Iwahashi H:** A new approach to species determination for yeast strains: DNA microarray-based comparative genomic hybridization using a yeast DNA microarray with 6000 genes. *Yeast* 2004, 21(4):351-365.
28. **Taboada EN, Acedillo RR, Luebbert CC, Findlay WA, Nash JHE:** A new approach for the analysis of bacterial microarray-based Comparative Genomic Hybridization: insights from an empirical study. *BMC Genomics* 2005, 6:78.
29. **Maydan JS, Flibotte S, Edgley ML, Lau J, Selzer RR, Richmond TA, Pofahl NJ, Thomas JH, Moerman DG:** Efficient high-resolution deletion

- discovery in *Caenorhabditis elegans* by array comparative genomic hybridization. **In.**; 2007.
30. **Gibson G:** Microarrays in ecology and evolution: a preview. *Molecular Ecology* 2002, 11:17-24.
 31. **Riley M, Serres MH:** Interim report on genomics of *Escherichia coli*. *Annual Review of Microbiology* 2000, 54:341-411.
 32. **Pearson BM, Pin C, Wright J, I'Anson K, Humphrey T, Wells JM:** Comparative genome analysis of *Campylobacter jejuni* using whole genome DNA microarrays. *FEBS Letters* 2003, 554(1-2):224-230.
 33. **Medlin LK, Barker GLA, Campbell L, Green JC, Hayes PK, Marie D, Wrieden S, Vaultot D:** Genetic characterisation of *Emiliana huxleyi* (Haptophyta). *Journal of Marine Systems* 1996, 9(1-2):13-31.
 34. **Iglesias-Rodriguez MD, Saez AG, Groben R, Edwards KJ, Batley J, Medlin LK, Hayes PK:** Polymorphic microsatellite loci in global populations of the marine coccolithophorid *Emiliana huxleyi*. *Molecular Ecology Notes* 2002, 2(4):495-497.
 35. **Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer ELL:** The Pfam Protein Families Database. *Nucleic Acids Research* 2002, 30(1):276-280.
 36. **MacFarlane R, Bhattacharya D, Singh U:** Genomic DNA microarrays for *Entamoeba histolytica*: Applications for use in expression profiling and strain genotyping. *Experimental Parasitology* 2005, 110:196-202.
 37. **Salama N, Guilemin K, McDaniel TK, Sherlock G, Tompkins L, Falkow S:** A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *PNAS* 2000, 97:14668-14673.
 38. **Kato-Maeda M, Rhee JT, Gingeras TR, Salamon H, Drenkow J, Smittipat N, Small PM:** Comparing genomes within the species *Mycobacterium tuberculosis*. *Genome Research* 2001, 11(4):547-554.
 39. **Broekhuijsen M, Larsson P, Johansson A, Bystrom M, Eriksson U, Larsson E, Prior RG, Sjostedt A, Titball RW, Forsman M:** Genome-Wide DNA Microarray Analysis of *Francisella tularensis* Strains Demonstrates Extensive Genetic Conservation within the Species but Identifies Regions

- That Are Unique to the Highly Virulent *F. tularensis* subsp. *tularensis*. *J Clin Microbiol* 2003, 41(7):2924-2931.
40. Thierstein HR, Geitzenauer KR, Molfino B, Shackleton NJ: Global synchronicity of late Quarternary coccolith datum levels: Validation by oxygen isotopes. *Geology* 1977, 5:400-404.
 41. Bratbak G, Levasseur M, Michaud S, Cantin G, Fernández E, Heimdal BR, Haldal M: Viral activity in relation to *Emiliana huxleyi* blooms: a mechanism of DMSP release? *Marine Ecology Progress Series* 1995, 128:133-142.
 42. Frada M, Probert I, Allen MJ, Wilson WH, de Vargas C: The "The Cheshire Cat" escape strategy of the coccolithophore *Emiliana huxleyi* in response to viral infection. *Proceedings of the National Academy of Sciences* 2008, 105(41):15944-15949.
 43. Thyrrhaug R, Larsen A, Brussaard CPD, Bratbak G: Cell cycle dependent virus production in marine phytoplankton. *Journal of Phycology* 2002, 38(2):338-343.
 44. Fuhrman J: Genome sequences from the sea. *Nature* 2003, 424(6952):1001-1002.
 45. Petrov DA: Evolution of genome size: new approaches to an old problem. *Trends in Genetics* 2001, 17(1):23-28.
 46. Saez AG, Probert I, Geisen M, Quinn P, Young JR, Medlin LK: Pseudocryptic speciation in coccolithophores. *PNAS* 2003, 100(12):7163-7168.
 47. Norkin LC: Virus receptors: implications for pathogenesis and the design of antiviral agents. *Clin Microbiol Rev* 1995, 8(2):293-315.
 48. Baranowski E, Ruiz-Jarabo CM, Domingo E: Evolution of Cell Recognition by Viruses. *Science* 2001, 292(5519):1102-1105.
 49. Ejrnaes M, Filippi CM, Martinic MM, Ling EM, Togher LM, Crotty S, von Herrath MG: Resolution of a chronic viral infection after interleukin-10 receptor blockade. *J Exp Med* 2006.
 50. Erbar S, Diederich S, Maisner A: Selective receptor expression restricts Nipah virus infection of endothelial cells. *Virology Journal* 2008, 5(1):142.

51. **Ren RB, Costantini F, Gorgacz EJ, Lee JJ, Racaniello VR:** Transgenic mice expressing a human poliovirus receptor: a new model for poliomyelitis. *Cell* 1990, 63:353-362.
52. **Resnick D, Pearson A, Krieger M:** The SRCR superfamily - A family reminiscent of the Ig superfamily. *Trends Biochem Sci* 1994, 19:5-8.
53. **Manning G, Plowman GD, Hunter T, Sudarsanam S:** Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci* 2002, 27:514-520.
54. **Monick MM, Staber JM, Thomas KW, Hunninghake GW:** Respiratory Syncytial Virus Infection Results in Activation of Multiple Protein Kinase C Isoforms Leading to Activation of Mitogen-Activated Protein Kinase. *J Immunol* 2001, 166(4):2681-2687.
55. **Pickart CM, Eddins MJ:** Ubiquitin: structures, functions, mechanisms. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 2004, 1695(1-3):55-72.
56. **Woelk T, Sigismund S, Penengo L, Polo S:** The ubiquitination code: a signalling problem. *Cell Division* 2007, 2(1):11.
57. **Quinn P, Bowers RM, Zhang YY, Wahlund TM, Fanelli MA, Olszova D, Read BA:** cDNA Microarrays as a Tool for Identification of Biomineralization Proteins in the Coccolithophorid *Emiliana huxleyi* (Haptophyta). *Applied and Environmental Microbiology* 2006, 72(8):5512-5526.
58. **Marsh ME:** Biochemistry and Molecular Biology : Regulation of CaCO₃ formation in coccolithophores. *Comparative Biochemistry and Physiology Part B* 2003, 136(4):743-754.
59. **Paasche E:** A review of the coccolithophorid *Emiliana huxleyi* (Prymnesiophyceae), with particular reference to growth, coccolith formation, and calcification-photosynthesis interactions. *Phycologia* 2002, 40(6):503-529.
60. **Ozaki N, Sakuda S, Nagasawa H:** A novel highly acidic polysaccharide with inhibitory activity on calcification from the calcified scale "coccolith" of a coccolithophorid alga, *Pleurochrysis haptoneoformis*. *Biochemical and Biophysical Research Communications* 2007, 357(4):1172-1176.

61. **Young JR:** Functions of coccoliths. **In: *Coccolithophores*. Edited by Winter A, Siesser WG. Cambridge: Cambridge University Press; 1994: 63-82.**
62. **Wahlund TM, Hadaegh AR, Clark R, Nguyen B, Fanelli M, Read BA:** Analysis of Expressed Sequence Tags from Calcifying Cells of Marine Coccolithophorid (*Emiliana huxleyi*). ***Marine Biotechnology* 2004, 6(3):278-290.**
63. **Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M *et al*:** TM4: a free, open-source system for microarray data management and analysis. ***BioTechniques* 2003, 34(2):374-378.**
64. **Hollander M, Wolfe DA:** Nonparametric statistical inference. **New York: John Wiley & Sons; 1973.**
65. **Sneath PHA:** The application of computers to taxonomy. ***Journal of General Microbiology* 1957, 17:201-226.**
66. **Ward JHj:** Hierarchical Grouping to Optimize an Objective Function. ***J Am Stat Assoc* 1963, 58(236-244).**

Figure legends

Figure 1 - Similarity plot of the 15 tested *E. huxleyi* strains and the two out-groups *G. oceanica* and *I. galbana* compared to the reference (*E. huxleyi* CCMP1516).

E. huxleyi strains and out-groups plotted in a 2D-map of amounts of log₂-ratios above 0.48696 (x-axis) and below -0.8270 (y-axis), respectively. The number of the log₂-ratios above the threshold of 0.48696 indicates increased copy number. The number of the log₂-ratios below the threshold of -0.8270 indicates reduced copy number, deletion, or low homology to the reference strain (*E. huxleyi* CCMP1516). The closer a strain or species is to the origin of the graph the higher the similarity to the reference strain.

Figure 2 - Dendrogram of all 16 *E. huxleyi* strains and the two out-groups *G. oceanica* and *I. galbana* based on the CGH data for all genes derived from the ANOVA analysis.

The dendrogram was produced by taking Euclidean distances of the CGH data (log₂ ratios) in a neighbor-joining tree (n = 1.000 bootstrap iterations) in R with the ape-package.

Figure 3 - Heat map of the Wilcoxon test analysis concerning virus susceptibility of the 15 *E. huxleyi* strains.

Strains were grouped according to their virus susceptibility (Tab. 1). Significance was determined as p-value < 0.01. From the resulting log₂-ratios of these significant signals a heat map was generated by hierarchical clustering. The heat map identifies overall high (white) and low (red) signal intensities. Virus-susceptible strains are indicated as green leafs in the hierarchical clustering above the heat map. The reference strain *E. huxleyi* CCMP1516 was excluded in the analysis but drawn in the heat map to visualize the variation in log₂-self-self-ratios.

Figure 4 - Heat map of the Wilcoxon test analysis concerning the morphology/morphotypes of the 15 *E. huxleyi* strains.

Strains were grouped according to their morphology (Tab. 3). From the resulting log₂-ratios of these significant signals a heat map was generated by hierarchical clustering. The heat map identifies overall high (white) and low (red) signal intensities. Strains which possess coccoliths are indicated as green leafs in the hierarchical clustering above the heat map. The reference strain *E. huxleyi* CCMP1516 was excluded in the analysis but drawn in the heat map as described in figure 3.

Tables

Table 1 - Virus susceptibility and resistance of *E. huxleyi* strains derived from Allen *et al.* [20]

<i>Emiliana huxleyi</i> host strain	<i>Emiliana huxleyi</i> virus (EhV) strain								
	86	84	88	163	201	205	202	208	207
92 (English Channel)	-	-	-	-	-	-	-	-	-
92A (English Channel)	-	-	-	-	-	-	-	-	-
92D (English Channel)	-	-	-	-	-	-	-	-	-
92E (English Channel)	-	-	-	-	-	-	-	-	-
92F (English Channel)	+	+	+	+	+	+	+	+	+
CCMP379 (Unknown)	-	-	-	-	-	-	-	-	-
CCMP374 (Gulf of Maine)	+	+	+	+	+	+	+	+	+
CCMP373 (Sargasso Sea)	-	-	-	-	-	-	-	-	-
12-1 (Sargasso Sea)	-	-	+	+	+	+	+	+	+
CCMP1516 (South Pacific)	+	+	+	-	+	+	+	+	+
Van 556 (North Pacific)	-	-	-	-	-	-	-	-	-
CH 24/90 (North Atlantic)	-	-	-	-	-	-	-	+	-
CH 25/90 (North Atlantic)	-	-	-	-	-	-	-	-	-
L (Oslo Fjord)	+	+	+	+	+	-	+	+	+
NZEH (South Pacific)	-	-	-	-	-	-	-	-	-
EH2 (South Pacific)	+	+	+	-	+	+	+	+	+

+, culture lysis; -, no evidence of lysis after 14 days of viral infection cultures were not lysed and considered to be non-susceptible to the virus strain [20].

Table 2 - Antibiotic treatment mixture

Antibiotic	Concentration in culture [mg/mL]
Ampicillin	0.05
Gentamycin	0.003
Streptomycin	0.025
Chloramphenicol	0.001
Ciprofloxacin	0.010

Table 3 - Isolation sites and morphology of the 16 *E. huxleyi* strains

<i>Emiliana huxleyi</i> strain	coccoliths	Collection site
92 (English Channel)	N	49°19N 07°26W
	N	50.1669N 4.2504W 1mile west of the
92A (English Channel)		Eddystone
92D (English Channel)	Y	50°02N 4°22W
92E (English Channel)	Y	49°52N 06°12W 2m depth
92F (English Channel)	Y	49°52N 06°12W 2m depth
	N	50.1669N 4.2504W 1mile west of the
CCMP379 (= 92A)		Eddystone
CCMP374 (Gulf of Maine)	N	42.5000N 69.0000W Gulf of Maine (5 meters)
CCMP373 (Sargasso Sea)	N	32.1667N 64.5000W
12-1 (Sargasso Sea)	Y	32.0000N 62.0000W (50 meter depth)
CCMP1516 (South Pacific)	Y	02.6667S 82.7167W (surface)
Van 556 (North Pacific)	N	49°05N 144°40W
CH 24/90 (North Atlantic)	Y	57°20N 01°09E
CH 25/90 (North Atlantic)	Y	57°26N 6°13E
L (Oslo Fjord)	N	60°N 11°E
NZEH (South Pacific)	Y	Big Glory Bay, NZ
EH2 (South Pacific)	Y	Great Barrier ,Reef

Y, strain possesses coccoliths; N, strain has no coccoliths.

Supplementals**Table S1 – Identified genes of *E. huxleyi* by a Wilcoxon-Test in respect to virus susceptibility and virus resistance.**

Excel-file including array, gene and protein- ID, html-link to the genome website of the reference strain *E. huxleyi* CCMP1516, description and function of the identified genes and the log₂-ratios means of the 16 *E. huxleyi* strains.

Table S2 – Identified genes of *E. huxleyi* by a Wilcoxon-Test in respect the existence of coccoliths.

Excel-file including array, gene and protein- ID, html-link to the genome website of the reference strain *E. huxleyi* CCMP1516, description and function of the identified genes and the log₂-ratios means of the 16 *E. huxleyi* strains.

Figure 1

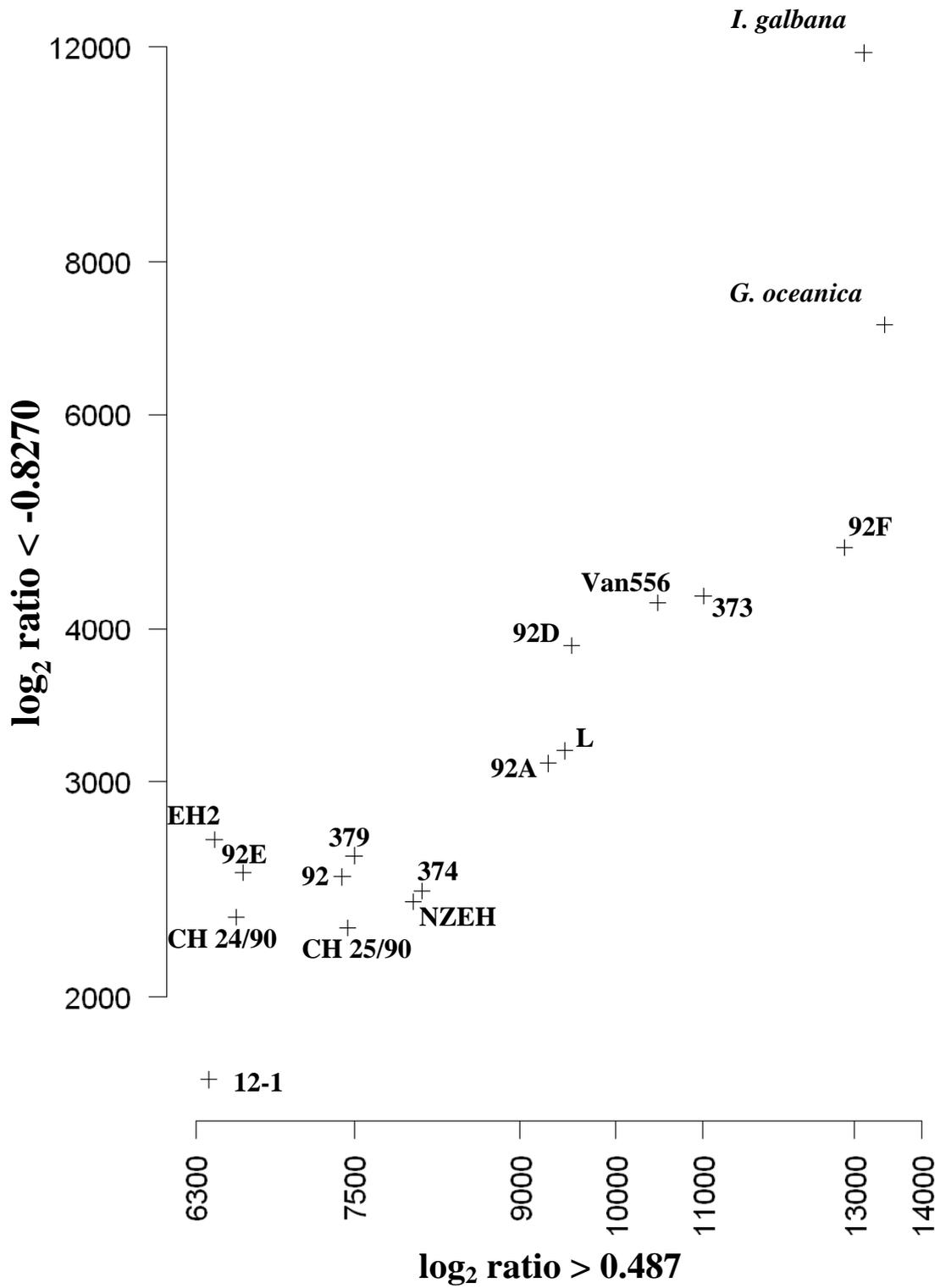


Figure 2

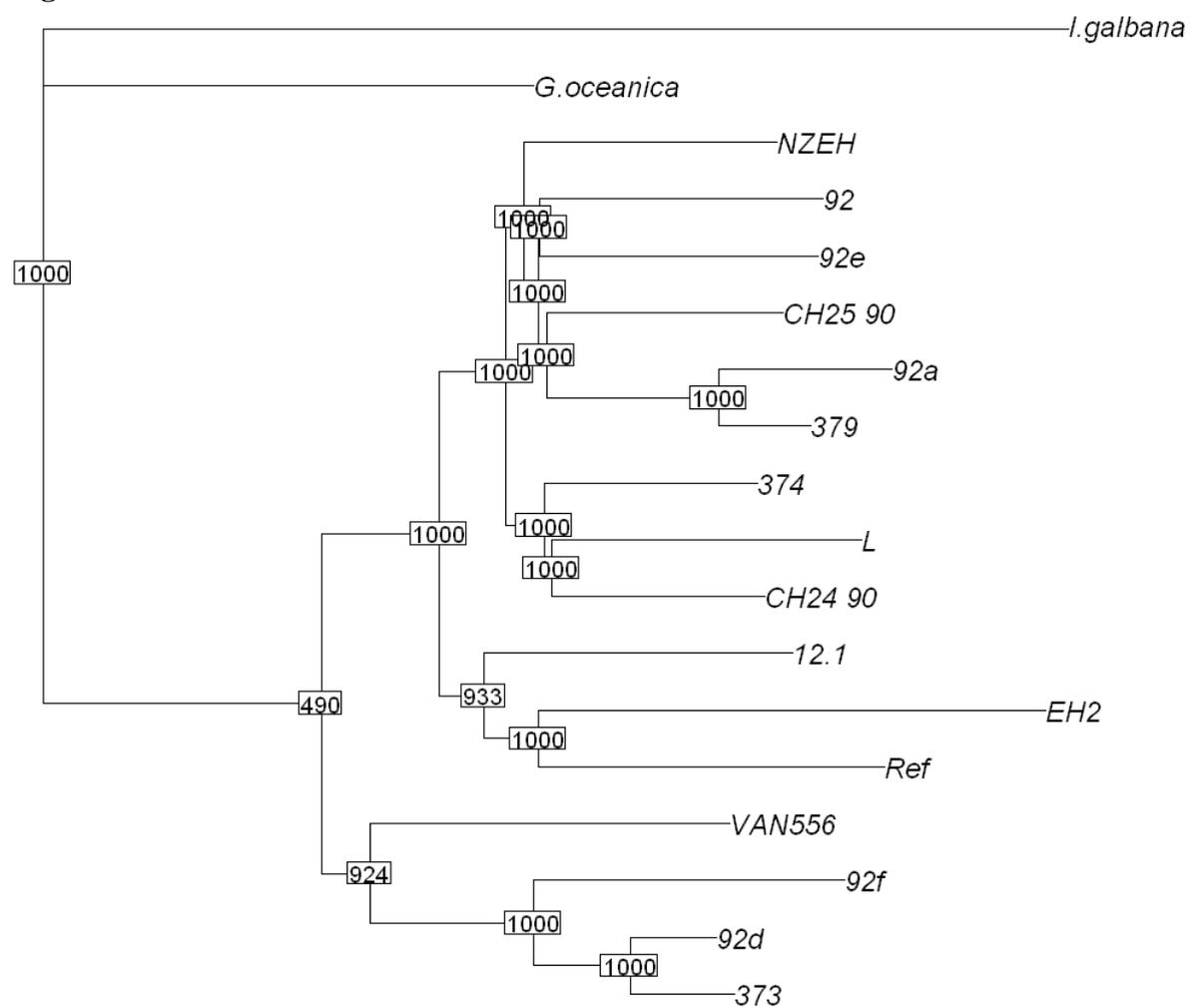
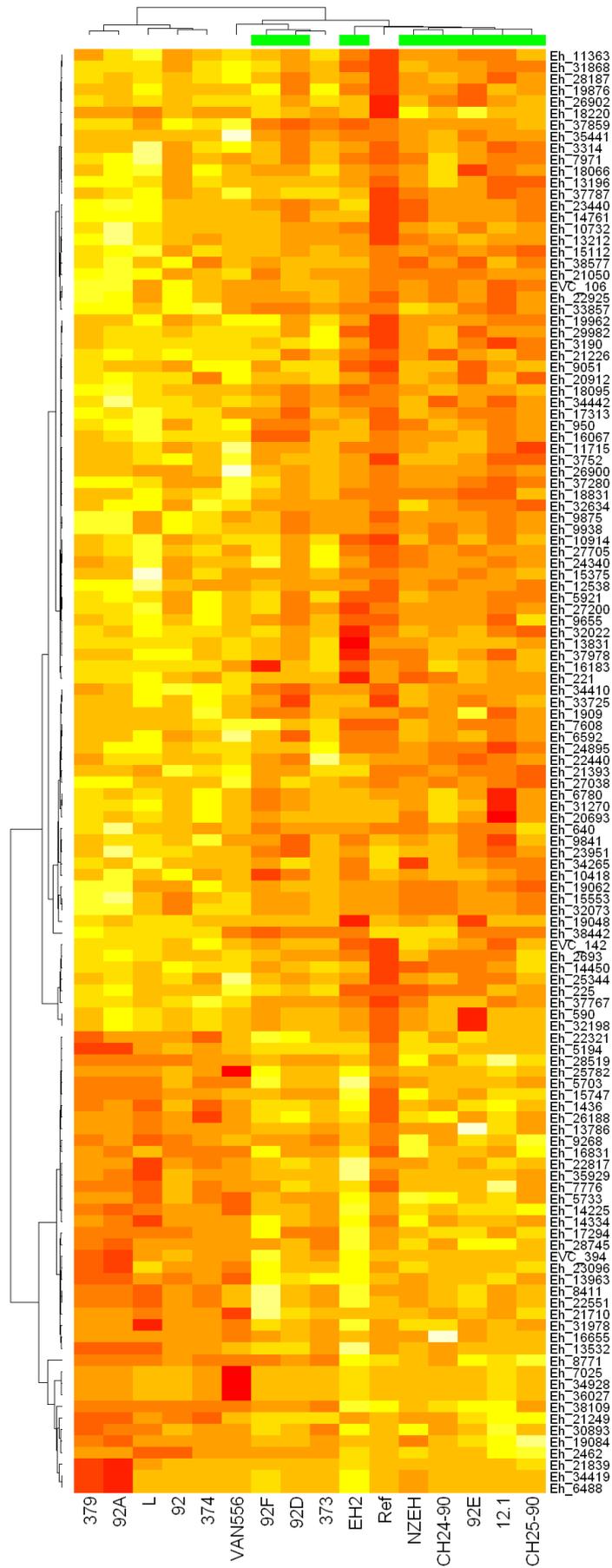


Figure 4



3 GENERAL DISCUSSION

The publications presented in this thesis focused on the identification of genes involved in the virus infection of *Emiliana huxleyi* by taking advantage of EST libraries and DNA microarrays. Furthermore, genomic differences of several *E. huxleyi* strains from different geographic origin were determined by comparative genomic hybridizations concerning virus susceptibility and morphology. Experiments yielded in a diverse range of new information on *E. huxleyi*. Perspectives for future research that arise from this thesis are given at the end.

3.1 EFFECTS OF VIRAL INFECTION ON *EMILIANA HUXLEYI*

Emiliana huxleyi is of great importance for nutrient and biogeochemical cycles of sulphur and carbon in the marine environment due to its ability to form immense blooms. Lytic viral infection of this alga leads to termination of blooms and therefore influences the global climate. For this reason it was of particular interest to get more knowledge about this complex alga and especially into the host-virus interaction.

Publication I provides the first cDNA libraries of a host-virus interaction in the marine microalgae community, a first view into the gene expression throughout a virus infection of *E. huxleyi* by taking advantage of ESTs and the basis for **publication II**. The results of **publication I** and **II** identified many genes involved in

the host's response to viral infection. The differential expression of certain genes during viral infection suggests their involvement in the interaction between the host and the virus. This makes them suitable targets for further investigation. **Publication I** and **II** investigated EST libraries from *E. huxleyi* during viral infection. Three cDNA libraries generated 6, 12 and 24 hours post infection with EhV-86 (in this thesis referred to as T6, T12 and T24) were compared to a library from an uninfected culture (T0) by sequencing, clustering and manual annotation. Since the genome sequences for both, the virus and the host were available, it was possible to determine the precise number of ESTs from both in all four libraries, respectively. Results of **publication I** and **II** provide insights into the infection mechanisms of the virus EhV-86 in *E. huxleyi*. Furthermore, changes in gene expression levels, assessed through two-colour oligonucleotide microarrays, were compared with the gene frequency in the EST libraries (**publication II**). The use of ESTs coupled with microarray analysis has shown that it is a powerful tool to study gene expression of an organism under different conditions.

The results of **publication I** show that by the sequencing of only around 90 ESTs per library a trend was already recognizable. The genes of *E. huxleyi* involved in photosynthesis were down-regulated for the benefit of an increased transcription and translation. Less than 5% ESTs were of viral origin at 6 (T6) and 12 (T12) hours post viral infection. In addition, at 24 hours (T24) post infection only 10% of the mRNA was of host origin. **Publication II** confirmed the tendencies of **publication I** and discovered more genes involved in the host's response to viral infection. The results demonstrated the power of EST libraries and DNA microarrays to obtain data on gene expression and regulation during viral infection.

The most prominent effect of the viral infection on the host transcriptome is the change in the expression of genes involved in photosynthesis, transcription and translation, glycolysis, fatty acid metabolism, and protein degradation (Fig. 1). For viral replication the biosynthetic machinery of the host cell must be modified for the benefit of viral proteins. The virus drastically changes host transcription already after 6 hours, when only 38 (3.9%) viral ESTs are present (**publication II**). Hence, the up-regulation of genes related to protein and ATP synthesis in the mitochondria, genes involved in glycolysis and fatty acid synthesis, and genes involved in RNA-synthesis and translation of the host could be an indication for a boost in the expression of viral

proteins which require large amounts of energy. Another indication for enhanced viral protein expression is the down-regulation of host genes involved in photosynthesis, protein degradation, signalling, pigment synthesis, RNA processing, the citric acid cycle, and protein import to the mitochondria. In general, both organelles of *E. huxleyi* seem to be down-regulated for the benefit of an increased transcription and translation used for viral replication.

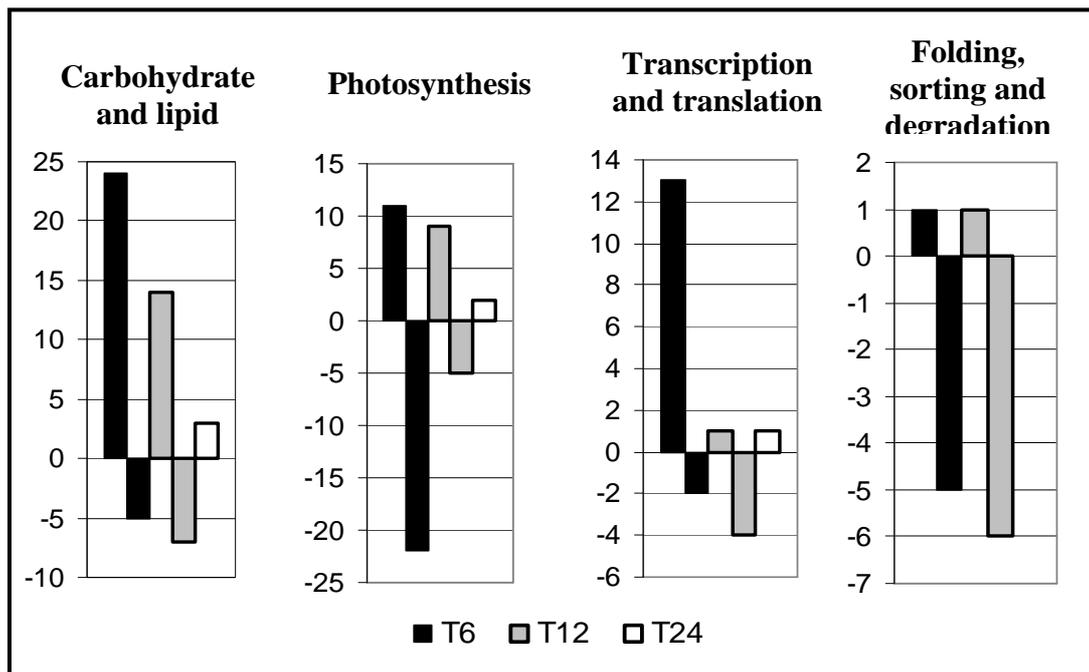


Figure 1. Differential expression of genes involved in carbohydrate and lipid metabolism (including glycolysis and fatty acid metabolism), photosynthesis, transcription and translation, and folding, sorting and degradation during viral infection. Positive values of barplots indicates the number of significantly up-regulated genes, whereas negative values shows the down-regulated genes at 6 (black bars), 12 (grey bars) and 24 hours (white bars) post infection in comparison to a healthy culture.

The break-down of photosynthesis is of particular importance to photosynthetic organisms. During viral infection of *E. huxleyi*, EST and microarray results revealed significant reduction of genes involved in photosynthesis. Prior infection 17.2% of ESTs were related to genes involved in photosynthesis. After 6, 12 and 24 hours viral infection the percentage of photosynthesis-related ESTs decreased to 2.5, 3.3, and 0.5%, respectively. The observation of a reduction of photosynthesis

during viral infection has been previously reported for *E. huxleyi* cultures by measuring photochemical capacity (F_V/F_M), carotenoids and chlorophyll composition, and intracellular reactive oxygen species (ROS) production (Evans *et al.*, 2006, Llewellyn *et al.*, 2007).

ESTs of *E. huxleyi* related to transcription and translation increased from 3.4% at T0 to 25.0% at T6 and 12.4% at T12 suggesting a favoured synthesis of viral proteins. The changes of host transcription already after six hours viral infection, when only 38 (3.9%) viral ESTs are present, could be indicated by the occurrence of a putative viral DNA-dependent RNA polymerase II subunit (EHV064). RNA polymerase II is essential for transcription of viral genes. In addition, at T6 a putative viral protease (EHV349) was identified which might be responsible for the degradation of host proteins. The down-regulation of host genes involved in protein degradation at T6 and T12 such as ubiquitin indicate the takeover of the host machinery. A viral HNH endonuclease family protein (EHV093) was found at T12 which could lead to the degradation of host DNA. Furthermore, a putative viral proliferating cell nuclear antigen (EHV440) was identified and which could help in DNA binding during transcription of viral genes. The occurrence of a putative viral fatty acid desaturase (EHV061) could be involved in membrane lysis, i.e. the lysis of host cells.

A previous study has shown that host cells release virus particles between 4 and 48 h after infection while remaining intact (Allen & Wilson, 2006). Furthermore, it has been demonstrated that coexistence of host and virus is possible and that both can replicate during infection (Thyrhaug *et al.*, 2003). Supporting these findings, microarray results showed an up-regulation of several genes involved in photosynthesis during viral infection, which indicates, that at least some cells were intact and perform photosynthesis. In addition, the low abundance of viral ESTs at T12 (2.3%) indicates that perhaps many host cells remained intact and only a few infections took place. However, the change to 89.2% viral ESTs at T24 suggests that between 12 and 24 hours post infection the virus took over the transcriptional machinery of the entire *E. huxleyi* population.

At T24, a putative viral major capsid protein (EHV085) whose function is well known and defined in viral systems (Allen *et al.*, 2008), and a putative viral DNA ligase (EHV158) which could be used for DNA replication were identified.

These two genes could be involved in the packaging and the following release of new viruses. One copy of a Longevity-assurance (LAG1) family protein (EHV014) was identified which is involved in the ceramide synthesis. Ceramide can act in regulating apoptosis suggesting that LAG1 could be involved in the lysis of host cells. Another indication for the releasing of new viruses could be the up-regulation of host genes involved in the exchange with vacuoles such as Vacuolar ATP synthase catalytic subunit A and V-ATPase subunit d at T6 and T12. These suggestions are in accordance with previous studies (Wilson *et al.*, 2005, Allen *et al.*, 2006a, Allen & Wilson, 2006). It was shown that at 4 h post infection viruses just started being released. However, at 33 h post infection the host cell is still in a steady stage of virus releasing and re-infection until the host finally lyses after around 48 h.

- the virus drastically changed the host transcriptome already six hours post infection
- host genes involved in photosynthesis were down-regulated during viral infection
- host genes involved in energy production like glycolysis and fatty acid synthesis were up-regulated during viral infection
- after 24 hours post infection only 10% of the mRNA was of host origin
- between 12 and 24 hours post infection the virus took over the transcriptional machinery of the entire *E. huxleyi* population

3.2 BIODIVERSITY IN SEVERAL STRAINS OF *E. HUXLEYI*

The influence of viruses is well recognized on marine geochemical cycles by regulation of host populations. As stated above *Emiliana huxleyi* plays an important role in global biogeochemical cycles and its blooms are often terminated by viruses. But the algae survives the termination by escaping the virus through life-cycle transition (Frada *et al.*, 2008). In addition, previous studies have reported different genome sizes among different morphotypes of *E. huxleyi* from different geographical

regions via DNA microsatellites and restriction fragment length polymorphism (RFLP) analysis (Medlin *et al.*, 1996, Iglesias-Rodriguez *et al.*, 2002). The same studies revealed the presence of different ecotypes of *E. huxleyi* potentially with differences in genome organization in response to environmental conditions or to potential threats, such as viral infections. Furthermore, an example for a connection between genetic variation and virus susceptibility has been demonstrated (Schroeder *et al.*, 2002). It has been shown that virus resistant strains of *E. huxleyi* display a higher DMSP-lyase activity than strains that are susceptible to virus infection. One reason for the different enzyme activities could be variations in the expression of the gene coding for the enzyme due to either a change in transcriptional regulation or a change of the copy number in the genomes. So far, research was focused on 18S rRNA, microsatellites and a limited number of functional genes. To assess role of ecological diversification, virus susceptibility, and morphology (e.g. formation of coccoliths) in determining intra-species genetic differences, whole genome analysis is required.

Publication III describes the first attempt to apply microarray-based comparative genomic hybridization (CGH) on *E. huxleyi* strains. The goal of **publication III** was the detection of genetic diversity and of genes possibly related to virus susceptibility and morphology of *E. huxleyi* strains from different geographic origin. Genomic DNA of 15 different *E. huxleyi* strains was compared by co-hybridization with the sequenced strain CCMP1516. The two species *Gephyrocapsa oceanica* and *Isochrysis galbana* were taken as out-groups. The relative copy number of each gene transcript was determined by the signal intensity of the two samples described by the log₂-ratio (LR = sample/reference). The results of **publication III** revealed the genetic distance between *E. huxleyi* strains and the two out-groups *G. oceanica* and *I. galbana* based on cluster analysis of the log₂-ratios (LRs). A self-versus-self hybridization was used to determine the threshold for the identification of genetic divergence. Divergent genes concordant with reduced copy numbers, deletion or nucleotide divergence were below the threshold of LR < -0.8270. Gene duplications were indicated by LRs above 0.48696. The genetic distances between strains of *E. huxleyi* are in accordance with previous reports that demonstrated different genome sizes among different morphotypes of *E. huxleyi* from different

geographical regions based on DNA microsatellites and restriction fragment length polymorphism (RFLP) analysis (Medlin *et al.*, 1996, Iglesias-Rodriguez *et al.*, 2002).

CGH has been extensively utilized to elucidate genetic diversity mainly in bacterial systems like *Helicobacter pylori*, *Campylobacter jejuni*, *Entamoeba histolytica*, *Francisella tularensis*, *Mycobacterium tuberculosis*, (Salama *et al.*, 2000, Kato-Maeda *et al.*, 2001, Broekhuijsen *et al.*, 2003, Pearson *et al.*, 2003, MacFarlane *et al.*, 2005) but also in the eukaryotic systems of yeast (Watanabe *et al.*, 2004). Microarray analysis had indicated a limited genetic variation within the species and strains. Strain comparisons showed differential hybridization between 0.17 and 16.7% of the gene transcripts. The genetic diversity increased at most up to 90% within the subspecies. In contrast, the results of **publication III** revealed huge genetic variation between 27% and 57% within the species and up to 69% to the older genus *G. oceanica* (Fig. 2). As *E. huxleyi* has evolved from *G. oceanica* only 268.000 years ago (Thierstein *et al.*, 1977) and became dominant around 70.000 years ago, this high genetic diversity could indicate that *E. huxleyi* is still in its evolutionary radiation.

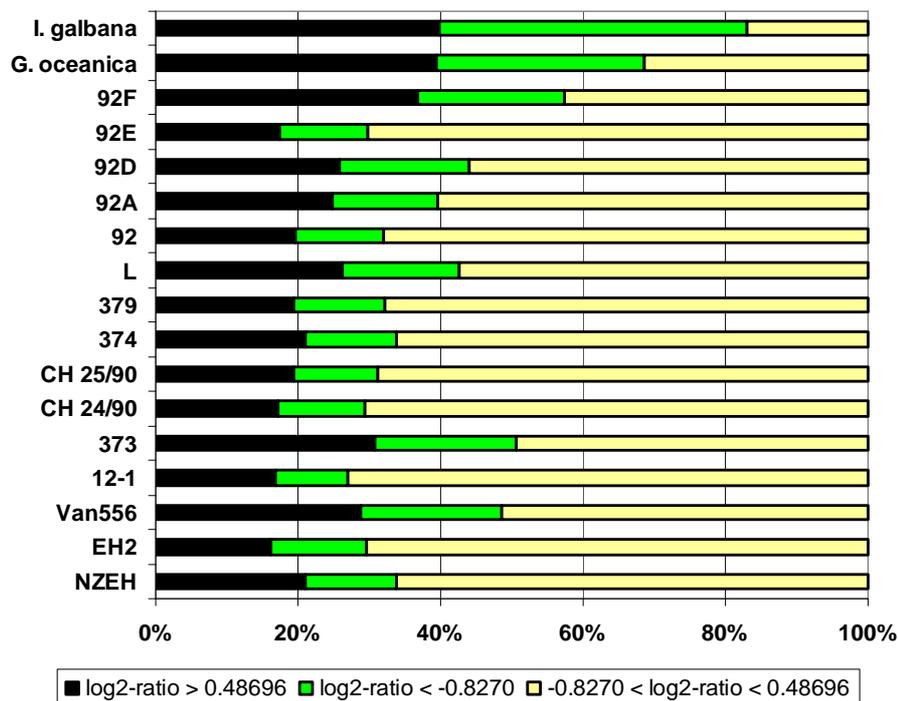


Figure 2. Genetic diversity in per cent of the 15 different *E. huxleyi* strains and the two outgroups *G. oceanica* and *I. galbana* in comparison to the reference strain *E. huxleyi* CCMP1516. The number of the log2-ratios above the threshold of 0.48696 indicates increased copy number. The number of the log2-ratios below the threshold of -0.8270 indicates reduced copy number, deletion, or low homology to the reference strain. Log2-ratios between the two thresholds indicate no significant difference between sample and reference.

The highest genetic variation amongst the strains in comparison to the reference was observed in the case of strain 92F (57.4%). This strain is virus susceptible and possesses the ability to produce coccoliths, both suggesting a higher similarity to the reference strain (CCMP1516). The geographic origin of the strains could be an explanation for the high genomic deviation from the reference. The reference strain was isolated near the coast of Ecuador whereas 92F was obtained from the English Channel. Another strain collected in the English Channel, 92E, showed 70.3% similarity to the reference strain. Both English Channel strains possess coccoliths but show different virus susceptibility. As the reference strain is virus susceptible, the genetic differences between these two strains could not be caused by virus susceptibility. The genomic comparisons support earlier findings of blooms being dominated by a succession of different populations (Bratbak *et al.*, 1995). During a bloom the community composition of *E. huxleyi* is affected by viruses in the role of a population controlling factor. One specific population is decimated by viruses making blooms of succeeding populations possible (Bratbak *et al.*, 1995). Affected populations induce life-cycle transition to a haploid (1n) stage to escape viral infection as the haploid phase has been demonstrated to be resistant to viral infection (Frada *et al.*, 2008). At the time of isolation 1992 during a bloom in the English Channel, the strain 92F was in a haploid stage while strain 92E was in a diploid (2n) stage. This suggests that 92F was escaping viral infection whereas 92E was blooming due to resistance to specific viruses. Reasons for the genetic distance between these two strains could be still differences in the morphotypes, genome size or ecological strategies (Thyrhaug *et al.*, 2002).

So far, the strain 92E is regarded as virus resistant against 9 different EhVs. As viruses are the most abundant biological entities in the ocean (Fuhrman, 1999, Suttle, 2000, Wommack & Colwell, 2000) it is likely that viruses specific for this strain exist but have not been isolated yet.

However, more strains of this rather young species (Saez *et al.*, 2003) should be analyzed before drawing definite conclusions on their genetic variation.

3.2.1 VIRUS SUSCEPTIBILITY OF *EMILIANA HUXLEYI*

To elucidate whether a lack of certain genes, copy number changes or sequence divergence between reference and tester strain may explain the different biological properties of virus susceptibility a Wilcoxon rank-sum test was applied (**publication III**). The test compared the log₂-ratios (LR) between groups of samples to identify genes regarding virus susceptibility. The reference strain (CCMP1516) and the two out-groups *G. oceanica* and *I. galbana* were excluded in this analysis. Strains were grouped according to their virus susceptibility. Prior to analysis, all mean signals that failed to produce a positive result above the threshold of the reference (LR > 0.48696) in at least one of the 16 different *E. huxleyi* strains (excluding *G. oceanica* and *I. galbana*) were removed from the data set. This resulted in 21371 signal combinations. Significance was defined as p-value < 0.01. The resulting genes were manually analysed by BLAST searches against the NCBI non-redundant protein database (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>, NCBI-nr) and the SwissProt database (<http://www.expasy.org/sprot/>) and were compared with Pfam (Bateman *et al.*, 2002). Among others, two membrane receptors were found to be different between the susceptible and the resistant strains: one scavenger receptor protein and one receptor L domain-containing protein. The first step of virus infection involves attachment of virus particles to host-specific cell surface receptors (Norkin, 1995, Baranowski *et al.*, 2001). This prepares the way for the viruses to enter the host cell. Once inside the host cell, viruses utilize the host machinery in order to enhance the efficiency of its replication process. Consequently, the expression of the receptor on the outer surface of the host is a major determinant of the route of entry of the virus into the host and of the patterns of virus spread and pathogenesis in the host (Norkin, 1995). Viruses have evolved to exploit these receptors to gain entry into cells. Previous studies have demonstrated that the lack of receptor expression restrict virus entry (Ren *et al.*, 1990, Ejrnaes *et al.*, 2006, Erbar *et al.*, 2008). The identified scavenger receptor is a transmembrane glycoprotein and reminiscent of members of the immunoglobulin (Ig) superfamily (Resnick *et al.*, 1994). Members of this family are known to be involved in entry of more than one virus into cells ((Baranowski *et al.*, 2001) and references therein). The L domain of the second identified membrane receptor is also found in insulin receptor (IR) which is closely related to members of

the tyrosine-kinase receptor superfamily. Members of this family play a role in different cellular processes, including division, proliferation, apoptosis, and differentiation (Manning *et al.*, 2002). Moreover, it has been shown that protein kinases are activated by viral infection (Monick *et al.*, 2001), suggesting that the L domain-containing membrane receptor could be involved in virus susceptibility or infection. Different virus susceptibility could be due to differences in copy numbers or modifications of these two receptors. Therefore, the identified receptors are suitable targets for further investigations regarding virus susceptibility.

The occurrence of an ubiquitin and an ubiquitin-conjugation enzyme E1 might indicate its involvement in virus infection. **Publication II** combined an EST study with gene expression analysis by using microarrays. Results had indicated the down-regulation of three ESTs related to the ubiquitin protein family and the up-regulation of two of them during viral infection (**publication II**). Ubiquitin and its relatives regulate processes in eukaryotic cells by covalent attachment to other cellular proteins, thereby changing the stability, localization, or activity of the target protein (Pickart & Eddins, 2004). The most prominent function of ubiquitin is the mediated proteolysis of labelled target proteins. Moreover, ubiquitin modifications are also involved in virus budding (Woelk *et al.*, 2007) indicating the importance of ubiquitin and its relatives for virus susceptibility or infection.

It would be of great importance to show that there are significant differences in the degree of variation in the genes associated with ubiquitin. However, more strains should be analyzed and other quantitative methods like qPCR should be applied before drawing definite conclusions on its involvement in virus infection.

- CGH revealed a huge intra-species diversity in *E. huxleyi*
- high genetic diversity between two strains from the same geographic origin suggests difference in morphotypes, genome size or ecological strategy
- CGH made it possible to identify genes in relation to virus susceptibility
- identification of two membrane receptors, possibly playing a key role in virus susceptibility
- identification of proteins related to ubiquitin (also found in the EST study) indicating their possible involvement in virus infection

3.3 PERSPECTIVES OF FUTURE RESEARCH

In summary, the advantages of an EST approach for simultaneous discovery and identification of host and viral genes involved in viral infection have been demonstrated. Moreover, CGH based on microarrays was proven extremely useful for phylogenetic reconstruction and pinpointing single gene differences between closely related strains of *Emiliana huxleyi* with respect to virus susceptibility and morphology, i.e. existence of coccoliths or not.

While the aim of **publication I** was to provide a first insight into the host-virus interaction of *E. huxleyi*, the aim of **publication II** was to broaden the basis of available sequence information. **Publication II** has also shown that complementing this approach with microarray analysis enables the detection of even more subtle changes in gene expression. Viral infection affects the transcriptional machinery of *E. huxleyi* within a few hours by decreasing the expression of genes involved in photosynthesis and protein degradation at the benefit of fatty acid metabolism, glycolysis, and transcription and translation. The expression of *E. huxleyi* and EhV-86 genes changed significantly between 12 and 24 hours after infection, indicating further functional investigations during this infection period. Quantitative RT-PCR (qPCR) could be used to follow the expression of identified genes during the infection period mentioned before with higher sampling resolution (e.g. every hour). Since the infection process is not synchronized between single cells, bulk samples from any given time point have no resolution power for the infection stage on a single cell level. However, up to now little is known about the processes involved in the viral infection of *E. huxleyi*. But the numbers of highly expressed but functionally uncharacterized sequences have the potential of yet unknown proteins relevant in viral infection. Therefore future investigations should regard the quantification of relative transcript abundances by using qPCR to validate microarray analysis. Another consideration would be the study of ratio of variable to mean fluorescence (F_V/F_M) and the effect of abiotic factors like nutrient availability, pH or CO₂ during viral infection.

In **publication III** microarray-based CGH was successfully applied to elucidate genetic diversity among different strains of *E. huxleyi* of different geographic origin. As a strain specific microarray can only tell what kind of genes are

present or not in relation to the genome studied, the construction of an additional microarray of a resistant strain would be appropriate. However, results revealed that up to 57% of the genes showed a pattern of hybridization concordant with deletion, nucleotide divergence or gene duplication within the species compared to the reference strain *E. huxleyi* CCMP1516. One reason of the genetic differences between strains could be differences in the morphotypes and genome size. Therefore, further investigations should regard measurements of genome size by flow cytometry. Moreover, a Wilcoxon rank-sum test was used to compare log₂-ratios between groups of samples to identify genes related virus susceptibility and morphology.

Among others, the two membrane receptors and the ubiquitin-related proteins that possibly play a role in virus infection are suitable targets for further investigations. qPCR can be used to determine the real copy number in each genome. Future work should also include the use of microarrays transcript profiling experiments and knock-out mutants to focus on the expression of the identified key genes. To identify more genes regarding virus susceptibility, gene expression analysis should be considered during viral infection, e.g. after 4 hours viral infection. Furthermore, recent findings revealed a novel virus-escaping strategy of *E. huxleyi* during blooms (Frada *et al.*, 2008). Virus mediated termination of *E. huxleyi* blooms induces life-cycle transition of affected populations. Hence, further investigations should also focus on the life-cycle stage. The proposed further investigation outlined above can extend our understanding of virus susceptibility and viral infection of *E. huxleyi*.

4 REFERENCES

- ALLEN, M.J., FORSTER, T., SCHROEDER, D.C., HALL, M., ROY, D., GHAZAL, P. & WILSON, W.H. (2006a). Locus-specific gene expression pattern suggests a unique propagation strategy for a giant algal virus. *Journal of Virology* **80**(15): 7699-7705.
- ALLEN, M.J., HOWARD, J.A., LILLEY, K.S. & WILSON, W.H. (2008). Proteomic analysis of the EhV-86 virion. *Proteome Science* **6**: 11.
- ALLEN, M.J., MARTINEZ-MARTINEZ, J., SCHROEDER, D.C., SOMERFIELD, P.J. & 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., HOLDEN, M.T.G. & WILSON, W.H. (2006b). Evolutionary History of the Coccolithoviridae. *Molecular Biology and Evolution* **23**(1): 86-92.
- ALLEN, M.J., SCHROEDER, D.C. & WILSON, W.H. (2006c). Preliminary characterisation of repeat families in the genome of EhV-86, a giant algal virus that infects the marine microalga *Emiliana huxleyi*. *Archives of Virology* **151**(3): 525-535.
- ALLEN, M.J. & WILSON, W.H. (2006). The coccolithovirus microarray: an array of uses. *Brief Funct Genomic Proteomic* **5**(4): 273-279.
- ANDREAE, M.O. (1990). Ocean-atmosphere interactions in the global biogeochemical sulfur cycle. *Mar Chem* **30**(1-29).

- BALCH, W.M., HOLLIGAN, P.M., ACKLESON, S.G. & VOSS, K.J. (1991). Biological and optical properties of mesoscale coccolithophore blooms in the Gulf of Maine. *Limnology and Oceanography* **36**(4): 629-643.
- BALCH, W.M., HOLLIGAN, P.M. & KILPATRICK, K.A. (1992). Calcification, photosynthesis and growth of the bloom-forming coccolithophore, *Emiliana huxleyi*. *Continental Shelf Research* **12**(12): 1353-1374.
- BARANOWSKI, E., RUIZ-JARABO, C.M. & DOMINGO, E. (2001). Evolution of Cell Recognition by Viruses. *Science* **292**(5519): 1102-1105.
- BATEMAN, K.-H., Birney, E., CERRUTI, L., DURBIN, R., ETWILLER, L., EDDY, S.R., GRIFFITHS-JONES, S., HOWE, K.L., MARSHALL, M & SONNHAMMER, E.L.L. (2002). The Pfam Protein Families Database. *Nucleic Acids Research* **30**(1): 276-280.
- BAUMANN, K.-H., BÖCKEL, B. & FRENZ, M. (2004). Coccolith contribution to South Atlantic carbonate sedimentation. In *Coccolithophores: From Molecular Processes to Global Impact* (THIERSTEIN, H.R. & YOUNG, J.R. editors), pp. 367-402. Springer, Berlin Heidelberg New York.
- BIDLE, K.D., HARAMATY, L., BARCELOS E RAMOS, J. & FALKOWSKI, P. (2007). Viral activation and recruitment of metacaspases in the unicellular coccolithophore, *Emiliana huxleyi*. *Proceedings of the National Academy of Sciences* **104**(14): 6049-6054.
- BILLARD, C. (1994). Life cycles. In *The Haptophyte Algae* (GREEN, J.C. & LEADBEATER, B.S.C. editors), pp. 167-186. Clarendon Press, Oxford.
- BONALDO, M.F., LENNON, G. & SOARES, M.B. (1996). Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Research* **6**(9): 791-806.
- BOUCK, A. & VISION, T. (2007). The molecular ecologist's guide to expressed sequence tags. *Molecular Ecology* **16**: 907-924.
- BOWLER, C., ALLEN, A.E., BADGER, J.H., GRIMWOOD, J., JABBARI, K., KUO, A., MAHESWARI, U., MARTENS, C., MAUMUS, F., OTILLAR, R.P., *et al.* (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**(7219): 239-244.
- BOWN, P.R., LEES, J.A. & YOUNG, J.R. (2004). Calcareous nannoplankton evolution and diversity through time. In *Coccolithophores: From Molecular*

- Processes to Global Impact* (THIERSTEIN, H.R. & YOUNG, J.R. editors), pp. 481-508. Springer, Berlin Heidelberg New York.
- BRAND, L.E. (1994). Physiological ecology of marine coccolithophores. In *Coccolithophores* (WINTER, A. & SIESSER, W.G. editors), pp. 39-49. Cambridge University Press, Cambridge.
- BRATBAK, G., EGGE, J.K. & HELDAL, M. (1993). Viral mortality of the marine alga *Emiliania huxleyi* (Haptophyceae) and termination of algal blooms. *Marine Ecology Progress Series* **93**: 39-48.
- BRATBAK, G., LEVASSEUR, M., MICHAUD, S., CANTIN, G., FERNÁNDEZ, E., HEIMDAL, B.R. & HELDAL, M. (1995). Viral activity in relation to *Emiliania huxleyi* blooms: a mechanism of DMSP release? *Marine Ecology Progress Series* **128**: 133-142.
- BROEKHUIJSEN, M., LARSSON, P., JOHANSSON, A., BYSTROM, M., ERIKSSON, U., LARSSON, E., PRIOR, R.G., SJOSTEDT, A., TITBALL, R.W. & FORSMAN, M. (2003). Genome-Wide DNA Microarray Analysis of *Francisella tularensis* Strains Demonstrates Extensive Genetic Conservation within the Species but Identifies Regions That Are Unique to the Highly Virulent *F. tularensis* subsp. *tularensis*. *J. Clin. Microbiol.* **41**(7): 2924-2931.
- BROWN, C.W. & YODER, J.A. (1993). Blooms of *Emiliania huxleyi* (Prymnesiophyceae) in surface waters of the Nova Scotian Shelf and the Great Bank. *Journal of Plankton Research* **15**(12): 1429-1438.
- BRUSSAARD, C.P.D. (2004). Viral Control of Phytoplankton Populations - a Review. *The Journal of Eukaryotic Microbiology* **51**(2): 125-138.
- BRUSSAARD, C.P.D., KUIPERS, B. & VELDHUIS, M.J.W. (2005). A mesocosm study of *Phaeocystis globosa* population dynamics: I. Regulatory role of viruses in bloom control. *Harmful Algae* **4**(5): 859-874.
- BUESSELER, K.O., LAMBORG, C.H., BOYD, P.W., LAM, P.J., TRULL, T.W., BIDIGARE, R.R., BISHOP, J.K.B., CASCIOTTI, K.L., DEHAIRS, F., ELSKENS, M., *et al.* (2007). Revisiting Carbon Flux Through the Ocean's Twilight Zone. *Science* **316**(5824): 567-570.
- CASTBERG, T., THYRHAUG, R., LARSEN, A., SANDAA, R.-A., HELDAL, M., VAN ETTEN, J.L. & BRATBAK, G. (2002). Isolation and characterization of

- a virus that infects *Emiliana huxleyi* (Haptophyta). *Journal of Phycology* **38**(4): 767-774.
- CHARLSON, R.J., LOVELOCK, J.E., ANDREAE, M.O. & WARREN, S.G. (1987). Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature* **326**(6114): 655-661.
- DE LA ROCHA, C.L. & DEPAOLO, D.J. (2000). Isotopic evidence for variations in the marine calcium cycle over the Cenozoic. *Science* **289**: 1176-1178.
- DERISI, J.L., IYER, V.R. & BROWN, P.O. (1997). Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. *Science* **278**(5338): 680-686.
- DERISI, J.L., IYER, V.R. & BROWN, P.O. (1998). *The MGuide: A complete guide to building your own microarrayer*. Stanford University, Stanford, CA.
- DUPONT, S., WILSON, K., OBST, M., SKÖLD, H., NAKANO, H. & THORNDYKE, M.C. (2007). Marine ecological genomics: when genomics meets marine ecology. *Marine Ecology Progress Series* **332**: 257-273.
- EDVARSEN, B., EIKREM, W., GREEN, J.C., ANDERSEN, R.A., STAAV, S.Y.M.-V.D. & MEDLIN, L.K. (2000). Phylogenetic reconstructions of the Haptophyta inferred from 18S ribosomal DNA sequences and available morphological data. *Phycologia* **39**(1): 19-35.
- EISEN, M.B. & BROWN, P.O. (1999). DNA arrays for analysis of gene expression. *Methods in Enzymology* **303**: 179-205.
- EJRNAES, M., FILIPPI, C.M., MARTINIC, M.M., LING, E.M., TOGHER, L.M., CROTTY, S. & VON HERRATH, M.G. (2006). Resolution of a chronic viral infection after interleukin-10 receptor blockade. *J. Exp. Med.*
- ERBAR, S., DIEDERICH, S. & MAISNER, A. (2008). Selective receptor expression restricts Nipah virus infection of endothelial cells. *Virology Journal* **5**(1): 142.
- EVANS, C., MALIN, G., MILLS, G.P. & WILSON, W.H. (2006). Viral infection of *Emiliana huxleyi* (Prymnesiophyceae) leads to elevated production of reactive oxygen species. *Journal of Phycology* **42**(5): 1040-1047.
- FALKOWSKI, P.G., BARBER, R.T. & SMETACEK, V. (1998). Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science* **281**(5374): 200-206.

- FALKOWSKI, P.G., KATZ, M.E., KNOLL, A.H., QUIGG, A., RAVEN, J.A., SCHOFIELD, O. & TAYLOR, F.J.R. (2004). The Evolution of Modern Eukaryotic Phytoplankton. *Science* **305**(5682): 354-360.
- FALKOWSKI, P.G., LAWS, E.A., BARBER, R.T. & MURRAY, J.W. (2003). Phytoplankton and Their Role in Primary, New, and Export Production. In *Ocean Biogeochemistry: The Role of the Ocean Carbon Cycle* (FASHAM, M.J.R. editor), Springer-Verlag, Berlin Heidelberg New York.
- FALKOWSKI, P.G. & RAVEN, J.A. (2007). *Aquatic Photosynthesis: Second Edition*. Princeton University Press, Princeton, New York.
- FIELD, C.B., BEHRENFELD, M.J., RANDERSON, J.T. & FALKOWSKI, P. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science* **281**(5374): 237-240.
- FRADA, M., PROBERT, I., ALLEN, M.J., WILSON, W.H. & DE VARGAS, C. (2008). The "Cheshire Cat" escape strategy of the coccolithophore *Emiliana huxleyi* in response to viral infection. *Proceedings of the National Academy of Sciences* **105**(41): 15944-15949.
- FUHRMAN, J.A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature* **399**(6736): 541-548.
- GEISEN, M., BILLARD, C., BROERSE, A.T.C., CROS, L., PROBERT, I. & YOUNG, J.R. (2002). Life cycle associations involving pairs of holococcolithophorid species: Intraspecific variation or cryptic speciation? *Eur. J. Phycol.* **37**: 531-550.
- GENTRY, T.J., WICKHAM, G.S., SCHADT, C.W., HE, Z. & ZHOU, J. (2006). Microarray Applications in Microbial Ecology Research. *Microbial Ecology* **52**(2): 159-175.
- GIBSON, G. (2002). Microarrays in ecology and evolution: a preview. *Molecular Ecology* **11**: 17-24.
- GOBLER, C.J., HUTCHINS, D.A., FISHER, N.S., COSPER, E.M. & SANUDO-WILHELMY, S. (1997). Release and bioavailability of C, N, P, Se, and Fe following viral lysis. *Limnol. Oceanogr.* **42**: 1492-1504.
- GOODWIN, P.H., OLIVER, R.P. & HSIANG, T. (2004). Comparative analysis of expressed sequence tags from *Malva pusilla*, *Sorghum bicolor*, and *Medicago*

- truncatula* infected with *Colletotrichum* species. *Plant Science* **167**(3): 481-489.
- GREEN, J.C., COURSE, P.A. & TARRAN, G.A. (1996). The life-cycle of *Emiliana huxleyi*: A brief review and a study of relative ploidy levels analysed by flow cytometry. *Journal of Marine Systems* **9**(1-2): 33-44.
- GROSSMAN, A.R. (2005). Paths toward Agal Genomics. *Plant Physiology* **137**: 410-427.
- HACKETT, J., SCHEETZ, T., YOON, H., SOARES, M., BONALDO, M., CASAVANT, T. & BHATTACHARYA, D. (2005). Insights into a dinoflagellate genome through expressed sequence tag analysis. *BMC Genomics* **6**(1): 80.
- H AidAR, A.T. & THIERSTEIN, H.R. (2001). Coccolithophore dynamics off Bermuda (N. Atlantic). *Deep Sea Research Part II* **48**: 1925-1956.
- HAY, W.W. (2004). Carbonate fluxes and calcareous nannoplankton. In *Coccolithophores: From Molecular Processes to Global Impact* (THIERSTEIN, H.R. & YOUNG, J.R. editors), pp. 509-528. Springer, Berlin Heidelberg New York.
- HILL, R.W., A., W.B., COTTRELL, M.T. & DACEY, J.W.H. (1998). Virus-mediated total release of dimethylsulfoniopropionate from marine phytoplankton: a potential climate process. *Aquat. Microb. Ecol.* **14**: 1-6.
- HOLLIGAN, P.M., FERNÁNDEZ, E., AIKEN, J., BALCH, W.M., BOYD, P., BURKILL, P.H., FINCH, M., GROOM, S.B., MALIN, G., MULLER, K., *et al.* (1993). A biogeochemical study of the coccolithophore, *Emiliana huxleyi*, in the North Atlantic. *Global Biogeochemical Cycles* **7**(4): 879-900.
- HOLLIGAN, P.M., VIOLLIER, M., HARBOUR, D.S., CAMUS, P. & CHAMPAGNE-PHILIPPE, M. (1983). Satellite and ship studies of coccolithophore production along a continental shelf edge. *Nature* **304**(5924): 339-342.
- HSIANG, T. & GOODWIN, P.H. (2003). Distinguishing plant and fungal sequences in ESTs from infected plant tissues. *Journal of Microbiological Methods* **54**(3): 339-351.
- IGLESIAS-RODRIGUEZ, M.D., SAEZ, A.G., GROBEN, R., EDWARDS, K.J., BATLEY, J., MEDLIN, L.K. & HAYES, P.K. (2002). Polymorphic

- microsatellite loci in global populations of the marine coccolithophorid *Emiliana huxleyi*. *Molecular Ecology Notes* **2**(4): 495-497.
- ITO, T. & FOLLOWS, M.J. (2003). Upper ocean control on the solubility pump of CO₂. *Journal of Marine Research* **61**: 465-489.
- JACOBSEN, A., BRATBAK, G. & 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. & BRATBAK, G. (2002). Flow cytometric analysis of an *Emiliana huxleyi* bloom terminated by viral infection. *Aquatic Microbial Ecology* **27**(2): 111-124.
- JANTASURIYARAT, C., GOWDA, M., HALLER, K., HATFIELD, J., LU, G., STAHLBERG, E., ZHOU, B., LI, H., KIM, H., YU, Y., *et al.* (2005). Large-Scale Identification of Expressed Sequence Tags Involved in Rice and Rice Blast Fungus Interaction. *Plant Physiol.* **138**(1): 105-115.
- KATO-MAEDA, M., RHEE, J.T., GINGERAS, T.R., SALAMON, H., DRENKOW, J., SMITTIPAT, N. & SMALL, P.M. (2001). Comparing genomes within the species *Mycobacterium tuberculosis*. *Genome Research* **11**(4): 547-554.
- KELLER, M.D. (1989). Dimethyl sulfide production and marine phytoplankton: the importance of species composition and cell size. *Biol Oceanogr* **6**: 375-382.
- KLAVENESS, D. (1972). *Coccolithus huxleyi* (Lohmann) Kamptner. I. Morphological investigations on the vegetative cell and the process of coccolith formation. *Protistologica* **8**: 335-346.
- KORE-EDA, S., CUSHMAN, M.A., AKSELROD, I., BUFFORD, D., FREDRICKSON, M., CLARK, E. & CUSHMAN, J.C. (2004). Transcript profiling of salinity stress responses by large-scale expressed sequence tag analysis in *Mesembrythemum crystallinum*. *Gene* **341**: 83-92.
- LAWRENCE, J.E., CHAN, A.M. & SUTTLE, C.A. (2002). Viruses causes lysis of the toxic bloom-forming alga, *Heterosigma akashiwo* (Raphidophyceae), are widespread in coastal sediments of British Columbia, Canada. *Limnol. Oceanogr.* **47**: 545-550.

- LAWRENCE, J.E. & SUTTLE, C.A. (2004). Effect of viral infection on sinking rates of *Heterosigma akashiwo* and its implications for bloom termination. *Aquat. Microb. Ecol.* **37**: 1-7.
- LAWTON, J.H. (1994). What do species do in ecosystem? *Oikos* **71**(3): 367-374.
- LEUNG, Y.F. & CAVALIERI, D. (2003). Fundamentals of cDNA microarray data analysis. *Trends in Genetics* **19**(11): 649-659.
- LI, X., GU, W., MOHAN, S. & BAYLINK, D.J. (2002). DNA microarrays: their use and misuse. *Microcirculation* **9**(1): 13-22.
- LIDIE, K.B., RYAN, J.C., BARBIER, M. & VAN DOLAH, F.M. (2005). Gene Expression in Florida Red Tide Dinoflagellate *Karenia brevis*: Analysis of an Expressed Sequence Tag Library and Development of DNA Microarray. *Mar Biotechnol* **7**(5): 481-493.
- LIPSHUTZ, R.J., FODOR, S.P., GINGERAS, T.R. & LOCKHART, D.J. (1999). High density synthetic oligonucleotide arrays. *Nature Genetics* **21**: 20 - 24.
- LISS, P.S., HATTON, A.D., MALIN, G., NIGHTINGALE, P.D. & TURNER, S.M. (1997). Marine sulphur emissions. *Philosophical Transactions of the Royal Society B: Biological Sciences* **352**(1350): 159-169
- LLEWELLYN, C.A., EVANS, C., AIRS, R.L., COOK, I., BALE, N. & WILSON, W.H. (2007). The response of carotenoids and chlorophylls during virus infection of *Emiliana huxleyi* (Prymnesiophyceae). *Journal of Experimental Marine Biology and Ecology* **344**(1): 101-112.
- MACFARLANE, R., BHATTACHARYA, D. & SINGH, U. (2005). Genomic DNA microarrays for *Entamoeba histolytica*: Applications for use in expression profiling and strain genotyping. *Experimental Parasitology* **110**: 196-202.
- MALIN, G., LISS, P.S. & TURNER, S.M. (1994). Dimethyl sulphide: production and atmospheric consequences. In *The Haptophyte Algae* (GREEN, J.C. & LEADBEATER, B.S.C. editors), pp. 303–320. Clarendon Press, Oxford.
- MALIN, G., TURNER, S.M. & LISS, P.S. (1992). Sulfur: The plankton/climate connection. *Journal of Phycology* **28**(5): 590-597.
- MALIN, G., WILSON, W.H., BRATBAK, G., LISS, P.S. & MANN, N.H. (1998). Elevated production of dimethylsulfide resulting from viral infection of cultures of *Phaeocystis pouchetii*. *Limnol. Oceanogr.* **43**: 1389-1393.

- MANNING, G., PLOWMAN, G.D., HUNTER, T. & SUDARSANAM, S. (2002). Evolution of protein kinase signaling from yeast to man. *Trends Biochem. Sci.* **27**: 514-520.
- MARSH, M.E. (2003). Biochemistry and Molecular Biology : Regulation of CaCO₃ formation in coccolithophores. *Comparative Biochemistry and Physiology Part B* **136**(4): 743-754.
- MARTINÉZ, J.M., SCHROEDER, D.C., LARSEN, A., BRATBAK, G. & WILSON, W.H. (2007). Molecular Dynamics of *Emiliana huxleyi* and Cooccurring Viruses during Two Separate Mesocosm Studies. *Applied and Environmental Microbiology* **73**(2): 554-562.
- MATSUZAKI, M., MISUMI, O., SHIN-I, T., MARUYAMA, S., TAKAHARA, M., MIYAGISHIMA, S.-Y., MORI, T., NISHIDA, K., YAGISAWA, F., NISHIDA, K., *et al.* (2004). Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **428**(6983): 653-657.
- MEDLIN, L.K., BARKER, G.L.A., CAMPBELL, L., GREEN, J.C., HAYES, P.K., MARIE, D., WRIEDEN, S. & VAULOT, D. (1996). Genetic characterisation of *Emiliana huxleyi* (Haptophyta). *Journal of Marine Systems* **9**(1-2): 13-31.
- MERCHANT, S.S., PROCHNIK, S.E., VALLON, O., HARRIS, E.H., KARPOWICZ, S.J., WITMAN, G.B., TERRY, A., SALAMOV, A., FRITZ-LAYLIN, L.K., MARÉCHAL-DROUARD, L., *et al.* (2007). The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science* **318**(5848): 245-250.
- MIDDELBOE, M. & JORGENSEN, N.O.G. (2006). Viral lysis of bacteria: an important source of dissolved amino acids and cell wall compounds. *J. Mar. Biolog. Assoc. UK* **86**(03): 605-612.
- MIDDELBOE, M., JORGENSEN, N.O.G. & KROER, N. (1996). Effects of viruses on nutrient turnover and growth efficiency of non-infected marine bacterioplankton. *Appl. Environ. Microbiol.* **62**: 1991-1997.
- MIDDELBOE, M. & LYCK, P.G. (2002). Regeneration of dissolved organic matter by viral lysis in marine microbial communities. *Aquat. Microb. Ecol.* **27**(187-194).
- MONICK, M.M., STABER, J.M., THOMAS, K.W. & HUNNINGHAKE, G.W. (2001). Respiratory Syncytial Virus Infection Results in Activation of

- Multiple Protein Kinase C Isoforms Leading to Activation of Mitogen-Activated Protein Kinase. *J Immunol* **166**(4): 2681-2687.
- MORSE, J.W. & MACKENZIE, F.T. (eds.) (1990). *Geochemistry of Sedimentary Carbonates*. Elsevier, Amsterdam.
- MUNOZ, E., BOGARAD, L. & DEEM, M. (2004). Microarray and EST database estimates of mRNA expression levels differ: The protein length versus expression curve for *C. elegans*. *BMC Genomics* **5**(1): 30.
- NAGASAKI, K., ANDO, M., IMAI, I., ITAKURA, S. & ISHIDA, Y. (1994a). Virus-like particles in *Heterosigma akashiwo* (Raphidophyceae) : a possible red tide disintegration mechanism. *Mar. Biol* **119**(2): 307-312.
- NAGASAKI, K., ANDO, M., ITAKURA, S., IMAI, I. & ISHIDA, Y. (1994b). Viral mortality in the final stage of *Heterosigma akashiwo* (Raphidophyceae) red tide. *J. Plankton Res.* **16**(11): 1595-1599.
- NAGASAKI, K. & YAMAGUCHI, M. (1997). Isolation of a virus infectious to the harmful bloom causing microalga *Heterosigma akashiwo*. *Aquat Microb Ecol* **13**: 135-140.
- NORKIN, L.C. (1995). Virus receptors: implications for pathogenesis and the design of antiviral agents. *Clin. Microbiol. Rev.* **8**(2): 293-315.
- NOZAKI, H., TAKANO, H., MISUMI, O., TERASAWA, K., MATSUZAKI, M., MARUYAMA, S., NISHIDA, K., YAGISAWA, F., YOSHIDA, Y., FUJIWARA, T., *et al.* (2007). A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga *Cyanidioschyzon merolae*. *BMC Biology* **5**(1): 28-35.
- PAASCHE, E. (2002). A review of the coccolithophorid *Emiliana huxleyi* (Prymnesiophyceae), with particular reference to growth, coccolith formation, and calcification-photosynthesis interactions. *Phycologia* **40**(6): 503-529.
- PEARSON, B.M., PIN, C., WRIGHT, J., I'ANSON, K., HUMPHREY, T. & WELLS, J.M. (2003). Comparative genome analysis of *Campylobacter jejuni* using whole genome DNA microarrays. *FEBS Letters* **554**(1-2): 224-230.
- PICKART, C.M. & EDDINS, M.J. (2004). Ubiquitin: structures, functions, mechanisms. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1695**(1-3): 55-72.

- PIMM, S.L., RUSSEL, G.J., GITTLEMAN, J.L. & BROOLS, T.M. (1995). The Future of Biodiversity. *Science* **269**(5222): 347-350.
- RAFALSKI, J.A., HANAFEY, M., MIAO, G.H., CHING, A., LEE, J.M., DOLAN, M. & TINGEY, S. (1998). New experimental and computational approaches to the analysis of gene expression. *Acta Biochimica Polonica* **45**(4): 929-934.
- REN, R.B., COSTANTINI, F., GORGACZ, E.J., LEE, J.J. & RACANIELLO, V.R. (1990). Transgenic mice expressing a human poliovirus receptor: a new model for polioimyelitis. *Cell* **63**: 353-362.
- RESNICK, D., PEARSON, A. & KRIEGER, M. (1994). The SRCR superfamily - A family reminiscent of the Ig superfamily. *Trends Biochem. Sci.* **19**: 5-8.
- REUSCH, T.B.H., VERON, A.S., PREUSS, C., WEINER, J., WISSLER, L., BECK, A., KLAGES, S., KUBE, M., REINHARDT, R. & BORNBERG-BAUER, E. (2008). Comparative Analysis of Expressed Sequence Tag (EST) Libraries in the Seagrass *Zostera marina* Subjected to Temperature Stress. *Mar Biotechnol* **10**: 297-309.
- RILEY, M. & SERRES, M.H. (2000). Interim report on genomics of *Escherichia coli*. *Annual Review of Microbiology* **54**: 341-411.
- ROST, B. & RIEBESELL, U. (2004). Coccolithophores and the biological pump: responses to environmental changes. In *Coccolithophores: From Molecular Processes to Global Impact* (THIERSTEIN, H.R. & YOUNG, J.R. editors), pp. 99-125. Springer, Berlin Heidelberg New York.
- RUDD, S. (2003). Expressed sequence tags: alternative or complement to whole genome sequences? *Trends in Plant Science* **8**(7): 321-329.
- SAEZ, A.G., PROBERT, I., GEISEN, M., QUINN, P., YOUNG, J.R. & MEDLIN, L.K. (2003). Pseudo-cryptic speciation in coccolithophores. *PNAS* **100**(12): 7163-7168.
- SALAMA, N., GUILÉMIN, K., MCDANIEL, T.K., SHERLOCK, G., TOMPKINS, L. & FALKOW, S. (2000). A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *PNAS* **97**: 14668-14673.
- SANDAA, R.A., HELDAL, M., CASTBERG, T., THYRHAUG, R. & 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.

- SCHENA, M., SHALON, D., DAVIS, R.W. & BROWN, P.O. (1995). Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science* **270**(5235): 467-470.
- SCHMITT, A.O., SPECHT, T., BECKMANN, G., DAHL, E., PILARSKY, C.P., HINZMANN, B. & ROSENTHAL, A. (1999). Exhaustive mining of EST libraries for genes differentially expressed in normal and tumour tissues. *Nucl. Acids Res.* **27**: 4251-4260.
- SCHROEDER, D.C., OKE, J., HALL, M., MALIN, G. & WILSON, W.H. (2003). Virus succession observed during an *Emiliana huxleyi* bloom. *Applied and Environmental Microbiology* **69**(5): 2484-2490.
- SCHROEDER, D.C., OKE, J., MALIN, G. & WILSON, W.H. (2002). Coccolithovirus (Phycodnaviridae): Characterisation of a new large dsDNA algal virus that infects *Emiliana huxleyi*. *Archives of Virology* **147**(9): 1685-1698.
- SHIMIZU, Y. (1996). Microalgal metabolites: a new perspective. *Annual Review of Microbiology* **50**: 431-465.
- SMAYDA, T.J. (1997). What is a bloom? A commentary. *Limnol. Oceanogr.* **42**(5 part 2): 1132-1136.
- SMAYDA, T.J. (2000). Ecological features of harmful algal blooms in coastal upwelling ecosystems. *South African Journal of Marine Science/Suid-Afrikaanse Tydskrif vir Seewetenskap* **22**: 219-253.
- SMETACEK, V. (1999). Diatoms and the ocean carbon cycle. *Protist* **150**: 25-32.
- SOARES, M.B., BONALDO, M.F., JELENE, P., SU, L., LAWTON, L. & EFSTRATIADIS, A. (1994). Construction and characterization of a normalized cDNA library. *Proceedings of the National Academy of Sciences of the United States of America* **91**(20): 9228-9232.
- STEFELS, J., STEINKE, M., TURNER, S., MALIN, G. & BELVISO, S. (2007). Environmental constraints on the production and removal of the climatically active gas dimethylsulphide (DMS) and implications for ecosystem modelling. *Biogeochemistry* **83**(1-3): 245-275.
- SUKHANOVA, I.N. & FLINT, M.V. (1998). Anomalous blooming of coccolithophorids over the Eastern Bering Sea Shelf. *Oceanology* **38**(4): 502-505.

- SUTTLE, C.A. (2000). Ecological, evolutionary, and geochemical consequences of viral infection of cyanobacteria and eukaryotic algae. In *Viral Ecology* (HURST, C.J. editor), pp. 247-296. Academic Press, New York.
- SUTTLE, C.A. (2005). Viruses in the sea. *Nature* **437**(7057): 356-361.
- SUTTLE, C.A. (2007). Marine viruses - major players in the global ecosystem. *Nat Rev Micro* **5**(10): 801-812.
- THIERSTEIN, H.R., GEITZENAUER, K.R., MOLFINO, B. & SHACKLETON, N.J. (1977). Global synchronicity of late Quaternary coccolith datum levels: Validation by oxygen isotopes. *Geology* **5**: 400-404.
- THYRHAUG, R., LARSEN, A., BRUSSAARD, C.P.D. & BRATBAK, G. (2002). Cell cycle dependent virus production in marine phytoplankton. *Journal of Phycology* **38**(2): 338-343.
- THYRHAUG, R., LARSEN, A., THINGSTAD, T.F. & BRATBAK, G. (2003). Stable coexistence in marine algal host-virus systems. *Marine Ecology Progress Series* **254**: 27-35.
- VAN CAPELLEN, P. (2003). Biomineralization and Global Biogeochemical Cycles. *Reviews in Mineralogy and Geochemistry* **54**(1): 357-381.
- VAN ETTEN, J.L., GRAVES, M.V., MÜLLER, D.G., BOLAND, W. & DELAROCHE, N. (2002). Phycodnaviridae - large DNA algal viruses. *Archives of Virology* **147**(8): 1479-1516.
- VAN ETTEN, J.L., LANE, L.C. & MEINTS, R.H. (1991). Viruses and Viruslike Particles of Eukaryotic Algae. *Microbiological Reviews* **55**(4): 586-620.
- VAN ETTEN, J.L. & MEINTS, R.H. (2003). Giant viruses infecting algae. *Annual Review of Microbiology* **53**: 447-494.
- VAN STRAALLEN, N.M. & ROELOFS, D. (2006). *An Introduction to Ecological Genomics*. Oxford University Press, Oxford.
- VENTELON-DEBOUT, M., NGUYEN, T.T.H., WISSOCQ, A., BERGER, C., LAUDIE, M., PIÉGU, B., COOKE, R., GHESQUIÈRE, A., DELSENY, M. & BRUGIDOU, C. (2003). Analysis of the transcriptional response to Rice Yellow Mottle Virus infection in *Oryza sativa indica* and *japonica* cultivars. *Molecular Genetics and Genomics* **270**(3): 253-262.
- VOLK, T. & HOFFERT, M.I. (1985). Ocean carbon pumps: Analysis of relative strengths and efficiencies in ocean-driven atmospheric CO₂ changes. In *The*

- Carbon Cycle and Atmospheric CO₂: Natural variations archean to present. Geophys. Monogr. Ser 32* (SUNDQUIST, E.T. & BROECKER, W.S. editors), pp. 99-110. AGU, Washington, D.C.
- WATANABE, T., MURATA, Y., OKA, S. & IWAHASHI, H. (2004). A new approach to species determination for yeast strains: DNA microarray-based comparative genomic hybridization using a yeast DNA microarray with 6000 genes. *Yeast* **21**(4): 351-365.
- WEINBAUER, M.G. (2004). Ecology of prokaryotic viruses. *FEMS Microbiology Reviews* **28**(2): 127-181.
- WESTBROEK, P., BROWN, C.W., BLEIJSWIJK, J.V., BROWNLEE, C., BRUMMER, G.J., CONTE, M., EGGE, J., FERNANDEZ, E., JORDAN, R., KNAPPERTSBUSCH, M., *et al.* (1993). A model system approach to biological climate forcing. The example of *Emiliana huxleyi*. *Global and Planetary Change* **8**(1-2): 27-46.
- WESTBROEK, P., HINTE, J.E.V., BRUMMER, G.-J., VELDHUIS, M., BROWNLEE, C., GREEN, J.C., HARRIS, R. & HEIMDAL, B.R. (1994). *Emiliana huxleyi* as a key to biosphere-geosphere interactions. In *The Haptophyte Algae* (GREEN, J.C. & LEADBEATER, B.S.C. editors), pp. 321-334. Clarendon Press, Oxford.
- WILHELM, S.W. & SUTTLE, C.A. (1999). Viruses and nutrient cycles in the sea. *BioScience* **49**: 781-788.
- WILSON, W.H., SCHROEDER, D.C., ALLEN, M.J., HOLDEN, M.T.G., PARKHILL, J., BARRELL, B.G., CHURCHER, C., HAMLIN, N., MUNGALL, K., NORBERTCZAK, H., *et al.* (2005). Complete Genome Sequence and Lytic Phase Transcription Profile of a Coccolithovirus. *Science* **309**(5737): 1090-1092.
- WILSON, W.H., TARRAN, G.A., SCHROEDER, D., COX, M., OKE, J. & MALIN, G. (2002). Isolation of viruses responsible for the demise of an *Emiliana huxleyi* bloom in the English Channel. *J. Mar. Biol. Assoc. UK* **82**(03): 369-377.
- WINTER, A., JORDAN, R.W. & ROTH, P.H. (1994). Biogeography of living coccolithophores in ocean waters. In *Coccolithophores* (WINTER, A. &

- SIESSER, W.G. editors), pp. 161-177. Cambridge University Press, Cambridge.
- WINTER, A. & SIESSER, W.G. (1994). *Coccolithophores*. Cambridge Univ. Press, New York.
- WOELK, T., SIGISMUND, S., PENENGO, L. & POLO, S. (2007). The ubiquitination code: a signalling problem. *Cell Division* **2**(1): 11.
- WOMMACK, K.E. & COLWELL, R.R. (2000). Virioplankton: Viruses in Aquatic Ecosystems. *Microbiology and Molecular Biology Reviews* **64**(1): 69-114.
- YE, R.W., WANG, T., BEDZYK, L. & CROKER, K.M. (2001). Applications of DNA microarrays in microbial systems. *J. Microbiol. Meth* **47**(3): 257-272.

5 SUMMARY

This thesis aimed at the identification of genes involved in the host-virus interaction of the coccolithophore *Emiliana huxleyi* and the virus EhV-86 by taking advantage of EST libraries and DNA microarrays. Microarray-based comparative genomic hybridization (CGH) was used to investigate the genetic variation of several *E. huxleyi* strains and to identify genes with respect to virus susceptibility and morphology, e.g. formation of coccoliths.

Analysis of expressed sequence tags (ESTs) was performed to gain insights into the host-virus interaction of *E. huxleyi*. Three complementary DNA (cDNA) libraries generated 6, 12 and 24 h post viral infection were compared to a library from an uninfected culture by sequencing, clustering and manual annotation of randomly selected ESTs. At first, a preliminary set of 60-90 ESTs from each library were annotated to get an overview of gene expression changes that occur during viral infection of *E. huxleyi*. BLAST-searches of the sequenced genome of the virus (EhV-86) were used to identify viral genes. Results of this small sample probe show already a trend towards down-regulation of genes involved in photosynthesis of *E. huxleyi* for the benefit of an increased transcription and translation for viral replication. At 6 (T6) and 12 (T12) hours post viral infection the algal transcriptome changed significantly although only 3-4 viral transcripts were present. In addition, at 24 hours (T24) post infection only 10% of the mRNA was of host origin. Viral transcripts identified at T24 encode proteins involved in protein degradation, nucleic acid degradation, transcription and replication.

As a next step, 1100-1500 ESTs per library were sequenced and annotated. Results confirmed the previous tendencies and discovered more genes involved in the host's response to viral infection. Furthermore, two-colour oligonucleotide microarrays were used to verify the gene expression results of the ESTs. A total of 4480 ESTs were assembled into 1871 clusters of which 223 clusters were of viral origin. A putative function could be assigned to 35% of the host clusters and to 20% of the viral clusters.

In addition, microarray expression analysis indicated that 231 out of 565 oligonucleotides of *E. huxleyi* changed their expression level in at least one time point. Results suggest that viral infection affects the transcriptional machinery of *E. huxleyi* within a few hours by decreasing the expression of genes involved in photosynthesis and protein degradation at the benefit of fatty acid metabolism, glycolysis, and transcription and translation. The expression of *E. huxleyi* and EhV-86 genes changed significantly between 12 and 24 hours after infection.

The results provide insights into the infection mechanisms of the virus EhV-86 in *E. huxleyi* and demonstrate the power of EST libraries and DNA microarrays to obtain data on gene expression and regulation during viral infection.

Microarray-based comparative genomic hybridization (CGH) was applied to investigate genomic diversity of 16 *E. huxleyi* strains of different geographic origin and to identify genes related to virus susceptibility and morphology. The microarray consisted of 565 genes derived from the former EST study of *E. huxleyi* strain CCMP1516 and 37880 genes from the ongoing genome project of the same strain. *Gephyrocapsa oceanica* and *Isochrysis galbana* were taken as out-groups. A total of 32395 gene transcripts showed significant hybridization patterns and were used to elucidate genetic diversity. Hybridization intensities were compared to determine the relative copy number of each gene transcript. Comparisons with the sequenced *E. huxleyi* strain CCMP1516 revealed that 27% (8740 genes) to 57% (18581 genes) of the genes showed a pattern of hybridization concordant with deletion, nucleotide divergence or gene duplication within the species and up to 83% (26881 genes) between the genera. The largest variation was observed among the species for *E. huxleyi* strain 92F. Regarding variation with respect to virus susceptibility and morphology the most abundant genes with known function were associated with

metabolism, transport, and transcription and translation. In addition, two membrane receptors and two proteins related to ubiquitin were identified which show significant differences between virus susceptible and resistant strains.

The results obtained by using CGH demonstrate that this method is appropriate to compare the gene content of different *E. huxleyi* strains. CGH was successfully applied to identify genes related to virus susceptibility and morphology. Among others, the membrane receptors and the ubiquitin-related proteins that possibly play a role in virus infection deserve further attention.

6 ZUSAMMENFASSUNG

Ziel dieser Arbeit war es mittels EST-Bänken und DNA-Microarrays Gene zu identifizieren, die in der Wirts-Virus-Interaktion der Coccolithophoride *Emiliana huxleyi* und dem Virus EhV-86 involviert sind. Mehrere *E. huxleyi* Stämme wurden auf ihre genomische Variation mit Hilfe von Microarray-basierenden Hybridisierungen untersucht. Des Weiteren zielte der genomische Vergleich darauf ab, Gene in Bezug auf Virusanfälligkeit und Morphologie (z.B. Aufweisen von Coccolithen) zu identifizieren.

Um Einblick in die Wirts-Virus-Interaktion von *E. huxleyi* zu erhalten wurden *expressed sequence tags* (ESTs) analysiert. Drei cDNA-Bänke wurden nach 6, 12 und 24 Stunden Virusinfektion hergestellt und mit einer Bank einer uninfizierten Kultur verglichen. Dazu wurden zufällig ausgewählte ESTs sequenziert, gruppiert und manuell annotiert. Um einen Überblick der Genexpressionsveränderungen von *E. huxleyi* während der Virusinfektion zu bekommen, wurden anfänglich nur 60-90 ESTs pro Bank annotiert. Die Virus-Gene wurden mittels BLAST-Suche gegen das sequenzierte Virus-Genom (EhV-86) identifiziert. Trotz einer solch kleinen Stichprobe konnte ein Trend Richtung Herunterregulierung von Genen der Photosynthese ausgemacht werden. Außerdem wurde eine höhere Anzahl an Genen der Transkription und Translation von *E. huxleyi* nachgewiesen, die vermutlich zur Replikation des Virus verwendet werden. Obwohl jeweils nur 3-4 virale Gene vorhanden waren, veränderte sich das Transkriptom der Alge nach 6 (T6) und 12 (T12) Stunden Virusinfektion signifikant. Nach 24 Stunden (T24) Infektion stammten nur noch 10% der mRNA vom Wirt. Unter den identifizierten viralen Gene in T24

befinden sich Proteine die vermutlich für den Abbau der Wirts-DNA und -proteine zuständig, sowie für die Transkription und Replikation des Virus verantwortlich sind. Darauf folgend wurden pro Bank 1100-1500 ESTs sequenziert und annotiert. Die neuen Ergebnisse bestätigten die vorherigen Tendenzen und wiesen noch mehr Gene, die in der Wirts-Virus-Interaktion beteiligt sind, nach. Insgesamt 4480 ESTs wurden in 1871 Gruppen (Cluster) assembliert von denen 223 viralen Ursprungs sind. Zu 35% der Wirts-Cluster und zu 20% der viralen Cluster konnten mögliche Funktionen zugeordnet werden. Um die Genexpressionsergebnisse der ESTs zu verifizieren wurden des Weiteren Oligonucleotid-Microarrays in einer klassischen ‚A versus B‘ Konfiguration benutzt. Hierbei zeigten 231 der 565 Oligonucleotide von *E. huxleyi* bei mindestens einem Zeitpunkt eine Veränderung im Expressionsmuster. Die Ergebnisse beider Methoden deuten darauf hin, dass innerhalb weniger Stunden Virusinfektion der Transkriptionsapparat von *E. huxleyi* beeinflusst wird. Dabei wird die Expression von Genen der Photosynthese und des Proteinabbaus zum Vorteil eines erhöhten Fettsäure Metabolismus, Glykolyse, Transkription und Translation reduziert. Zwischen 12 und 24 Stunden Virusinfektion veränderte sich die Expression von *E. huxleyi* und EhV-86 signifikant.

Die Ergebnisse geben einen Einblick in den Infektionsmechanismus des Virus EhV-86 in *E. huxleyi* und veranschaulichen das Potential von EST-Bänken und DNA-Microarrays Genexpressionsdaten während einer Virusinfektion zu erzielen.

Um die genomische Diversität von 16 *E. huxleyi* Stämme verschiedenen geographischen Ursprungs zu untersuchen wurden Microarray-basierende Hybridisierungen mit genomischer DNA durchgeführt. Des Weiteren wurde mit dieser Methode versucht Gene in Bezug auf Virusanfälligkeit und Morphologie zu identifizieren. Der Microarray bestand aus 565 Genen der vorherigen EST-Studie des *E. huxleyi* Stammes CCMP1516 und 37880 Genen aus dem laufenden Genomprojekt des gleichen Stammes. *Gephyrocapsa oceanica* und *Isochrysis galbana* wurden als Außengruppen verwendet. Die Hybridisierungen ergaben insgesamt 32395 signifikante Signalkombinationen der Gentranskripte und wurden für die Aufklärung der genetischen Diversität benutzt. Um die relative Kopienanzahl von jedem Gen zu ermitteln, wurden die Hybridisierungsintensitäten der Teststämme und Außengruppen mit denen des sequenzierten Referenz-Stammes *E. huxleyi* CCMP1516 verglichen.

Innerhalb der Spezies ergab der Vergleich zur Referenz Unterschiede von 27% (8740 Gene) bis 57% (18581 Gene) und bis zu 83% (26881 Gene) beim Vergleich mit den Außengruppen. Bei diesen Gen Unterschieden handelt es sich um Deletionen, Divergenzen in der Nukleinsäure oder Gen Duplikationen. Der *E. huxleyi* Stamm 92F zeigte innerhalb des Stammvergleiches die größte Variation zum Referenz-Stamm. In Bezug auf Virusanfälligkeit und Morphologie wurden die meisten Gene mit bekannter Funktion dem Metabolismus, dem Transport, der Transkription und Translation zugeordnet. Außerdem konnten zwei Membranrezeptoren und zwei Ubiquitin-ähnliche Proteine identifiziert werden, die signifikante Unterschiede zwischen virusanfälligen und –resistenten Stämmen aufwiesen.

Die Analyse des genomischen Vergleichs ergab, dass diese Methode sehr gut zur Untersuchung des Gen Gehaltes von verschiedenen *E. huxleyi* Stämmen geeignet ist. Des Weiteren konnte diese Methode erfolgreich für die Identifizierung von Genen in Bezug auf Virusanfälligkeit und Morphologie angewendet werden. Unter anderem konnten dabei Membranrezeptoren und Ubiquitin-ähnliche Proteine identifiziert werden, die möglicherweise eine Rolle in der Virusinfektion von *E. huxleyi* spielen und somit für weitere Untersuchungen in Betracht zu ziehen sind.

7 DANKSAGUNG

Prof. Dr. Dieter Wolf-Gladrow und Prof. Dr. Allan Cembella danke ich ganz herzlich für die Begutachtung meiner Arbeit und Prof. Dr. Kai Bischof für die kurzfristige Übernahme des 3. Prüfers. Dr. Klaus Valentin und Dr. Katja Metfies danke ich für die Einwerbung von Geldmitteln und die Bereitstellung meines Themas. Vor allem Klaus danke ich sehr herzlichst für die Unterstützung in der letzten Panik-Phase, das entgegengebrachte Vertrauen diese Arbeit fertig zu kriegen und die Möglichkeit, noch ein bisschen länger am AWI zu bleiben. Mein besonderer Dank gilt Dr. Uwe John, dessen anregende Diskussionen und ansteckende Begeisterung mich bei meiner Arbeit enorm unterstützt haben und ohne die ich diese Arbeit nicht fertig gebracht hätte. Ganz herzlich und allerliebste möchte ich mich bei den Mitgliedern und Ehemaligen der Meereisgruppe bedanken: Erika Allhusen, Maddalena Bayer, Gerhard Dieckmann, Birte Gerdes, Niko Hoch, Andreas Krell, Christiane Uhlig und unseren alten Diplomanden Burkhard, Janni, Jan, Sandra und Susan. Danke für die "Adoption", für Anleitung und Diskussion von und über Methoden, Gespräche nicht nur beruflicher Art, für Frühstückssorgien und leckere Geburtstagskuchen, für lustige Kohltouren, Weihnachtsmarktbesuche und unvergessliche Abende, sei es beim Grillen oder bei Mongolischen Feuertöpfen. Special thanks go to Willie Wilson, for the opportunity to participate at a mesocosm experiment in Bergen. Ein großes Dankeschön geht auch an die AG Medlin und AG Cembella für ein tolles Arbeitsklima und guten Ratschlägen aber auch Beschäftigungen außerhalb des AWI: Andrea, Bank, Christine, Helga, Ines, Karina, Linda, Michael, Monica, Nina, Sabine, Sara, Shinya, Sonja, Steffi und Sylke.

Besonders bei Kerstin Töbe möchte ich mich bedanken, für ihr Zuhören, Diskutieren und Beraten, an guten wie an schlechten Tagen. Meinen Korrekturlesern Gerald Langer, Sven Kranz und schon erwähnte einen herzlichen Dank und für die ein oder andere Referenz.

Meiner Familie danke ich von ganzem Herzen für deren Unterstützung in allen Lebenslagen. Danke, dass ihr mir diese Chance gegeben habt!

Des Weiteren möchte ich meinen Freunden und Berlinern danken, dass ihr mir vergessene oder falsch gratulierte Geburtstage verzeiht, fürs einfach Dasein wenn man einen zum Zuhören oder Vollabern braucht(e), und für Geduld und Toleranz, das Freundschaften einiges aushalten können. Ich verspreche Euch, meine Besuche werden jetzt wieder zahlreicher werden!

Und zum Schluss kommt immer das Beste: ich danke dir mein Herz, dass es dich gibt, für deine Unterstützung und Ausdauer besonders in der extrem langen Stress-Phase, für die Bereicherung meines Lebens, für die Erkenntnis und Abwechslung das es auch neben der Arbeit ein Leben gibt, ohne dich hätte ich das alles nicht geschafft.