Ecological and functional biodiversity in a marine algal-virus system: genotypes, phenotypes and their ecological significance

by

Jozef I. Nissimov, BSc., MSc

A thesis submitted to the University of Nottingham
for the degree of

DOCTOR OF PHILOSOPHY

Department of Plant and Crop Sciences
University of Nottingham

In Collaboration with
The Plymouth Marine Laboratory (UK)

October 2013

ABSTRACT

Coccolithoviruses are large dsDNA viruses infecting the cosmopolitan calcifying marine phytoplankton *Emiliania huxleyi*. Therefore they are instrumental components of algal bloom demise and thus agents contributing significantly to biogeochemical cycling in the oceans. Several coccolithovirus strains exist in culture and have been used so far to study the co-evolutionary arms-race between them and their unicellular host in naturally occurring or induced blooms in the North Atlantic Ocean and the fjords of Norway. However, little is known of their distribution in non-bloom conditions, their natural diversity in times of reduced infectivity rates, and the role of functionally important genes found in natural coccolithovirus communities. Even less is known about their genetic differences and the phenotypic consequences of these differences on their infection dynamics. Hence here a three dimensional approach was undertaken, during which the genomes of several coccolithovirus strains were analysed, their diversity in the global ocean characterised, and their phenotypic properties as seen from their infection dynamics with their host established.

It was revealed that although coccolithoviruses share a common subset of core genes, they differ in a large proportion of their genomic material, as seen from the presence and/or absence of large sub-clusters of functionally unknown genes. Moreover, a gene that encodes for a phosphate scavenging mechanism (phosphate permease) was truncated from the genome of the Norwegian isolate EhV-99B1 but not from any other strain, while a gene encoding for the virulence factor sialidase was truncated only in the genomes of the English Channel strains isolated in 2001. The discovery of an additional gene that is potentially involved in the regulation of sphingosine and ceramide intermediates during the *de novo* virus encoded sphingolipid biosynthesis pathway was

also intriguing, and the extent of gene homology to host genes (i.e. almost 13% of the analyzed genomes) highlighted the importance of horizontal gene transfer events in the co-evolution between algal hosts and their viruses. Secondly, it was established that virus competition over its resource, the host-cell, is fierce and that during host co-infection, some viruses (i.e. EhV-207) were superior to others (i.e. EhV-86) in their quicker utilisation of the host metabolic machinery and possibly shorter latency period within the infected cells. The biogeochemical and evolutionary implications of these distinct phenotypic properties are far reaching as in the environment there would be hundreds of different virus strains fighting over a few dominant hosts, with "losers" and "winners" coming and going from a particular niche, affecting the recirculation of nutrients and carbon at different rates.

Finally, by community fingerprinting cococlithoviruses in the global ocean with phylogenetic markers (major capsid protein) and functional markers (serine palmitoyltransferase) it was discovered that the diversity of these viruses increase with depth, and that the 3D structure of the SPT protein (involved in the propagation of host cell death) differs among strains, dictated by a variant amino acid linker region between the two domains of the protein, LCB1 and LCB2, potentially influencing the efficiency of the virus encoded sphingolipid biosynthesis pathway.

This study is an important first step in understanding the role of coccolithoviruses, their evolution, their functional characteristics, and the possible implications of the latter to biogeochemical cycling and global climate and primary production predictions.

List of contents

Page

1. INTRODUCTION.....1

1.1. Putting the algal-virus system into perspective.....1

1.2. The coccolithophore *Emiliania huxleyi* and its distribution.....3

1.2.1. The influence of coccolithophores on global climate.....5

1.3. Viruses at a glance.....8

1.4.1. Viruses in the oceans.....10

1.4.2. The "Viral Shunt" and "killing the winner".....11

1.5. *Phycodnaviridae*: viruses of phytoplankton.....15

1.5.1. Coccolithoviruses.....16

1.5.2. General characteristics of Emiliania huxleyi viruses.....17

1.5.2.1 Emiliania huxleyi virus- 86, 163, and 99B1.....17

1.5.3. Emiliania huxleyi virus 86: genomics, proteomics and life cycle.....18

1.5.4. Horizontal gene transfer in coccolithoviruses.....21

1.5.5. Programmed Cell Death and Apoptosis.....23

1.5.6. Molecular markers for phylogeny.....24

1.5.6.1. Current status of coccolithovirus diversity.....26

1.6. The Plymouth Marine Laboratory virus collection and the "new viruses".....27

1.7. Thesis objectives.....28

2. MATERIALS AND METHODS.....30

2.1. GENERAL MATERIALS.....30

2.1.1. Laboratory consumables and reagents.....30

2.1.2. Origin of *Emiliania huxleyi* strain.....31

2.1.3. Media preparation and algal cultivation.....31

2.1.4. Coccolithovirus isolates.....32

2.1.5. Virus lysate stock generation.....32

2.2. GENERAL METHODS.....34

2.2.1. Analytical Flow Cytometry of *Emiliania huxleyi*.....34

2.2.2. Analytical Flow Cytometry of *Emiliania huxleyi* viruses.....34

2.2.3. Isolation of coccolithoviruses.....35

2.3. MOLECULAR METHODS.....37

2.3.1. Primers.....37

2.3.2.1. General PCR protocol.....43

2.3.2.1.1. PCR of four coccolithovrises.....43

2.3.3. Agarose gel electrophoresis.....43

2.3.4. DNA fragment sequencing.....44

2.3.5. Virus total genomic DNA extraction for full genome sequencing.....44

2.3.5.1. Primary lysate generation for full genome sequencing.....44

2.3.5.2. Coccolithovirus lysate concentration by PEG precipitation and purification by caesium chloride gradient.....45

2.3.5.3. Genomic DNA extraction from purified coccolithovirus concentrates.....46

2.3.6. DNA extraction from environmental and experimental samples.....47

2.4. BIOINFORMATICS.....48

2.4.1. Sequence alignments of environmental sequences.....48

2.4.2. Whole genome alignment.....48

2.4.3. Genome comparison and general statistics.....48

3. COCCOLITHOVIRUSES: "GENOMIC THIEVES AND METABOLIC THUGS"- A GENOME COMPARISON.....49

3.1. INTRODUCTION.....49

3.1.1. Genome comparison: the power of bioinformatics.....51

3.1.2. The Marine Microbial Initiative.....53

3.2. MATERIALS AND METHODS.....55

3.2.1. Virus maintenance.....55

3.2.2. Coccolithovirus strains for sequencing.....55

3.2.3. Virus concentration, DNA isolation and extraction for sequencing.....55

3.2.4. Genome sequencing, assembly and annotation.....55

3.2.5. Whole genome reconstruction and alignment.....56

3.2.6. Transmission Electron Microscopy (TEM).....57

3.3. RESULTS.....58

3.3.1. Morphology and virion structure.....58

3.3.2. Genomic analysis.....59

3.3.2.1. Sequencing statistics.....59

3.3.2.2. DNA pol, MCP and COGs phylogeny.....60

3.3.2.3. Gene counts and genome size.....63

3.3.3. Differences and similarities.....67

3.3.3.1. Unique genes and homology to EhV-86.....67

3.3.3.2. Virus encoded tRNAs.....68

3.3.3.3 Genes associated with sphingolipid biosynthesis.....69

3.3.4. Homology to host genes.....72

3.3.5. Notable absences of genes in the sequenced viruses.....73

3.3.6. Inferences from proteomic analysis of the EhV-86 virion.....74

3.4. DISCUSSION.....77

3.4.1. Structural characteristics and membrane protein composition: possible implications……...77

3.4.2. Evolution and classification……..78

3.4.3. Sphingolipid biosynthesis genes……...82

3.4.4. Single gene and tRNAs variability……86

3.4.5. HGT events with the host……..90

3.5. CONCLUSIONS: WHAT THE FUTURE HOLDS FOR EHVS……..92

4. INFECTION DYNAMICS OF COCCOLITHOVIRUSES: THE VIRUS "FIGHT CLUB"……94

4.1. INTRODUCTION.....94

4.2. MATERIALS AND METHODS.....98

4.2.1. Culture conditions and setup.....98

4.2.2. Probe design and optimisation.....100

4.2.3. PCR with qPCR primers for EhV-86 and EhV-207.....100

4.2.4. Sampling procedure.....101

4.2.5. Analytical Flow Cytometry (AFC) of host and virus abundance.....101

4.2.6. DNA extraction from top phase and pellet samples.....101

4.2.7. Real-time PCR.....102

4.3. RESULTS.....104

4.3.1. PCR primer specificity.....104

4.3.2. Analytical Flow Cytometry (AFC) analysis of the virus "fight club".....105

4.3.2.1. *E. huxleyi* CCMP 2090: host abundance.....105

4.3.2.2. AFC analysis of free VLPs.....108

4.3.3. qPCR analysis of the virus "fight club".....111

4.3.3.1. Calibration curve efficiencies of the qPCR "fight club" sample analysis.....111

4.3.3.2. Virus copy number of “fight club” versus “solo” cultures.....112

4.3.3.3. Virus copy number of free and attached VLPs.....113

4.4. DISCUSSION.....118

4.4.1. The losers and winners of the virus “fight club” – a numbers game.....119

4.4.2. The ecological significance of the virus “fight club”.....122

4.4.3. The evolutionary significance of the virus “fight club”.....126

4.5. CONCLUSIONS.....128

5. ENVIRONMENTAL COCCOLITHOVIRUS BIODIVERSITY AND ITS POSSIBLE IMPLICATIONS ON NICHE ASSOCIATION AND SPHINGOLIPID METABOLISM.....129

5.1. INTRODUCTION.....129

5.2. MATERIALS AND METHODS.....133

5.2.1. Atlantic Meridional Transect cruise (AMT).....133

5.2.1.1. AMT-20 sample collection and stations.....133

5.2.2. The Western Channel Observatory (WCO) time series.....135

5.2.2.1. WCO sample collection.....136

5.2.3. DISCO cruise sample collection (1999).....137

5.2.4. DNA Extraction, PCR and DGGE analysis.....137

5.2.5. Bioinformatic analysis.....141

5.2.5.1. MCP and SPT 3D protein modelling.....141

5.2.6. Analytical Flow Cytometry (AFC) of coccolithophores and their viruses.....142

5.3. RESULTS.....143

5.3.1. Coccolithophore and coccolithovirus AFC.....143

5.3.2. Coccolithovirus Major Capsid Protein diversity.....144

5.3.3. Structural implications of MCP diversity.....148

5.3.4. Coccolithovirus serine palmitoyltransferase (SPT) diversity.....151

5.3.5. Structural implications of SPT diversity.....151

5.4. DISCUSSION.....157

5.5. CONCLUSIONS.....164

6. COCCOLITHOVIRUS ISOLATION AND GENOTYPIC FINGERPRINTING.....166

6.1. INTRODUCTION.....166

6.1.1. Microarrays and their application.....167

6.1.2. Polymerase chain reaction and its applications.....168

6.2. MATERIALS AND METHODS.....169

6.2.1. Isolation of new coccolithoviruses.....169

6.2.2. Genomic DNA isolation and purification from four new coccolithoviruses...169

6.2.3. Coccolithovirus microarray chips.....170

6.2.4. Coccolithovirus genomic DNA labelling.....170

6.2.5. Coccolithovirus DNA microarray hybridization.....171

6.2.6. Acquisition, analysis and storage of data.....171

6.2.7. PCR fingerprinting of coccolithoviruses.....172

6.3. RESULTS.....173

6.3.1. Sample screening and virus isolation.....173

6.3.2. Genomic DNA extraction and assessment.....173

6.3.3. Microarray analysis.....174

6.3.4. PCR fingerprinting with EhV-86 primers.....177

6.4. DISCUSSION.....183

6.4.1. Lack of hybridisation: true or false-negative?.....185

6.5. CONCLUSIONS.....187

7. SUMMARY AND FUTURE STUDIES.....188

7.1. Future Research.....193

REFERENCES.....198

APPENDICES.....215

PUBLISHED ARTICLES.....242

Guide to figures

Page

Fig. 1.1. Phylogenies of major algal groups.....2

Fig. 1.2. Satellite images of coccolithophore blooms captured by NASA.....4

Fig. 1.3. Conceptual diagram of the climate feedback loop.....8

Fig. 1.4. Virus mediated catalysis of biogeochemical cycling.....12

Fig. 1.5. Schematic hypothetic representation of the proposed life-cycle.....21

Fig. 3.1. Transmission electron micrograph of the nine coccolithophore virions.....58

Fig. 3.2. Multiple Sequence Alignment of the DNA pol (DNA polymerase).....60

Fig. 3.3. Neighbour joining tree of 42 different coccolithovirus MCP sequences.....62

Fig. 3.4. Genome cluster profile based on the predicted Clusters of Orthologous Groups.....63

Fig. 3.5. Multiple whole genome alignment of nine EhVs.....65

Fig. 3.6. Reconstruction of EhVs based on the complete backbone of EhV-86.....66

Fig. 3.7. Multiple sequence alignment of the serine palmitoyltransferase.....71

Fig. 3.8. Multiple sequence alignment of the predicted alkaline phytoceramidase.....71

Fig. 3.9. CLUSTLAW alignment of the previously expressed ehv455 (sialidase) in EhV-86 and its homologue in the genome of EhV- 208.....75

Fig. 3.10. Theoretical representation of the deletion of a small section of genes in the viruses isolated in 2001.....79

Fig. 3.11. Hypothetical *de novo* sphingolipid biosynthesis pathway.....85

Fig. 4.1. Infors Minifors chemostat bioreactor.....98

Fig. 4.2. Culture flasks in which the virus "fight club" experiment.....99

Fig. 4.3. Gel electrophoresis of PCR products conducted with primers specific to EhV-86 and EhV-207.....104

Fig. 4.4. *E. huxleyi* CCMP 2090 average cells mL^{-1}.....106

Fig. 4.5. AFC density plot analysis at t0, t8, t9 and t13 of *E. huxleyi* hosts.....107

Fig. 4.6. Average density mL^{-1} (of three replicates) of EhV-86.....109

Fig. 4.7. Analytical Flow Cytometry density plot analysis at t8, t9 and t13 of VLPs...110

Fig. 4.8. Calibration curve for the qPCR amplification.....111

Fig. 4.9. Combined (free-floating and cell-associated) VCN averages.....113

Fig. 4.10. Cell-associated (A) and free floating (B) VCN averages (± SD).....115

Fig. 4.11. EhV-86 (A) and EhV-207 (B) virus copy number averages (± SD).....115

Fig. 4.12. EhV diversity in a mesocosm study in a Norwegian fjord.....124

Fig. 5.1. The cruise track of the AMT-20 on board the RRS James Cook.....134

Fig. 5.2. Vertical profiles of 65 sampling stations.....135

Fig. 5.3. Western Channel Observatory station L4, south of Plymouth, UK.....136

Fig. 5.4. AMT-20 sample collection through a 24 arm sampling rosette.....140

Fig. 5.5. Average coccolithophore abundance along the AMT-20.....143

Fig. 5.6. Number of major capsid protein bands in each sample on the AMT-20.....145

Fig. 5.7. The number of MCP bands in each sample detected by DGGE.....145

Fig. 5.8. Evolutionary relationships of 58 MCP gene fragment sequences.....147

Fig. 5.9. CLUSTALW alignment of 67 (50 bp long) fragments of the MCP.....149

Fig. 5.10. CLUSTALW alignment of 16 amino acid fragment sequences of the.....150

Fig. 5.11. 3D model of the hypothetical structure of the full MCP protein.....150

Fig. 5.12. Evolutionary relationships of 31 SPT gene fragment sequences.....153

Fig. 5.13. 3D models of four different hypothetical structures of the SPT protein.....155

Fig. 5.14. Secondary structure predictions and CLUSTALW alignment.....156

Fig. 6.1. CsCl gradient from four lysates of the newly isolated coccolithoviruses.....174

Fig. 6.2. Visual inspection of the isolated DNA from the lysate concentrates of EhV-18, EhV-145, EhV-156 and EhV-164.....174

Fig. 6.3. Positive hybridisations of the four new coccolithovirus isolates.....176

Fig. 6.4. PCR fingerprints of four newly isolated coccolithoviruses.....178

Fig. 7.1. (Upper left) Geographical locations (characterised by different mean annual sea surface temperatures, ranging from 21°C in red to 8°C in blue).....196

Fig. App.1. The seven tRNAs present in the genomes of coccolithoviruses.....241

Guide to tables

Page

Table 1.1. Classification and general features of Emiliania huxleyi viruses.....18

Table 2.1. A list of reagents and their constituents used.....30

Table 2.2. Growth media for the *Emiliania huxleyi* CCMP 2090.....32

Table 2.3. EhV strains in the PML virus collection.....33

Table 2.4. Primer sets used in Chapter 4 and Chapter 5.....38

Table 2.5. PCR fingerprinting forward and reverse primers of 96 EhV-86 gene fragments.....39

Table 3.1. Sequencing statistics of seven newly sequenced coccolithoviruses.....59

Table 3.2. General characteristics of the completely or near completely sequenced coccolithoviruses to date.....64

Table 3.3. Unique protein coding sequences of nine EhVs..... 67

Table 3.4. Genes predicted to encode for tRNAs.....69

Table 3.5. Summary of the total number of genes or protein coding sequences (with and without function prediction) of nine sequenced coccolithoviruses……...72

Table 3.6. The presence (+) of homologous genes in the newly sequenced EhVs.....74

Table 3.7. 28 proteins previously identified in the virion of EhV-86……..….76

Table 4.1. The virus "fight club" culture conditions A-G.....102

Table 4.2. Free, unattached, average virus like particles (VLPs).....112

Table 6.1. Positive hybridisation or/and detection by PCR amplification.....179

Table App1. CLUSTALW alignment of a 209 bp fragment sequence.....215

Table App2. The 468 protein coding sequences in EhV-86…….….217

Table App.3. EhV-84 homology analysis to the draft genome (including introns) of *E. huxleyi* 1516……229

Table App.4. EhV-86 homology analysis to the draft genome (including introns) of *E. huxleyi* 1516……230

Table App.5. EhV-88 homology analysis to the draft genome (including introns) of *E. huxleyi* 1516……..…..231

Table App.6. EhV-99B1 homology analysis to the draft genome (including introns) of *E. huxleyi* 1516……...232

Table App.7. EhV-201 homology analysis to the draft genome (including introns) of *E. huxleyi* 1516……... 233

Table App.8. EhV-202 homology analysis to the draft genome (including introns) of *E. huxleyi* 1516 ……..234

Table App.9. EhV-203 homology analysis to the draft genome (including introns) of *E. huxleyi* 1516……...235

Table App.10. EhV-207 homology analysis to the draft genome (including introns) of *E. huxleyi* 1516…….....236

Table App.11. EhV-208 homology analysis to the draft genome (including introns) of *E. huxleyi* 1516…….237

List of symbols and abbreviations

<	Less than/lower than
>	More than/higher than
~	Approximately
µ	Specific growth rate
µL	Micro litre
µm	Micro metre
AA	Acrylic acid
aa (s)	Amino acid(s)
ACT	Artemis Comparison Tool
AFC	Analytical flow cytometer
AMT	Atlantic Meridional Transect
aPHC	Alkaline phytoceramidase
bp	Base pairs
BLAST	Basic Local Alignment Search Tool
BLASTn	Nucleotide Basic Local Alignment Search Tool
BLASTp	Protein Basic Local Alignment Search Tool
BSA	Bovine Serum Albumin
CCMP	Center for Culture of Marine Phytoplankton
CCN	Cloud condensation nuclei
COG(s)	Clusters of Orthologous Groups of proteins
CsCl	Caesium chloride
CDS(s)	Coding Sequence (s)
CDD	Conserved Domain Dataset
CTD	Conductivity Temperature Density
d^{-1}	Per day
DOM	Dissolved Organic Matter
DMS	Dimethylsulphate
DMSP	Dimethylsulphoniopropionate
DGGE	Denaturing Gradient Gel Electrophoresis
(ds)DNA	(double stranded) deoxyribonucleic acid
DNA pol	DNA polymerase
EC	English Channel
EDTA	Ethylenediaminetetraacetic acid
EhV	Emiliania huxleyi virus
e.g.	In Latin: exempli gratia (for example)
EM	Electron microscopy
ESTs	Expressed sequence tags
EtBr	Ethidium Bromide
EtOH	Ethanol
g	Gram
G	Gravitational force

GTE	Glucose-Tris-EDTA
h	Hour
HGT	Horizontal gene transfer
hGSL	Host derived glycosphingolipid
i.e.	In Latin: id est (that is)
IMG/ER	Integrated Microbial Genomes-Expert Review platform
kDa	Kilo Dalton
Kbp	Kilo base pair
km	Kilometre
L	Litre
LCB	Long chain base
m	Metre
min	Minute
M	Molar
MIGS	Minimum Information about a Genome Sequence
MCP	Major Capsid Protein
m.w.	Molecular weight
N	Nitrate
N/A	Not applicable or not available
ng	Nano-gram
nm	Nano-metre
nt	Nucleotides
NF	Norwegian Fjord
NCLDV	Nucleo-Cytoplasmic Large DNA Virus
MOI	Multiplicity of infection
ORF	Open reading frame
P	Phosphate
PP	Phosphate permease
PBS	Phosphate buffer saline
PCD	Programmed cell death
PCR	Polymerase chain reaction
PML	Plymouth Marine Laboratory
POM	Particulate organic matter
PEG	Polyethylene glycol
qPCR	Quantitative (real time) PCR
rDNA	Ribosomal DNA
RNA	Ribonucleic acid
rpm	Revolutions per minute
RRS	Royal Research Ship
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SSC	Saline-sodium citrate
TAE	Tris-acetate-EDTA
TE	Tris-EDTA
rRNA	Ribosomal ribonucleic acid

tRNA	Transfer ribonucleic acid
TBLASTN	Basic local alignment search tool utilising a translated nucleotide database
UV	Ultra violet
VCN	Virus copy number
vGSL	Virally derived glycosphingolipid
VLP (s)	Virus like particle (s)
w/v	Weight to volume
WCO	Western Channel Observatory

Acknowledgements

Writing an acknowledgments section is a bit like writing a thank you speech at the annual Oscars. The list can go on and on and there is never enough time or room to thank everyone enough. But there is however always enough room on the page to thank the most important people, the ones that made this research possible and also in the process made my time as a junior researcher in the last 4 years incredibly enjoyable. This task is made even easier when on your final thesis draft under the previously empty Acknowledgments section it is suddenly written in your supervisors' words: "I'd like to thank Mike for being the best looking and most fantastic PhD supervisor ever". Leaving the looks aside, he is of course right in every single way in stating what was obvious to everyone else: that he is indeed a fantastic supervisor. Often when fellow PhD students speak of their supervisors they don't sound very happy, and they always complain of how their supervisors were never there for them and how little they cared about their projects. However this is the complete opposite of what I have experienced and the last thing I can ever say about my supervisors. I was very lucky and fortunate at the same time to have two of the most incredible supervisors that one can wish for. When I first started my PhD everyone told me that I have a good balance of supervisors; i.e. between a good cop (Susan Kimmance) and a bad cop (Mike Allen). It turned out that they were both good cops in their own way, allowing me to develop the project and take it into a direction of my own interest (and in doing so; "boldly go where no coccolithovirus researcher has gone before").

On one side you had Mike; a young researcher and a leading scientist that transformed and innovated the algae and virus research group within PML in the last 10 years and

whose success speaks for itself. Mike was always there for me, and even in days during which I knocked on his office door on more than a dozen occasions (an in days when he obviously woke on the wrong side of the bed), he never sent me away and his door was always open to me. This I believe is the strongest asset that a student can wish for in their supervisors' character. When I expressed my concerns about a particular experiment or a laboratory approach Mike always advised me wisely on how to improve my chances of success (often by multiplying the complexity of an experiment 10 fold) and how to think around problems (often a complex diagram on a blue roll paper). At the end he would always say with a smile: "What could possibly go wrong?!" Remarkably due to his correct guidance throughout this PhD not much did go wrong, and for the majority of my time I had a smooth ride. I will never forget his master Yoda impressions during the microarray experiments and will never forget his "intense" field work days out on the local estuary and the "tough" times experienced by us both on our way to a workshop in the Netherlands. But more than his professionalism and ability to always be there for me, Mike was a great friend and one that I am glad to have met. I know that I will always be able to rely on Mike regardless of whether it is tomorrow or in 50 years time. His human qualities are rare in today's world. He is just superb in every possible way.

On the other side you also got Susan. The best part of my PhD as I see it was the incredible opportunity to participate in field work in the Arctic, Atlantic and Antarctic. Susan understood that doing a PhD is not all about just pipetting colourless liquids from one tube into another and staring at agarose gel images from dusk till dawn. She knew from her own experience how important it is for someone during their PhD to experience the world and see as much of it as possible, whilst they are still young and

full of energy. Susan made my dreams of reaching remote geographical locations into a reality. She provided constant support for the argument of why I should participate in field work and why it is important for me to go. In addition, she was also a great companion to travel with and work with in the Arctic. Almost everything I know today about field work and the extremely difficult task of preparing the logistics for such endeavours and applying yourself in the field derives from Susan. Moreover, although she dedicated herself into creating a family in the last two years of my PhD and was busier than ever before, she still managed to always advice and help in interpreting results, manuscript preparation and even analysing samples on the AFC with me and conducting the often dull routine of sampling during the virus "fight club" experiment of this thesis. Susan knew how important it is to have a second pair of hands and was always there when I needed her.

But there are also many others that made this research possible and contributed to who I am now as a person and also as a researcher. It was indeed the incredible Marine Microbiology lecture series by Colin Munn at the University of Plymouth during my undergraduate degree that inspired me to undertake such research and exposed me to the field of marine viruses. I wish to everyone in their life time to experience an incredible teacher such as Colin. His enthusiasm towards the subject and his guidance over the years will forever be an inspiration to me. Others that deserve to be on this page include: Paul Rooks (our laboratory technician) who was always there in the lab with his incredible and sometimes unusual tunes to cheer up the day, and Charlotte Worthy who was incredibly helpful in teaching me some of the basics of virus laboratory practice within PML and finding my way around at the beginning when I initially started my Masters degree at PML. Having gone through a thesis submission process at

Nottingham, Charlie was also very helpful in providing guidance on the final submission stages of this thesis. Others such as Simon Thomas, Bonnie Laverock, Karen Weynberg, and Emma Ransom, were also there to advice on molecular techniques and in doing so made it all look less scary than it actually was. In addition Mark Jones made the analysis of thousands of environmental samples possible by providing assistance in the DNA extraction of these samples, their PCR amplification and in teaching me the basics of DGGE. Many of the chapters in this thesis are a direct consequence of his help and for this I am also forever grateful. This research would not have been possible without the support of the crew members of the RRS James Cook, the BODC that allowed access to its metadata during the AMT-20 cruise, everyone involved in the acquisition of samples from the WCO in the last 15 years, and the crew members of the PML Quest. Also, I would like to thank Prof Gary Carvalho (Bangor University) and Prof Martin Broadley (University of Nottingham) for serving as my external and internal examiners respectively, for showing an interest in this thesis and providing a final and crucial feedback on the science conducted during my PhD.

I would like to thank also my family and friends for their support love and friendship over the years. My old friends are still there while I have made many new ones. The old ones were always interested in my progress and kept in touch despite the geographical distance, while the new ones always provided much entertainment on Friday nights in the local Yard Arm pub after work and also with endless amounts of outdoor fun, full of laughter and unforgettable moments. But one cannot finish an acknowledgement section without thanking the most important people in his life, his family. My Mom and Dad were always there for me, and have provided an endless support without which I would not have succeeded. They were always proud of me and encouraged me from an early

stage to succeed and be driven the way I am today. There are not enough words in this world to express the enormous gratitude and love I have towards them and I hope that I will one day be able to give my own children at least a fraction of what they gave to me over the years.

Last but not least, I would like to express my gratitude and endless love to Maeve, my partner in life, for always being there for me, never judging, and always being patient with me, even in times when I was very much unbearable. Thank you for standing beside me and keeping me fed with yummy food; your love knows no boundaries. I couldn’t do any of this without you.

Author's declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Graduate Committee.

This study was funded with the aid of a NERC PhD student research grant, the Oceans 2025 program, and the Gordon and Betty Moore Foundation grant for the sequencing of viruses.

List of published peered reviewed articles and manuscripts derived directly from this PhD thesis (articles are attached at the end of the thesis):

Nissimov, J.I., Worthy, C.A., Rooks, P., Napier, J.A., Kimmance, S.A., Henn, M.R., Ogata, H. and Allen, M.J. Draft genome sequence of the coccolithovirus EhV-84. *Standards in Genomic Sciences*, 2011, 5**,** 1-11.

Nissimov, J.I., Worthy, C.A., Rooks, P., Napier, J.A., Kimmance, S.A., Henn, M.R., Ogata, H. and Allen, M.J. Draft Genome Sequence of the Coccolithovirus Emiliania huxleyi Virus 203. *Journal of Virology*, 2011, 85 (24), 13468–13469.

Nissimov J.I., Worthy, C.A., Rooks, P., Napier, J.A., Kimmance, S.A., Henn, M.R., Ogata, H. and Allen, M.J. Draft genome sequence of the coccolithovirus EhV-202. *Journal of Virology*, 2012, 86 (4), 2380-2381.

Nissimov, J.I., Worthy, C.A., Rooks, P., Napier, J.A., Kimmance, S.A., Henn, M.R., Ogata, H. and Allen, M.J. Draft genome sequence of four coccolithoviruses: EhV- 88, EhV- 201, EhV-207 and EhV-208, *Journal of Virology*, 2012, 86 (5), 2896-2897.

Nissimov, J.I., Jones, M., Napier, J.A., Munn, C.B., Kimmance, S.A. and Allen, M.J. Functional inferences of environmental coccolithovirus biodiversity, 2013, ViroSin, 1995-820X.

Nissimov, J.I., Worthy, C.A., Rooks, P., Napier, J.A., Kimmance, S.A. and Allen, M.J. Coccolithoviruses: 'genomic thieves and metabolic thugs' *Manuscript in preparation*.

Nissimov, J.I., Napier, J.A., Kimmance, S.A. and Allen, M.J. The Coccolithovirus fight-club' *Manuscript in preparation*.

Nissimov, J.I., Napier, J.A., Kimmance, S.A., and Allen, M.J. Draft genome sequence of four coccolithoviruses: EhV- 18, EhV- 145, EhV-156 and EhV-164. *Manuscript in preparation*.

Other published peered reviewed articles derived from work conducted during this PhD:

Sperling M., Piontek J., Gerdts G., Wichels A., Schunck H., Roy A.-S, La Roche J., Gilbert J., **Nissimov J.I.**, Bittner L., Romac S., Riebesell U., Engel A. Effect of elevated CO2 on the dynamics of particle-attached and free-living bacterioplankton communities in an Arctic fjord. *Biogeosciences,* 2013, 10, 181-191.

Roy, A.-S., Gibbons, S.M., Schunck, H., Owens, S., Caporaso, J.G., Sperling, M., **Nissimov, J. I**., Romac, S., Bittner, L., Riebesell, U., LaRoche, J., and Gilbert, J.A. Ocean acidification shows negligible impacts on high-latitude bacterial community structure in coastal pelagic mesocosms. *Biogeosciences*, 2012, 10, 555–566.

Allen, M.J., Tait, K., Mühling, M., Weynberg, K., Bradley, C., Trivedi, U., Gharbi, K., **Nissimov, J.**, Mavromatis, K., Jensen, C.N., Grogan, G., and Ali S.T. Genome Sequence of *Stenotrophomonas maltophilia* PML168, Which Displays Baeyer-Villiger Monooxygenase Activity' *Journal of Bacteriology*, 2012, 194 (17), 4753-4754.

Scientific seminars and conferences at which work from this thesis was presented:

2010 Society for General Microbiology (SGM) spring conference in Edinburgh, UK. Poster presentation: "Host-virus dynamics within the *Emiliania huxleyi* system: the effects of host fitness, phosphate availability and virus strain variability".

2010 International Symposium on Microbial Ecology 13 in Seattle (ISME13), USA. Oral presentation: "Ecological and functional biodiversity in a marine algal virus system".

2011 University of Nottingham, annual postgraduate symposium. Poster presentation: "The genetic diversity of Coccolithoviruses and their biogeography".

2011 British Antarctic Survey, 10th Antarctic Funding Initiative Workshop. Oral presentation: "The genetic diversity of Coccolithoviruses and their biogeography".

2011 Aquatic Virus Workshop 6 (AVW6), Netherlands. Oral presentation: "Coccolithoviruses: same, same but different".

2013 Association for the studies of limnology and oceanography (ASLO), New Orleans, USA. Oral presentation: "Protein fold differences in the coccolithovirus-encoded serine palmitoyltransferase (SPT) and its possible implications for the demise of *Emiliania huxleyi*"

1. INTRODUCTION

In this thesis an attempt was made for the first time to understand the functional relevance of the genetic diversity displayed by viruses infecting the eukaryotic algal species *Emiliania huxleyi*; to determine the dynamic interactions of these viruses with their unicellular host; and to extrapolate the consequences of these interactions to global biogeochemical cycling in the worlds' oceans. In this chapter the role of marine phytoplankton and their co-occurring viruses will be explained with an emphasis on elucidating the ecological importance of the *Emiliania huxleyi*-virus system.

1.1. Putting the algal-virus system into perspective

Planet Earth is referred to by many as the Blue Planet. This is because more than 70% of the Earths' surface is covered by water. However, many plant biologists and marine microbiologists will disagree with the above statement. Surprisingly, half of our planet can be considered green, i.e. consists of either photosynthetic (primary producers) organisms visible to the naked eye (terrestrial trees and bushes, seaweeds, sea grasses, kelps, and corals) or of photosynthetic microorganisms that vary in shape and size, and are invisible as individuals to the naked eye and can be detected only by either microscopy and/or the use of specific molecular techniques, or when in vast populations remotely from afar. This invisible fraction includes small bacteria such as cyanobacteria (e.g. *Prochlorococcus* and *Synechococcus*), and larger unicellular protists such as diatoms and haptophytes that comprise the diverse phytoplankton group (Partensky et al., 1999; Stanier and Cohen-Bazire, 1977; Thomas et al., 1978; Jordan and Chamberlain, 1997) (Figure 1.1).

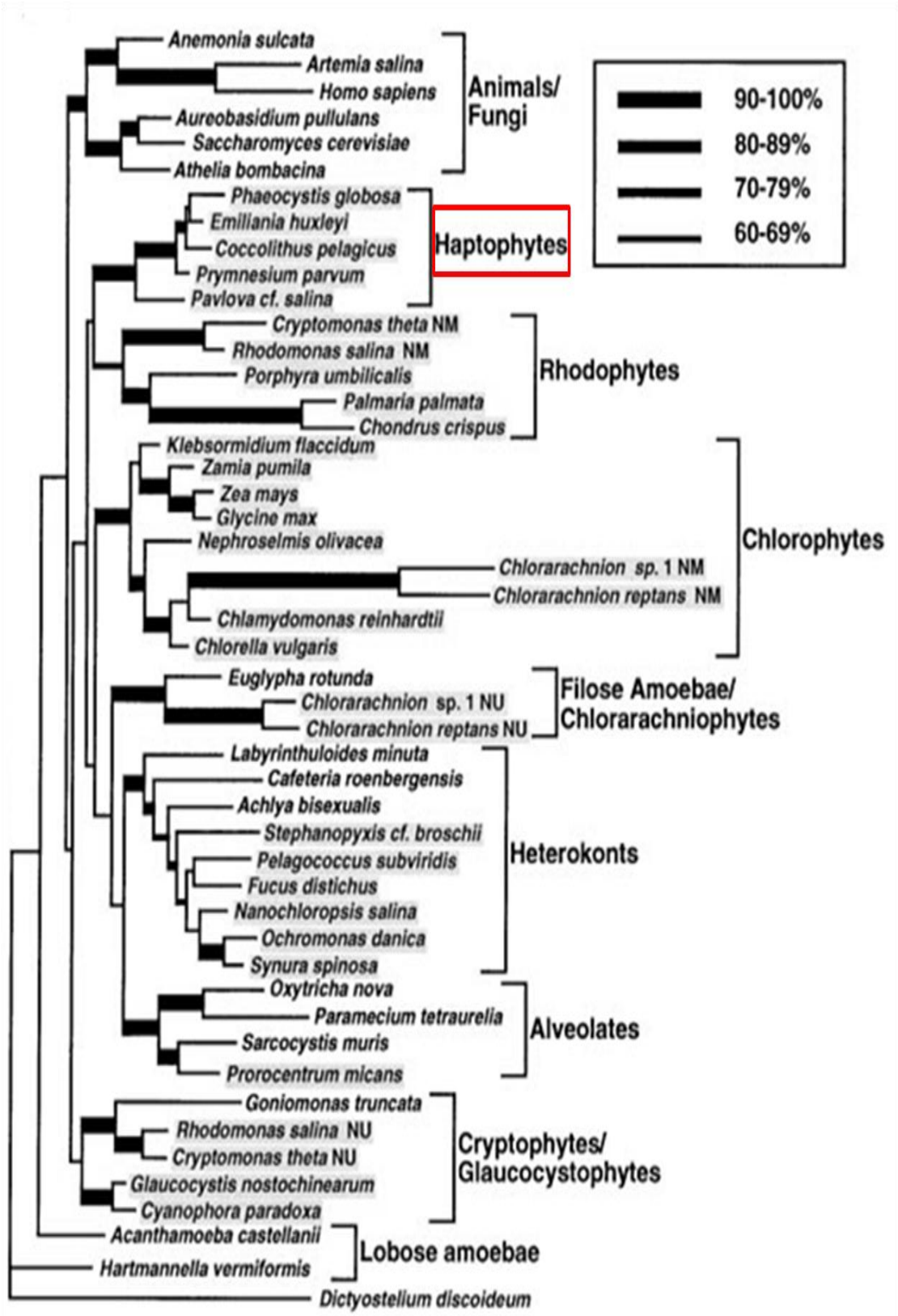


Fig. 1.1. Phylogenies of major algal groups based on the analysis of their small rDNA subunit by the maximum likelihood method (adapted from Battacharya and Medlin, 1998). In red is indicated the position of the Haptophyte algae, in which the coccolithophores and the sub species *E. huxleyi* are a part of. Photosynthetic taxa are highlighted in grey.

Viruses, at a fraction of the size of bacteria and plankton, are the most abundant biological entities on the planet. There are 10^{30} virus like particles in the worlds' oceans, or approximately 10^{7} per mL of seawater (Suttle, 2005). In fact viral infection of bacteria and phytoplankton is considered today as a major cause of marine microbe mortality in that it reduces global primary productivity, and acts as a major driving force in biogeochemical cycling, gene transfer events between hosts and viruses, and generally the evolution of life (Suttle, 2007). Several model systems have recently been used for the study of marine viruses, their hosts, and the interactions between the two (Seaton et al., 1995; Schroeder et al., 2002; Martinez et al., 2007; Marston et al., 2012). Perhaps one of the most relevant of these systems to biogeochemical cycling in oceans results in the termination of blooms of the cosmopolitan haptophyte alga *Emiliania huxleyi,* and is instigated by the viral genus *Coccolithoviridae* (large double stranded DNA viruses of the family *Phycodnaviridae*).

1.2. The coccolithophore *Emiliania huxleyi* and its distribution

The photosynthetic single celled calcifying marine plankton *Emiliania huxleyi* (*E. huxleyi*) is one of the most important primary producers on our planet; it dominates the planktos (at certain times of the year) (Feng et al., 2008) and has a worldwide distribution (Gattuso and Buddemeir, 2000*). E. huxleyi* belongs to the coccolithophore family within the *Haptophyte* genus of which there are more than 300 living species described (Jordan and Chamberlain, 1997). Coccolithophores produce calcium carbonate plates (known as coccoliths) intracellularly, which are expelled from the cell, and subsequently arranged around its exterior (Brownlee and Taylor, 2004). The chalky deposits (or detached coccoliths) that remain following cellular disintegration can be seen from space (Figure 1.2) and are found in sediments throughout the

worlds' coastal regions, and in fact, it is these algal calcium carbonate plates that form the modern day cliffs of Dover in the UK (Dupont et al., 1993).

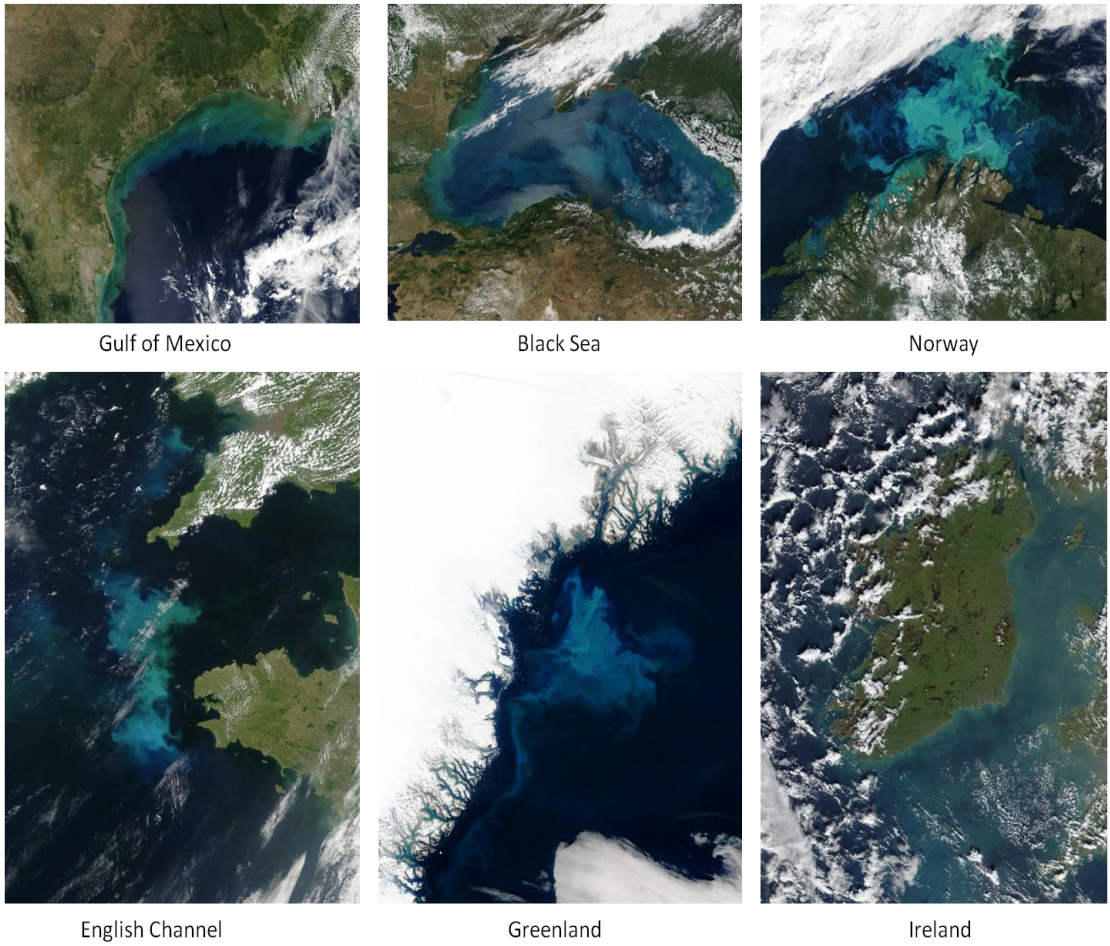


Fig. 1.2. Satellite images of coccolithophore blooms in the global ocean captured by NASA satellites (http://visibleearth.nasa.gov/).

The movement of coccolithophores in the water column and the ecological niche which they occupy is governed by the ocean currents, the availability of nutrients (i.e. phosphate and nitrate), and appropriate temperature and light levels for photosynthesis (Brown, 1995). Once these conditions are favourable for coccolithophore growth, one or more species (the most dominant of which is *E. huxleyi*) form extensive blooms (often visible from space), where the algal cellular density can typically reach 10^7 cells per mL in the sun-lit part of the water column (usually the upper 200 m) (Honjo, 1976; Boeckel and Baumann, 2008). The bio-geographical range of *E. huxleyi* and

other related coccolithophorid species is diverse and they can usually be found at low abundance throughout the global ocean (Eynaud et al., 1999; Poulton et al., 2010; Shutler et al., 2012). *E. huxleyi* has been found globally, for instance: in the North Atlantic (Parke and Dixon, 1976; Iglesias-Rodriguez et al., 2006), English Channel (Wilson et al., 2002; Schroeder et al., 2002), Pacific (Iglesias-Rodriguez et al., 2006), Southern Ocean, (Cook et al., 2011), Mediterranean Sea (Vilicic et al., 2009), Black Sea (Caraus, 2003), in inshore waters such as Norwegian fjords (Iglesias-Rodriguez et al., 2006; Frada et al., 2006; Janknegt et al., 2009; Guo et al., 2012) and around sub-Antarctic and Antarctic islands (Findlay et al., 2005).

1.2.1. The influence of coccolithophores on global climate

The importance of annual algal blooms and their ecological relevance is almost indisputable. There has been increasing evidence recently that coccolithophores such as *E. huxleyi* play a crucial role in the regulation of our climate by being the central component of several self regulating mechanisms (Charlson et al., 1987; Rijssel and Gieskes, 2002; Franklin et al., 2010). The physiological and morphological characteristics of the coccolithophorid surface, allows them to display some unique properties. Due to the presence of calcium carbonate plates on the outer surface of the cells, the calcite allows coccoliths to scatter more light than they absorb and the rate at which light is reflected back into space increases during blooms (Moore et al., 2012). As a consequence of this decrease in radiative forcing the ocean cools down. This may eventually have a negative effect on the growth of coccolithophores because of self-shading (as they and other phytoplankton and bacteria below the bloom need light for growth) decreasing their abundance, thus serving indirectly as a positive feedback on the warming of the ocean (since there are less cells to reflect the light back to space

and therefore more light available for other photosynthetic organisms at depth). In addition, the increasing cellular density of the upper layer of the ocean during a bloom can act as a cover that prevents, to a certain extent, the penetration of light into the deeper layers of the ocean, serving as an additional negative feedback mechanism.

The construction of calcium carbonate plates in coccolithophores also plays a crucial role in the net concentration of free carbon dioxide on our planet. Coccoclithophores are primary producers and therefore transform CO_2 into organic matter via photosynthesis. In this aspect, all phytoplankton including the coccolithophores are the same. Coccolithophores however also remove carbon from the global carbon cycle in an inorganic form through the formation of their calcium carbonate chalk liths; yet, crucially, the carbonate chemistry of calcification in coccolithophores actually produces CO_2 through the following conversion: $Ca^{2+} + 2HCO_3^- \rightarrow CaCO_3 + H_2O + CO_2$ (Raven and Crawford, 2012).

Finally, coccolithophores may also play an important role in climate regulation by influencing cloud albedo (Figure 1.3) due to the release of the sulphur compounds dymethyl sulphide (DMS) and dimethylsulfoniopropionate (DMSP) (Charlston et al., 1987). The concentration of these compounds in *E. huxleyi* is at least 100 times greater than in diatoms (Keller et al., 1989) and is released when the cells are dying due to factors such as nutrient starvation (Tilman et al., 1982); virus induced mortality (Bratbak et al., 1993), zooplankton grazing (Bratbak et al., 1995; Levasseur et al., 1996; Kirchman, 1999), or programmed cell death (Bidle et al., 2007). Due to the mixing of the upper ocean with the atmosphere and the sheer force of winds and waves, these compounds are released into the atmosphere where they eventually form

cloud condensation nuclei (CCN) that fuel the development of clouds. The formation of extra clouds increases the backscatter of light into space, the result of which is the decrease in light penetration into the atmosphere and ocean. As a consequence phytoplankton becomes limited by light and DMS and DMSP production is inhibited. Work still continues today to try and fully understand the mechanisms of this complicated feedback system that involves biological and physical and chemical factors alike. The main criticism to this hypothesis comes from the lack of quantitative understanding of particle development processes in the marine-atmosphere boundary layer and the lack of evidence that links global CCN concentrations to DMS fluxes, suggesting that the role of DMS in climate control, as suggested in the CLAW hypothesis (the acronym derives from the first letter of the authers surnames: Charlson, Lovelock, Andreae and Warren), might be low (Ayers and Cainey, 2007; Woodhouse et al., 2010).

Hence, coccolithophores and the agents that contribute to their mortality may be part of a self-regulating feedback mechanism on climate control and primary production, where a potential shift in biotic or abiotic conditions, whether natural or anthropogenic, can have a profound effect on the outcome of a bloom, its role in biogeochemical cycling, carbon export to the deep ocean, and their bioavailability to upper trophic levels as a food source in the marine environment.

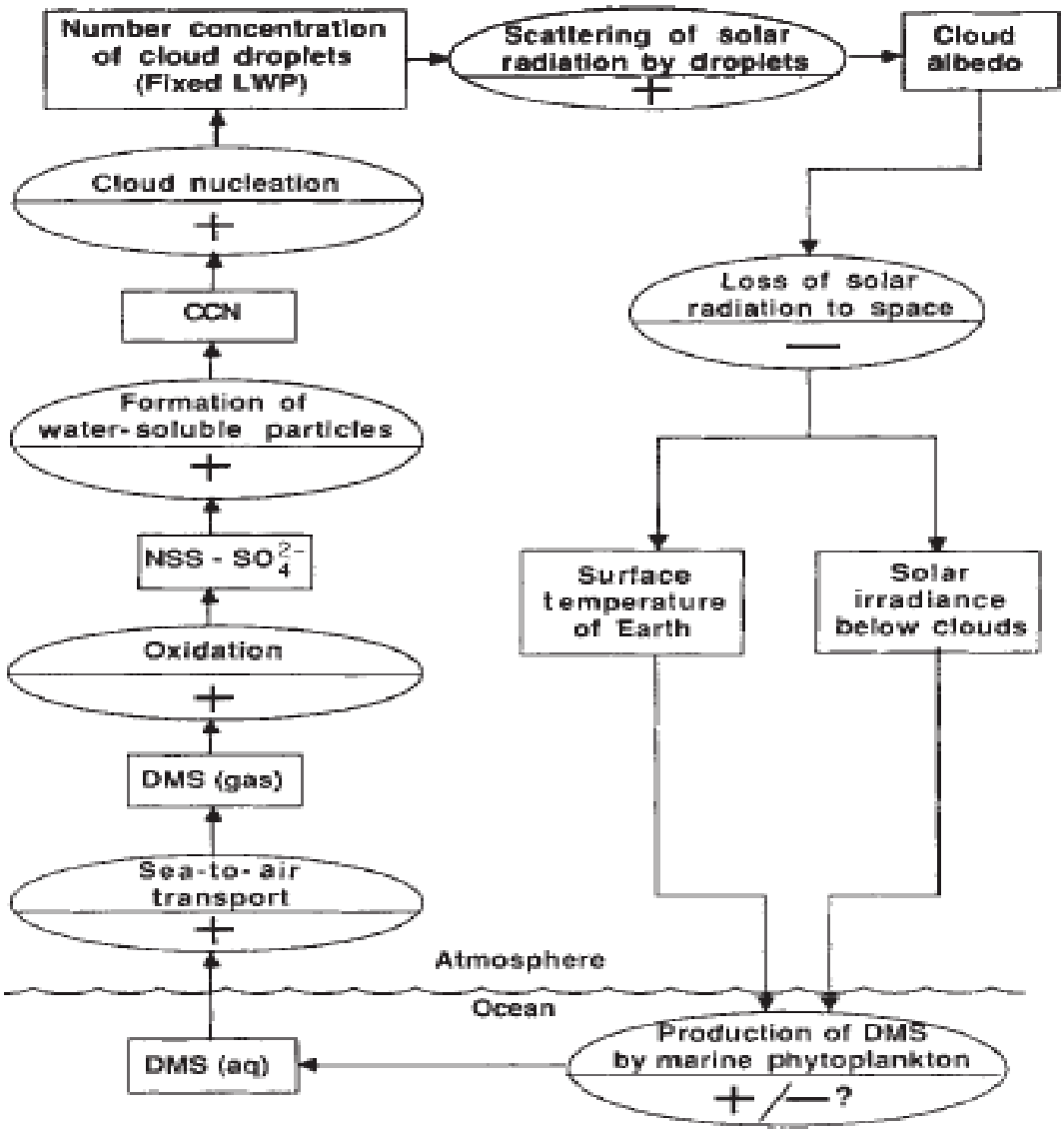


Fig. 1.3. Conceptual diagram of the climate feedback loop proposed in the CLAW hypothesis, adapted from Charlson et al. (1987), in which positive and negative effects are marked by a + or a -.

1.3. Viruses at a glance

Viruses (in Latin virus=poison) are the smallest biological entities on our planet. They consist of a nucleic acid genome (single or double stranded DNA or RNA) enclosed by a simple protein coat known as a capsid (Carter and Saunders, 2007). Some viruses are enclosed by a bi-layer lipid membrane envelope, a feature that often helps in the protection of the viral core from ultraviolet degradation and physical damage. In some viruses the envelope also aids in attachment and adsorption of the virus into the

infective host cell (Allen and Wilson, 2011). Viruses are obligate intracellular parasites; therefore the host is essential for virus replication as they require the host cellular machinery and metabolism for their own propagation.

The initial contact of viruses to their hosts occurs by passive diffusion during which they use the exposed cellular structures of their hosts in order to attach themselves and enter into the cells. There are three main types of virus replication. Lytic infection during which viruses attach to host cells and introduce their nucleic acids into the cells, followed by the production of a large number of virus progenies that are eventually released by bursting the cell and as a consequence killing it. The second is chronic infection in which the release of virus progeny is not lethal, i.e. the host cells release these newly formed viruses by budding over several generations. The third is lysogeny in which the viral nucleic acids are integrated into the genome of the host and become part of it. It replicates as an integrated genetic material (sometimes referred to as a prophage) with the replication and reproduction of the host until its lytic infection is triggered by stress to the host (Fuhrman, 1999).

The life cycle of most viruses can be generalised in the following steps: binding to receptors on the membrane of a target cell via its own receptors on the envelope or capsid, entry into the cell and un-coating of its envelope (or injection of its DNA into the cytoplasm of its host), the expression of virus genes, the replication of the viral genome, the assembly of the newly formed virions in the host cytoplasm, and finally the release from the cell of the newly assembled virions by either bursting the cell envelope or budding (Allen and Wilson, 2011; Carter and Saunders, 2007). For unicellular organisms the last step often results in the lysis and death of the host cell.

The study of viruses (and subsequently the majority of scientific knowledge on their life strategies and the cellular mechanisms by which they infect, replicate and persist) has been acquired in the last 100 years mainly by investigating medically important disease mediators in humans such as Smallpox (Behbehani, 1983), Polio (Tebbens et al., 2013), Influenza (Zimmer and Burke, 2009), Hepatitis (Hoofnagle et al., 2012), and HIV (Hladik and McElrath, 2008) and to a lesser extent commercially important viruses such as Tobacco Mosaic virus (Scholthf, 2004) and Rabies (Bourhy et al., 1999). However we now know that viruses are not just cellular parasites and disease mediators. They play an active role in the evolution of life and impact biotic and abiotic factors in the environment (Suttle, 2005; Breitbart and Rohwer, 2005). With regards to the viruses under study in this thesis, it is now well established that *E. huxleyi* specific viruses (EhVs) are often instrumental in the termination of *E. huxleyi* blooms, thereby affecting the abundance and diversity of *E. huxleyi* populations and the biogeochemical cycling of carbon and sulphur in the marine environment. By infecting bacterial and microalgal populations (such as the coccolithophores), viruses directly control the local abundance of these microbes, and thus the amount of nutrients recycled through the marine trophic web (i.e. dissolved organic matter: DOM and particulate organic matter: POM), the release of climate regulation compounds (i.e. DMS and DMSP), and the export of carbon into the deep ocean (Weitz and Wilhelm, 2012).

1.4.1. Viruses in the oceans

Viruses are now considered to contain the largest pool of genetic variability on Earth, and have inspired many to embark on oceanic voyages in order to explore the mysteries hidden in the four letter code of marine virus genomes. Several expeditions

in recent years have embarked on long haul missions around the worlds' oceans to try to detect the presence and impact of viruses via means of metagenomics, metatranscriptomics, and phylogenetic molecular based fingerprinting techniques. For instance, during the Sorcerer I and II expeditions (Yooseph et al., 2007), researchers filtered large quantities of seawater from different depths around the globe and then used culture independent techniques to capture the diversity of marine viruses and gain an insight into their biogeography. During another research cruise, on board the French vessel TARA (Karsenti et al., 2011) scientists size fractioned the water samples in order to capture the diversity from larger zooplankton at the top of the food chain, all the way to the smallest fraction at the bottom, the viruses. A different approach was undertaken in two other cruises, the first in 1999 in the North Atlantic (Wilson et al., 2002; Martinez et al., 2012) and the second in the North Atlantic near Iceland (Rowe et al., 2011). Both applied specific molecular techniques (polymerase chain reaction [PCR] and Denaturing Gradient Gel Electrophoresis [DGGE]) followed by single gene genomics in order to map and characterise the distribution of coccolithophores and their co-occurring coccolithovirus communities. They revealed that the phylogenetic diversity of coccolithoviruses in natural systems is greater than that observed in the laboratory, and that this diversity has been so far under-represented.

1.4.2. The "Viral Shunt" and "killing the winner"

Although viruses in the marine environment have many ecological roles, the major one that has been attributed to them is their role as "viral lubricators" through the "Viral Shunt" (Wilhelm and Suttle, 1999; Suttle, 2007). This process refers to the retainment of nutrients at lower trophic levels in the marine food web. When

phytoplankton cells are grazed, energy is available to higher trophic levels, whereas, viral-mediated mortality forces the food web towards a more regenerative system in which DOM and POM recycling occurs (Figure 1.4), stimulating bacterial production (Wilhelm and Suttle, 1999). Sequestration of materials in viruses, bacteria and dissolved matter may lead to better retention of nutrients in the sunligh zone in virus-infected systems, because more material remains in these small non-sinking forms. This may be particularly important for nutrients that potentially limit productivity (such as N, P and Fe), which are relatively concentrated in bacteria compared to eukaryotes.

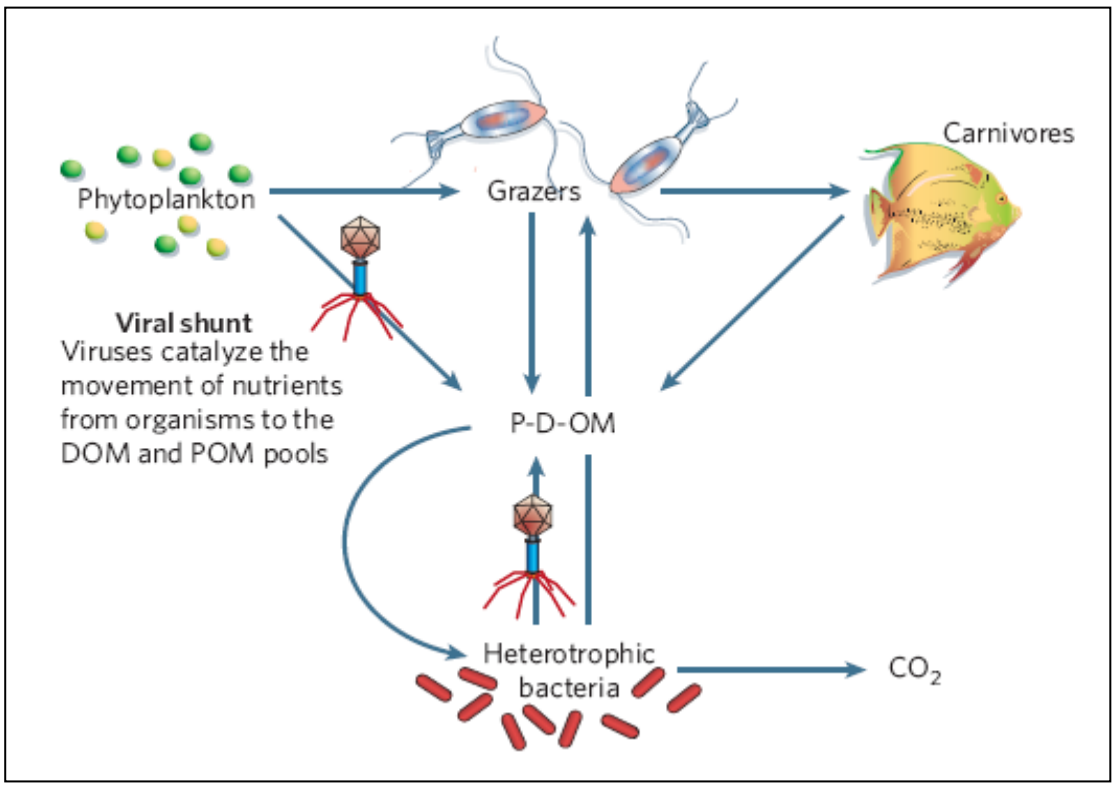


Fig. 1.4. Virus mediated catalysis of biogeochemical cycling, adapted from Suttle (2005). Virus infection of phytoplankton, whether algal or bacterial, short circuit the flow of nutrients and carbon to higher trophic levels by lysing cells and causing the flux of these compounds to the pool of DOP and POM.

When viruses lyse bacterial or algal cells in the marine environment, the intracellular contents within the dying cells leaks out as particulate or dissolved organic material and enter the pool of DOM and POM. The two are distinguished based on their

filtration behaviour via polycarbonate filters. Organic material that is retained on filters with a pore size of 0.2 μm is POM while material that passes this pore size is the DOM. DOM is the largest pool of organic carbon in the ocean and perhaps even the largest in the biosphere, available to heterotrophic bacteria (which use carbon containing organic matter as a source of energy), cyanobacteria, eukaryotic phytoplankton, and protists (Ogawa and Tanoue, 2003).

The DOM pool includes amino acids and nucleic acids, usually recycled in the euphotic zone (the zone at which light is available for photosynthesis), whilst more rigid carbon-rich structures such as cell walls are exported to the deeper waters, where they can feed other microbes in the "microbial loop" (Azam et al., 1998; Pomeroy et al., 2007). Finally, virus activity also drives the formation of "marine snow" by releasing "sticky" components from within cells, thus influencing many biological and microscopic physical–chemical processes, and facilitating aggregation and sinking of material from the euphotic zone (Peduzzi and Weinbauer, 1993; Furhman, 1999).

Although viruses are usually species specific, there seems to be a "killing the winner" scenario, where the most abundant and active bacterial or phytoplankton populations at a given time within a given niche, are also the most susceptible to infection and eventually termination (Thingstad and Lignell, 1997). In such a model, viruses will prevent any single host species whose population size was determined by protozoan grazing, from dominating the whole community, resulting in the "killing the winner" scenario (Thingstad and Lignell, 1997) where high population density, characterised usually by lower diversity (i.e. the most abundant and active bacterial or phytoplankton population at a given time and a given niche is also the most

susceptible to infection) inevitably promotes the activity of viruses which rely on contact collisions for their replication.

The vast majority of evidence for this hypothesis (in the context of this thesis) has come from mesocosm and open ocean studies of *E. huxleyi* blooms and their associated viruses the *Emiliania huxleyi* viruses (EhVs) (Wilson et al., 2002; Martinez et al., 2007; Sorensen et al., 2009) in which the blooms are eventually terminated largely due to infection by these viruses. The hypothesis makes sense in both ecological and evolutionary terms in that when the bloom is at its peak the water is much denser and packed with cells closer to each other, therefore providing an easier, faster route for infection. The majority of the host population is killed by the co-occurring viruses and only a small fraction that is resistant to these infections survive, giving rise to the seed for the next bloom the following year or/and when conditions are favourable again. This small resistant population is the "winner" of the next bloom cycle and on an evolutionary time scale the viruses infecting these "winners" have to adapt and change each time in order to keep up and maintain their own replication and survival, consistent with the Red Queen hypothesis proposed by Van Valen (1973). In addition, the "viral-host stable co-existence" theory proposed by Thyrhaug et al. (2003) suggests that the co-existence of viruses and their hosts allows the recovery of host populations following viral infection. This is due to an efficient feedback mechanism in which the viral lysis during the peak of an infection cycle results eventually in the low abundance of hosts and as a consequence a reduction in the rate of infection of these hosts by the viruses. Hence the function of viruses in the "microbial loop" and the "viral shunt" goes beyond the scope of biogeochemistry, into the realm of co-evolutionary host parasite arms race for survival.

1.5. *Phycodnaviridae*: viruses of phytoplankton

The *Phycodnaviridae* family consists of viruses that infect either marine or freshwater single celled micro-alga. There are currently six genera in the *Phycodnaviridae*: Chlorovirus (PBCV-1, infecting freshwater single-celled green algae), Phaeovirus (EsV-1, infecting multi-cellular filamentous brown algae), Prasinovirus (MpV-SP1 and OtV-1, infecting single- celled green algae), Prymnesiovirus (CbV-PW1, infecting single-celled marine algae), Raphidovirus (HaV01, infecting single-celled harmful algal bloom species) and Coccolithovirus (EhV-86, infecting single-celled red algae). All the viruses in this family have an icosahedral virion structure and are similar in their general morphological characteristics (Dunigan et al., 2006). Although their particle size range is 100 to 200 nm in diameter, genetically, these viruses are very different, and variability exists even between virus strains that infect the same host organism (Allen et al., 2006a, 2006b; Fitzgerald et al., 2007a, 2007b, 2007c; Weynberg et al., 2009; Nissimov et al., 2011a, 2011b, 2012a, 2012b; Pagarete et al., 2013; Jeanniard et al., 2013).

The first algal virus was isolated from *Micromonas pusilla* in the late 70s (Mayer and Taylor, 1979). With the discovery and development of phylogenetic markers such as the virus encoded DNA polymerases (Knopf, 1998) and genes encoding for capsid protein structure (Mannige and Brooks, 2010), it became possible to not only distinguish different viruses within the same family but also different strains of the same virus (Rowe et al., 2011; Martinez et al., 2012). Astonishingly, in 20 years of research thereafter of the *Phycodnaviridae*, it was observed that although morphologically similar there is very little similarity on the genomic scale (i.e. genetic structure and content) of members of this family. For instance only 14 genes were

identified to be common to viruses infecting PBCV-1, EsV-1 and EhV-86 (Wilson et al., 2009; Dunigan et al., 2006), predicted to encode for D5-type ATPase, DNA polymerase, ATPase, helicase, capsid protein, thiol-oxidoreductase, D6R-type helicase, protein kinase, VLTF2-Like transcription factor, proliferating cell nuclear antigen, ribonucleotide reductase (large and small subunit), A494R-like uncharacterized protein, and thioredoxin.

1.5.1. Coccolithoviruses

The coccolithoviruses, a group of viruses that infect coccolithophores have been shown in the last two decades to be a major cause of *E. huxleyi* bloom termination in the English Channel (Wilson et al., 2002a), Norwegian fjords (Martinez et al., 2007), Gulf of Main (Rowe et al., 2011), and the North Sea (Wilson et al., 2002b). The coccolithoviruses studied to date are all large dsDNA viruses and belong to the NCLDV (Nucleo-Cytoplasmic Large DNA viruses) class of viruses within the *Phycodnaviridae* family (Table 1.1) (Dunigan et al., 2006). They have the potential to terminate a 100,000 km^2 bloom in a matter of days (Holligan et al., 1983). The coccolithoviruses have icosahedral virions, are lytic (latent period of 4-6 h, Dunigan et al., 2006), with a replication cycle in the nucleus of their algal host and assembly of new virions in the cytoplasm (Mackinder et al., 2009). The exact adsorption mechanism of these enveloped viruses into the cell has not been fully established however it is believed that they enter the cell either by means of endocytosis that is followed by fusion of their envelope (and capsid) with the vacuole membrane of their host or by direct fusion with the plasma membrane of the cell (Mackinder et al., 2009). The release from the cell is by budding, potentially through the action of lipid rafts, and the burst size is 400-1000 viruses per infected cell (Castberg et al., 2002).

1.5.2. General characteristics of Emiliania huxleyi viruses

Emiliania huxleyi viruses (EhVs) are a recently discovered group of coccolithoviruses that infect the marine single celled coccolithophorid alga *Emiliania huxleyi*. All EhVs known to date, have been isolated from either the English Channel or the Norwegian Fjord of Raunefjorden, and they all have a 160-180 nm diameter icosahedral structure, with a genome of approximately 400 kbp (Wilson et al., 2005). The abundance of these giant viruses (or Giruses as some virologists have named them due to their enormous size) typically reach 10^5- 10^7 per mL in natural seawater under bloom conditions (Martinez et al., 2007; Sorensen et al., 2009) and 10^8 -10^9 per mL under laboratory culture (personal observations). They are bigger than some small bacteria, and can be visualised both by electron microscopy and also by fluorescence staining techniques such as Analytical Flow Cytometry.

1.5.2.1 Emiliania huxleyi virus- 86, 163, and 99B1

The model coccolithovirus strain EhV-86 (AJ890364) was isolated in 1999 from a coccolithophore bloom in the English Channel (Wilson et al., 2002a) and its genome was sequenced in its entirety in 2005 to reveal a circular genome of 407,339 bp (Wilson et al., 2005). Two further strains, EhV-99B1 and EhV-163 were isolated in 1999 and 2000 respectively from a Norwegian Fjord and have also had their genomes sequenced (Allen et al., 2006b; Pagarete et al., 2013). An initial comparison of EhVs has shown that EhV-86, EhV-163 and EhV-99B1 differ in a very small number of genes (Allen et al., 2006b; Pagarete et al., 2013), despite the fact that these strains were isolated from geographically distinct locations.

So far the most notable differences observed between strains relate to the presence or absence of a gene encoding a phosphate permease (ehv117 in EhV-86) and other genes with no current assigned function. ehv117 has been replaced (at the same locus) in all Norwegian originating strains examined to date by a putative endonuclease encoding gene (Allen et al., 2006b). Potentially, both gene products could play a role in determining intracellular phosphate availability. One by scavenging from the external environment (permease) and the other from the degradation of host genomic material (endonuclease). This however has not been confirmed, and laboratory studies to test this theory are required.

1.5.3. Emiliania huxleyi virus 86: genomics, proteomics and life cycle

So far, *Emiliania huxleyi* virus 86 (EhV-86) has been extensively used as a model system to study the complex interactions between a virus and a host in the laboratory on a genetic and proteomic level (Allen et al., 2006b, 2006c, 2008, 2011, Allen and Wilson, 2006). EhV-86 has a 170-175 nm diameter icosahedral capsid; the largest genome known for a virus in this family (Wilson et al., 2005). Interestingly, the full genome sequence of this model strain revealed that at least four of its genes encode proteins involved in sphingolipid biosynthesis (Wilson et al., 2005; Monier et al., 2009) and have been suggested to actively promote metacaspase activity in the host cell (Bidle et al., 2007).

This finding was of great importance because sphingolipid biosynthesis leads to the formation of ceramide (Merrill, 2002), a proven suppressor of cell growth (Obeid et al., 1993) and inducer of programmed cell death (PCD) and apoptosis (Mao and Obeid, 2008). The coccolithoviruses have acquired these genes from the host by means of horizontal gene transfer (HGT) (Pagarete et al., 2009), and have been

suggested by some to manipulate programmed cell death through the activity of these gene products (Monier et al., 2009). Thus, coccolithoviruses may induce and actively recruit host metacaspases as part of their replication strategy (Bidle et al. 2007).

Table 1.1. Classification and general features of Emiliania huxleyi viruses (EhVs), adapted from SIGS (adapted from Nissimov et al., 2011b).

Parameter	**Phylogeny**
Classification	Domain: *Viruses, dsDNA viruses, no RNA stage* Class: *NCLDV (Nucleo-Cytoplasmic Large DNA)* Family: *Phycodnaviridae* Genus: *Coccolithovirus* Species: Emiliania huxleyi virus
Virion shape	Icosahedral
Habitat	Oceanic, Coastal
Biotic relationship	Obligate intracellular parasite of *Emiliania huxleyi*
Pathogenicity	Lytic virus of *Emiliania huxleyi*
Particle diameter	150-200 nm
Genome size	138-420 kbp

In addition, proteomic analysis of the virion structure revealed many viral proteins with potential roles associated with viral budding, signalling, anti-oxidation, adsorption and host range determination (Allen et al., 2008). 80% of the proteins encoded by EhV-86 are unique and have no matches in the public protein database. There is a set of approximately 50 ‘core’ virus genes in the NCLDV group that usually encode for genome replication, expression and virion structure, and only 25 of these are in the genome of EhV-86 (Wilson et al., 2005). The presence of a large novel promoter region in the genome of EhV-86, i.e. a region of unique promoter

elements with no homology to the public domain that are associated with the first transcriptional phase of viral CDSs, together with the presence of virus encoded RNA polymerase suggests that the expression of many EhV transcripts occurs in the cytoplasm and that they have their own transcriptional machinery.

To explain this further Allen et al. (2006c) showed that the transcription of virus genes during infection by EhV-86 is bi-phasal. The first stage was shown to be dominated by the expression of a group of CDSs 1 h post infection of which each CDS in this group was associated with a novel promoter element within the genome of EhV-86, while the rest of the viral CDSs were expressed 2-4 hours post infection in what was described as a secondary stage. Hence the authors hypothesised that the viral nucleic acids targeted the host nucleus immediately upon infection in order to transcribe the first stage genes by the aid of the host encoded RNA polymerase that recognised the viral promoter elements (described as early nucleus transcription), followed by the second stage transcription in the cytoplasm by the aid of the virus encoded RNA polymerase (described as late cytoplasm transcription). More evidence to support this theory is required hence the current knowledge is merely circumstantial. A schematic representation of the described life cycle of coccolithoviruses, is seen in Figure 1.5.

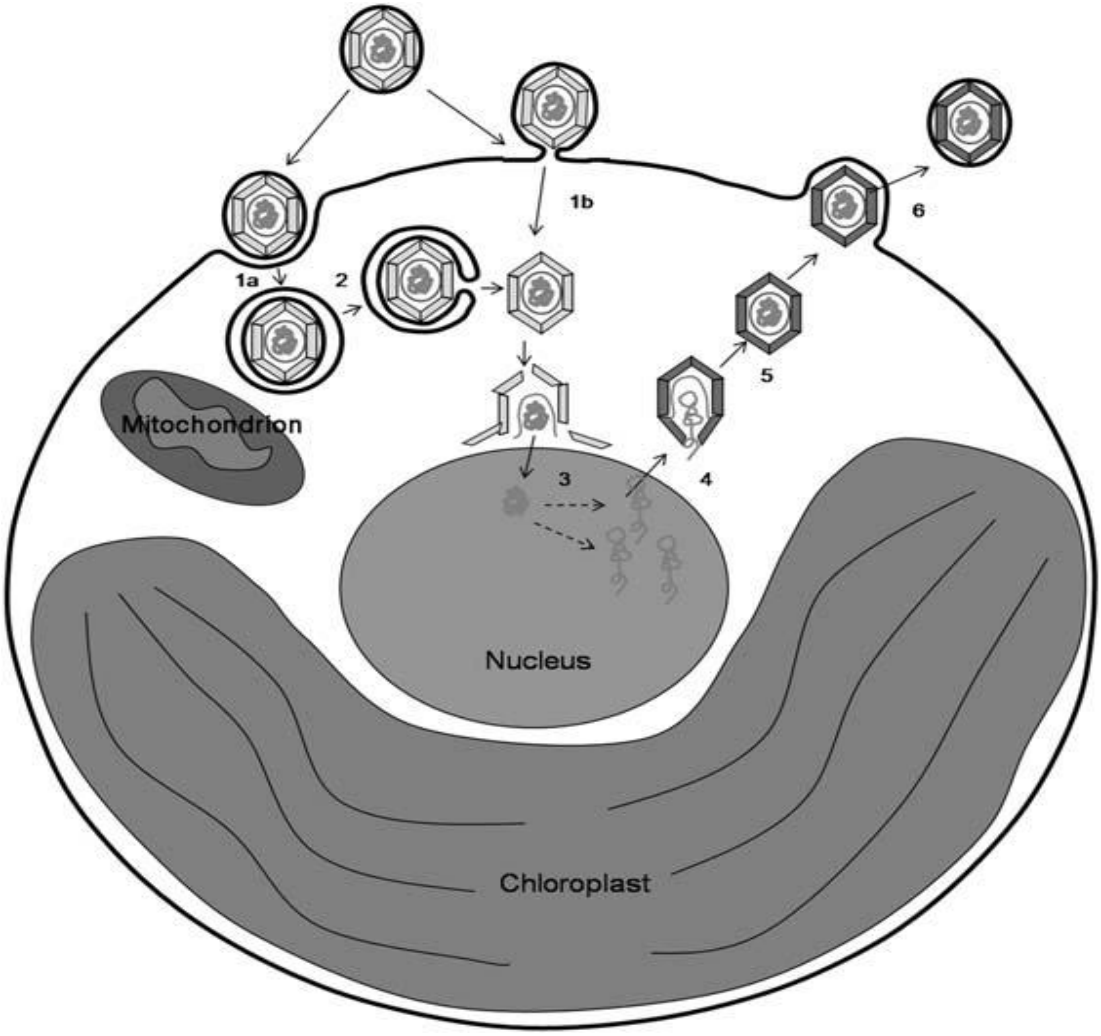


Fig. 1.5. Schematic hypothetic representation of the proposed life cycle of the coccolithovirus EhV-86 adapted from Mackinder et al. (2009). The enveloped virus enters the host either by endocytosis (step1a) followed by fusion of its envelope with the cell vacuole membrane (step2) or by fusing its envelope with the host plasma membrane (step 1b). The viral genome is released after the capsid breakdown (step 3) post the rapid capsid encapsulation of the nucleoprotein. Then the viral genome enters the nucleus in which promoter sequences are expressed by the host encoded RNA polymerase. The rest of the viral genes are expressed within the cytoplasm in which also the capsid assembly of new virions occurs (step 4). The newly assembled viruses are transported to the plasma membrane (step 5) from which they are released by budding (step 6).

1.5.4. Horizontal gene transfer in coccolithoviruses

Horizontal Gene Transfer (HGT) is a molecular or/and cellular process by which genetic information is transferred between either multi-cellular organisms or microbes and their associated viruses (Koonin et al., 2001; Gogarten and Townsend, 2005; Liu et al., 2010; McDaniel et al., 2010; Vogan and Higgs, 2011) by means that are independent of normal reproductive strategies (i.e. not parent to offspring). In the case

of viruses this can be simply described as: the acquisition of genetic material function from either the host or from other co-infecting viruses. For example, during the replication phase, host DNA can sometimes become incorporated into the DNA of the virus, resulting in its stable acquisition and integration into the virus genome.

In bacteria, HGT is largely responsible for the reoccurrence of diseases and resistance to antimicrobial agents (Martinez, 2009; Warnes et al., 2012), while bacterial viruses (phages) were shown to possess similar genes to those present in their infective host (Canchaya et al., 2003), potentially also acquired by HGT. One such example is the *Prochlorococcus* cyanobacteria (or blue green alga) and its phage (Lindell et al., 2004). The phage has at least four genes involved in photosynthesis, which derive directly from its host (Sullivan et al., 2005). Moreover, it was also shown that "satellite" viruses or virophages (viruses that "infect" other viruses and hinder their successful replication within the host cell) of Mimivirus (a virus that infects amoebae) have genes that are similar to genes found in the giant Mimivirus and Mamavirus (La Scola et al., 2008; Claverie and Abergel, 2009).

In eukaryotes, HGT has been shown extensively in plant pathogens and fungi (Richards et al., 2011; Renner and Bellot, 2012), however the most interesting transfer of genes relevant to algal-virus systems and this study occurs between *Emiliania huxleyi* and the coccolithoviruses. However, the total number of homolog EhV encoded genes among different coccolithovirus strains to their host is currently unknown.

1.5.5. Programmed Cell Death and Apoptosis

Programmed cell death (PCD) refers to a process in cells during which a series of coordinated intracellular events causes the eventual death of the cell. This process results from specific gene expression and is mediated by caspase (and metacaspase) enzymes. The result of this series of biochemical reactions is a cell in the state of apoptosis, characterised by nuclear organelle and DNA fragmentation, cell shrinkage, loss of membrane integrity and the creation of apoptotic bodies of cellular content (Suderman et al., 2008). In many cases environmental stress such as nutrient limitation (Bidle and Bender, 2008; Thamatrakoln et al., 2012) and enforced darkness (Segovia and Berges, 2009); as well as virus infection (Bidle et al., 2007; Vardi et al., 2012) can induce the intracellular signalling of PCD and the activation of caspase (a precursor of cell death). For instance the rapid demise of dinoflagellate blooms in the worlds' oceans has been attributed to PCD that follows environmental stress such as energy limitation for growth (i.e. nutrient limitation), the result of which is lack of cellular division and in some cases production of reactive oxygen species that promote cell death (Franklin et al., 2006).

Apoptotic PCD is essential for cellular life, termination of abnormal cells, and has a role in development and immunity in multi-cellular and single celled organisms (Hedrick et al., 2010). In the medical realm there are many examples of viruses that impact the PCD pathway. For instance, it was established that virulent strains of influenza A and B induced structural changes typical of those observed in cells undergoing apoptosis (Hinshaw et al., 1994). Similarly, it was shown that respiratory pathogens, such as the human coronaviruses that invade the central nervous system, influence the induction of host initiated cell death by using specific virus encoded

glycoproteins (Favreau et al., 2012). However there are still many uncertainties with regards to viruses and their apoptotic effect in marine microalgae (Franklin et al., 2006). For instance, EhVs have been implicated as apoptosis mediators in *E. huxleyi* infected cells (Bidle et al., 2007; Wilson et al., 2009). EhV infection results in: rapid cellular component degradation, reduction in the photosynthetic activity and efficiency of the cells (Evans et al., 2006; Llewellyn et al., 2007), and up-regulation of metacaspase expression (Bidle et al., 2007). In addition, the genomes of EhVs contain genes for the biosynthesis of sphingolipids and ceramide (see previous section), also known as a promoters of PCD and apoptosis via the sphingolipid pathway and ceramide signalling. Hence the presence of these PCD promoting genes in the genome of EhVs is surprising, given that it is in the virus "best interest" to maintain a healthy host cell for as long as possible in order to enable its own successful replication. It was proposed by Vardi et al. (2012) that these genes are important in the ability of EhVs to manipulate the host PCD machinery and as a consequence regulate the host-virus interactions and the fate of the infected host cells.

1.5.6. Molecular markers for phylogeny

When attempting to study the diversity of a particular group that is phylogenetically related, or in trying to distinguish different strains of the same species in an environmental DNA sample, it is of great importance to choose genes that are relatively conserved in the target organism that is under study. For instance, the 16S rDNA gene that codes for the 16S ribosomal RNA subunit of prokaryotic ribosomes, is used extensively for the phylogenetic study of Bacteria and Archaea (Woese and Fox, 1977). In viruses however not a single gene is common across all viruses, hence the difficulty in using such a marker. There are however conserved genes within

different lineages of viruses. DNA polymerase (DNA pol) for instance is used to distinguish viruses in the *Phycodnaviridae* family (Gimenes et al., 2012), while the major capsid protein (MCP) gene has also been used to study the phylogeny of coccolithoviruses (Martinez, et al., 2012; Rowe et al., 2011).

Thus the selection of marker genes such as the viruses encoded MCP and DNA pol can be very useful if attempting to study the general diversity and distribution of viruses in the environment. However, the use of conserved genes for phylogeny in the case of viruses helps little if one aims at investigating the functional characteristics or functional diversity of a virus population of a known ecological impact such as the coccolithoviruses. The use of these markers will reveal who is in a particular niche but not necessarily what they are doing in this niche and whether their actions as seen from RNA and trancriptomic data are related to the conditions they are found in and adapted to.

For instance, transcriptomic data that reveals the functional characteristics of members of the *Coccolithoviridae* family are now available from laboratory studies during which the researchers used microarray methods during the infection cycle of a susceptible host (Allen and Wilson, 2006; Allen et al., 2006c, 2007). However, relating these data to the functional characteristics of coccolithoviruses in the natural environment would be difficult by solely using relatively conserved phylogenetic markers that change little from one strain to another on a spatial and temporal scale.

1.5.6.1. Current status of coccolithovirus diversity

The use of MCP as a marker for coccolithovirus diversity has nevertheless proven useful in fingerprinting coccolithovirus communities in environmental samples, revealing the succession of EhVs during blooms in the ocean and in demonstrating their phylogenetic diversity. For instance, the MCP analysis of nucleotide sequence samples collected across 6000 km of open ocean revealed that the diversity of EhV genotypes was underrepresented by strains capable of infecting an *E. huxleyi* host in the laboratory and that across the sampled transect many MCP genotypes were represented only once while others were represented multiple times (Rowe et al., 2011). This indicated that location specific distinctions of EhV genotypes in the natural environment exist and that these distinctions can be observed by the use of the MCP marker gene. Also, in two separate msocosm studies in a Norwegian fjord (experiments in large enclosures of natural water), it was shown that it is possible to trace the propagation of an infection and the succession of EhVs by using the MCP marker gene, and gain an insight into the genotypic diversity of EhV communities during the propagation of these induced blooms (Schroeder et al., 2003; Martinez et al., 2007; Sorensen et al., 2009). Remarkably, it was possible by the use of the MCP to establish that although the EhV diversity was high during both mesocosm studies at the beginning of the blooms, the diversity decreased towards the end of the blooms, dominated by the same two EhV genotypes in both mesocosm experiments. Hence the MCP marker also revealed that some EhVs have an advantage over others during a bloom and that their succession potential differs.

In a different study, due to the conservation of DNA sequence, the MCP marker gene was used to elucidate the distinct relatives of the modern coccolithovirus, by

coccolithovirus community fingerprinting sediment record samples from the Black Sea that span more than 7000 years into the past (Coolen, 2011). By looking at the diversity of fragments of the MCP gene in these sediments, correlating the data with the diversity of the host from the same samples, and the integration of paleoenvironmental data, the author revealed that although the same group of coccolithoviruses and hosts persisted for centuries, a major shift in hydrologic and nutrient regimes 2500 years ago selected for a new group of coccolithoviruses.

Hence, although the MCP marker enables the possibility to assess host virus interactions and their phylogenetic diversity in the modern ocean and in the distant past, it does not reveal the diversity in their functional characteristics; i.e. it does not reveal the reasons behind why some coccolithovirus strains are better in bloom termination demise than others and what gives them a fitness advantage over others. Therefore, using more variable genes such as the virus encoded serine palmitoyltrnasferase (SPT), that encodes for a specific function in a metabolic pathway of proven importance during infection, is essential in order to understand if the functional diversity of coccolithoviruses is conserved or varies among strains from different geographical settings and among strains isolated from the same location but at different times. Thus the approach taken in this thesis was to use in conjunction both of these markers in order to study the phylogeny, distribution and function of coccolithoviruses in samples from a variety of temporal and spatial settings.

1.6. The Plymouth Marine Laboratory virus collection and the "new viruses"

The Plymouth Marine Laboratory (PML) virus collection is a growing collection of viruses that have been isolated from the English Channel and various coastal waters

around the UK during the last 15 years. Water samples are routinely tested against potential algal hosts and suspected infections/inhibitions are diluted to extinction and purified before storage for future analysis (Chapter 6). Currently there are in culture viruses infecting *Ostreococcus tauri* and also more than 15 different virus strains capable of infecting one or more strains of *E. huxleyi*.

1.7. Thesis objectives

In order to learn more about these fascinating coccolithoviruses, their diversity, and their interactions with their algal hosts (and ultimately, their impact on marine food web dynamics and biogeochemical cycling), this thesis attempts to answer the following compelling questions:

A) What is the genomic diversity of *Emiliania huxleyi* viruses, and are there substantial functional differences between coccolithovirus strains isolated from different locations and different times?

B) Are there strain specific differences in the infection dynamics of EhVs and how do these differences manifest during infection?

C) What is the phylogenetic and functional diversity of coccolithoviruses in different geographical settings, and what is their distribution during non-bloom conditions in the Atlantic Ocean?

To address objective **(A)**, seven coccolithoviruses that originate from water concentrated samples in the PML archive were sent for full genome sequencing at the Broad Institute in the US, were annotated, and compared to the reference genomes of EhV-86 and EhV-99B1 (Chapter 3). Additionally, four new viruses were isolated,

analysed by microarray and PCR, and sent for sequencing at the NERC Biomolecular Facility in Liverpool, UK (Chapter 6). To address objective **(B)**, the infection dynamics (i.e. phenotypic characteristics during infection) of two coccolithoviruses with a known genomic difference, EhV-86 and EhV-207 were investigated and compared by means of Analytical Flow Cytometry (AFC), and quantitative Polymerase Chain Reaction (qPCR) (Chapter 4). Finally objective **(C)** was addressed by: i) participating in a research expedition to the Atlantic Ocean and obtaining DNA samples from various depths at different geographical locations, and ii) by investigating DNA samples previously collected in the North Sea and from the English Channel, for phylogenetic and functional diversity of coccolithoviruses (by means of PCR, DGGE and eventually capillary sequencing) (Chapter 5).

2. MATERIALS AND METHODS

2.1. GENERAL MATERIALS

2.1.1. Laboratory consumables and reagents

Consumables and reagents for general laboratory practice were obtained from Invitrogen, Sigma-Aldrich, Promega, Fisher–Scientific and Qiagen. The ultrapure 18 Ω water (MQ) used in this study was obtained from a Millipore water purification unit.

Table 2.1. A list of reagents and their constituents used throughout this PhD research.

Reagent	**Constituents**
50 × TAE (Severn Biotech LTD)	242 g Tris, 57.1 mL Glacial Acetic Acid, 100 mL 0.5 M EDTA (pH 8) + MQ to final volume of 1 L
1 × TAE	200 mL of 50 × TAE in 10 L of MQ
TE buffer (10:1, pH 8.0)	1 mM EDTA pH 8.0, 10 mM Tris HCl
Ethidium Bromide (Sigma)	10 mg / mL H_2O
DNA loading dye	0.05 % Bromophenol blue, 50 % Sucrose; 10 mM EDTA pH 8.0
DGGE loading dye	80% glycerol, 20% dH_2O (MQ), 0.25% bromophenol blue, 0.25% xylene cyanol
GTE buffer	50 mM glucose, 25 mM Tris HCl, 10 mM EDTA
1 × Lysis buffer	20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM Na_2EDTA 1 mM EGTA, 1% Triton, 2.5 mM sodium pyrophosphate 1 mM beta-glycerophosphate, 1 mM Na_3VO_4, 1 µg/mL leupeptin
1 × PBS (pH 7.2)	137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 mM KH_2PO_4

2.1.2. Origin of *Emiliania huxleyi* strain

Emiliania huxleyi (Lohmann) Hay and Mohler 2090 was originally obtained from the Provasoli-Guillard Centre for the cultivation of Marine Phytoplankton (CCMP) in Maine, USA. It was isolated on the 5th of July 1991 from the South Pacific (02.66 S', 82.71 W') near the coast of South America (https://ncma.bigelow.org/node/1/strain/CCMP2090). Its optimal temperature for growth ranges between 18 -22°C. The cell length of the alga usually ranges from 3-7 µm and most cells lack external scales or calcium carbonate plates (i.e. coccoliths). This strain has been maintained in the Plymouth Marine Laboratory culture collection in the UK since the late 90's. The strain used in this study is from this collection.

2.1.3. Media preparation and algal cultivation

All *E. huxleyi* CCMP 2090 cultures were grown in an f/2 media (Table 2.2). In order to prepare the f/2 media for the algal cultures, fresh seawater was collected weekly from the L4 site south west of Plymouth, located 10 km offshore (50°15'N; 04°13'W). The filtered sea water was then pre-filtered through a 1.6 µm Sartorious filtration system (Millipore GFF filters) and then 30 kDa filtered in the laboratory using tangential flow filtration (section 2.2.3), and autoclaved at 126°C for 20 min. Seawater was then aged for several months at 16°C in the dark, and finally distributed to 1 L sterile glass bottles with added 0.2 µm syringe filtered and autoclaved nutrients (Phosphate, Nitrate, Vitamins and Trace Metals), and stored in a dark cool place (16°C) until used. *Emiliania huxleyi* strain CCMP 2090 was cultivated in f/2 nutrient media in a controlled temperature cabinet or room under a light/dark cycle of 16/8 h respectively, at a temperature of 18°C throughout all the experiments in this thesis.

Table 2.2. Growth media for the *Emiliania huxleyi* CCMP 2090 stock and experiment cultures.

Medium	Components
f/2 (minus Si)	884 µM $NaNO_3$, 36 µM $NaH_2PO_4 \cdot H_2O$, 11.7 µM Fe EDTA$\cdot 6H_2O$, 0.9 µM $MnCl_2 \cdot 4H_2O$, 12 µM $Na_2EDTA \cdot 2H_2O$, 0.04 µM $CuSO_4 \cdot 5H_2O$, 0.03 µM $NaMoO_4 \cdot 2H_2O$, 0.08 µM $ZnSO_4 \cdot 7H_2O$, 0.05 µM $CoCl_2 \cdot 6\ H_2O$, 0.37 nM vitamin B_{12}, 2 nM biotin, 0.3 µM Thiamine HCl

2.1.4. Coccolithovirus isolates

The virus strains used in this study (Table 2.3) were obtained from either the Plymouth Marine Laboratory (PML) virus collection library in the UK, originally isolated during natural blooms from either the English Channel or the Raunefjorden fjord in Western Norway, or from various waters around the UK.

2.1.5. Virus lysate stock generation

For each EhV strain a continuous batch of fresh viral lysate was prepared every three to six months and prior to experiments, in order to ensure that the viruses were kept fresh at all times and had a high amount of infective virions within them. For this, 500 mL of an exponentially growing *E. huxleyi* CCMP 2090 culture (~4×10^6 mL^{-1}) was infected with 5-10 mL of a current virus lysate stock. After 14 days, or when the culture was lysed completely (complete loss of pigmentation, assessed using analytical flow cytometry, section 2.2.2) the new lysates were filtered (via 0.2 µm pore size filters, Millipore Express) to remove *E. huxleyi* cell debris and the filtrate (i.e. filtered lysate) stored in the dark at 4°C until required.

Table 2.3. EhV strains in the PML virus collection and their isolation information. EC- English Channel; NF- Norwegian Fjord, LM- Lossiemouth, England; F- Fife shore, Scotland, NIA-no information available.

Strains	Location	Isolation date	Lat / Long	Depth (m)	Reference
EhV-84	EC	26/07/1999	50°15'N, 04°13'E	15	Nissimov et al., 2011a (this study)
EhV-86	EC	30/07/1999	50º13'N, 04º09'W	15	Wilson et al., 2002
EhV-88	EC	26/07/1999	50°15'N, 04°13'W	5	Nissimov et al., 2012a (this study)
EhV-201	EC	27/07/2001	49°56'N, 04°19'W	2	Nissimov et al., 2012a (this study)
EhV-202	EC	27/07/2001	50°00'N, 04°18'W	15	Nissimov et al., 2012b (this study)
EhV-203	EC	27/07/2001	50°00'N, 04°18'W	15	Nissimov et al., 2011b (this study)
EhV-207	EC	01/08/2001	50°15'N, 04°13'W	5	Nissimov et al., 2012a (this study)
EhV-208	EC	01/08/2001	50°15'N, 04°13'W	5	Nissimov et al., 2012a (this study)
EhV-163	NF	20/07/2000	60º20'N, 05º20' E	surface	Allen et al., 2006b
EhV-99B1	NF	07/1999	60º20'N, 05º20' E	surface	Pagarete et al., 2013
EhV-18	EC	2008	50°15'N, 04°13'W	Surface	This study
EhV-145	LM	2008	51°72'N, 03°28'W	Surface	This study
EhV-156	EC	2009	50°15'N, 04°13'W	Surface	This study
EhV-164	F	2008	NIA	Surface	This study

2.2. GENERAL METHODS

2.2.1. Analytical Flow Cytometry of *Emiliania huxleyi*

For the enumeration of cellular density and abundance of *Emiliania huxleyi*, sub-samples of one mL were removed from the culture flasks and assessed by analytical flow cytometry (AFC) using a FACScan Flow Cytometer (Beckton Dickinson, Oxford, UK) with a 15 mW Argon laser exciting 488 nm with a standard filter setup, for one minute at flow rate of ~75 µL min^{-1}. To reduce coincidence during analysis, dense cultures were 10-fold serial diluted in 0.2 µm filtered seawater according to their density. Mean red cell fluorescence (proxy for chlorophyll) and side and forward light scatter data from each sample were analysed and recorded on CellQuestPro™ software. The flow rate was calibrated everyday with Milli-Q water for five minutes. Flow Cytometric data was analysed with WinMDI 2.9 (2000) software.

2.2.2. Analytical Flow Cytometry of *Emiliania huxleyi* viruses

Viral abundance was also determined using AFC, following the protocol of Brussaard (2000). Sub-samples of one mL were taken from either freshly made lysate stocks or from infected cultures during an experiment, and fixed with 50% glutaraldehyde in final concentration of 0.5 %, for 30 minutes at 4°C. The samples were then snap frozen in liquid nitrogen and stored at -80°C until analysis. Before analysing the samples on the Flow Cytometer the samples were defrosted in a 35°C water bath. 50 µL from each sample was diluted with 450 µL of sterile and filtered TE 10:1 buffer at pH 8 to give a 10-fold dilution. Samples were further diluted (up to 500-fold) if the concentration was subsequently deemed too high for an accurate measurement as observed with the Flow Cytometer. The samples were then transferred to 5 mL Flow Cytometry tubes (BD Biosciences) to which 5 µL of SYBR Green I was added at a

final dilution of 5×10^{-5} the commercial stock. Tubes were then incubated in a dark 80°C water bath for 10 min and then cooled in the dark at room temperature for five min prior to analysis on the FACScan (Beckton Dickinson, Oxford, UK) Flow Cytometer for one min at a flow rate of ~11 µL min^{-1}. *Emiliania huxleyi* viruses (EhVs) were discriminated from other particles using green fluorescence versus side scatter. Data files were analysed using the CellQuestPro™ software.

2.2.3. Isolation of coccolithoviruses

Water was collected weekly from the L4 time series at the Western Channel Observatory south of Plymouth in the English Channel, UK, for the last 15 years and from other locations around the UK coast. Before concentrating one L subsamples from this water by tangential flow filtration (TFF) (Suttle et al., 1991) 100 fold through a 100,000 molecular weight cut off polyethersulfone membrane filtration devise (Vivaflow 200), the water was pre-filtered via 0.45 µm filters (Millipore membrane filters). Subsequently the TFF resulted in 30 mL volume concentrated filtrates. These concentrates were stored in the PML L4 sample archive for future use. Then, exponentially growing *E. huxleyi* CCMP 2090 cultures in sterile 48 well plates were assayed using the dilution to extinction method as previously described by Wilson et al. (2002) and modified in this study. Briefly, 100 µL aliquots from each concentrated library sample were added to 900 µL of the *E. huxleyi* host. Following host lysis, i.e. loss of coloration due to cellular disintegration, samples were removed and serially diluted from 10^{-1} to 10^{-10} in UV filter sterilized Milli-Q water and once again 100 µL added to 900 µL of the *E. huxleyi* host. The well plates were stored until analysis in a controlled temperature (CT) room under a light/ dark cycle of 16/8 at ~18°C and light intensity of ~90 µM photons m^{-2} s^{-1}. The highest dilution at which a

host culture was cleared in the well plates was further diluted to 10^{-10} in the same way twice more. Each time the highest dilution at which infection was observed was removed and further diluted as above. Then aliquots of one mL were taken out from the last concentration/infection round and added to one L of exponentially growing *E. huxleyi* CCMP 2090 for the generation of the first lysate of the newly isolated viruses (known as the primary lysate). Once lysed, the infected cultures were 0.2 µm filtered (Millipore membrane filters, cellulose acetate/ cellulose nitrate), analysed by AFC for the quantification of free EhV VLPs and stored in the dark at 4°C.

2.3. MOLECULAR METHODS

2.3.1. Primers

The oligonucleotide primers for the detection of the virally encoded major capsid protein (MCP) of EhVs have been used successfully in previous studies (Rowe et al., 2011; Martinez et al., 2007, 2012) and were therefore used in this study. New primers were manually designed and used for the detection of the serine palmitoyltransferase (SPT) gene ehv050 after the CLUSTALW multiple alignments of the full SPT gene sequences from ten different EhV laboratory isolates (Table 2.3). The two sets of molecular probes were used to evaluate the genetic diversity of EhVs in environmental samples and the presence of functionally important genes such as SPT in natural assemblages (Chapter 5). All primers in this thesis were manufactured by Sigma-Aldrich®.

Strain specific primers were also designed for *Emiliania huxleyi* virus strain 86 (EhV-86) and *Emiliania huxleyi* virus strain 207 (EhV-207) for qPCR analysis during the virus "fight club" experiments (Chapter 4). Their specificity was validated experimentally by PCR amplification (Chapter 4). In addition, 96 primer pairs (Table 2.5), previously designed to amplify EhV-86 and EhV-163 genes in a shotgun survey for the identification of novel functions for biotechnological exploitation (personal communication with Dr. Mike Allen) were also used in Chapter 6. All the proposed primers in Table 2.4 and 2.5 were analysed for primer dimers, annealing temperature, hair pins and other possible PCR artefacts by the OligoAnalyzer 3.1 platform (IDT Integrated DNA Technologies).

Table 2.4. Primer sets used in Chapter 4 and Chapter 5 of this study.

Primer	**Sequence**	**Fragment size (bp)**	**Reference**
MCP-F1	5'-GTCTTCGTACCAGAAGCACTCGCT-3'	284	Schroeder et al., (2003)
MCP-R1	5'-ACGCCTCGGTGTACGCACCCTCA-3'		
MCP-F2-GC	5'CGCCCGGGGCGCGCCCCGGGCGGGGCGGGGGCACGGGGGGTTCGCGCTCGAGTCGAT C-3'	135	Schroeder et al., (2003)
MCP-R2	5'-GACCTTTAGGCCAGGGAG-3'		
SPT-F1	5'-GTTGGATATCCCGCAACACC-3'	703	This study
SPT-R1	5'-CAATGTCGCCAATGTTGGC-3'		
SPT-F2-GC	5'CGCCCGGGGCGCGCCCCGGGCGGGGCGGGGGCACGGGGGGGAATCTCGCGCGCG-3'	335	This study
SPT-R2	5'-CGCGGTCCACATGTACC-3'		
SPT-F2	5'-GGAATCTCGCGCGCG-3'		
qPCR(EhV-86)-F	5'-GCACAACTTTCAACAATTCG-3'	209	This study
qPCR(EhV-86)-R	5'-TCAGCTCAACTTTTGGATCA-3'		
qPCR(EhV-207)-F	5'-CATAGGGTTGGCAATATTCA-3'	353	This study
qPCR(EhV-207)-R	5'-TTCGAAACAACTTGGTCAAC-3'		

Table 2.5. PCR fingerprinting forward and reverse primers of 96 EhV-86 gene fragments used in Chapter 6. The primers include restriction sites at the ends as these were initially designed for cloning.

Agarose lane	Target name	Gene product name	Forward primer	Reverse primer
1	ehv018	putative endonuclease	CCAGGGACCAGCAATGGGTATTAATGGGCTATCTCGTC	GAGGAGAAGGCGCGTTACTTTCGTTTCTGGCCCACAG
2	ehv039		CCAGGGACCAGCAATGCAAGAGGCATATTATAATTATGCAC	GAGGAGAAGGCGCGTTAGTTTATGGTGTAATCTCGAAGGTG
3	ehv104	putative helicase	CCAGGGACCAGCAATGTTTGCGAAATGTATTTCAAAACCATTG	GAGGAGAAGGCGCGTTAAAGCTTTCCTATTTTTTTCTTTAATGCAG
4	ehv152	putative DNA-binding protein	CCAGGGACCAGCAATGGAATCACTAAACGTTGTAAAATTTCTG	GAGGAGAAGGCGCGTTACGCTGAAACCTTTTGTCGTTTTG
5	ehv307		CCAGGGACCAGCAATGGCACCCAATACGGAACC	GAGGAGAAGGCGCGTTAGGATGGCGTCTGTGACAG
6	ehv358	putative thioredoxin	CCAGGGACCAGCAATGGAGCTCAGCATGAAGATGG	GAGGAGAAGGCGCGTTACTTCTCAATAAACTCCTTAAAACTATCC
7	ehv434	putative DNA-directed RNA polymerase II subunit	CCAGGGACCAGCAATGTTTAGCAGCATAGAAGATGCCAC	GAGGAGAAGGCGCGTTACTAAAATTCTAATGTGGCGGAAATATGAG
8	ehv465	putative thioredoxin protein	CCAGGGACCAGCAATGATCTTTATAATCGCACCGGTC	GAGGAGAAGGCGCGTTAAATTGACTCATCGACGAATTGTAAC
9	ehv087		CCAGGGACCAGCAATGTCTACATCTCAGGATATCAAGAC	GAGGAGAAGGCGCGTTATCAAGACACTTCCATATCGACTAC
10	ehv150		CCAGGGACCAGCAATGGACGAATTTCATACTACAGTAG	GAGGAGAAGGCGCGTTATTGTTGGGTATTGTGTTCGG
11	ehv298		CCAGGGACCAGCAATGGCCACCGTCGGACAAAAAAG	GAGGAGAAGGCGCGTTATCACGCATCAATCATCTCCTGGATC
12	ehv233		CCAGGGACCAGCAATGTTTAAAATGAAGCACTGCTGCAATC	GAGGAGAAGGCGCGTTACTATACCGTAAAAAGGCTTAGGTATG
13	ehv020	putative proliferating cell nuclear antigen	CCAGGGACCAGCAATGTTTGAGTGCAAAGGAAATGCCGGAC	GAGGAGAAGGCGCGTTACAGATCGTCTATCTTAGGCGCCAG
14	ehv040		CCAGGGACCAGCAATGAAATTCCACTGGCCGACTG	GAGGAGAAGGCGCGTTACTAAGTATCCATATATTGATCAGCATTC
15	ehv105	transcription factor S-II (TFIIS) family protein	CCAGGGACCAGCAATGTCACTATTTACCAAAAATCCCCTAAATAC	GAGGAGAAGGCGCGTTACTGCACTTTCCACTTTTGATTACAC
16	ehv158	putative DNA ligase	CCAGGGACCAGCAATGTTTTCGGGGTATCGTTCTTATCAAG	GAGGAGAAGGCGCGTTAGTTAACCGACGCTTGTGATTGATAC
17	ehv308		CCAGGGACCAGCAATGGTCACCCAAGTCAATCATATG	GAGGAGAAGGCGCGTTATCAATTATTCATCTTATAAAACCAAGTCAAAC
18	ehv363	putative esterase	CCAGGGACCAGCAATGCTTCATAAATTTGAAACCCATCG	GAGGAGAAGGCGCGTTACTAGCACTTGATAATGTGTTCTTTG
19	ehv440	putative proliferating cell nuclear antigen	CCAGGGACCAGCAATGTCTAATCAATCAATTACAGTTACATTTGAAG	GAGGAGAAGGCGCGTTATCAATCAATGTTTCTGGGAGTAATAAAAAGTG
20	ehv117a		CCAGGGACCAGCAATGTTTAGCAGCATAGAAGATGCCAC	GAGGAGAAGGCGCGTTACTAAAATTCTAATGTGGCAGAAATATGAG
21	ehv106		CCAGGGACCAGCAATGTCTGTTGTTCGGACATCATTG	GAGGAGAAGGCGCGTTAAACTATTTCCCCCAACACGTTTG
22	ehv323		CCAGGGACCAGCAATGTCTGCCGCCACGCTCAG	GAGGAGAAGGCGCGTTACCAAACGGGACGACCAACAAAAG

23	ehv300		CCAGGGACCAGCAATGCTTCCGAATCAGATGAACCAG	GAGGAGAAGGCGCGTTACACATCACAGTACAAACCAACGAC
24	ehv235		CCAGGGACCAGCAATGCCTACCCATAATCGTACCGTC	GAGGAGAAGGCGCGTTACTAATCTACAATTACGTCATCATCAGTTG
25	ehv022	phosphoglycerate mutase family	CCAGGGACCAGCAATGATAGTGCGTTTTATTCGACATGG	GAGGAGAAGGCGCGTTATTCGAGAGTTTGATCCAGTGTATAC
26	ehv041	putative endonuclease	CCAGGGACCAGCAATGTCTGGCATCATATATTACATCC	GAGGAGAAGGCGCGTTAAAATATAAAATACGGATTGTATGGCATC
27	ehv109	OTU-like cysteine protease	CCAGGGACCAGCAATGACATCAACCGCATCCCGTATAAG	GAGGAGAAGGCGCGTTATTTTTTTACTAATGAGTCATAATGACCGGAC
28	ehv166	putative RING finger protein	CCAGGGACCAGCAATGTCCGTTGGTTTCTGTGCCGTTTG	GAGGAGAAGGCGCGTTAATCTACATGTACACGAAGTACTTCAGTG
29	ehv325		CCAGGGACCAGCAATGAATAGACTTGGATATATTGCGATTGGTTTTAC	GAGGAGAAGGCGCGTTATCACTGCATACATGGCACGCGTTC
30	ehv393	DnaJ domain-containing protein	CCAGGGACCAGCAATGCAGTACATCACACAAAGTGTAC	GAGGAGAAGGCGCGTTAGATGAAACCAAACCAACTGC
31	ehv444	putative DNA topoisomerase	CCAGGGACCAGCAATGTCTGACCAAAAGATGCAGGTC	GAGGAGAAGGCGCGTTACTTTCTAGCTTTCTTTGGCTTGG
32	ehv434a		CCAGGGACCAGCAATGCAGGCAGATCCAAAAGAAGGTG	GAGGAGAAGGCGCGTTAGCTGTGCATGGAAAGTCGG
33	ehv107	Superfamily I DNA and RNA helicases	CCAGGGACCAGCAATGTCAAAATACAGTTTATCATCTCAACAAC	GAGGAGAAGGCGCGTTATAACTTCAATAACTCTAATGTTTTAACTGG
34	ehv290		CCAGGGACCAGCAATGGCCTCCCCGATTGCGATC	GAGGAGAAGGCGCGTTACTTAAGAGCCATGAGACGAGAC
35	ehv304		CCAGGGACCAGCAATGGTCCCATCCACCGCTGAC	GAGGAGAAGGCGCGTTAATTGTTAATTACGAAAACACCCGACCCCAC
36	ehv265		CCAGGGACCAGCAATGGTCTCTGCGATGCTGACGATG	GAGGAGAAGGCGCGTTACTAATTCACAACAACTGCCCTTGACG
37	ehv023	putative deoxycytidylate deaminase	CCAGGGACCAGCAATGGACTTTACAAAGTGTATGTCTTG	GAGGAGAAGGCGCGTTATCAAAGTAAAAAAGTAAGTAATCTATACGTC
38	ehv042		CCAGGGACCAGCAATGGGTACTGGAACCTCGCGAC	GAGGAGAAGGCGCGTTACGACGTCACGTGGGCTCG
39	ehv128	ERV1/ALR family protein	CCAGGGACCAGCAATGCCCAACAAAACAGACGAGAC	GAGGAGAAGGCGCGTTACTTTTTATACTTTCTATACCCAACCAC
40	ehv175		CCAGGGACCAGCAATGGAGAAAGAGGAAAGTCTCACTG	GAGGAGAAGGCGCGTTAGAATTCTGGCGGCTCGTTTAG
41	ehv340		CCAGGGACCAGCAATGAAGAACTTTTCATCTGACACATCTG	GAGGAGAAGGCGCGTTATTGATACTGATCCAAATAACTAGACG
42	ehv397	putative deoxyuridine 5'-triphosphate nucleotidohydrolase	CCAGGGACCAGCAATGTCCTCAACCACTACCAATCTATATG	GAGGAGAAGGCGCGTTACATACCAGTACTACCAAACCCATTAG
43	ehv447	putative serine protease	CCAGGGACCAGCAATGTGTTATAATATAAACATGTTTCTATTCATAAC	GAGGAGAAGGCGCGTTACTATGTACGCAATGAATACTTCTTG
44	ehv434b		CCAGGGACCAGCAATGCTACAACAAAGTTTCAAGTGC	GAGGAGAAGGCGCGTTACACCATGTTAAACTCTGTGTTTTC
45	ehv141	PIF1 helicase	CCAGGGACCAGCAATGTCGCTAACCGTATGTCAGCAG	GAGGAGAAGGCGCGTTACACACAATAAAAGTTAACGACTTTTGGATC
46	ehv293		CCAGGGACCAGCAATGGCTTTCGTCGACAACACGTTTTC	GAGGAGAAGGCGCGTTACATCGCAATTGGGACACCG
47	ehv305		CCAGGGACCAGCAATGACGACGATGTCTTCGGTTC	GAGGAGAAGGCGCGTTAATACTCCATGTAAATAGCAAATAACC
48	ehv271		CCAGGGACCAGCAATGCAGAACATCTCAAATAATCTCCTTGATAC	GAGGAGAAGGCGCGTTAAAATCCACCCCCACCCCCAC

49	ehv028	putative lipase	CCAGGGACCAGCAATGTTTTTCATTATAACGGCAATAACG	GAGGAGAAGGCGCGTTATAGTAACATGCGTTTTGCATATTCC
50	ehv050	putative serine palmitoyltransferase	CCAGGGACCAGCAATGTACACGGCCGTATTCATATGTTTATG	GAGGAGAAGGCGCGTTACTATGACAGATACTCAATCATAAACTTGAC
51	ehv131		CCAGGGACCAGCAATGGGCAAAAAGACGTCCAAACC	GAGGAGAAGGCGCGTTATCACCATGTATCTCTTCTTACGATTATTG
52	ehv184	putative DNA-binding protein	CCAGGGACCAGCAATGGGGAAGCGTCTACGTC	GAGGAGAAGGCGCGTTACGGACATACTTCACATGTAAAAATTG
53	ehv346		CCAGGGACCAGCAATGTTAGAAAATCTACCATTAACAATTGACC	GAGGAGAAGGCGCGTTACATTAAAAGTGGGTGATTCTGACC
54	ehv401	putative ribonuclease	CCAGGGACCAGCAATGAACCGTGATGAAGAACTTCTAAC	GAGGAGAAGGCGCGTTATCATTGTGTAAACTTTTTAATTGGAGCATATG
55	ehv451	putative protein kinase	CCAGGGACCAGCAATGTCAACTAATTCATATGTATCACAAATAATC	GAGGAGAAGGCGCGTTAACAAGTTCTTAATTTTGGATTAATACACAAC
56	ehv019	Protein kinase domain	CCAGGGACCAGCAATGTTTGAAAATGCAGATATCGTATCG	GAGGAGAAGGCGCGTTAATCTTTTGTTAACCGAGTAAGCAAC
57	ehv142		CCAGGGACCAGCAATGTCTGACACAAATGTTAAGCCATTATC	GAGGAGAAGGCGCGTTACTATTCGGTCTTCATTAGATCTCTGAC
58	ehv294		CCAGGGACCAGCAATGGCTTTCGGCAAGATGGCAATTTC	GAGGAGAAGGCGCGTTACTAAAAAGGAAGGGGAAGTTTCTTG
59	ehv348	alkylated DNA repair protein	CCAGGGACCAGCAATGGATTTCAAAGAAATGTTTGAAAACGAAC	GAGGAGAAGGCGCGTTAAGTGTAATACGAACGTATTGTAAGG
60	ehv274		CCAGGGACCAGCAATGACCAGCTTTTACAATGTGTCATTCTG	GAGGAGAAGGCGCGTTACGCTAGATAACCCCTAGCCAC
61	ehv034		CCAGGGACCAGCAATGAGTATCAAGGATCGCGTATTTG	GAGGAGAAGGCGCGTTATCAAACACAGTCCTCAAAAACCTC
62	ehv072	putative DNA-binding protein	CCAGGGACCAGCAATGGATGCAGTTCCAATTACTCG	GAGGAGAAGGCGCGTTAAAAACTACTTAATACAATTTTTGCATTTGG
63	ehv133	putative ATP-dependent protease proteolytic subunit	CCAGGGACCAGCAATGCTATATAAATCCAATGCGTATGTTTCTG	GAGGAGAAGGCGCGTTACTAAAGGAGTTCAGTGGGGAGTAG
64	ehv189		CCAGGGACCAGCAATGTTTAGTGCAGCTCTCGGAG	GAGGAGAAGGCGCGTTACTACATGTTTGCATACATGCCATC
65	ehv349	putative protease	CCAGGGACCAGCAATGGAGTATACGCATAATAAATGTTTTATACC	GAGGAGAAGGCGCGTTATCAATACTCTACACGAGCGGTG
66	ehv402	putative protein kinase	CCAGGGACCAGCAATGGGACGAACAAAAAAGCGTATTG	GAGGAGAAGGCGCGTTACACAATGAAGAGTGGTTTGTAGTG
67	ehv453	putative mRNA capping enzyme	CCAGGGACCAGCAATGTTGTCAGATTGTAAAACTGTATTATACAAC	GAGGAGAAGGCGCGTTACTATTTCTTCCTAAACTCACGTATTTC
68	ehv024		CCAGGGACCAGCAATGACGGATCAAGTGGCAAAGAAAC	GAGGAGAAGGCGCGTTACTATGGTTCAAACTTGGCCAATTCAC
69	ehv220		CCAGGGACCAGCAATGGGTGATCATAATGAAATAAATGTGC	GAGGAGAAGGCGCGTTACTAATTTGCATCACATGGGCCATAC
70	ehv295		CCAGGGACCAGCAATGAACTCGGCTCAGGTTTCTTTTTATC	GAGGAGAAGGCGCGTTACTAATTAAGACGGGGGTCGATAAAG
71	ehv218		CCAGGGACCAGCAATGGACGCGCATACAATGACATTC	GAGGAGAAGGCGCGTTACTAATGAGAATTCGAATTATAAGCAAACTC
72	ehv279		CCAGGGACCAGCAATGGAAGCCACGAAACGCACCATC	GAGGAGAAGGCGCGTTAATTAACATATTGATCAATAAAGTCATCTACCTG
73	ehv037		CCAGGGACCAGCAATGGCGTGTCTTGCGGAAAAGTC	GAGGAGAAGGCGCGTTATCCAAACCTGTTCAATTCATATTGAG
74	ehv093	HNH endonuclease family protein	CCAGGGACCAGCAATGTACAACTATCGGTCACGCCGTC	GAGGAGAAGGCGCGTTAATAATAATAAAATTCCTTATATTCATCCAC

75	ehv136	putative nucleic acid-binding protein	CCAGGGACCAGCAATGTCTCAGTCGCATCTTCCTG	GAGGAGAAGGCGCGTTATCACAGACTAACCGTTGTTGTATTC
76	ehv230	putative endonuclease	CCAGGGACCAGCAATGACAAGGATAAATGTAGTACCAACTAC	GAGGAGAAGGCGCGTTAATTATTAAGTCTGTCAGTAATTCTAGCTC
77	ehv356	DNA or RNA helicases of superfamily II	CCAGGGACCAGCAATGCCAAATAGAATTTATAGACTGTATTTG	GAGGAGAAGGCGCGTTACTATGAAATGCCAAGATAATGTGCAATG
78	ehv430	putative helicase	CCAGGGACCAGCAATGTTTCTCCAAGAATATCATTCCGAAC	GAGGAGAAGGCGCGTTACTATTCTAGTTTTTGGCGTTTATTAGAAAC
79	ehv455	putative sialidase	CCAGGGACCAGCAATGATGACAGGTGTACTATACTTGTTTTATG	GAGGAGAAGGCGCGTTACTAGTTTAGTTTATATAGTACAATGTCTGTAG
80	ehv058	Superfamily II DNA/RNA helicases, SNF2 family	CCAGGGACCAGCAATGGACCCGGTTACCGATACTG	GAGGAGAAGGCGCGTTACTATACTGCATACTCACTATAATGATTTTTAAC
81	ehv302		CCAGGGACCAGCAATGGCTGCCATTATGAAGAAAACCAC	GAGGAGAAGGCGCGTTAAACACTCATGTGTCCAGTGTAATAATAC
82	ehv296		CCAGGGACCAGCAATGGCCGCCTCGACCCAGATG	GAGGAGAAGGCGCGTTATCAAACGCGGTCATACGGATAATCGAG
83	ehv219		CCAGGGACCAGCAATGGTTAAAAAGATTGCAAAACAAGTGAAG	GAGGAGAAGGCGCGTTATCACATCACATTATAGCTAACTAATATGTTG
84	ehv280		CCAGGGACCAGCAATGCCAGCCATCGCAAAGAACCAAAAG	GAGGAGAAGGCGCGTTAACAGATACCCCATCCCCCCGCAG
85	ehv038		CCAGGGACCAGCAATGGCGTGTGCAGTACAATACAATTTAATTG	GAGGAGAAGGCGCGTTATCAAACAACCGGTTTATTTGCCGCTC
86	ehv101	putative hydrolase	CCAGGGACCAGCAATGGTATATGAAGCGATTGATGATACTC	GAGGAGAAGGCGCGTTAAATTGACATATCAGATTTGGATAATTGTTTC
87	ehv151	putative serine protease	CCAGGGACCAGCAATGTTGTTGTTACTTGTTAAGTGTATGTATG	GAGGAGAAGGCGCGTTATCATGCAGCTAAATTACGATTCC
88	ehv301		CCAGGGACCAGCAATGGAAAACAATCTAAGAGAGTTTATTGTTG	GAGGAGAAGGCGCGTTACCCCATTTCAGGAAAATAATATAGC
89	ehv357		CCAGGGACCAGCAATGGCAAACTTTAATAATGACAATGCATATG	GAGGAGAAGGCGCGTTACTATTTGGCATTAAGTTTTGCTTCG
90	ehv431	putative thymidylate kinase	CCAGGGACCAGCAATGTACAGCACCCCATACAAATCC	GAGGAGAAGGCGCGTTATCACATGTGCTTATTCACAAACTTATCAG
91	ehv461		CCAGGGACCAGCAATGTCAGTTTCAAGTAATATTGCCGGGCGTAG	GAGGAGAAGGCGCGTTAGTTATAACATGTATCTACCTTGGAATTTTTCG
92	ehv084		CCAGGGACCAGCAATGCAGACCAATCACGAAAATGTTATG	GAGGAGAAGGCGCGTTAGTTTGAATATAACTTTAAAAACATATCACGAG
93	ehv322		CCAGGGACCAGCAATGGCCAGCCAATCGGTTCC	GAGGAGAAGGCGCGTTACTACGACATTCCATATCCGACAGTG
94	ehv297		CCAGGGACCAGCAATGACCACCGCATCGAACATGAAC	GAGGAGAAGGCGCGTTACTCCTGGGAGATTGCAATCTCATC
95	ehv161		CCAGGGACCAGCAATGAAATTTACATCTATACTCGCATACAC	GAGGAGAAGGCGCGTTACTATATGATTTTGGAAATGTATCTTGC
96	ehv289		CCAGGGACCAGCAATGCCTGCCATGAAGCGCACG	GAGGAGAAGGCGCGTTACTACAACGCCGCCGAATTTGACG

2.3.2.1. General PCR protocol

All PCR reactions in this study were conducted in a VWR JENCONS Uno Thermal Cycler in 25 µL final volume reactions (unless otherwise stated). PCR reactions were set as follows: 1 µL of virus lysate template (or ~50 ng μL^{-1} of extracted DNA) was mixed with 5 µL of 1 × PCR reaction buffer (Promega), 1.5 µL of 25 mM $MgCl_2$, 0.1 µL of *Taq* DNA polymerase (Promega), 2 µL of 10 µM of each primer (Table 2.4), 1.25 µL of 2 mM dNTPs and DNA-free molecular biology grade water (Sigma-Aldrich). Cycle conditions are described in Chapter 4.2.3 and Chapter 5.2.4.

2.3.2.1.1. PCR of four coccolithoviruses

The PCR reactions for four newly isolated coccolithoviruses were conducted in 96 well plates in final reaction volumes of 50 µL that consisted of 1 µL of template DNA, 0.5 µL forward and reverse primers (20 µM concentrations, Table 2.5), 5 µL $MgCl_2$, 6 µL dNTPs, 10 µL PCR buffer, and the rest 25.6 µL dH_2O. Cycle conditions are described in Chapter 6.2.7.

2.3.3. Agarose gel electrophoresis

The extracted DNA from the environmental samples in Chapter 5 was assessed using a NanoDrop2000 (Thermo Scientific) in order to quantify and qualify the total DNA concentration before the molecular fingerprint analysis study. Amplified DNA gene fragments of the two coccolithoviruses used in the fight club experiment in Chapter 4, from each PCR step in Chapter 5, and amplified DNA of gene fragments of four new coccolithoviruses in Chapter 6 were also assessed by agarose gel electrophoresis. For this ultrapure agarose at a concentration of 1% was dissolved in either 20 or 150 mL of 1 × TAE buffer by heating to boiling point in a microwave and allowed to cool

down for 1-2 min. Ethidium Bromide (EtBr, Sigma) at a final concentration of 0.01 % was added and the solution was then poured into a gel tray to solidify at room temperature for 30 min. A 100 bp ladder (Promega) was used in all the gels as a marker. Agarose gels with a volume of 20 mL were run for 30 min and gels with a volume of 150 mL were run for 1 h, both at 100 V, and then visualised on a UV transilluminator (Syngene GeneGenius) and photographed with a GeneSnap system.

2.3.4. DNA fragment sequencing

All DNA fragment sequencing in Chapter 5 was conducted by the LGC Genomics sequencing centre in the UK (www.lgcgenomics.com) using the LifeTech (former Applied Biosystems) BigDye version 3.1 sequencing mix. After purification of the sequencing reactions from unincorporated terminators by gel filtration (Centri –Pur 96, EMP biotech Berlin) the samples were run on a ABI 3730 XL instrument using POP7 polymer and standard run conditions for up to 1,000 nt read length.

2.3.5. Virus total genomic DNA extraction for full genome sequencing

2.3.5.1 Primary lysate generation for full genome sequencing

Emiliania huxleyi strain CCMP 2090 was grown in 1 L cultures (f/2 nutrient media) in the laboratory under a light/dark cycle of 16/8 h respectively, at a temperature of 16°C. Once the cultures reached mid exponential growth (i.e. ~4×10^6 mL^{-1}), they were infected with lysate stocks of the following virus strains: EhV-84, EhV-88, EhV-201, EhV-202, EhV-203, EhV-207, and EhV-208, at an M.O.I. of 1:1. Infection, host death and viral production were confirmed by AFC and visual loss in coloration and pigmentation of the infected cultures. Fresh virus lysates were produced by filtering the 1 L lysed cultures through 0.2 µm pore 47 mm diameter Durapore filters

(Millipore). The same procedure was applied on the four newly isolated coccolithoviruses described in Chapter 6.

2.3.5.2. Coccolithovirus lysate concentration by PEG precipitation and purification by caesium chloride gradient

Viruses were concentrated by polyethylene glycol (PEG) precipitation, subjected to a caesium chloride (CsCl) gradient and the DNA from this gradient was then extracted as described by Allen et al. (2008). Briefly, to each 1 L filtered lysate were added 58.4 g Sodium Chloride (NaCl) and these were gently dissolved with the aid of a magnetic stirrer. The solutions were then left to stand on ice for at least 1 h before being centrifuged for 15 min at 5,000 × g. The supernatant was retained and PEG 6000 (Fisher-Scientific, Leicester, UK) was added to the supernatant of each solution at a final concentration of 10 %, allowed to dissolve, and then incubated at 4ºC over night. The solutions were then centrifuged at 6,000 g for 25 min at 4ºC, the supernatant discarded and the tubes containing the pellet left to dry upside down for 15 min at room temperature. Subsequently, the virus pellets were resuspended in 12 mL of 1 × PBS buffer.

The virus particles of each virus strain were then purified by a CsCl gradient centrifugation. For this, the resulting 12 mL PEG concentrates were subdivided into 4 × 3 mL sub-concentrate samples. The 3 mL sub-samples were individually placed in 15 mL Falcon tubes to which CsCl was added in the following amounts: 0.03 g, 0.40095 g, 1.2375 g, and 2.04675 g. The tubes were mixed gently until the CsCl was dissolved, and starting with the densest concentration carefully layered into an ultraclear Beckman ultracentrifugation tube; i.e. each 3 mL layer corresponding to a

final volume of 1.0 g mL^{-1}, 1.1 mL^{-1}, 1.3 g mL^{-1}, and 1.4 g mL^{-1}. The tubes with the layered gradients were then centrifuged at 25,000 rpm for 3 h at 15°C and the resulting purified virus (represented by a visible white band in the middle of the centrifuged gradient tubes) was extracted with a syringe and subjected to a genomic DNA extraction.

2.3.5.3. Genomic DNA extraction from purified coccolithovirus concentrates

The purified coccolithovirus concentrates from the CsCl gradient were subjected to a phenol/chloroform genomic DNA extraction procedure. 0.5 mL of concentrated and purified virus lysates were placed into sterile tubes in a heating block at 90°C for 1 min and then transferred onto ice for a further minute, repeating this three times. Then to the tubes containing the concentrates were added 20 µL 0.5 M EDTA pH 8.0 at a final concentration of 20 mM, 5 µL proteinase K at a final concentration of 50 µg mL^{-1}, and 25 µL of 10% SDS at a final concentration of 0.5%, and incubated in a water bath for 1 h at 65°C. After the incubation the tubes were transferred onto ice and gently mixed with 60 µL of phenol. Then 500 µL of chloroform/isoamyl alcohol (24:1) was added to the tubes, mixed gently, after which they were centrifuged at 10,000 rpm for 5 min. After removing the top phase into a clean Eppendorf microfuge tube and adding 500 µL of 7.5 M ammonium acetate, they were left at room temperature for 30 min. A centrifugation step at 10,000 rpm for 15 min followed after which the supernatants were placed into a clean 2 mL Eppendorf tube. To these tubes 1 mL of 100% ethanol (ETOH) was added, leaving them to precipitate for 3 h at 4°C, and then centrifuging them at 13,000 rpm for 30 min after which the supernatants were removed and discarded. Finally, the pellets were washed by centrifugation for 10 min with 500 µL of 70% ETOH and air dried over night. The genomic DNA pellets

were resuspended in 50 µL of TE buffer, quantified by spectrophotometer (i.e. NanoDrop) and a 1% agarose gel electrophoresis prior to being sent for sequencing.

2.3.6. DNA extraction from environmental and experimental samples

The 0.2 µm filters (Millipore membrane filters, cellulose acetate/ cellulose nitrate) collected during the AMT-20 cruise (Chapter 5) and the samples that originated from the sampling of the L4 time series in the English Channel were subjected to a total DNA phenol/chloroform extraction procedure. Each filter (i.e. sample) was halved, so that half could be snap frozen at -80°C for future analysis if necessary. The remaining filters were placed on Petri dishes, cut further into smaller easily dissolvable pieces and then placed into 2 mL Eppendorf tubes. Each tube was vortexed with 300 µL of GTE buffer, 100 µL of lysozyme (10 mg mL^{-1}) and 100 µL of 0.5 M EDTA pH8. After incubating at room temperature on an Agilent bioanalyser chip-shaker at 1000 rpm for 1.5 h, the tubes were incubated for further 10 min with an added 200 µL of 10% SDS (w/v). They were then vortexed with 500 µL of phenol and centrifuged for 10 min at 13,000 rpm in order to separate the top aqueous layer from the bottom phenol layer. The top layers (~500 µL) were transferred to clean 2 mL Eppendorf tubes containing 500 µL of 24:1 chloroform/isoamyl alcohol. These were inverted gently and centrifuged at 13,000 rpm for 10 min. After transferring the top layer again into clean 2 mL Eppendorf tubes they were incubated for 30 min at room temperature with added 350 µL of 7.5 M NH_4-acetate, and then centrifuged at 13,000 rpm for 15 min. The supernatant was then removed into new 2 mL Eppendorf tubes and the DNA precipitated as described previously in this chapter (Section 2.3.5.3). The phenol/chloroform part of the described procedure was also applied on the samples generated during the virus fight club experiment in Chapter 4.

2.4. BIOINFORMATICS

2.4.1. Sequence alignments of environmental sequences

The alignment and phylogenetic reconstruction of sequences from NCBI and the newly received sequences from the sequencing centre were aligned using the MEGA4 (Version 4.0.2) multiple sequence alignment software. Primers sequences were deleted and the fragments made the same length in order to decrease bias that can arise from potential gaps and sequences with different lengths.

2.4.2. Whole genome alignment

Whole genome alignment of the genomes sequences of EhV-84, EhV-86, EhV-88, EhV-201, EhV-202, EhV-203, EhV-207, EhV-208 and EhV-99B1 was performed by uploading the FASTA file of each into the multiple genome alignment software-MAUVE (version 2.3.1). Before the upload of the files each contig in each FASTA file was blasted against the complete genome of EhV-86 and its position on the EhV-86 backbone visualised by the Artemis Comparison Tool software. Then the contigs were reordered in the correct order manually and reassembled into a new correctly orientated FASTA sequence file.

2.4.3. Genome comparison and general statistics

The GenBank files of each fully sequenced viral genome together with the GenBank files of the sequenced genomes of EhV-86 and EhV-99B1 were uploaded into the IMG/ER analysis platform for additional annotation, gene prediction, comparison and general statistic information. The platform was used in order to detect single gene homology among the genomes, predicted COGs (Clusters of Orthologous Gropus of proteins), candidate enzyme predictions and internal BLASTn and BLASTp analysis.

3. COCCOLITHOVIRUSES: GENOMIC THIEVES AND METABOLIC THUGS- A GENOME COMPARISON

3.1. INTRODUCTION

Viruses catalyse the biologically driven processes of nutrient cycling between trophic levels (Suttle, 2005) and their relentless activity is essential for life on Earth. All living biological entities are susceptible to virus infection, including the oceanic microalgae which play a fundamental and crucial role in global primary productivity. It is perhaps ironic that the cellular destruction mediated by virus infection can often be made more efficient through the acquisition, modification and utilisation of genetic machinery which has often been stolen from a relative of the cell currently under attack (Moreau et al., 2010; Monier et al., 2011). Indeed, by arming themselves with the greatest pool of genetic diversity on the planet, virus mediated horizontal gene transfer (HGT) has had and continues to have a profound impact on the evolution of life.

To date, despite the advent of culture independent high throughput sequencing approaches and techniques, the invisible abundance and the genetic diversity of the estimated 10^{30} viruses present in our oceans has remained fundamentally unexplored. This is largely due to: A) the difficulty of applying total DNA sequencing approaches such as metagenomics on entire virus communities where there is not a single gene common to all viruses (as opposed to bacteria and archaea) and B) due to a holistic approach that has been undertaken by some researchers that over simplifies environmental microbial genetics, their diversity and possibly their phenotypic characteristics. Many studies today rely heavily on computer models in which

researchers over simplify biological systems by the input of several confounding parameters in to a model that aims at mimicking the natural environment or a real biological system and its interaction with other biotic and abiotic factors alike. In the case of marine viruses, unfortunately, many studies lack a complete understanding of virus-host dynamics, and fail to grasp that specific groups of viruses, can act and interact with their hosts differently depending on their genomic potential and the conditions to which they and their hosts are exposed to, something that can be revealed only by single genome analysis and comparison.

Hence, in order to gain a true understanding of marine viruses, and more importantly their role in the oceans, researchers need to go back to the basics and undertake a reductionist approach, where a single component (or in this case a single group of viruses) within a marine system is used as an experimental model. In simple terms one cannot understand how a watch works without going into the detailed mechanics of its structure, and furthermore one will not be able to explain why some watches are better than others just by looking at them and stating that they are different. In this chapter an attempt is made to undertake a reductionist approach towards the study of marine viruses, by near completely sequencing the genomes of seven previously unstudied coccolithoviruses, their comparison to previously sequenced strains (i.e. EhV-86 and EhV-99B1), and the analysis of possible HGT events with their algal host. Due to the lack of available genomic data on the host strain *Emiliania huxleyi* CCMP 2090, for the comparative part of this study, a different strain was used for which a preliminary genome sequence was available; i.e. *Emiliania huxleyi* CCMP 1516. These two strains are believed to be the same, with 2090 being "naked" (i.e. no coccoliths) and 1516 being calcified (personal communications with Dr. Mike Allen).

3.1.1. Genome comparison: the power of bioinformatics

It is now possible with the aid of next generation genome sequencing technologies, user friendly advanced computational tools, and relatively simple algorithms, to study microorganisms on a genetic level and compare the predicted function of specific genes and their translated protein sequences to actual functions observed both in the laboratory and in the field. The development of these tools for the broader community has allowed access for researchers that are not experts in a programming language. Astonishingly, what 30 years ago was only realistically achievable during an entire PhD project (for instance the sequencing of several hundred nucleotide base pairs) can now be done in several weeks or even days (the output of which is several hundred Gigabytes of data). Often (depending on the available funds), the genetic sequence of an organism (or in this study a virus) comes to the researcher in a form that is easy to work with. The assembly of the genome and the preliminary annotation is performed by the sequencing centre and all that is left for the researcher working with the genome is to determine the quality of the sequence, its correct orientation and annotation, detect potential errors in the sequence or/and the predicted functions of the genes. This can now be easily achieved by uploading the information provided by the sequencing centre to one of the online genome analysis platforms (i.e. MGRAST for metagenomes and IMG/ER for bacterial and viral genomes) and investigate manually for errors by using one of the following freely available softwares: Artemis, Artemis Comparison Tool (ACT), and MAUVE. These platforms can provide additional annotation of predicted genes, allowing the researcher to detect potential errors, amend these if necessary, and find predicted functions and potential homologs in other sequenced genomes in the public domain.

Such a comparative approach is undertaken nowdays on a routine basis when virologists seek to uncover the mysteries of virus genomes and unravel the differences between virus strains from distinct ecological niches. While the comparison of the sequenced genomes of EhV-86 and its counterparts EhV-163 (Allen et al., 2006b) and EhV-99B1 (Pagarete et al., 2013) revealed the degree of genomic similarities that exist between these three coccolithovirus strains, genomic differences were also unveiled; i.e. the gene coding for phosphate permease and the type and number of tRNA genes. Elsewhere, a comparative analysis of *Prochlorococcus* bacteriophages (podoviruses and myoviruses) revealed that they share a number of core genes with T4 and T7 bacteriophages (Sullivan et al., 2005), a possible indication of their evolutionary co-ancestry. Moreover, cyanobacterial genes, such as the photosynthetic genes psbA and hliP (which help maintaining the photosynthesis during infection) and the talC gene (which facilitate during infection alternative routes for carbon metabolism), were also found in the genome of these bacteriophages, further indicating the importance of comparative genomics and the discovery of new HGT pathways with ecological significance.

The sequencing/computational approach of several model systems as described here is essential in modern microbiology, molecular biology and biological oceanography. It helps in our understanding of the evolution of viruses, their changes and adaptations in the course of history, the possible implications of environmental conditions on these viruses and their local specific adaptations (through the analysis of lost or gained genes), their co-evolutionary arms race with their hosts, and the significance of their genomic potential on their infection dynamics with their hosts and as a consequence the effect on biogeochemical and carbon cycling in the marine environment. It is only

through a comparative approach that one can determine functional differences between two or more laboratory model virus strains and enable the design of an accurate experimental approach with for instance new molecular markers, increasing the coverage and the success of a study. Finally, such an approach enables to test hypothesis derived from the genomic potential of viruses in the laboratory and extrapolate this into field studies and experiments.

3.1.2. The Marine Microbial Initiative

The Gordon & Betty Moore Foundation's Marine Microbiology Initiative (MMI) aims to generate new knowledge about the composition, function, and ecological role of the microbial communities that serve as the basis of the ocean's food webs and that facilitate the flow of nitrogen, carbon, and energy in the ocean (http://www.moore.org/marine-micro.aspx). In an effort to understand the ecology and evolution of marine viruses and to explore the diversity and ecological roles of bacteriophage and algal virus communities in particular, the Broad Institute (Harvard and MIT research institute in the USA) collaborated with MMI and researchers whose sequencing nominations were chosen by the Marine Phage, Virus, and Virome Selection Committee. During the propagation of the research conducted during this PhD the opportunity to be part of this initiative presented itself and it was decided to attempt to sequence EhVs from the PML virus collection. The basis for the selection for sequencing of the new EhV strains was the current consensus of their global importance in the demise of *E. huxleyi* blooms (Bratbak et al., 1995; Wilson et al., 2002), the horizontal gene transfer events observed in previously sequenced EhVs (Monier et al., 2009; Pagarete et al., 2009), the metabolic potential displayed by their large genome size (Wilson et al., 2005, 2009) and their manipulation of signaling

pathways such as programmed cell death in their host organism (Bidle et al., 2007, 2011; Vardi et al., 2009, 2012).

Here, a complete genome classification and a set of features for the following strains: EhV-84, EhV-88, EhV-201, EhV-202, EhV-203, EhV-207 , EhV-208 and EhV-99B1 are presented. The reconstruction of these virus genomes based on the EhV-86 reference genome is described and the similarities and the differences of these genomes to themselves and others in the database are illustrated. The aim of this chapter is to provide a new genomic insight into these viruses, and unveil the scope of HGT between them and their hosts, with the hope to increase the current knowledge of their evolution and functional potential in the world's oceans.

3.2. MATERIALS AND METHODS

3.2.1. Virus maintenance

A continuous batch of fresh viral lysate was prepared every three to six months for each EhV in the PML culture library in order to ensure that the viruses were kept fresh at all times with a high amount of infective virions within them (Details in Section 2.1.5).

3.2.2. Coccolithovirus strains for sequencing

The genomes of the coccolithovirus strains: EhV-84 (JF974290), EhV-88 (JF974310), EhV-201 (JF974311), EhV-202 (HQ634145), EhV-203 (JF974291), EhV-207 (JF974317) and EhV-208 (JF974318) are deposited in Genbank. Two other previously studied strains, EhV-86 (AJ890364) and EhV-99B1 (FN429076) were included here for the comparative analysis. The newly sequenced strains were kept in culture at the PML virus collection from the time of isolation.

3.2.3. Virus concentration, DNA isolation and extraction for sequencing

Prior to their sequencing the viruses were concentrated by PEG precipitation, subjected to a CsCl gradient and their DNA extracted, as previously described in Section 2.3.5.

3.2.4. Genome sequencing, assembly and annotation

The genomes of strains: EhV-84, EhV-88, EhV-201, EhV-202, EhV-203, EhV-207 and EhV-208 were sequenced using the 454 FLX pyrosequencing technology platform (Roche/454, Branford, CT, USA) by the Broad Institute in the US. Library

construction and sequencing were performed as previously described (Henn et al., 2010). General protocols for library construction can be found at: www.broadinstitute.org/annotation/viral/Phage/Protocols.html.
Sequencing statistics can be seen in Table 3.2.

De novo genome assembly of resulting reads was performed using the Newbler v2.3 assembly software package and genes were identified using the Broad Institute's Automated Phage Annotation Protocol as previously described (Henn et al., 2010). In short, evidence based and *ab initio* gene prediction algorithms where used to identify putative genes followed by construction of a consensus gene model using rules-based evidence approach. Gene models where manually checked for errors such as in-frame stops, very short proteins, splits, and merges. Additional gene prediction analysis and functional annotation was performed manually on the Artemis software and within the Integrated Microbial Genomes – Expert Review platform (Markowitz et al., 2009).

3.2.5. Whole genome reconstruction and alignment

The annotated EhV genomes were used for whole genome alignments. The Artemis Comparison Tool (ACT) was used for BLASTn pairwise alignment and visualisation of the full genome sequences (Carver et al., 2005). MAUVE 2.0, a program designed for the identification of conserved genomic DNA regions, their presence and/or rearrangement was used in order to observe the variable regions in each genome (Darling et al., 2010). Once contigs were identified they were manually converted and placed at their appropriate location on the genome based on the backbone of the sequenced genome of EhV-86.

3.2.6. Transmission Electron Microscopy (TEM)

Viral particles of each EhV were purified from the lysate of *E. huxleyi* strain CCMP 2090. Each EhV strain was concentrated using a CsCl gradient (Chapter 2, Allen et al., 2008). The samples were then fixed for 15 minutes using 50:50 (v/v) Trumps fixative pH 7.2, for 4 hours at room temperature and stored at 4°C and washed with sterile distilled water. Samples were then negatively stained and prepared for imaging with transmission electron microscopy following the protocol by Ackermann and Heldal (2010) using Cu-grids (200 mesh) supported with carbon coated formvar film. Observations of viral particles were performed on the JSM2011 Fas TEM (JOEL UK) electron microscope with an UltraScan 1000 camera (GATAN UK) for imaging. The TEM imaging part of this study was done in collaboration with Dr. Charlotte Worthy at the Rothamstead Research institute in the UK. The author prepared all EhV samples ready for imaging at Rothamsted Research by Dr. Worthy.

3.3. RESULTS

3.3.1. Morphology and virion structure

Transmission Electron Microscopy (TEM) was used to assess the particle sizes and visible morphological features of nine EhVs (Figure 3.1). The coccolithoviruses imaged here show that the coccolithovirus particle sizes are in the range of 180-205 nm in diameter. All the EhVs contained an electron dense region at their core (DNA) surrounded by an icosahderal capsid and external lipid envelope, in addition to the presence of an internal lipid bi-layer membrane within their glycoprotein shell (Figure 3.1).

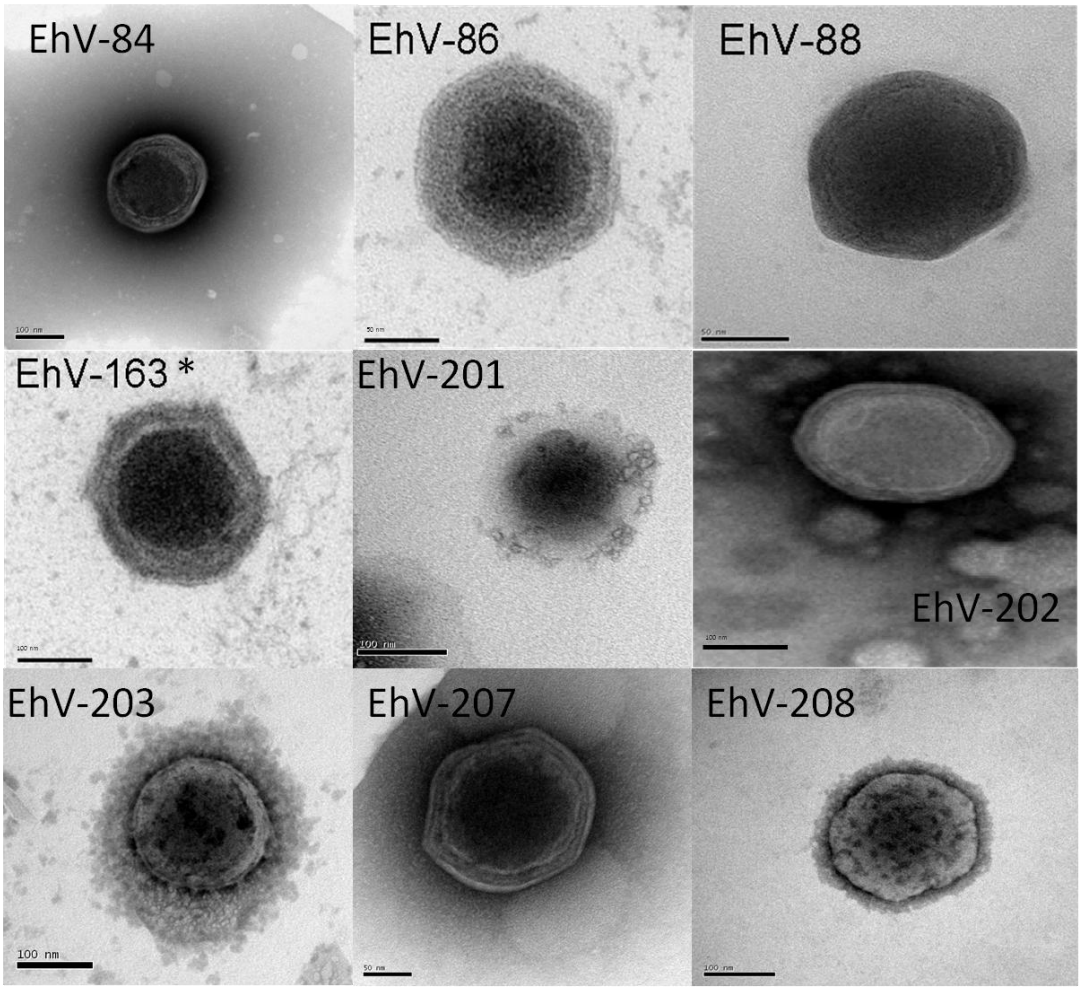


Fig. 3.1. Transmission electron micrograph of the nine coccolithophore virions compared in this study with a scale of 50-100 nm. * There are currently no images available of EhV-99B1 but due to its close resemblance to EhV-163 (Dr. Mike Allen and Dr. Charlotte Worthy, personal communications) an image of the latter is shown.

The general morphological characteristics of EhV-86 have been previously reported (Castberg et al.. 2002; Dunigan et al., 2006; Wilson et al., 2009) and remarked to be similar to the morphology of the PBCV-1 within the *Phycodnaviridae* family (a dsDNA virus infecting *Chlorella sp*.), hence in this respect all EhVs in this study are also structurally similar to PBCV-1. The most evident morphological difference between EhVs and others in the *Phycodnaviridae* family such as PBCV-1 is that the coccolithoviruses are enveloped, whereas PBCV-1 is not.

3.3.2. Genomic analysis

3.3.2.1. Sequencing statistics

The finishing quality of the sequenced genomes by the Broad Institute in the US were all higher than 99% and the number of contigs to be analysed and reordered based on the EhV-86 backbone genome ranged from 7-17. These contigs were at different lengths, a summary can be seen in Table 3.1.

Table 3.1. Sequencing statistics of seven newly sequenced coccolithoviruses, as received from the sequencing centre in the US.

	EhV-84	EhV-88	EhV-201	EhV-202	EhV-203	EhV-207	EhV-208
Finishing quality	>99%	>99%	>99%	>99%	>99%	>99%	>99%
Number of contigs	9	8	7	12	7	16	17
Average contig size (bp)	43980	49575	58100	33868	57131	26274	24083
Largest contig size (bp)	97445	128083	125476	137441	142770	139159	85876
Assembly size (using large contigs)	395820	396598	406701	406416	399920	420391	409403
Assembly coverage ("peak depth")	36.16	67.13	79.38	35.67	41.89	33.72	84.39
Total number of reads used	28526	74782	78268	31250	71325	33894	85422
Sequencing platform	454	454	454	454	454	454	454

3.3.2.2. DNA pol, MCP and COGs phylogeny

Maximum likelihood phylogenetic analysis of available DNA polymerase gene sequences (DNA pol), one of the viral kingdom's phylogenetic markers (equivalent to 16S rRNA sequences in bacteria) indicated that the English Channel strains clustered separately from the Norwegian strains, with the exception of EhV-202 that seemed to be the most different (Figure 3.2). In addition, amongst the English Channel strains there seemed to be sub-clustering between strains isolated in 1999 and those isolated in 2001. Other algal viruses such as *Paramecium bursaria chlorella* virus (PBCV-1), *Micromonas pusilla* virus SP1 (MpV-SP1), *Chrysochromulina brevifilum* virus PW1 (CbV-PW1), *Ectocarpus siliculosus* virus 1 (EsV-1), *Heterosigma akashiwo* virus 01 (HaV-01) are included here as an additional reference and they cluster outside the EhV Genus.

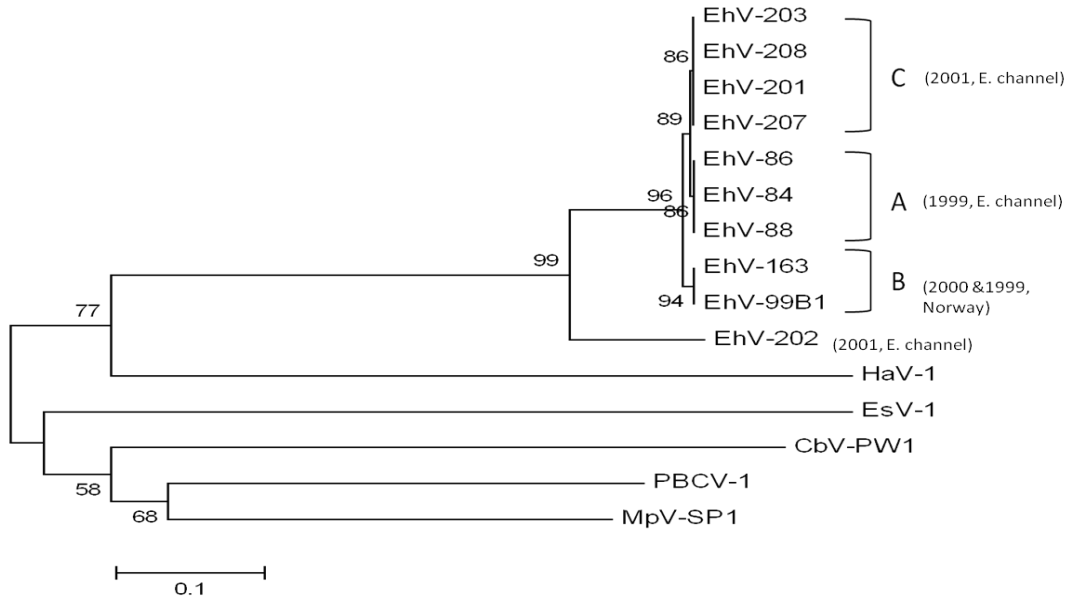


Fig. 3.2. Multiple Sequence Alignment of the DNA pol (DNA polymerase) gene of ten coccolithoviruses (EhVs) and five other algal viruses. A- EhVs isolated from the English Channel in 1999, B- EhVs isolated from a Norwegian Fjord in 1999 and 2000, and C- EhVs isolated from the English Channel in 2001. The evolutionary history was inferred using the Neighbour-Joining method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analysed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches when greater than 50%. The evolutionary distances were computed using the Maximum Composite Likelihood method (and are in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option) (Felsenstein, 1985; Saitou and Nei, 1987; Tamura et al., 2004). Phylogenetic analyses were conducted in MEGA4 (Tamura et al., 2007).

A 209 bp fragment of the MCP gene of EhV-99B1, EhV-84, EhV-86, EhV-88, EhV-201, EhV-202, EhV-203 and EhV-207 were aligned with 34 recently discovered "ancient" MCP sequences (Table App1) that were recovered from sediments in the Black Sea by Coolen et al. (2011). The DNA sequence alignment showed great diversity, characterized by three sub-clustered groups of EhV MCPs (Figure 3.3). Two "ancient" sequences clustered together with EhVs isolated in 2001 (Cluster C) and 14 "ancient" sequences clustered with EhVs isolated in 1999 (Clusters A+B). Surprisingly there were also 18 "ancient" MCP sequences that sub-clustered separately into a "new" cluster of EhVs, designated as cluster D. Unfortunately the full genome sequencing of EhV-208 resulted in an incomplete MCP gene sequence, due to the MCP gene sequence being at the end of a contig. The result of this was a partial sequence of this gene and it was decided not to include this in the analysis and construction of the phylogenetic tree seen below.

With regards to these two marker genes (i.e. MCP and DNA pol), the two were different in their degree of conservation among strains. With the exception of EhV-202 (due to its high degree of dissimilarity), on average DNA pol (ehv030) was more conserved on the nucleotide level than the amino acid level, while MCP (ehv085) was less conserved on the nucleotide level and more on the amino acid level. The average % of identity for DNA pol among strains was 99.14 % (±0.69 SD) on the nucleotide level and 98 % (±0 SD) on the amino acid level, whereas the average % of identity for MCP was 96.28 % (±0.75 SD) on the nucleotide level and 100 % (±0 SD) on the amino acid level. Finally, a genome cluster profile based on the predicted Clusters of Orthologous Groups of proteins (COGs) within the IMG/ER platform (each COG is a group of several proteins that are inferred to be orthologs or/and are direct

evolutionary counterparts), confirmed the sub-clustering of the newly sequenced virus strains to two distinct groups, with EhV-202 being the most distinct of them all (Figure 3.4).

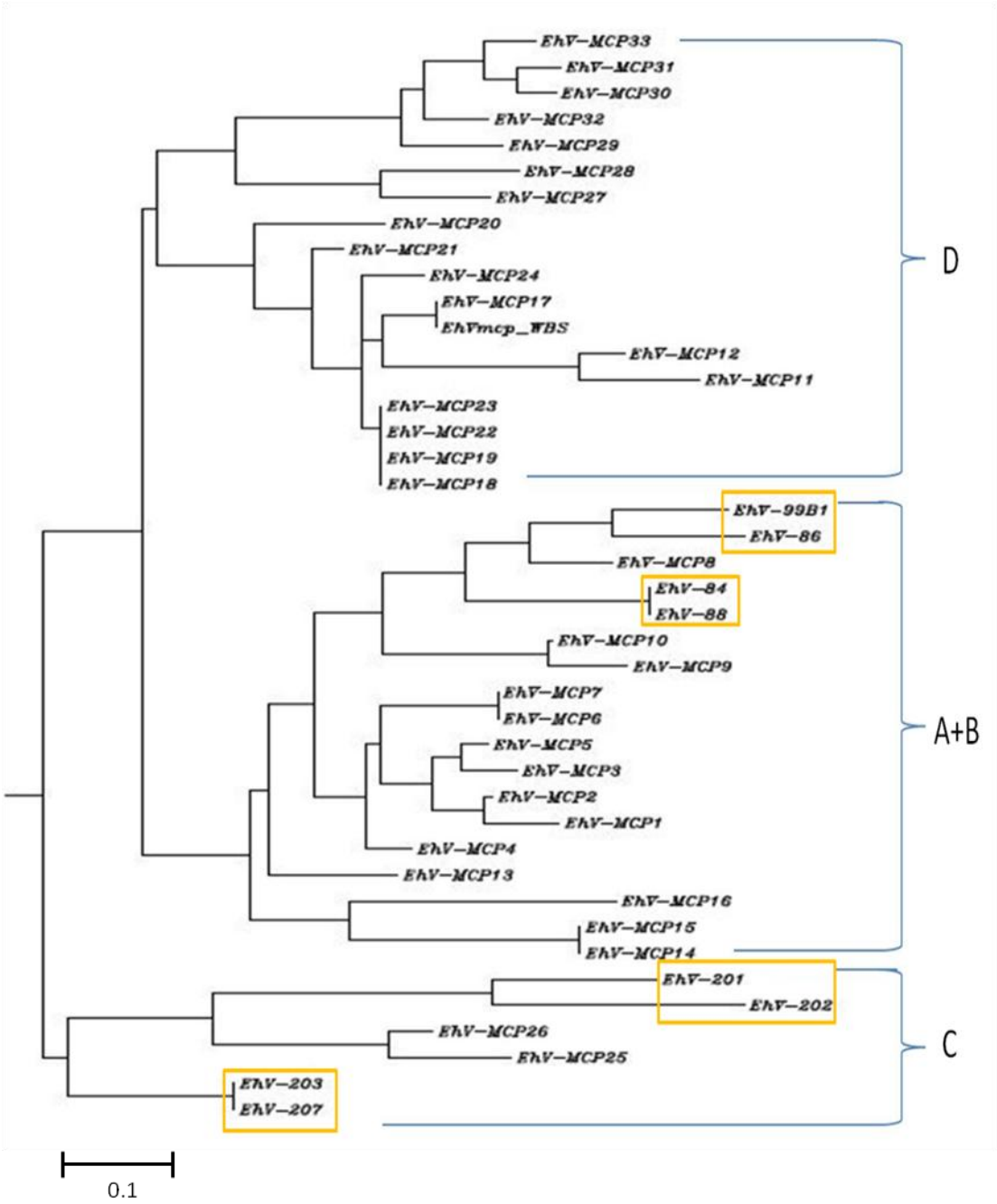


Fig. 3.3. Neighbour joining tree of 42 different coccolithovirus MCP sequences (209 bp in length) from three different geographical locations: MCP 1-33 and MCP_WBC were recovered from sediments in the Black Sea (data obtained from NCBI), EhV-99B1 was from a Norwegian fjord, and EhV-84 – EhV-207 were from the English Channel. The A-D grouping is based on the similarity of their MCP gene fragment. In yellow are boxed the EhV strains that are in culture, were studied in this thesis and have their complete or near-complete genome sequenced and annotated.

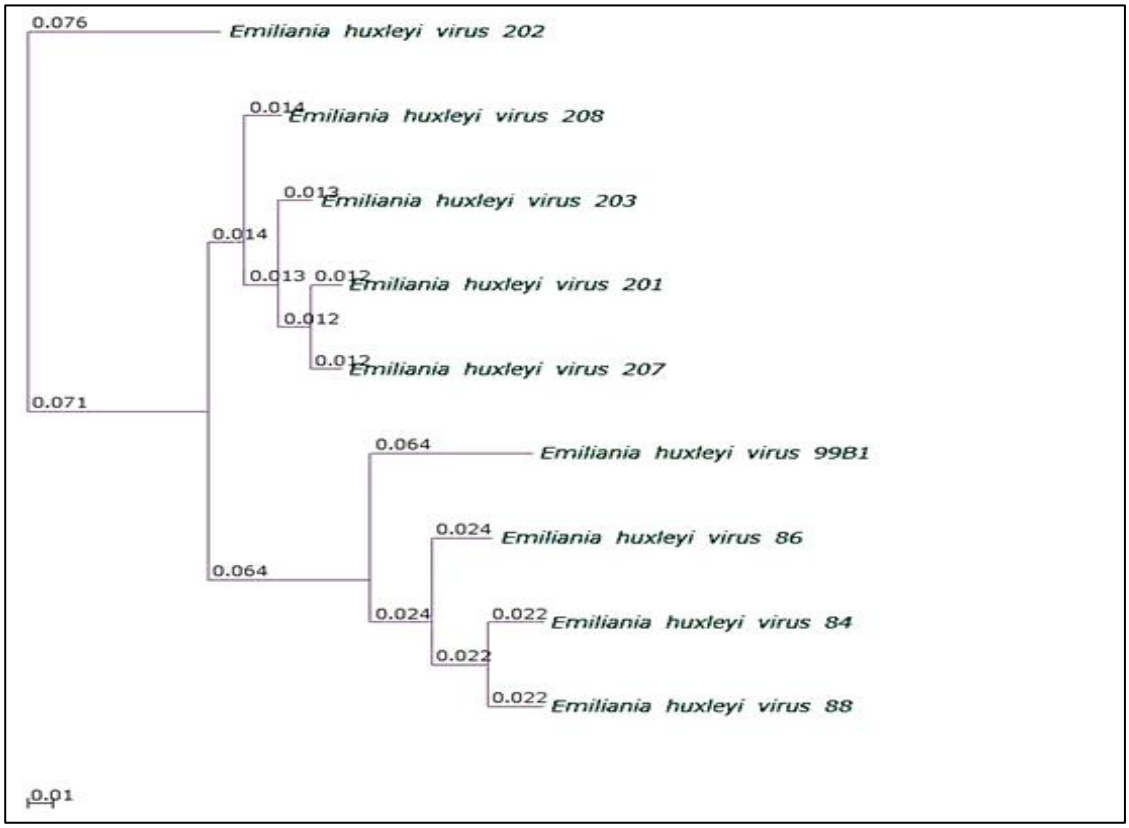


Fig. 3.4. Genome cluster profile based on the predicted Clusters of Orthologous Groups of proteins (COGs) of nine coccolithoviruses with the build in IMG/ER platform using the PhyloXML algorithm.

3.3.2.3. Gene counts and genome size

Before further analysis was performed contigs of all strains were re-arranged based on the EhV-86 assembled backbone and the genome reconstruction was performed manually with the assistance of the Artemis Comparison Tool (ACT) software and MAUVE (Carver et al., 2005; Darling et al., 2010). Then, the newly sequenced genomes and the two reference genomes EhV-86 and EhV-99B1 were uploaded into the IMG/ER online platform (Markowitz et al., 2009) for additional annotation, analysis and comparison. In this comparison EhV-163 was excluded due to an incomplete sequence of its genome (only 80% sequenced) and an already existing report (Allen et al., 2006b). The near complete genome sequencing and the additional IMG/ER platform analysis indicated that the genome length of the EhVs under study ranged from 396,620-421,891 bp and that they had a GC content of 40.12- 40.49 %.

The genomes contained 451- 488 coding sequences (CDSs) with an average of ~90 protein coding genes with functional prediction, including 3-6 tRNA genes. A set of general features for the sequenced EhVs is presented in Table 3.2.

Table 3.2. General characteristics of the completely or near completely sequenced coccolithoviruses to date.

Virus	Genes	Total bases	CDS	Coding Bases	Genes with function prediction	tRNA	GC (%)
EhV-86	478	407339	472	369157	90	5	40.18
EhV-99B1	451	376759	444	333400	90	6	40.04
EhV-84	486	396620	482	334463	85	4	40.17
EhV-88	480	397298	475	357803	90	5	40.18
EhV-201	457	407301	451	363714	89	6	40.46
EhV-202	488	407516	485	352215	93	3	40.30
EhV-203	470	400520	464	364178	91	6	40.12
EhV-207	479	421891	473	371313	93	6	40.49
EhV-208	461	411003	455	348386	90	6	40.42

All genomes appeared to be highly conserved in all reading frames (Figure 3.4 & Figure 3.5) with a relatively high degree of homology. However from the genome reconstruction it was also clear that all the viruses varied in a ~80,000 bp long section located in the middle of each genome. This section was characterised by numerous rearrangements, inversions, insertions, deletions and truncations (Figure 3.5 & Figure 3.6).

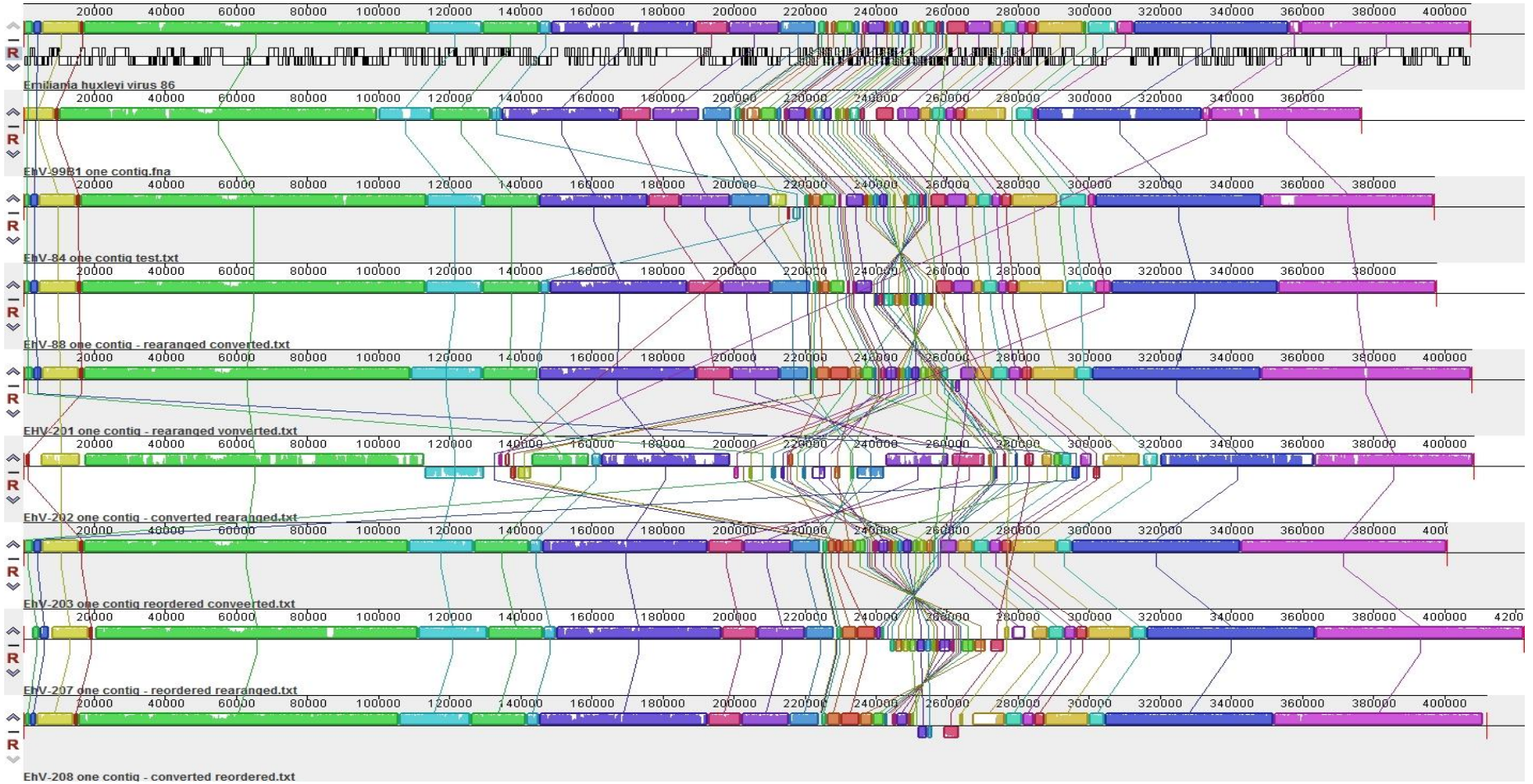


Fig. 3.5. Multiple whole genome alignment of nine EhVs (EhV-84 to EhV-208). Collinear blocks and their order are given for each genome as a series of coloured blocks. The lines between the LCBs represent each set of homologous sequences and their position, relative to the EhV-86 reference genome (the very top sequence).

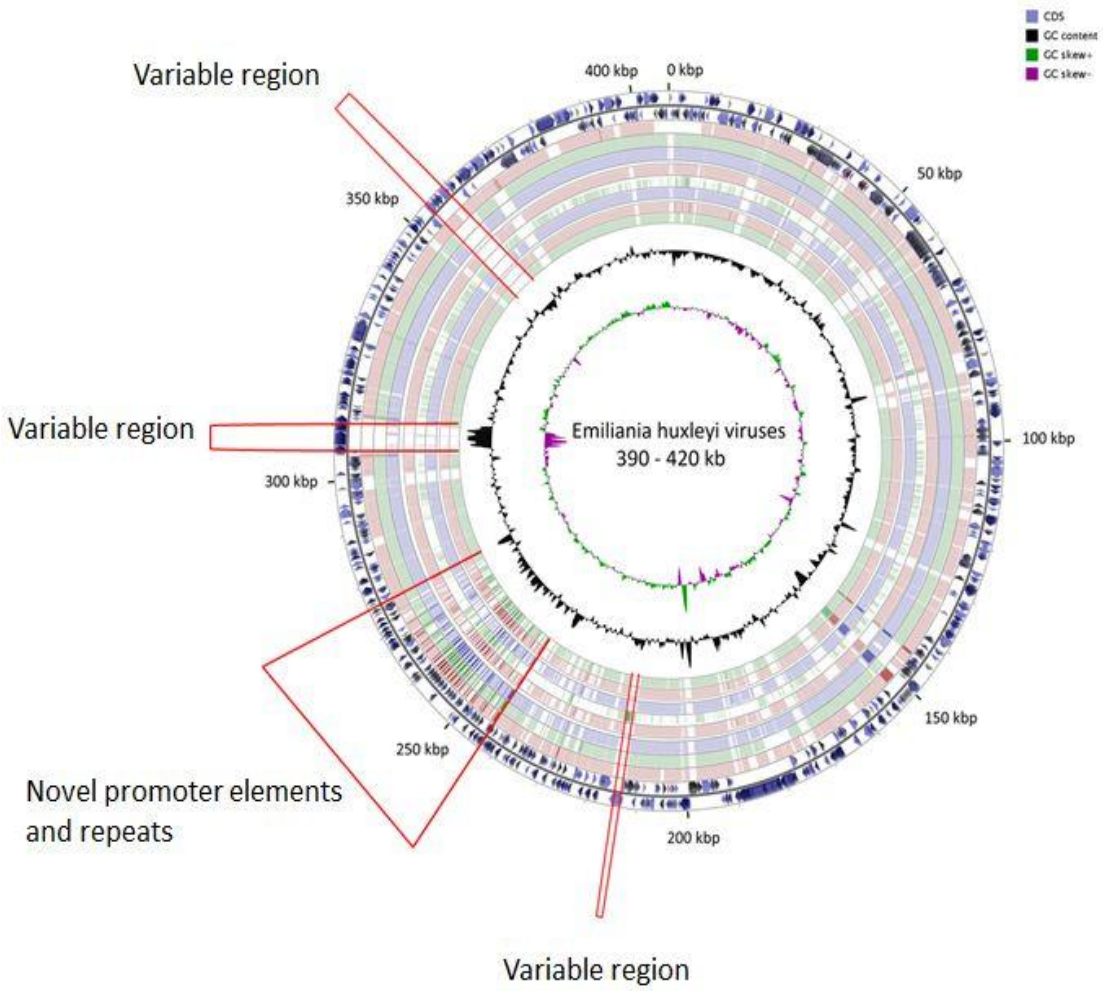


Fig. 3.6. Reconstruction of EhVs based on the complete backbone of EhV-86. From outside in: circle 1 and 2 are the forward and reverse strands of EhV-86 with marked CDSs, circles 3-10 are the genomes of EhV-99B1, EhV-84, EhV-88, EhV-201, EhV-202, EhV-207, and EhV-208 respectively, circle 11 is the GC content and circle 12 is the GC skew. In each circle the white gaps represents areas of the genome which are missing or were not assembled in the final versions of the genomes in comparison to the backbone of EhV-86.

3.3.3. Differences and similarities

3.3.3.1. Unique genes and homology to EhV-86

The majority of protein coding sequences were the same across the EhV gene pool, with CDSs in some strains showing more than one hit to counterparts in the genome of EhV-86 (Table App2). However there were some sequences or/and potential genes that were unique to each genome with no homology to any of the other genomes investigated (Table 3.3). In order to check for unique CDSs/predicted protein coding genes the online IMG/ER platform with the BLASTp alignment algorithm was used. In the setup of the alignment algorithm the cut-off point of the minimum % of identity and the E value were changed to 10% and e^{-5} respectively.

Table 3.3. Unique protein coding sequences of nine EhVs, based on the build-in BLASTP alignment algorithm of the phylogenetic profiler in the IMG/ER analysis platform.

CDS	**EhV-99B1**	**EhV-84**	**EhV-86**	**EhV-88**	**EhV-201**	**EhV-202**	**EhV-203**	**EhV-207**	**EhV-208**
hypothetical proteins	5	27	3	2	1	54	3	7	6
putative endonuclease*	1								
putative membrane proteins	3								
putative transposase	1								
pseudo genes				2	2		2	2	2
ADP ribose pyrophosphatase						1			
LPS biosynthesis protein						1			
Polyubiquitin						1			
predicted protein- tyrosine phosphatise						1			
putative membrane proteins									
zinc finger protein							1		
Total	**10**	**27**	**3**	**4**	**3**	**58**	**6**	**9**	**8**

*similar to ehv117a in EhV-163.

EhV-99B1 encoded five hypothetical proteins, one putative endonuclease (435 bp, 144 aa) with 99.5 % identity to ehv117a in EhV-163, three putative membrane

proteins and one putative transposase (1923 bp, 640 aa) unique to its genome. In contrast, EhV-86 had only three unique CDSs, all of which were hypothetical proteins (ehv236, ehv290, and ehv343), whereas EhV-84 had 27 unique hypothetical proteins. The EhV-88 genome had four unique genes, two of which encode hypothetical proteins, and two were pseudo genes. EhV-201 had only one unique hypothetical protein and two pseudo genes on its genome. EhV-203 encoded three unique hypothetical proteins, two pseudo genes and one zinc finger protein (600 bp, 199 aa). EhV-207 had seven unique hypothetical proteins and two pseudo genes; and EhV-208 had six unique hypothetical proteins and two pseudo genes encoded on their genomes. By far the most different genome with the highest number of unique CDSs was the genome of EhV-202. It encoded 54 unique hypothetical proteins, one ADP ribose pyrophosphate (432 bp, 143 aa), one lipopolysaccharide **(**LPS) biosynthesis protein (804 bp, 267 aa), one polyubiquitin (243 bp, 80 aa), and one predicted protein encoding for tyrosine phosphatase (501 bp, 166 aa).

3.3.3.2. Virus encoded tRNAs

A set of predicted tRNAs is shown in Table 3.4. EhV-84 was predicted to encode four tRNA genes (Arg, Asn, Gln and Ile), whereas EhV-86 had five tRNAs (Arg, Asn, Gln, Ile and Leu). The Ile tRNA of EhV-84 varied dramatically containing a 26 base intron insertion (99 bases in EhV-84; 73 bases in EhV-86) (Figure App1-E). EhV-88 was predicted to encode five tRNA genes (Arg, Asn, Gln, Ile and Lys) and had the same intron in its Ile tRNA as observed in EhV-84. EhV-99B1 was predicted to encode six tRNA genes (Arg, Asn, Gln, Ile, Leu and Lys). EhV-201, EhV-203, EhV-207 and EhV-208 were predicted to encode six tRNA genes (Arg, Asn Gln, Glu, Leu, and Lys). Their tRNA Leu was 18 bp shorter than in EhV-86 (85 and 103 bp respectively) and there was evidence of base deletion (Figure App1-F). They were similar in only

14% of their sequence. EhV-202 was predicted to encode three tRNA genes (Arg, Asn and Gln). The tRNA Asn in EhV-202 was 15 bases longer than in EhV-86 (i.e. 89 and 74 bp respectively), due to an intron in the Asn tRNA of EhV-202 (Figure App1-B). The tRNA Gln in EhV-202 resembled the Gln in EhV-86. It was 71 bp long as opposed to 72 in EhV-86 and they shared a bp identity of only 78% (Figure App1-C). Despite the variation in tRNA content and morphotypes there was no correlation between the number of amino acid codon types in a particular genome and the presence, absence or type of tRNAs found in each genome.

Table 3.4. Genes predicted to encode for tRNAs and their presence (+) or absence in each sequenced genome. "Essential" and "non essential" tRNAs as proposed in this study are highlighted in yellow and green respectively.

tRNA	EhV-86	EhV-99B1	EhV-84	EhV-88	EhV-201	EhV-202	EhV-203	EhV-207	EhV-208
Arg	+	+	+	+	+	+	+	+	+
Asn	+	+	+	+	+	+	+	+	+
Gln	+	+	+	+	+	+	+	+	+
Glu					+		+	+	+
Ile	+	+	+	+					
Leu	+	+			+		+	+	+
Lys		+		+	+		+	+	+

3.3.3.3 Genes associated with sphingolipid biosynthesis

The sphingolipid biosynthesis pathway encoding genes: ehv031 (sterol desaturase), ehv050 (serine palmitoyltransferase), ehv077 (transmembrane fatty acid elongation protein) and ehv079 (lipid phosphate phosphatase) were all present in the newly sequenced genomes. These genes varied only slightly from their homologues in the EhV-86 genome. The % of nucleotide identity of ehv031 among strains (in comparison with EhV-86) ranged from 82 – 98 %, ehv050 ranged from 98 – 100 %, ehv077 ranged from 82 – 100 %, and ehv079 ranged from 78 – 100 %, with the genes

in the genome of EhV-202 being the most different with the lowest % of identity. A multiple sequence alignment of the SPT gene (ehv050) revealed that there were at least three distinct versions of this gene among strains (Figure 3.7). The SPT sequence of EhV-202 clustered outside these three groups.

Interestingly, the following CDSs: ENVG00245 (EhV-84), ehv329 (EhV-86), EOVG00319 (EhV-88), EPVG00153 (EhV-201), EXVG00143 (EhV-202), ELVG00153 (EhV-203), EQVG00147 (EhV-207), ERVG00100 (EhV-208) and ehv329 (EhV-99B1) were all predicted as a gene encoding an alkaline phytoceramidase (aPHC) based on a Pfam hit (protein family domain 05875) with an HMM (Hidden Markov Model) score of 27, an E value of $1.2e^{-06}$ and a percentage alignment of 90.26. This protein family consists of several ceramidases, which are enzymes involved in the regulation of cellular levels of ceramides, sphingoid bases, and their phosphates. A multiple sequence alignment of both the nucleotide and the amino acid sequences showed the sub-clustering of the gene into three groups with the exception of EhV-202 which was again an outlier to the three groups (Figure 3.8).

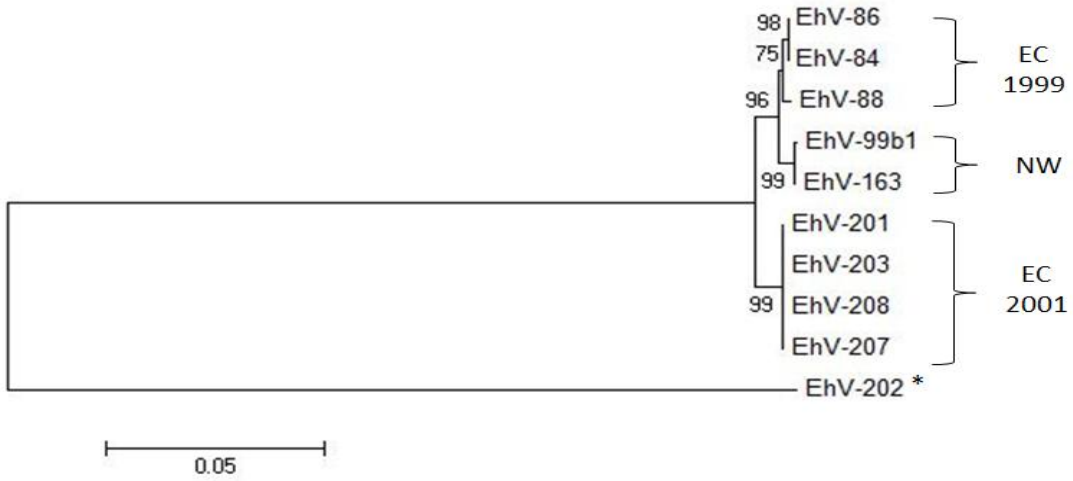


Fig. 3.7. Multiple sequence alignment of the serine palmitoyltransferase (SPT, ehv050) gene of ten coccolithoviruses (EhVs) from the English Channel (EC) and a Norwegian fjord (NW). * The English Channel strain EhV-202 clusters outside these three groups. The evolutionary history was inferred using the Neighbour-Joining method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analysed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches when greater than 50%. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option) (Felsenstein, 1985; Saitou and Nei, 1987; Tamura et al., 2004). Phylogenetic analyses were conducted in MEGA4 (Tamura et al., 2007).

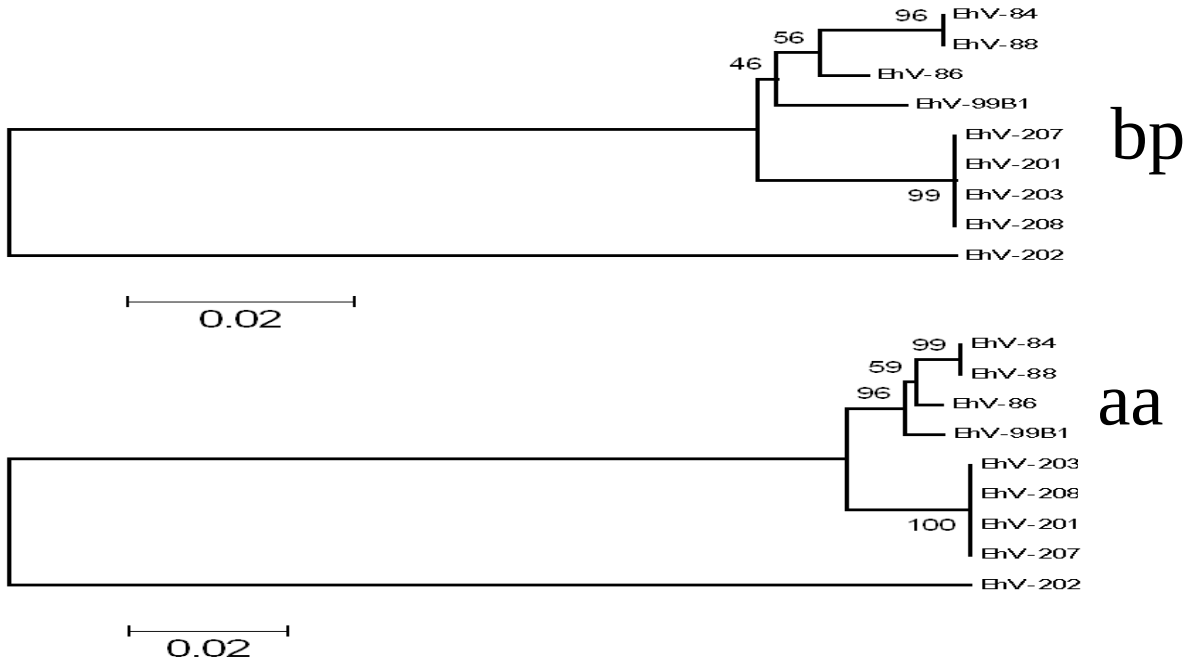


Fig. 3.8. Multiple sequence alignment of the predicted alkaline phytoceramidase (aPHC) encoding CDSs (804 bp) and their amino acid translation (267 aa).The evolutionary history was inferred using the Neighbour-Joining method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analysed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches when greater than 50%. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option) (Felsenstein, 1985; Saitou and Nei, 1987; Tamura et al., 2004). Phylogenetic analyses were conducted in MEGA4 (Tamura et al., 2007).

3.3.4. Homology to host genes

On average approximately 62 EhV protein coding sequences (or ~13% of the analyzed genomes) had a degree of homology to counterparts in the sequenced host genome of *Emiliania huxleyi* CCMP 1516. In most of the EhV genomes the gene with the highest similarity to a homologue in *E. huxleyi* encoded for a phosphate permease. EhV-202 was an exception as it had a 242 bp (80 aa) long gene predicted to encode for polyubiquitin with 96 % identity to a similar sequence in *E. huxleyi* CCMP 1516 (Table 3.5). A detailed description of the genes with predicted functions and their identity to counterparts in *E. huxleyi* can be seen in Table App3-App11. Nearly half of these genes had no assigned function or were predicted as genes encoding for hypothetical or membrane proteins and are therefore not shown in these tables. The majority of genes with high identity across the coccolithovirus genomes were genes encoding for deaminases, helicases, ribonucleoside-diphosphate reductases, and Lectin C-type domain proteins.

Table 3.5. Summary of the total number of genes or protein coding sequences (with and without function prediction) of nine sequenced coccolithoviruses that have homologues or exhibit a degree of similarity to counterpart sequences in the draft genome of *E. huxleyi* CCMP 1516. The analysis was performed in the build in BLASTp algorithm of the IMG/ER platform.

Genome	Total number of homologues	With assigned/ predicted function	Highest % identity to homologues in *E. huxleyi* 1516
EhV-84	54	38	phosphate permease (86.4 %)
EhV-86	66	44	phosphate permease (84.6 %)
EhV-88	57	38	phosphate permease (86.4 %)
EhV-201	60	35	phosphate permease (86.4 %)
EhV-202	66	51	polyubiquitin (96.0 %)
EhV-203	65	36	phosphate permease (86.4 %)
EhV-207	62	36	phosphate permease (86.4 %)
EhV-208	64	36	phosphate permease (86.4 %)
EhV-99B1	63	44	ribonucleoside-diphosphate reductase (66.8%)
Average:	~62	~40	

3.3.5. Notable absences of genes in the sequenced viruses

53 out of 66 protein coding genes with function predictions in the genome of EhV-86 have been initially observed to be expressed during infection of *E. huxleyi* (Wilson et al., 2005) and hence are believed to be important during infection. Eight of these expressed genes proved to be absent in one or more strain of the newly sequenced EhVs (Table 3.6). The endonuclease encoding gene ehv041 was only present in EhV-99B1 and EhV-86. Like in the other sequenced Norwegian strain (i.e. EhV-163), EhV-99B1 had the phosphate permease gene (ehv117) missing, and instead replaced by a 435 bp long fragment of that gene encoding a putative endonuclease. The DNA ligase gene ehv158 was missing together with four other genes in the same locus from EhV-84 and EhV-201. One of these other genes was ehv160 encoding a serine protease. This gene was missing in EhV-201 and also in EhV-84. The gene ehv184 which encodes a DNA binding protein was found only in EhV-86. A lectin protein encoded by ehv346 was absent in EhV-202, as was ehv363 (encoding an esterase). Outside of EhV-86, the helicase encoded by ehv430 was only found in EhV-99B1, and was replaced by two smaller hypothetical genes in EhV-84 and absent from the rest of the genomes. Finally, a sialidase encoded by ehv455 was truncated in all the English Channel strains isolated in 2001 and present only in the strains isolated in 1999. Although an alignment of this truncation can be seen for only EhV-86 and EhV-208 in Figure 3.9, the same applied for the other strains where this truncation had occurred.

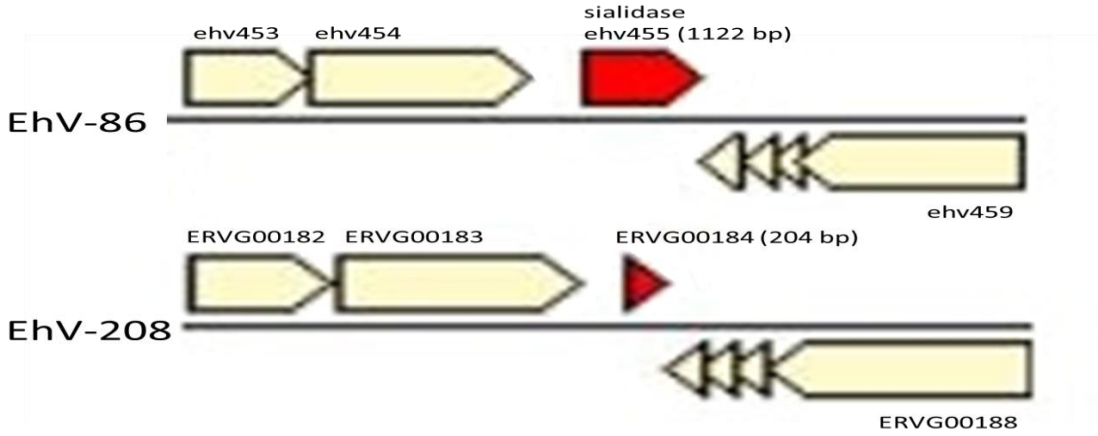
Table 3.6. The presence (+) of homologous genes in the newly sequenced EhVs with a previously assigned and expressed function in EhV-86 during a microarray infection experiment (Wilson et al., 2005).

Function	CDS	EhV-99B1	EhV-84	EhV-88	EhV-201	EhV-202	EhV-203	EhV-207	EhV-208
endonuclease	ehv041	+							
phosphate permease	ehv117*		+	+	+	+	+	+	+
DNA ligase	ehv158	+		+		+	+	+	+
serine protease	ehv160	+		+		+	+	+	+
DNA binding protein	ehv184								
lectin protein	ehv346	+	+	+	+		+	+	+
esterase	ehv363*	+	+	+	+		+	+	+
helicase	ehv430	+							
sialidase	ehv455	+	+	+					

* Expression of this gene was not confirmed during infection of *E. huxleyi* by EhV-86 in the study by Wilson et al. (2005) but later microarray studies have shown some level of hybridisation of this gene (Allen et al., 2007) .

3.3.6. Inferences from proteomic analysis of the EhV-86 virion

Previously, a proteomic analysis (by LC-MS/MS) revealed that the virion of EhV-86 is composed of at least 28 proteins (Allen et al., 2008). Many of these now have an assigned or a predicted function and the majority of their genes have one or more homologues in the newly sequenced genomes in this study (Table 3.7). Out of these 28, ten were absent in one or more genomes (i.e. ehv015, ehv034, ehv036, ehv038, ehv067, ehv100, ehv189, ehv191, ehv301, and ehv356), and six had more than one homologue in many of the other genomes (i.e. ehv034, ehv035, ehv038, ehv060, ehv189, and ehv191).




```
EhV-86_ehv455          ATGATGACAGGTGTACTATACTTGTTTTATGATTTGATTGCATTTCATATAATAAGTTTG
EhV-208_ERVG00184      ---ATGACAGGTGTACTATACTTGTTTTATGATTTTATTGCATTTCATATAATAAGTTTG
                          ******************************** ************************
EhV-86_ehv455          TGCAATACAGTTATTCAATGGTACAATACAGTTATGGATATTTGTTTTGGTAAAACTGGT
EhV-208_ERVG00184      TGCATTACAGTTGTTCAATGGTACAATACAATTATGGATATTTGTTTTGGTAAAACTGGT
                       **** ******* ***************** ****************************
EhV-86_ehv455          ATTATGTTTGGTGCACCTAGAGATATCGTAACGATATACAAACGCAACGAAGATGATTAT
EhV-208_ERVG00184      ATTATGTTTGGTGCACCTAGAGATATTGTAACGATATACAAACGCAACGAAGATGATTAT
                       ************************* **********************************
EhV-86_ehv455          TATACCTATCGTGTCCCGTCATTAGTTGTTGTTGGAAATGGGGACATGTTATGTTTTGCT
EhV-208_ERVG00184      CAA--CTAAACTATTTTAGAACGCTTTGA------------------------------
                        *   ***   * *      *    ***
EhV-86_ehv455          GAAGCTCGTAAATATTCATCATACGACCGCGGAACAGTTGGACTAGCAGTAAAACGTTCG
EhV-208_ERVG00184      ------------------------------------------------------------

EhV-86_ehv455          TATGATGGTGGAGATACTTGGGGTGATATTCACATCCTATTTGAAGCGGAAGGGATTACA
EhV-208_ERVG00184      ------------------------------------------------------------

EhV-86_ehv455          TATGGAAATCCAACGACTGTATATGCAAATGGTGCACTACATCTATTCTTCTGTAAGAAC
EhV-208_ERVG00184      ------------------------------------------------------------

EhV-86_ehv455          AACAAGCAAGTATACTACATGAGATCAATGGATAGTGCCGCGCATATATGGATAAAACCT
EhV-208_ERVG00184      ------------------------------------------------------------

EhV-86_ehv455          ACTGATATTACAACCGAAATTATCGGAATGTCAGAAGCATCTGTCTGGATTGCAACCGGT
EhV-208_ERVG00184      ------------------------------------------------------------

EhV-86_ehv455          CCAGGATCTGCATTGTATAACAATAATCGATTAATCGTACCAGTAAACTATGATGGTGGG
EhV-208_ERVG00184      ------------------------------------------------------------

EhV-86_ehv455          TCTGCTACATTTTATTCAGATACATATGGTCAGACATGGTCGGTGTCTAACAAAATCAAT
EhV-208_ERVG00184      ------------------------------------------------------------

EhV-86_ehv455          ATGGGTAATGAATGTCAAGTGATCCGGATTGGAGGAGATGCACTAATGATTAATTGTCGA
EhV-208_ERVG00184      ------------------------------------------------------------

EhV-86_ehv455          TACAATACTACAAATGGAAGGTCAACTACTCTTGATACAATCACATCTCCTCGTATTATT
EhV-208_ERVG00184      ------------------------------------------------------------

EhV-86_ehv455          ACATATTCGTCCGACAATGGGTCTACATGGGAAGATGTACATATGAAGTGGCATATGAAT
EhV-208_ERVG00184      ------------------------------------------------------------

EhV-86_ehv455          ACAAAATATGGAAGATCTACATGCCATGGGTCAATGATTTCATTAAAGACTGGTATGTTG
EhV-208_ERVG00184      ------------------------------------------------------------

EhV-86_ehv455          TATATCGGTCCAGAAAATAAAATTATACGTGCTGGGTTGTCAGTATGGTATTCTGTCATG
EhV-208_ERVG00184      ------------------------------------------------------------

EhV-86_ehv455          GGATCCAGATGGGAAAAAATATTAACAATTGATAATGGATTTTCGGCATATTCAACTATA
EhV-208_ERVG00184      ------------------------------------------------------------

EhV-86_ehv455          ACAACACTTGGTGATATGCATTTTATTGCATACGAACGTGGATGTGCATTTGATGGAATG
EhV-208_ERVG00184      ------------------------------------------------------------

EhV-86_ehv455          AAACCATATACTACAGACATTGTACTATATAAACTAAACTAG
EhV-208_ERVG00184      ----------------------------------------
```

Fig. 3.9. CLUSTALW alignment of the previously expressed ehv455 (sialidase) in EhV-86 and its homologue in the genome of EhV- 208 (ERVG00184). The 1122 bp long gene was truncated in all EhVs isolated in 2001 and instead remains a 204 bp fragment of that gene. The truncation is in a region that is otherwise mostly conserved and does not show a great degree of variability amongst the genomes.

Table 3.7. 28 proteins previously identified in the virion of EhV-86, their function, and the possible number of homologues gene copies of each in the genomes of EhV-86, EhV-99B1, EhV-84, EhV-88, EhV-201, EhV202, EhV-203, EhV-207 and EhV-208, based on the BLASTp sequence alignment of IMG/ER (max E value: 1e-05, min. percent identity: 40%).

Gene	Function	EhV-86	EhV-99B1	EhV-84	EhV-88	EhV-201	EhV-202	EhV-203	EhV-207	EhV-208
ehv015	putative membrane protein	1	0	1	1	1	1	1	1	1
ehv034	putative membrane protein	2	0	2	2	1	0	1	1	1
ehv035	similar to SMC2 protein	2	2	2	2	2	2	2	2	2
ehv036	HlyD family secretion protein	1	0	1	1	1	1	1	1	1
ehv038	putative membrane protein	2	0	2	2	2	0	2	2	2
ehv055	putative membrane protein	1	1	1	1	1	1	1	1	1
ehv060	lectin protein	2	2	1	1	3	4	3	3	3
ehv067	hypothetical protein	1	1	1	1	1	0	1	1	1
ehv085	major capsid protein	1	1	1	1	1	1	1	1	1
ehv100	predicted protein	1	1	1	1	1	0	1	1	1
ehv149	hypothetical protein	1	1	1	1	1	1	1	1	1
ehv168	hypothetical protein	1	1	1	1	1	1	1	1	1
ehv175	Putative serine/threonine protein kinase	1	1	1	1	1	1	1	1	1
ehv182	diaminopomelate decarboxylase	1	1	1	1	1	1	1	1	1
ehv189	pol-like protein	1	1	0	1	1	0	1	1	1
ehv191	putative membrane protein	2	2	0	2	3	1	3	3	2
ehv195	putative membrane protein	1	1	1	1	1	1	1	1	1
ehv200	putative membrane protein	1	1	1	1	1	1	1	1	1
ehv250	GCN5-related N-acetyltransferase	1	1	1	1	1	1	1	1	1
ehv301	NB-ARC domain containing protein	1	1	0	1	0	0	0	0	0
ehv325	envelope glycoprotein	1	1	1	1	1	1	1	1	1
ehv333	CRISPR-associated protein, Csel family	1	1	1	1	1	1	1	1	1
ehv340	putative fimbrial associated sortase-like protein	1	1	1	1	1	1	1	1	1
ehv356	DNA or RNA helicase of superfamily II	1	1	0	1	1	1	1	1	1
ehv358	putative thioredoxin	1	1	1	1	1	1	1	1	1
ehv442	conserved hypothetical protein	1	1	1	1	1	1	1	1	1
ehv454	hemocyanin isoform I	1	1	1	1	1	1	1	1	1
ehv461	fatty acid/ phospholipid synthesis protein	1	1	1	1	1	1	1	1	1

3.4. DISCUSSION

3.4.1. Structural characteristics and membrane protein composition: possible implications

The structure and diameter range of all viruses in this study was similar (180-205 nm), consistent with previous observations of members of this virus family (Dunigan et al., 2006). However it was intriguing to determine that although the viruses looked the same their genetic composition varied, including among genes that have been implicated to encode for virion capsid and membrane proteins. The numerous membrane proteins of the viral lipid layer, and the possible diversity in virion protein composition among strains, as seen from the proteomic comparison in this study to previous work conducted by Allen et al. (2008), suggests that the initial attachment strategies of some strains might be different.

In their original work, Allen et al. (2008) identified 28 proteins associated with the EhV-86 virion. These proteins are believed to be involved in the attachment and fusion of the virus to the host cellular membrane, before being endocytosised in their entirety into the cellular matrix, as well as in roles such as viral budding, the caspase pathway, signaling, anti-oxidation, and host range determination (Allen et al., 2008). In this study some of the genes that encode for these proteins had more than one homologue (or copy) in the newly sequenced coccolithovirus genomes. Hence it is possible that the amount of proteins associated with the virion membrane, their type and subsequently their function may vary among strains and could play a crucial role in determining the actual size of the virions (virion diameter range of 180-205 nm in diameter) and as a consequence the surface area available for attachment and adsorption into the host cell.

Moreover, it has been hypothesized previously that once internalized, the EhV virion is contained within a vacuole and is able to utilize internal microtubules to position itself near to the endoplasmic reticulum and nucleus for un-coating and replication of its DNA (Mackinder et al., 2009; Vardi et al., 2009; Kegel et al., 2010). It is possible that virus membrane protein composition and the number of gene copies that encode for different variations of these proteins are evolutionary adaptations that determine the host range of different virus strains and their initial movement within the infected host cell for transcriptionional initiation.

3.4.2. Evolution and classification

The genomic comparison of the closely related coccolithoviruses described here clearly showed that full genome sequencing of a large number of viruses, infecting the same host organism, is essential for our understanding of how these viruses have evolved and in revealing the genetic potential hidden in their gene pool. More importantly, the findings here highlight again the imperative question of how we classify a strain and at what point and degree of genomic and proteomic difference it becomes a new species (that is if viruses were to be included in the species concept). However, when the key genomic features are parallel among strains of the same organism (or virus) and their placement on a genome is consistent one cannot argue the concept of species, even in the case of EhV-202, in which the majority of the genome was less than 90% similar to the rest of the EhV genomes analyzed in this study. Large sections of repeats and shifts of entire clusters of genes around the genome, as seen from the aligned genomes here suggest that this is a more difficult task when dealing with viruses and that perhaps their evolution is more dynamic than previously thought.

Interestingly, both microarray data (Allen et al., 2007) and the genome analysis here indicate that the strains isolated in 2001 have no homologues to the EhV-86 genes ehv084 and ehv086. However the gene in between them- ehv085, which encodes for the virus MCP is (of course) present in all of these strains. This suggests that at some point in their evolutionary past a small genomic section (in this case comprising of three genes) has been replaced (or inserted), and that at the same time the current MCP gene present in these strains was acquired from another coccolithovirus, possibly during a multiple virus strain co-infection scenario (Figure 3.10). Such rapid and dynamic evolution, during which genes and entire clusters are lost or exchanged between the array of virus strains in the natural environment (some of which probably co-infect hosts at the same time) may be beneficial to them in the long term in keeping with the pace of host evolution and adaptation.

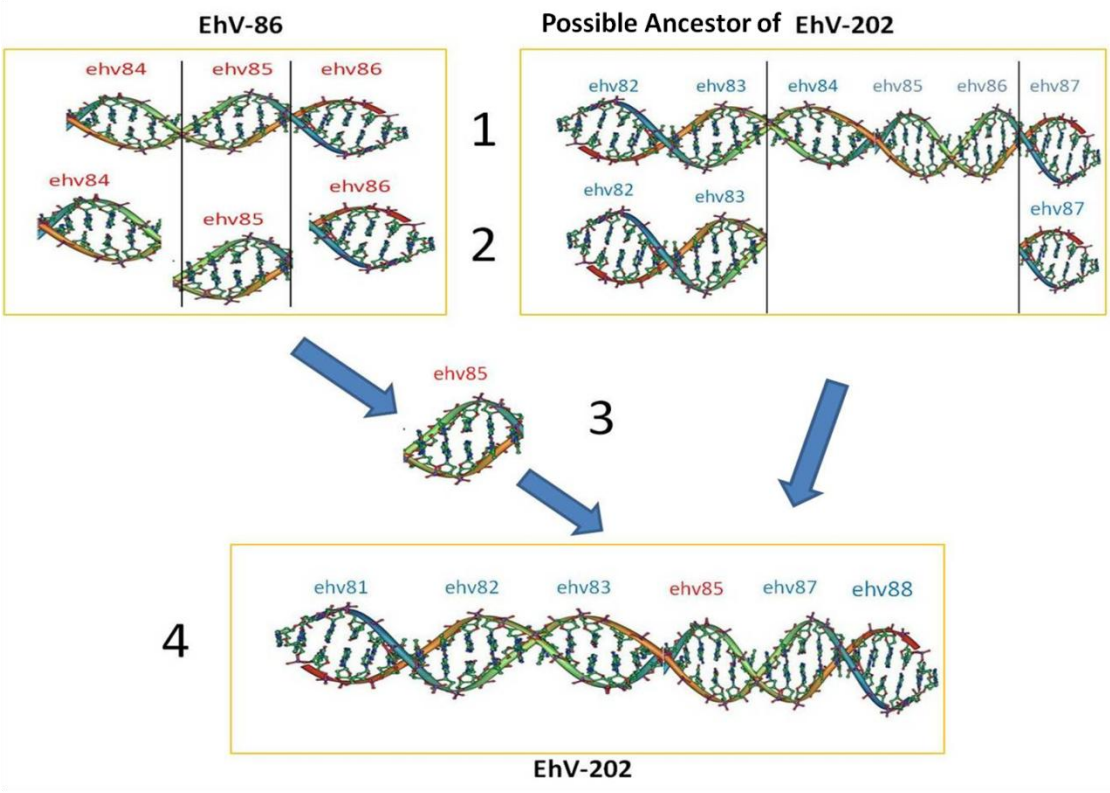


Fig. 3.10. Theoretical representation of the possible replacement of a small section of genes in the viruses isolated in 2001 (step 1 and 2) and the acquisition of an MCP encoding gene (step 3 and 4) from an ancestral EhV-86 like virus during a simultaneous co-infection and replication within an *E. huxleyi* host.

The phylogenetic analyses in this study clearly divide the coccolithoviruses into sub clusters, whether through the investigation of genes used for phylogeny or the analysis of functional genes. In fact one might conclude that in a sense, given that the genome of EhV-202 was so different on both a DNA and on an amino acid level, and had the highest amount of potentially unique protein coding sequences, is a strain with a more distant common ancestor than the rest of the strains in this study. Furthermore, EhV-202 may have a subtly different infection strategy to the rest. For instance, in phages LPS encoding genes (also predicted in the genome of EhV-202) are believed to be involved and used during infection in the establishment of prophage and the alternation of the cell-surface composition of their infected host, thus inhibiting the ability of other phages to attach to a potential host (Sullivan et al., 2005). If correct, such a feature may aid EhV-202 during infection and make it advantageous over other EhV strains. Other protein coding genes predicted to encode ADP ribose pyrophosphatase, tyrosine phosphatase and polyubiquitin were also found in the genome of EhV-202 and were absent from the other strains. The last is a protein known to be encoded by some viruses in order to enable viral replication by interfering with the host ubiquitin pathway. Many viruses, including members of the *Orthopoxvirus* family require a functional ubiquitine-proteosome system, essential for the degradation of un-needed or damaged proteins (by breaking the peptide bonds of proteins). In contrast, virally encoded polyubiquitin homologs of *Poxviruses* may also inhibit the ubiquitine proteosome system by breaking the linkage that targets proteins for degradation by the proteasome system (Gao and Luo, 2006; Randow and Lehner, 2009).

In addition, it seems that the predicted LPS gene has been recently acquired by HGT

event from the host as it was identical in 96% to a homologue in the genome of *E. huxleyi* 1516. In the virus the gene has no introns, suggesting that an mRNA intermediate and a possible retro-virus might have been involved in its acquisition from the host. Such genomic "thievery" by EhV-202 could potentially prove to be beneficial for the manipulation of metabolic pathways such as transcription and replication, thus establishing EhV-202 as the biggest "thug" of the current crop of coccolithovirus isolates. Nevertheless, regardless of its genomic differences, EhV-202 infects successfully and replicates within the same host as the rest of the viruses studied here, suggesting that the absence of these genes in the rest of the viruses in this study is neither essential nor even necessary for a successful infection and replication. What has driven the acquisition of these genes remains a mystery.

With regards to the phylogenetic marker genes DNA pol and MCP, although useful in the classification of viruses into groups and families, help very little in the understanding of selection and evolution in coccolithoviruses. However by using them in conjunction with ancient DNA sequences from the geological record (Coolen, 2011) and comparing them to sequences obtained from "modern" strains it is possible to fit the sequenced strains in this study into a historical frame hierarchy. It was seen that the MCP sequences cluster into three defined groups, i.e. the "modern" sequences were not constrained to one group and fitted into two out of the three possible groups, suggesting that in the past (at least with regards to the Black Sea) there were viral strains similar to the ones that were found in English Channel waters in 1999 and 2001. The fact that the alignment of the MCP fragment of "ancient" and "modern" EhVs and its % of identity ranged from only 92-100% (sequence alignment identity) also indicates that there is indeed to some extent, at least in this fragment of the MCP

gene, a high degree of conservation, that is present regardless of geographical location, time of isolation, or even functional characteristic differences among strains.

However, from the variation seen in other genes among strains in this comparative study, one should ask the questions of whether these marker genes can provide useful information with regards to the evolutionary arms race between the virus and its host, the selection pressures that these viruses were exposed to along their evolutionary history (variable environmental conditions, host resistance, shift in dominance of hosts, etc.) and the HGT events with the host that have occurred in the short and long evolutionary past of coccolithoviruses. Hence it can be confirmed that DNA pol and MCP are useful markers for the classification of *Phycodnaviruses* and EhVs, but less suitable for evolutionary analysis as a whole when the main purpose of the study is function and its effect on host-virus dynamics and niche biogeochemistry.

3.4.3. Sphingolipid biosynthesis genes

Genes encoding for a sphingolipid pathway have been implicated in the modulation of the apoptotic pathways of their host (Michaelson et al., 2010, and as previously discussed in Chapter 1). In coccolithoviruses putative novel sphingolipids have been suggested to be biosynthesized during the viral infection process. It was proposed (Pagarete et al., 2009) that such virus-specific sphingolipids (such as the ones described by Bidle et al., 2009) could play a role in virion release. In this theory the formation of sphingolipid rafts are promoted by the virus to occur in greater number in the host's outer plasma membrane in order to become focal points for viral budding and release of virions. As the viral genome is unable to complete the synthesis of sphingolipids without the aid of the host's cellular machinery it is likely that each

gene involved in the pathway has been taken from the host in individual HGT events to promote such a process. This could account for the piecemeal acquisition of the pathway by the coccolithoviruses whereby any improvement in the extent of lipid rafts in the outer membrane could offer a significant advantage to the infecting virus, and is not reliant on the presence of a complete pathway. Indeed, no coccolithovirus actually has the complete pathway for sphingolipid biosynthesis encoded on its genome (a homologue for 3-sphinganine reductase is missing), therefore there is an existing and established reliance on this host encoded function.

Despite small changes in nucleotide and amino acid sequence, the sphingolipid biosynthetic machinery encoded within all the EhV genomes in this study appears to be conserved. Whether this was the case in the evolutionary history of the coccolithoviruses remains to be seen. An in depth investigation of the coccolithovirus encoded sphingolipid gene diversity in samples from sediment records, such as the ones extracted and analyzed by Coolen in the Black Sea (2011) would allow to not only determine the conservation status of these genes but also reveal at what point each and one of these genes were acquired and under what selection pressure if any this has occurred.

The discovery of an additional gene potentially involved in sphingolipid metabolism in this study: the enzyme alkaline phytoceramidase (aPHC) is also surprising. The putative function of this gene has not been discovered in a previous analysis of the full genomes of EhV-86 and EhV-99B1, and the partially sequenced EhV-163. Its discovery today might be due to improved BLAST algorithms and the availability of more nucleotide and amino acid sequences in the public domain with a known

function than there were seven years ago when the first coccolithovirus was sequenced. The possible presence of such a gene raises interesting questions with regards to the acquisition of this pathway in coccolithovirus genomes and the ability of coccolithoviruses to control cellular levels of ceramide through the biosynthesis of their own ceramide controlling enzymes such as aPHC. The gene predicted to encode for the enzyme aPHC does not have a homologue in the genome of *E. huxleyi* and it is not clear where this gene was "stolen" from or what its origin might be. Its possible function however can be put into the context of the virus encoded sphingolipid biosynthesis pathway. The end product of this pathway is a ceramide-like sphingolipid and is regarded as an important structural and signaling molecule in *E. huxleyi* cells (as previously discussed). Given that ceramide functions in regulating and proliferating PCD and apoptosis in eukaryotic cells (Mao and Obeid, 2008), the presence of a virally encoded ceramidase could be therefore advantageous to the virus as its level of expression and translation may influence the level of available ceramide and its further transformation and breakdown to sphingosine and sphingosine 1-phospates (Figure 3.11). Furthermore, having the ability to control the levels of ceramide in the host cell will also allow to a certain extent the control over the biosynthesis of sphingosine 1-phosphate; a molecule which has many functions. For instance, in yeast it suppresses growth and enhances the tolerance of a cell to heat stress (Mao and Obeid, 2008).

Hence, ceramides and their regulation through enzymes such as ceramidase are critical in achieving a balance between the lipid intermediates within the sphingolipid biosynthesis pathway. An imbalance in one or more intermediates can result in an altered and abnormal physiological state of the eukaryotic host cell, impairing as a

consequence its fitness and chances to fight virus infection. In the *E. huxleyi*-virus system, the ecological significance of this pathway manipulation could be far reaching: viral induced control of infection dynamics during a bloom has global consequences for biochemical cycling in locations where hosts and viruses persist, and also over geological scales. The deposition of calcite in sediments as seen from cores analyzed in the Black Sea (Coolen, 2011) is largely a result of these host-virus interactions and the active pump of calcium into the deep ocean post virus induced death and host demise.

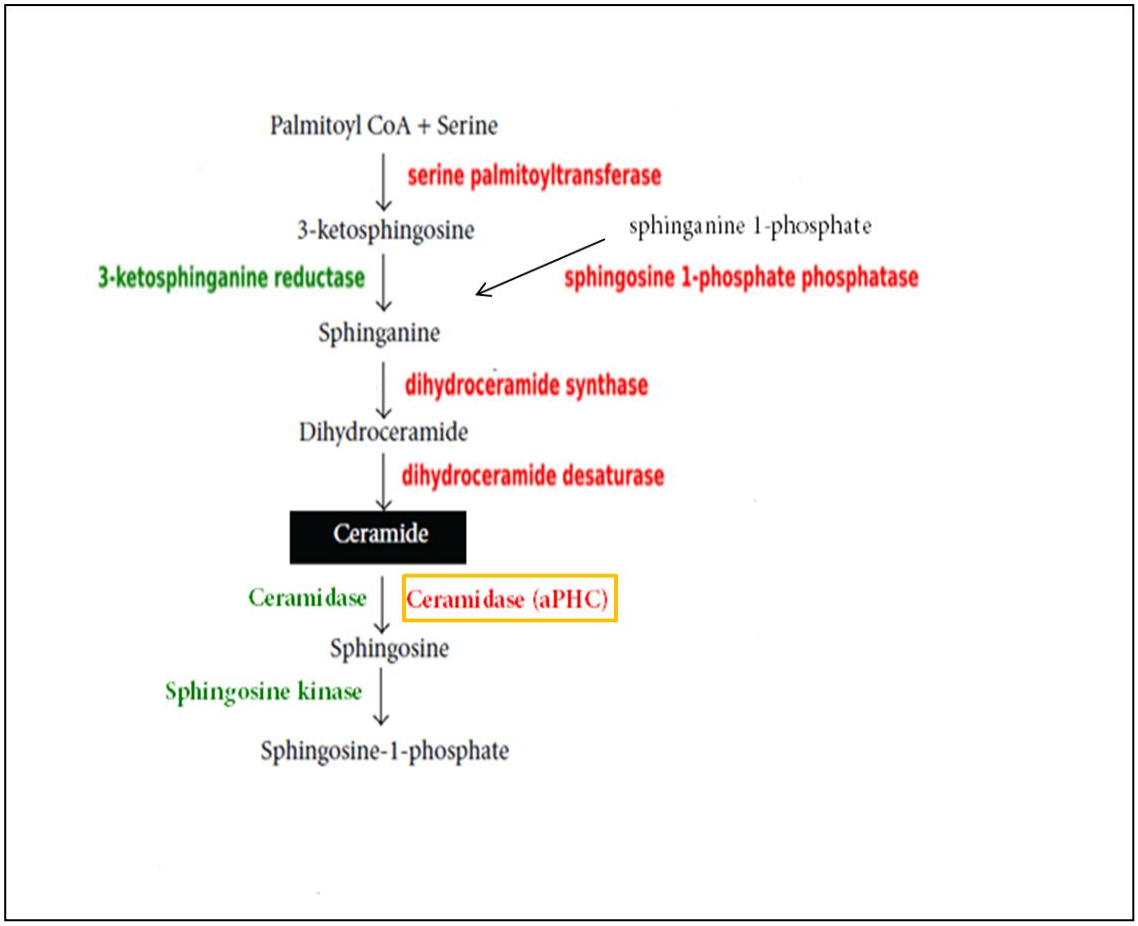


Fig. 3.11. Hypothetical *de novo* sphingolipid biosynthesis pathway encoded by the EhVs (adapted and modified from Monier et al., 2009). In red are the enzymes that are encoded in the genome of the EhVs and were stolen from the host via HGT (with the exception of aPHC); in green are enzymes encoded only in the genome of the *E. huxleyi* host; and in the yellow box is the newly identified enzyme alkaline phytoceramidase that was predicted to be encoded by the EhVs.

3.4.4. Single gene and tRNAs variability

The reasons behind the variability of some genes, the susceptibility of some of these to change, and their observed absence from a particular EhV genome are puzzling. Whilst MCP (or at least the fragments studied here) has changed little with time, other genes are either absent completely from some EhV strains or have been truncated to lose their original function. For instance, the gene predicted to encode for the enzyme sialidase (sometimes also named neuraminidase) was truncated in all the English Channel strains isolated in 2001. This gene has been shown to be an important virulence factor in influenza and other viruses, as it helps in the release of the viruses from the host cell (Roggentin et al., 1993; Taylor, 1996; Hughes et al., 2000; Schauer and Kamerling, 2011). In the *E. huxleyi*-virus context, the virally encoded sialidase was indeed expressed during a lytic infection of EhV-86 (Wilson et al., 2005). However, whether coccolithovirus strains with or without sialidase display distinct phenotypes during an infection due to the presence or absence of sialidase remains to be determined. Moreover, it is impossible to conclude whether the fact that only viruses without the sialidase encoding gene were isolated in 2001 (as opposed to the ones with a sialidase encoding genes isolated in 1999) is due to the isolated strains being the most common at that particular time in the bloom (due to an actual change in the EhV community in the English Channel over time), or more importantly (but highly unlikely) a microevolutionary event in which this feature has been lost in all EhV strains between the years 1999 and 2001. It could simply be that the EhV strains isolated in 2001 correlated with a specific dominant strain of *E. huxleyi* at the time of isolation which influenced the occurrence of a dominant EhV community that lacked particular genomic features such a sialidase. Another possible explanation for this could be that strains similar to the ones isolated in 1999 have not disappeared but have been translocated to a different geographical region, due to the high degree of

circulation and mixing of the coastal waters in the English Channel and their association with a particular *E. huxleyi* strain or strains that are no longer present in this environment. With regards to possible environmental selection pressure that might have driven the loss of this gene in some of these virus strains, it is not currently known whether physicochemical factors influence or impact the operation of the sialidase enzyme. So far the only report dealing with this refers to the sensitivity and stability of virus encoded sialidase to alternations in pH in influenza-A viruses (Takahashi et al., 2003).

The absence of other genes with a known or predicted function in some of these virus strains is also puzzling. The gene encoding for a phosphate permease (PP) in the genome of EhV-86 (i.e. ehv117) has been deleted not only from the genome of EhV-163 but also from the genome of EhV-99B1 (also shown in Pagarete et al., 2013). In the genome of *E. huxleyi* CCMP 1516 the PP encoded protein was suggested to be used in order to increase the uptake of environmentally available phosphorus from the aquatic environment for essential cellular functions (Monier et al., 2011). The presence of this gene on some coccolithovirus genomes (English Channel strains) suggests an evolutionary fitness advantage for these isolates at low concentrations of environmental phosphate. If the host is able to increase the concentration of bio-available phosphate due to the presence of this coccolithovirus gene product then the host may be less limited by phosphate and hence viral infection would also be more successful (as phosphate is essential for many biochemical pathways within cells including virion production and genome replication) (Bratbak et al., 1993). Although phosphate limitation does not stop infection it has indeed been shown to reduce the virus burst size (Bratbak et al., 1993; Jacquet et al., 2002; Danovaro et al., 2011). It is

not however, clear why these genes are found in the genomes of the English Channel isolates in this study, given that the English Channel, the environment from which these isolates originate from is rarely limited by phosphate (Smyth et al., 2009).

The detection of an endonuclease in EhV-163 and EhV-99B1 at the same locus which is absent from the "English" isolates studied here provides some indication of placement in the early hierarchy of the EhV lineage. The loss of PP in the Norwegian isolates suggests that the "English" strains are ancestral to the Norwegian ones. Although the specific function of the endonuclease in these strains has not been identified the general function of endonucleases is known to be the repair of damaged or mismatched nucleotides or their degradation to free phosphate within them for other functions (Van Etten et al., 1991). Perhaps the deletion of this gene in the "Norwegian" strains suggests a level of allopatric speciation and the next stage of their evolution in which the PP gene product is not required. The reasons behind its replacement by a putative endonuclease are currently unknown, and future studies will have to concentrate on linking the presence and absence of genes with environmental parameters in order to conclude whether the observed differences are indeed a result of adaptation to local environments where the physicochemical conditions differ.

One should refrain from over interpreting these differences due to the unknown phenotypic characteristics of EhV-163 and EhV-99B1 during active infection in comparison to their "English" relatives. Comparing the infection dynamics of: "shared" (multiple strains infecting the same host at the same time) versus individual virus strains (single strains infecting the same host at the same time), "Norwegian" versus "English" strains, and the careful monitoring of gene expression of important

genes via either microarrays or qPCR methods, will determine whether differences in the presence/absence of certain genes actually affect infection dynamics, the successful infection and replication by different coccolithovirus strains, and thus the abundance of the host and its subsequent demise (some of these issues will be dealt with in Chapter 4).

Whenever evolutionary studies are performed, one should always keep in mind that some features could be due to random selection or simply due to chance. This is most likely more than true in the case of the tRNAs composition observed in this study. Although there appeared to be a group of "essential" tRNAs present in all the EhV genomes in this study, there were also some that can be described as "non essentials", i.e. lacking in one or more genome. tRNAs are a useful arsenal of weapons that many viruses possess. By encoding their own tRNAs and "adapting" them to their codon composition, viruses bypass the reliance on host tRNAs. For instance PBCV-1 encodes 11 tRNAs, even though its genome size is only 330 kb long and predicted to encode for 366 proteins (Li et al., 1997), much less than the coccolithoviruses. In PBCV-1 tRNAs play a crucial role in the synthesis of the viral proteins by supplementing the lack of particular tRNA in the host genome. However this almost random distribution of the "non essential" tRNAs in the coccolithovirus genomes may influence their adaptability potential to different host strains in which certain tRNAs are present in minor amounts, and in doing so enhancing the translation efficiency of coccolithovirus genes. Thus a large number of virally encoded tRNAs can be seen also as an advantageous characteristic. In this case, if judged solely by its tRNAs composition, EhV-202 would appear as a poor competitor with potentially smaller host range as it only encodes for three tRNAs, whereas EhV-86 for instance encodes

for six tRNAs. A future investigation of the host range of these different viruses and the in-depth analysis of their use of their tRNAs during infection would be a natural next step in the study of these viruses.

3.4.5. HGT events with the host

To date, the only genes of known function acquired from the host via HGT events by coccolithoviruses are the genes involved in the sphingolipid biosynthesis pathway and the gene encoding for PP. However the discovery here that on average 13% of the coccolithovirus genes under investigation share homology with genes in the draft genome sequence of *E. huxleyi* 1516 (Read et al,, 2013) is striking and raises interesting questions with regards to not only the co-evolution of coccolithophores and their viruses but also the evolution and establishment of the eukaryotic cell.

There are two possible scenarios that can explain the large degree of homology seen here. The first and most likely scenario is through the occasional flow and incorporation of host genes into the genomes of their viruses upon simultaneous virus and host replication (i.e. HGT and lateral gene flow). If one is to accept that viruses need to "steal" genomic material from their hosts in order to co-evolve with them and ensure their own evolution and survival, then in the case of the coccolithoviruses this means that the homologous genes seen at present have been gradually acquired from the host through a long history of co-evolutionary arms race. Following acquisition some of these genes have been also liable to loss, as seen from the loss in PP gene in the Norwegian isolates. The second scenario, one that is somewhat controversial and not fully accepted by the majority of evolutionary biologists is the "viral eukaryogenesis" hypothesis proposed recently by Bell (2011), in which was claimed

that the origin of the ancestral nucleus is in fact a large, complex DNA virus that established a continual presence in the cytoplasm of archaea and eventually evolved to be an eukaryotic cell (with features similar to archaea and DNA viruses alike). If this hypothesis is to be extrapolated here then this will mean the opposite; i.e. the coccolithoviruses have been losing genes, and drifting away from their ancestral DNA virus.

Regardless of which theory one chooses to support, it is difficult to ignore the fact that large portions of genomic material are shared between the coccolithoviruses in this study and their host. Whatever the metabolic and evolutionary driving forces behind the sharing of material, it is a crucial feature and recurring theme in the study of co-evolution of algae and its viruses. Moreover, there is no doubt that in the future with the advances in the prediction of genes, proteins and functions of microbes, the knowledge of the current "unknowns" in the genomes of the ecologically relevant coccolithoviruses will also increase, and the identity of many hypothetical protein encoding genes that currently overwhelm the few of known function will be revealed. This will allow us to better understand the reasons and functional relevance behind the acquisition of large chunks of host sequence in the viral genomes.

3.5. CONCLUSIONS: WHAT THE FUTURE HOLDS FOR EHVS

The analysis presented here is a complex comparative study of marine viruses with an established ecological function. As an important first step, this study highlights the enormous genetic diversity hidden in coccolithoviruses and illustrates the importance of using well characterized model virus strains in the study of evolutionary relationships between different virus strains, host-virus co-evolution, and the possible role of biogeography in shaping modern marine virus genotypes.

Although mostly similar, the genomes examined here also presented features that were unique among strains, with genetic material lost and gained in various strains. Most importantly, this study shows that a core set of genes, some encoding for a near complete biochemical pathway are conserved; regardless of the geographical origin of virus isolation in which they are found or the time at which these viruses were isolated. Such conservation provides crucial evidence for the importance of particular biochemical pathways and metabolic processes (such as the virally encoded sphingolipid biosynthesis pathway) and their manipulation during infection of coccolithophore blooms in the natural environment.

The majority of genomes analyzed here are ‘only’ drafts at approximately 99% completion. Their completion in the future could provide even more information as there may be important features that were missed in the gaps between the sequenced contigs of each genome during the sequencing process. The repetitive regions which are difficult to sequence in particular, could hold vital clues to genome replication and transcriptional control. Nevertheless, the information that is provided from these drafts is more than sufficient for the generation of new hypotheses and the beginning of a

series of experimental approaches in the laboratory. Now that the genotypes of these strains are known it is possible to establish a direct link to their phenotypes, i.e. their infection dynamics and structural characteristics; with the ultimate goal of integrating this information into ecological models of primary production, bloom dynamics and nutrient recycling in the worlds' oceans.

This study and the algal virus system it describes is an excellent example of thievery of genetic material by viruses on a scale never observed before in the marine realm of virology. At first glance these viruses appear as nothing more than thugs, manipulating and coercing their hosts' biochemical machinery, selfishly to their own advantage. However they may not look so fierce if their actions are judged in an evolutionary perspective during which they drive not only their own evolution, but also the evolution and diversification of their hosts through a thoroughly unique co-evolutionary arms race for survival. It is unclear however how the genomic differences observed in this Chapter and other studies on coccolithoviruses to date influence the infection dynamics of these viruses with their hosts and whether different virus strains exhibit different infection dynamics. Therefore the potential phenotypic differences between two coccolithovirus strains are explored in Chapter 4.

4. INFECTION DYNAMICS OF COCCOLITHOVIRUSES: THE VIRUS "FIGHT CLUB"

4.1. INTRODUCTION

To date, coccolithoviruses have been shown to be a major cause for the demise of blooms in both temperate and sub temperate oceanic regions (Bratbak et al., 1995; Martinez et al., 2007; Brussaard et al., 2008; Sorensen et al., 2009). Their role in regulating coccolithophore phytoplankton assemblages has been firmly established through large scale environmental studies where researchers have monitored the virus-induced decline of blooms in both the open ocean and continental shelf regions, and induced semi-natural blooms in large enclosures of seawater (mesocosms) in Norwegian fjords (Wilson et al., 2002; Martinez et al., 2007, 2012). During these studies the main parameters under investigation were host and virus abundance using flow cytometry and sometimes molecular fingerprinting to establish the extent of virus infection, host biomass reduction, and subsequently the identity of the virus strains present at the time of sampling. With the exception of a laboratory study that investigated the susceptibility of different *Emiliania huxleyi* host strains to infection by a single coccolithovirus strain (Bidle and Kwityn, 2012), the majority of data regarding host-virus dynamics in this system comes from these former field studies, where the populations of host and virus were examined as a whole (i.e. by means of AFC where distinction between different strains during infection dynamics could not be made) and strain specific differences were not investigated.

Although useful, these previous studies do not reveal the single strain variability differences of coccolithoviruses with regards to infection strategies and mechanisms.

Within a coccolithophore bloom, the success of coccolithoviruses will vary and will be influenced most likely by the type of host strains present (Wilson et al., 2002a; Martinez et al., 2007, 2012; Rowe et al., 2011; Coolen, 2011). There will be a pool of different virus strains and many host strains (some of which will be more dominant than others). Given that it was already established in Chapter 3 that there are clear genetic differences between coccolithovirus isolates (i.e. there are genes that are either present or absent in some coccolithovirus strains), it is likely that within a mixed community of viruses which are all co-infecting during bloom conditions, subtle variations in their phenotypic properties, i.e. infection dynamics, may have a profound impact on the overall composition of the host community.

The re-occurrence of coccolithoviruses and their persistence in the environment and co-occurrence with a host population can be explained by the stable co-existence theory proposed by Thyrhaug et al. (2003) in which it was proposed that the co-existence of algae and viruses is possible due to an efficient feedback mechanism that reduces the rate of infection when the host abundance is low due to initial virus infection and population demise. Surprisingly, the authors concluded that this feedback mechanism was due to the development of resistance in the host populations under study to virus infection, which allowed the cultures to eventually recover due to phenotypic plasticity of the hosts, and not due to a phenotypic change in the infecting viruses.

However the role of viral phenotypic differences and its effect on the algal population needs to be investigated further. Whether viral phenotypic differences are driven and influenced by variations in host resistance and susceptibility to infection, or due to a

natural variation in the infection dynamics strategies employed by different coccolithovirus strains (due to their enormous genetic potential hidden in their large genomes) remains to be determined. Recent work by Thomas et al. (2011) showed that green algal species such as *Bathycoccus* sp., *Micromonas* sp. and *Ostreococcus tauri* can potentially develop resistance if exposed to the same virus strain over a period of time in the laboratory. They proposed that host resistance occurred by either spontaneous mutations that are selected for during viral infection or by an as yet unknown chemical signal from infective cells to the rest of the population that causes a shift in the population towards resistant phenotypes. The latter was indeed shown during infection of *E. huxleyi* in which virus-susceptible diploid cells were observed to undergo meiosis post virus infection to produce haploid cells that were resistant to virus infection (Frada et al., 2008). Furthermore, in the *E. huxleyi*–virus system it was also observed that apoptosis triggered by the secretion of virally encoded sphingolipids also might infer resistance in some cells (Pagarete et al., 2009) by acting as a "chemical alarm" for infection to the rest of the population (Vardi et al., 2009; Bidle and Vardi, 2011).

The exact mechanisms through which resistance occurs are unclear. It is however believed that resistance to viral adsorption might be conferred by mutations in the host cells that include the structural modification of cellular receptors on the algal surface and the decrease in the number or loss in particular receptor sites vital for virus attachment (Grimsley et al., 2012). The occurrence of resistance however comes at a cost to the host cells in the form of reduced growth rates (Thyrhaug et al., 2003) and the possible increase in their sensitivity to other viruses (Bidle and Vardi, 2011).

In the laboratory, some coccolithophores have been shown to be susceptible to one or more coccolithovirus strains (Bidle and Kwityn, 2012, and personal observations), hence like in any other biological system, competition between viruses for infection and successful replication exists. From an evolutionary perspective the viruses that are the most competent and efficient infection agents will be the ones from which future generations or coccolithovirus strain types derive, in any given ecological niche.

Hence the focus in this chapter will shift from the traditional study of host phenotypic plasticity and resistance to the study of virus phenotypic differences and their manifestation during infection. It is only through an experimental approach in the laboratory that these phenotypic differences between coccolithovirus strains can be observed and documented. In this chapter infection dynamic experiments were performed with two strains that were shown in Chapter 3 to differ in part of their genomic material: EhV-86 and EhV-207. For simplicity, it was decided that the focus would be to identify any potential differences exhibited by the two viruses during infection, and not consider host variability here. Therefore a single strain of host algae was chosen as a candidate for infection: *E. huxleyi* CCMP 2090. The main goal was to compare the infection dynamics exhibited by EhV-86 and EhV-207, and assess their impact on host growth, which was revealed by competitive interactions of the two viruses during a virus "fight club" experiment. Strain specific primers were designed for each coccolithovirus, and the abundances of virus like particles (VLPs) were monitored for one week post- infection by means of qPCR and Analytical Flow Cytometry (AFC).

4.2. MATERIALS AND METHODS

4.2.1. Culture conditions and setup

E. huxleyi CCMP 2090 was grown in a 2 L f/2 medium (Section 2.1.3), stirred and aerated in an Infors Minifors chemostat bioreactor (Figure 4.1) at a light and dark cycle of 16:8 hours respectively, temperature of 18°C, and white light intensity of ~86 µM photons m^{-2} s^{-1}, prior to the beginning of the experiment. Once the *E. huxleyi* in the chemostat was at a cellular density of 1.5×10^6 mL^{-1} (i.e. beginning of exponential growth), 100 mL were distributed into 12 polystyrene sterile tissue culture flasks (CellStar) (Figure 4.2).

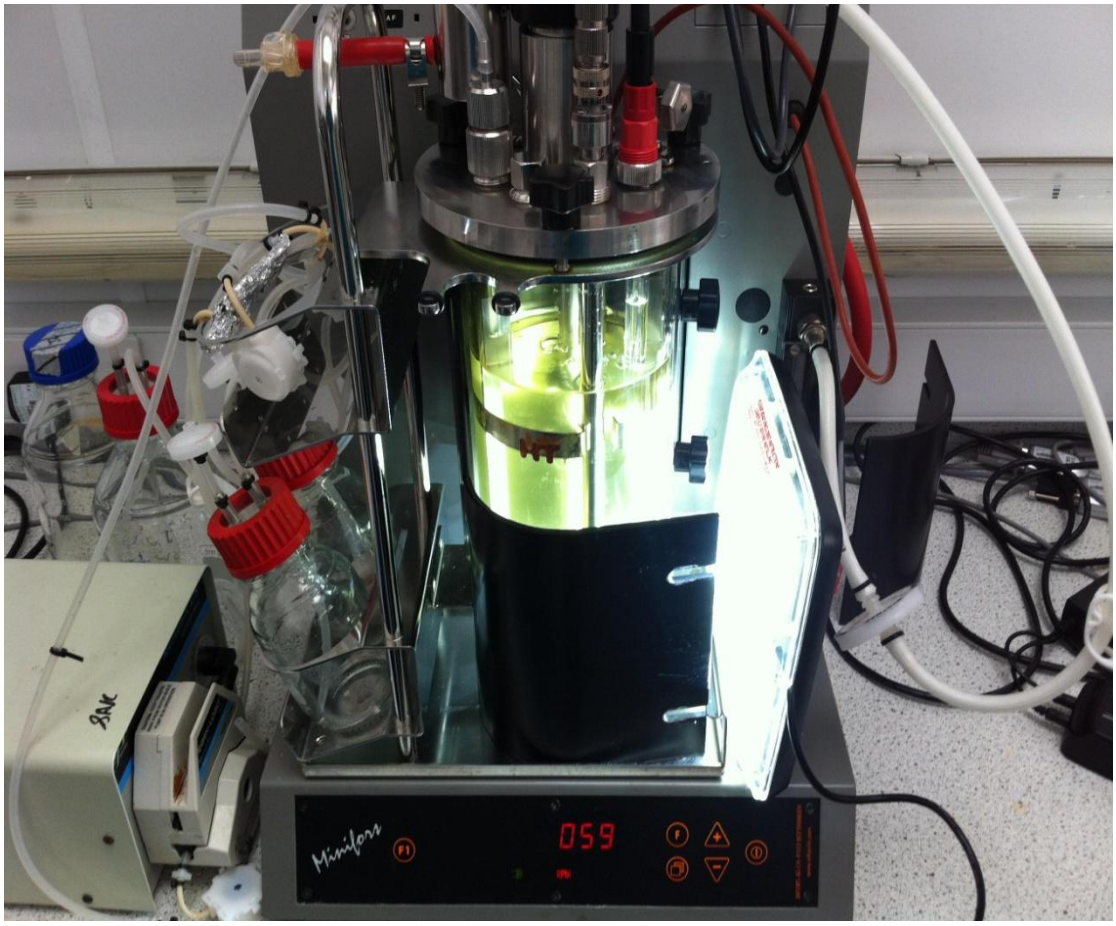


Fig. 4.1. Infors Minifors chemostat bioreactor in which the *E. huxleyi* CCMP 2090 biomass was generated prior to inoculation of the culture flasks during the "fight club" experiment.

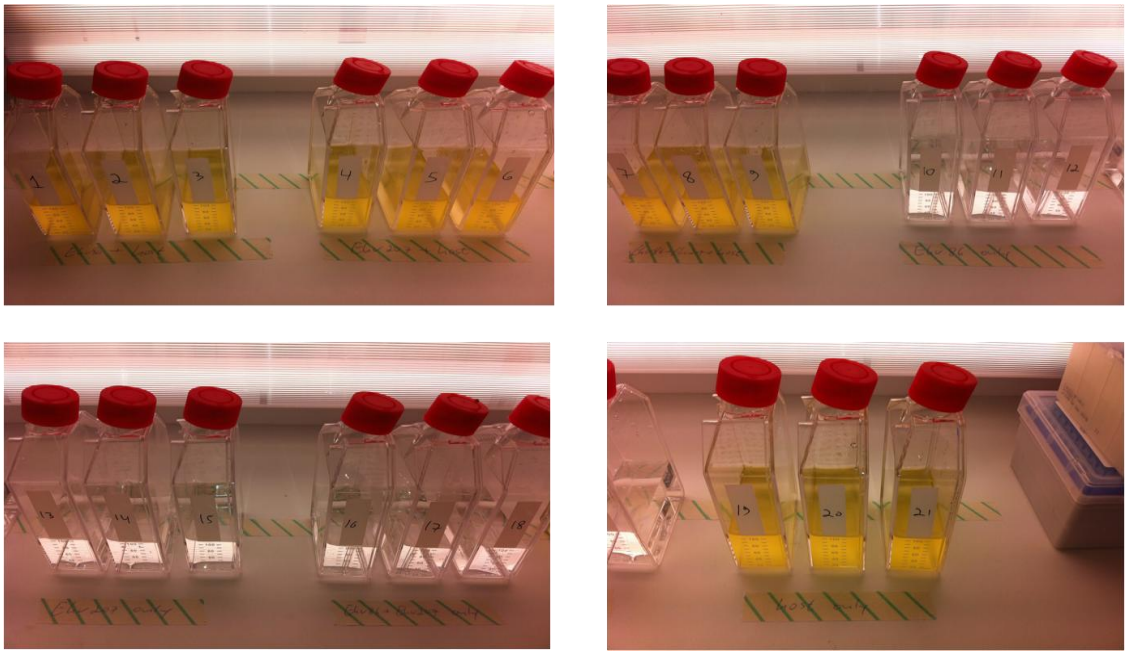


Fig. 4.2. Culture flasks in which the virus "fight club" experiment was conducted, containing aliquoted 100 mL of exponentially growing *E. huxleyi* CCMP 2090 from the Infors Minifors chemostat bioreactor.

The 12 algal subculture flasks and 9 other culture flasks that contained only f/2 were placed in a culture cabinet under the same conditions as in the chemostat and left to acclimatise for 48 h (Figure 4.2). On the day of the experiment (t0) the flasks in the culture cabinet were organized into seven different culture conditions (A-G) as described in Table 4.1. Ten minutes before the first sampling point (t0), cultures 1-3 (A) were inoculated with fresh EhV-86 lysate with a multiplicity of infection (MOI) of 1:1, cultures 4-6 (B) were inoculated with EhV-207 with a MOI of 1:1, and cultures 7-9 (C) were inoculated with both EhV-86 and EhV-207 at an overall MOI of 1:1, with a 1:1 ratio of each virus. Cultures 10-12 (D) were virus controls and contained no host (only f/2), and were inoculated with the same volume of EhV-86 lysate as cultures 1-3. Likewise, cultures 13-15 (E) contained only f/2 and were inoculated with the same volume of EhV-207 lysate as cultures 4-6, and cultures 16-18 (F) contained only f/2 and were inoculated with the same volume of both EhV-86 and EhV-207 lysates as in culture 7-9. Cultures 19-21 (G) were the negative controls containing host

with lysates of EhV-86 and EhV-207, that were inactivated by autoclaving at 126°C, followed by filter sterilization through a 0.2 µm sterile cellulose acetate syringe filters (Gilson), and overnight exposure to UV light, prior to their addition to the control flasks.

4.2.2. Probe design and optimisation

The selection for specific primer probes for the two viruses EhV-86 and EhV-207 was performed on the IMG/ER analysis platform, in which the protein coding genes of EhV-86 and EhV-207 were compared using BLASTp with a maximum E value of $1e^{-05}$ and minimum percent identity of 40%. Following this analysis, EhV-86 and EhV-207 strain specific genes were identified (i.e. ehv290 [510 bp] and EQVG00465 [1059 bp], respectively) and selected for the design of PCR primers. The design of the primers was as described in Section 2.3.1 and the final sequence lengths that were to be amplified by the PCR/qPCR were 209 bp (for EhV-86) and 353 bp (for EhV-207) long. Primers were ordered from and supplied by the Sigma-Aldrich Company Ltd.

4.2.3. PCR with qPCR primers for EhV-86 and EhV-207

In order to first establish that the designed primers were strain specific, a standard PCR (Section 2.3.2.1) was performed on fresh lysate stocks of EhV-86 and EhV-207 with the primers for ehv290 (i.e. qPCR(EhV-86)-F and qPCR(EhV-86)-R) and EQVG00465 (i.e. qPCR(EhV-207)-F and qPCR(EhV-207)-R). The PCR was conducted in triplicates and the set up conditions were as described in Section 2.3.2.1. The PCR cycle included an initial denaturation step at 95°C for 3 min, followed by 34 cycles at 95°C, 60°C and 74°C for 30, 60 and 90 sec respectively, and a final cycle at 95°C, 60°C and 74°C for 30 sec, 5 min and 5 min respectively.

4.2.4. Sampling procedure

Once experimental conditions were set up in the flasks the cultures were left for ten minutes so that the viruses could acclimatise to the culture conditions and establish initial attachments/infections with the host cells. Then 3 mL of sample were taken out from each at the following times: 0 (t0), 1 h (t1), 2 h (t2), 3 h (t3), 4 h (t4), 8 h (t5), 12 h (t6), 24 h (t7), 48 h (t8), 72 h (t9), 96 h (t10), 120 h (t11), 144 h (t12), and 168 h (t13). The 3 mL were further divided into three sub-samples. 1 mL was fixed with 0.5% glutaraldehyde for the enumeration of host abundance using AFC and 1 mL was spun down at a speed of × 13,000 rpm for 30 sec and the top phase removed into a new tube, for the analysis of free and attached VLPs by AFC. The pellet was re-suspended in 1 mL of DNA free water and both were fixed in 1% glutaraldehyde. The same was performed with the last subsample of 1 mL with the exception of the addition of the glutaraldehyde as these samples were used for qPCR. All samples were snap frozen in LN_2 and stored at -80°C for further analysis.

4.2.5. Analytical Flow Cytometry (AFC) of host and virus abundance

All samples were analysed in bulk after the conclusion of the experiment. Samples were defrosted at room temperature and then analysed using flow cytometry following the protocols described in Section 2.2.1 and 2.2.2. The two EhV strains could not be distinguished by AFC in the combined EhV-86+EhV-207 cultures (as they gave the same green fluorescence signature).

4.2.6. DNA extraction from top phase and pellet samples

After product amplification was not detected during a test qPCR with a subset of samples from the "virus fight club" experiment it was decided to troubleshoot the reaction conditions. It was established that most likely PCR inhibitors in the raw

samples were the cause for the lack in qPCR amplification. Hence it was decided to perform a phenol-chloroform DNA extraction on all the samples from the experiment intended for qPCR (as described in Section 2.3.6).

Table 4.1. The virus "fight club" culture conditions A-G and their initial volume* at the beginning of the experiment (t0). Each culture condition was performed in triplicates (i.e. a total of 21 f/2 cultures containing *E. huxleyi* CCMP 2090 at a cellular density of 1.5×10^6 mL^{-1}). + indicates the presence of EhV-86, EhV-207, or host in a given culture, – indicates inactivated virus.

Culture	Volume of f/2 (mL)*	EhV-86	EhV-207	*E. huxleyi* host	Experimental culture condition
1	100	+		+	
2	100	+		+	A
3	100	+		+	
4	100		+	+	
5	100		+	+	B
6	100		+	+	
7	100	+	+	+	
8	100	+	+	+	C
9	100	+	+	+	
10	100	+			
11	100	+			D
12	100	+			
13	100		+		
14	100		+		E
15	100		+		
16	100	+	+		
17	100	+	+		F
18	100	+	+		
19	100	-	-	+	
20	100	-	-	+	G
21	100	-	-	+	

4.2.7. Real-time PCR (i.e. qPCR)

Real-time PCR assays on extracted DNA samples were carried out in optical-grade 96-well plates in an ABI PRISM®7000 Sequence Detection System (Applied Biosystems, UK) with the Qiagen Quantifast Sybr Green PCR kit (with a ready to use master mix). The calibration curve (or standards) for the qPCR to which each "fight club" DNA sample was compared to consisted of triplicates of the serial dilutions (10^-

^{1} to 10^{-10}) of amplified and gel extracted PCR products of ehv290 (for EhV-86) and EQVG00465 (for EhV-207) at initial DNA concentrations of 65.7 ng μL^{-1} and 57.4 ng μL^{-1} respectively (i.e. DNA copy number of 2.91 x 10^{11} and 1.51 x 10^{11} respectively). The calibration curve diluted samples were loaded each time at the same plate as the DNA samples from the "fight club" experiment in order to reduce bias occurring due to small shifts in fluorescence signal from one 96 well plate to another.

The reactions of both standards and samples consisted of 12.5 µL of Sybr Green master mix, 0.5 µL of primer qPCR(EhV-86)-F or qPCR(EhV-207)-F (at a final conc. of 0.2 µM), 0.5 µL of primer qPCR(EhV-86)-R or qPCR(EhV-207)-R (at a final conc. of 0.2 µM), 1 µL of template DNA (sample, standard, or NTC [no template control]) and 10.5 µL of RNAse free water (final volume of 25 µL). The thermal cycling conditions (on the ABI PRISM 7000 cycler) were as follows: an initial cycle of 95ºC for 10 min followed by 40 cycles at 95ºC for 30 sec and 60ºC for 30 sec.

The automated generation of the calibration curve by the ABI PRISM 7000 Sequence detection system allowed the logarithmic plotting of each standard concentration against the cycle number at which the detected fluorescence signal increased above the threshold value-CT. Then, the Sequence Detection System software calculated the target gene DNA copy number (or concentration) from the CT value obtained for each of the samples with unknown concentration. The calibration curve slope was used in order to determine the reaction efficiency (E) using the following equation: $E= -1+10^{(-1/slope)}$. For instance, if E equalled to 1 then this meant a 100% product doubling in each amplification cycle.

4.3. RESULTS

4.3.1. PCR primer specificity

PCR amplification of EhV-86, EhV-207 and EhV-86/EhV-207 lysates (A,B and C) revealed that the newly designed qPCR primers were indeed strain specific (Figure 4.3). When each set of primers was used in a sample from which the target strain was not present (EhV-86 or EhV-207), amplification was not detected; i.e. the amplification was specific to the set of primers used and only samples that had the target strain produced a PCR amplicon. This test experiment, in triplicate, also revealed that the amplified products of the EhV-86 and EhV-207 specific targets, correspond to the size that was predicted from the genomic investigation of these genes; i.e ~200 and ~353 bp respectively.

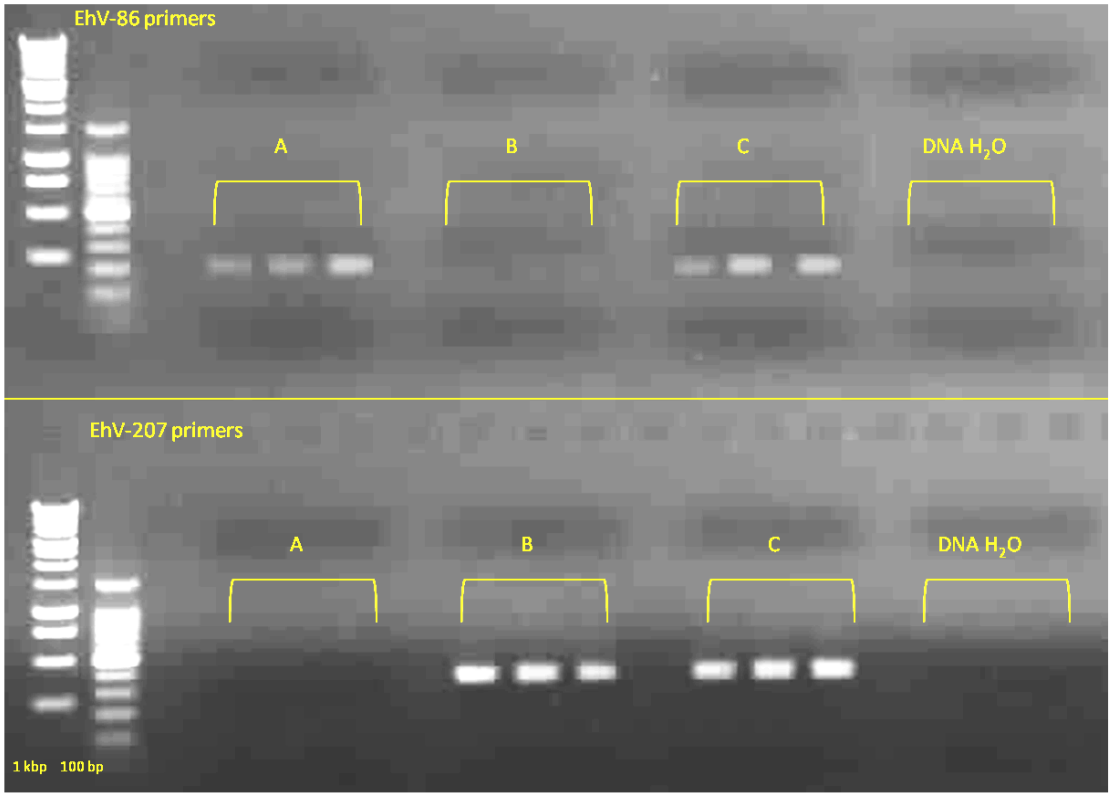


Fig. 4.3. Gel electrophoresis of PCR products conducted with primers specific to EhV-86 and EhV-207. The products are amplified regions from subsamples from lysates taken from three different culture conditions: EhV-86 infected host (A), EhV-207 infected host (B), and host infected simultaneously by both EhV-86 and EhV-207 (C). In the last two lanes to the right (top and bottom) are the control DNA free water samples.

4.3.2. Analytical Flow Cytometry (AFC) analysis of the virus "fight club"

4.3.2.1. *E. huxleyi* CCMP 2090: host abundance

The AFC approach was undertaken in order to record both host cellular abundances as virus propagation progressed and to monitor host demise over time. In general, there were no significant differences in the host cellular abundance between cultures infected by EhV-86 (condition A; 2.98×10^6 mL^{-1}; SD $\pm$ 1.41×10^5), EhV-207 (condition B; 2.56×10^6 mL^{-1}, SD $\pm$ 3.10×10^5) or combined EhV-86+EhV-207 (condition C; 2.41×10^6 mL^{-1}, SD $\pm$ 2.36×10^5) in the first two days post the initial infection (i.e. host abundance decrease was not observed and the host populations continued to grow) (Figure 4.4). However, by day 3, both the EhV-207 and EhV-86+EhV-207-infected cultures crashed dramatically and three days post-infection there was ~96 % loss of cells in these cultures (i.e. condition B; 6.10×10^4 mL^{-1}, SD $\pm$ 1.25×10^4, condition C; 1.35×10^5 mL^{-1}, SD $\pm$ 4.49×10^4). The crash was simultaneous between these two culture conditions as can be seen in Figure 4.4. In contrast during the same time period (first two days post infection), there was a 5.85% increase in *E. huxleyi* abundance in the EhV-86 infected cultures (i.e. condition A; 2.98×10^6 mL^{-1}, SD $\pm$ 1.41×10^5). It was only after the fourth day that the EhV-86 infected cultures showed a similar rate of host density reduction (i.e. 2.63×10^6 mL^{-1}, SD $\pm$ 1.65×10^5), by which point the cellular density of the EhV-207 and EhV-86+EhV-207 infected cultures were at an average low of 2.66×10^3 mL^{-1} (SD $\pm$ 1.03×10^3) and 1.64×10^3 mL^{-1} (SD $\pm$ 3.42×10^2) respectively (Figure 4.4). This was also evident from the *E. huxleyi* density plot from AFC analysis (Figure 4.5) in which the mean red fluorescence (proxy for chlorophyll and host abundance) 72 h post-infection was much lower (lower red fluorescence and side scatter) in the EhV-207 and EhV-86+EhV-207-infected cultures than in the EhV-86 infected ones.

By the end of the experiment (seven days post infection, t13), the majority of the *E. huxleyi* host population in the EhV-86 infected cultures was lysing and/or dying (as can be seen in Figure 4.5 and 4.6. from the decrease in the amount of cells with higher red fluorescence). However, there was still a small relatively healthy (i.e. higher red fluorescence) host population remaining as the average host abundance in these cultures was 2.77×10^5 mL^{-1} (SD ± 6.89×10^4) as opposed to: 2.52×10^2 (SD ± 3.36×10^1) and 1.18×10^3 (SD ± 1.24×10^3) cells mL^{-1} in the EhV-207 and EhV-86+EhV-207 infected cultures respectively (not seen on the plot due to the logarithmic scale of the graph). With regards to the control cultures that contained inactivated coccolithovirus virions (condition G). *E. huxleyi* abundance was still high by the end of the experiment and cellular density was 4.32×10^6 mL^{-1} (SD ± 3.04×10^5) (Figure 4.4).

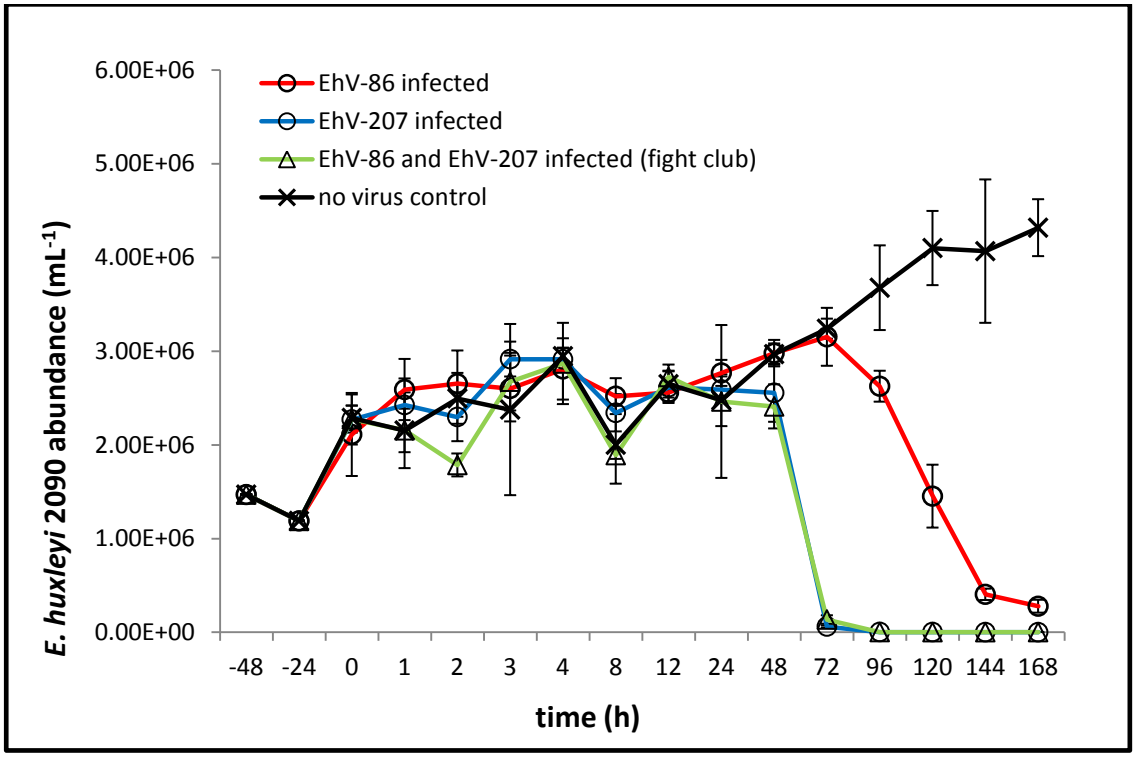


Fig. 4.4. *E. huxleyi* CCMP 2090 average cells mL^{-1} (of three replicates, error bars are ±standard deviations) following infection by EhV-86 (culture condition A- red line) EhV-207 (culture condition B- blue line), and combined EhV-86 and EhV-207 (culture condition C- green line): control cultures that contained inactivated viruses (black line). The first measurements of the cultures were taken two days before the addition of the virus stocks (i.e. t-2 or -48 h). Cell density was enumerated using AFC (Red fluorescence *vs* side scatter).

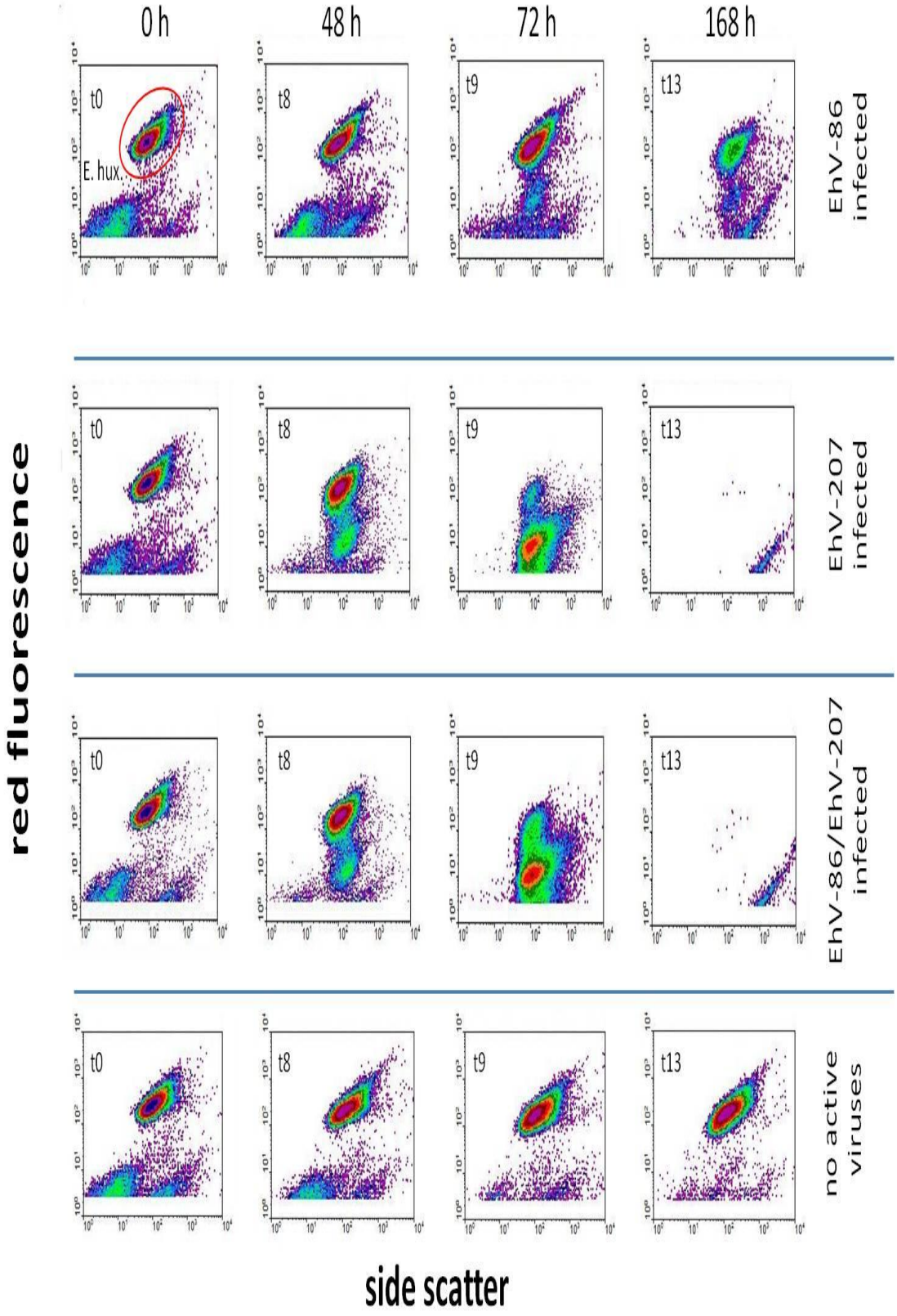


Fig. 4.5. AFC density plot analysis at t0, t8, t9 and t13 of *E. huxleyi* hosts (cellular red fluorescence [y axis] *vs* side scatter [x axis] ranging from 10^0 to 10^4). On the density plot, each dot or point represents an individual cell that has passed through the instrument. Pink/red hotspots indicate large numbers of cells. The subsamples for this analysis were from cultures infected by either EhV-86 ("solo"), EhV-207 ("solo") or combined EhV-86+EhV-207 ("fight club").

4.3.2.2. AFC analysis of free VLPs

As observed by AFC analysis, viral production rates from the infected *E. huxleyi* cells were similar in cultures infected by EhV-207 and EhV-86+EhV-207 throughout the experiment, with the exception of 1 h post-infection (t1), where the amount of free EhV-86 and EhV-207 virus like particles (VLPs) in the combined virus condition (C) was significantly lower than the amount of free EhV-86 or EhV-207 VLPs in the single treatments (A and B); i.e. 4.89×10^5 mL^{-1} (SD ± 9.87×10^4) as opposed to 6.57×10^5 (SD ± 5.87×10^4) and 7.75×10^5 mL^{-1} (SD ± 7.05×10^4) respectively ($t= 0.0164$, $P<0.05$) (Figure 4.6). Then, two hours post infection (t2), the average amount of free VLPs in all culture conditions decreased to an average of 2.91×10^5 mL^{-1} (SD ± 1.42×10^4). Thus by two hours post-infection almost half (45%) of the inoculated VLPs in all culture conditions were either attached to the host cell receptors or had penetrated into the cells for replication.

The first round of mass virion release from the infected cells in all culture conditions was 3 h post-infection (as seen in Figure 4.6.). At 8 h post-infection there were more free VLPs in the EhV-207 and EhV-86+EhV-207 infected cultures than in the EhV-86 infected ones; i.e. 8.73×10^5 (SD ± 2.54×10^4), 9.10×10^5 (SD ± 3.18×10^4), and 7.18×10^5 VLPs mL^{-1} (SD ± 5.18×10^4) respectively. The trend of lower and perhaps slower virion production/release from the EhV-86 infected cultures continued until the end of the experiment 7 days post-infection where the amount of free VLPs in these cultures (5.19×10^7, SD ± 3.88×10^6) was on average ~92% less than in the EhV-207 (6.58×10^8, SD ± 1.09×10^8) or EhV-86+EhV-207 (7.04×10^8, SD ± 8.55×10^7) infected cultures ($t=0.0003$, $P<0.05$) (Fig. 4.6). Interestingly, the maximum number of

VLPs in both EhV-207 and EhV-86/EhV-207 infected cultures was reached 48 h post infection (Figure 4.6 & 4.7).

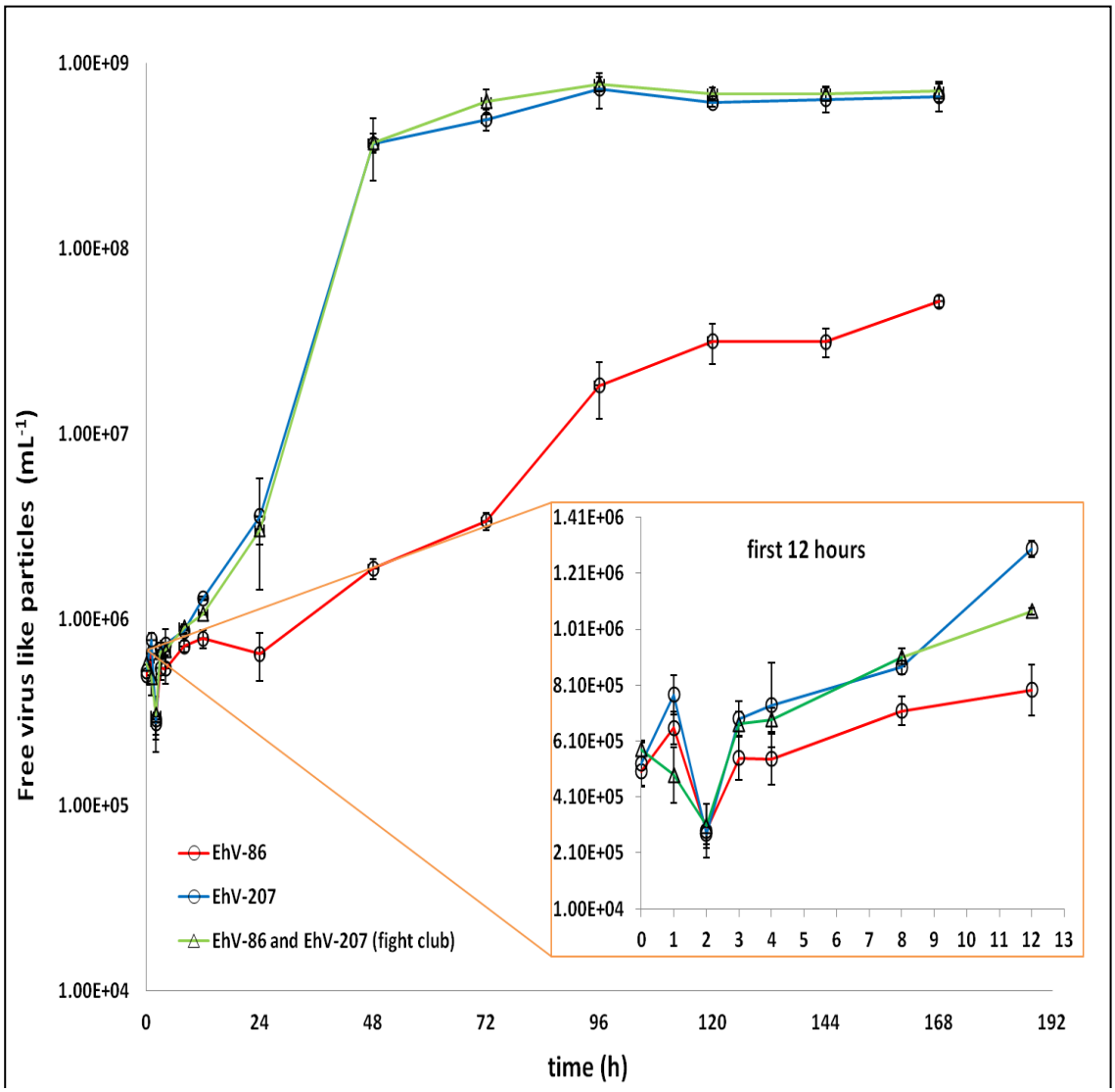


Fig. 4.6. Average density mL^{-1} (of three replicates) of EhV-86 (culture condition A- red line) EhV-207 (culture condition B- blue line) and combined EhV-86 and EhV-207 (culture condition C- green line) VLPs. The controls are not seen here as they were below the level of detection. The free VLPs density was enumerated using AFC (green fluorescence *vs* side scatter).

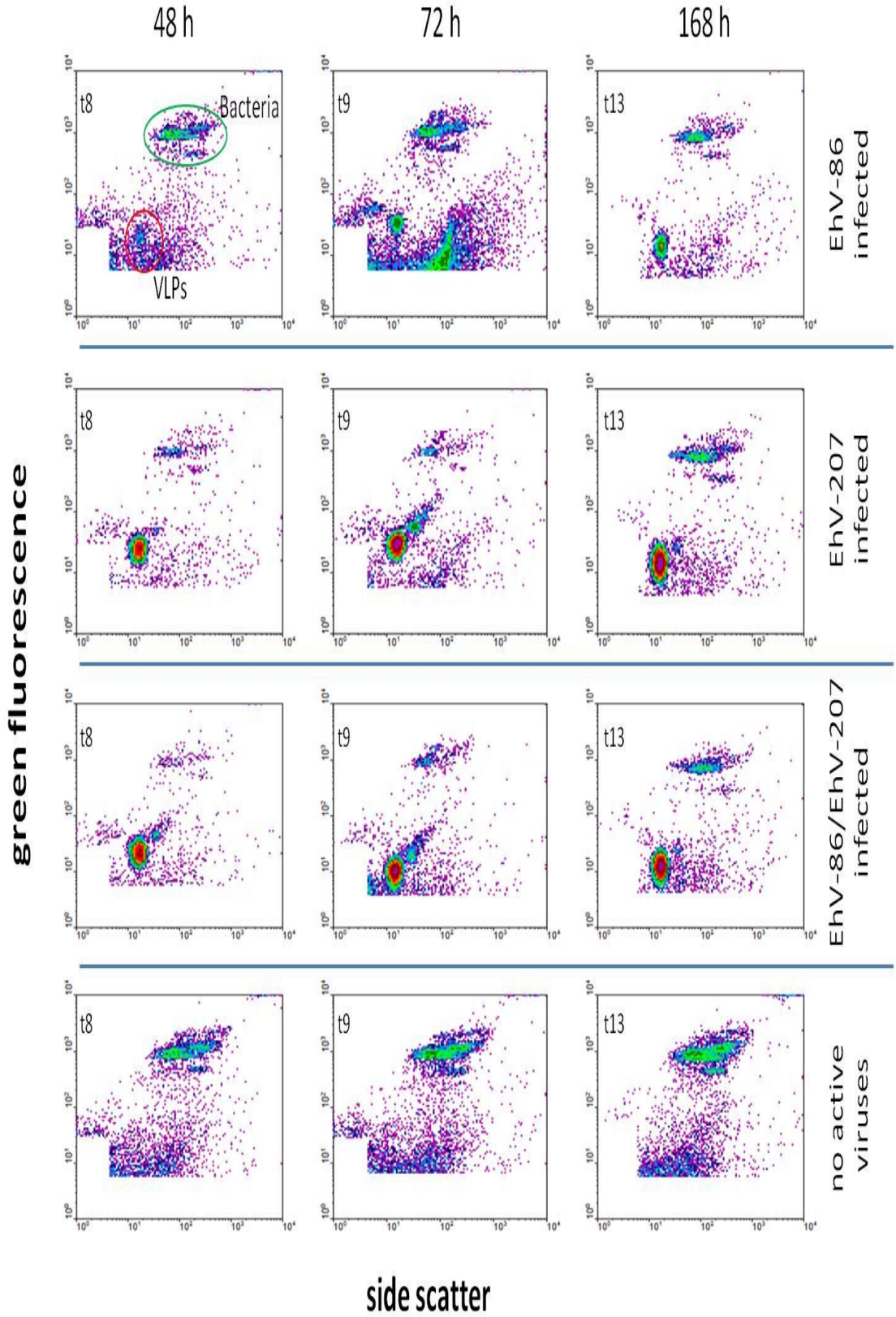


Fig. 4.7. AFC density plot analysis at t8, t9 and t13 of VLPs (green fluorescence [y axis] *vs* side scatter [x axis] ranging from 10^0 to 10^4). On the density plot, each dot or point represents an individual cell or virus that has passed through the instrument. Pink/red hotspots indicate large numbers of bacterial cells or coccolithoviruses. Subsamples for this analysis were from cultures infected by either EhV-86 ("solo"), EhV-207 ("solo") or combined EhV-86+EhV-207 ("fight club").

4.3.3. qPCR analysis of the virus "fight club"

4.3.3.1 Calibration curve efficiencies of the qPCR "fight club" sample analysis

The calibration curves (Figure 4.8) that were established in order to confirm the optimum conditions for the qPCR analysis indicated that under the described PCR conditions, the serial dilutions of the known concentrations of ehv290 (for EhV-86) and EQVG00465 (for EhV-207) PCR generated DNA was log-linear for both with a correlation coefficence (R^2) of 0.99. The efficiency ($E= -1+10^{(-1/slope)}$) of the reactions were 102.21 % and 88.70 % respectively. With the analysis of each plate of samples from the "fight club" experiment the established dilutions of the above comparative DNA concentrations were also re-analysed in order to allow accurate quantification of the "fight club" samples in relation to these known DNA concentrations.

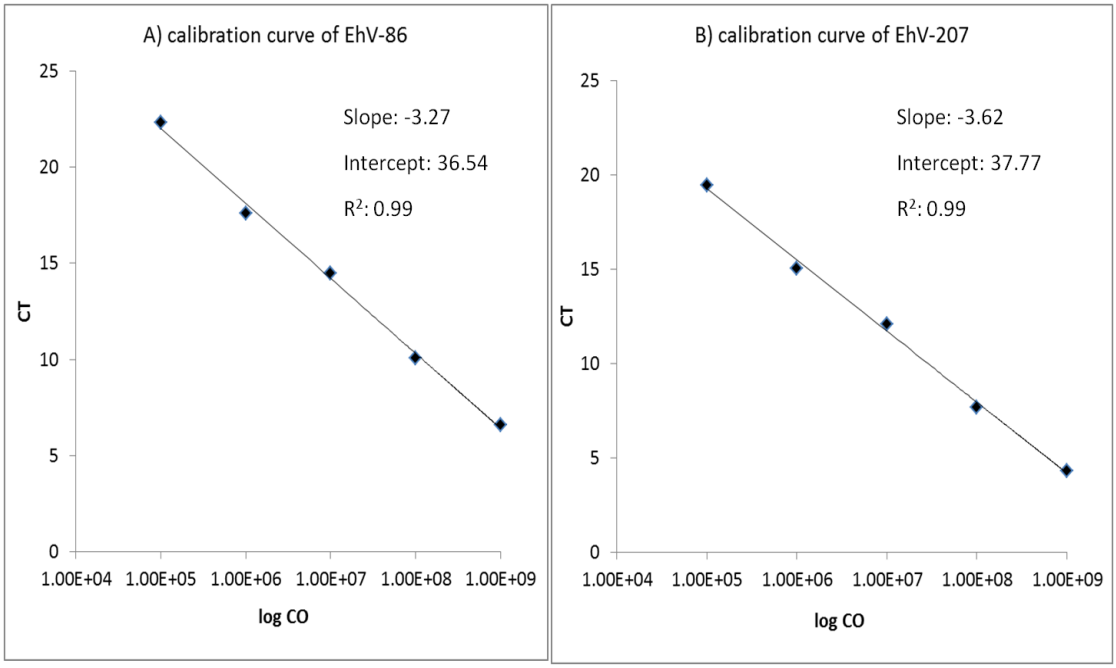


Fig. 4.8. Calibration curve for the qPCR amplification of known amounts of purified DNA of EhV-86 (ehv290) and EhV-207 (EQVG00465). CT= cycle number, (log CO= known concentrations of purified EhV-86 (A) and EhV-207 (B) DNA products.

4.3.3.2. Virus copy number of "fight club" versus "solo" cultures

A comparison of the amount of VLPs detected by AFC as opposed to those detected by the qPCR technique revealed that there was a 100-fold decrease in the amount of virus copy number (VCN) detected by the qPCR method, most likely due to the loss of viruses during the phenol-chloroform DNA extraction step (Table 4.2). This reduction in apparent virus abundance was consistent throughout the experiment and was thus taken into consideration when interpreting the qPCR results of the virus "fight club" experiment. In addition, the resolution of the qPCR was accurate only 12 h post infection at VCN higher than 10^3 mL^{-1}, hence the qPCR results of t0 – t5 were below the detection levels and are not shown here.

Table 4.2. Free, unattached, average virus like particles (VLPs) number and virus copy number (VCN) per mL, detected by analytical flow cytometry (AFC) and quantitative polymerase chain reaction (qPCR) respectively, 12 h and 168 h post-infection of cultures infected by either EhV-86 or EhV-207.

Analysis type / **time (h)**	**AFC**	**qPCR**	**AFC**	**qPCR**
	EhV-86 (solo free)	**EhV-86 (solo free)**	**EhV-207 (solo free)**	**EhV-207 (solo free)**
12	**7.92×10^5**	**1.40×10^3**	**1.30×10^6**	**2.14×10^4**
168	**5.19×10^7**	**4.51×10^5**	**6.58×10^8**	**6.71×10^6**

Regardless of whether the two viruses were co-infecting or solo-infecting a host population in culture, the two strains exhibited different infection dynamics, and the qPCR results revealed more than what was initially observed by the AFC analysis. Throughout the experiment, in the "solo" cultures EhV-207 exhibited faster infection dynamics than EhV-86 and replicated to produce new VLPs quicker; i.e. the combined VCN (free virus particles + virus particles that were either attached to the host cells or were inside the cells) was higher (Figure 4.9). In the "fight club" cultures, the presence of EhV-86 did not affect the infection dynamics of EhV-207, and the latter produced VLPs in a similar rate to the "solo" cultures that were infected only by EhV-

207. In contrast, EhV-86 was affected by the presence of EhV-207 in the "fight club" cultures and the amount of EhV-86 VCN at the end of the experiment in these cultures was 1000 fold less than in the "solo" cultures that had only EhV-86; i.e. 6×10^3 (SD ± 2.86×10^3) and 4×10^6 (SD ± 1.28×10^6) mL^{-1} respectively. Thus, EhV-207 was not only a faster strain with regards to its infection dynamics but was also a superior strain that out-competed its EhV-86 rival when they were both present in a host culture (at similar initial ratios).

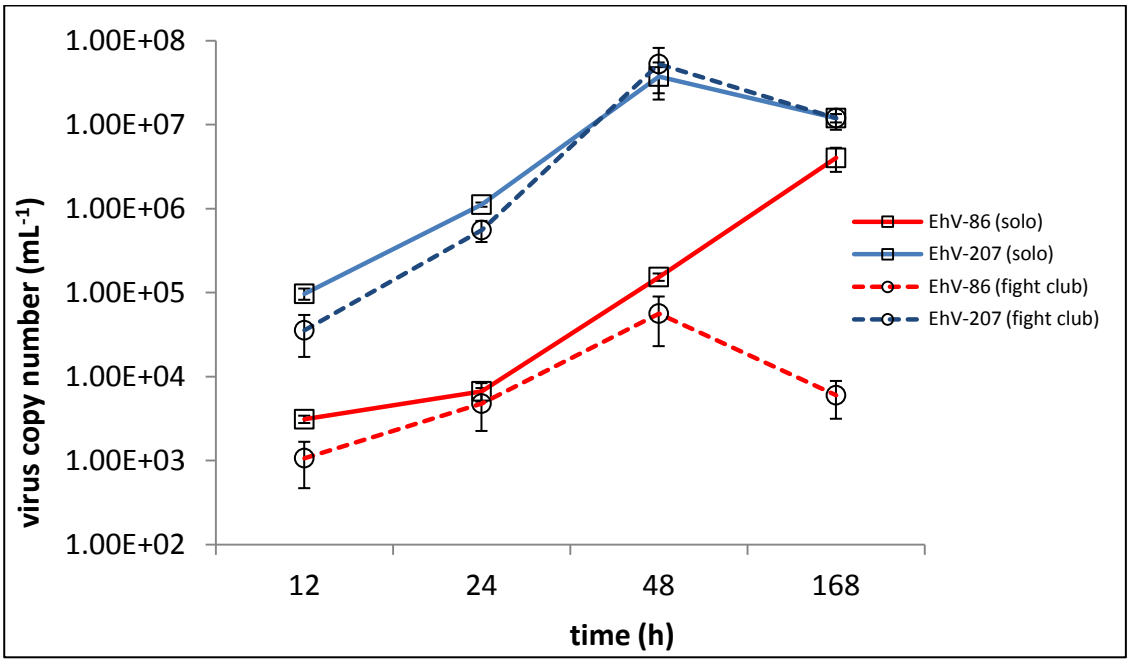


Fig. 4.9. Combined **(**free-floating and cell-associated) VCN averages (±SD) of subsamples from cultures in triplicates infected by either EhV-86 ("solo" cultures), EhV-207 ("solo" cultures) or combined EhV-86 and EhV-207 ("fight club" cultures), 12 h, 24 h, 48 h and 168 h post infection (i.e. t6, t7, t8, and t13 respectively); performed with qPCR strain specific primers for the discrimination of one coccolithovirus strain from the other.

4.3.3.3. Virus copy number of free and attached VLPs

In order to gain a deeper understanding of the interactions that have occurred during this experiment it was also necessary to understand the genome replication rates of each strain. Hence the qPCR analysis was also used to determine the abundance of EhV-86 and EhV-207 virus copy number (VCN) at any given time, by analysing those free in solution (either released/unadsorbed virions) and those associated with the

pelleted fraction, with the assumption that the VCN in the pelleted fraction represent viruses or synthesized viral genomes within the cells (infectious) or virus particles attached to the cell receptors (infecting). The presence of EhV-86 in the "fight club" cultures did not affect the amount of cell-associated or free EhV-207, as the VCN of both in each fraction in the "fight club" cultures was similar to the VCN of the cell-associated and free EhV-207 VLPs in the EhV-207 "solo" cultures (Figure 4.10). For instance, 48 h post infection the average cell-associated EhV-207 VCNs in the "fight club" and "solo" cultures were 5.00×10^7 (SD ± 2.83×10^7) and 4.47×10^7 (SD ± 1.52×10^7) respectively, and the average free EhV-207 VCNs in the "fight club" and "solo" cultures were 3.89×10^6 (SD ± 2.62×10^5) and 5.25×10^6 (SD ± 7.07×10^4) respectively. Interestingly, throughout the experiment (t6-t13), the amount of cell-associated VCN of EhV-207 in both "fight club" and "solo" culture conditions was higher than the amount of free EhV-207 VCN, with the exception of 168 h post infection (t13) where the amount of free and cell-associated EhV-207 VCN was on average similar: i.e. 6.90×10^6 (SD ± 2.66×10^5) and 5.59×10^6 (SD ± 1.01×10^6) respectively (Figure 4.11 B).

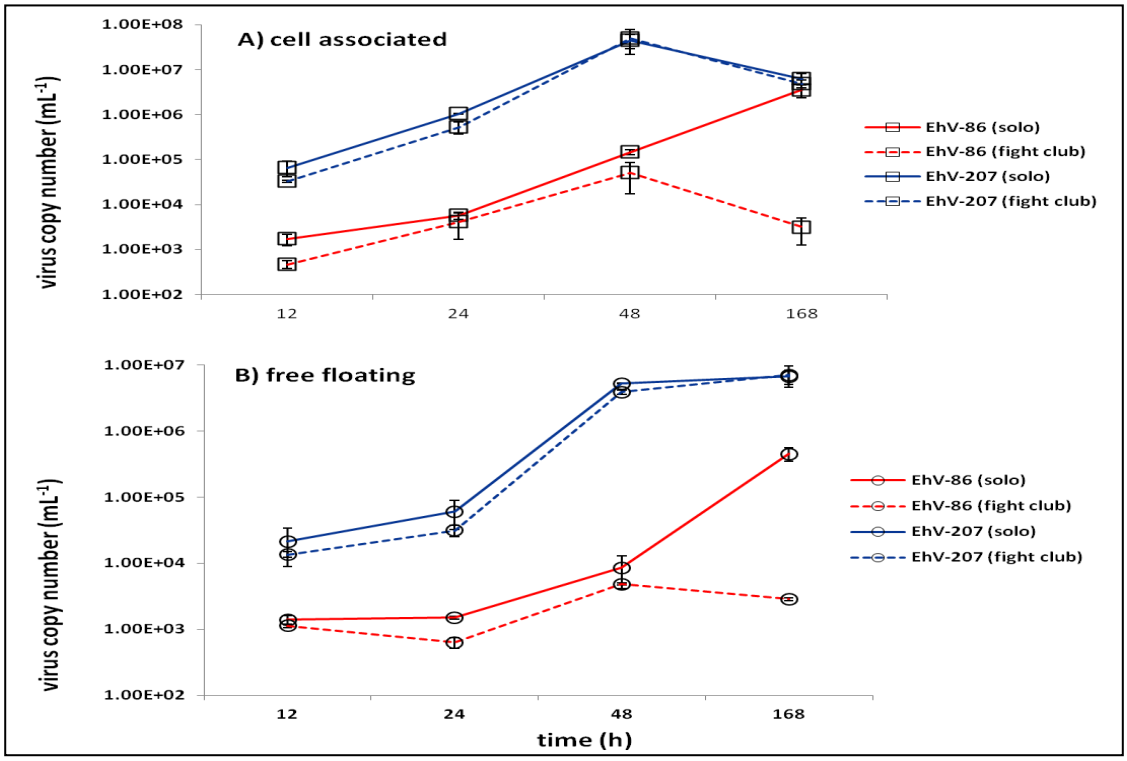


Fig. 4.10. Cell-associated (A) and free floating (B) VCN averages (±SD) of subsamples from cultures in triplicates infected by either EhV-86 ("solo" cultures), EhV-207 ("solo" cultures) or combined EhV-86 and EhV-207 ("fight club" cultures), 12 h, 24 h, 48 h and 168 h post infection (i.e. t6, t7, t8, and t13 respectively); performed with qPCR strain specific primers for the discrimination of one coccolithovirus strain from the other.

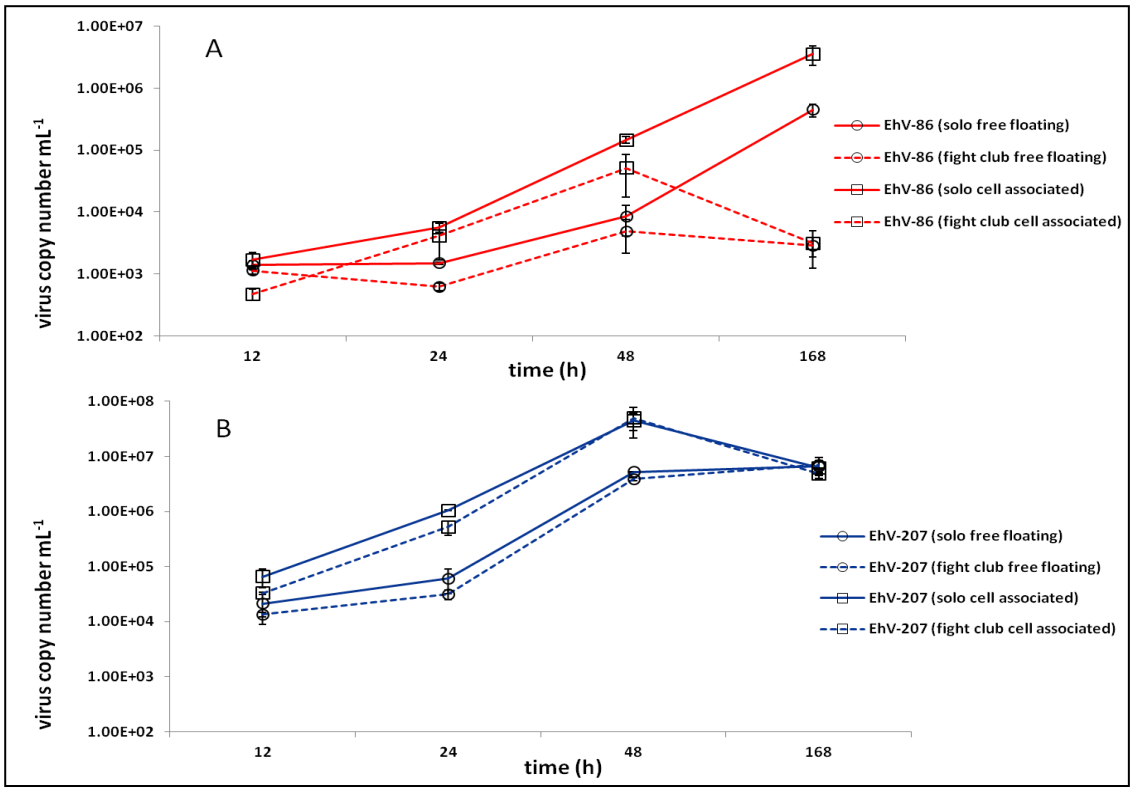


Fig. 4.11. EhV-86 (A) and EhV-207 (B) virus copy number averages (±SD) of subsamples from "solo" (full lines) and "fight club" (dashed lines) cultures in triplicates, 12 h, 24 h, 48 h and 168 h post infection (i.e. t6, t7, t8, and t13 respectively); performed with qPCR strain specific primers for the discrimination of one coccolithovirus strain from the other.

In contrast, the presence of EhV-207 in the "fight club" cultures affected the VCN of both cell-associated and free EhV-86 (Figure 4.10 & 4.11 A). For instance, at t6 and t8 (i.e. 12 h and 48 h post-infection) the VCNs of cell-associated EhV-86 in the "solo" cultures were 1.71×10^3 (SD ± 5×10^2) and 1.47×10^5 (SD ± 1.87×10^4) respectively and the VCNs of the cell-associated EhV-86 in the "fight club" cultures were 4.70×10^2 (SD ± 9.72×10^1) and 5.14×10^4 (SD ± 3.42×10^4) respectively. In simpler terms the VCN of the cell-associated EhV-86 in the "solo" cultures were ×3.6 and ×2.8 higher respectively than the EhV-86 cell-associated VCN in the "fight club" cultures at the same time points (i.e. 72 % and 65 % less cell-associated EhV-86 genomes in the "fight club" cultures respectively, Figure 4.10 A). Also, at these same time points (i.e. 12 h and 48 h post-infection), the VCNs of free EhV-86 in the "solo" cultures were 1.40×10^3 (SD ± 2.27×10^2) and 8.52×10^3 (SD ± 4.37×10^3) respectively, and the VCNs of the free EhV-86 in the "fight club" cultures were 1.14×10^3 (SD ± 1.52×10^2) and 4.87×10^3 (SD ± 2.71×10^3) respectively. In simpler terms, the VCN of the free EhV-86 in the "solo" cultures were × 1.2 and × 1.7 higher respectively than the EhV-86 free VCN in the "fight club" cultures at the same time points (i.e. 18 % and 42 % less free EhV-86 genomes in the "fight club" cultures respectively, Figure 4.10 A).

Although at the end of the experiment (i.e. 168 h post infection) the amount of cell-associated EhV-86 in the EhV-86 "solo" cultures was similar to the amount of cell-associated EhV-207 in the EhV-207 "solo" and EhV-207 "fight club" cultures (Figure 4.10 A), there was a dramatic decrease in the amount of cell-associated EhV-86 copy number in the "fight club" cultures, with an average low of 3.13×10^3 mL^{-1} (SD ± 1.88×10^3), more than × 1000 times less than the number of EhV-207 in these same

cultures (Figure 4. 10 B). Finally, the amount of free EhV-86 at the end of the experiment was lower than the amount of free EhV-207 in both the "solo" and "fight club" cultures (Figure 4.10 B); i.e. EhV-86 "solo"- 4.51×10^5 (SD $\pm$ 1.06×10^5), EhV-207 "solo"- 6.71×10^6 (SD $\pm$ 1.67×10^6), EhV-86 "fight club"- 2.87×10^3 (SD $\pm 9.78 \times 10^2$), and EhV-207 "fight club"- 7.09×10^6 (SD $\pm$ 2.54×10^6).

4.4. DISCUSSION

The design of strain specific genetic markers enabled the author to distinguish the virus production of these two virus strains in simultaneously infected *E. huxleyi* host cultures, determine the competitive interactions of these viruses over the host, and identify potential differences in their infection dynamics. The qPCR primers were designed after the full genome of EhV-86 was BLAST searched against the full genome of EhV-207. The result of this homology search (also validated by PCR) revealed that each primer was strain specific and not present in any of the other EhV strains in the IMG/ER database.

To date, differences in infection dynamics of different coccolithovirus strains have not been observed and this is the first time where these dynamics are investigated in detail in controlled laboratory conditions. In addition, this is the first time that the competitive interactions between strains of marine viruses; an issue usually ignored by marine virologists, has been studied. When infecting in combination, EhV-207 was not affected by the presence of EhV-86 whereas EhV-86 was, and a significant reduction in free and attached EhV-86 VLPs was seen two days post the initial infection. This means that when co-infecting, the EhV-86 strain was “beaten” and its persistence in this experimental setup was under threat. The significance of the results here are potentially crucial and even fundamental in the understanding of how viruses interact with their hosts and each other, and in establishing the outcome of an infection in the environment and its significance for biogeochemical cycling.

4.4.1. The losers and winners of the virus "fight club" – a numbers game

Throughout the experiment EhV-207 was superior to EhV-86 in that it was a faster replicating virus that killed the host population much quicker, regardless of whether it was co-infecting with EhV-86 in the same experimental environment or not. Hence even without considering the interactions within the "fight club" cultures, EhV-207 could potentially be deemed a "superior" strain to EhV-86, at least with regards to their infection of *E. huxleyi* CCMP 2090 under these laboratory conditions. The dramatic increase in EhV-207 VLPs 24 h post infection in the "solo" and "fight club" cultures and the subsequent decrease in host population density thereafter indicates that at this point there were more EhV-207 than EhV-86 virions available for further infection. Two possible explanations support and can provide clarification to this. The first is that for some currently unknown reason either EhV-86 remained latent within the host cells for longer than EhV-207 or/and that EhV-207 unpacked, replicated, assembled, and re-packed its genome into new virions within the host cells much quicker than its EhV-86 rival (presumably through the altered expression of shared genes or the expression of unique genes), and subsequently released these quicker from the infected cells.

This can explain why 8 h post-infection, there were already more free EhV-207 VLPs than EhV-86 VLPs in the "solo" cultures. If this is to be extrapolated onto the "fight club" cultures then this would mean that 8 h post-infection EhV-207 already had a numerical advantage over EhV-86 (due to their principal different infection strategies and infection dynamics characteristics). Hence past this point, EhV-207 was most likely responsible for the demise of the majority of the host cells within these "fight club" cultures, leaving EhV-86 less hosts for its own replication. The second

explanation, one where the final outcome was the same, is that the amount of VLPs produced per infected host cell was much higher when hosts were infected by EhV-207 than EhV-86.

This might have been either due to different levels of resistance within the host population to these two virus strains or due to some fundamentally different unknown genetic characteristic of the two viruses. With regards to the later, EhV-207 has an extra tRNA not found in the genome of EhV-86 and also 49 genes (of which 47 have no assigned or predicted function) that have no homologs with EhV-86. Among these are two genes predicted to encode for glycosyl-transferases. Although rare in viruses, glycosyl-transferase encoding genes have been previously reported in bacteriophages, poxviruses, herpesviruses and baculoviruses (Markine-Goriaynoff et al., 2004). In some bacteriophages, they have the ability to modify the virus DNA in order to protect it from host restriction endonucleases, and in *Chlorella* viruses such as PBCV-1 they have been implicated in the synthesis of glycan components of the virus major capsid protein (Zhang et al., 2007). Hence the presence of these genes could be beneficial in EhV-207 and aid in its faster genomic assembly and capsid construction prior to release from the infected cells. Regardless of the mechanisms involved, EhV-86 appeared to be a poor competitor to EhV-207 and lost the battle over infection and replication when co-infecting with EhV-207, evident by the much lower number of free and attached EhV-86 VLPs in the "fight club" cultures (as opposed to the EhV-86 "solo" ones).

Another aspect that is worth considering during the analysis of these complicated interactions is that most likely there have been occasions during which a single host

cell was simultaneously infected by both virus strains. In this case, an internal intra-cellular battle over the host cellular metabolic machinery would have occurred between the two viruses. Such a scenario could explain the lower amount of total free VLPs in the "fight club" cultures in comparison to the free EhV-207 "solo" ones 12 hrs post infection (AFC data), and the higher total amount of free VLPs in these cultures in comparison to the EhV-86 "solo" ones (AFC data). Evidently, this might have been the tipping point at which EhV-86 lost the battle to EhV-207. The qPCR results showed that at this time point indeed there were less cell-associated or attached EhV-86 VLPs in the "fight club" cultures than in the EhV-86 "solo" ones.

Such a scenario although theoretically possible is maybe less likely due to the mutual exclusion theory proposed by Luria and Delbruck (1943) in which it was proposed that a virus particle that infects first a host cell alters it to the extent of which a second infection by another virus is unlikely. This was indeed shown to be the case in the co-infection of *Chlorella* by two closely related viruses PBCV-1 and NY-2A (Greiner et al., 2009). In this previous study, infection by the PBCV-1 virus depolarised the host cell membrane to exclude further infections by NY-2A. However the genomic comparison in Chapter 3 suggests that genes have been indeed transferred between closely related virus strains, possibly a result of co-infection and retrovirus involvement. Hence it is not currently known whether mutual exclusion mechanisms such as the ones observed in *Chlorella* exist in coccolithoviruses, but it is worth considering and investigating further.

Finally, comparing the virus copy number (VCN) of free to the VCN of attached EhVs revealed an aspect never observed before in the study of coccolithovirus

replication within the host cell. The fact that at each time point the cell-associated VCN was higher than the free and detached VCN suggests that not all the newly synthesized genomes were packed into virions and released from the cells. This indicates that there is a currently unknown factor that limits the amount of virus progeny produced by the two viruses per an infected host cell and that there is an unknown mechanism that determines the final amount of new viruses produced. The limiting factor could be that there is simply not enough "raw" material in the cells to produce a large quantity of capsids that will encapsulate all the newly produced genomes. Alternatively it could be that there is a limited amount of virions that can be encapsulated by the disintegrating host membrane and that there is an inadequate amount of lipid rafts for budding from the host cell. Either way this is an important observation that might explain to a certain extent the observed differences between free and attached virus copy number. If EhV-207 is better at utilising the raw material and produces more lipid rafts within the infected host, this will provide it with an advantage over EhV-86.

4.4.2. The ecological significance of the virus "fight club"

To date, interactions within algal blooms have been seen as two dimensional "battles" between algal and/or bacterial hosts *vs* their viruses (or versus their zooplankton predators). However, a three dimensional "battle" of hosts *vs* viruses *vs* other viruses is something that needs to be taken into account when trying to understand current microbial interactions and the role of marine microorganisms as bio-geochemical catalysts.

Although only observed here in a laboratory experiment, the results of this study can be extrapolated to natural systems, where indeed there is mounting evidence that such processes occur. For instance, during coccolithophore blooms in the North Sea and a Norwegian fjord, there was more than one *E. huxleyi* genotype at any given time and at least an order of magnitude more coccolithovirus (or EhV) genotypes (Wilson et al., 2002b; Martinez et al., 2007, 2012). In a recent mesocosm experiment in the Norwegian fjord (Sorensen et al., 2009), it was observed that the number of distinct EhV genotypes decreased with the propagation of an *E. huxleyi* infection; i.e. early on during the development of the bloom the EhV community was more diverse and there were more distinct EhV genotypes (detected by DGGE) then towards the end of the bloom (Figure 4.12). Hence the observations in this Chapter together with the field studies strongly suggest that coccolithovirus competition is fierce. Moreover, if the phenotypic characteristics of coccolithoviruses vary (as can be seen from the "fight club" experiment) and the outcome of an infection is influenced by the competitive interactions between different viral strains, then the propagation of an infection in a natural environment (and as a consequence the demise of the host population), will influence the rates of carbon and nutrient cycling (Wilhelm and Suttle, 1999), the export of carbon into the deep ocean (Thingstad and Lignell, 1997; Brussaard et al., 2008, Evans et al., 2009) and the potential release of climatically active compounds such as DMS and DMSP from the dying host cells (Evans et al., 2007). This finding is crucial to biogeochemical models, which are currently based on the incorporation of data of only one host and one virus, as different strains (as seen from this study) will affect phytoplankton populations at different rates and therefore the rates of biogeochemical cycling will also be influenced.

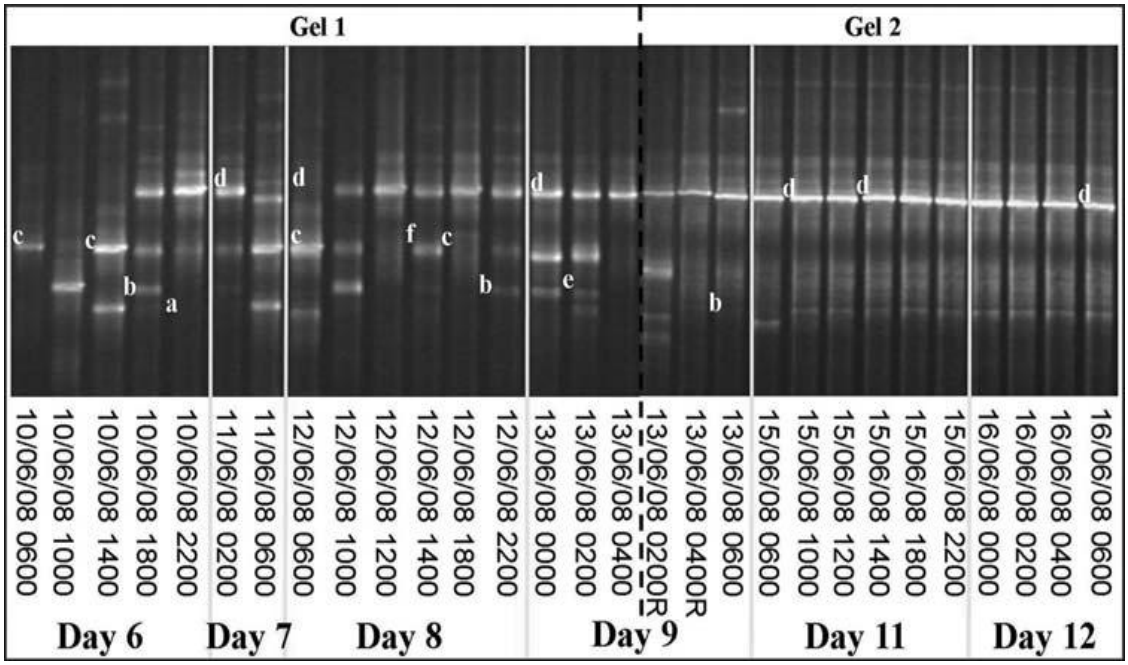


Fig. 4.12. EhV diversity in a mesocosm study in a Norwegian fjord shown by DGGE analysis of the coccolithovirus encoded MCP gene, adapted from Sorensen et al. (2009). The letters indicate distinct MCP bands confirmed by sequencing.

In a rapidly changing marine environment, whether due to anthropogenically induced or naturally occurring physicochemical stress factors such as nutrient limitation, variations in salinity levels, ocean acidification (i.e. carbonate chemistry alternations), and increased sea surface temperature, it is important to understand the phenotypic diversity changes that will occur within a microbial community and how these changes will affect globally important processes such as biogeochemical cycling and re-cycling. In a theoretical selection pressure scenario imposed by any of the physicochemical stress factors mentioned above, some coccolithophore genotypes in a given niche will survive and adapt to the changing conditions better than others. In turn, coccolithoviruses that display the most efficient infection and replication phenotypes (in this laboratory based instance, EhV-207 like strains) that correspond to the most dominant host species or strain, will efficiently utilize the host biochemical

machinery and propagate faster than coccolithoviruses that display slower infection dynamics (for instance EhV-86 like strains).

If there are only two infection dynamics phenotypes similar to the ones displayed by EhV-207 and EhV-86 (although in the natural environment there would be many more different EhV genotypes displaying distinct phenotypes not tested in this study), and the abundance of coccolithoviruses that possess these characteristics are in equal ratios initially, then the coccolithoviruses that have the EhV-86 infection dynamic-like phenotype (i.e. the losers of the "fight club" experiment) will be outcompeted (in this round of infection) by those that have the EhV-207 infection dynamic-like phenotype (i.e. the winners of the virus "fight club"). The ecological significance under the described scenario would be a localised short-time rapid increase in the rate of carbon export and recycling of nutrients, and the decrease in the total amount of nutrients and carbon recycled, due to the faster infection dynamics displayed by the temporarily dominant EhV-207 like strains. Alternatively, if EhV-86 like genotypes were to be hypothetically the "winners" of these competitive interactions, then this would have meant that the slower infection kinetics displayed by their infection–dynamics phenotypes will allow the host population to reach higher densities, the result of which would be also an increase in the total amount of carbon export and nutrient recycling but on a longer time scale. Hence, the outcome of a coccolithovirus fight club will influence the timing of host-population crash, and as a consequence the timing and amount of carbon and nutrients recirculated into the deeper ocean and within the microbial food web.

4.4.3. The evolutionary significance of the virus "fight club"

The evolutionary significance of virus competition is the fuelling of the co-evolutionary arms race between the host and its virus. During the described competition scenario, if the two strains were the only ones present in a given environmental niche, then an increase in the fitness of EhV-207 would have resulted in the decrease in fitness of EhV-86 to the point of its extinction or near extinction, meaning that eventually EhV-86 like genotypes (and as a consequence,phenotypes) would have disappeared from this niche and the local environment would have been dominated by coccolithovirus strains with EhV-207 like phenotypes. Viruses possessing these phenotypes, will infect the most dominant coccolithophore species and/or strains in consistence with the "killing the winner" hypothesis (Thingstad and Lignel, 1997), essentially transforming the host "winners" to "losers" with time. Then, the new host "winners" will most likely be a sub-population that is resistant to EhV-207 like strains but possibly more sensitive to EhV-86 like strains (or other similarly low abundance genotypes that have optimal infection strategies for these new host "winners"). This fits within the virus-host stable co-existence theory in which was hypothesised that the indeed it is to some extent the phenotypic plasticity of the algal hosts and their ability to recover post virus infection that makes the co-existence of these hosts and their viruses possible both on short and also on evolutionary time scales (Thyrhaug et al. 2003).

The emergence of novel viruses with niche specific characteristics for infection, are to a large extent, a result of these competitive interactions. For instance, in plants, the occurrence of more than one RNA virus and their environmental association with their host is common (Roossinck, 2005). If these viruses are similar to one another then they will be in direct competition, whilst if they are not similar then they will not be.

Indeed it was shown that plant RNA virus evolution occurs due to "survival of the fittest" scenario, during which closely related viruses increased the positive selection of some of these viruses over others (Roossinck and Palukaitis, 1995). Hence it is right to conclude that the competitive interactions between the closely related coccolithoviruses EhV-86 and EhV-207 will not only drive the fitness and evolution of their hosts but also their own.

Finally, the competitive interactions displayed here by coccolithoviruses raise exciting questions with regards to the ability of these different strains to evolve strategies for the utilisation of new resources; i.e. the infection of a new host. Recently it was shown that resource competition between bacteriophages (i.e. competition over host availability for infection and replication) promoted the evolution of novel bacteriophage phenotypes with the ability to utilise new hosts, suggesting that this sort of competition was essential for driving the evolution of host range expansion (Bono et al., 2013). Although coccolithoviruses are fundamentally different from bacteriophages with regards to their rate of mutation (much slower), a similar resource competition over a larger time scale may explain the large amount of susceptible infectious hosts to some coccolithovirus strains (personal communications with Dr. Worthy from Rothamstead Research in the UK, and personal observations). It may be that indeed this is what drives the transformation and emergence of generalist coccolithovirus morphs, capable of infecting an array of new hosts.

4.5. CONCLUSIONS

The study of the intense competitive interactions within the virus “fight club” revealed a complex picture. Whether a particular coccolithovirus strain (and all its associated genomic, proteomic, and metabolomic characteristics) will proliferate in the environment will most likely be determined by its ability to co-evolve with one or more host genotypes and outcompete other viruses less fit in a particular environment or during a particular situation. One can only imagine the complexity of the interactions described here in naturally occurring blooms where at any given time there would be many different coccolithophore genotypes (and subsequently phenotypes) and an even larger number of distinct coccolithoviruses. Many will display distinct infection dynamics beyond the scope and results of this chapter. Further large scale characterisation of coccolithovirus genotypes and their observed phenotypic characteristics during infection, and the analysis of their phylogenetic and functional biodiversity (as seen in Chapter 5) will shed a new light in the evolution of these “clever” molecular parasites and their place in microbial oceanography and biogeochemical cycling of the oceans.

5. ENVIRONMENTAL COCCOLITHOVIRUS BIODIVERSITY AND ITS POSSIBLE IMPLICATIONS ON NICHE ASSOCIATION AND SPHINGOLIPID METABOLISM

5.1. INTRODUCTION

So far coccolithoviruses have been studied in mesocosm systems (Martinez et al., 2007), natural open ocean blooms (Wilson et al., 2002a; Rowe et al., 2011), and in laboratory based experiments (Vardi et al., 2006; Pagarete et al., 2001, 2009; Frada et al., 2008; Bidle and Kwityn, 2012). With the recent sequencing of laboratory isolates, glimpses into the diversity of these viruses at the genetic level have been observed (Wilson et al., 2005; Allen et al., 2006b; Nissimov et al., 2011a, 2011b, 2012a, 2012b; Pagarete et al., 2012; and Chapter 3). Genomic analysis of coccolithoviruses isolated from different geographical locations has shown that they differ at a number of genomic loci; however the functional relevance of this has yet to be determined experimentally and an insight of the differences in infection dynamics of different coccolithovirus strains was provided in Chapter 4.

Problems have arisen when extrapolating the diversity characterised in the laboratory to naturally occurring virus communities. Viral isolates studied in the laboratory often represent the most abundant virus strains present at the time of isolation (often dependant on the most abundant host at that time), and are biased towards isolates capable of infecting established laboratory strains of *Emiliania huxleyi* and may or may not be environmentally relevant. Indeed, during natural conditions in the oceans, many different viral strains compete with each other for infection and replication (also seen in laboratory studies, see Chapter 4). Some will be more successful than others. The 'winner' is determined by a plethora of transient environmental conditions,

creating an ever-changing landscape to adapt and evolve to. It is this intense selection pressure that contributes to the diverse pool of genes observed in the handful of 'model' strains characterised to date. However, reliance on a limited number of strains to infer ecological functional relevance of a group often ignores the diversity and variation found in the natural environment.

Traditionally, the DNA polymerase gene (DNA pol) is used to study the diversity and phylogeny of phycodnaviruses (algal viruses) (Chen et al., 1996). In recent years the gene encoding the major capsid protein (MCP) has also been used as an alternative marker, capable of distinguishing phylogenetic differences on a strain level. Several research cruises have used these markers to observe the diversity of coccolithoviruses in natural blooms. The first study looked at the temporal succession of *Emiliania huxleyi* and their viruses during the propagation of a natural bloom in the North Sea in 1999 (Martinez et al., 2012), whilst another cruise in the North Atlantic between Iceland and the UK in 2005 focused mainly on the distribution of coccolithoviruses, their location specific distinctions and their clustering with the use of the MCP marker gene (Rowe et al., 2011).

Despite these efforts, there are many questions that still remain unanswered. For example, it is not currently known how the viruses persist during non-bloom periods. With the exception of coccolithovirus sequences extracted from Black Sea sediments (Coolen 2011), the diversity of the coccolithoviruses in non-bloom conditions is poorly understood. Given the harsh conditions to which viruses are exposed to in their natural environment (i.e. UV degradation, attachment to non host organic particles and sinking), it is somewhat surprising that infection, and the resulting bloom termination,

occurs regularly and reliably on a yearly basis. Yet, perhaps, the most important questions left unanswered (in this and the majority of virus systems under laboratory study) is what is the functional relevance of the observed biodiversity, and how does it impact on the ecology of the virus community and their interaction with their hosts?

In this Chapter the aim is to investigate both the biogeographic and temporal distribution of coccolithoviruses and their diversity with the established MCP marker, whilst also targeting a previously discussed gene whose protein is of known metabolic function during infection, serine palmitoyltransferase (SPT). SPT is the first and rate limiting enzyme in the *de novo* sphingolipid biosynthesis pathway and is encoded on the viral genome (Wilson et al., 2005; Han et al., 2006; Monier et al., 2009). It has been implicated in the formation of lipid rafts and virus release during infection (Pagarete et al., 2009), and is even considered to be involved in the mass termination of coccolithophore blooms via the propagation of programmed cell death (PCD) of its host (Vardi et al., 2009; Bidle and Vardi, 2011; Vardi et al., 2012). SPT gene expression has been observed during infection of *E. huxleyi* under both laboratory and natural conditions, and the enzyme's activity fully characterised (Allen et al., 2006c; Han et al., 2006; Pagarete et al., 2009).

Here, the two genes (encoding MCP and SPT) are used as markers for phylogeny and functionality in a study attempting to assess both spatial and temporal variability, using an archive of DNA samples collected during a cruise in the Atlantic in 2010, a coccolithophore cruise in the North Sea in 1999, and samples collected weekly during a seven year period from the Western Channel Observatory in the English Channel near Plymouth, UK. By obtaining samples from a variety of locations and time points,

and using the two marker genes, the current understanding of the classification of these viruses and their distribution is improved, and an insight into their functional biodiversity and ecological relevance is revealed.

5.2. MATERIALS AND METHODS

5.2.1. Atlantic Meridional Transect cruise (AMT)

The Atlantic Meridional Transect is a research program that has been running since 1995 along a 13,500 km transect from the UK to the southern parts of the Atlantic Ocean on board various UK funded research vessels (Robinson et al., 2009). The aims of the cruise tracks undertaken over the years are to enable biogeochemical measurements in the South and North Atlantic gyres, as well as gain understanding of the nature of ecological variability of phytoplankton and bacteria, their role in carbon cycling, and the exchange of active gases between the surface of the ocean and the atmosphere in continental shelf regions near Northern Europe, Southern Africa and Southern Latin America. During these transects core measurements such as salinity, pH, temperature, nutrient composition, chlorophyll levels, DOM and POM measurements are made with an attempt to elucidate the various biotic and abiotic changes that occur in this large oceanic basin over time. To date, the AMT program has contributed more than 200 peer reviewed articles and many graduate students in the realm of marine sciences have had the opportunity to obtain data during these cruises and incorporate their research with the core measurements taken in individual or collaborative studies. Hence I participated in the annual AMT cruise (i.e. AMT-20) from October to November 2010 and obtained samples that increased the spatial scale of the coccolithovirus diversity aspect of my PhD research subject.

5.2.1.1. AMT-20 sample collection and stations

Seawater was collected twice a day (before dawn and at solar noon) through a Conductivity Temperature and Density instrument (CTD) at stations along the Atlantic Meridional Transect-20 (AMT-20) cruise track (www.amt-uk.org) (Figure

5.1), between Southampton, UK and Chile, over a 6 week period from 12th October to 25th November 2010 (AMT-20 cruise: www.amt-uk.org/cruises/amt20). In total, 325 DNA samples were taken from 65 stations from five depths (Figure 5.2), each depth corresponding to 97%, 55%, 33%, 14% and 1% light penetration. 10 L of seawater from each sampling depth were filtered through a 0.2 µm Millipore nitrocellulose membrane filter (47 mm), the filters snap-frozen in liquid nitrogen and then stored at -80°C for the extraction of DNA and the molecular analysis of coccolithovirus diversity. Two further 1 mL sub-samples from each depth were fixed in 0.5% glutaraldehyde for Analytical Flow Cytometry (AFC) of coccolithophores and coccolithoviruses.

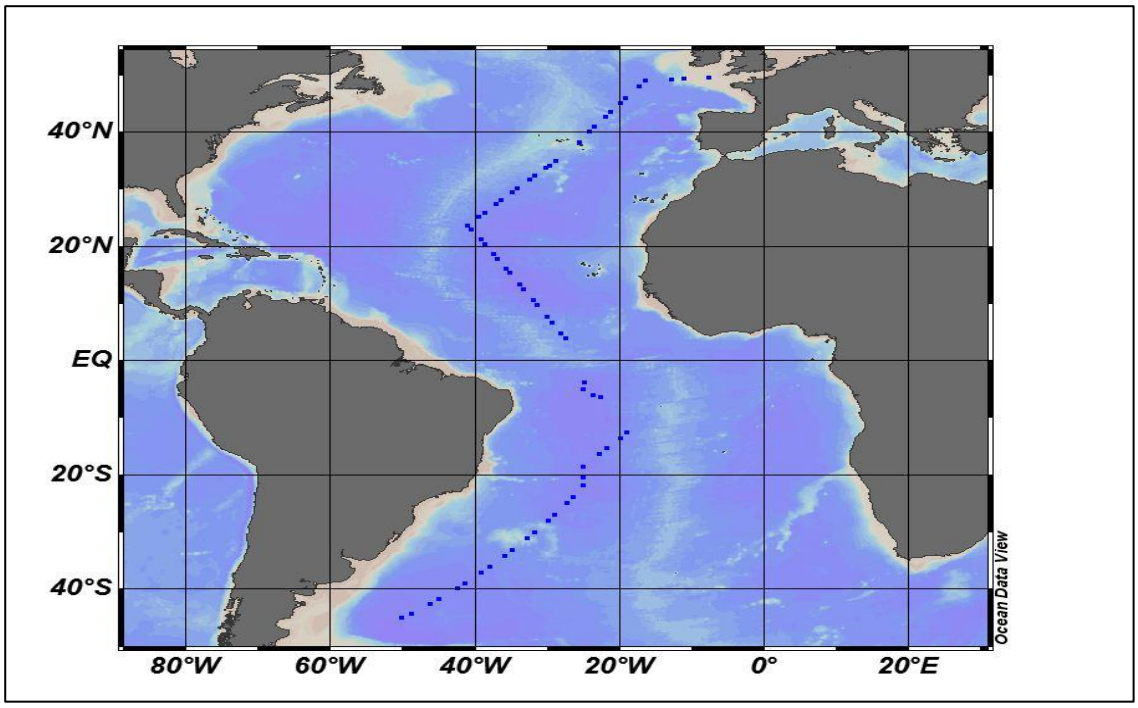


Fig. 5.1. The cruise track of the AMT-20 on board the RRS James Cook (Oct-Nov 2010) with the 65 CTD stations sampled marked as blue dots. Several gaps in the sampling exist due to occasional malfunction of the CTD sampling rosette and limiting weather conditions.

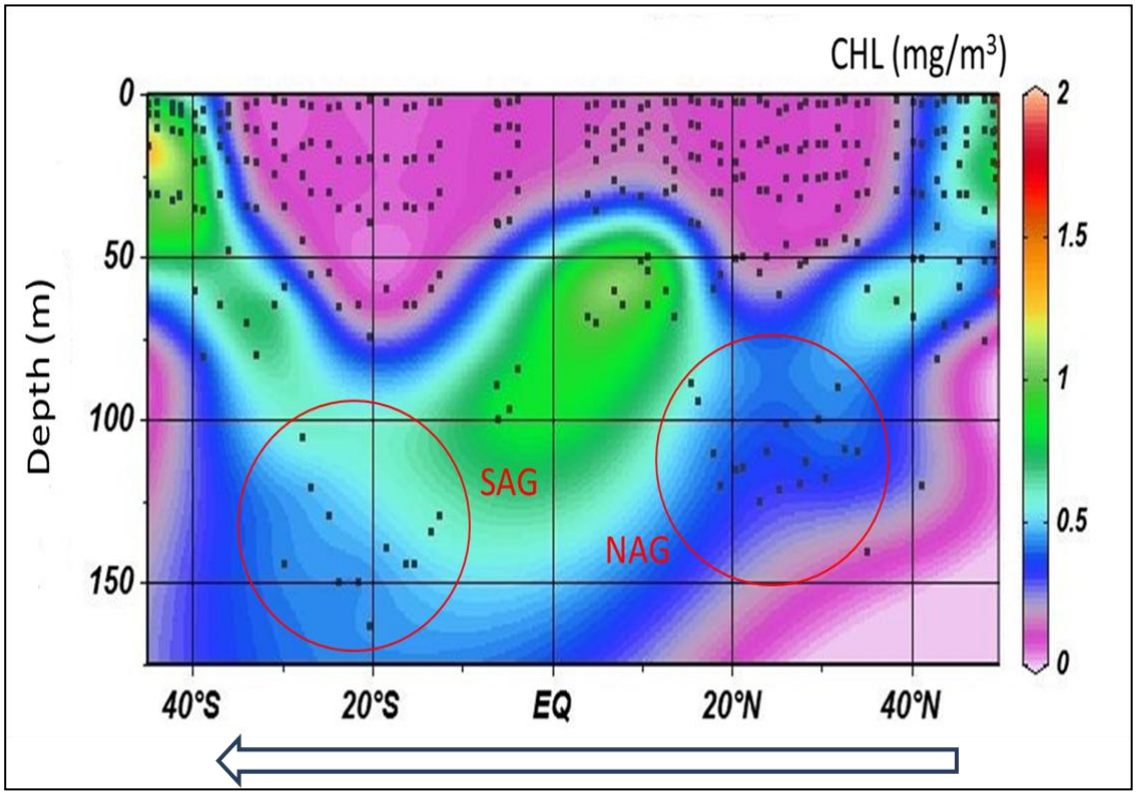


Fig. 5.2. Vertical profiles of 65 sampling stations over a crude chlorophyll (CHL) map of the annual Atlantic Meridional Transect (AMT-20) cruise. CTD profiles at each station were obtained and the Deep Chlorophyll Maximum (DCM) region, which often corresponded to 1% light penetration levels was also the deepest sample obtained for this study. The red circles indicate the North Atlantic Gyre (NAG) and the South Atlantic Gyre (SAG) and the blue arrow is the direction of the cruise. The raw oceanographic data was obtained during the cruise and analysed by the BODC (British Oceanographic Data Centre) before it was incorporated into this study.

5.2.2. The Western Channel Observatory (WCO) time series

Marine biological and oceanographic time series stations have existed for over 100 years in the Atlantic and the Pacific Ocean (Fuhrman et al., 2006; Robinson et al., 2009; Malmstrom et al., 2010). Today almost every coastal laboratory that specialises in marine sciences and has access to small vessels has created a time series for the observation of biological and oceanographic trends over time; an important aspect in climate change studies and ecosystem function observations. One such station is the Western Channel Observatory (WCO) time series station- L4 (Figure 5.3), located 10 km south of Plymouth Sound in the Southern part of the UK (50° 15.00' N, 4° 13.02' W, www.westernchannelobservatory.org.uk). It is one of the oldest series of

observations in the English Channel, with samples and routine core measurements being collected since the beginning of the 20th century. Around the L4 station, *Emiliania huxleyi* blooms are a common feature occurring almost annually typically during summer months when conditions are most suitable for growth (Widdicombe et al., 2010), with a low abundance of coccolithophores usually observed throughout the year (Smyth et al., 2010). With additional funding during the last 20 years, this station has turned into a fundamental tool for global climate change prediction research, the modelling of biogeochemical cycles, phytoplankton primary productivity observations, zooplankton composition analysis, and more recently studies on microbially regulated processes and their genetic and metabolic biodiversity.

Figure 5.3. Western Channel Observatory (WCO) station L4, south of Plymouth, UK.

5.2.2.1. Western Channel Observatory (WCO) sample collection

For the samples that originated at the L4 site and were used in this Chapter, weekly seawater samples of one L in volume were taken when possible between 2001 and 2007 from surface waters by a CTD sampling rosette. The collected water was brought to the laboratory as soon as possible and was filtered through 0.45 µm

Millipore nitrocellulose membrane filters (47 mm). The filters were then snap-frozen in liquid nitrogen and stored at -80°C for future molecular analysis. The result of this time series sampling effort were a further 117 total DNA surface seawater samples available for the molecular analysis of coccolithoviruses.

5.2.3. DISCO cruise sample collection (1999)

An additional 144 extracted total DNA samples were obtained from the Plymouth Marine Laboratory DNA archive. These samples were originally collected in June, 1999 on board the RRS Discovery during a phytoplankton bloom located in the North Sea (East to West from -2.0° to 4.0° and North to South from 61.0° to 51.0°). The diversity of coccolithoviruses and coccolithophores found during this bloom using detection of the MCP gene fragment have been shown elsewhere (Martinez et al., 2012), and therefore was not repeated in this study. Here, an SPT diversity study was performed on the North Sea samples.

5.2.4. DNA Extraction, PCR and DGGE analysis

All samples (except the already processed DISCO samples) were subjected to a total genomic DNA extraction following an adapted phenol-chloroform protocol as described by Schroeder et al. (2005) (Section 2.3.6.). Extracted DNA samples were subjected to a two-step nested PCR, DGGE fingerprinting, and a third step PCR prior to sequencing. A detailed representation of the method used during the AMT cruise and thereafter can be seen in Figure 5.4. Primers and reaction volumes for the detection of the coccolithovirus Major Capsid Protein (MCP) by DGGE have been described previously (Schroeder et al., 2002; Martinez et al., 2007, 2012; and Section 2.3.1 and 2.3.2.1). Primers for the detection of the coccolithovirus serine

palmitoyltransferase (SPT) gene were designed manually following the multiple DNA alignment of SPT from nine fully sequenced coccolithovirus genomes (EhV-84, EhV-86, EhV-88, EhV-99B1, EhV-201, EhV-202, EhV-203, EhV-207 and EhV-208). All PCR reactions were conducted in a VWR JENCONS Uno Thermal Cycler in 25 μL final volume as described in Section 2.3.2.1.

For the first step of the nested PCR reaction with the first set of MCP primers, the cycle conditions included an initial denaturation step at 95°C for 3 min, followed by 34 cycles at 95°C, 60°C and 74°C for 30, 60 and 90 sec respectively, and a final cycle at 95°C, 60°C and 74°C for 30 sec, 5 min and 5 min respectively. Only samples that gave a band when visualised by agarose electrophoresis after the first step were subjected to the second PCR step. The cycle conditions in the second step of the nested PCR reaction with the second set of MCP primers were the same as the first step with the exception of the annealing temperature which was 55°C.

For the first step of the nested PCR reaction with the first set of SPT primers, the cycle conditions included an initial denaturation step at 95°C for 5 min, followed by an annealing gradient of decreasing temperatures of 66-56°C (i.e. 1 min at each temperature), 34 cycles at 95°C, 56°C and 72°C for 30, 60 and 90 sec respectively, and a final cycle at 95°C, 56°C and 72°C for 30 sec, 5 min and 5 min respectively. Again, only samples that gave a band when visualised by agarose electrophoresis after the first step were subjected to the second PCR step. The cycle conditions in the second step of the nested PCR reaction with the second set of SPT primers were the same as the first step.

All the samples that gave positive bands after the second steps were subjected to a DGGE fingerprinting. The DGGE was performed using an Ingeny PhorU-2 system. 15 µL of nested PCR product (from PCR step 2) was applied directly onto an 8% w/v polyacrylamide gel (acrylamide /N,N'-methylene bisacrylamide, 37:1, w/w) in 1 × TAE buffer (40 mM Tris pH7.4, 20 mM NaAcetate, 1 mM Na_2EDTA). A 30 to 60 % linear denaturing gradient was formed using 20% and 80% denaturants (100% denaturant being 7 M urea and 40% v/v formamide). 20 µL of marker (composed of the single product from the nested PCR of ten laboratory strains) was used in the first well of each gel. For SPT samples, electrophoresis was performed at a constant voltage of 100V and a temperature of 60ºC for 17 hrs. For MCP samples a constant voltage of 200V was used for 3.5 hrs.

Following electrophoresis the gels were stained for 1 h in Milli-Q water containing 1 µg mL^{-1} ethidium bromide then de-stained in Milli-Q water for 1 h, visualised on a UV transilluminator (Syngene GeneGenius) and photographed using the Syngene GeneSnap software. Bands of interest were excised and incubated in 30 µL of DNA water at 4°C overnight and then 2 µL were used as a template for a final third step PCR (prior to sequencing) following all the conditions of the second steps described above (using MCP-F2-GC and MCP-R2 or SPT-F2 [GGAATCTCGCGCGCG] and SPT-R2) with the exception of the number of cycles used which was increased to 45.

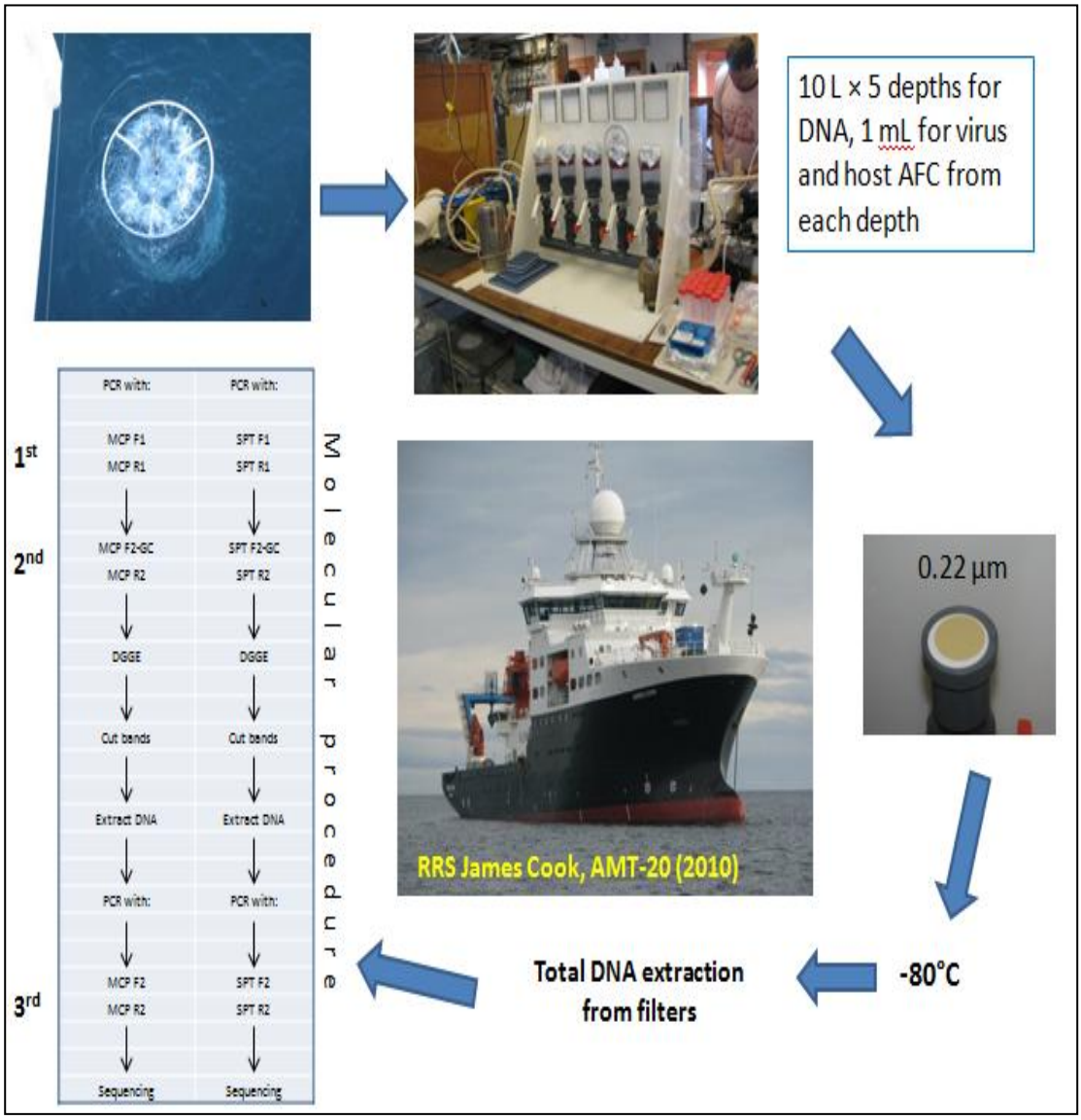


Fig. 5.4. AMT-20 sample collection through a 24 arm sampling rosette (upper left) on board the RRS James Cook and the processing of samples, from filtration to molecular fingerprinting by PCR and DGGE.

Sanger sequencing on the amplified samples after the 3rd PCR step was performed by the LGC Genomics Sequencing Centre (www.lgcgenomics.com). Briefly, the LifeTech (former Applied Biosystems) BigDye version 3.1 sequencing mix was used for cycle sequencing. After the purification of the sequencing reactions by gel filtration (Centri –Pur 96, EMP biotech Berlin) the samples were run on an ABI 3730 XL instrument using POP7 polymer and standard run conditions for up to 1,000 nt read length.

5.2.5. Bioinformatic analysis

Coccolithovirus isolate MCP and SPT sequences retrieved from GenBank and the new environmental sequences (submitted to GenBank under accession numbers: AB738836, AB738837, AB738838, AB738839, AB738840, AB738841, AB738842, AB738843, AB738844, HE970437, HE970438, HE970439 and HE970440) were aligned using the MEGA4 (version 4.0.2) multiple sequence alignment software (Tamura et al., 2007). Prior to CLUSTALW alignment the primer sequences were removed and the sequences were made the same length in order to decrease bias that can arise from potential gaps and sequences with different lengths.

5.2.5.1. MCP and SPT 3D protein modelling

Translated MCP and SPT gene sequences from EhV-86 and EhV-99B1 were modelled in order to determine their predicted secondary and tertiary protein structures and to compare these to the protein folds of the new MCP and SPT sequences obtained from the environmental samples in this Chapter. All full and partial sequences were uploaded into the Phyre2 protein fold recognition server (Kelley and Sternberg, 2009), and the final models of SPT and MCP were conducted by Phyre2 using seven and six templates respectively in order to maximise confidence, percentage identity and alignment coverage. The resulting theoretical models were uploaded into the 3D protein structure analysis software: Jmol (Hanson et al., 2013), Swiss PdBViewer (Guex and Peitsch, 1997), and Astex (Hartshorn, 2002).

5.2.6. Analytical Flow Cytometry (AFC) of coccolithophores and their viruses

Coccolithophore and virus abundances were determined using a FACScan Flow Cytometer (Beckton Dickinson, Oxford, UK) as described by Brussaard et al. (2000) and in Chapter 2.2. Data files were analysed using the WinMDI 2.9 and CellQuestPro™ software.

5.3. RESULTS

5.3.1. Coccolithophore and coccolithovirus AFC

Analytical flow cytometric analysis revealed a very small number of coccolithophores across the Atlantic transect, indicating that there was no obvious coccolithophorid bloom during the AMT-20 cruise (Figure 5.5). The highest abundance of coccolithophores was seen at the beginning of the transect in the first few days, where the average cell number was 35-50 mL^{-1}. Then the number of cells decreased and during the majority of the remaining transect there were on average 10 cells mL^{-1}. The number of cells peaked again to a maximum of 35 mL^{-1} towards the end of the Atlantic transect near the South East part of Latin America, as expected in regions of higher productivity such as the Argentinean and Chilean continental shelves.

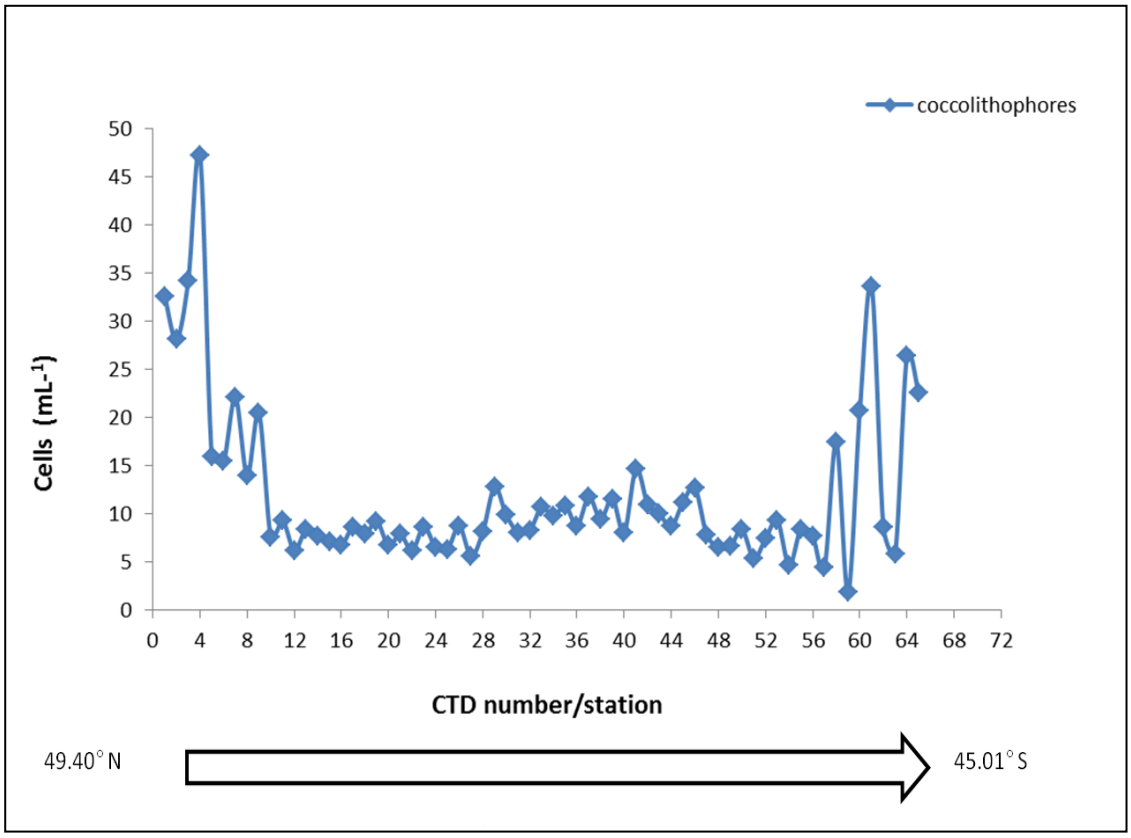


Figure 5.5. Average coccolithophore abundance along the AMT-20 transect, North to South. Coccolithophores were identified using their red fluorescence *vs* side scatter profiles during AFC analysis.

5.3.2. Coccolithovirus Major Capsid Protein diversity

Coccolithovirus MCP product was successfully amplified in a quarter (80 samples) of the 325 samples collected during the AMT-20 transect. The majority of positive samples displayed multiple distinct bands by DGGE analysis, indicating that there was more than one coccolithovirus strain (or genotype) present in these samples. The MCP fragments were detected at locations along the entire transect (Figure 5.6), even in areas where host presence was not expected and was only observed by AFC in small numbers (Figure 5.4); i.e. in the gyres. The number of distinct bands observed by DGGE analysis decreased around the North and South Atlantic Gyres, as phytoplankton abundance also decreased in these areas. The highest numbers of distinct MCP bands were consistently detected at the Deep Chlorophyll Maximum (DCM) region of each sampling station along the transect, and there was a significant positive correlation between increased depth and number of distinct MCP bands observed per sample (Pearson correlation, r=0.53) (Figure 5.7).

MCP products were also amplified successfully in 103 out of 117 L4 time series samples (i.e. 83%). Hence, at the L4 sampling station coccolithovirus MCPs were detected in all years (2000-2007), throughout each year. In the analysis by Martinez et al. (2012) of the DISCO North Sea research cruise samples, out of a total number of 144 samples, they successfully amplified an MCP product from 132 samples (i.e. 91%).

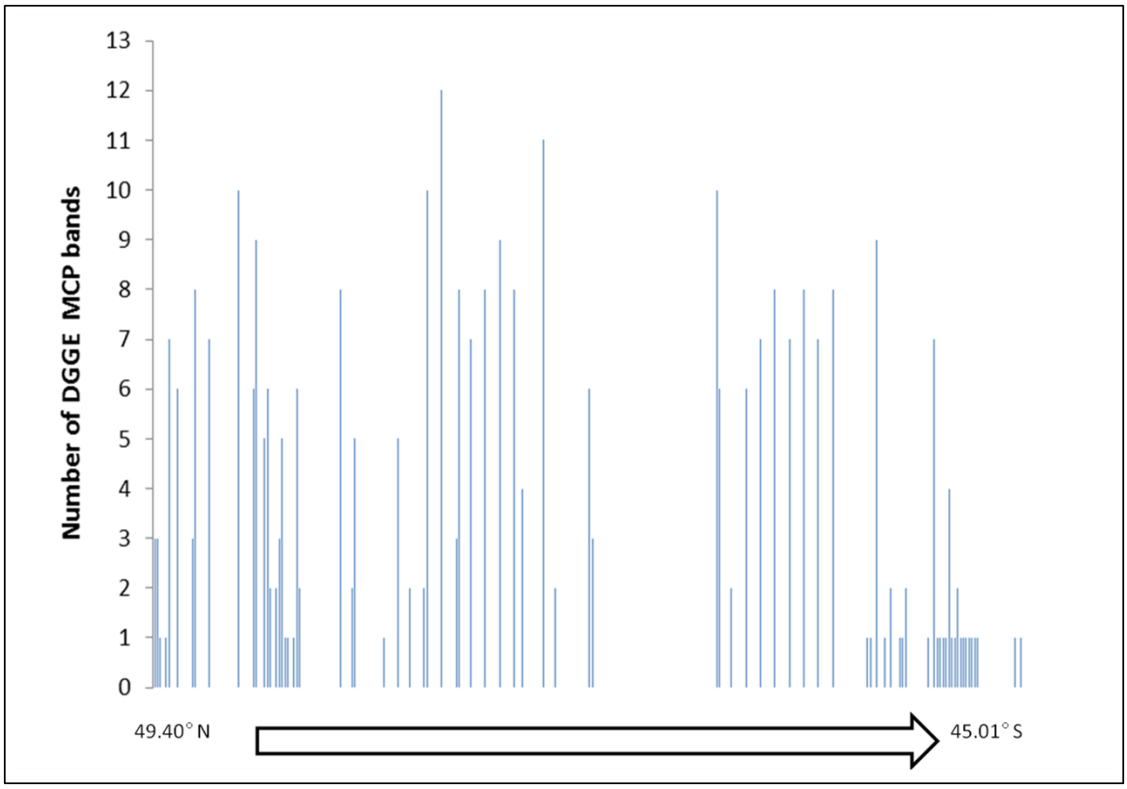


Fig. 5.6. Number of MCP bands in each sample on the AMT-20 transect, North to South. The detection of bands on the DGGE gels was after the second step of the nested PCR.

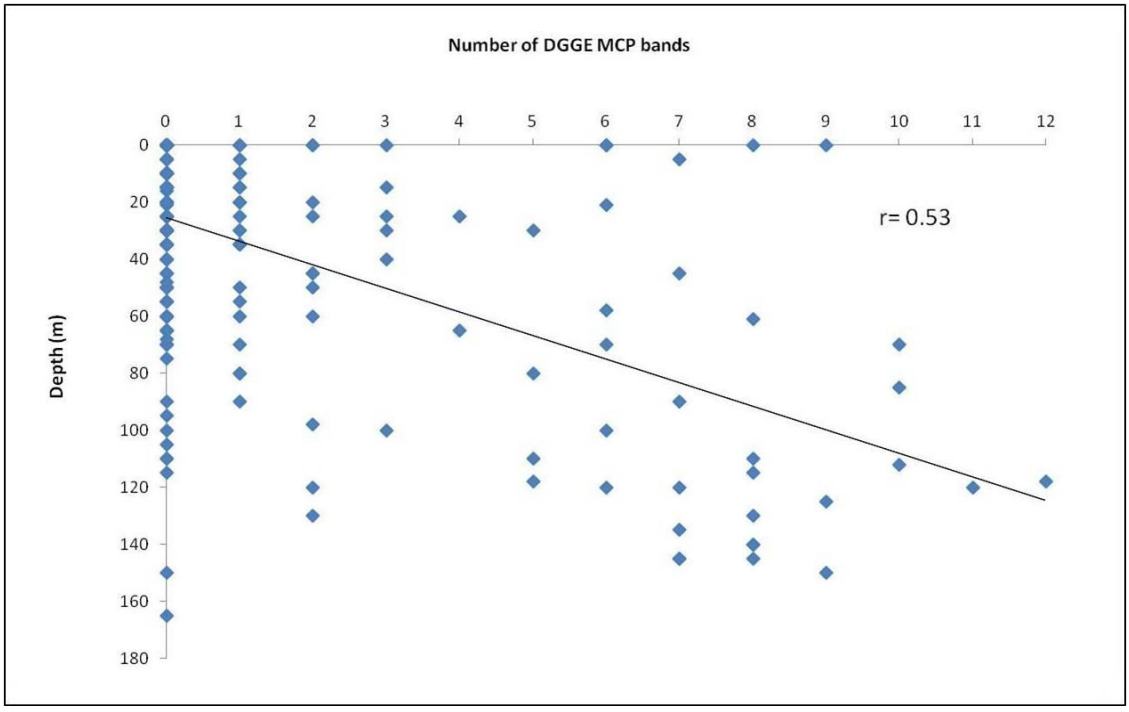


Fig. 5.7. The number of MCP bands in each sample detected by DGGE on the AMT-20 transect cruise. There was a positive correlation of the number of distinct MCP bands with increasing depth (Pearson, r=0.53). 245 samples had no bands and 80 samples had one or more bands, increasing with depth.

Distinct MCP bands were successfully extracted and sequenced from the DGGE gels. AMT samples yielded 33 unique MCP bands, and L4 samples a further 25. The new sequences were used in combination with nine MCP gene fragments from the EhV isolates described in Chapter 3 for multiple sequence alignment and a phylogenetic analysis. Although DGGE analysis was based upon a 135 bp fragment, restrictions in sequencing methodology allowed the reliable generation of sequence from only a 50 bp region of the MCP gene. Most AMT-20 and L4 sequences clustered into one of five sub-clades of coccolithoviruses: denoted as A, B, C, D or E, alongside the established laboratory isolates (Figure 5.8). However, nine new genotypes were discovered (and subsequently submitted to GenBank): AMT2b-3 [DDBJ: AB738836], AMT273-1 [DDBJ: AB738837], AMTt1-2 [DDBJ: AB738838], L4/23.04.2001/c [DDBJ: AB738839], L4/12.07.2006/a [DDBJ: AB738840], L4/04.06.2007/a [DDBJ: AB738841], L4/02.01.2006 [DDBJ: AB738842], L4/02.09.2002 [DDBJ: AB738843], and L4/01.07.2003 [DDBJ: AB738844].

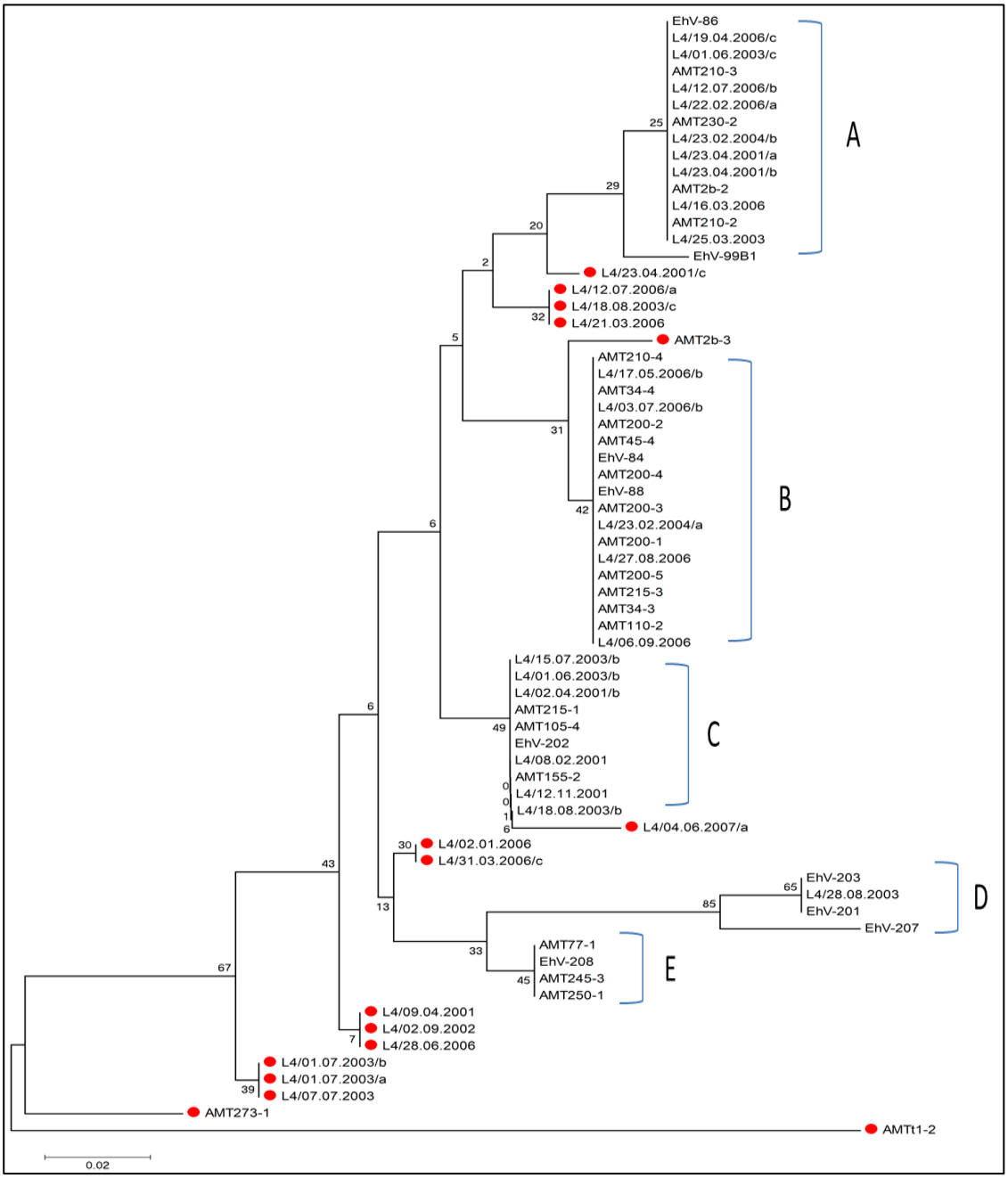


Fig. 5.8. Evolutionary relationships of 58 MCP gene fragment sequences and further nine sequences from laboratory EhV isolates**.** The sequences were extracted from DGGE bands from the AMT-20 transect and the L4 time series samples, the nine sequences of laboratory EhV strains were obtained from the NCBI database. In red are indicated novel sequences that cluster outside of the five already known clades of EhV MCPs (A=EhV-86 and EhV-99B1 like, B= EhV-84 and EhV-88 like, C= EhV-202 like, D=EhV-201, EhV-203 and EhV-207 like, and E=EhV-208 like). The evolutionary history was inferred using the Neighbour-Joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch length = 0.54677874 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10000 replicates) is shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site (Tamura et al., 2004). Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 50 positions (bp) in the final dataset. Phylogenetic analyses were conducted in MEGA4 (Tamura et al., 2007).

5.3.3. Structural implications of MCP diversity

Despite the short MCP fragment the analysis was based upon, the coccolithoviruses displayed high diversity with nucleotide polymorphisms observed at 14 of the 50 nucleotide locations under analysis (Figure 5.9), which corresponded to just three amino acid changes when translated (Figure 5.10). The majority of the novel environmental sequences differed from the EhV isolate sequences at one amino acid position; e.g. a single phenylalanine replaced by serine, or valine replaced by isoleucine. The AMT-20 sequence AMTt1-2 differed in two amino acids from the known EhV isolate sequences; alanine replaced by serine, and valine replaced by isoleucine. Using the full length MCP from EhV-99B1, the 3D protein structure of the predicted major capsid protein was modelled in order to reveal possible structural differences resulting from MCP sequence diversity. 74% of the final model of the EhV-99B1 MCP protein (Figure 5.11) was modelled with an accuracy of >90%, based on the following six protein structures on the Protein Databank in a decreasing order of alignment coverage and % of confidence: 1J5Q, 1M4X 1M3Ya1, 1M3Ta2, 3SAM, and 2W0C. The MCP protein of EhV-99B1 was modelled again several times with the newly obtained sequences from the AMT-20 transect and L4 time series, incorporated into its sequence and replacing the original 50 bp sequence in this gene.

When modelled in 3D, the coccolithovirus diversity observed within the variant sequences was predicted to have negligible, if any, impact on overall tertiary structure (data not shown) of the EhV-99B1 MCP. Indeed, even the potentially greatest change between the small, polar serine and the large, hydrophobic phenylalanine in the incorporated sequences had no impact whatsoever on the predicted structures. This was most likely due to the high degree of structural conservation, regardless of their

primary sequence, observed in all major capsid proteins and virions studied to date (Bamford et al., 2005; Krupovic and Bamford, 2008, 2011; Abrescia et al., 2012).

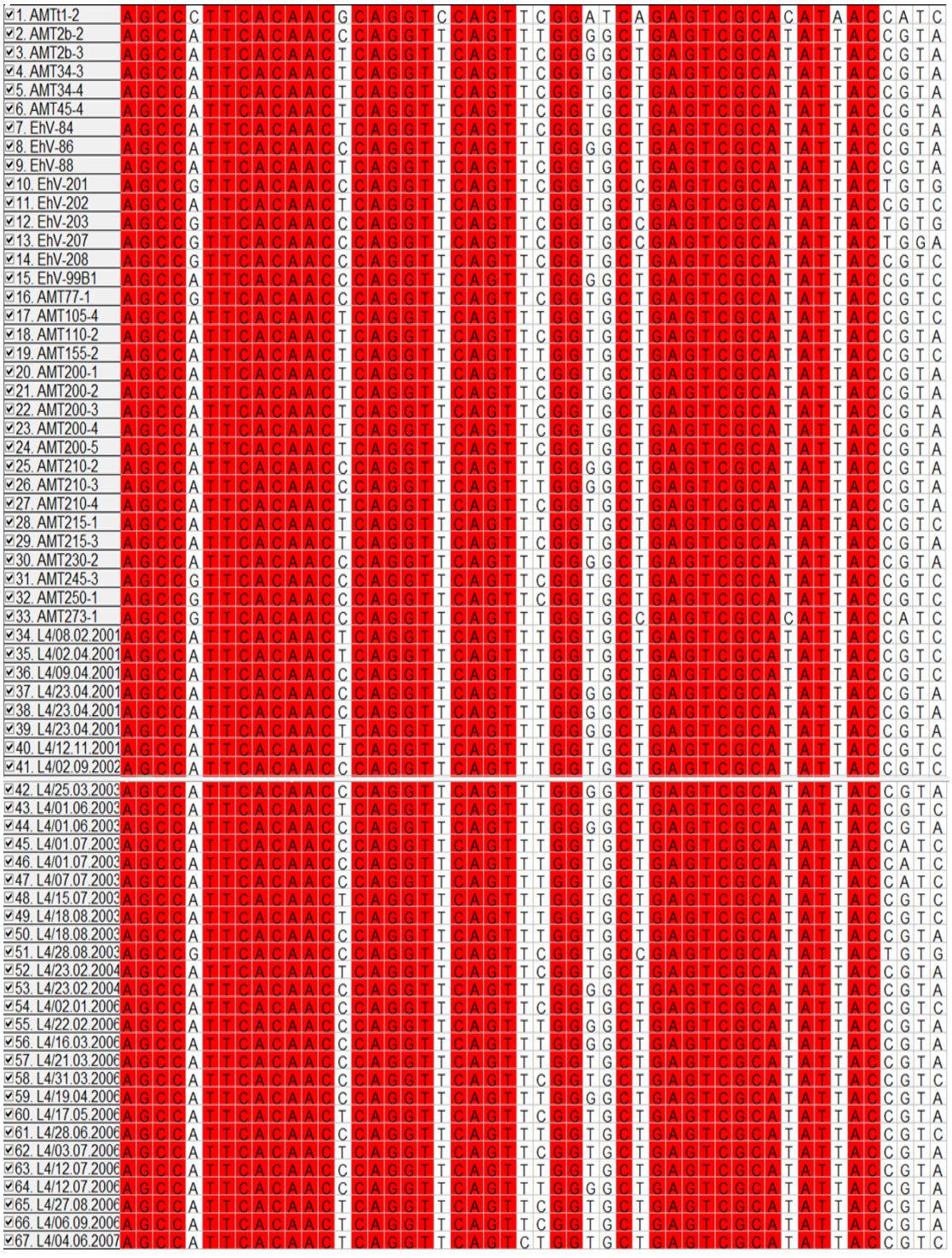


Fig. 5.9. CLUSTALW alignment of 67 (50 bp long) fragments of the MCP gene from AMT-20, L4 and laboratory EhV strains. In red are highlighted conserved areas along these sequences.

	P	F	T	T	Q	V	Q	F	G	A	E	S	H	I	T	V
1. AMT2b-3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
2. AMT273-1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	I
3. AMTt1-2	.	.	.	.	.	.	.	.	.	S	.	.	.	.	.	I
4. L4/23.04.2001/c	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
5. L4/12.07.2006/a	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
6. L4/04.06.2007/a	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.
7. L4/02.01.2006	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
8. L4/02.09.2002	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
9. L4/01.07.2003/a	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	I
10. EhV-84	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
11. EhV-86	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
12. EhV-88	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
13. EhV-201	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
14. EhV-202	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
15. EhV-203	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
16. EhV-207	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
17. EhV-208	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
18. EhV-99B1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Fig. 5.10. CLUSTALW alignment of 16 amino acid fragment sequences of the MCP gene from AMT-20, L4 and laboratory EhV strains. In yellow are highlighted conserved areas along these sequences and a consensus sequence is shown at the top of the table.

Fig. 5.11. 3D model of the hypothetical structure of the full MCP protein (496 aa) encoded by the laboratory isolate EhV-99B1.

5.3.4. Coccolithovirus serine palmitoyltransferase (SPT) diversity

Despite the success with the MCP marker, SPT gene fragments could not be amplified in any of the AMT-20 samples. However, 335 bp products of the coccolithovirus SPT fragment gene were amplified in 41 L4 samples (35%) from a total of 117. Of the 144 DISCO cruise samples, 65 (45%) successfully yielded an SPT product. In total 14 and 16 distinct bands were successfully extracted and sequenced from the L4 and DISCO SPT DGGE gels, respectively. Phylogenetic analysis of a sequenced 207 bp region of these products revealed that most L4 and DISCO sequences clustered into one of four sub-genotypes of the SPT gene: A, B, C, or D, and were similar to SPT fragment sequences obtained from characterised EhV isolates (Figure 5.12). However at least four new genotypes were discovered: DIS/29.06.1999/80m [EMBL: HE970437], DIS/23.06.1999/100m/a [EMBL: HE970438] (E); L4/16.09.2002 [EMBL: HE970439] and L4/15.07.2003 [EMBL: HE970440] (F).

5.3.5. Structural implications of SPT diversity

The amplified region of the SPT gene targets the most variable section of this gene among laboratory isolates and corresponds to the "linker" region between the two domains of the SPT protein: LCB1 and LCB2. Based on full genomic sequences available, the entire SPT from EhV-99B1 and EhV-86 were 3D modelled (Figure 5.13 A & B). 89% of the final models of the SPT protein were modelled with an accuracy of >90%, based on the following seven protein structures on the Protein Databank in a decreasing order of alignment coverage and % of confidence: 3LWS, 1BS0, 3HQT, 3A2B, 2BWN, 3TQX, 2W8W. In order to discover whether the new environmental SPT fragment sequences can affect the predicted structure of the entire SPT protein with its two subunits, the original "linker" of the EhV-99B1 SPT protein was replaced

and substituted by these environmental sequences. Then these new hypothetical amino acid sequences were re-modelled in the same way that the original EhV-99B1 and EhV-86 SPT sequences were modelled.

From the secondary structure prediction (Figure 5.14) and the 3D models it was clear that the “linker” region of the SPT protein has the potential to differ structurally among strains and also influence the entire fold of the SPT protein. The two domains LCB1 and LCB2 of the protein are potentially closer to each other in EhV-99B1 than in EHV-86 (Figure 5.13) and this was most likely due to two amino acid changes in the linker between the two domains; i.e. in EhV-86 tyrosine and glutamic acid were both replaced by aspartic acid. The changes from tyrosine to aspartic acid affected the position and the length of the first alpha helix in the sequenced region, resulting in a shorter helix in EhV-99B1 (Figure 5.14).

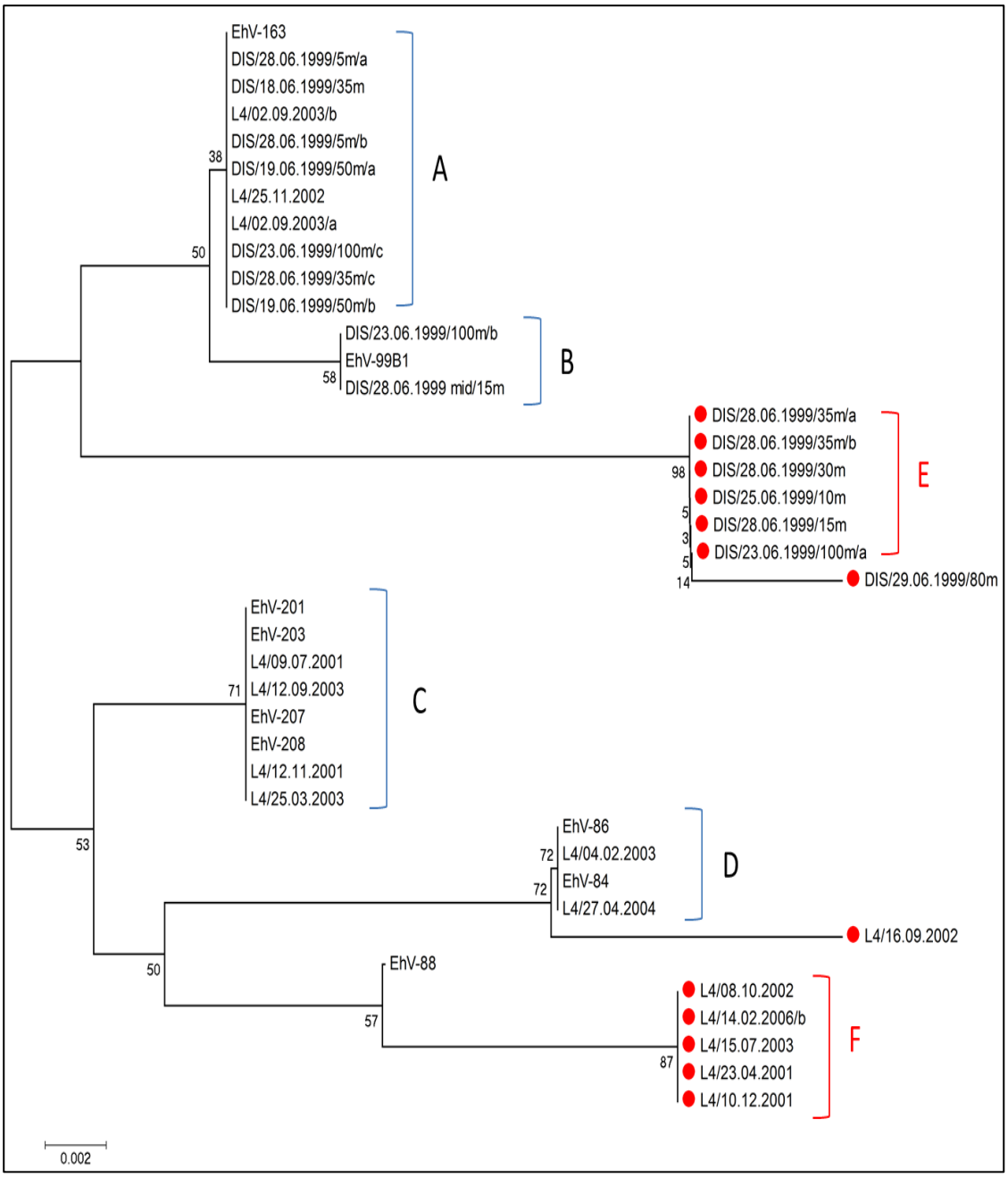


Fig. 5.12. Evolutionary relationships of 31 SPT gene fragment sequences. The sequences were extracted from DGGE bands from the L4 time series and the DISCO cruise samples, the nine sequences of laboratory EhV strains were obtained from the NCBI database. In red are indicated novel sequences which cluster outside of the four already known clades of EhV SPTs (A=EhV-163 like, B=EhV-99B1 like, C=EhV201, EhV-203, EhV-207 and EhV-208 like; and D=EhV-84 and EhV-86 like). The evolutionary history was inferred using the Neighbour-Joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch length = 0.08538966 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10000 replicates) is shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site (Tamura et al., 2004). Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 207 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 (Tamura et al., 2007).

In addition there were also potential differences between the modelled fold of the protein when L4/16.09.2002 and L4/15.07.2003, and the DIS/23.06.1999/100m and DIS/29.06.1999/80m sequences were incorporated into the EhV-99B1 sequence. The first alpha helix in the L4/16.09.2002 and L4/15.07.2003 predicted structures of the incorporated sequences were predicted to be longer than the same helix in the original EhV-99B1, EhV-86, DIS/23.06.1999/100m and DIS/29.06.1999/80m structures (Figure 14). The helix was 11 aa long and had an altered position. This was most likely a result of an amino acid (aa) change immediately after the helix, i.e. proline, serine and tyrosine instead of proline, serine and aspartic acid. The sequence PSY has a net neutral charge while the sequence PSD is negatively charged. Regardless of this, the predicted folds of the L4/16.09.2002 and L4/15.07.2003 derived sequences were more similar to the fold of the EhV-86 SPT than the fold of EhV-99B1, and the folds of DIS/23.06.1999/100m and DIS/29.06.1999/80m structures were more similar to the fold of the EhV-99B1 than the EhV-86 version (Figure 5.13).

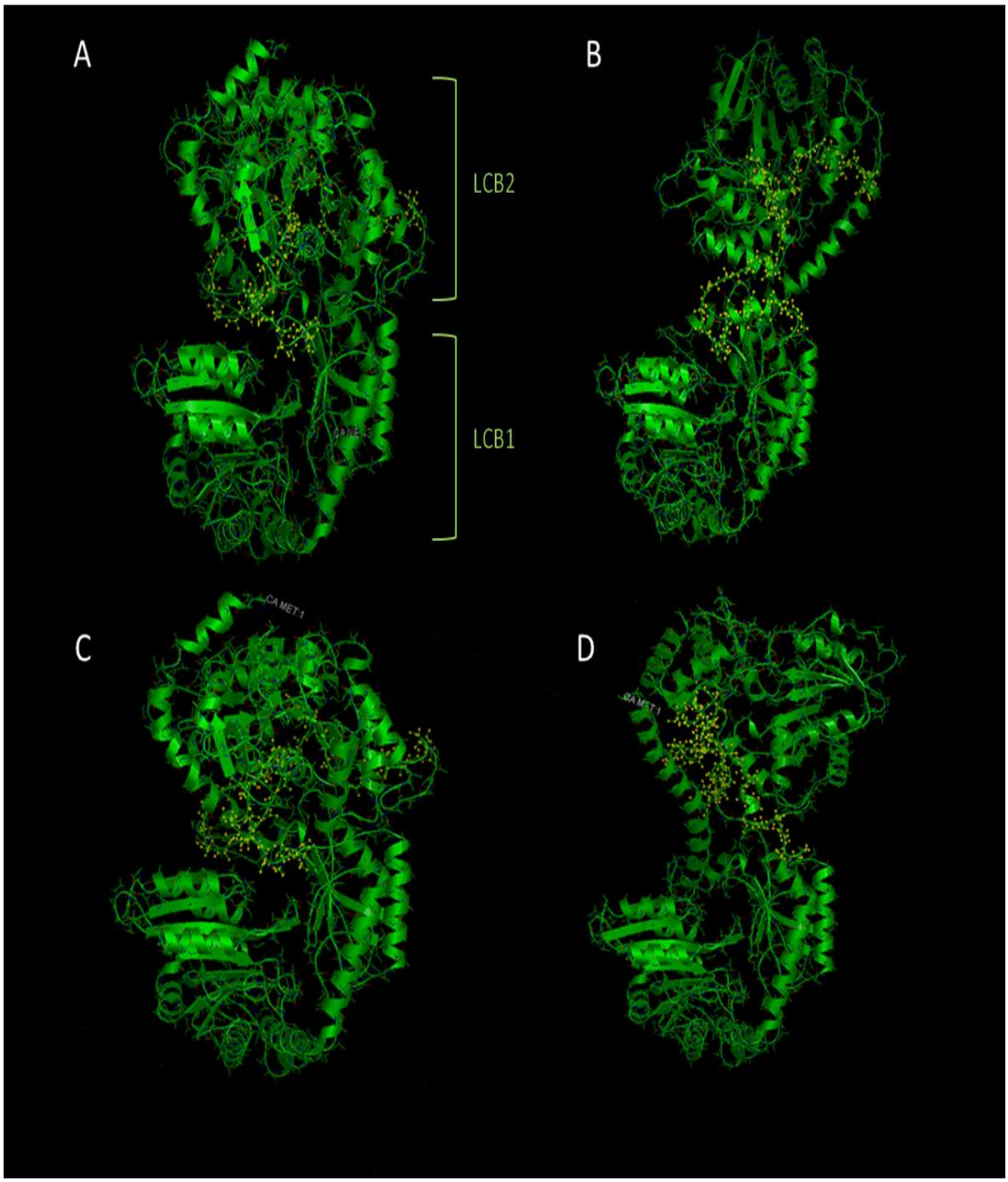


Fig. 5.13. 3D models of four different hypothetical structures of the SPT protein: A) laboratory strain EhV-99B1 isolated originally from a Norwegian Fjord, B) laboratory strain EhV-86 isolated originally from the English Channel, C) a 207 bp (or 69 aa) long sequence from the DISCO cruise in the North Sea incorporated and replacing an original 207 bp long sequence in the full SPT gene of EhV-99B1, and D) a 207 bp (or 69 aa) long sequence from the L4 sampling station in the English Channel incorporated and replacing an original 207 bp sequence in the full SPT gene of EhV-99B1. In yellow are highlighted the amplified regions between the two domains LCB1 and LCB2 (i.e. the linker between the two domains) and in green are seen the predicted beta sheets and alpha helixes of the SPT proteins.

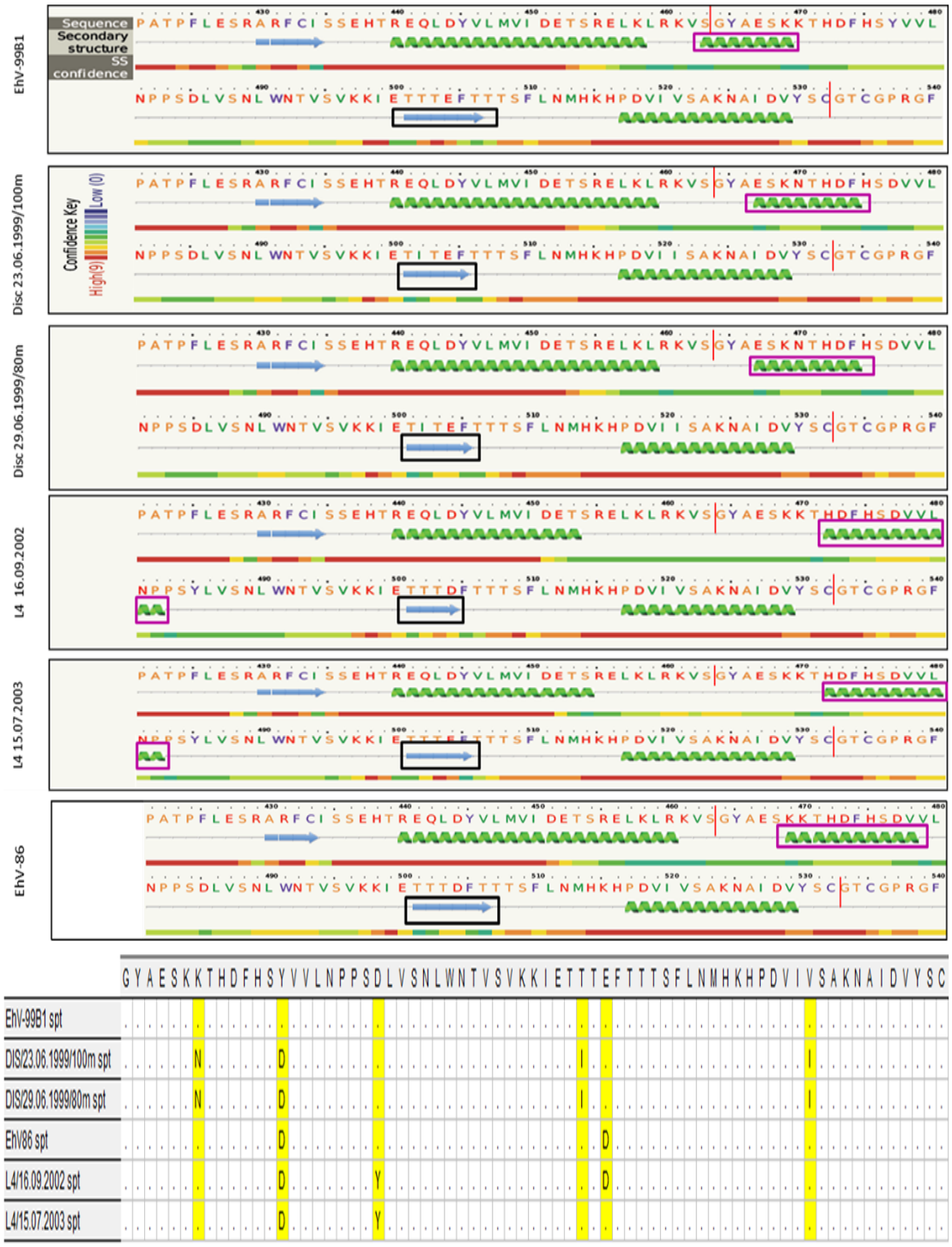


Fig. 5.14. Secondary structure predictions and CLUSTALW alignment of: amplified 69 amino acid (207 bp long) regions of the SPT protein of four environmental sequences and two laboratory strains, located between the two domains LCB1 and LCB2. Alpha helixes, beta sheets and a confidence scale bar of the prediction are shown below each sequence. The variable regions of the predicted structures of each sequence are highlighted in the black and purple boxes and also in yellow in the alignment table.

5.4. DISCUSSION

Previously, studies that have looked at the diversity of natural populations of coccolithoviruses and/or their infection kinetics with their hosts have relied upon the detection of VLPs via direct enumeration by AFC, occasionally in combination with MCP molecular fingerprinting where and if applicable (Schroeder et al., 2002; Martinez et al., 2007, 2012; Sorensen et al., 2009). However, these studies depended on high numbers of host cells and their associated virus community. Here, with the exception of the samples collected during the DISCO cruise which were during a coccolithophore bloom, the focus was on environmental samples where there was no obvious coccolithophore population present at the time of sampling or coccolithophores were present but in very small numbers (such as the AMT-20 samples).

Until now, there has been little if any information on the "invisible" abundance of coccolithoviruses under non-bloom conditions, a state in which the coccolithophore host is found during the majority of its life cycle in the world's oceans. In support of the "everything is everywhere" hypothesis (de Wit and Bouvier, 2006), the MCP marker revealed that coccolithoviruses were indeed present along the entire Atlantic transect and were not restricted to locations where host abundances were high. Even in locations where the coccolithophore abundance was lower than 10 mL^{-1} at least one dominant coccolithovirus genotype was detectible by DGGE fingerprinting. This highlights the importance of using the right sampling approach when host and virus numbers are low. The captured diversity in this study was successful largely due to the filtration of large quantities of seawater through filters with very small pore sizes

(although this can be also achieved as effectively by ultra-filtration and virus concentration protocols not covered in this thesis).

Although the environmental physicochemical parameters recorded at the time of sampling (i.e. the core measurements collected during the CTD sampling) were not linked to the molecular data obtained in this study (future more comprehensive studies will need to be undertaken), several ecologically relevant observations can still be made. The first is that the number of distinct MCP bands representing different coccolithoviruses increased with increasing depth. This raises interesting questions of the ability of this group of viruses to "survive" in times of reduced activity and rate of infection during non-bloom conditions. When distances between host cells are small, viral infection is rife and the host population can be rapidly decimated, as seen under *E. huxleyi* bloom conditions (Schroeder et al., 2003). However, viruses must "survive" the lean times when host numbers are sparse. Exposed to, and at the whim of, the environment, viruses will quickly perish, especially if they have no host to protect them during times of high UV irradiation or in locations where the days are longer and the virions are exposed to photo-damage for longer periods of time (Wilhelm et al., 1998; Jacquet and Bratbak, 2003). A possible mechanism for evading such stress factors could be their association with the deep chlorophyll maximum (DCM) region, where the typical percentage of light penetrating is around 1% and the infection and virus production turnover is likely somewhat slower as coccolithophore hosts require sufficient light for their replication and subsequent bloom formation. Although in principle free floating viruses in the water column have no way of controlling the depth that they are at, the fact that on the Atlantic Ocean transect their diversity increased with depth towards the DCM region could be due to the lack of sufficient

hosts for attachment and adsorption and as a consequence their subsequent sinking and attachment to non-host material in the form of "marine snow". Hence, indirectly the depths could harbour the diverse reservoirs of genetic potential exploited under bloom conditions overhead, in an 'infection from below' dynamic, which is reliant on vertical mixing of the water column. Most likely both host and virus are transported up to the surface together and when the host increases in numbers at the beginning of a bloom and its biomass is high enough, active large scale infection can begin. Such a hypothesis can be supported by the observation that most samples from the DCM region at the AMT-20 transect had a larger number of MCP sequences than the samples from shallower depths, and thus a higher diversity of coccolithovirus genotypes.

The second ecologically-relevant observation from this study is that coccolithoviruses appear less diverse with regards to their phylogenetic classification than previously thought, at least with regards to the markers used, evident by the detection of sequences across the Atlantic transect with 100% similarity to sequences from laboratory EhV strains. Most MCP sequences clustered into one of the already established coccolithovirus clades and only MCP sequences obtained from a few locations exhibited potentially new coccolithovirus genotypes. The former can be also said about the L4 station where a small number of new coccolithovirus genotypes were also present several years after the original isolation of the laboratory EhV strains from the same location. Further support for the somewhat smaller phylogenetic diversity of coccolithoviruses comes from the previous MCP analysis conducted by Martinez et al. (2012) on the DISCO North Sea samples and from the analysis conducted by Rowe et al. (2011) in the North Atlantic where it was also shown that

the MCP gene fragments in these distinct locations were similar to the ones found in laboratory EhVs isolated originally in 1999 at the English Channel L4 station. This however needs to be taken with a slight caution as the phylogenetic analysis in this Chapter was based on sequences that were only 50 bp long and do not represent the entire MCP gene sequence. In addition there might have been a slight bias towards strains similar to the ones upon for which these markers were developed. Given that the markers were designed based on established laboratory EhV strains it is possible that this is why they also only detected similar strains, and that genotypes with a more diverse sequence might have been missed.

Nevertheless, proven and established molecular markers such as the MCP gene are useful when the aim is to study natural samples where the preliminary information about the sites of sampling and potential habitat range of the host is limited. However, MCP is limited when attempting to make functional inferences based directly on the phylogenetic diversity measurements. Therefore, a new genetic marker was developed here, and an attempt was made to make such functional inferences. While the SPT marker lacked the sensitivity and/or spatial coverage of the established coccolithovirus MCP marker, structural (and thus potentially functional) differences were inferred based on the information obtained from using it (i.e. the 3D models and their possible ecological significance). A possible future step in their investigation would be to express these different SPT types and assay the precise enzyme function and their performance during infection. The later approach is likely to be successful due to the already characterised EhV-86 SPT enzyme in the laboratory (Han et al., 2006).

The lack of amplification of SPT from any of the AMT samples was intriguing. The failure to detect the SPT gene from the AMT samples might have been a consequence of detection limits, in that the primers that were used were just not sensitive enough for these samples in which coccolithovirus numbers were presumably very low (they were not detected by AFC). Alternatively, the SPT gene might have been absent or present in a more divergent form in these samples that was undetectable by the designed SPT probes. To date, all coccolithoviruses that have been isolated and sequenced have been shown to harbour the SPT gene, regardless of their geographical source of isolation (Wilson et al., 2005; Allen et al., 2006b; Nissimov et al., 2011a, 2011b, 2012a, 2012b; Pagarete et al., 2013) hence the absence of this gene in the Atlantic based coccolithovirus communities is unlikely.

The SPT gene in coccolithoviruses is the product of a horizontal gene transfer event from the host, and is unique in being the only SPT biochemically characterised that is a fusion between the LCB1 and LCB2 domains (Han et al., 2006; Monier et al., 2009). Since the target of the primers was the "linker" region between these domains, SPT may still be present in the AMT coccolithovirus community, but as two distinct genes, yet to be linked up. Future characterisation of the individual LCB1 and LCB2 domains from both host and virus systems may shed some light on this theory. Nevertheless, the 3D models of the coccolithovirus SPT protein suggests that the conformation of the overall protein structure has the potential to be altered when the "linker" section is different, even if this difference is a change in only two or three amino acids. Based on these models, LCB2 and LCB1 are further away from each other in the EhV-86 encoded SPT than in the EhV-99B1 version of the predicted protein, and further apart still in some of the environmental isolates (the "linker" section is the most diverse and

the one influencing the entire fold of the two domains as seen in the sequence substitution exercise above and a multiple sequence alignment of this gene). Given that SPT is the first and rate limiting step in the sphingolipid biosynthesis pathway (Han et al., 2006; Michaelson et al., 2010), the distances between, for example, the pyridoxal phosphate binding lysine located on the LCB2 domain and the cysteine residue in the glycine motif on the LCB1 domain, may be crucial in determining the rate and speed of co-factor binding to the SPT enzyme. If this is indeed the case then the somewhat different amino acid sequence between the two domains demonstrated in this Chapter could influence dramatically the rate of this reaction, the catalytic capability of the enzyme and as a consequence the rate of the infection cycle. An expansion of the work conducted by Han et al. (2006) in which they expressed the EhV-86 encoded SPT in yeast would be a natural next step in the study of SPT diversity and its impact on infection dynamics and *E. huxleyi*-virus interactions.

Indeed, the identification of clear infection dynamic differences in Chapter 4 between closely related coccolithovirus strains, and previous work conducted during some of my previous research (data not published or shown) where I showed clear differences in infection dynamics between EhV-86 and EhV-V1 (the latter being a relative of EhV-99B1 that has the same SPT confirmation), suggests that 'similar' virus strains can differ with regards to at least one aspect of the infection or/and replication cycle. Clearly, no firm conclusions as to the exact genetic or metabolic basis for the differences observed can be made. However, the observation that the virus isolates in the virus "fight club" experiment managed to infect and produce virions in the infected host cultures while exhibiting different infection dynamics puts the

coccolithovirus encoded SPT enzyme and the pathway in which it is involved in as a potential candidate for these observed differences.

The SPT enzyme, as previously discussed, is the crucial rate limiting step in the sphingolipid biosynthesis pathway that appears to be of utmost importance during infection (Bidle et al., 2007; Pagarete et al., 2009; Vardi et al., 2009; Bidle and Vardi 2011; Pagarete et al., 2011; Vardi et al., 2012). The implications of this could be important with regards to biogeochemical cycling in the oceans where nutrients are being constantly re-circulated via the "microbial loop" and "viral shunt" into different trophic levels and carbon export to the deep ocean following viral-induced *E. huxleyi* bloom termination (Suttle 2005; 2007; Brussaard et al., 2008; Falkowski et al., 2008). Previously, addition of viruses to seawater collected in the field led to 78 % reduction in primary productivity, indicating the crucial importance of viruses as regulator factors of phytoplankton communities and as a consequence biogeochemical cycling (Suttle et al., 1990, 1992). There is currently no specific data available in the literature that can speculate on the exact contribution of coccolithoviruses to the precise amount of carbon export post viral infection however the number is most likely very high at particular times of the year in which coccolithophores are the dominant phytoplankton species. Hence one can only speculate the specific implications of infection by viruses that vary in their infection strategies and the impact this might have on reducing primary productivity and the consequences of this reduction.

5.5. CONCLUSIONS

The use of phylogenetic markers such as the MCP gene are useful in the mapping of coccolithoviruses on a spatial scale, and the investigation of genotype diversity on a temporal scale, even when the number of virus particles is low and their detection impossible by non-molecular based techniques. However, an investigation of a conserved gene for phylogeny cannot reveal much about the functional characteristics of a given community of viruses or particular strains. Therefore, additional markers such as the SPT gene can be used in conjunction with the MCP marker which not only provides additional diversity information, but also provides crucial insights into the function and metabolic potential of the viruses under study. Moreover, an investigation of functionally important genes and their 3D theoretical modelling can provide an additional insight into the importance of these proteins, their possible mode of operation, and raise new hypotheses that can be experimentally tested in the future through infection dynamics experiments with coccolithovirus strains that exhibit these SPT differences in combination with qPCR expression profiling and microarrays. The results here provide further evidence that, the often ignored, intra-family biodiversity could have real and measurable impact on community and ecosystem composition of ecologically important groups of marine viruses.

Implicated in both cellular signalling and structural roles, virally encoded SPT control over sphingolipid production appears crucial during infection (Han et al., 2006; Bidle et al., 2007; Pagarete et al., 2009; Vardi et al., 2009; Michaelson et al., 2010; Bidle and Vardi 2011; Vardi et al., 2012). Given that a distinct battle to control the expression of the sphingolipid pathway occurs during natural bloom conditions and mesocosms (Pagarete et al., 2009), the identified potential structural difference in

SPT, and the determination that strains harbouring distinct SPTs do indeed display different infection phenotypes needs further investigation. Whether the phenotypes displayed are related directly to SPT function could not be determined in this study, and the evidence at this stage is merely circumstantial. However this study takes the first steps in the development of a new marker with a functional relevance. The computer modelling of the SPT protein and its observed diversity, together with a laboratory approach undertaken by many to establish its function during infection suggest it is of crucial importance. The real power of such markers will be revealed only after future studies, in which the activity of variant SPTs and their enzyme kinetics are established.

The genetic manipulation of virus genomes in the future will hopefully allow these questions to be addressed. For now, one needs to be content with the knowledge that coccolithovirus diversity exists, it is measurable within the natural environment and, crucially, this genetic diversity is translated into environmentally relevant phenotypes and contributes to and impacts upon global ecosystem functioning. This study represents the start of the journey to link biodiversity measurements to ecosystem function.

6. COCCOLITHOVIRUS ISOLATION AND GENOTYPIC FINGERPRINTING

6.1. INTRODUCTION

Global phytoplankton communities and their assemblages change on both a temporal and spatial scale and as a consequence their co-occurring and co-evolving viruses also differ (Dandonnea et al., 2006; Boeckel and Baumann, 2008; Malmstrom et al., 2010; Gimenes et al., 2012). This change could be either the 'disappearance' of certain strains (host and/or viral), the shift in dominance between strains due to local adaptations to varied environmental conditions, or a genetic change during which viruses or their hosts gain or lose particular features in order to enhance their own survival and future success. The latter, although a rare event, is inevitable in evolutionary terms due to the sheer number of players involved (i.e. phytoplankton, bacteria and viruses). This is particularly true in the case of biological entities such as viruses in which the genomes are relatively small and simple, and their lifestyle is short with high generation turnover over small periods of time.

Given that it was already established in Chapter 3 that coccolithoviruses from different geographical settings and those isolated from the same location several years apart differ in their genomic composition through full genome sequencing, the question over the genetic potential, size and diversity of the coccolithovirus gene pool remains. Hence it is essential to isolate new viruses to expand our knowledge of this interesting virus family. Indeed, the isolation of the last coccolithovirus was more than a decade ago, and its comparison to other coccolithovirus strains isolated several years after and the observed differences reported in Chapter 3, suggest that there is the

potential for different coccolithovirus communities and specific isolates to display significant genetic variation over the existing isolates.

To achieve this, a large scale screening of water samples collected at the L4 time series station in the English Channel and from various locations around the UK was performed against the potential infection of *E. huxleyi* CCMP 2090 in the attempt to isolate new coccolithoviruses. A microarray based study and a PCR fingerprint approach of newly isolated coccolithovirus isolates was performed to provide a snapshot of their genomic potential.

6.1.1. Microarrays and their application

Established microarray based techniques have been utilised in order to investigate the expression of genes through the analysis of messenger RNA transcripts. In medical virology, microarrays have also been used in the discovery of viruses that are difficult to culture through traditional techniques. For instance a retrovirus in prostate tumours was discovered by virus chips that were designed to detect known virus groups (Urisman et al., 2006), while viruses associated with respiratory tract infections in children were also detected by the utilisation of DNA microarrays (Chiu et al., 2008).

Microarrays (usually immobilised DNA based oligonucleotide probes arrayed onto a solid surface) have also been adopted in the marine realm of virus research. For instance they were used to show for the first time the gene expression of PBCV-1 during infection throughout its life cycle (Yanai-Balser et al., 2010) and Allen et al. (2007) used genomic DNA hybridisation based on an array designed for the EhV-86 genome in order to reveal the genomic diversity among different coccolithovirus isolates. In the latter study the EhV-86 array was used as a tool to investigate the

genomic potential of the coccolithovirus family, its characterization, and study the process of infection by EhVs through transcriptional data, revealing and demonstrating the full power of the microarray. Although microarrays are currently thought less accurate than so-called next generation sequencing (be it 'digital transcriptomics or full genome sequencing), and should be used in combination with other molecular biology techniques, in this chapter they were used in order to provide a preliminary insight into the diversity of new coccolithoviruses (prior to their future genome sequencing) due to the constrains of time and funds.

6.1.2. Polymerase chain reaction and its applications

The polymerase chain reaction (PCR) was developed by Kary Mullis in 1985 (Mullis and Faloona, 1987). Its applications in molecular biology span from the cloning of DNA for sequencing and phylogenetic and functional studies of species, to the diagnosis of infectious diseases and the establishment of genetic fingerprints of microorganisms in different macro and micro environments such as the oceans and the human associated microbiota. The basics of this method are the exposure of DNA to a variation of thermal cycling and the use of specific enzymes such as DNA polymerase for its exponential replication. Its exposure to repeated cycles of cooling and heating for the separation of the two DNA strands, the use of enzymes and the addition of short DNA fragments (i.e. primers that correspond to the mirror image of the ends of the target sequences that one wants to investigate) are a central component in the selective study of any DNA sequence of interest. There are many versions and methods for PCR, but the most common one, and the one used in this chapter, is a straight forward amplification of target sequences that belong to genes of a known or predicted function.

6.2. MATERIALS AND METHODS

6.2.1 Isolation of new coccolithoviruses

Water (1 L in volume) was collected by a CTD sampling rosette weekly from a time series site (L4) located ~10 km off the coast of Plymouth, UK and also from various coastal locations in English and Scottish waters during the last 15 years. Once the water was collected it was filtered and concentrated via tangential flow filtration system (TFF) as described in Section 2.2.3. The concentrates were then stored at 4°C in the dark until future use. In 2011, concentrates from these samples were screened against an exponentially growing culture of *Emiliania huxleyi* CCMP 2090 for signs of infection and clonal isolates of new viruses were subsequently obtained, as previously described in Chapter 2.2.3. In total ~300 concentrates were analysed, equating to 10 years of weekly sampling. In addition, two more concentrates, one from the coast of Fife in Scotland and the other from the coast of Lossiemouth in England were also screened.

6.2.2. Genomic DNA isolation and purification from four new coccolithoviruses

The concentration of the virus lysates, the purification and extraction of their DNA for the microarrays, PCRs and full genome sequencing, was as described in Chapter 2.3.5. Before the microarray analysis the DNA was quantified via a Nanodrop and its quality assessed on a 2 % agarose gel electrophoresis. Then half of the DNA was used for the microarray analysis and the other half was sent for next generation 454 genome sequencing at the NBAF (NERC Biomolecular Facilities) centre in Liverpool, UK. The results of the latter are not shown in this chapter as at the time of writing these samples were still in the sequencing centre and the results were pending.

6.2.3. Coccolithovirus microarray chips

The microarray chips for the analysis of coccolithoviruses were designed by Allen et al. (2007) and were also previously used in the study by Pagarete et al. (2011). Briefly, the arrays were constructed based on the complete genome sequence of the coccolithovirus EhV-86. Each array chip included a total of 3571 gene probes of which 2271 matched *E. huxleyi* ESTs and 1300 matched EhV-86 and EhV-163 genomic sequences (corresponding to at least two probes per predicted EhV CDS). Each array chip contained five technical replicates (i.e. five meta-grids) with seven subgrids on each consisting of 12 rows and 13 columns of spots representing different gene probes.

6.2.4. Coccolithovirus genomic DNA labelling

The labelling of the template DNA was as described by Allen et al. (2007). Briefly, the extracted and purified genomic DNA of the four new coccolithoviruses was used as template in random-primed polymerization reactions by directly incorporating fluorescent analogues of dCTP (Cy3-dCTP). In a reaction volume of 30 µL, 2.5 µg of random Promega hexamers were mixed with 500 ng of template DNA, denatured at 95°C for 2 min and then cooled on ice. Then 4 µL of × 10 Klenow buffer (NEB), 1 µL of dNTP mix (2.5 mM dGTP, dATP and dTTP), 1 µL of Amersham Cy3-dCTP and 20 Units of Klenow (large fragment of DNA polymerase I, exonuclease-free; NEB) were added to the reaction mixtures. The mixtures were then incubated at 37°C for 16 h, after which the reactions were stopped by their heating at 70°C for 10 min. Finally they were cleaned and purified by using a Centricon column (YM30, Millipore).

6.2.5. Coccolithovirus DNA microarray hybridization

Labelled DNA was hybridized to the EhV-86 based microarray slides as previously described by Allen et al. (2007). Briefly, before hybridization, the slides were incubated for 1 h at 65°C in a prehybridization solution (× 3 SSC, 0.2% SDS, 1% BSA). Then the slides were rinsed in distilled water followed by isopropanol and then spun down in a centrifuge at 200 g. Then to the samples were added 7.5 µL of × 20 SSC and 1 µL of 10% SDS in a total volume of 50 µL. They were then denatured for 2 min at 95°C and left to cool to 65°C. Finally the samples were applied on to the microarray slides, covered with an Erie Scientific Lifter Slip (22 × 50 mm), placed for hybridisation at 65°C for 12 h, after which they were washed in 1 × SSC, 0.1% SDS, and then dried by centrifugation at 200 g.

6.2.6. Acquisition, analysis and storage of data

Microarray slides were scanned using an Axon GenePix 4100A Microarray scanner. Fluorescent spot intensities were quantified using the ImaGene 9 (BioDiscovery) software. For each array, a threshold intensity was chosen in order to indicate the occurrence of positive hybridization of the template DNA to the microchip probes on the slides, ranging from 700-900 (arbitrary units). The positive hybridization of the template CDSs (or DNA coding sequences) of EhV-18, EhV-145, EhV-156, and EhV-164 was represented by a single spot on each and one of the five meta-grids on four subsequent array slides. Only probes that had signals above the threshold (i.e. successfully hybridised) in three out of the five replicated spots were considered as a positive presence of the corresponding CDS. The absence or/and the lack of detection of a particular CDS in a given strain did not definitively indicate its absence, only a

lack of successful hybridisation which could also be caused by variation in primary DNA sequence, inserts or deletions.

6.2.7. PCR fingerprinting of coccolithoviruses

Genomic DNA was also subjected to PCR fingerprinting by using 96 primer pairs (Section 2.3.1, Table 2.3), that were previously designed to amplify EhV-86 and EhV-163 genes in a shotgun survey for the identification of novel functions for biotechnological exploitation (Dr Mike Allen, personal communications). The PCR was conducted in 96 well plates in a final volume of 50 µL reaction mixes as previously described in Section 2.3.2.1.1. The cycle conditions included an initial denaturation step at 95°C for 3 min, followed by 34 cycles at 95°C, 58°C and 74°C for 30, 60 and 90 sec respectively, and a final cycle at 95°C, 58°C and 74°C for 30 sec, 5 min and 5 min respectively.

6.3. RESULTS

Four potentially new coccolithoviruses were isolated as a result of screening the L4 concentrates time series. The genomic DNA of these isolates was screened on an array that was previously designed based on the full genome sequence of EhV-86 and a PCR amplification of homologous genes within these viruses was performed by using a set of primers designed for the amplification of gene fragments of EhV-86. The EhV-86 array chips were used here as references and hence the hybridisation principle of the array was employed here only to show the presence of genes that are similar to those in the genome of EhV-86 in a presence/absence manner.

6.3.1. Sample screening and virus isolation

Out of more than 300 samples screened against an *E. huxleyi* CCMP 2090 host only four lysed the host and therefore exhibited infection during the initial well plate analysis. All four were successfully diluted to extinction through three rounds of infection (complete lysis of the host culture occurred within one week), concentrated and purified by a CsCl gradient as described in Chapter 2.3.5 (Figure 6.1) and successfully produced primary lysate stocks (established by AFC as ~2-4 × 10^8 VLPs mL^{-1}) for future full genome sequencing, microarray analysis and PCR fingerprinting. The new coccolithoviruses were named EhV-18, EhV-145, EhV-156 and EhV-164 and their source of isolation can be seen in Section 2.1.5, Table 2.3.

6.3.2. Genomic DNA extraction and assessment

The extracted DNA from the purified concentrated lysates of the four new viruses was visualised by a 1% agarose gel electrophoresis (Figure 6.2) and quantified by a NanoDrop, to reveal that each DNA sample had ~60 ng μL^{-1}. Half of this DNA was

sent to a sequencing centre in Liverpool for complete genome sequencing and the other half was used for the microarrays and PCR fingerprinting.

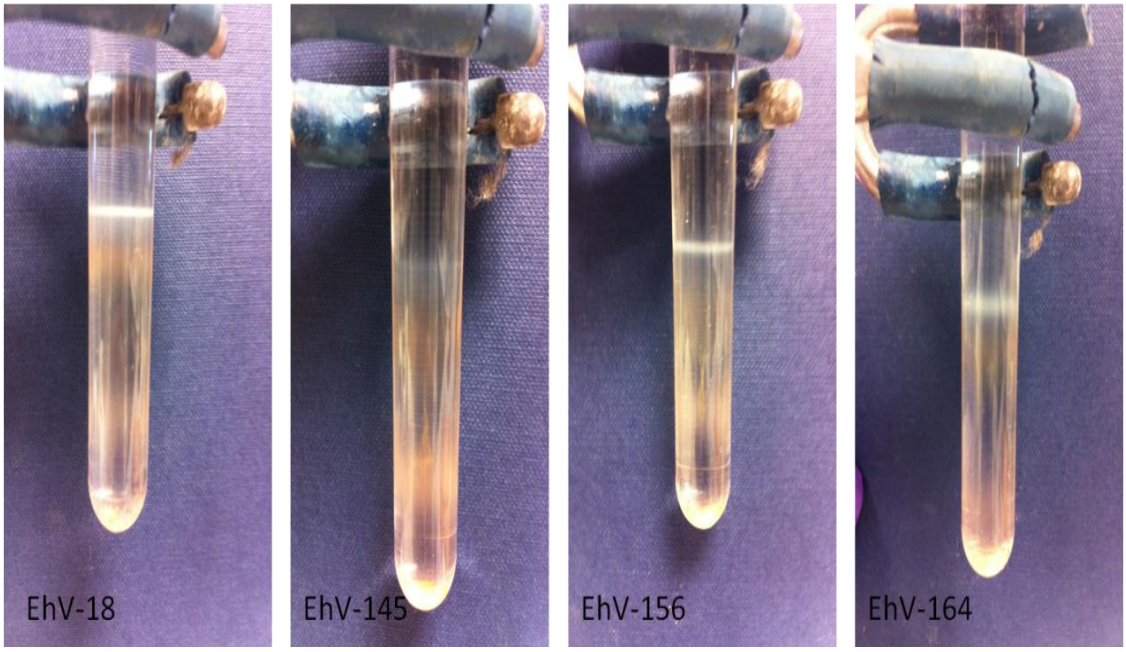


Fig. 6.1. CsCl gradient from four lysates of the newly isolated coccolithoviruses. The white band in each column is the concentrated virus fraction that was used for total genomic DNA extraction.

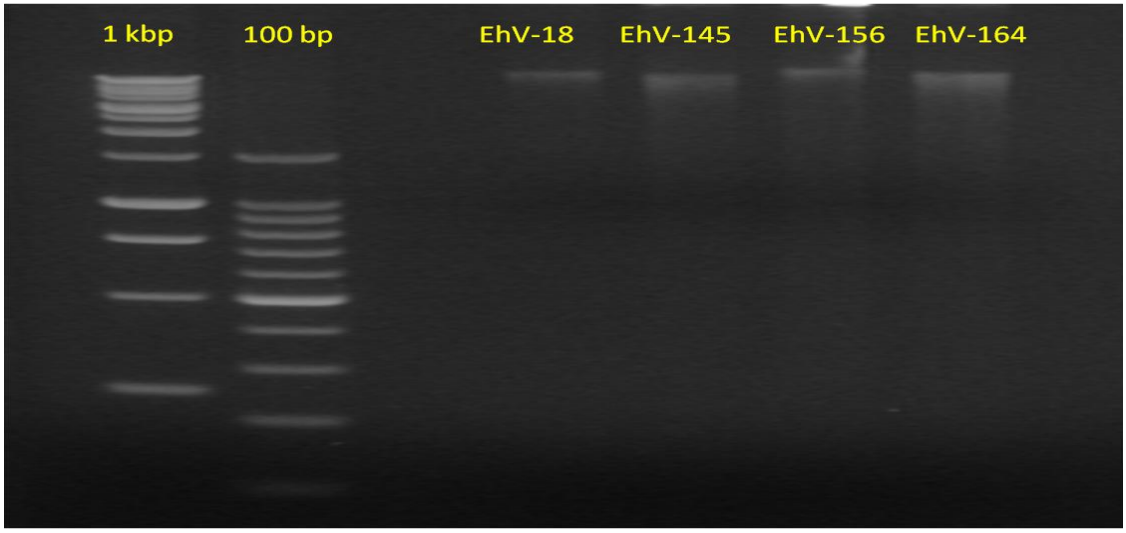


Fig. 6.2. Visual inspection of the isolated DNA from the lysate concentrates of EhV-18, EhV-145, EhV-156 and EhV-164.

6.3.3. Microarray analysis

Successful DNA hybridizations to the EhV-86 arrays were obtained for the newly isolated coccolithoviruses (Figure 6.3). EhV-18, EhV-145, EhV-156 and EhV-164 hybridised to 72, 86, 76 and 272 EhV-86 CDSs probes respectively (Table 6.2). 170

CDSs gave negative hybridisations (in at least three out of the five metagrid replicates) and are presumably either absent or sufficiently different in sequence in the strains tested. Of the 472 genes of EhV-86, 20 gave positive hybridisations in all four isolates, including those genes with a known function such as ehv058 (Superfamily II DNA/RNA helicases, SNF2 family), ehv061 (putative fatty acid desaturase), ehv128 (ERV1/ALR family protein), ehv361 (putative serine protease) and ehv463 (predicted RNA-binding protein homologous to eukaryotic snRNP).

The four strains exhibited three different profiling groups based on the positive and negative detection of hybridised CDSs, with EhV-18 and EhV-156 exhibiting the most similar hybridisation profiles. The hybridisation profile of EhV-145 was more similar to the profile of EhV-164 than the other two strains, with the latter exhibiting the highest similarity to the reference strain EhV-86 upon which the array was designed.

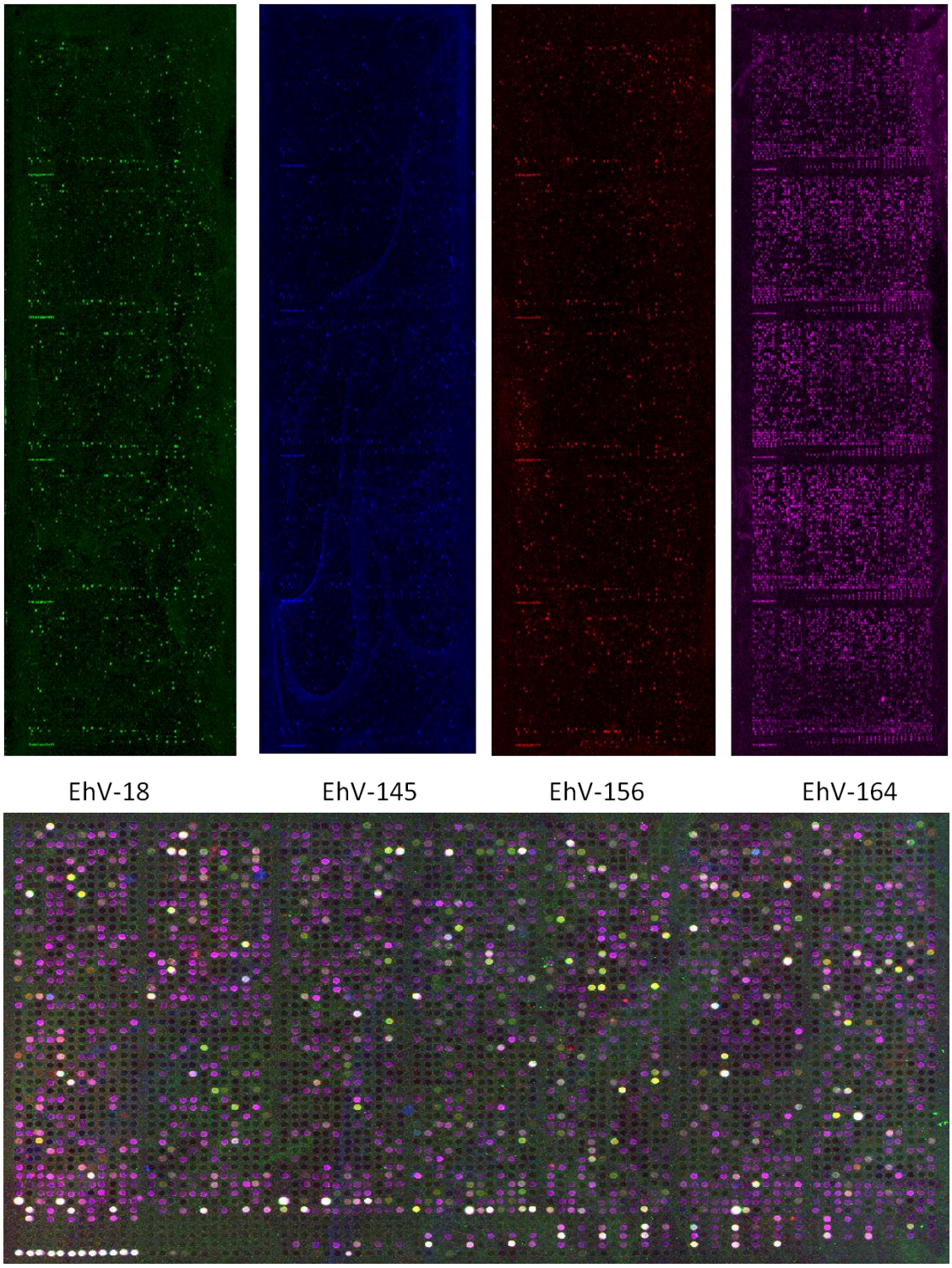


Fig. 6.3. Positive hybridisations of the four new coccolithovirus isolates shown as either green, blue, red or purple spots (top), and an example of one metagrid combining an overlay of the four (bottom). In the latter bright yellow spots are genes that exhibited hybridisation in all four strains.

With regards to the genes involved in the *de novo* sphingolipid biosynthesis pathway, ehv031 (sterol desaturase) and ehv050 (serine palmitoyltransferase) were only

detected in EhV-164, ehv077 (transmembrane fatty acid elongation protein) was detected only in EhV-156, and ehv079 (lipid phosphate phosphatase) was not detected at all in any of the strains. Furthermore, the gene encoding for a putative phosphate permease was also not detected in any of the strains, whereas the gene encoding for a putative sialidase was detected only in EhV-145 and EhV-164.

6.3.4. PCR fingerprinting with EhV-86 primers

The PCR profiling of the four coccolithoviruses revealed in support of the microarray data that EhV-145 and EhV-164 were the most similar isolates to the reference EhV-86, with the highest amount of positive amplifications; i.e. many of the positively hybridised genes on the arrays also showed positive amplifications with the PCR fingerprinting in those two strains (Figure 6.4). In addition, positive amplification was also observed of several gene fragments of an unknown function in EhV-156 and EhV-18 that otherwise showed no array hybridisation (Table 6.1).

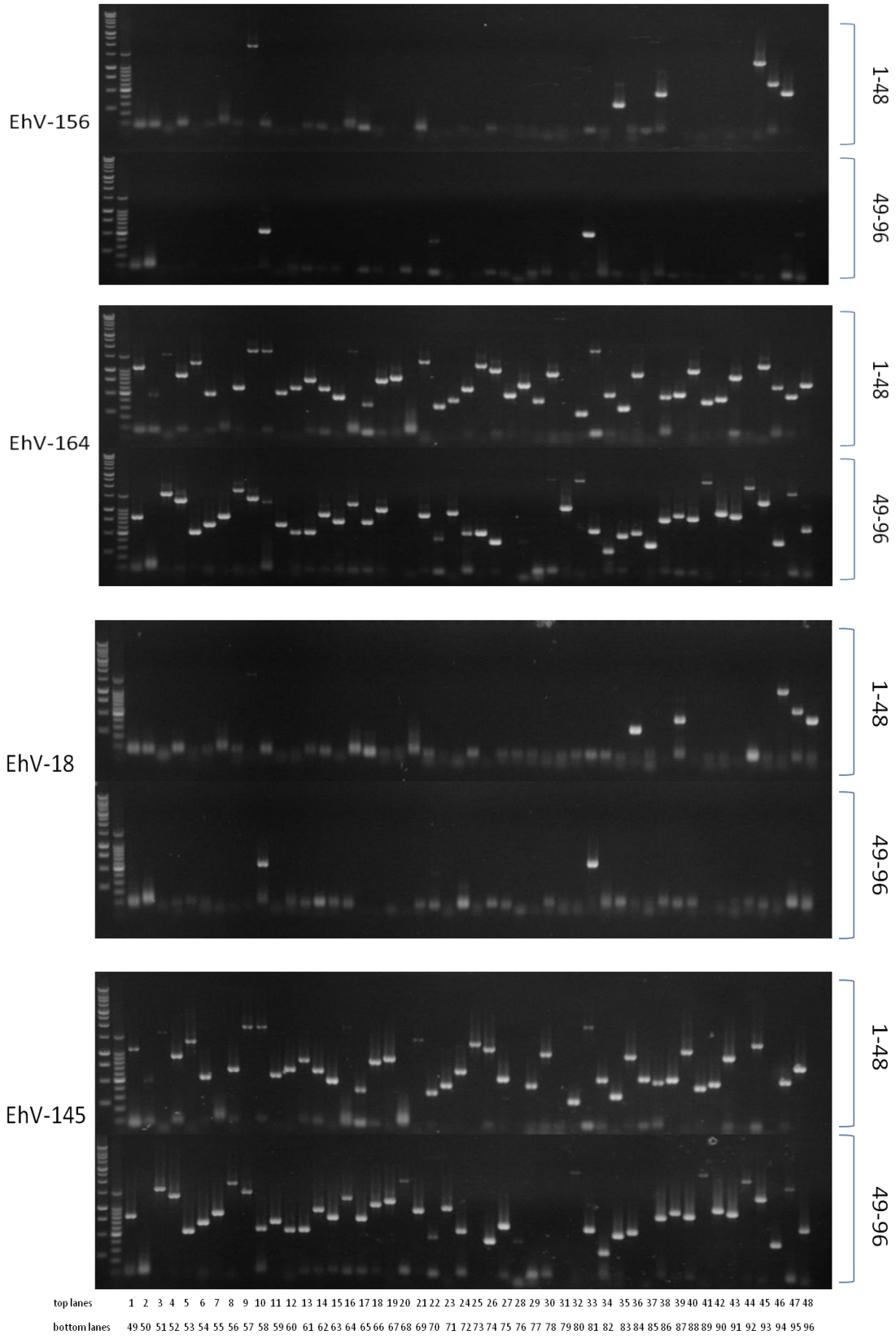


Fig. 6.4. PCR fingerprints of four newly isolated coccolithoviruses on a 1% agarose gel electrophoresis. The first two lanes to the left of each image are a 1 kbp and 100 bp molecular weight markers and the rest of the lanes (left to right) are the 96 amplified products, with the top gel in each image corresponding to products with primer sets 1-48 and the bottom gel primers corresponding to sets 49-96. Detailed information of the primers used in each lane can be seen in Table 2.5, Section 2.3.1.

Table 6.1. Positive hybridisation or/and detection by PCR amplification of 468 CDSs in the genomes of four newly isolated viruses, clustered into three profiles based on their similarity to each other and to the reference genome of EhV-86, with EhV-18 being the most different and EhV-164 the most similar to EhV-86. Light grey: CDSs detected only by the array, dark gray: CDSs detected by only PCR, and black: CDSs detected by both the array and the PCR.

CDS	EhV-18	EhV-156	EhV-145	EhV-164	CDS	EhV-18	EhV-156	EhV-145	EhV-164
ehv001					ehv055				
ehv002					ehv056				
ehv003					ehv057				
ehv004					ehv058				
ehv006					ehv059				
ehv007					ehv060				
ehv008					ehv061				
ehv010					ehv062				
ehv011					ehv063				
ehv012					ehv064				
ehv015					ehv065				
ehv017					ehv067				
ehv018					ehv070				
ehv019					ehv072				
ehv020					ehv073				
ehv022					ehv075				
ehv023					ehv077				
ehv024					ehv078				
ehv026					ehv080				
ehv027					ehv081				
ehv028					ehv082				
ehv029					ehv084				
ehv030					ehv087				
ehv031					ehv092				
ehv032					ehv093				
ehv033					ehv095				
ehv034					ehv098				
ehv035					ehv101				
ehv037					ehv103				
ehv038					ehv104				
ehv039					ehv105				
ehv040					ehv106				
ehv041					ehv107				
ehv042					ehv109				
ehv043					ehv111				
ehv044					ehv113				
ehv045					ehv114				
ehv046					ehv115				
ehv047					ehv116				
ehv050					ehv118				
ehv053					ehv119				
ehv054					ehv121				

CDS	EhV-18	EhV-156	EhV-145	EhV-164	CDS	EhV-18	EhV-156	EhV-145	EhV-164
ehv122					ehv185				
ehv124					ehv188				
ehv125					ehv189				
ehv127					ehv190				
ehv128					ehv191				
ehv129					ehv193				
ehv130					ehv194				
ehv131					ehv196				
ehv133					ehv197				
ehv136					ehv199				
ehv138					ehv200				
ehv139					ehv201				
ehv140					ehv202				
ehv141					ehv203				
ehv142					ehv205				
ehv143					ehv206				
ehv145					ehv207				
ehv146					ehv208				
ehv147					ehv209				
ehv150					ehv210				
ehv151					ehv211				
ehv152					ehv213				
ehv153					ehv214				
ehv154					ehv215				
ehv156					ehv216				
ehv158					ehv218				
ehv160					ehv219				
ehv161					ehv220				
ehv162					ehv222				
ehv163					ehv224				
ehv164					ehv225				
ehv166					ehv226				
ehv168					ehv227				
ehv170					ehv228				
ehv172					ehv229				
ehv174					ehv230				
ehv175					ehv231				
ehv176					ehv233				
ehv177					ehv234				
ehv178					ehv235				
ehv180					ehv238				
ehv183					ehv239				
ehv184					ehv240				

CDS	EhV-18	EhV-156	EhV-145	EhV-164	CDS	EhV-18	EhV-156	EhV-145	EhV-164
ehv241					ehv293				
ehv242					ehv294				
ehv243					ehv295				
ehv244					ehv296				
ehv245					ehv297				
ehv246					ehv298				
ehv247					ehv299				
ehv248					ehv300				
ehv251					ehv301				
ehv252					ehv302				
ehv255					ehv303				
ehv256					ehv304				
ehv258					ehv305				
ehv259					ehv306				
ehv260					ehv307				
ehv261					ehv308				
ehv262					ehv309				
ehv263					ehv311				
ehv264					ehv312				
ehv265					ehv313				
ehv266					ehv314				
ehv267					ehv315				
ehv268					ehv316				
ehv269A					ehv317				
ehv269					ehv318				
ehv271					ehv320				
ehv272A					ehv321				
ehv272					ehv322				
ehv273					ehv323				
ehv274					ehv324				
ehv275					ehv325				
ehv279					ehv326				
ehv280					ehv327				
ehv281					ehv331				
ehv282					ehv334				
ehv283					ehv335				
ehv284					ehv340				
ehv286					ehv344				
ehv287					ehv345				
ehv288					ehv346				
ehv289					ehv347				
ehv290					ehv348				
ehv291					ehv349				

CDS	EhV-18	EhV-156	EhV-145	EhV-164	CDS	EhV-18	EhV-156	EhV-145	EhV-164
ehv350					ehv411				
ehv351					ehv412				
ehv352					ehv413				
ehv353					ehv417				
ehv354					ehv424				
ehv356					ehv425				
ehv357					ehv426				
ehv358					ehv429				
ehv360					ehv430				
ehv361					ehv431				
ehv363					ehv432				
ehv364					ehv433				
ehv365					ehv434				
ehv370					ehv435				
ehv371					ehv437				
ehv373					ehv439				
ehv374					ehv440				
ehv378					ehv441				
ehv379					ehv442				
ehv380					ehv443				
ehv381					ehv444				
ehv384					ehv445				
ehv385					ehv446				
ehv386					ehv447				
ehv387					ehv448				
ehv388					ehv450				
ehv389					ehv451				
ehv390					ehv452				
ehv391					ehv453				
ehv392					ehv455				
ehv393					ehv457				
ehv394					ehv460				
ehv395					ehv461				
ehv396					ehv462				
ehv397					ehv463				
ehv398					ehv464				
ehv399					ehv465				
ehv400					ehv466				
ehv401					ehv467				
ehv402					ehv468				
ehv403									
ehv407									
ehv410									
Profile	**Prof 1**		**Prof 2**	**Prof 3**		**Prof 1**		**Prof 2**	**Prof 3**

6.4. DISCUSSION

In this chapter the genomic diversity of newly isolated coccolithoviruses (strains EhV-18, EhV-145, EhV-156 and EhV-164) was shown through two sequencing independent molecular techniques: PCR and microarrays. Each method has its own technical limitations; i.e. they are not as accurate as a total genomic sequencing approach. However, together they offer an efficient and rapid way to characterise coccolithovirus genomic diversity in an economical fashion. While the arrays were useful in categorising the new strains into separate profiles, they were biased towards strains most similar to the reference genome EhV-86 for which the microarrays were designed, evident by the fact that more EhV-145 and EhV-164 gene fragments hybridised to the arrays than EhV-18 and EhV-156 ones. The oligo probes on the arrays were just 75 bases long and if significant variation occurred in this region, hybridisation efficiency would have been affected. As a consequence even if the genes that gave a negative hybridisation did exist in the genomes of the newly isolated viruses, a sequence identity of <90% would have resulted in lack or partial hybridisation to the arrays (Allen and Wilson, 2006, Allen et al., 2007). To explain this further, a sequence that was different in eight or more bases would have not been detected by the arrays.

Nevertheless, the fact that so many EhV-18 and EhV-156 CDSs were not detected by the array suggests that these two viruses are genetically very different from the rest of the coccolithovirus strains studied in this thesis. What is even more surprising is that the geographical origin of isolation of these two viruses is the same as the origin of isolation of EhV-86 and yet they are so different, while EhV-145 and EhV-164 are from a much northern location and yet are more similar to EhV-86. This can be seen

as a strong evidence for the "everything is everywhere, but the environment selects" hypothesis and also as an additional evidence of the enormous diversity hidden within the *Coccolithoviridae* family. In this theory, the Dutch microbiologist Bass Becking originally suggested for the first time that closely related microbes have the potential to be distributed globally and that in particular environments the reason they are often not seen is because they are in very low numbers due to specific selection pressures imposed by the environment (i.e. most of the microbial biodiversity is hidden) or and possibly due to the limits in our ability to detect them with available techniques (de Wit and Thierry Bouvier, 2006). However due to the lack of environmental data on the isolation sites of some of these strains and the current lack in understanding of the impact environmental triggers (and selection pressures) have in shaping coccolithovirus communities it would be difficult to concur whether this hypothesis stands in the case of the virus diversity and its comparison in this chapter.

Many genes that were "missed" by the arrays were picked up by PCR, in particular those within strains EhV-145 and EhV-164. The products targeted by the primers for the specific genes tested ranged from 200-1700 bp and were therefore affected less by any substitutions, insertions or deletions outside the priming regions. Although the PCR products were not purified, cloned and sequenced, and therefore the size of the products were estimated only visually, it was a useful method that validated the results of the arrays. The limitation of this method comes from the availability of primer sets (which was limited to just 96 genes), and the target virus upon which they were designed. A brief inspection of the amplification potential of the 96 primers used in this study on the genome of EhV-202, another very different coccolithovirus (based on the data in Chapter 3), indicates that none of these PCR primers would have

resulted in a successful amplification of gene fragments found on the genome of EhV-202. This indicates that EhV-18 and EhV-156 are most likely as diverse and distinct as EhV-202.

6.4.1. Lack of hybridisation: true or false-negative?

Currently all fully sequenced coccolithoviruses were shown to have the four genes encoding for the *de novo* sphingolipid biosynthesis pathway mentioned throughout this thesis (Wilson et al., 2005; Nissimov et al., 2011a, 2011b, 2012a, 2012b; Pagarete et al., 2013). Furthermore these genes are expressed during infection and are therefore considered as an important feature during *E. huxleyi*-coccolithovirus dynamics (Pagarete et al., 2009; Vardi et al., 2009, 2012). Hence it is somewhat surprising that none of these genes were detected in EhV-18 and EhV-156, and that only some of these genes were detected by the microarray in EhV-145 and EhV-165. Several explanations can be put forward. The first is that these genes are not as crucial to coccolithovirus infection and host demise as previously suggested, evident by the fact that *E. huxleyi* cultures could be successfully infected by these latter virus strains. However given the amount of literature that demonstrates the importance of sphingolipid biosynthesis genes during infection it could simply be a matter of methodological bias during one or more of the hybridisation steps or that the arrays that were used in this study have exceeded their half life and were partially damaged.

Another more likely explanation could be that these strains were more similar to an ancestral coccolithovirus in which only part of this pathway has been acquired from the host. Further evidence for this comes from samples collected during the AMT-20 cruise (see Chapter 5) in which there was no detection of the SPT gene in any of the

samples collected on the Atlantic transect, even in samples where there was indeed a positive detection of EhV genotypes. A gradual acquisition of these genes during the *Coccolithoviridae* evolutionary history and the possible substantial divergence of coccolithoviruses at the nucleotide might explain their lack of detection here and in the Atlantic Ocean samples analysed in Chapter 5. Although not fully assembled and annotated yet, some preliminary data from the sequencing of EhV-18 in the sequencing centre in Liverpool is already available (not shown in this study). It suggests that the SPT encoded enzyme in EhV-18 is encoded by to separate genes, one encoding for LCB2 and the other encoding for LCB1. If the two protein subunits in the SPT enzyme are yet to be linked in EhV-18 and other strains than this can also explain the lack of their detection by the methods applied here and in Chapter 5.

A possible future expansion of this work should undoubtedly be the design of primers for the detection of all the genes involved in the *de novo* coccolithovirus encoded sphingolipid biosynthesis pathway. If applied on samples from a sedimentary record in conjunction with genes for EhV and host phylogeny, in a similar fashion to the approach undertaken by Coolen in the Black Sea (Coolen, 2011), it would be possible to discover when these genes were acquired, which one of the four was acquired first, and possibly gain an insight into what evolutionary selection pressure drove their acquisition from their host in the first place.

6.5. CONCLUSIONS

The fingerprints of the newly isolated strains highlighted the potential of using standard techniques such as PCR and microarrays in conjunction for the rapid assessment of viruses isolated from a variety of water samples originating in distinct geographical settings, when funds are limited and equipment for these methods is within a researcher's reach. The presence of a particular gene was confirmed by the positive hybridisation to an oligo probe on the array and this was the most important result that one can take home from this fingerprinting effort. In contrast, the absence of genes cannot be confirmed with the methods used here as negative hybridisations do not necessary mean a lack of a particular gene; i.e. the absence of evidence in this case was not evidence of absence. Full genomic sequencing will ultimately settle the matter definitively. In Chapter 3 was shown that each and every one of the seven newly sequenced genomes harbours a set of unique genes distinct to all others. However, novel genes within these four new coccolithoviruses could not be detected by the arrays, as the chips and oligo probes were designed based upon the known gene complement of EhV-86. Nevertheless, both approaches confirmed that strains EhV-18 and EhV-156 in particular are likely to be remarkably different to all known coccolithoviruses to date, highlighting the enormous and the very much still undiscovered genetic diversity hidden in the genomes of marine viruses The full genomic sequencing of these viruses that is currently in progress will reveal the extent to which these viruses are different from others in culture and reveal how diverse these new viruses actually are.

7. SUMMARY AND FUTURE STUDIES

In order to learn more about these fascinating coccolithoviruses (or EhVs), a three step approach was undertaken. The first aimed at providing phenotypic information with regards to the "behaviour" of different coccolithoviruses through infection dynamics experiments of several EhV strains with their host. During this part, techniques such as analytical flow cytometry (for host enumeration and virus production measurements) and qPCR (for absolute DNA copy numbers of different viruses as an indicator for virus succession post and during infection) were applied. It was shown that coccolithoviruses characterised by a somewhat different genomic potential can compete with each other and display distinct infection dynamic strategies. This discovery is crucial as knowledge of the rate of production of infectious viruses is essential to our current understanding of how these viruses influence the extent of host reduction, and hence, the flow of dissolved organic matter (such as phosphorus, nitrogen, and carbon) available to fuel microbial food webs in the oceans. This work demonstrated how the evolutionary diversity of coccolithovirus genotypes has the potential to affect infection dynamics. Hence, here for the first time was shown the distinct differences in the rate of infection between different coccolithovirus strains that were isolated at different times from a single geographical location and that their rate of production is dependent on competitive interactions within a coccolithovirus community. Moreover, the data generated is of great importance as it can be integrated generally into marine-ecosystem models and specifically into food-web models in the future in order to discover whether the flow of nutrients (and the amount of nutrients) will be available to lower or higher trophic levels in a particular marine environment following virus infection. The phenotypic differences of diverse

coccolithovirus strains unveiled in this thesis are crucial as currently most models incorporate only a single virus and do not take into account that these viruses can affect their hosts differently.

Although traditional laboratory experiments are important in "behavioural" studies of viruses, they provide only a glimpse of the invisible genomic diversity of viruses and, as a consequence, their ability to successfully infect and co-evolve with their hosts. Darwinian evolution theory dictates that the phenotypic properties of biological entities derive from their genomic potential. Hence, the second aspect of this work was to use computational tools in the analysis of the genomes of viruses infecting *E. huxleyi*. This included the sequencing of seven EhV genomes, their annotation, and comparison to reference genomes in the public domain, and the microarray and PCR fingerprinting of four newly isolated coccolithoviruses. Fundamental differences in the genomic composition of EhV strains isolated from different geographical settings and also between strains isolated at the same location but at different times were found.

A key finding of this research was the discovery of a large proportion of genes potentially acquired from the host via horizontal gene transfer events (~13% of their genome) and the presence or absence of genes known to be important as virulence factors during infection in some viruses (i.e. sialidase). In addition all the fully sequenced genomes had the genes that code for the *de novo* sphingolipid biosynthesis pathway with the exception of the newly isolated viruses where microarray and PCR fingerprinting failed to detect some of these genes (the reason for which is currently unknown).

The last part of this research was to validate the laboratory-based and comparative genomics results to coccolithovirus diversity within natural systems. Indeed, the third (and final) aspect was to look at the aforementioned phenotypic and genotypic features in natural virus communities in the global ocean. Samples collected from the Western Channel Observatory time-series site (L4) in the English Channel during the last 15 years and samples collected previously from the North Sea were analysed. In addition, environmental samples were also obtained from the Atlantic Ocean in order to increase the spatial coverage of this study. EhV community fingerprinting was conducted by the use of traditional molecular biology techniques such as PCR, DGGE, and sequencing, with well-established molecular marker genes for EhV phylogeny (major capsid protein; MCP) and for infection/function (serine palmitoyltransferase; SPT). Indeed this was the first time that these two markers were used in conjunction for phylogenetic and functional diversity of EhVs on a large temporal and spatial scale. The results of the sample analysis were remarkable. Although successful amplification of the MCP gene (=EhV identity/phylogeny) was obtained for all (or most) samples, amplification of the SPT gene was only successful in a small number of samples, suggesting that some EhV populations had an altered SPT gene framework. This was particularly surprising given that all sequenced EhV strains to date have the SPT gene residing in their genomes. This raised fundamentally interesting questions with regards to the functional diversity of the virally encoded sphingolipid biosynthesis pathway, its importance during infection, and its possible absence or divergence in some strains.

To shed more light on the potential importance of the virus encoded sphingolipid pathway, the tertiary structure of the fold associated with the multi domain SPT

encoded protein from both laboratory coccolithovirus strains and the new environmental SPT sequences was modelled. The fold of the SPT protein (and its two subunits LCB1 and LCB2) is critically important for the binding of substrates such as myristoyl Co-A and/or palmitoyl Co-A to the active sites of this enzyme (Han et al., 2006; Vardi et al, 2009, 2012; Monier et al., 2009; Michaelson et al,, 2010; Bidle and Vardi, 2011). Strikingly, the modelled SPT protein fold was quite different among strains, with the SPT fold of North Sea sequences being more similar to that of Norwegian-derived EhV isolates, rather than those isolated in the English Channel. It is not currently clear however whether environmental cues in these different locations have influenced the evolution of this gene, the diversity seen in its 3D structure, or the actual acquisition of this pathway in the coccolithovirus evolutionary history.

Nearly half of all proteomes studied to date are dominated by multi-domain proteins or consist of several protein subunits. This number is proportionally higher in eukaryotes, so it was not surprising that the virus encoded SPT, which has its origins from its *E. huxleyi* eukaryotic host, is a multi-domain protein. What was particularly notable and perhaps surprising was the observed difference in the predicted fold of the two SPT subunits among EhV strains. In proteins, it is not unusual to see differences in the length and composition of the "linker" connecting two or more domains. The "linker" anchors the two domains together, establishing structural and functional assembly of multi-domain proteins, and it has been shown previously that altering the length of linkers connecting subunits can affect the stability of a protein, rates of folding and protein orientation (Coen, 1996; Cohen. 2011).

The intriguing and recent findings of an intricate "chemical warfare" and ever-present co-evolutionary arms race between the algal hosts and their viruses (Bidle et al., 2007, 2011; Vardi et al., 2009), the fact that small molecular parasites such as coccolithoviruses can impact on global processes such as climate regulation and the recirculation of carbon and other nutrients within marine environments (van Rijssel and Gieskes, 2002) have proved an ongoing motivation for the majority of work conducted in this thesis.

In the *E. huxleyi*-virus system one of the chemical agents driving this warfare, and as a consequence to a certain extent aspects of biogeochemical cycling in the ocean, are glycosphingolipids (GSLs), which are synthesised within the host cell and whose biosynthetic genes are encoded either by the host or by the virus (as a result of horizontal gene transfer events). As previously discussed the first and rate limiting enzyme in the sphingolipid biosynthesis pathway is the serine palmitoyltransferase (SPT) enzyme (Han et al., 2006). Thus by utilising its own altered SPT copy (and associated biosynthetic pathway), coccolithoviruses can hijack the lipid biosynthesis machinery of the cell, control the viral replication process and possibly manipulate the host cell by triggering and regulating the programmed cell death (PCD) pathway of the alga (Bidle et al., 2007). Therefore it is possible that it is the fold of the virus encoded SPT protein that determines its preference of myristoyl Co-A over palmitoyl Co-A co-factors, the production rate of virally encoded glycosphingolipids (vGSLs), and as a consequence, critically regulates the speed of infection displayed by different EhV virus strains.

7.1. Future Research

Perhaps expectedly for a PhD thesis, at the end of this research there are more unanswered questions than there were at the beginning. It is still unknown what more than 80% of the hypothetical genes in the EhV genomes code for and it is still unclear whether the ones that have a predicted function translate into proteins with a real ecological relevance. For instance, some of the EhV genomes have a gene with a predicted function for sialidase and phosphate permease, while others only have a partial sequence of these genes left (Nissimov et al., 2011a, 2011b, 2012a, 2012b; Allen et al., 2006b; Pagarete et al., 2013). Given that sialidase is an important virulence factor in other viruses (von Itzstein et al., 1993), and phosphate permease has been previously implicated as a scavenging mechanism for the acquisition of phosphate (Wilson et al., 1995), it is unknown what the potential presence or absence of these genes (and many other genes with an unknown function) would mean with regards to the infection dynamics of EhVs with their hosts.

Another key observation that remains unanswered is the ability of different EhV strains to infect their host and how this is related to host sensitivity/resistance and, more broadly, how it manifests on larger ecological scales. In addition, the exact physicochemical conditions that drive the selection of successful coccolithophore genotypes in the environment and in turn the composition and diversity of their co-occurring viruses are also unknown. There are niche specific adaptations of different coccolithophores, and not all species are found throughout the global ocean (Tyrrell and Merico, 2004; Langer et al., 2009; Charalampopoulou et al., 2011). With predicted future climate change scenarios it is inevitable that some species will do better than others. This, in turn, will influence the type of virus community that is

associated with the host “winners”. Viruses that are capable of using the host machinery for sphingolipid biosynthesis in the algal host will most likely proliferate. Hence it is important to know whether the fold of the SPT protein as displayed by different virus strains plays a function in the ability to infect and co-evolve alongside successful hosts.

Some of the future goals of researchers in the field of coccolithoviruses and their hosts should be to try and link molecular ecology and biochemistry with evolutionary processes and ecological success. By merging molecular, biochemical, and physiological techniques (PCR, cloning, sequencing, qPCR, flow cytometry, plaque assays and HPLC–MS), it should be possible to answer the following fundamental questions:

1) Does the fold of the SPT enzyme in different EhV strains play a role in the rate of vGSLs biosynthesis and the rate of the EhV infection?

2) Is this predicted fold representative of the true conformation, and does it impact the ability of the SPT enzyme to bind and process its substrate in different algal strains?

3) Do host populations use this inherent variability in the virally encoded SPT enzyme to evade or reduce infection?

4) Are variations in the predicted SPT fold influenced by environmental conditions, and if they are what is the significance of this fold to host cell fate and ultimately biogeochemical cycling?

5) Can it affect the rate of infection dynamics during blooms and can it consequently affect the rate of carbon export into the deep ocean and the recirculation of nutrients?

6) What is the relationship (if any) between the observed coccolithovirus genetic diversity and environmental physicochemical conditions in which they are found?

7) Is coccolithovirus evolution mainly driven by environmental cues or is it mainly influenced by the "need" to keep up with the evolution of their hosts?

In order to put the findings of this thesis into the context of future research a general hypothesis that combines aspects of all the future research points mentioned above can be seen in Figure 7.1. During a theoretical selection pressure scenario imposed by for instance physicochemical stress (perhaps a consequence of increased sea surface temperature), a specific coccolithophore genotype (C) survives and adapts to the new conditions better than A or AB (Figure 7.1). In turn, coccolithoviruses that have the right conformation and protein fold of their virally encoded SPT enzyme (CB or C) will efficiently utilize the myristoyl Co-A substrate within the host cells, produce vGSLs, and propagate faster than coccolithoviruses that have a slightly different active site and protein fold (A and AB). If all four viruses are present in the above scenario in equal ratios then the coccolithoviruses that have the A and AB SPT structure (i.e. "losers") will be outcompeted by those that have the SPT structure C and CB (i.e. "winners"), the result of which would be a more rapid turnover of carbon and nutrients. However, if A and AB are the most dominant and abundant ones they will not be able to propagate fast within the coccolithophore population (C), essentially allowing the host cells to take back control over their own sphingolipid biosynthesis machinery, the result of which would be a reduction in vGSLs, an increase in host GSLs, slower infection dynamics within a bloom, and reduced export of carbon and recycling of nutrients (a trend that will continue until coccolithoviruses with the C and CB SPT protein confirmation become the dominant genotypes). Hence,

in an ecological context, only coccolithoviruses that possess the most effective SPT configuration in relation to the corresponding hosts in a given environment will be the ones to propagate and influence carbon fluxes and nutrient recycling.

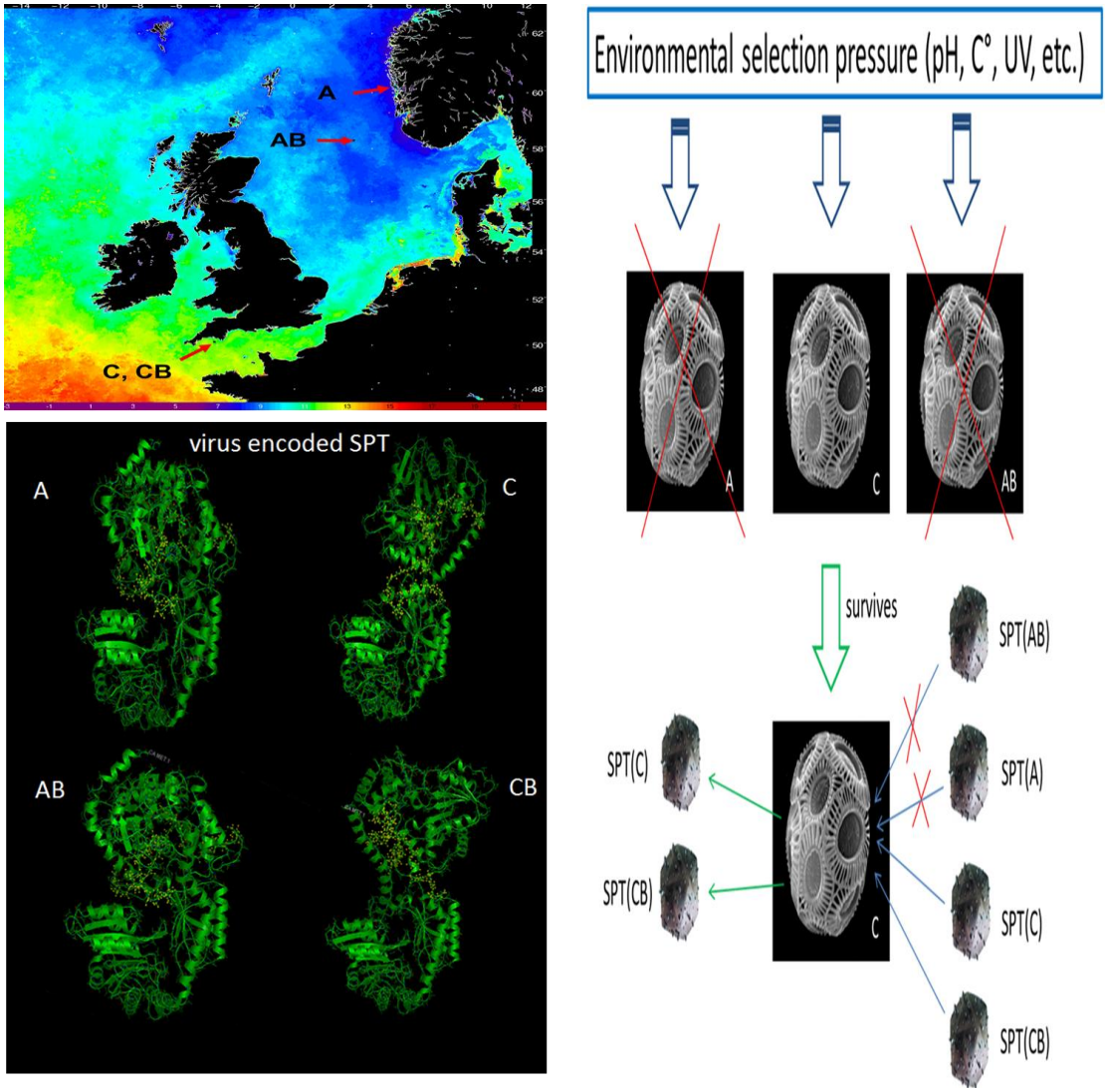


Fig. 7.1. (Upper left) Geographical locations (characterised by different mean annual sea surface temperatures, ranging from 21°C in red to 8°C in blue) from which different coccolithovirus genotypes and consequently different versions of the virus encoded SPT protein were obtained. (Bottom left): 3D models of four different structures of coccolithovirus encoded SPT proteins: A) laboratory strain EhV-99B1 isolated originally from a Norwegian Fjord, AB) sequence from a cruise in the North Sea, C) laboratory strain EhV-86 isolated originally from the English Channel, and CB) sequence from the L4 sampling station in the English Channel. Highlighted in yellow are the amplified regions of the "linker" between the two domains of the SPT protein: LCB1 and LCB2.

Finally, now that the genomic potential and diversity of a large number of coccolithovirus isolates is known, an important aspect of any future research would be to characterise the infection dynamics phenotypes of these isolates in the laboratory and relate this to potential differences in their expression profiles during infection. Furthermore, phylogentic and functional diversity studies of coccolithoviruses need to be expanded to other geographical locations as currently there is no data available on coccolithoviruses in the vast majority of the worlds' ocean where they are likely to persist; i.e. the Pacific Ocean, Southern Ocean, Mediterranean Sea, and Black Sea. The research in this thesis provides an important first step and a framework of marine research of the ecological and functional biodiversity in a marine algal-virus system of ecological importance, through the analysis of coccolithovirus genotypes, phenotypes and their ecological significance.

REFERENCES

Abrescia, N.G., Bamford, D.H., Grimes, J.M. and Stuart, D.I. Structure unifies the viral universe. *Annual Review of Biochemistry*, 2012, 81, 795-822.

Ackermann, H.W. and Heldal, M. Basic electron microscopy of aquatic viruses. In: Wilhelm, S.W., Weinbauer, M.G., and Suttle, C.A., eds., *Manual of Aquatic Viral Ecology*, ASLO, 2010, pp 182-192.

Allen, M.J., Schroeder, D.C, Holden, M.T.G. and Wilson, W.H. Evolutionary history of the Coccolithoviridae. *Molecular Biology and Evolution,* 2006a, 23, 86-92.

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

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

Allen, M.J. and Wilson, W.H. The coccolithovirus microarray: an array of uses. *Briefings in Functional Genomics & Proteomics*, 2006, 5, 273–279.

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

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

Allen, M.J., Lanzén, A. and Bratbak, G. Characterisation of the coccolithovirus intein. *Marine Genomics*, 2011, 4, 1–7.

Allen, M.J. and Wilson, W.H. Viruses of algae and Mimivirus. In: Acheson, N.A., ed., *Fundamentals of Molecular Virology*. 2nd edition. Wiley: Hoboken, New Jursey, USA, 2011, pp. 325-340.

Ayers, G.P. and Cainey, J.M. The CLAW hypothesis: a review of the major developments. *Environmental Chemistry*, 2007, 4, 366-374.

Azam, F., Fenchel, T. and Field, J. The ecological role of water-column microbes in the sea. *Marine Ecology Progress Series*, 1983, 10, 257–263.

Azam F., Fandino L., Grossart H. and Long R. Microbial loop: its significance in oceanic productivity and global change. *Rappid Communications of Internatioanl. Mer Médit,* 1998, 35, 1-3.

Bamford, D.H., Grimes, J.M. and Stuart, D.I. What does structure tell us about virus evolution? *Current Opinion in Structural Biology*, 2005, 15, 655-663.

Behbehani, A.M. The smallpox story: life and death of an old disease. *Microbiological Reviews*, 1983, 47, 455-509.

Bell, P.J. Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? *Journal of Molecular Evolution*, 2001, 53, 251–256.

Bidle, K.D., Haramaty, L., Ramos, J.B. and Falkowski, P. Viral activation and recruitment of metacaspases in the unicellular coccolithophore, Emiliania huxleyi. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104, 6049–6054.

Bidle, K.D. and Bender, S.J. Iron starvation and culture age activate metacaspases and programmed cell death in the marine diatom,Thalassiosirapseudonana. *Eukaryotic Cell*, 2008, 7, 223–236.

Bidle, K.D. and Vardi, A. A chemical arms race at sea mediates algal host-virus interactions. *Current Opinion in Microbiology*, 2011, 14, 449–457.

Bidle, K.D. and Kwityn, C.J. Assessing the role of caspase activity and metacaspase expression on viral susceptibility of the coccolithophore, Emiliania huxleyi (Haptophyta). *Journal of Phycology*, 2012, 48, 1079-1089.

Boeckel, B. and Baumann, KH. Vertical and lateral variations in coccolithophore community structure across the subtropical frontal zone in the South Atlantic Ocean. *Marine Micropaleontology*, 2008, 67, 255–273.

Bono, L.M., Gensel, C.L., Pfennig, D.W. and Burch, C.L. Competition and the origins of novelty: experimental evolution of niche-width expansion in a virus. *Biology Letters*, 2013, 9, rsbl 20120616.

Bourhy, H., Kissi, B., Audry, L., Smreczak, M., Sadkowska-Todys, M., Kulonen, K., Tordo, N., Zmudzinski, J.F. and Holmes, E.C. Ecology and evolution of rabies virus in Europe. *The Journal of General Virology*, 1999, 80, 2545–2557.

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

Bratbak, G., Levasseur, M., Michaud, S., Cantin, G., Fernández, E., Heimdal B.R. and Heldal M. Viral activity in relation to Emiliania huxleyi blooms: a mechanism of DMSP release? *Marine Ecology Progress Series*, 1995, 128, 133–142.

Brown, C.W. Global distribution of Coccolithophore blooms. *Oceanography*, 1995, 8, 59–60.

Brownlee, C. and Taylor, A.R. Calcification in coccolithophores: a cellular perspective. In: Thierstein, H.R. and Young, J.R., eds., *Coccolithophores: From the Molecular Processes to Global Impact*, Springer, Heidelberg, 2004. pp. 30-49.

Breitbart, M. and Rohwer, F. Here a virus, there a virus, everywhere the same virus? *Trends in Microbiology*, 2005, 13, 278–284.

Brussaard, C.P. D., Marie D. and Bratbak G. Flow cytometric detection of viruses. *Journal of Virological Methods*, 2000, 85, 175–182.

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

Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M-L. and Brüssow, H. Phage as agents of lateral gene transfer. *Current Opinion in Microbiology*, 2003, 6, 417–424.

Cărăuş I. The algae of Romania. Report in: *Studii şi Cercetări, Universitatea Bacău, Biologie 7*, 2002, pp. 1-694.

Carter, J. and Saunders, V. Virology: Principles and Applications. Wiley, Chichester: England, 2007, pp. 31-102.

Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M-A., Barrell, B.G. and Parkhill, J. ACT: the Artemis Comparison Tool. *Bioinformatics*, 2005, 21, 3422–3423.

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

Charalampopoulou, A., Poulton, A.J., Tyrrell, T. and Lucas, M.I. Irradiance and pH affect coccolithophore community composition on a transect between the North Sea and the Arctic Ocean. *Marine Ecology Progress Series*, 2011, 431, 25−43.

Charlson, R.J., Lovelock, J.E, Andreae, M.O. and Warren, S.G. Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature*, 1987, 326, 655-661.

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

Chiu, C.Y., Urisman, A., Greenhow, T.L., Rouskin, S., Yagi, S., Schnurr, D., Wright, C., Drew, W.L., Wang, D., Weintrub, P.S., Derisi, J.L. and Ganem, D. Utility of DNA microarrays for detection of viruses in acute respiratory tract infections in children. *The Journal of Paediatrics*, 2008, 153, 76–83.

Claverie, J-M. and Abergel, C. Mimivirus and its virophage. *Annual Review of Genetics*, 2009, 43, 49–66.

Cook, S.S., Whittock, L., Wright, S.W. and Hallegraeff, G.M. Photosynthetic pigment and genetic differences between two Southern Ocean morphotypes of *Emiliania huxleyi* (Haptophyta). *Journal of Phycology*, 2011, 47, 615-626.

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

Dandonneau, Y., Montel, Y., Blanchot, J., Giraudeau, J. and Neveux, J. Temporal variability in phytoplankton pigments, picoplankton and coccolithophores along a transect through the North Atlantic and tropical southwestern Pacific. *Deep Sea Research Part I: Oceanographic Research Papers*, 2006, 53, 689–712.

Danovaro, R., Corinaldesi, C., Dell'anno, A., Fuhrman, J.A., Middelburg J.J., Noble R.T. and Suttle C.A. Marine viruses and global climate change. *FEMS Microbiology Reviews*, 2011, 35, 1–42.

Darling, A.E., Mau, B. and Perna, N.T. ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *Public Library of Science One*, 2010, 5, e11147.

de Wit, R. and Bouvier, T. Everything is everywhere, but, the environment selects; what did Baas Becking and Beijerinck really say? *Environmental Microbiology*, 2006, 8, 755-758.

Duintjer, T.R.J., Pallansch, M.A., Chumakov, K.M., Halsey, N.A., Hovi, T., Minor, P.D., Modlin, J.F., Patriarca, P.A., Sutter, R.W., Wright, P.F., Wassilak, S.G.F., Cochi, S.L., Kim, J-H. and Thompson, K.M. Review and assessment of poliovirus immunity and transmission: synthesis of knowledge gaps and identification of research needs. *Risk Analysis*, 2013, 33, 606–646.

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

Dupont, J., Collins, M. and Lafite, R. Annual variations in suspended particulate matter within the Dover Strait. *Oceanologica Acta*, 1993, 16, 507–516.

Dyhrman, S. and Haley, S. Long serial analysis of gene expression for gene discovery and transcriptome profiling in the widespread marine coccolithophore Emiliania huxleyi. *Applied and Environmental Microbiology*, 2006, 72, 252–260.

Evans, C., Malin, G., Mills, G.P. and Wilson, W.H. Viral Infection of Emiliania Huxleyi (Prymnesiophyceae) Leads To Elevated Production of Reactive Oxygen Species. *Journal of Phycology*. 2006, 42, 1040–1047.

Eynaud, F., Giraudeau, J., Pichon, J-J. and Pudsey, C.J. Sea-surface distribution of coccolithophores, diatoms, silicoflagellates and dinoflagellates in the South Atlantic Ocean during the late austral summer 1995. *Deep Sea Research Part I: Oceanographic Research Papers*, 1999, 46, 451–482.

Falkowski, P.G., Fenchel, T. and Delong, E.F. The microbial engines that drive Earth's biogeochemical cycles. *Science*, 2008, 320, 1034-1039.

Favreau, D.J., Meessen-Pinard, M., Desforges, M. and Talbot, P.J. Human coronavirus-induced neuronal programmed cell death is cyclophilin d dependent and potentially caspase dispensable. *Journal of Virology*, 2012, 86, 81–93.

Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 1985, 39, 783–791.

Feng, Y., Warner, M.E., Zhang, Y., Sun, J., Fu, F-X., Rose, J.M. and Hutchins D.A. Interactive effects of increased pCO2, temperature and irradiance on the marine coccolithophore Emiliania huxleyi (Prymnesiophyceae). *European Journal of Phycology*, 2008, 43, 87–98.

Findlay, C.S., Young, J.R. and Scott F.J. Haptophyta: Order Coccolithophorales. In: Scott, F.J. and Marchant, H.J., ed., *Antarctic Marine Protists*. Canberra & Hobart: Australian Biological Resources Study (Australian Antarctic Division), 2005, pp 276-294.

Fitzgerald, L.A., Graves, M.V., Li, X., Feldblyum, T., Hartigan, J. and Van Etten, J.L. Sequence and annotation of the 314-kb MT325 and the 321-kb FR483 viruses that infect Chlorella Pbi. *Virology*, 2007a, 358, 459–471.

Fitzgerald, L.A., Graves, M.V., Li, X., Feldblyum, T., Nierman, W.C. and Van Etten, J.L. Sequence and annotation of the 369-kb NY-2A and the 345-kb AR158 viruses that infect Chlorella NC64A. *Virology*, 2007b, 358, 472–484.

Fitzgerald, L.A, Graves, M.V., Li, X., Hartigan, J., Pfitzner, A.J.P., Hoffart, E. and Van Etten, J.L. Sequence and annotation of the 288-kb ATCV-1 virus that infects an endosymbiotic chlorella strain of the heliozoon Acanthocystis turfacea. *Virology*, 2007c, 362, 350–361.

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

Frada, M.J., Bidle, K.D., Probert, I. and De Vargas, C. In situ survey of life cycle phases of the coccolithophore Emiliania huxleyi (Haptophyta). *Environmental Microbiology*, 2012, 14, 1558–1569.

Franklin, D.J., Brussaard, C.P.D. and Berges, J.A. What is the role and nature of programmed cell death in phytoplankton ecology? *European Journal of Phycology*, 2006, 41, 1–14.

Franklin, D., Steinke, M., Young, J., Probert, I. and Malin, G. Dimethylsulphoniopropionate (DMSP), DMSP-lyase activity (DLA) and dimethylsulphide (DMS) in 10 species of coccolithophore. *Marine Ecology Progress Series*, 2010, 410, 13–23.

Fuhrman, J.A. Marine viruses and their biogeochemical and ecological effects. *Nature*, 1999, 399, 541–548.

Fuhrman, J.A., Hewson, I., Schwalbach, M.S., Steele, J.A., Brown, M.V. and Naeem, S. Annually reoccurring bacterial communities are predictable from ocean conditions. *Proceedings of the National Academy of Sciences of the United States of America*, 2006, 103, 13104–13109.

Fuszard, M.A., Wright, P.C. and Biggs, C.A. Comparative quantitative proteomics of prochlorococcus ecotypes to a decrease in environmental phosphate concentrations. *Aquatic Biosystems*, 2012, 8: 7.

Gao, G. and Luo, H. The ubiquitin-proteasome pathway in viral infections. *Canadian Journal of Physiology and Pharmacology*, 2006, 84, 5–14.

Gattuso, J. and Buddemeier, R.W. Calcification and CO2. *Nature*, 2000, 407, 311–313.

Gimenes, M.V., Zanotto, P.M.D.A., Suttle, C.A., Da Cunha, H.B. and Mehnert, D.U. Phylodynamics and movement of Phycodnaviruses among aquatic environments. *The ISME Journal*, 2012, 6, 237–47.

Gogarten, J.P. and Townsend, J.P. Horizontal gene transfer, genome innovation and evolution. *Nature Reviews Microbiology*, 2005, 3, 679–687.

Grimsley N.H., Thomas R., Kegel J.U., Jacquet S., Moreau H. and Desdevises Y. Genomics of Algal Host – Virus Interactions. *Advances in Botanical Research*, 2012, 64, 344-382.

Guex, N. and Peitsch, M.C. SWISS-MODEL and the Swiss- Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis*, 1997, 18, 2714-2723.

Guo, J., Lapi, S., Ruth, T.J. and Maldonado, M.T. The effects of iron and copper availability on the copper stoichiometry of marine phytoplankton. *Journal of Phycology*, 2012, 48, 312-325.

Han, G., Gable, K., Yan, L., Allen, M.J., Wilson, W.H., Moitra, P., Harmon, J.M. and Dunn, T.M. Expression of a novel marine viral single-chain serine palmitoyltransferase and construction of yeast and mammalian single-chain chimera. *The Journal of Biological Chemistry*, 2006, 281, 39935–39942.

Hanson, RM, Prilusky, J, Renjian, Z, Nakane, T. and Sussman, J.L. JSmol and the next-generation web-based representation of 3D molecular structure as applied to Proteopedia. *Israel Journal of Chemistry*, 2013, 53, 207-216.

Hartshorn, M.J. AstexViewer: a visualisation aid for structure-based drug design. *Journal of Computer-Aided Molecular Design*, 2002, 16, 871–881.

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

Hecky, R.E. and Kilham, P. Nutrient limitation of phytoplankton in freshwater and marine environments : A review of recent evidence on the effects of enrichment. *Limnology and Oceanography*, 1988, 33, 796–822.

Hedrick, S.M., Ch'en, I.L. and Alves, B.N. Intertwined pathways of programmed cell death in immunity. *Immunological Reviews*, 2010, 236, 41–53.

Henn, M.R., Sullivan, M.B., Stange-Thomann, N., Osburne, M.S., Berlin, A.M., Kelly, L., Yandava C., Kodira C., Zeng Q., Weiand M., Sparrow T., Saif S., Giannoukos, G., Young, S.K., Nusbaum, C., Birren, B.W. and Chisholm, S.W. Analysis of high-throughput sequencing and annotation strategies for phage genomes. *Public Library of Science One*, 2010, 5, e9083.

Hinshaw, V.S., Olsen, C.W., Dybdahl-sissoko, N. and Evans, D. Apoptosis: a mechanism of cell killing by influenza A and B viruses. *Journal of Virology*, 1994, 68, 3667–3673.

Hladik, F. and McElrath, M.J. Setting the stage: host invasion by HIV. *Nature Reviews Immunology*, 2008, 8, 447–457.

Holligan, P.M., Viollier, M., Harbour, D.S., Camus, P. and Champagnephilippe, M. Satellite and ship studies of coccolithophore production along a continental shelf edge. *Nature*, 1983, 304, 339–342.

Honjo, S. Coccoliths: production, transportation and sedimentation. *Marine Micropaleontology*, 1976, 1, 65-79.

Hoofnagle, J.H., Nelson, K.E. and Purcell, R.H. Hepatitis E. *The New England Journal of Medicine*, 2012, 367, 1237–1244.

Hughes, M.T., Matrosovich, M., Rodgers, M.E., McGregor, M. and Kawaoka, Y. Influenza A viruses lacking sialidase activity can undergo multiple cycles of replication in cell culture, eggs, or mice. *Journal of Virology*, 2000, 74, 5206–5212.

Iglesias-Rodriguez, M.D., Schofield, O.M., Batley, J., Medlin, L.K. and Hayes, P.K. Intraspecific genetic diversity in the marine coccolithophore *Emiliania huxleyi* (Prymnesiophyceae): the use of microsatellite analysis in marine phytoplankton population studies. *Journal of Phycology*, 2006, 42, 526-536.

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

Janknegt, P.J., De Graafe, C.M., Van de Poll, W.H., Visser, R.J.W., Rijstenbil, J.W. and Buma, A.G.J. Short-term antioxidative responses of 15 microalgae exposed to excessive irradiance including ultraviolet radiation. *European Journal of Phycology*, 2009, 44, 525-539.

Jeanniard, A., Dunigan, D.D., Gurnon, J.R., Agarkova, I.V., Kang, M., Vitek, J., Duncan, G., McClung, O.W., Larsen, M., Claverie, J-M., Van Etten, J.L. and Blanc, G. Towards defining the chloroviruses: a genomic journey through a genus of large DNA viruses. *BMC Genomics*, 2013, 14: 158.

Jordan, R. and Chamberlain, A. Biodiversity among haptophyte algae. *Biodiversity and Conservation*, 1997, 6, 131-152.

Karl, D.M. and Tien, G. Temporal variability in dissolved phosphorus concentrations in the subtropical North Pacific Ocean. *Marine Chemistry*, 1997, 56, 77–96.

Karsenti, E., Acinas, S.G., Bork, P., Bowler, C., De Vargas, C., Raes, J., Sullivan, M., Arendt, D., Benzoni, F., Claverie, J-M., Follows, M., Gorsky, G., Hingamp, P., Iudicone, D., Jaillon, O., Kandels-Lewis, S., Krzic, U., Not, F., Ogata, H., Pesant, S., Reynaud, E.G., Sardet, C., Sieracki, M.E., Speich, S., Velayoudon, D., Weissenbach, J. and Wincker, P. A. Holistic approach to marine eco-systems biology. *Public Library of Science Biology*, 2011, 9, e1001177.

Kegel, J.U., Blaxter, M., Allen, M.J., Metfies, K., Wilson, W.H. and Valentin, K. Transcriptional host–virus interaction of Emiliania huxleyi (Haptophyceae) and EhV-86 deduced from combined analysis of expressed sequence tags and microarrays. *European Journal of Phycology*, 2010, 45, 1–12.

Keller, M.D., Bellows, W.K. and Guillard, R.R.L. Dimethyl sulphide production in marine phytoplankton. In: E.S. Saltzman and Cooper, W. J., ed., *Biogenic Sulphur in the Marine Environment*. ACS symposium series (393), American Chemical Society. Washingdton D.C., 1989, 167-182.

Kelley, L.A. and Sternberg, M.J.E. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols*, 2009, 4, 363–371.

Kirchman, D.L. Phytoplankton death in the sea. *Nature*, 1999, 398, 293–294.

Knopf, C.W. Evolution of viral DNA-dependent DNA polymerases. *Virus Genes*, 1998, 16, 47–58.

Koonin, E.V., Makarova, K.S. and Aravind, L. Horizontal gene transfer in prokaryotes: Quantification and Classification. *Annual Reviews in Microbiology*, 2001, 55, 709-742.

Krupovic, M. and Bamford, D.H. Virus evolution: how far does the double beta-barrel viral lineage extend? *Nature Reviews Microbiology*, 2008, 6, 941-948.

Krupovic, M. and Bamford, D.H. Double-stranded DNA viruses: 20 families and only five different architectural principles for virion assembly. *Current Opinion in Virology*, 2011, 1, 118-124.

Langer, G., Nehrke, G., Probert, J., Ly, J. and Ziveri, P. Strain specific responses of *Emiliania huxleyi* to changing seawater carbonate chemistry. *Biogeosciences*, 2009, 6, 2637−2646.

Larsen, J.B., Larsen, A., Thyrhaug, R., Bratbak, G. and Sandaa, R-A. Marine viral populations detected during a nutrient induced phytoplankton bloom at elevated pCO_2 levels. *Biogeosciences Discussions*, 2007, 4, 3961–3985.

Larsen, J.B., Larsen, A., Thyrhaug, R., Bratbak, G. and Sandaa, R. Response of marine viral populations to a nutrient induced phytoplankton bloom at different pCO 2 levels. *Biogeoscineces*, 2008, 5, 523–533.

La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., Merchat, M., Suzan-Monti, M., Forterre, P., Koonin, E. and Raoult, D. The virophage as a unique parasite of the giant mimivirus. *Nature*, 2008, 455, 100–104.

Levasseur, M., Michaud, S., Egge, J., Cantin, G., Nejstgaard, J.C., Sanders, R., Fernandez, E., Solberg, P.T., Heimdal, B. and Gosselin, M. Production of DMSP and DMS during a mesocosm study of an Emiliania huxleyi bloom: influence of bacteria and Calanus finmarchicus grazing. *Marine Biology*, 1996, 126, 609–618.

Li, Y., Lu, Z., Sun, L., Ropp, S., Kutish, G.F., Rock, D.L. and Van Etten, J.L. Analysis of 74 kb of DNA located at the right end of the 330-kb chlorella virus PBCV-1 genome. *Virology*, 1997, 237, 360–377.

Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F. and Chisholm, S.W. Transfer of photosynthesis genes to and from Prochlorococcus viruses. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101, 11013–11018.

Liu, H., Fu, Y., Jiang, D., Li, G., Xie, J., Cheng, J., Peng, Y., Ghabrial, S. A. and Yi, X. Widespread horizontal gene transfer from double-stranded RNA viruses to eukaryotic nuclear genomes. *Journal of Virology*, 2010, 84, 11876–11887.

Llewellyn, C.A., Evans, C., Airs, R.L., Cook, I., Bale, N. and Wilson, W.H. The response of carotenoids and chlorophylls during virus infection of Emiliania huxleyi (Prymnesiophyceae). *Journal of Experimental Marine Biology and Ecology*, 2007, 344, 101–112.

Luria, S.E. and Delbruck, M. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, 1943, 28, 491–511.

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

Malmstrom, R.R., Coe, A., Kettler, G.C., Martiny, A.C., Frias-Lopez, J., Zinser, E.R. and Chisholm, S.W. Temporal dynamics of Prochlorococcus ecotypes in the Atlantic and Pacific oceans. *The ISME Journal*, 2010, 4, 1252–1264.

Mannige, R.V. and Brooks, C.L. Periodic table of virus capsids: implications for natural selection and design. *Public Library of Science One*, 2010, 5, e9423.

Mao, C. and Obeid, L.M. Ceramidases: regulators of cellular responses mediated by ceramide, sphingosine, and sphingosine-1-phosphate. *Biochimica et Biophysica Acta*, 2008, 1781, 424–434.

Markine-Goriaynoff, N., Gillet, L., Van Etten, J.L., Korres, H., Verma, N., Vanderplasschen, A. Glycosyltransferases encoded by viruses. *Journal of General Virology*, 2004, 85, 2741–2754.

Markowitz, V.M., Mavromatis, K., Ivanova, N.N., Chen, I-M.A, Chu, K. and Kyrpides, N.C. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics*, 2009, 25, 2271–2278.

Marston, M.F., Pierciey, F.J., Shepard, A., Gearin, G., Qi, J., Yandava, C., Schuster, S.C., Henn, M.R. and Martiny, J.B.H. Rapid diversification of coevolving marine Synechococcus and a virus. *Proceedings of the National Academy of Sciences of the United States of America*, 2012, 109, 4544–4549.

Martinez, J.L. The role of natural environments in the evolution of resistance traits in pathogenic bacteria. *Proceedings in Biological Sciences*, 2009, 276, 2521–2530.

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

Martínez, J.M., Schroeder, D.C., Larsen, A., Bratbak, G. and Wilson, W.H. Molecular dynamics of Emiliania huxleyi and cooccurring viruses during two separate mesocosm studies. *Applied and Environmental Microbiology*, 2007, 73, 554–562.

Martiny, A.C., Huang, Y. and Li, W. Occurrence of phosphate acquisition genes in Prochlorococcus cells from different ocean regions. *Environmental Microbiology*, 2009, 11, 1340–1347.

Mayer, J.A. and Taylor, F.J.R. A virus which lyses the marine nanoflagellate Micrornonas pusilla. *Nature*, 1979, 281, 299-301.

McDaniel, L.D., Young, E., Delaney, J., Ruhnau, F., Ritchie, K.B. and Paul, J.H. High frequency of horizontal gene transfer in the oceans. *Science*, 2010, 330, 50.

Merrill, A.H. De novo sphingolipid biosynthesis: a necessary, but dangerous, pathway. *The Journal of Biological Chemistry*, 2002, 277, 25843–25846.

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

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

Monier, A., Welsh, R.M., Gentemann, C., Weinstock, G., Sodergren, E., Armbrust, E.V., Eisen, J. A. and Worden, A.Z. Phosphate transporters in marine phytoplankton and their viruses: cross-domain commonalities in viral-host gene exchanges. *Environmental Microbiology*, 2011, 14, 162-176.

Moore, T.S., Dowell, M.D. and Franz, B.A. Detection of coccolithophore blooms in ocean color satellite imagery: A generalized approach for use with multiple sensors. *Remote Sensing of Environment*, 2012, 117, 249–263.

Moreau, H., Piganeau, G., Desdevises, Y., Cooke, R., Derelle, E. and Grimsley, N. Marine prasinovirus genomes show low evolutionary divergence and acquisition of protein metabolism genes by horizontal gene transfer. *Journal of Virology*, 2010, 84, 12555–12563.

Mullis, K. and Faloona, F. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in Enzymology*, 1987, 155, 335–350.

Nissimov, J.I., Worthy, C.A, Rooks, P., Napier, J.A, Kimmance, S.A., Henn, M.R., Ogata, H. and Allen, M.J. Draft genome sequence of the Coccolithovirus Emiliania huxleyi virus 203. *Journal of Virology*, 2011a, 85, 13468–13469.

Nissimov, J.I., Worthy, C.A., Rooks, P., Napier, J. A., Kimmance, S.A., Henn, M.R., Ogata, H. and Allen, M.J. Draft genome sequence of the coccolithovirus EhV-84. *Standards in Genomic Sciences*, 2011b, 5, 1–11.

Nissimov, J.I., Worthy, C.A., Rooks, P., Napier, J. A., Kimmance, S.A., Henn, M.R., Ogata, H. and Allen, M.J. Draft Genome Sequence of Four Coccolithoviruses: Emiliania huxleyi Virus EhV-88, EhV-201, EhV-207, and EhV-208. *Journal of Virology*, 2012a, 86, 2896–2897.

Nissimov, J.I., Worthy, C.A., Rooks, P., Napier, J.A., Kimmance, S.A., Henn, M.R., Ogata, H. and Allen, M.J. Draft Genome Sequence of the Coccolithovirus Emiliania huxleyi Virus 202. *Journal of Virology*, 2012b, 86, 2380–2381.

Obeid, L.M., Linardic, C.M., Karolak, L.A. and Hannun, Y.A. Programmed cell death induced by ceramide. *Science*, 1993, 259, 1769–1771.

O'Brien, V. Viruses and apoptosis. *Journal of General Virology*. 1998, 79, 1833–1845.

Ogawa, H. and Tanoue, E. Dissolved organic carbon in oceanic waters. Journal of Oceanography, 2003, 59, 129–147.

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

Pagarete, A., Lanzén, A., Puntervoll, P., Sandaa, R-A., Larsen, A., Larsen, J.B., Allen, M.J. and Bratbak, G. Genomic sequence and analysis of EhV-99B1, a new coccolithovirus from the Norwegian fjords. *Intervirology*, 2013, 56, 60–66.

Parke, M. and Dixon, P.S. Check-list of British marine algae - third revision. *Journal of the Marine Biological Association of the United Kingdom*, 1976, 56, 527-594.

Partensky, F., Hess, W.R. and Vaulot, D. Prochlorococcus, a Marine Photosynthetic Prokaryote of Global Significance Prochlorococcus. *Microbiology and Molecular Biology Reviews*, 1999, 63, 106-127.

Peduzzi, P. and Weinbauer, M.G. Effect of concentrating the virus-rich 2–200-nm size fraction of seawater on the formation of algal flocs (marine snow). *Limnology and Oceanography*, 1993, 38, 1562–1565.

Pomeroy, L.R, Williams, PJ., Azam F., Hobbie J.E. The microbial loop. *Oceanography*, 2007, 20, 28–33.

Poulton, A.J., Charalampopoulou, A., Young, J.R., Tarran, G.A.., Lucas, M.I. and Quartly, G.D. Coccolithophore dynamics in non-bloom conditions during late summer in the central Iceland Basin (July-August 2007). *Limnology and Oceanography*, 2010, 55, 1601–1613.

Raven, J. and Crawfurd, K. Environmental controls on coccolithophore calcification. *Marine Ecology Progress Series*, 2012, 470, 137–166.

Randow, F. and Lehner, P.J. Viral avoidance and exploitation of the ubiquitin system. *Nature Cell Biology*, 2009, 11, 527–534.

Read, B.A., Kegel, J., Klute, M.J., Kuo, A., Lefebvre, S.C., et al. Pan genome of the phytoplankton Emiliania underpins its global distribution. *Nature*, 2013, 1–5.

Renner, S.S. and Bellot, S. Horizontal Gene Transfer in Eukaryotes: Fungi-to-Plant and Plant-to-Plant Transfers of Organellar DNA. In: Bock, R. and Knoop, V., eds., *Genomics of Chloroplasts and Mitochondria, Advances in Photosynthesis and Respiration* 35, 2012, pp. 223–235.

Richards, T.A., Soanes, D.M., Jones, M.D.M., Vasieva, O., Leonard, G., Paszkiewicz, K., Foster, P.G., Hall, N. and Talbot, N.J. Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. *Proceedings of the National Academy of Sciences of the United States of America*, 2011, 108, 15258–15263.

Robinson, C., Holligan, P., Jickells, T. and Lavender, S. The Atlantic Meridional Transect Programme (1995–2012). *Deep Sea Research Part II: Topical Studies in Oceanography*, 2009, 56, 895–898.

Roggentin, P., Schauer, R., Hoyer, L.L. and Vimr, E.R. The sialidase superfamily and its spread by horizontal gene transfer. *Molecular Microbiology*, 1993, 9, 915–921.

Roossinck, M.J. and Palukaitis, P. Genetic analysis of helper virus-specific selective amplification of cucumber mosaic virus satellite RNAs. *Journal of Molecular Evolution*, 1995, 40, 25-29.

Roossinck, MJ. Symbiosis versus competition in plant virus evolution. *Nature reviews. Microbiology*, 2005, 3, 917–924.

Rowe, J.M., Fabre, M-F., Gobena, D., Wilson, W.H. and Wilhelm, S.W. Application of the major capsid protein as a marker of the phylogenetic diversity of Emiliania huxleyi viruses. *FEMS Microbiology Ecology*. 2011, 76, 373–380.

Saitou, N. and Nei, M. The Neighbor-joining Method : A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution*, 1987, 4, 406–425.

Schauer, R. and Kamerling, J.P. The Chemistry and Biology of Trypanosomal trans-Sialidases: Virulence Factors in Chagas Disease and Sleeping Sickness. *Chembiochem : a European Journal of Chemical Biology*, 2011, 12, 2246-2264.

Scholthof, K-B.G. Tobacco mosaic virus: a model system for plant biology. *Annual Review of Phytopathology*, 2004, 42, 13–34.

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

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

Seaton, C.G.R., Lee, K. and Rohozinski, J. Photosynthetic Shutdown in Chlorella NC64A associated with the infection cycle of Paramecium bursaria Chlorella Virus-1. *Plant Physiology*, 1995, 108, 1431–1438.

Segovia, M. and Berg**e**s, J.A. Inhibition of caspase-like activities prevents the appearance of reactive oxygen species and dark-induced apoptosis in the unicellular chlorophyte Dunaliella tertiolecta. *Journal of Phycology*, 2009, 45, 1116–1126.

Shutler, J.D., Land, P.E., Brown, C.W., Findlay, H.S., Donlon, C.J., Medland, M., Snooke, R. and Blackford, J.C. Coccolithophore surface distributions in the North Atlantic and their modulation of the air-sea flux of CO_2 from 10 years of satellite Earth observation data. *Biogeosciences Discussions*, 2012, 9, 5823–5848.

Smyth, T.J., Fishwick, J.R., AL-Moosawi, L., Cummings, D.G., Harris, C., Kitidis, V., Rees, A., Martinez-Vicente, V. and Woodward, E.M.S. A broad spatio-temporal view of the Western English Channel observatory. *Journal of Plankton Research*, 2009, 32, 585–601.

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

Sowell, S.M., Abraham, P.E., Shah, M., Verberkmoes, N.C., Smith, D.P., Barofsky, D.F. and Giovannoni, S.J. Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. *The ISME Journal*, 2011, 5, 856–865.

Stanier, R.Y. and Cohen-Bazire, G. Phototrophic prokaryotes: the cyanobacteria. *Annual Review of Microbiology*, 1977, 31, 225-274.

Suderman, R.J., Pruijssers, A.J. and Strand, M.R. Protein tyrosine phosphatase-H2 from a polydnavirus induces apoptosis of insect cells. *The Journal of General Virology*, 2008, 89, 1411–1420.

Sullivan, M.B., Coleman, M.L., Weigele, P., Rohwer, F. and Chisholm, S.W. Three Prochlorococcus cyanophage genomes: signature features and ecological interpretations. *Public Library of Science Biology*, 2005, 3, e144.

Suttle, C.A., Chan, A.M. and Cottrell, M.T. Infection of phytoplankton by viruses and reduction of primary productivity. *Nature*, 1990, 347, 467–469.

Suttle, C.A., Chan, A.M., Cottrell, M.T. and Chan,Y.M. Use of ultrafiltration to isolate viruses from seawater which are pathogens of marine phytoplankton. *Applied Environmental Microbiology*, 1991, 57, 721-726.

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

Suttle, C.A. Viruses in the sea. *Nature*, 2005, 437, 356–361.

Suttle, C.A. Marine viruses-major players in the global ecosystem. *Nature Reviews Microbiology*, 2007, 5, 801–812.

Takahashi, T., Suzuki, T., Hidari, K.I-P.J., Miyamoto D. and Suzuki Y. A molecular mechanism for the low-pH stability of sialidase activity of influenza A virus N2 neuraminidases. *FEBS Letters*, 2003, 543, 71–75.

Tamura, K., Nei, M. and Kumar, S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences*, 2004, 101, 11030–11035.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, 2007, 24, 1596–1599.

Taylor, G. Sialidases: structures, biological significance and therapeutic potential. *Current Opinion in Structural Biology*, 1996, 6, 830–837.

Thamatrakoln, K., Korenovska, O., Niheu, A.K. and Bidle, K.D. Whole-genome expression analysis reveals a role for death-related genes in stress acclimation of the diatom Thalassiosira pseudonana. *Environmental Microbiology*, 2012, 14, 67–81.

Thingstad, T. and Lignell, R. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquatic Microbial Ecology*, 1997, 13, 19–27.

Thingstad, T. and Mantoura, R.F.C. Titrating excess nitrogen content of phosphorous-deficient eastern Mediterranean surface water using alkaline phosphatase activity as a bio-indicator. *Limnology and Oceanography: Methods*, 2005, 3, 94–100.

Thomas, W.H., Dodron, A.N., and Reid, F.M.H. Diatom productivity compared to other algae in natural marine phytoplankton assemblages. *Journal of Phycology*, 1978, 14, 250-253.

Thomas R., Grimsley N., Escande M-L., Subirana L., Derelle E. and Moreau H. Acquisition and maintenance of resistance to viruses in eukaryotic phytoplankton populations. *Environmental Microbiology*, 2011, 13, 1412–1420.

Thyrhaug, R., Larsen, A., Thingstad, T. and Bratbak G. Stable coexistence in marine algal host-virus systems. *Marine Ecology Progress Series*, 2003, 254, 27–35.

Tilman, D., Kilham, S.S. and Kilham, P. Phytoplankton community ecology: The role of limiting nutrients. *Annual Review of Ecology and Systematics*, 1982, 13, 349-372.

Tyrrell, T. and Merico, A. Emiliania huxleyi: bloom observations and the conditions that induce them. In: Thierstein, H.R. and Young, J.R., eds., *Coccolithophores: from molecular processes to global impact*, Springer, Verlag, 2004, pp. 75–97.

Urisman, A., Molinaro, R. J., Fischer, N. and Plummer, S. J. Identification of a novel gammaretrovirus in prostate tumors of patients homozygous for R462Q RNASEL variant. *Public Library of Scineces Pathogens*, 2006, 2, e25.

Van Etten, J.L., Lane, L.C. and Meints, R.H. Viruses and viruslike particles of eukaryotic algae. *Microbiological Reviews*, 1991, 55, 586–620.

Van Mooy, B.A.S, Fredricks, H.F., Pedler, B.E., Dyhrman, S.T., Karl, D.M., Koblízek, M., Lomas, M.W., Mincer, T.J., Moore, L.R., Moutin, T., Rappé, M.S. and Webb, E.A. Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. *Nature*, 2009, 458, 69–72.

Van Rijssel, M. and Gieskes, W.W. Temperature, light, and the dimethylsulfoniopropionate (DMSP) content of Emiliania huxleyi (Prymnesiophyceae). *Journal of Sea Research*, 2002, 48, 17–27.

Van Valen, L. A new evolutionary law. *Evolutionary Theory*, 1973, 1, 1-30.

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

Vardi, A, Haramaty, L., Van Mooy, B.A.S., Fredricks, H.F., Kimmanc,e S.A. and Larsen, A. Host – virus dynamics and subcellular controls of cell fate in a natural coccolithophore population. *Proceedings of the National Academy of Sciences of the United States of America*, 2012, 109, 19327–19332.

Vilicic, D., Djakovac, T., Buric, Z. and Bosak, S. Composition and annual cycle of phytoplankton assemblages in the northeastern Adriatic Sea. *Botaica Marina*, 2009, 52, 291-305.

Vogan, A.A. and Higgs, P.G. The advantages and disadvantages of horizontal gene transfer and the emergence of the first species. *Biology Direct,* 2011, 6: 1.

Von Itzstein, M., Wu, W.Y., Kok, G.B. Pegg, M.S., Dyason, J.C., Jin, B., Phan, V., Smythe, M.L., White, H.I. and Oliver, S.W. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*, 1993, 363, 418-423.

Warnes, S., Highmore, C. and Keevil, C. Horizontal Transfer of Antibiotic Resistance Genes on Abiotic Touch Surfaces: Implications for Public Health. *mBio,* 2012, 3, e00489-12.

Weitz, J.S. and Wilhelm, S.W. Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biology Reports,* 2012, 4:17.

Weynberg, K.D., Allen, M.J., Ashelford, K., Scanlan, D.J. and Wilson, W.H. From small hosts come big viruses: the complete genome of a second Ostreococcus tauri virus, OtV-1. *Environmental Microbiology*, 2009, 11, 2821–2839.

Widdicombe, C.E., Eloire, D., Harbour, D., Harris, R.P. and Somerfield P.J. Long-term phytoplankton community dynamics in the Western English Channel. *Journal of Plankton Research*, 2010, 32, 643–655.

Wilhelm, S.W. and Suttle, C.A. Viruses and Nutrient Cycles in the Sea aquatic food webs. *BioScience*, 1999, 49, 781-788.

Wilson, C.A., Eiden, M.V., Anderson, W.B., Lehel, C. and Olah, Z. The dual-function hamster receptor for amphotropic murine leukemia virus (MuLV), 10A1 MuLV, and gibbon ape leukemia virus is a phosphate symporter. *Journal of virology*. 1995, 69, 534–537.

Wilson, W.H., Tarran, G.A., Schroeder, D.C., Cox, M., Oke, J. and Malin, G. Isolation of viruses responsible for the demise of an Emiliania huxleyi bloom in the English Channel. *Journal of the Marine Biological Association of the UK*, 2002a, 82, 369–377.

Wilson, W.H., Tarran, G. and Zubkov, M.V. Virus dynamics in a coccolithophore-dominated bloom in the North Sea. *Deep Sea Research Part II: Topical Studies in Oceanography,* 2002b, 49, 2951–2963.

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. and Ghazal, P. Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. *Science,* 2005, 309, 1090–1092.

Wilson, W.H., Van Etten, J.L. and Allen, M.J. The Phycodnaviridae: the story of how tiny giants rule the world. *Current Topics in Microbiology and Immunology*, 2009, 328, 1–42.

Woese, C.R. and Fox, G.E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America*, 1977, 74, 5088–5090.

Woodhouse, M.T., Carslaw, K.S., Mann, G.W., Vallina, S.M., Vogt, M., Halloran, P.R. and Boucher, O. Low sensitivity of cloud condensation nuclei to changes in the sea-air flux of dimethyl-sulphide. *Atmospheric Chemistry and Physics*, 2010, 10, 7545–7559.

Wu, J., Sunda, W., Boyle, E.A, Karl, D.M. Phosphate Depletion in the Western North Atlantic Ocean. *Science*, 2000, 289, 759–762.

Yanai-Balser, G.M., Duncan, G.A., Eudy, J.D., Wang, D., Li, X., Agarkova, I.V., Dunigan, D.D. and Van Etten, J.L. Microarray analysis of Paramecium bursaria chlorella virus 1 transcription. *Journal of Virology*, 2010, 84, 532–542.

Yau, S., Lauro, F.M., DeMaere, M.Z., Brown, M.V., Thomas, T., Raftery, M.J., Andrews-Pfannkoch, C., Lewis, M., Hoffman, J.M., Gibson, J.A. and Cavicchioli, R. Virophage control of antarctic algal host-virus dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 2011, 108, 6163–6168.

Yooseph, S., Sutton, G., Rusch, D.B., Halpern, A.L., Williamson, S.J., Remington, K., Eisen, J.A., Heidelberg, K.B., Manning, G., Li, W., Jaroszewski, L., Cieplak, P., Miller, C.S., Li, H., Mashiyama, S.T., Joachimiak, M.P., Van Belle, C., Chandonia, J-M., Soergel, D.A., Zhai, Y., Natarajan, K., Lee, S., Raphael, B.J., Bafna, V., Friedman, R., Brenner, S.E., Godzik, A., Eisenberg, D., Dixon, J.E., Taylor, S.S., Strausberg, R.L., Frazier, M. and Venter, J.C. The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *Public Library of Science Biology*, 2007, 5, e16.

Young, L.S., Dawson, C.W. and Eliopoulos, A.G. Viruses and apoptosis. *British Medical Bulletin*, 1997, 53, 509–521.

Zhang, Y., Xiang, Y., Van Etten, J.L. and Rossmann, M.G. Structure and function of a Chlorella virus encoded glycosyltransferase. *Structure*, 2007, 15, 1031–1039.

Zimmer, S.M. and Burke, D.S. Historical perspective-Emergence of influenza A (H1N1) viruses. *The New England Journal of Medicine*, 2009, 361, 279–285.

Appendices

Table App1. CLUSTALW alignment of a 209 bp fragment sequence of the MCP gene of 42 different EhVs. In red: base pair substitutions amongst the different strains.

```
EhV-99B1        CAGGTTCAGTTTGGGGCTGAGTCGCATATTACCGTCAATCGTCAAGGCGATTTGCTATCGTGGATGTACCTCAAGATTGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTT
EhV-86          CAGGTTCAGTTTGGGGCTGAGTCGCATATTACCGTAAATCGCCAAGGCGATTTGCTATCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTT
EhV-MCP8        CAGGTTCAGTTTGGGGCTGAGTCGCATATTACCGTCAATCGCCAAGGCGATTTGCTATCGTGGATGTACCTCAAGATTGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-84          CAGGTTCAGTTCGGTGCTGAGTCGCATATTACCGTAAATCGCCAAGGCGATTTGCTATCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-88          CAGGTTCAGTTCGGTGCTGAGTCGCATATTACCGTAAATCGCCAAGGCGATTTGCTATCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP10       CAGGTTCAGTTTGGTGCTGAGTCGCATATTACCGTTAATCGTCAAGGCGATTTGCTATCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTT
EhV-MCP9        CAGGTTCAGTTTGGTGCTGAGTCGCATATTACCGTTAATCGTCAAGGCGATTTGCTATCGTGGATGTACCTCAAGATTGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTT
EhV-MCP7        CAGGTTCAGTTCGGTGCTGAGTCGCATATTACCGTCAACCGTCAAGGCGATTTGCTATCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP6        CAGGTTCAGTTCGGTGCTGAGTCGCATATTACCGTCAACCGTCAAGGCGATTTGCTATCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP5        CAGGTTCAGTTTGGTGCTGAGTCGCATATTACGGTCAACCGGCAAGGCGATTTGCTATCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP3        CAGGTTCAGTTTGGTGCTGAGTCGCATATTACGGTCAACCGTCAAGGCGATTTGCTATCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP2        CAGGTTCAGTTTGGTGCTGAGTCGCATATTACGGTCAACCGTCAAGGCGATTTGCTATCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP1        CAGGTTCAGTTTGGTGCTGAGTCGCATATTACGGTCAACCGTCAAGGCGATTTGCTATCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP4        CAGGTTCAGTTTGGTGCTGAGTCGCATATTACCGTCAACCGGCAAGGCGATTTGCTATCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP13       CAGGTTCAGTTTGGGGCTGAGTCGCATATTACCGTCAATCGCCAAGGCGATTTGCTATCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP15       CAGGTTCAGTTCGGTGCTGAGTCGCATATTACCGTTAATCGCCAGGGCGATTTGCTATCGTGGATGTACCTTAAGATCATCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP14       CAGGTTCAGTTCGGTGCTGAGTCGCATATTACCGTTAATCGCCAGGGCGATTTGCTATCGTGGATGTACCTTAAGATCATCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP16       CAGGTTCAGTTCGGTGCTGAGTCGCATATTACGGTCAACCGCCAGGGCGATTTTCTATCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP19       CAGGTTCAGTTTGGTGCTGAGTCGCATATTACCGTCAATAGGCAAGGCGACTTGCTTTCGTGGATGTACCTCAAGATCATCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP18       CAGGTTCAGTTTGGTGCTGAGTCGCATATTACCGTCAATAGGCAAGGCGACTTGCTTTCGTGGATGTACCTCAAGATCATCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP22       CAGGTTCAGTTTGGTGCTGAGTCGCATATTACCGTCAATAGGCAAGGCGACTTGCTTTCGTGGATGTACCTCAAGATCATCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP23       CAGGTTCAGTTTGGTGCTGAGTCGCATATTACCGTCAATAGGCAAGGCGACTTGCTTTCGTGGATGTACCTCAAGATCATCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP17       CAGGTTCAGTTTGGTGCTGAGTCGCATATTACCGTCAATAGGCAAGGCGACTTGCTTTCGTGGATGTACCTTAAGATCATCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhVmcp_WBS      CAGGTTCAGTTTGGTGCTGAGTCGCATATTACCGTCAATAGGCAAGGCGACTTGCTTTCGTGGATGTACCTTAAGATCATCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP12       CAGGTTCAGTTTGGGGCTGAGTCGCATATTACCGTCAATAGGCAAGGCGATTTGCTATCGTGGATGTACCTCAAGATCATCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP11       CAGGTTCAGTTTGGGGCTGAGTCGCATATTACCGTCAATAGGCAAGGCGATTTGCTATCGTGGATGTACCTTAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP24       CAGGTTCAGTTTGGTGCTGAGTCGCATATTACCGTCAATAGGCAAGGCGACTTGCTTTCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP21       CAGGTTCAGTTTGGTGCTGAGTCGCATATTACCGTCAATAGGCAAGGCGACTTTCTTTCGTGGATGTACCTCAAGATCATCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP20       CAGGTTCAGTTCGGTGCTGAGTCGCATATTACCGTCAATAGGCAAGGCGACCTTCTTTCGTGGATGTACCTCAAGATCATCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP31       CAGGTTCAGTTCGGTGCTGAGTCGCATATTACCGTCAACCGTCAGGGCGACCTTCTTTCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP30       CAGGTTCAGTTCGGTGCTGAGTCGCATATTACCGTCAACCGTCAGGGCGACCTTCTTTCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP33       CAGGTTCAGTTTGGTGCTGAGTCGCATATTACCGTCAACCGTCAGGGCGACCTTCTTTCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP32       CAGGTTCAGTTCGGTGCTGAGTCGCATATTACCGTCAACCGTCAAGGCGACCTTCTTTCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP29       CAGGTTCAGTTTGGTGCTGAGTCGCATATTACGGTCAACCGTCAAGGCGACCTTCTTTCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP28       CAGGTTCAGTTCGGTGCTGAGTCGCATATTACCGTCAATAGTCAGGGTGACCTTCTATCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP27       CAGGTTCAGTTTGGTGCTGAGTCGCATATTACCGTCAATAGGCAGGGTGACCTTCTTTCGTGGATGTACCTCAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-203         CAGGTTCAGTTTGGTGCTGAGTCGCATATTACCGTCAATAGGCAAGGCGACTTGCTATCGTGGATGTACCTTAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-207         CAGGTTCAGTTTGGTGCTGAGTCGCATATTACCGTCAATAGGCAAGGCGACTTGCTATCGTGGATGTACCTTAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-201         CAGGTTCAGTTCGGTGCCGAGTCGCATATTACTGTGAATAGGCAAGGCGACCTTCTTTCGTGGATGTACCTTAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-202         CAGGTTCAGTTCGGTGCCGAGTCGCATATTACTGTGAATAGACAGGGTGACCTTCTATCGTGGATGTACCTTAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP26       CAGGTTCAGTTCGGTGCCGAGTCGCATATTACCGTCAATCGTCAAGGCGATTTGCTATCGTGGATGTACCTTAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
EhV-MCP25       CAGGTTCAGTTCGGTGCCGAGTCGCATATTACCGTCAATCGTCAAGGCGATCTTCTATCGTGGATGTACCTTAAGATCGTCCTCCCTGGCCTAAAGGTCCAGAACCAGGCGGACACCGTC
                *********** ** ** ************** ** **  * ** ** **  * ** ************** *****  ****************************************
```

```
EhV-99B1        CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGACGTATCGCACGTCCTTCCATACATTGAGGGTGC
EhV-86          CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGACGTATCGCACGTCCTTCCATACATTGAGGGTGC
EhV-MCP8        CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGATGTATCGCACGTCCTTCCATACATTGAGGGTGC
EhV-84          CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGACGTATCGCACGTCCTTCCATACATTGAGGGTGC
EhV-88          CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGACGTATCGCACGTCCTTCCATACATTGAGGGTGC
EhV-MCP10       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP9        CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP7        CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP6        CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP5        CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP3        CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP2        CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP1        CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGCACGTCCTTCCATACATTGAGGGTGC
EhV-MCP4        CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP13       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP15       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP14       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP16       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP19       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP18       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP22       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP23       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP17       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhVmcp_WBS      CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGC-CGTCCTCCCATACATTGAGGGTGC
EhV-MCP12       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP11       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP24       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP21       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP20       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP31       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP30       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP33       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP32       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP29       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCACTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP28       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP27       CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGATGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-203         CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGACGTATCGCACGTCCTTCCATACATTGAGGGTGC
EhV-207         CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGACGTATCGCACGTCCTTCCATACATTGAGGGTGC
EhV-201         CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGACGTATCGCACGTCCTTCCATACATTGAGGGTGC
EhV-202         CAGCCAACTCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGACGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP26       CAGCCAACCCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGACGTATCGCACGTCCTCCCATACATTGAGGGTGC
EhV-MCP25       CAGCCAACCCAGCAGTCGTTCGCGTCGCTCGACAACGACGTCGCTGCGCAGGCTGACGTATCGCACGTCCTCCCATACATTGAGGGTGC
                ******** ******************************** ************* ******* ****** ****************
```

Table App2. The 468 protein coding sequences in EhV-86 and the number of homolog copies of each in the genomes of EhV-99B1, EhV-84, EhV-88, EhV-201, EhV-202, EhV-203, EhV-207 and EhV-208, based on the BLASTP sequence alignment of IMG/ER (max E value: 1e-05, min. percent identity: 20%) The four tRNAs of EhV-86 are not shown here.

Gene name	Product Name	EhV-99B1	EhV-84	EhV-88	EhV-201	EhV-202	EhV-203	EhV-207	EhV-208
ehv001	Recombination endonuclease VII.	1	0	0	2	1	2	1	1
ehv002	putative membrane protein	0	1	1	1	0	1	2	1
ehv003	putative membrane protein	0	1	1	1	1	1	1	1
ehv004	hypothetical protein	1	2	2	2	2	2	2	2
ehv005	putative membrane protein	0	1	1	1	1	1	1	1
ehv006	putative membrane protein	0	1	1	1	1	1	2	0
ehv007	hypothetical protein	0	1	1	1	2	1	1	1
ehv008	putative membrane protein	0	1	1	1	1	1	2	1
ehv009	hypothetical protein	0	1	1	1	1	1	1	1
ehv010	putative membrane protein	1	1	1	1	0	1	1	1
ehv011	hypothetical protein	1	1	1	1	1	1	1	1
ehv012	hypothetical protein	1	1	1	1	1	0	1	1
ehv013	putative membrane protein	1	1	1	1	0	1	1	0
ehv014	Longevity-assurance (LAG1) family protein	1	1	1	1	1	1	0	1
ehv015	putative membrane protein	0	1	1	1	1	1	1	1
ehv016	putative membrane protein	1	1	1	1	1	1	1	1
ehv017	hypothetical protein	2	2	2	2	2	2	2	1
ehv018	putative endonuclease	1	1	1	1	1	1	1	1
ehv019	Protein kinase domain.	1	0	1	1	1	1	1	1
ehv020	putative proliferating cell nuclear antigen	1	1	1	1	1	1	1	1
ehv021	putative serine protease	4	4	4	4	3	4	4	4
ehv022	phosphoglycerate mutase family protein	1	1	1	1	1	1	2	0
ehv023	putative deoxycytidylate deaminase	1	1	1	1	1	1	1	2
ehv024	hypothetical protein	1	1	1	0	0	0	0	0
ehv025	putative membrane protein	1	1	1	1	1	1	1	1
ehv026	ribonucleoside-diphosphate reductase small chain	1	1	1	1	1	1	1	1
ehv027	hypothetical protein	1	1	1	1	1	1	1	1
ehv028	putative lipase	1	1	1	1	1	1	1	1
ehv029	putative membrane protein	1	1	1	1	1	1	1	1
ehv030	putative DNA polymerase delta catalytic subunit	1	1	1	1	1	1	1	1
ehv031	putative sterol desaturase	1	1	1	1	1	1	1	1
ehv032	putative membrane protein	2	2	2	2	2	2	2	2
ehv033	putative membrane protein	2	2	2	2	2	2	2	2
ehv034	putative membrane protein	0	3	3	2	1	2	2	3
ehv035	putative membrane protein	2	2	2	2	2	2	2	2

ehv036	putative membrane protein	0	1	1	1	1	1	1	1
ehv037	putative membrane protein	0	2	2	2	0	2	2	2
ehv038	putative membrane protein	0	2	2	2	0	2	2	2
ehv039	hypothetical protein	1	1	1	1	1	1	1	1
ehv040	putative membrane protein	1	1	1	0	0	0	0	0
ehv041	putative endonuclease	1	1	1	1	1	1	1	1
ehv042	hypothetical protein	0	1	1	0	1	0	0	0
ehv043	hypothetical protein	2	1	1	1	1	1	1	1
ehv044	hypothetical protein	1	1	1	0	0	1	0	0
ehv045	putative membrane protein	1	1	1	1	0	1	1	1
ehv046	putative membrane protein	0	1	1	1	0	1	1	1
ehv047	hypothetical protein	1	1	1	1	1	1	1	1
ehv048	putative membrane protein	1	1	1	1	1	1	1	1
ehv049	hypothetical protein	1	1	1	1	1	1	1	1
ehv050	putative serine palmitoyltransferase	1	1	1	1	1	1	1	1
ehv051	hypothetical protein	1	1	1	2	0	2	2	2
ehv052	hypothetical protein	1	1	1	0	0	0	0	0
ehv053	hypothetical protein	1	1	1	1	1	1	1	1
ehv054	hypothetical protein	1	1	1	1	1	1	1	1
ehv055	putative membrane protein	2	2	2	2	3	2	2	2
ehv056	Methyltransferase domain.	2	2	2	2	2	2	2	2
ehv057	hypothetical protein	1	1	1	1	1	1	1	1
ehv058	Superfamily II DNA/RNA helicases, SNF2 family	1	1	1	1	1	1	1	1
ehv059	putative membrane protein	1	1	1	1	1	1	1	1
ehv060	putative lectin protein	3	1	2	3	2	3	3	2
ehv061	putative fatty acid desaturase	1	1	1	1	1	1	1	1
ehv062	putative membrane protein	1	1	1	1	1	1	1	1
ehv063	hypothetical protein	1	1	1	1	1	1	1	1
ehv064	putative DNA-dependent RNA polymerase II largest subunit	1	2	1	1	2	2	2	1
ehv065	hypothetical protein	1	1	1	1	1	1	1	1
ehv066	putative membrane protein	1	0	1	1	1	1	1	1
ehv067	hypothetical protein	1	1	1	1	1	1	1	1
ehv068	hypothetical protein	1	1	1	1	1	1	1	1
ehv069	putative membrane protein	1	1	1	1	1	1	1	1
ehv070	hypothetical protein	1	1	1	1	1	1	1	1
ehv071	hypothetical protein	1	1	1	1	1	1	1	1
ehv072	putative DNA-binding protein	1	1	1	1	1	1	1	1
ehv073	hypothetical protein	1	1	1	1	1	1	1	1
ehv074	putative membrane protein	1	1	1	1	1	1	1	1
ehv075	putative membrane protein	1	1	1	1	1	1	1	1
ehv076	putative membrane protein	1	1	1	1	1	1	1	1
ehv077	putative transmembrane fatty acid elongation protein	1	1	1	1	1	1	1	1
ehv078	putative membrane protein	2	2	2	2	3	2	2	2

ehv079	putative lipid phosphate phosphatase	1	1	1	1	1	1	1	1
ehv080	putative membrane protein	1	1	1	1	1	1	1	1
ehv081	putative membrane protein	1	1	1	1	0	1	1	1
ehv082	putative membrane protein	1	0	1	1	1	1	1	1
ehv083	putative membrane protein	0	1	1	2	3	2	1	1
ehv084	hypothetical protein	0	0	1	0	0	0	0	0
ehv085	putative major capsid protein	1	1	1	1	1	0	1	1
ehv086	hypothetical protein	1	1	1	0	0	0	0	0
ehv087	hypothetical protein	1	0	1	1	1	1	0	1
ehv088	putative membrane protein	0	1	1	1	1	1	1	1
ehv089	putative membrane protein	1	2	1	1	1	1	1	1
ehv090	hypothetical protein	1	1	1	1	1	1	1	1
ehv091	putative membrane protein	2	2	2	2	2	2	2	2
ehv092	hypothetical protein	1	1	1	1	0	1	1	1
ehv093	HNH endonuclease family protein	1	1	1	1	1	1	1	1
ehv094	putative membrane protein	1	1	1	1	1	1	1	1
ehv095	putative membrane protein	1	1	1	1	1	1	1	1
ehv096	hypothetical protein	1	1	1	1	1	1	1	1
ehv097	putative membrane protein	1	1	1	1	1	1	1	0
ehv098	hypothetical protein	1	1	1	1	1	1	1	1
ehv099	putative membrane protein	1	1	1	1	1	1	1	1
ehv100	putative membrane protein	1	1	1	1	0	1	1	1
ehv101	putative hydrolase	2	2	2	2	1	2	2	2
ehv102	putative membrane protein	1	1	1	1	1	1	1	1
ehv103	putative vesicle-associated membrane protein	1	1	1	1	1	1	1	1
ehv104	putative helicase	1	1	1	1	1	1	1	1
ehv105	transcription factor S-II (TFIIS) family protein	1	1	1	1	1	1	1	1
ehv106	hypothetical protein	0	1	1	1	1	1	1	1
ehv107	Superfamily I DNA and RNA helicases	1	1	1	1	1	1	1	1
ehv108	putative DNA-directed RNA polymerase subunit	1	1	1	1	1	1	1	1
ehv109	OTU-like cysteine protease	1	1	1	1	1	1	1	1
ehv110	putative RING finger protein	1	1	1	1	1	1	1	1
ehv111	hypothetical protein	1	1	1	1	1	1	1	1
ehv112	hypothetical protein	1	1	1	1	1	1	1	1
ehv113	bifunctional dihydrofolate reductase-thymidylate synthase	1	1	1	1	1	1	1	1
ehv114	putative membrane protein	1	1	1	1	1	1	1	1
ehv115	hypothetical protein	0	1	1	1	1	1	1	1
ehv116	putative membrane protein	1	1	1	1	1	1	1	1
ehv117	putative phosphate permease	0	1	1	1	1	1	1	1
ehv118	putative membrane protein	1	1	1	1	1	1	1	1
ehv119	putative membrane protein	1	1	1	1	1	1	1	1
ehv120	hypothetical protein	1	1	1	1	1	1	1	1
ehv121	hypothetical protein	1	1	1	1	1	1	1	1

ehv122	ATPases of the AAA+ class	1	1	1	1	1	1	1	1
ehv123	hypothetical protein	1	1	1	1	1	1	1	1
ehv124	hypothetical protein	1	1	1	1	1	1	1	1
ehv125	hypothetical protein	0	1	1	0	0	0	0	0
ehv127	hypothetical protein	1	1	1	0	0	0	0	0
ehv128	ERV1/ALR family protein	1	1	1	1	1	1	1	1
ehv129	hypothetical protein	0	1	1	1	0	1	1	1
ehv130	hypothetical protein	1	1	1	1	0	1	1	1
ehv131	putative membrane protein	1	1	1	1	1	1	1	1
ehv132	hypothetical protein	1	1	1	1	1	1	1	1
ehv133	putative ATP-dependent protease proteolytic subunit	1	1	1	1	1	1	1	1
ehv134	hypothetical protein	1	1	1	1	1	1	1	1
ehv135	putative membrane protein	1	1	1	1	1	1	1	1
ehv136	putative nucleic acid-binding protein	1	1	1	1	1	1	1	1
ehv137	putative membrane protein	1	1	1	0	0	0	0	0
ehv138	Recombination endonuclease VII	1	0	1	2	1	2	1	1
ehv139	AhpC/TSA family	1	1	1	1	1	1	1	1
ehv140	hypothetical protein	1	1	1	1	1	1	1	1
ehv141	PIF1 helicase	1	1	1	1	1	1	1	1
ehv142	hypothetical protein	1	1	1	1	1	1	1	1
ehv143	putative membrane protein	1	1	1	1	1	1	1	1
ehv144	hypothetical protein	1	1	1	1	1	1	1	1
ehv145	hypothetical protein	1	1	1	1	1	1	1	1
ehv146	putative membrane protein	1	1	1	1	0	1	1	1
ehv147	hypothetical protein	1	1	1	1	2	1	1	1
ehv148	hypothetical protein	1	1	1	1	1	1	1	1
ehv149	putative membrane protein	1	1	1	1	1	1	1	1
ehv150	hypothetical protein	1	1	1	1	1	1	1	1
ehv151	putative serine protease	4	4	4	3	4	4	4	4
ehv152	putative DNA-binding protein	1	1	1	1	1	1	1	1
ehv153	hypothetical protein	1	1	1	1	1	1	1	1
ehv154	Orthopoxvirus protein of unknown function (DUF830)	2	2	2	2	2	2	2	2
ehv155	putative membrane protein	1	1	1	1	1	1	1	1
ehv156	hypothetical protein	1	1	1	1	1	1	1	1
ehv157	putative membrane protein	1	0	1	1	0	1	1	1
ehv158	putative DNA ligase	1	0	1	0	1	1	1	1
ehv159	putative membrane protein	0	0	1	0	1	1	1	1
ehv160	putative serine protease	1	1	1	1	1	2	2	2
ehv161	putative membrane protein	0	0	1	0	1	1	1	1
ehv162	hypothetical protein	1	1	1	1	1	1	1	1
ehv163	hypothetical protein	1	1	1	1	1	1	1	1
ehv164	putative membrane protein	1	0	1	1	1	1	1	1
ehv165	putative membrane protein	1	1	1	1	1	1	1	1
ehv166	putative RING finger protein	1	1	1	1	1	1	1	1

ehv167	putative DNA-directed RNA polymerase subunit	1	1	1	1	1	1	1	1
ehv168	putative membrane protein	1	1	1	1	1	1	1	1
ehv169	hypothetical protein	1	1	1	1	1	1	1	1
ehv170	putative membrane protein	1	1	1	1	1	1	1	1
ehv171	putative membrane protein	1	0	1	1	1	1	1	1
ehv172	putative membrane protein	0	1	1	0	1	0	0	0
ehv173	putative membrane protein	0	1	1	1	1	1	1	1
ehv174	hypothetical protein	1	1	1	1	1	1	1	1
ehv175	hypothetical protein	2	2	2	2	2	2	2	2
ehv176	putative membrane protein	1	1	1	1	1	1	1	1
ehv177	putative membrane protein	2	2	2	2	2	2	2	2
ehv178	hypothetical protein	1	1	1	1	1	1	1	1
ehv179	Major Facilitator Superfamily protein	1	1	1	1	1	1	1	1
ehv180	putative membrane protein	1	1	1	1	1	1	1	1
ehv181	putative membrane protein	1	1	1	1	1	1	1	1
ehv182	putative membrane protein	1	1	1	1	1	1	1	1
ehv183	hypothetical protein	1	0	1	1	0	1	1	1
ehv184	putative DNA-binding protein	1	0	0	0	0	0	0	0
ehv185	putative membrane protein	1	1	1	1	2	1	1	1
ehv185A	hypothetical protein	0	1	1	0	0	0	0	0
ehv186	hypothetical protein	1	1	1	1	1	1	1	1
ehv187	putative membrane protein	1	1	1	1	0	1	1	1
ehv188	putative membrane protein	0	0	1	1	1	1	1	1
ehv189	hypothetical protein	1	0	1	1	1	1	1	1
ehv190	putative membrane protein	1	0	1	1	1	1	1	1
ehv191	putative membrane protein	2	0	2	3	1	3	3	2
ehv192	putative membrane protein	5	1	3	3	3	3	3	3
ehv193	putative membrane protein	1	1	1	1	1	1	1	1
ehv194	hypothetical protein	1	1	1	1	1	1	1	1
ehv195	putative membrane protein	1	1	1	1	1	1	1	1
ehv196	putative membrane protein	1	1	1	1	1	1	1	1
ehv197	putative membrane protein	1	1	1	1	1	1	1	1
ehv198	hypothetical protein	1	1	1	1	1	1	1	1
ehv199	putative membrane protein	1	1	1	1	1	1	1	1
ehv200	putative membrane protein	1	1	1	2	2	2	2	2
ehv201	hypothetical protein	1	1	1	1	1	1	1	1
ehv202	hypothetical protein	1	1	1	1	1	1	0	1
ehv203	putative membrane protein	1	1	1	1	1	1	1	1
ehv204	putative membrane protein	2	1	1	1	1	1	1	1
ehv205	hypothetical protein	1	1	1	1	1	1	1	1
ehv206	putative membrane protein	1	1	1	1	1	1	1	1
ehv207	putative membrane protein	1	1	0	0	0	0	0	0
ehv208	hypothetical protein	1	1	1	1	1	1	1	1
ehv209	hypothetical protein	1	1	1	1	1	1	1	1

ehv210	hypothetical protein	1	1	1	1	1	1	1	1
ehv210A	hypothetical protein	1	0	0	0	0	0	0	0
ehv211	hypothetical protein	1	1	1	1	1	1	1	1
ehv212	hypothetical protein	1	0	1	1	0	1	1	1
ehv213	putative membrane protein	1	0	0	0	0	0	0	0
ehv214	putative membrane protein	0	1	1	1	1	1	1	1
ehv215	hypothetical protein	1	1	1	1	1	1	1	1
ehv216	hypothetical protein	1	1	1	1	1	1	1	1
ehv217	hypothetical protein	1	1	1	1	1	1	1	1
ehv218	hypothetical protein	1	1	1	1	1	1	1	1
ehv219	hypothetical protein	1	0	1	1	1	1	1	1
ehv220	hypothetical protein	2	2	2	2	3	2	2	2
ehv221	hypothetical protein	1	1	1	1	1	1	1	1
ehv222	putative membrane protein	1	1	1	1	1	1	1	1
ehv223	hypothetical protein	1	1	1	1	1	1	1	1
ehv224	hypothetical protein	0	0	0	0	1	0	0	0
ehv225	hypothetical protein	0	1	0	1	2	1	1	1
ehv226	hypothetical protein	2	2	2	2	2	2	2	1
ehv227	putative membrane protein	1	1	1	0	0	0	0	0
ehv228	hypothetical protein	1	1	1	1	1	1	1	1
ehv229	putative membrane protein	1	1	1	1	1	1	1	1
ehv230	DNA-(apurinic or apyrimidinic site) lyase / Pyrimidine dimer DNA glycosylase	1	1	1	1	1	1	1	1
ehv231	hypothetical protein	1	1	1	1	0	1	1	1
ehv232	putative membrane protein	1	1	1	1	1	1	1	1
ehv233	hypothetical protein	1	1	1	1	1	1	1	1
ehv234	hypothetical protein	2	2	2	2	1	2	2	2
ehv235	hypothetical protein	4	5	4	4	3	3	4	3
ehv236	hypothetical protein	0	0	0	0	0	0	0	0
ehv237	hypothetical protein	1	0	0	0	0	0	0	0
ehv238	putative membrane protein	1	0	0	0	0	0	0	0
ehv239	hypothetical protein	1	1	1	1	1	1	1	1
ehv240	putative membrane protein	1	1	1	1	1	1	1	1
ehv241	hypothetical protein	1	1	0	1	1	1	1	1
ehv242	hypothetical protein	1	1	1	0	0	0	0	0
ehv243	putative membrane protein	1	0	1	1	0	1	1	1
ehv244	PDZ domain (Also known as DHR or GLGF)	1	0	0	0	1	0	0	0
ehv245	hypothetical protein	1	1	1	1	1	1	1	1
ehv246	putative lectin protein	1	1	1	1	1	1	1	1
ehv247	hypothetical protein	1	1	1	1	1	1	1	1
ehv248	hypothetical protein	1	1	1	1	3	1	1	1
ehv249	hypothetical protein	1	1	1	1	1	1	1	1
ehv250	putative membrane protein	1	1	1	1	1	1	1	1
ehv251	hypothetical protein	1	1	1	0	1	0	0	0

ehv252	hypothetical protein	1	1	1	1	2	1	1	1
ehv253	hypothetical protein	1	0	0	1	0	1	1	1
ehv254	hypothetical protein	1	1	1	1	1	1	1	1
ehv255	hypothetical protein	1	0	0	0	0	0	0	0
ehv256	hypothetical protein	2	1	1	1	1	1	1	1
ehv257	hypothetical protein	1	0	0	0	0	0	0	0
ehv258	hypothetical protein	1	1	1	1	1	1	1	1
ehv259	putative membrane protein	0	2	2	1	0	1	1	1
ehv260	hypothetical protein	1	1	1	1	1	1	1	1
ehv261	hypothetical protein	1	1	1	1	0	1	1	1
ehv262	hypothetical protein	1	1	1	1	1	1	1	1
ehv263	hypothetical protein	1	0	0	1	1	2	2	2
ehv264	hypothetical protein	1	1	1	1	0	1	1	1
ehv265	hypothetical protein	1	1	2	2	3	1	2	2
ehv266	hypothetical protein	4	4	4	4	3	3	4	3
ehv267	hypothetical protein	1	1	1	0	0	0	0	0
ehv268	putative membrane protein	1	1	1	1	1	1	1	1
ehv269	hypothetical protein	1	1	1	0	1	0	0	0
ehv269A	putative membrane protein	1	0	0	0	0	0	0	0
ehv270	putative membrane protein	1	1	1	0	0	0	0	0
ehv271	putative membrane protein	1	1	1	0	0	0	0	0
ehv272	hypothetical protein	4	4	4	1	6	3	4	3
ehv272A	hypothetical protein	1	1	1	1	0	1	1	1
ehv273	hypothetical protein	1	0	0	0	1	0	0	0
ehv274	hypothetical protein	1	1	1	1	1	1	1	1
ehv275	hypothetical protein	0	1	1	0	1	0	0	0
ehv276	hypothetical protein	1	1	1	1	1	1	1	1
ehv277	putative membrane protein	1	0	1	1	0	1	1	0
ehv278	putative membrane protein	1	1	1	0	0	0	0	0
ehv279	hypothetical protein	1	1	1	1	0	1	1	1
ehv280	hypothetical protein	2	1	1	1	1	1	1	1
ehv281	hypothetical protein	1	1	1	1	2	1	1	1
ehv282	hypothetical protein	2	2	2	2	1	2	2	2
ehv283	hypothetical protein	1	1	1	1	0	1	1	1
ehv284	hypothetical protein	1	1	1	1	1	1	1	1
ehv285	hypothetical protein	2	2	2	0	2	0	1	0
ehv286	hypothetical protein	0	0	1	0	0	0	0	0
ehv287	hypothetical protein	1	1	1	1	0	1	1	1
ehv288	hypothetical protein	3	2	2	2	2	2	3	1
ehv289	hypothetical protein	1	1	1	1	1	1	1	0
ehv290	hypothetical protein	0	0	0	0	0	0	0	0
ehv291	hypothetical protein	3	3	3	1	5	3	4	3
ehv292	hypothetical protein	3	3	3	1	5	3	4	3
ehv293	hypothetical protein	4	4	4	4	3	3	4	3
ehv294	hypothetical protein	1	1	1	1	1	0	1	1

ehv295	hypothetical protein	1	1	1	1	1	0	1	0
ehv296	hypothetical protein	1	0	0	0	1	0	0	0
ehv297	hypothetical protein	2	1	1	1	2	1	2	1
ehv298	hypothetical protein	0	1	1	0	1	0	1	1
ehv299	hypothetical protein	1	1	1	0	2	0	1	1
ehv300	hypothetical protein	1	1	1	0	0	0	1	1
ehv301	hypothetical protein	1	0	1	1	0	0	0	1
ehv302	hypothetical protein	1	2	2	1	4	1	3	2
ehv303	putative membrane protein	1	0	0	0	0	0	0	0
ehv304	hypothetical protein	1	1	1	0	1	0	0	0
ehv305	hypothetical protein	1	1	1	0	1	0	1	1
ehv306	hypothetical protein	1	1	1	0	2	0	1	2
ehv307	hypothetical protein	1	1	2	2	0	2	2	1
ehv308	hypothetical protein	1	1	1	1	0	1	1	2
ehv308A	hypothetical protein	1	0	0	0	0	0	0	0
ehv309	hypothetical protein	1	1	1	0	0	0	0	0
ehv310	putative membrane protein	0	2	2	1	1	1	1	2
ehv311	hypothetical protein	1	1	1	1	1	1	2	2
ehv312	putative membrane protein	1	1	1	1	1	1	1	2
ehv313	hypothetical protein	2	1	1	2	2	2	2	0
ehv314	hypothetical protein	4	4	4	4	3	3	4	3
ehv315	putative membrane protein	1	1	1	1	0	1	1	2
ehv316	putative membrane protein	0	1	1	1	1	1	1	1
ehv317	hypothetical protein	1	1	1	1	0	1	1	2
ehv318	putative membrane protein	1	0	0	0	0	0	0	0
ehv319	putative membrane protein	1	1	1	1	0	1	1	0
ehv320	hypothetical protein	1	1	1	1	1	1	1	1
ehv321	putative membrane protein	1	0	0	0	1	0	0	0
ehv322	hypothetical protein	2	2	2	2	2	2	2	2
ehv323	hypothetical protein	1	1	1	1	1	1	1	1
ehv324	hypothetical protein	1	1	1	1	0	1	1	1
ehv325	putative membrane protein	1	1	1	1	1	1	1	1
ehv326	putative membrane protein	1	1	1	1	1	1	1	1
ehv327	hypothetical protein	1	1	1	1	1	1	1	1
ehv328	hypothetical protein	1	1	1	1	1	1	1	1
ehv329	putative membrane protein	1	1	1	1	1	1	1	1
ehv330	hypothetical protein	1	1	1	1	1	1	1	1
ehv331	hypothetical protein	1	1	1	1	1	1	1	1
ehv332	putative membrane protein	1	1	1	1	1	1	1	1
ehv333	putative membrane protein	1	1	1	1	1	1	1	1
ehv334	hypothetical protein	1	1	1	1	1	1	1	1
ehv335	putative membrane protein	1	1	1	1	1	1	1	1
ehv336	hypothetical protein	1	1	1	1	1	1	1	1
ehv337	hypothetical protein	1	1	1	0	0	2	2	0
ehv338	hypothetical protein	1	1	1	1	1	1	1	1

ehv339	hypothetical protein	1	1	1	1	1	1	1	1
ehv340	putative membrane protein	1	1	1	1	1	1	1	1
ehv341	putative membrane protein	1	1	1	1	1	1	1	1
ehv342	hypothetical protein	1	1	1	1	1	1	1	1
ehv343	hypothetical protein	0	0	0	0	0	0	0	0
ehv344	DNA-binding protein, stimulates sugar fermentation	0	1	1	1	1	1	1	1
ehv345	hypothetical protein	0	1	1	1	1	1	1	1
ehv346	hypothetical protein	1	1	1	1	0	1	1	1
ehv347	putative membrane protein	1	1	1	1	1	1	1	1
ehv348	Alkylated DNA repair protein	1	1	1	1	1	1	1	1
ehv349	putative protease	1	1	1	1	1	1	1	1
ehv350	hypothetical protein	3	3	3	3	3	3	3	3
ehv351	hypothetical protein	1	2	2	2	2	2	2	2
ehv352	hypothetical protein	1	1	1	1	1	1	1	1
ehv353	hypothetical protein	1	1	1	1	1	1	1	1
ehv354	hypothetical protein	1	1	1	1	1	1	1	1
ehv355	Phytanoyl-CoA dioxygenase (PhyH)	1	1	1	1	1	1	1	1
ehv356	DNA or RNA helicases of superfamily II	1	0	1	1	1	1	1	1
ehv357	hypothetical protein	1	1	1	0	0	0	0	0
ehv358	putative thioredoxin	1	1	1	1	1	1	1	1
ehv359	putative membrane protein	1	1	1	1	1	1	1	1
ehv360	hypothetical protein	1	1	1	1	1	1	1	1
ehv361	putative serine protease	2	3	3	3	3	3	3	3
ehv362	putative membrane protein	1	1	1	1	1	1	1	1
ehv363	putative esterase	3	3	3	2	1	2	2	2
ehv364	putative membrane protein	2	2	1	2	3	4	2	3
ehv365	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	2	2	2	1	0	1	1	1
ehv366	putative membrane protein	1	2	2	1	1	1	1	1
ehv367	Predicted metal-binding protein related to the C-terminal domain of SecA	1	1	1	1	1	1	1	1
ehv368	hypothetical protein	1	1	1	1	1	1	1	1
ehv369	hypothetical protein	1	1	1	1	1	1	1	1
ehv370	putative membrane protein	1	2	2	2	2	2	2	2
ehv371	putative membrane protein	1	1	1	1	1	1	1	1
ehv372	putative membrane protein	1	1	1	1	1	1	1	1
ehv373	hypothetical protein	1	1	1	1	1	1	1	1
ehv374	putative membrane protein	1	1	0	1	1	1	1	1
ehv375	putative membrane protein	1	1	1	1	1	1	1	1
ehv376	putative membrane protein	1	1	1	1	1	1	1	1
ehv377	putative membrane protein	1	1	1	0	1	1	1	0
ehv378	OB-fold nucleic acid binding domain	1	1	1	1	1	1	1	1
ehv379	putative membrane protein	1	1	1	1	1	1	1	1
ehv380	putative membrane protein	1	1	1	1	1	1	1	1

ehv381	putative membrane protein	1	1	1	1	1	1	1	1
ehv382	putative membrane protein	1	1	1	1	1	1	1	1
ehv383	putative membrane protein	1	1	1	1	1	1	1	1
ehv384	hypothetical protein	1	1	1	1	2	1	1	1
ehv385	hypothetical protein	1	1	1	1	1	1	1	1
ehv386	putative membrane protein	1	1	1	1	1	1	1	1
ehv387	hypothetical protein	1	1	1	1	1	1	1	1
ehv388	putative membrane protein	2	2	2	2	2	2	2	2
ehv389	hypothetical protein	1	1	1	1	1	1	1	1
ehv390	hypothetical protein	1	1	1	1	1	1	1	1
ehv391	hypothetical protein	1	1	1	1	1	1	1	1
ehv392	hypothetical protein	1	1	1	1	1	1	1	1
ehv393	DnaJ domain-containing prot.	1	1	1	1	1	1	1	1
ehv394	hypothetical protein	1	1	1	1	1	1	1	1
ehv395	CRAL/TRIO domain.	1	1	1	1	1	1	1	1
ehv396	putative membrane protein	1	1	1	1	1	1	1	1
ehv397	putative deoxyuridine 5'-triphosphate nucleotidohydrolase	1	1	1	1	1	1	1	1
ehv398	hypothetical protein	1	1	1	1	1	1	1	1
ehv399	putative DNA-directed RNA polymerase subunit	1	1	1	1	1	1	1	1
ehv400	Bacterial lipocalin	1	1	1	1	1	1	1	1
ehv401	RNase HII	1	1	1	1	1	1	1	1
ehv402	putative protein kinase	2	2	2	2	2	2	2	2
ehv403	Poxvirus Late Transcription Factor VLTF3 like./A2L zinc ribbon domain	1	1	1	1	1	1	1	1
ehv404	putative membrane protein	1	1	1	1	1	1	1	1
ehv405	hypothetical protein	1	1	1	1	1	1	1	1
ehv406	hypothetical protein	1	1	1	1	1	1	1	1
ehv407	putative membrane protein	1	1	1	1	1	1	1	1
ehv408	Orthopoxvirus protein of unknown function (DUF830)	2	2	2	2	2	2	2	2
ehv409	hypothetical protein	1	1	1	1	1	1	1	1
ehv410	hypothetical protein	1	1	1	1	1	1	1	1
ehv411	putative membrane protein	1	1	1	1	1	1	1	1
ehv412	putative membrane protein	1	1	1	1	1	1	1	1
ehv413	hypothetical protein	1	1	1	1	2	1	1	1
ehv414	hypothetical protein	1	1	1	1	1	1	1	1
ehv415	putative fatty acid desaturase	1	1	1	1	1	1	1	1
ehv416	putative membrane protein	0	1	1	0	0	0	0	0
ehv417	putative membrane protein	1	1	1	1	1	1	1	1
ehv418	hypothetical protein	1	1	1	1	1	1	1	1
ehv419	hypothetical protein	1	1	1	1	1	1	1	1
ehv420	hypothetical protein	1	0	0	0	0	0	0	0
ehv421	hypothetical protein	1	0	0	0	0	0	0	0
ehv422	putative membrane protein	1	1	1	0	0	0	0	0
ehv423	Protein of unknown function (DUF672)	1	0	0	0	0	0	0	0

ehv424	hypothetical protein	1	0	0	0	0	0	0	0
ehv425	putative membrane protein	1	1	1	1	1	1	1	1
ehv426	hypothetical protein	1	1	1	1	1	1	1	1
ehv427	putative membrane protein	1	1	1	1	1	0	1	1
ehv428	putative ribonucleoside-diphosphate reductase protein	1	1	1	1	2	1	1	1
ehv429	hypothetical protein	1	1	1	1	1	1	1	1
ehv430	putative helicase	1	0	0	0	0	0	0	0
ehv431	putative thymidylate kinase	1	1	1	1	1	1	1	1
ehv432	Methylase involved in ubiquinone/menaquinone biosynthesis	2	2	2	2	2	2	2	2
ehv433	putative membrane protein	1	1	1	1	1	1	1	1
ehv434	putative DNA-directed RNA polymerase II subunit	1	1	1	1	2	1	1	1
ehv435	hypothetical protein	1	1	1	1	1	1	1	1
ehv436	hypothetical protein	1	1	1	1	1	1	1	1
ehv437	hypothetical protein	1	1	1	1	1	1	1	1
ehv438	MYM-type Zinc finger with FCS sequence motif	1	1	1	1	1	1	1	1
ehv439	hypothetical protein	1	0	1	1	1	1	1	1
ehv440	proliferating cell nuclear antig.	1	1	1	1	1	1	1	1
ehv441	hypothetical protein	1	1	1	1	1	1	1	1
ehv442	putative membrane protein	1	1	1	1	1	1	1	1
ehv443	putative membrane protein	1	1	1	1	1	1	1	1
ehv444	putative DNA topoisomerase	1	1	1	1	1	1	1	1
ehv445	hypothetical protein	1	0	1	1	1	1	1	1
ehv446	hypothetical protein	1	1	1	1	1	1	1	1
ehv447	putative serine protease	3	3	3	3	3	3	3	3
ehv448	mRNA capping enzyme, beta chain	1	1	1	1	1	1	1	1
ehv449	hypothetical protein	1	1	1	1	1	1	1	1
ehv450	putative membrane protein	1	1	1	1	1	1	1	1
ehv451	putative protein kinase	1	1	1	1	1	1	1	1
ehv452	HMG (high mobility group) box	1	1	1	1	1	1	1	1
ehv453	putative mRNA capping enzyme	1	1	1	1	1	1	1	1
ehv454	putative membrane protein	1	1	1	1	1	1	1	1
ehv455	putative sialidase	1	1	1	1	0	1	1	1
ehv456	hypothetical protein	1	1	1	1	1	1	1	1
ehv457	putative membrane protein	1	1	1	1	1	1	1	1
ehv458	hypothetical protein	1	1	1	1	1	1	1	1
ehv459	putative nucleic acid independent nucleoside triphosphatase	1	2	1	1	2	2	1	1
ehv460	hypothetical protein	1	1	1	1	1	1	1	1
ehv461	putative membrane protein	1	1	1	1	1	1	1	1
ehv462	hypothetical protein	1	1	1	1	1	1	1	1
ehv463	Predicted RNA-binding protein homologous to eukaryotic snRNP	1	1	1	1	1	1	1	1
ehv464	putative membrane protein	0	1	1	1	1	1	1	1
ehv465	putative thioredoxin protein	0	1	1	1	1	1	1	1

ehv466	putative membrane protein	0	1	1	1	1	1	1	1
ehv467	putative membrane protein	0	1	1	1	1	1	1	1
ehv468	putative membrane protein	0	1	1	1	1	1	1	1

Table App.3. EhV-84 homology analysis to the draft genome (including introns) of *E. huxleyi* 1516 with the built in BLASTP algorithm on the IMG/ER online sequence analysis platform.

CDS	Protein function	Protein size (aa)	% Identity
ENVG00366	sodium/phosphate symporter	534	86.4
ENVG00070	ribonucleoside diphosphate reductase	746	72.3
ENVG00176	Lectin C-type domain	1994	69.4
ENVG00415	Alphaherpesvirus glycoprotein E	223	68.8
ENVG00139	ribonucleoside-diphosphate reductase small subunit	325	59
ENVG00131	flap endonuclease-1	358	56.9
ENVG00194	fatty-acyl elongase	322	56
ENVG00136	dCMP deaminase	173	55.7
ENVG00035	deoxyuridine 5'-triphosphate nucleotidohydrolase	148	55.2
ENVG00439	PIF1 helicase	420	51.1
ENVG00370	thymidylate synthase	480	50.4
ENVG00165	serine palmitoyltransferase	870	49.8
ENVG00400	DNA-directed RNA polymerase, subunit N (RpoN/RPB10)	86	44
ENVG00177	fatty acid desaturase	320	43.6
ENVG00413	major facilitator superfamily transporter protein	541	42.9
ENVG00145	Sterol desaturase	328	42.1
ENVG00259	DNA-binding protein, stimulates sugar fermentation	321	40.2
ENVG00181	DNA-directed RNA polymerase subunit beta	1055	40
ENVG00077	DNA-directed RNA polymerase subunit B	1156	39.9
ENVG00419	zinc finger protein	357	39.7
ENVG00127	longevity-assurance family protein	288	39.4
ENVG00038	Bacterial lipocalin	202	38.9
ENVG00074	thymidylate kinas	327	37.4
ENVG00147	TLC domain	153	36.8
ENVG00196	PAP2 superfamily	243	35.6
ENVG00054	delta 9 acyl-lipid fatty acid desaturase	258	35.5
ENVG00172	Methyltransferase domain.	238	35.4
ENVG00367	Sterol desaturase	327	34.3
ENVG00144	DNA polymerase	1012	33.7
ENVG00276	Protease	384	33.3
ENVG00039	ribonuclease HII	209	33.2
ENVG00180	RNA polymerase II largest subunit	418	33.1
ENVG00091	Protease	295	31.6
ENVG00087	topoisomersae II	1103	31.4
ENVG00134	Secreted trypsin-like serine protease	404	30.3
ENVG00429	Secreted trypsin-like serine protease	294	27.7
ENVG00095	Protein kinase domain	271	25.9

Table App.4. EhV-86 homology analysis to the draft genome (including introns) of *E. huxleyi* 1516 with the built in BLASTP algorithm on the IMG/ER online sequence analysis platform.

CDS	Protein function	Protein size (aa)	% Identity
ehv117	putative phosphate permease	534	84.62
ehv428	putative ribonucleoside-diphosphate reductase protein	746	67.28
ehv026	ribonucleoside-diphosphate reductase small chain	325	60
ehv158	putative DNA ligase	622	57.62
ehv023	putative deoxycytidylate deaminase	173	57.05
ehv077	putative transmembrane fatty acid elongation protein	285	55.73
ehv397	putative deoxyuridine 5'-triphosphate nucleotidohydrolase	148	55.17
ehv031	putative sterol desaturase	328	51.92
ehv061	putative fatty acid desaturase	320	49.37
ehv110	putative RING finger protein	278	49.06
ehv018	putative endonuclease	358	48.74
ehv141	PIF1 helicase	420	47.5
ehv050	putative serine palmitoyltransferase	870	44.93
ehv113	bifunctional dihydrofolate reductase-thymidylate synthase	480	44.32
ehv179	Major Facilitator Superfamily protein	541	44.21
ehv434	putative DNA-directed RNA polymerase II subunit	1156	43.8
ehv415	putative fatty acid desaturase	258	43.4
ehv103	putative vesicle-associated membrane protein	117	40.62
ehv431	putative thymidylate kinase	327	40.24
ehv344	DNA-binding protein, stimulates sugar fermentation	251	39.38
ehv014	Longevity-assurance (LAG1) family protein	288	39.35
ehv020	putative proliferating cell nuclear antigen	259	39.01
ehv400	Bacterial lipocalin	202	38.2
ehv064	putative DNA-dependent RNA polymerase II largest subunit	1468	37.9
ehv030	putative DNA polymerase delta catalytic subunit	1012	36.65
ehv056	Methyltransferase domain	238	36.17
ehv451	putative protein kinase	271	36.08
ehv058	Superfamily II DNA/RNA helicases, SNF2 family	867	35.54
ehv079	putative lipid phosphate phosphatise	236	33.59
ehv444	putative DNA topoisomerase	1103	33.53
ehv021	putative serine protease	404	33.19
ehv401	RNase HII (EC 3.1.26.4) (IMGterm)	209	33.18
ehv139	AhpC/TSA family	144	31.4
ehv447	putative serine protease	301	30.99
ehv348	Alkylated DNA repair protein	210	30.69
ehv361	putative serine protease	449	30.68
ehv160	putative serine protease	334	30.3
ehv151	putative serine protease	302	29.45
ehv395	CRAL/TRIO domain	215	28.08
ehv455	putative sialidase	373	26.97
ehv133	putative ATP-dependent protease proteolytic subunit	238	26.72
ehv154	Orthopoxvirus protein of unknown function (DUF830).	247	25.81
ehv408	Orthopoxvirus protein of unknown function (DUF830).	233	25.26
ehv104	putative helicase	527	20.41

Table App.5. EhV-88 homology analysis to the draft genome (including introns) of *E. huxleyi* 1516 with the built in BLASTP algorithm on the IMG/ER online sequence analysis platform.

CDS	Protein function	Protein size (aa)	% Identity
EOVG00119	phosphate permease	534	86.4
EOVG00179	ribonucleoside diphosphate reductase	746	72.8
EOVG00062	Lectin C-type domain	1994	69.4
EOVG00266	Alphaherpesvirus glycoprotein E	223	68.8
EOVG00025	ribonucleoside-diphosphate reductase small subunit	325	59
EOVG00472	DNA ligase 1	622	57.8
EOVG00017	Endonuclease	358	56.9
EOVG00079	transmembrane fatty acid elongation protein	322	56
EOVG00022	dCMP deaminase	173	55.7
EOVG00212	deoxyuridine 5'-triphosphate nucleotidohydrolase	148	55.2
EOVG00115	thymidylate synthase	480	50.4
EOVG00050	serine palmitoyltransferase	870	49.8
EOVG00463	DNA helicase	420	48.9
EOVG00251	DNA-directed RNA polymerase, subunit N (RpoN/RPB10)	86	44
EOVG00063	fatty acid desaturase	320	43.6
EOVG00264	Sugar phosphate permease	541	42.9
EOVG00031	Sterol desaturase	328	42.1
EOVG00333	DNA-binding protein, stimulates sugar fermentation	321	40.2
EOVG00066	RNA polymerase beta' subunit	1468	40
EOVG00174	DNA-directed RNA polymerase subunit B	1156	40
EOVG00269	zinc finger protein	357	39.7
EOVG00013	Protein transporter of the TRAM (translocating chain-associating membrane) superfamily, longevity assurance factor	288	39.4
EOVG00209	Bacterial lipocalin	202	38.9
EOVG00177	Thymidylate kinase	327	37.9
EOVG00081	PAP2 superfamily	243	35.6
EOVG00193	fatty acid desaturase	258	35.5
EOVG00057	Methyltransferase domain	238	35.4
EOVG00118	Sterol desaturase	327	34.3
EOVG00033	TLC domain	190	33.9
EOVG00030	DNA polymerase	1012	33.7
EOVG00350	Secreted trypsin-like serine protease	449	33.3
EOVG00208	ribonuclease HII	209	33.2
EOVG00160	Secreted trypsin-like serine protease	301	32.6
EOVG00164	DNA topoisomerase	1103	31.3
EOVG00020	Tryptase	404	30.3
EOVG00453	Secreted trypsin-like serine protease	294	27.7
EOVG00156	Protein kinase domain	271	25.1

Table App.6. EhV-99B1 homology analysis to the draft genome (including introns) of *E. huxleyi* 1516 with the built in BLASTP algorithm on the IMG/ER online sequence analysis platform.

CDS	Protein function	Protein size (aa)	% Identity
ehv428	ribonucleoside-diphosphate reductase protein	746	66.87
ehv434	DNA-directed RNA polymerase II subunit precursor	1507	63.64
ehv026	ribonucleoside-diphosphate reductase small chain	325	58.96
ehv158	DNA ligase	622	57.01
ehv077	transmembrane fatty acid elongation protein	285	55.73
ehv023	deoxycytidylate deaminase	173	55.03
ehv397	deoxyuridine 5'-triphosphate nucleotidohydrolase	149	54.02
ehv031	sterol desaturase	328	51.92
ehv061	fatty acid desaturase	320	49.37
ehv110	RING finger protein	278	49.06
ehv018	Endonuclease	358	48.74
ehv141	DNA helicase	420	46.25
ehv415	fatty acid desaturase	258	45.36
ehv050	putative serine palmitoyltransferase	870	44.78
ehv179	Major Facilitator Superfamily protein	528	44.4
ehv113	bifunctional dihydrofolate reductase-thymidylate synthase	480	44.32
ehv431	thymidylate kinase	327	40.24

Table App.7. EhV-201 homology analysis to the draft genome (including introns) of *E. huxleyi* 1516 with the built in BLASTP algorithm on the IMG/ER online sequence analysis platform.

CDS	Protein function	Protein size (aa)	% Identity
EPVG00115	phosphate permease	534	86.4
EPVG00251	ribonucleoside-diphosphate reductase	746	70.9
EPVG00058	Lectin C-type domain.	1990	64.7
EPVG00025	ribonucleoside-diphosphate reductase small subunit	325	59.5
EPVG00018	flap endonuclease-1	358	56.5
EPVG00075	fatty acid elongase	322	56.3
EPVG00023	deoxycytidylate deaminase	173	55.7
EPVG00218	deoxyuridine 5'-triphosphate nucleotidohydrolase	148	54
EPVG00048	serine palmitoyltransferase	870	50.5
EPVG00111	thymidylate synthase	480	50.4
EPVG00432	PIF1 helicase	420	48.9
EPVG00062	RNA polymerase subunit beta	1468	46.2
EPVG00422	DNA-directed RNA polymerase, subunit N (RpoN/RPB10)	86	44
EPVG00059	fatty acid desaturase	320	43.6
EPVG00410	Sugar phosphate permease	541	42.3
EPVG00168	DNA-binding protein, stimulates sugar fermentation	321	40.2
EPVG00014	Protein transporter of the TRAM (translocating chain-associating membrane) superfamily, longevity assurance factor	288	40
EPVG00257	RNA polymerase II	1156	39.9
EPVG00221	Bacterial lipocalin	202	38.9
EPVG00031	Sterol desaturase	328	37.1
EPVG00254	Thymidylate kinase	327	36.7
EPVG00053	Methylase involved in ubiquinone/menaquinone biosynthesis	238	35.9
EPVG00237	fatty-acid desaturase	258	35.5
EPVG00033	TLC domain	190	34.4
EPVG00114	Sterol desaturase	327	34.3
EPVG00030	DNA polymerase	1012	34
EPVG00021	serine protease	409	33.6
EPVG00184	Secreted trypsin-like serine protease	477	32.5
EPVG00399	Secreted trypsin-like serine protease	295	32
EPVG00222	RNase HII	209	31.8
EPVG00244	hypothetical protein	233	31.3
EPVG00403	DNA topoisomerase II	1103	31.3
EPVG00077	PAP2 superfamily	243	29.8
EPVG00442	Secreted trypsin-like serine protease	302	27.4
EPVG00395	Protein kinase domain	271	27.2

Table App.8. EhV-202 homology analysis to the draft genome (including introns) of *E. huxleyi* 1516 with the built in BLASTP algorithm on the IMG/ER online sequence analysis platform.

CDS	Protein function	Protein size (aa)	% Identity
EXVG00227	Polyubiquitin	80	96
EXVG00329	phosphate repressible phosphate permease	534	76.92
EXVG00045	ribonucleoside diphosphate reductase	491	69.67
EXVG00354	zinc finger protein	419	66.67
EXVG00276	ribonucleoside-diphosphate reductase small subunit	325	59.5
EXVG00222	fatty acid elongase	322	58.89
EXVG00303	DNA ligase	622	57.58
EXVG00278	deoxycytidylate deaminase	173	55.92
EXVG00046	ribonucleotide-diphosphate reductase	289	55.73
EXVG00078	deoxyuridine 5'-triphosphate nucleotidohydrolase	148	55.17
EXVG00271	Sterol desaturase	328	51.92
EXVG00319	PIF1 helicase	420	50
EXVG00241	fatty acid desaturase	320	49.37
EXVG00040	RNA polymerase II subunit 2	229	48.23
EXVG00190	Zinc finger, C3HC4 type (RING finger)	271	48.08
EXVG00284	flap endonuclease-1	358	47.06
EXVG00358	major facilitator superfamily transporter protein	544	44.38
EXVG00185	thymidylate synthase	480	43.93
EXVG00328	DNA-binding protein, stimulates sugar fermentation	301	41.86
EXVG00216	fatty acid desaturase	296	40.57
EXVG00252	serine palmitoyltransferase	867	40.06
EXVG00289	longevity-assurance family protein	288	39.74
EXVG00282	proliferating cell nuclear antigen	259	38.57
EXVG00043	thymidylate kinase	326	38.24
EXVG00238	DNA-directed RNA polymerase, subunit A	418	37.1
EXVG00247	Methylase involved in ubiquinone/menaquinone biosynthesis	238	36.7
EXVG00039	RNA polymerase II	916	36.55
EXVG00075	Bacterial lipocalin	202	36.52
EXVG00219	PAP2 superfamily	243	36
EXVG00272	DNA-directed DNA polymerase	1016	35.18
EXVG00237	RNA polymerase II subunit beta	1057	34.94
EXVG00197	Synaptobrevin	118	34.85
EXVG00299	Secreted trypsin-like serine protease	411	34.38
EXVG00181	Fatty acid hydroxylase superfamily	305	33.6
EXVG00028	DNA topoisomerase II	1103	33.53
EXVG00021	Protein kinase domain	271	33.33
EXVG00111	Secreted trypsin-like serine protease	435	32.71
EXVG00025	serine protease	295	32.66
EXVG00210	Sterol desaturase	205	32.6
EXVG00330	Predicted protein-tyrosine phosphatise	166	30.66
EXVG00245	Superfamily II DNA/RNA helicases, SNF2 family	867	30.25
EXVG00124	Alkylated DNA repair protein	210	30.05
EXVG00321	Redoxin	146	29.41
EXVG00080	CRAL/TRIO domain	215	29.27
EXVG00310	Secreted trypsin-like serine protease	289	29.18
EXVG00074	RNase HII (EC 3.1.26.4) (IMGterm)	150	28.57
EXVG00306	Orthopoxvirus protein of unknown function (DUF830)	250	26.85
EXVG00066	Orthopoxvirus protein of unknown function (DUF830)	233	26.8

EXVG00340	Protease subunit of ATP-dependent Clp proteases	232	25.58
EXVG00063	Glycosyltransferase (GlcNAc)	367	22.37
EXVG00196	DNA or RNA helicases of superfamily II	527	21.53

Table App.9. EhV-203 homology analysis to the draft genome (including introns) of *E. huxleyi* 1516 with the built in BLASTP algorithm on the IMG/ER online sequence analysis platform.

CDS	Protein function	Protein size (aa)	% Identity
ELVG00354	sodium/phosphate symporter	534	86.4
ELVG00052	ribonucleoside-diphosphate reductase	746	73
ELVG00228	Lectin C-type domain	1990	64.7
ELVG00194	ribonucleoside-diphosphate reductase small subunit	325	59.5
ELVG00433	DNA ligase	622	57.8
ELVG00187	flap endonuclease-1	358	56.5
ELVG00247	fatty acid elongase	322	56.3
ELVG00192	deoxycytidylate deaminase	173	55.7
ELVG00083	deoxyuridine 5'-triphosphate nucleotidohydrolase	148	54
ELVG00218	serine palmitoyltransferase	870	50.5
ELVG00358	thymidylate synthetase	480	50.4
ELVG00416	DNA helicase	420	48.9
ELVG00233	RNA polymerase subunit II	1053	46.2
ELVG00457	DNA-directed RNA polymerase, subunit N (RpoN/RPB10)	86	44
ELVG00229	fatty acid desaturase	320	43.6
ELVG00443	major facilitator superfamily transporter protein	541	42.3
ELVG00134	DNA-binding protein, stimulates sugar fermentation	321	40.2
ELVG00183	longevity-assurance family protein	288	40
ELVG00046	DNA-directed RNA polymerase subunit B	1156	39.9
ELVG00080	Bacterial lipocalin	202	38.9
ELVG00200	Sterol desaturase	328	37.1
ELVG00049	Thymidylate kinase	327	36.7
ELVG00223	Methylase involved in ubiquinone/menaquinone biosynthesis	238	35.9
ELVG00064	fatty acid desaturase	258	35.5
ELVG00202	TLC domain	190	34.4
ELVG00355	Sterol desaturase	327	34.3
ELVG00199	DNA-directed DNA polymerase	1012	34
ELVG00190	Protease	409	33.6
ELVG00232	DNA-directed RNA polymerase subunit A	418	33.1
ELVG00118	Serine protease	433	32.5
ELVG00026	Protease	295	32
ELVG00079	Ribonuclease HII	209	31.8
ELVG00030	Topoisomerase II	1103	31.3
ELVG00249	PAP2 superfamily	243	29.8
ELVG00426	Secreted trypsin-like serine protease	294	27.4
ELVG00022	Protein kinase domain	271	27.2

Table App.10. EhV-207 homology analysis to the draft genome (including introns) of *E. huxleyi* 1516 with the built in BLASTP algorithm on the IMG/ER online sequence analysis platform.

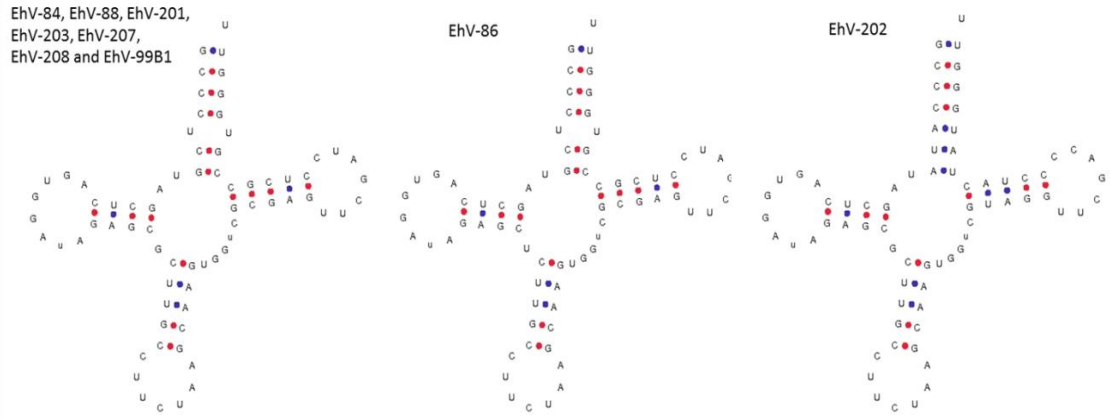
CDS	Protein function	Protein size (aa)	% Identity
EQVG00250	phosphate transporter	534	86.4
EQVG00048	ribonucleoside-diphosphate reductase	746	70.9
EQVG00306	Laminin G domain.	1520	68.3
EQVG00193	Lectin C-type domain	1990	64.7
EQVG00161	ribonucleoside-diphosphate reductase small subunit	325	59.5
EQVG00382	DNA ligase	622	57.8
EQVG00438	flap structure-specific endonuclease 1	358	56.5
EQVG00458	deoxycytidylate deaminase	173	56.4
EQVG00211	fatty acid elongase	322	56.3
EQVG00080	deoxyuridine 5'-triphosphate nucleotidohydrolase	148	54
EQVG00184	serine palmitoyltransferase	870	50.5
EQVG00246	thymidylate synthase	480	50.4
EQVG00399	PIF1 helicase	420	48.9
EQVG00198	DNA-directed RNA polymerase subunit beta	1053	46.2
EQVG00422	DNA-directed RNA polymerase, subunit N (RpoN/RPB10)	86	44
EQVG00194	fatty acid desaturase	320	43.6
EQVG00410	Sugar phosphate permease	541	42.3
EQVG00167	Sterol desaturase	305	41.8
EQVG00131	DNA-binding protein, stimulates sugar fermentation	321	40.2
EQVG00042	DNA-directed RNA polymerase subunit B	1156	39.9
EQVG00077	Bacterial lipocalin	202	38.9
EQVG00045	thymidylate kinase	327	36.7
EQVG00189	Methylase involved in ubiquinone/menaquinone biosynthesis	238	35.9
EQVG00061	fatty acid desaturase	258	35.5
EQVG00169	TLC domain	190	34.4
EQVG00249	Sterol desaturase	327	34.3
EQVG00166	DNA-directed DNA polymerase	1012	34
EQVG00441	Secreted trypsin-like serine protease	409	33.6
EQVG00197	DNA-directed RNA polymerase, beta' subunit/160 kD subunit	418	33.1
EQVG00115	Secreted trypsin-like serine protease	477	32.5
EQVG00023	Secreted trypsin-like serine protease	295	32
EQVG00076	ribonuclease H	209	31.8
EQVG00027	DNA topoisomerase II	1103	31.3
EQVG00213	PAP2 superfamily	243	29.8
EQVG00389	Secreted trypsin-like serine protease	302	27.4
EQVG00019	Protein kinase domain	271	27.2

Table App.11. EhV-208 homology analysis to the draft genome (including introns) of *E. huxleyi* 1516 with the built in BLASTP algorithm on the IMG/ER online sequence analysis platform.

CDS	Protein function	Protein size (aa)	% Identity
ERVG00232	phosphate permease	534	86.4
ERVG00004	ribonucleoside-diphosphate reductase	746	70.9
ERVG00134	Lectin C-type domain	1990	64.7
ERVG00166	ribonucleoside-diphosphate reductase small subunit	325	59.5
ERVG00168	deoxycytidylate deaminase	92	58.7
ERVG00331	DNA ligase I	622	57.8
ERVG00372	flap endonuclease 1	358	56.5
ERVG00117	fatty-acyl elongase	322	56.3
ERVG00367	deoxycytidylate deaminase	173	55.7
ERVG00035	deoxyuridine 5'-triphosphate nucleotidohydrolase	148	54
ERVG00143	serine palmitoyltransferase	870	50.5
ERVG00236	thymidylate synthase	480	50.4
ERVG00314	DNA helicase	420	48.9
ERVG00130	DNA-dependent RNA polymerase II largest subunit	1468	46.2
ERVG00348	DNA-directed RNA polymerase, subunit N (RpoN/RPB10)	86	44
ERVG00133	fatty acid desaturase	320	43.6
ERVG00360	Sugar phosphate permease	541	42.3
ERVG00085	DNA-binding protein, stimulates sugar fermentation	321	40.2
ERVG00211	DNA-directed RNA polymerase subunit B	1156	39.9
ERVG00376	longevity-assurance family protein	288	39.7
ERVG00032	Bacterial lipocalin	202	38.9
ERVG00001	thymidylate kinase	289	38
ERVG00160	Sterol desaturase	328	37.1
ERVG00138	Methylase involved in ubiquinone/menaquinone biosynthesis	238	35.9
ERVG00017	fatty acid desaturase	258	35.5
ERVG00158	TLC domain.	190	34.4
ERVG00233	Sterol desaturase	327	34.3
ERVG00161	DNA-directed DNA polymerase	1012	34
ERVG00369	Secreted trypsin-like serine protease	414	33.6
ERVG00069	Secreted trypsin-like serine protease	477	32.5
ERVG00192	Secreted trypsin-like serine protease	295	32
ERVG00031	ribonuclease H	209	31.8
ERVG00196	DNA topoisomerase II	1103	31.3
ERVG00115	PAP2 superfamily	243	29.8
ERVG00324	Secreted trypsin-like serine protease	302	27.4
ERVG00188	Protein kinase domain	271	27.2

tRNA Arg morphotypes

A

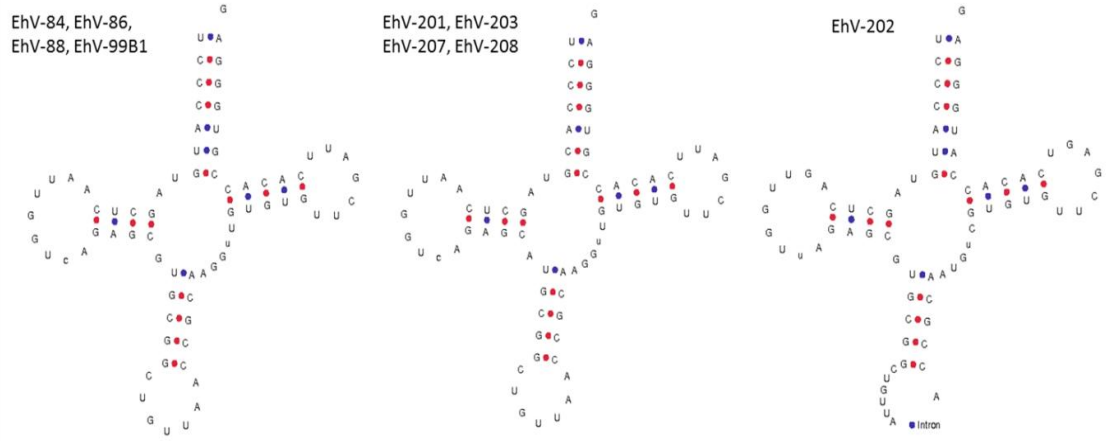



```
Arg-208     GCCCTCGTAGCTCAGTGGATAGAGCGCTTGCCTTCTAAGCAAGTGGTCGCGAGTTCGATC
Arg-99B1    GCCCTCGTAGCTCAGTGGATAGAGCGCTTGCCTTCTAAGCAAGTGGTCGCGAGTTCGATC
Arg-207     GCCCTCGTAGCTCAGTGGATAGAGCGCTTGCCTTCTAAGCAAGTGGTCGCGAGTTCGATC
Arg-203     GCCCTCGTAGCTCAGTGGATAGAGCGCTTGCCTTCTAAGCAAGTGGTCGCGAGTTCGATC
Arg-201     GCCCTCGTAGCTCAGTGGATAGAGCGCTTGCCTTCTAAGCAAGTGGTCGCGAGTTCGATC
Arg-88      GCCCTCGTAGCTCAGTGGATAGAGCGCTTGCCTTCTAAGCAAGTGGTCGCGAGTTCGATC
Arg-84      GCCCTCGTAGCTCAGTGGATAGAGCGCTTGCCTTCTAAGCAAGTGGTCGCGAGTTCGATC
Arg-86      GCCCTCGTAGCTCAGTGGATAGAGCTCTTGCCTTCTAAGCAAGTGGTCGCGAGTTCGATC
Arg-202     GCCCATATAGCTCAGTGGATAGAGCGCTTGCCTTCTAAGCAAGTGGTCGTAGGTTCGACC
            ****   ******************** ************************   ****** *

Arg-208     CTCGCCGTGGGTT
Arg-99B1    CTCGCCGTGGGTT
Arg-207     CTCGCCGTGGGTT
Arg-203     CTCGCCGTGGGTT
Arg-201     CTCGCCGTGGGTT
Arg-88      CTCGCCGTGGGTT
Arg-84      CTCGCCGTGGGTT
Arg-86      CTCGCCGTGGGTT
Arg-202     CCTACTATGGGTT
            *   *  ******
```

tRNA Asn morphotypes

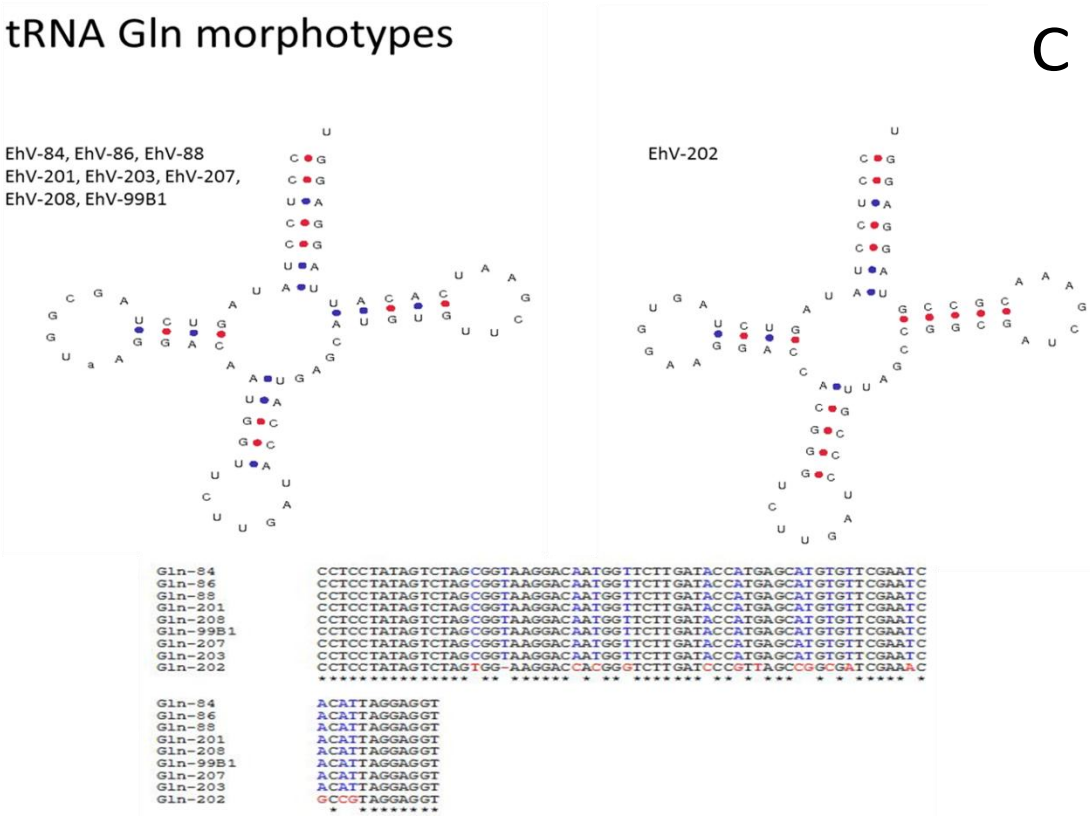
B



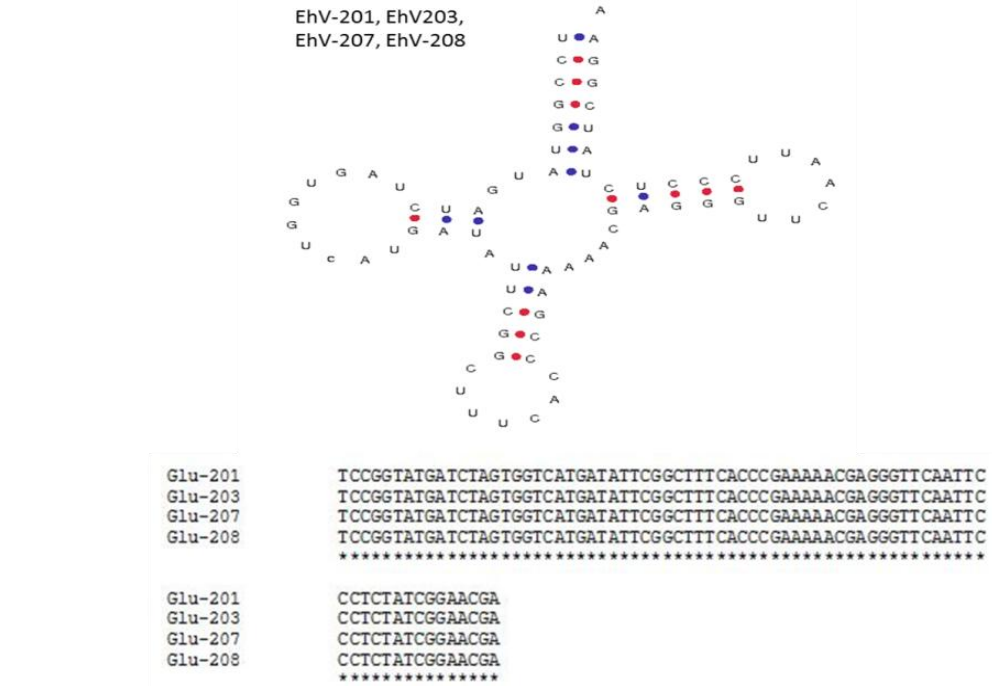

```
Asn-201     TCCCACGTAGCTCAATTGGTCAGAGCATGCGGCTGTTAACCGCAAGG------------
Asn-203     TCCCACGTAGCTCAATTGGTCAGAGCATGCGGCTGTTAACCGCAAGG------------
Asn-207     TCCCACGTAGCTCAATTGGTCAGAGCATGCGGCTGTTAACCGCAAGG------------
Asn-208     TCCCACGTAGCTCAATTGGTCAGAGCATGCGGCTGTTAACCGCAAGG------------
Asn-84      TCCCATGTAGCTCAATTGGTCAGAGCGTGCGGCTGTTAACCGCAAGG------------
Asn-86      TCCCATGTAGCTCAATTGGTCAGAGCGTGCGGCTGTTAACCGCAAGG------------
Asn-88      TCCCATGTAGCTCAATTGGTCAGAGCGTGCGGCTGTTAACCGCAAGG------------
Asn-99B1    TCCCATGTAGCTCAATTGGTCAGAGCGTGCGGCTGTTAACCGCAAGG------------
Asn-202     TCCCATGTAGCTCAGTTGGTTAGAGCGTGCGGCTGTTAACCGTAAGGACAGTGACCGCAA
            ***** ******** ***** ***** *************** ****

Asn-201     --TTGTGTGTTCGATTCACACCGTGGGAG
Asn-203     --TTGTGTGTTCGATTCACACCGTGGGAG
Asn-207     --TTGTGTGTTCGATTCACACCGTGGGAG
Asn-208     --TTGTGTGTTCGATTCACACCGTGGGAG
Asn-84      --TTGTGTGTTCGATTCACACCGTGGGAG
Asn-86      --TTGTGTGTTCGATTCACACCGTGGGAG
Asn-88      --TTGTGTGTTCGATTCACACCGTGGGAG
Asn-99B1    --TTGTGTGTTCGATTCACACCGTGGGAG
Asn-202     TGTCGTGTGTTCGAGTCACACCATGGGAG
              * ********** ******* ******
```

C

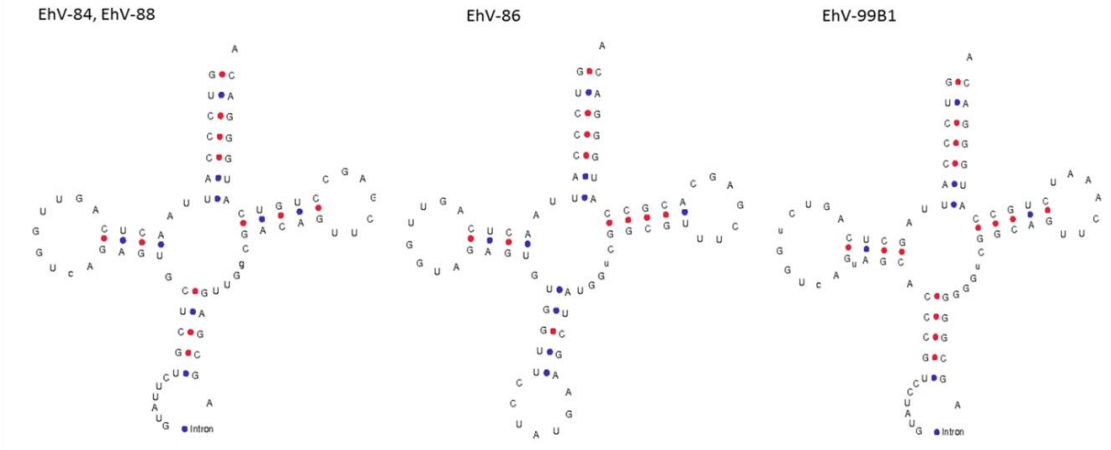

tRNA Gln morphotypes
EhV-84, EhV-86, EhV-88
EhV-201, EhV-203, EhV-207,
EhV-208, EhV-99B1
EhV-202
Gln-84 CCTCCTATAGTCTAGCGGTAAGGACAATGGTTCTTGATACCATGAGCATGTGTTCGAATC
Gln-86 CCTCCTATAGTCTAGCGGTAAGGACAATGGTTCTTGATACCATGAGCATGTGTTCGAATC
Gln-88 CCTCCTATAGTCTAGCGGTAAGGACAATGGTTCTTGATACCATGAGCATGTGTTCGAATC
Gln-201 CCTCCTATAGTCTAGCGGTAAGGACAATGGTTCTTGATACCATGAGCATGTGTTCGAATC
Gln-208 CCTCCTATAGTCTAGCGGTAAGGACAATGGTTCTTGATACCATGAGCATGTGTTCGAATC
Gln-99B1 CCTCCTATAGTCTAGCGGTAAGGACAATGGTTCTTGATACCATGAGCATGTGTTCGAATC
Gln-207 CCTCCTATAGTCTAGCGGTAAGGACAATGGTTCTTGATACCATGAGCATGTGTTCGAATC
Gln-203 CCTCCTATAGTCTAGCGGTAAGGACAATGGTTCTTGATACCATGAGCATGTGTTCGAATC
Gln-202 CCTCCTATAGTCTAGTGG-AAGGACCACGGGTCTTGATCCCGTTAGCCGGCGATCGAAAC
Gln-84 ACATTAGGAGGT
Gln-86 ACATTAGGAGGT
Gln-88 ACATTAGGAGGT
Gln-201 ACATTAGGAGGT
Gln-208 ACATTAGGAGGT
Gln-99B1 ACATTAGGAGGT
Gln-207 ACATTAGGAGGT
Gln-203 ACATTAGGAGGT
Gln-202 GCCGTAGGAGGT

D


tRNA Glu morphotypes
EhV-201, EhV203,
EhV-207, EhV-208
Glu-201 TCCGGTATGATCTAGTGGTCATGATATTCGGCTTTCACCCGAAAAACGAGGGTTCAATTC
Glu-203 TCCGGTATGATCTAGTGGTCATGATATTCGGCTTTCACCCGAAAAACGAGGGTTCAATTC
Glu-207 TCCGGTATGATCTAGTGGTCATGATATTCGGCTTTCACCCGAAAAACGAGGGTTCAATTC
Glu-208 TCCGGTATGATCTAGTGGTCATGATATTCGGCTTTCACCCGAAAAACGAGGGTTCAATTC
Glu-201 CCTCTATCGGAACGA
Glu-203 CCTCTATCGGAACGA
Glu-207 CCTCTATCGGAACGA
Glu-208 CCTCTATCGGAACGA

tRNA Ile morphotypes

E

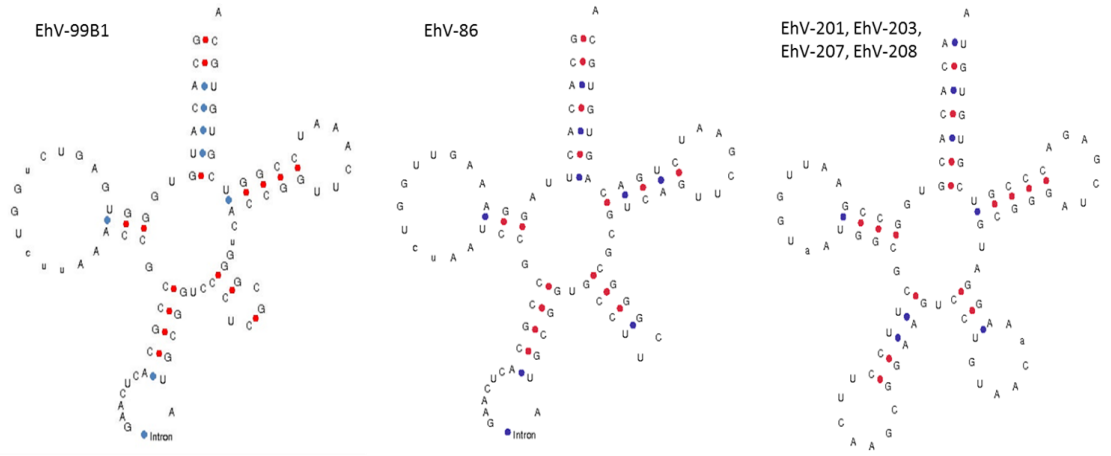



```
Ile-88      GTCCCATTAACTCAGT-TGGTCAG-AGTGCTCGTCTTATGAGATTGTCATAAAATCTGAG
Ile-99B1    GTCCCATTAGCTCAGTCTGGTCAGTAGCACCCGTCCTATGAGATTTCAATACAATCTAGG
Ile-84      GTCCCATTAACTCAGT-TGGTCAG-AGTGCTCGTCTTATGAGATTGTCATAAAATCTGAG
Ile-86      GTCCCATTAACTCAGT-TGGT-AG-AGTGTGGTTCCTATGAAGCTATGGT----------
            ********* ****** **** ** **       ** *****   *    *

Ile-88      ATAGCGAGTTGGCGACAGTTCGAGCCTGTCATGGGACACTA
Ile-99B1    CTAGCGGGGGGTCGGCAGTTCAAATCTGCCATGGGACA---
Ile-84      ATAGCGAGTTGGCGACAGTTCGAGCCTGTCATGGGACACTA
Ile-86      ----------CGGCGTTTCGAGCACGCCATGGGACACTA
                     ** *  *** *    * *********
```

tRNA Leu morphotypes

F




```
Leu-207     ACACACGTGGCCGAAT-TGGTAATGG--CGCTTCCTTCAAG---CGGAAGTCCTGTAACA
Leu-208     ACACACGTGGCCGAAT-TGGTAATGG--CGCTTCCTTCAAG---CGGAAGTCCTGTAACA
Leu-203     ACACACGTGGCCGAAT-TGGTAATGG--CGCTTCCTTCAAG---CGGAAGTCCTGTAACA
Leu-201     ACACACGTGGCCGAAT-TGGTAATGG--CGCTTCCTTCAAG---CGGAAGTCCTGTAACA
Leu-86      GCACACTTAGGAAAGT-TGGTCT-AATCCGCCGCACTCAAGGCATGTTAATCATGTGAGA
Leu-99B1    GCACATGTGGGTGAGTCTGGTCTTAAACCGCCGCACTCAAGACATGTTAATCATGTGAGA
             ****  * *   * * ****        ***  *  *****     *  * ** *** * *

Leu-207     AA--------------GGATGCGGGGATCGAGACCCGTCGTGTGTA
Leu-208     AA--------------GGATGCGGGGATCGAGACCCGTCGTGTGTA
Leu-203     AA--------------GGATGCGGGGATCGAGACCCGTCGTGTGTA
Leu-201     AA--------------GGATGCGGGGATCGAGACCCGTCGTGTGTA
Leu-86      TATGCGGTGCCTTCGGGCGCGTCAGTTCGAATCTGACAGTGTGCA
Leu-99B1    TATGCGGTCCCTCGCGGGTCACCGGTTCAAATCCGGTCGTGTGCA
             *           *       * ** *  *     ***** *
```

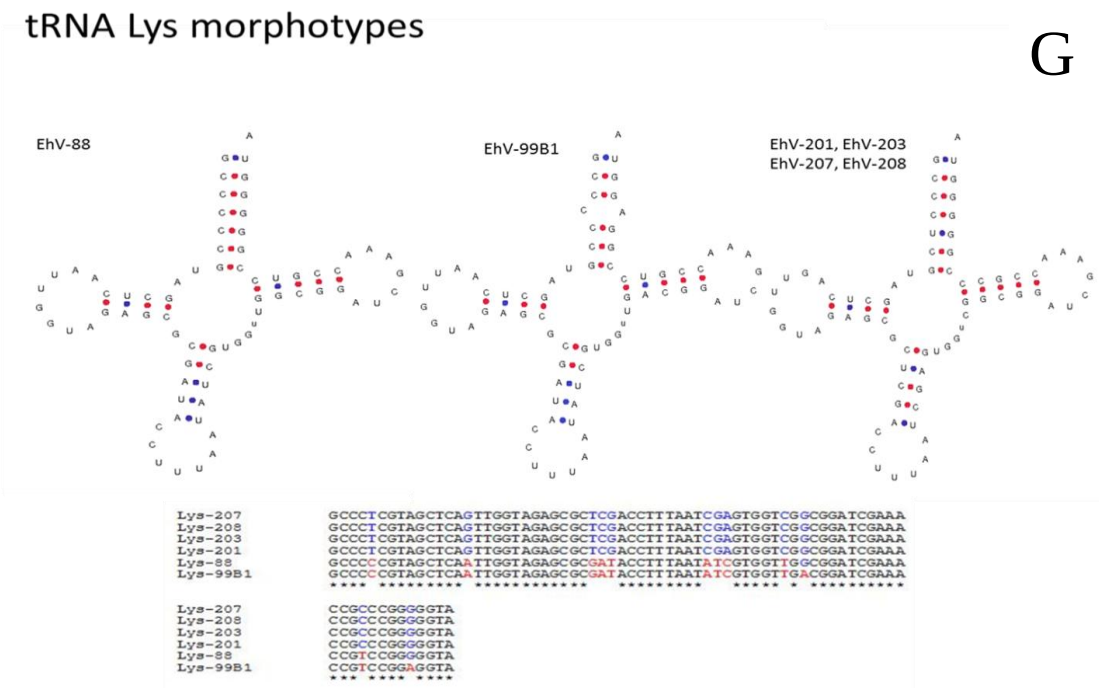


Fig. App.1. The seven tRNAs present in the genomes of coccolithoviruses, the various morphotypes of each tRNA (A-G) analyzed by the online tRNAscan-SE 1.21, and a nucleotide CLUSTALW sequence alignment of each. Red dots indicate G-C links and blue dots indicate A-U. In red and blue are also highlighted the differences in the sequence alignment of each tRNA.

Journal of Virology

Draft Genome Sequence of the Coccolithovirus Emiliania huxleyi Virus 203

Jozef I. Nissimov, Charlotte A. Worthy, Paul Rooks, Johnathan A. Napier, Susan A. Kimmance, Matthew R. Henn, Hiroyuki Ogata and Michael J. Allen

J. Virol. **2011, 85(24):13468. DOI: 10.1128/JVI.06440-11.**

JOURNAL OF VIROLOGY, Dec. 2011, p. 13468–13469 Vol. 85, No. 24
0022-538X/11/$12.00 doi:10.1128/JVI.06440-11

GENOME ANNOUNCEMENTS

Draft Genome Sequence of the Coccolithovirus *Emiliania huxleyi* Virus 203

Jozef I. Nissimov,[1] Charlotte A. Worthy,[1,2] Paul Rooks,[1] Johnathan A. Napier,[2] Susan A. Kimmance,[1] Matthew R. Henn,[3] Hiroyuki Ogata,[4] and Michael J. Allen[1]*

Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth PL1 3DH, United Kingdom[1]; Department of Biological Chemistry, Rothamsted Research, Harpenden, Herts AL5 2JQ, United Kingdom[2]; The Broad Institute of MIT and Harvard, Cambridge, Massachusetts[3]; and Structural and Genomic Information Laboratory, CNRS-UPR2589, Mediterranean Institute of Microbiology (IFR-88), Aix-Marseille University, 163 avenue de Luminy Case 934, FR-13288 Marseille, France[4]

Received 30 September 2011/Accepted 3 October 2011

The *Coccolithoviridae* are a recently discovered group of viruses that infect the marine coccolithophorid *Emiliania huxleyi*. *Emiliania huxleyi* virus 203 (EhV-203) has a 160- to 180-nm-diameter icosahedral structure and a genome of approximately 400 kbp, consisting of 464 coding sequences (CDSs). Here we describe the genomic features of EhV-203 together with a draft genome sequence and its annotation, highlighting the homology and heterogeneity of this genome in comparison with the EhV-86 reference genome.

Coccolithoviruses infect the coccolithophore *Emiliania huxleyi*, a cosmopolitan marine microalga which forms blooms that can cover up to 100,000 km^2 (10). Coccolithoviruses are a major cause of bloom termination, and their role in global biogeochemical cycling is gaining increasing attention (2). Coccolithovirus abundances can reach 10^7 ml^{-1} in natural seawater under bloom conditions and 10^8 to 10^9 ml^{-1} under laboratory culture (8). *Emiliania huxleyi* virus 86 (EhV-86), the model virus, was isolated in 1999 from the English Channel (50°13.79′N, 04°9.59′W), and was sequenced in its entirety in 2005 (9, 10). *Emiliania huxleyi* virus 203 (EhV-203) was isolated from the English Channel (50°00.36′N, 04°18.87′W) from a depth of 15 m on 27 July 2001 (1, 10). The icosahedral EhV-203 virion structure and morphology are similar to those of other coccolithoviruses and phycodnaviruses in general (11). Phylogenetic analysis of available major capsid protein (MCP) gene sequences indicates that the closest relatives to EhV-203 are EhV-201, EhV-202, and EhV-207 (1, 10).

EhV-203 genome sequencing, finishing, and annotation were performed by the Broad Institute. The genome was sequenced using the 454 FLX pyrosequencing technology platform (Roche/454, Branford, CT). Library construction and sequencing were performed as previously described (4). General protocols for library construction can be found at www.broadinstitute.org/annotation/viral/Phage/Protocols.html. *De novo* genome assembly of the resulting reads was performed using the Newbler v2.3 assembly software package (4). A total of 71,325 reads were produced and assembled into 7 contigs comprising 400,520 bp with a maximum contig length of 142,770 bp, average contig length of 57,131 bp, and an average coverage of approximately 40. Genes were identified using the Broad Institute's Automated Phage Annotation Protocol (4). Additional gene prediction analysis and functional annotation were performed within the Integrated Microbial Genomes—Expert Review (IMG-ER) platform (5).

General features of the EhV-203 genome sequence include a nucleotide composition of 40.12% G+C, a total of 464 predicted protein coding sequences (CDSs), and six tRNA genes (Arg, Asn, Gln, Glu, Leu, and Lys). Of the 464 CDSs annotated by IMG-ER, 91 have been annotated with functional product predictions. A total of 412 CDSs have homologues (>20% identity) within the EhV-86 genome (31 are 100% identical). Of the 52 CDSs unique to EhV-203, the majority are of unknown function, although two CDSs have homology to glycosyltranferases and one to a zinc finger protein.

Among the EhV-203 CDSs displaying the lowest similarity to their EhV-86 homologues are a lectin protein (63.62% identity) and two endonucleases (43.24% and 87.38% identity). The CDSs with the highest similarity (100% identity) include predicted DNA-directed RNA polymerase, DNA-binding protein, and transcription factor S-II and ERV1/ALR family proteins (9). EhV-203 encodes the same sphingolipid LCB biosynthetic machinery as EhV-86, with homologues for a sterol desaturase, serine palmitoyltransferase, transmembrane fatty acid elongation protein, lipid phosphate phosphatase, and two desaturases (3, 6, 7, 9). Like EhV-86, EhV-203 also lacks the critical sphingolipid LCB biosynthetic activity 3-ketosphinganine reductase (6). Further sequencing of related strains in the future will, no doubt, reveal more about the genetic and functional diversity of these environmentally important viruses.

* Corresponding author. Mailing address: Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth PL1 3DH, United Kingdom. Phone: 44 123 456 7890. Fax: 44 1752 633 101. E-mail: mija@pml.ac.uk.

Nucleotide sequence accession number. The nucleotide sequence for the draft genome sequence has been deposited in GenBank under accession no. JF974291.

This research was funded in part by the Gordon and Betty Moore Foundation through a grant to the Broad Institute (M.R.H.) and through the NERC Oceans 2025 program (M.J.A.). Sample G4014 was sequenced, assembled, and annotated at the Broad Institute. J.I.N. is supported by an NERC studentship, and C.W. is supported by a BBSRC Industrial CASE studentship sponsored by PML Applications. H.O. is supported by IGS/CNRS and ANR (grant no. ANR-09-PCS-GENM-218 and ANR-08-BDVA-003).

We thank Konstantinos Mavromatis from JGI who assisted with information regarding the IMG/ER platform and the Broad Institute's Genome Sequencing Platform, Finishing Team, and Annotation Team for their efforts to generate the genomic data. Jean Devonshire and the Centre for Bioimaging at Rothamsted provided technical support for transmission electron microscopy.

REFERENCES

1. **Allen, M. J., J. Martinez-Martinez, D. C. Schroeder, P. J. Somerfield, and W. H. Wilson.** 2007. Use of microarrays to assess viral diversity: from genotype to phenotype. Environ. Microbiol. **9:**971–982.
2. **Coyle, K. O., and A. I. Pinchuk.** 2002. Climate-related differences in zooplankton density and growth on the inner shelf of the southeastern Bering Sea. Prog. Oceanogr. **55:**177–194.
3. **Han, G. S., et al.** 2006. Expression of a novel marine viral single-chain serine palmitoyltransferase and construction of yeast and mammalian single-chain chimera. J. Biol. Chem. **281:**39935–39942.
4. **Henn, M. R., et al.** 2010. Analysis of high-throughput sequencing and annotation strategies for phage genomes. Plos One **5:**e9083.
5. **Markowitz, V. M., et al.** 2009. IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics **25:**2271–2278.
6. **Michaelson, L. V., T. M. Dunn, and J. A. Napier.** 2010. Viral trans-dominant manipulation of algal sphingolipids. Trends Plant Sci. **15:**651–655.
7. **Monier, A., et al.** 2009. Horizontal gene transfer of an entire metabolic pathway between a eukaryotic alga and its DNA virus. Genome Res. **19:** 1441–1449.
8. **Schroeder, D. C., J. Oke, M. Hall, G. Malin, and W. H. Wilson.** 2003. Virus succession observed during an Emiliania huxleyi bloom. Appl. Environ. Microbiol. **69:**2484–2490.
9. **Wilson, W. H., et al.** 2005. Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. Science **309:**1090–1092.
10. **Wilson, W. H., et al.** 2002. Isolation of viruses responsible for the demise of an Emiliania huxleyi bloom in the English Channel. J. Mar. Biol. Assoc. U. K. **82:**369–377.
11. **Wilson, W. H., J. L. Van Etten, and M. J. Allen.** 2009. The Phycodnaviridae: the story of how tiny giants rule the world. Curr. Top. Microbiol. Immunol. **328:**1–42.

Journal of Virology

Draft Genome Sequence of Four Coccolithoviruses: Emiliania huxleyi Virus EhV-88, EhV-201, EhV-207, and EhV-208

Jozef I. Nissimov, Charlotte A. Worthy, Paul Rooks, Johnathan A. Napier, Susan A. Kimmance, Matthew R. Henn, Hiroyuki Ogata and Michael J. Allen

J. Virol. **2012, 86(5):2896. DOI: 10.1128/JVI.07046-11.**

JVI Journals.ASM.org

GENOME ANNOUNCEMENT

Draft Genome Sequence of Four Coccolithoviruses: *Emiliania huxleyi* Virus EhV-88, EhV-201, EhV-207, and EhV-208

Jozef I. Nissimov,[a] Charlotte A. Worthy,[a,b] Paul Rooks,[a] Johnathan A. Napier,[b] Susan A. Kimmance,[a] Matthew R. Henn,[c] Hiroyuki Ogata,[d] and Michael J. Allen[a]

Plymouth Marine Laboratory, Plymouth, United Kingdom[a]; Department of Biological Chemistry, Rothamsted Research, Harpenden, Hertfordshire, United Kingdom[b]; The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA[c]; and Structural and Genomic Information Laboratory, Mediterranean Institute of Microbiology (IFR-88), Aix-Marseille University, Marseille, France[d]

The *Coccolithoviridae* are a group of viruses which infect the marine coccolithophorid microalga *Emiliania huxleyi*. The *Emiliania huxleyi* viruses (known as EhVs) described herein have 160- to 180-nm diameter icosahedral structures, have genomes of approximately 400 kbp, and consist of more than 450 predicted coding sequences (CDSs). Here, we describe the genomic features of four newly sequenced coccolithoviruses (EhV-88, EhV-201, EhV-207, and EhV-208) together with their draft genome sequences and their annotations, highlighting the homology and heterogeneity of these genomes to the EhV-86 model reference genome.

The coccolithoviruses infect *Emiliania huxleyi*, a globally distributed bloom-forming marine microalga. The abundance of *Emiliania huxleyi* viruses (EhVs) in natural seawater can typically reach 10^7 ml^{-1} in bloom conditions and 10^8 to 10^9 ml^{-1} under laboratory culture (6). Following the full sequencing of the model virus EhV-86 (isolated in 1999 from the English Channel) and the partial sequencing of EhV-163 (isolated from a Norwegian fjord the following year), we have recently undertaken to sequence the remaining English Channel isolates currently contained within the Plymouth Virus Collection (PVC) (1). The draft genomes of EhV-84, EhV-203, and EhV-202 have been described previously (4, 5). Here, we present the draft genomes of the remaining four coccolithoviruses in the PVC: EhV-88, EhV-201, EhV-207, and EhV-208.

EhV-88 was isolated from the English Channel (50°15′N/04°13′W) from a depth of 5 m in 1999, while EhV-201, EhV-207, and EhV-208 were isolated from the English Channel (49°56′N/04°19′W, 50°15′N/04°13′W, and 50°15′N/04°13′W, respectively) from a depth of 2 to 15 m in 2001 (1, 8). Their icosahedral virion structure and morphology are similar to those of other coccolithoviruses and of phycodnaviruses in general (9). Phylogenetic analysis of available major capsid protein (MCP) and DNA *pol* gene sequences indicates that the closest relatives to EhV-88 are EhV-84 and EhV-86 (4, 7), while EhV-201, EhV-207, and EhV-208 are closely related to EhV-203 (1, 8).

Genome sequencing, finishing, and annotation were performed by the Broad Institute. The genomes were sequenced using the 454 FLX pyrosequencing technology platform (Roche/454, Branford, CT). Library construction and sequencing were performed as previously described (2). General protocols for library construction can be found at http://www.broadinstitute.org/annotation/viral/Phage/Protocols.html. *De novo* genome assembly of resulting reads was performed using the Newbler v2.3 assembly software package (2). A total of 74,782, 78,268, 33,894, and 85,422 reads were produced and assembled into 8, 7, 16, and 17 contigs, comprising 396,598 bp, 406,701 bp, 420,391 bp, and 409,403 bp, for EhV-88, EhV-201, EhV-207, and EhV-208, respectively. Genes were identified using the Broad Institute's Automated Phage Annotation Protocol (2). Additional gene prediction analysis and functional annotation was performed within the Integrated Microbial Genomes-Expert Review (IMG-ER) platform (3).

General features of the genomes include nucleotide compositions of 40.18%, 40.46%, 40.49%, and 40.42% G+C and 475, 451, 473, and 455 predicted protein coding sequences (CDSs) for EhV-88, EhV-201, EhV-207, and EhV-208, respectively. EhV-201, EhV-207, and EhV-208 have six tRNAs (Arg, Asn, Gln, Glu, Leu, and Lys), while EhV-88 has five (Arg, Asn, Gln, Ile, and Lys). EhV-84 is most similar to the model virus EhV-86, encoding 231 CDSs with identical homologues in the EhV-86 genome. In contrast, EhV-201, EhV-207, and EhV-208 have just 26, 29, and 25 CDSs sharing 100% identity with their EhV-86 homologues, respectively. The majority of CDSs not shared with EhV-86 encode hypothetical proteins of unknown function. Those of predicted function include those encoding glycosyltransferase (EhV-201, EhV-207, and EhV-208), methyltransferase (EhV-88, EhV-207, EhV-208), and RNase (EhV-88 and EhV-201). The genomes of these viruses will provide new insights into coccolithovirus evolution and their coevolution and interaction with *Emiliania huxleyi*.

Nucleotide sequence accession numbers. The nucleotide sequence accession numbers for the draft genomes sequences have been deposited in GenBank under accession numbers JF974310, JF974311, JF974317, and JF974318.

ACKNOWLEDGMENTS

This research was funded in part by the Gordon and Betty Moore Foundation through a grant to the Broad Institute (M.R.H.) and through the NERC Oceans 2025 program (M.J.A.). J.I.N. is supported by an NERC studentship, and C.A.W. is supported by a BBSRC Industrial CASE studentship sponsored by PML Applications. H.O. is supported by IGS/CNRS and ANR (grants ANR-09-PCS-GENM-218, ANR-08-BDVA-003).

Received 8 December 2011 **Accepted** 9 December 2011

Address correspondence to Michael J. Allen, mija@pml.ac.uk.

doi:10.1128/JVI.07046-11

Samples G3265, G3266, G3249, and G3250 were sequenced, assembled, and annotated at the Broad Institute.

We thank Konstantinos Mavromatis from JGI who assisted with information regarding the IMG/ER platform and the Broad Institute's Genome Sequencing Platform, Finishing Team, and Annotation Team for their efforts to generate the genomic data. Jean Devonshire and the Centre for Bioimaging at Rothamsted provided technical support for transmission electron microscopy.

REFERENCES

1. **Allen MJ, Martinez-Martinez J, Schroeder DC, Somerfield PJ, Wilson WH.** 2007. Use of microarrays to assess viral diversity: from genotype to phenotype. Environ. Microbiol. **9**:971–982.
2. **Henn MR, et al.** 2010. Analysis of high-throughput sequencing and annotation strategies for phage genomes. PLOS One **5**:e9083.
3. **Markowitz VM, et al.** 2009. IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics **25**:2271–2278.
4. **Nissimov JI, et al.** 2011. Draft genome sequence of the coccolithovirus EhV-84. Stand. Genomic Sci. **5**:1–11.
5. **Nissimov JI, et al.** 2011. Draft genome sequence of the coccolithovirus *Emiliania huxleyi* virus 203. J. Virol. **85**:13468–13469.
6. **Schroeder DC, Oke J, Hall M, Malin G, Wilson WH.** 2003. Virus succession observed during an *Emiliania huxleyi* bloom. Appl. Environ. Microbiol. **69**:2484–2490.
7. **Wilson WH, et al.** 2005. Complete genome sequence and lytic phase transcription profile of a coccolithovirus. Science **309**:1090–1092.
8. **Wilson WH, et al.** 2002. Isolation of viruses responsible for the demise of an *Emiliania huxleyi* bloom in the English Channel. J. Mar. Biol. Assoc. U.K. **82**:369–377.
9. **Wilson WH, Van Etten JL, Allen MJ.** 2009. The Phycodnaviridae: the story of how tiny giants rule the world. Curr. Top. Microbio. Immunol. **328**:1–42.

Draft Genome Sequence of the Coccolithovirus Emiliania huxleyi Virus 202

Jozef I. Nissimov, Charlotte A. Worthy, Paul Rooks, Johnathan A. Napier, Susan A. Kimmance, Matthew R. Henn, Hiroyuki Ogata and Michael J. Allen

***J. Virol.* 2012, 86(4):2380. DOI: 10.1128/JVI.06863-11.**

JVI Journals.ASM.org

GENOME ANNOUNCEMENT

Draft Genome Sequence of the Coccolithovirus *Emiliania huxleyi* Virus 202

Jozef I. Nissimov,[a] Charlotte A. Worthy,[a,b] Paul Rooks,[a] Johnathan A. Napier,[b] Susan A. Kimmance,[a] Matthew R. Henn,[c] Hiroyuki Ogata,[d] and Michael J. Allen[a]

Plymouth Marine Laboratory, Plymouth, United Kingdom[a]; Department of Biological Chemistry, Rothamsted Research, Harpenden, Herts, United Kingdom[b]; The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA[c]; and Structural and Genomics Information Laboratory, CNRS-UPR2589, Mediterranean Institute of Microbiology (IFR-88), Aix-Marseille University, Marseille, France[d]

***Emiliania huxleyi* virus 202 (EhV-202) is a member of the *Coccolithoviridae*, a group of viruses that infect the marine coccolithophorid *Emiliania huxleyi*. EhV-202 has a 160- to 180-nm-diameter icosahedral structure and a genome of approximately 407 kbp, consisting of 485 coding sequences (CDSs). Here we describe the genomic features of EhV-202, together with a draft genome sequence and its annotation, highlighting the homology and heterogeneity of this genome in comparison with the EhV-86 reference genome.**

Emiliania huxleyi is a globally distributed marine microalgae which forms blooms that can cover up to 100,000 km^2 (10). Coccolithoviruses have been shown to be a major cause of bloom termination, and their role in global biogeochemical cycling is gaining increasing attention (3). Coccolithovirus abundances can reach 10^7 ml^{-1} in natural seawater under bloom conditions (8). *Emiliania huxleyi* virus 86 (EhV-86), the model virus, was isolated in 1999 from the English Channel (50°13.79′N/04°9.59′W) and was sequenced in its entirety in 2005 (9, 10). EhV-163, isolated from a Norwegian fjord, was partially sequenced in 2006 (2), and more recently, permanent draft genomes for EhV-84 and EhV-203 (both isolated in the English Channel) were obtained (6, 7). *Emiliania huxleyi* virus 202 (EhV-202) was isolated from the English Channel (50°00.36′N/04°18.87′W) from a depth of 15 m on 27 July 2001 (1, 10). Like other coccolithoviruses and phycodnaviruses in general, EhV-202 has an icosahedral virion structure (11). Phylogenetic analysis of available major capsid protein (MCP) gene sequences indicates that the closest relatives of EhV-202 are EhV-201, EhV-203, EhV-207, and EhV-208 (1, 10).

EhV-202 genome sequencing, finishing, and annotation were performed by the Broad Institute. The genome was sequenced using the 454 FLX pyrosequencing technology platform (Roche/454, Branford, CT). Library construction and sequencing were performed as previously described (4). General protocols for library construction can be found at www.broadinstitute.org/annotation/viral/Phage/Protocols.html. Genome assembly of resulting reads was performed using the Newbler v2.3 software package (4). A total of 31,250 reads were produced and assembled into 12 contigs comprising 407,516 bp, with a maximum contig length of 137,441 bp, an average contig length of 33,868 bp, and an average sequencing coverage of approximately 32.5 times (± 5.8). Genes were identified using the Broad Institute's automated phage annotation protocol (4). Additional gene prediction analysis and functional annotation were performed within the Integrated Microbial Genomes-Expert Review (IMG-ER) platform (5).

General features of the EhV-202 genome sequence include a nucleotide composition of 40.30% G+C, a total of 485 predicted protein coding sequences (CDSs), and three tRNA genes (Arg, Asn, and Gln). Of the 485 CDSs annotated by IMG-ER, 93 have been annotated with functional product predictions. Four hundred one CDSs have homologues (>26% identity) within the EhV-86 genome, yet only one shares 100% identity. EhV-202 is the most genetically distinct coccolithovirus strain sequenced to date. EhV-202 contains 83 unique CDSs; while the majority are of unknown function, two CDSs have homology to glycosyltransferases, one to a zinc finger protein, one to ADP ribose pyrophosphatase, one to lipopolysaccharide (LPS) biosynthesis protein, and one to polyubiquitin. Among the EhV-202 CDSs displaying the lowest similarities to their EhV-86 homologues is a putative phosphoglycerate mutase (62.30%), an ATPase (60.86%), and a CDS containing a PDZ domain (42.62%). The CDSs displaying the highest similarities to their EhV-86 counterparts include MCP (99.80%), sterol desaturase (99.51%), and DNA polymerase (92.13%).

Nucleotide sequence accession number. The draft genome sequence has been deposited in GenBank under accession number HQ634145.

ACKNOWLEDGMENTS

This research was funded in part by the Gordon and Betty Moore Foundation through a grant to the Broad Institute (M.R.H.) and through the NERC Oceans 2025 program (M.J.A.). J.I.N. is supported by a NERC studentship; C.A.W. is supported by a BBSRC Industrial CASE studentship sponsored by PML Applications. H.O. is supported by the IGS/CNRS and ANR (grants ANR-09-PCS-GENM-218 and ANR-08-BDVA-003).

We thank Konstantinos Mavromatis from JGI, who assisted with information regarding the IMG-ER platform, and the Broad Institute's genome sequencing platform, finishing team, and annotation team for their efforts to generate the genomic data. Jean Devonshire and the Centre for Bioimaging at Rothamsted provided technical support for transmission electron microscopy. Sample G4014 was sequenced, assembled, and annotated at the Broad Institute.

Received 21 November 2011 **Accepted** 28 November 2011

Address correspondence to Michael J. Allen, mija@pml.ac.uk.

doi:10.1128/JVI.06863-11

REFERENCES

1. **Allen MJ, Martinez-Martinez J, Schroeder DC, Somerfield PJ, Wilson WH.** 2007. Use of microarrays to assess viral diversity: from genotype to phenotype. Environ. Microbiol. **9**:971–982.
2. **Allen MJ, Schroeder DC, Donkin A, Crawfurd KJ, Wilson WH.** 2006. Genome comparison of two coccolithoviruses. Virol. J. **3**:15.
3. **Coyle KO, Pinchuk AI.** 2002. Climate-related differences in zooplankton density and growth on the inner shelf of the southeastern Bering Sea. Prog. Oceanogr. **55**:177–194.
4. **Henn MR, et al.** 2010. Analysis of high-throughput sequencing and annotation strategies for phage genomes. PloS One **5**:e9083.
5. **Markowitz VM, et al.** 2009. IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics **25**:2271–2278.
6. **Nissimov JI, et al.** 2011. Draft genome sequence of the coccolithovirus EhV-84. Stand. Genomic Sci. **5**:1.
7. **Nissimov JI, et al.** Draft genome sequence of the coccolithovirus *Emiliania huxleyi* virus 203. J. Virol. **85**:13468–13469.
8. **Schroeder DC, Oke J, Hall M, Malin G, Wilson WH.** 2003. Virus succession observed during an *Emiliania huxleyi* bloom. Appl. Environ. Microbiol. **69**:2484–2490.
9. **Wilson WH, et al.** 2005. Complete genome sequence and lytic phase transcription profile of a *Coccolithovirus*. Science **309**:1090–1092.
10. **Wilson WH, et al.** 2002. Isolation of viruses responsible for the demise of an *Emiliania huxleyi* bloom in the English Channel. J. Mar. Biol. Assoc. UK **82**:369–377.
11. **Wilson WH, Van Etten JL, Allen MJ.** 2009. The Phycodnaviridae: the story of how tiny giants rule the world. Curr. Top. Microbiol. Immunol. **328**:1–42.