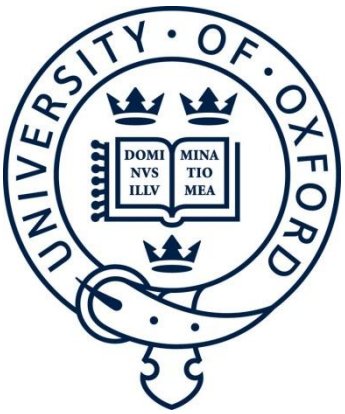
Genetic variability of the microalga *Emiliania huxleyi* (Haptophyta): a temporal and geographical study

Cecilia Balestreri

A thesis submitted to the University of Oxford

for the degree of Doctor of Philosophy



Wolfson College and Department of Earth Sciences
University of Oxford

In collaboration with

The Marine Biological Association of the UK

September 2015

DECLARATION OF AUTHORSHIP

Name: CECILIA BALESTRERI **Candidate number: 592092**

College: WOLFSON **Supervisors: Prof. Rosalind Rickaby**
Dr Declan Schroeder
Prof. Colin Brownlee

Title of thesis:

Genetic variability of the microalga *Emiliania huxleyi* (Haptophyta): a temporal and geographical study

Word count: 41,873

Please tick to confirm the following:

I have read and understood the University's disciplinary regulations concerning conduct in examinations and, in particular, the regulations on plagiarism. ☑

I have read and understood the Education Committee's information and guidance on academic good practice and plagiarism. ☑

The thesis I am submitting is entirely my own work except where otherwise indicated. ☑

It has not been submitted, either partially or in full, for another Honour School or qualification of this University, or for a qualification at any other institution. ☑

I have clearly indicated the presence of all material I have quoted from other sources, including any diagrams, charts, tables or graphs. ☑

I have clearly indicated the presence of all paraphrased material with appropriate references. ☑

I have acknowledged appropriately any assistance I have received in addition to that provided by my supervisors. ☑

I have not copied from the work of any other candidate. ☑

I have not used the services of any agency providing specimen, model or ghostwritten work in the preparation of this thesis. ☑

I agree to retain an electronic copy of this work until the publication of my final examination result, except where submission in hand-written format is permitted. ☑

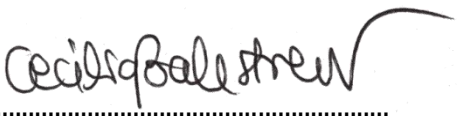
I agree to make any such electronic copy available to the examiners should it be necessary to confirm my word count or to check for plagiarism. ☑

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 financed by Natural Environment Research Council (NERC) and The Marine Biological Association of the UK (MBA, Plymouth) and carried out in collaboration with MBA.

Relevant scientific seminars and conferences were regularly attended at with work was often presented; external institutions were visited for consultation purposes and several papers prepared for publication.

A list of conference attended and publications is found in Appendix III.

Candidate's signature:  Date: ….8/02/2016…..

Acknowledgements

I am extremely grateful to my supervisor Dr Declan Schroeder for his guidance, presence and encouragement during the last four years. He has been a role model and has supported me in time of need. I am also deeply grateful to my other supervisors; Prof. Colin Brownlee for his expertise, advice and help, and Prof. Ros Rickaby for her kind suggestions and help in dealing with department bureaucracy. A special acknowledgement also goes to Dr Andrea Highfield, for her enormous help and patience, for being my mentor in the lab as well as being a good friend.

I want to thank my family and my boyfriend Matthew, for their support, encouragement and their love. They had to deal with me during the most stressful parts of my doctorate, but never left my side.

Another special thanks for my good friends and lab mates Leela Chakravarti, Gideon Mordecai and Flavia Flaviani. This PhD wouldn't have been the same without you. We had great fun and I really appreciated the time we spent together.

I also want to acknowledge Dr El Mahdi Bendif, Dr Joana Projecto-Garcia, Sarah Magozzi, Angela Ward, Matt Hall, Dr Grazyna Durak, Dr Rowena Stern, Dr Priscilla Licandro, Dr Karim Hahn-Lüchmann, Dr Consuelo Carbonell-Moore, Tingting Shi, Dr Lizeth Aveñdano, Dr Manal Al-Kandari, Dr Joanna Schroeder and Dr Stacy Krueger-Hadfield, for their help, advice and friendship.

A number of people helped and advised me in these four years, a great appreciation goes to Dr Jeremy Young, Prof. Toby Tyrrell, Dr Renee Lee, Dr Barney Balch, Dr Glen Tarran, Dr Sally Thorpe, Dr Tim Smyth, Claire Widdicombe and Glenn Harper.

Thank you to all the participants and the crew members of the cruises D366, RR1202 and JR271. They turned three working trips into three unforgettable adventures.

Finally, but not less important, I want to thank all the MBA staff members, researchers and students. They are great people and at the MBA I always felt like at home.

Contents

List of abbreviation and acronyms

ACC – Antarctic Circumpolar Current

AMOVA – Analysis of molecular variance

ARB - ARB project (latin, "arbor"= tree)

AVHRR – Advance Very High Resolution Radiometer

blastn – Nucleotide BLAST (Basic Local Alignment Search Tool)

bp – base pair

CBD - Convention on Biological Diversity

CMM – Coccolith Morphology Motif

CT – Cycle Threshold

CTD – Conductivity Temperature Depth

DMS - Dimethyl sulphide

DMSP - Dimethylsulphoniopropionate

DNA - Deoxyribonucleic acid

e.g. – exempli gratia

EMBL - European Molecular Biology Laboratory

FD – Fold Difference

FSC – Forward Side Scatter

gcm^{-3} – grams/ cubic centimetre

GEOPS – Generic Earth Observation Processing System

GPA – Calcium binding protein gene

hclust – Hierarchical cluster analysis

Hi Di formamide - Highly deionized formamide

HPLC high - performance liquid chromatography

i.e. – id est

IPCC - Intergovernmental Panel on Climate Change

kV – kilo volt

kYa - Thousand Years Ago

LD cycle – Light/dark cycle

LSU - Large Subunit of the nuclear ribosomal RNA gene complex

m^2s – squared metre * second

Ma – Million Years

MANOVA – Multivariate Analysis of Variance

MBA – Marine Biological Association of the UK

mL – milli litre

MLGs – Multilocus genotypes

mM – milli molar

MODIS - Moderate Resolution Imaging Spectroradiometer

mRNA – Messenger RNA

mtDNA – Mitochondrial DNA

NASA – National Atmospheric and Space Administration

nB – *E. huxleyi* bloom period

NCBI - National Center for Biotechnology Information

NEODAAS – NERC Earth Observation Data Acquisition and Analysis Service

NERC - Natural Environment Research Council

ng – nano gram

NGS – Next Generation Sequencing

NHF – Net Heat Flux

NLSST – Non-linear SST

nm – nano metre

NOAA – National Oceanic and Atmospheric Administration

NPP - Net Primary Production

nT – Total number of sampling times

OSTIA – Operational SST and Ice Analysis

pCO_2 – Partial pressure CO_2

PCR – Polymerase Chain Reaction

pmol – pico mole

qPCR - Quantitative PCR or real-time PCR

R/V – Research Vessel

rbcL - Ribulose bisphosphate carboxylase large chain

rDNA - Ribosomal DNA

RNA - Ribonucleic acid

rRNA - Ribosomal ribonucleic acid

RuBisCO - Ribulose-1,5-bisphosphate carboxylase/oxygenase

SEM – Scanning Electron Microscope

sp. - species

SST - Sea Surface Temperature

SSU - Small subunit of the nuclear ribosomal RNA gene complex

Taq – Polymerase enzyme GoTaq Flexi

UK – United Kingdom

UKOA – UK Ocean Acidification

Var. – Variety

Ver. – Version

WCO – Western Channel Observatory

WEC – Western English Channel

µL – micro litre

µm – micro metre

µmol – micro mole

List of tables

Page

List of figures

Page

List of supplementary figures

Page

Thesis abstract

Genetic variability of the microalga *Emiliania huxleyi* (Haptophyta): a temporal and geographical study

*By **Cecilia Balestreri**, Wolfson College*

For the Degree of Doctor of Philosophy, September 2015

The Earth's climate is changing at a pace that was never observed before, and this may result in species migration to new habitats or, more drastically, to extinction. Nevertheless, certain species which have a fast turnover might evolve and become resilient to the effects of a rapidly changing environment. This resiliency-scenario better applies to species with large population size and rapid generation times, such as the coccolithophore *Emiliania huxleyi,* which plays a fundamental role in the marine ecosystem since it produces calcium carbonate coccoliths and it is responsible of circa 80% of carbonate precipitation in seawater. *E. huxleyi* shows both morphological and genetic intraspecific variability, with A and B being its two main coccolith morphological types, corresponding to its two main genotypes characterised by coccolith morphology motif (CMM) I & II respectively.

Here I present the results of a temporal and geographical study, aimed to investigate the extant diversity of *E. huxleyi* in both the Northern and Southern hemispheres and its genetic standing stock over a 6-year time period. I found that only genotypes CMM I & CMM IV persist and dominate throughout the years, and they are both representative of *E. huxleyi* A morphotype, which dominates North-eastern Atlantic, Western English Channel (WEC) and North Sea *E. huxleyi* bloom events.

Additionally, my study confirms the genetic variability in the global *E. huxleyi* community and reveals that the intraspecific variability is defined by geophysical features, with *E. huxleyi* morphotype A being dominant in temperate regions (10-18 ºC), while morphotypes B & B/C as defined by CMM II & IIb, respectively, being dominant in polar regions (<10 ºC).

Finally my results show that cultures maintained in laboratory collections might not be representative of real extant stocks, and therefore may not necessarily describe the genetic composition of wild biogeographic populations.

CHAPTER 1

General introduction

1.1 Phytoplankton Biodiversity

Biodiversity is a concept which embraces many aspects of the life on our planet, related to different habitats, communities, species and genotypes (Medlin and Kooistra, 2010). It can be described as the total genetic information on Earth or in any part of it and it can be distinguished from diversity, which is constituted by the components that are active and abundant at one particular time and space (Pedrós-Alió, 2006).

The ocean efficiency to recover from perturbation, maintain water quality and provide food is diminished by biodiversity loss (Worm *et al.*, 2006). The recent rapid climate change has modified the population structure and the biogeographic ranges of many intertidal indicator species (e.g. Mieszkowska *et al.*, 2006), as well as the response of the marine pelagic community, leading to a discrepancy between trophic levels and functional groups (e.g. Edwards and Richardson, 2004).

The Convention on Biological Diversity (CBD) was created in 2002 to achieve a reduction in the biodiversity loss by the year 2010. In areas where biodiversity is affected by increasing pressure-symptoms, the indicators of the state of biodiversity, such as extinction risk, community composition and population dynamics, and habitat extent and condition exhibited decline and no reductions in rate (Butchart *et al.*, 2010). The declines affected vertebrates and habitat specialist birds; shorebird populations worldwide; extent of forest; mangroves; seagrass beds; and the condition of coral reefs. None showed considerable reductions in the rate of decline.

1.1.1 The diversity-stability debate

Biodiversity is closely related to ecosystem functioning, emerging as one of the most controversial research areas in ecology over the last two decades (Loreau *et al.*, 2002). The diversity-stability debate divided scientists between the supporters of the view that ecosystem-stability is enhanced by the presence of diverse communities and their challengers who believed that community dynamics could be destabilized by diversity (McCann, 2000). This debate played a key role in understanding whether the loss of inter- and intra- species variability will affect the functioning of ecosystems (Cardinale *et al.*, 2012).

The Gaia-hypothesis speculates that the Earth's ecosystem is resistant and resilient over long time scales and at global spatial scales, whether or not the present organisms living in it will persist over such scales (Free and Barton, 2007). Nevertheless, it is of fundamental importance to assess whether and how the ecological processes which control the fluxes of energy, nutrients and organic matter through the environment will be negatively altered by the loss of biodiversity.

1.1.2 The importance of phytoplankton

Life evolved predominantly in the oceans (Sagan, 1961; Baross and Hoffman, 1985), and their role in the Earth's health is crucial. Planktonic communities underpin many biogeochemical processes on a global scale (Cotner and Biddanda, 2002; Allison and Martiny, 2008), and they are the base of the marine food web. Many studies conducted in the mid-1990s suggested that biomass production and nutrient cycling respond strongly to changes in biological diversity (Cardinale *et al.*, 2012). Furthermore, key physical processes such as changing sea surface temperatures and

upwelling events are factors that can impact the biodiversity of certain regions determining a variation in the primary productivity.

Global surface temperature has increased by about 0.2°C per decade in the past 30 years (Hansen *et al.*, 2006), and the elevated temperatures have resulted in changes in phytoplankton community structure (Feng *et al.*, 2009). The phenology (annually recurring life cycle events such as the timing of migrations and flowering) of many organisms may be impacted, resulting in decoupling of trophic interactions and altered food-web structure, ultimately leading to changes in ecosystem-level where the success of higher trophic levels depends on the synchronization with the planktonic production (Edwards and Richardson, 2004).

On land and in the oceans oxygenic photosynthesis is responsible for the production of virtually all organic matter (Field *et al.*, 1998). The availability of iron, nitrogen and light control the variation of both spatial and seasonal variation in photosynthetic biomass in the oceans. Consequently, regional and repetitive distributions of net primary production (NPP) are affected by physical (e.g. rain precipitation, photosynthetically active radiation, ocean circulation, and water-column stratification) and biological processes (e.g. species composition and their interactions in the communities)(Vitousek *et al.*, 1997; Field *et al.*, 1998).

Furthermore, phytoplankton convert dissolved CO_2 into organic carbon in the surface ocean and the fate of this carbon may be to be eventually rereleased to the atmosphere as respiratory CO_2 or exported into deeper oceans. The atmospheric CO_2 concentration is currently around 380 ppm, the highest it has reached in the last 20 million years (Berner, 1990) and according to the projections of future climate from forcing scenario experiments (IS92a of the IPCC 1996) the CO_2 concentration will double by around year 2060 (Cubasch et al., 2001). At the current pH level the

prevalent form of the carbon in the seawater is bicarbonate, $HCO3^-$ (Fabry et al., 2008). The dissolution of increasing CO_2 concentrations, with the consequent rise in hydrogen ions concentration and a reduced carbonate ion concentration, according to the reactions

$$CO_2 + H_2O \Leftrightarrow H_2CO_3$$
$$H_2CO_3 \Leftrightarrow H^+ + HCO_3^-$$
$$HCO_3^- \Leftrightarrow H^+ + CO_3^{2-}$$

is leading to a lower pH level in the surface ocean. This reduction in the pH value and the consequent alterations in fundamental chemical balances has been called 'ocean acidification' and it has concerned scientists in the last decades, because it has an impact first on seawater chemical speciation and biogeochemical cycles of many elements and compounds and subsequently potentially on marine organisms, especially those characterised by the calcification processes (Doney *et al.*, 2009; Guinotte and Fabry, 2008).

In order to understand the effects of changing climate on species diversity and resilience, it is necessary to assess the current biodiversity and understand the mechanisms underlying ecosystem dynamics. Assessing biodiversity is also of fundamental importance to understand how ecosystems are regulated and to evaluate productivity and other ecosystem services (Cardinale et al., 2012). Phytoplankton can be divided into functional groups according to how they influence the silica, carbon or nitrogen cycles. They are loosely classified as silicifiers, calcifiers and diazotrophs, respectively, which play specific roles in the marine ecosystem and a number of studies have demonstrated how these groups are sensitive to increasing CO_2 concentrations (Kroeker *et al.*, 2013, 2010; Beaufort *et al.*, 2011; Rost *et al.*, 2008; Fabry *et al.*, 2008) and to climate change.

Calcifiers are directly impacted by ocean acidification, as calcium carbonate dissolves in acidic conditions. Calcifiers are represented by numerous groups in the oceans, both in the animal kingdom and among the protists:

- *Foraminifera* are ameboid protists, the majority marine, which produce a calcium carbonate shell, some becoming quite elaborate in structure (Kennett and Srinivasan, 1983).
- *Dinoflagellates* are flagellate protists and some genera, such as *Ensiculifera*, *Pentapharsodinium* and *Scrippsiella* include species which have calcified or partially calcified resting cyst wall; although the walls are not exclusively calcareous (Head, 1996).
- *Calcifying algae*, such as the unicellular marine coccolithophores (Haptophyta division), the multicellular marine macrophyte Corallina spp. and Halimeda spp. and the freshwater alga Chara (Charophyta division) produce calcium carbonate and play a key role in the global carbon circle (Brownlee and Taylor, 2002).
- *Calcareous sponges* (phylum Porifera, class Calcarea) are characterized by spicules made from calcium carbonate in the form of calcite or aragonite (Barnes, 1963). Aragonite in seawater dissolves more readily than calcite (Guinotte and Fabry, 2008).
- *Stony corals* (phylum Cnidaria, class Anthozoa, order Scleractinia) are marine animals formed by a founding polyp which settles on the seabed and starts to secrete a calcium carbonate skeleton to protect its soft body, creating a large colony (Bourne, 1900).

- *Barnacles* (phylum Arthropoda, subphylum Crustacea) are marine arthropods, which tend to live in shallow and tidal waters. Their external structure is strengthened by calcareous plates (Anderson, 1994).
- *Sea snails* (phylum Mollusca, class Gastropoda) are marine gastropods with shells (Bouchet et al., 2005).
- *Bivalves* (phylum Mollusca, class Bivalvia), such as clams, mussels, oysters and scallops, constitute a class of marine and freshwater molluscs that have laterally compressed bodies enclosed by a shell. Shells may contain both aragonite and calcite in separate monomineralic layers, or they may be totally made of aragonite (Kennedy *et al.*, 1969).
- *Echinoderms* (phylum Echinodermata), such as brittle stars, starfish and sea urchins, have a mesodermal skeleton composed of calcareous plates or ossicles, composed mineralogically of a crystal of calcite (Bäuerlein et al., 2009).
- *Teleosts or bony fish* (phylum Chordata, superclass Osteichthyes) form calcium carbonate crystals as the result of absorbing water across the intestine from the surrounding salted seawater (Grosell, 2014). Estimates suggested that bony fish produce 40 million–110 million tonnes of calcium carbonate per year, and this range accounts for 3%–15% of the estimated total (Kwok, 2009).

Although there are no accurate calculations of the relative contributions of these organisms on the global calcium carbonate production, the pelagic foraminifera and coccolithophores together are likely to represent the main producers. It has been estimated that the coccolithophores represent up to 50% of all current global oceanic

calcium carbonate production (Beardall and Raven, 2013; Brownlee and Taylor, 2002).

A study by Liu *et al.*, (2009) demonstrated that the haptophyte algal division, which includes the coccolithophores, dominates the chlorophyll a-normalized phytoplankton standing stock in modern oceans. Their findings show that the evolution of complex eukaryotic cells is a crucial force in the functioning of the biosphere and that, ecologically, they may have substantially impacted the oceanic carbon pump (Liu *et al.*, 2009).

1.1.3 Coccolithophores diversity

Coccolithophores form one important group of the Haptophyta phylum: they are characterised by beautiful calcium carbonate structures, they are major sediment formers and indicators of palaeoceanographic change, as stated in a study by Young and his colleagues (2003), who redefined the coccolithophores taxonomy, which is composed of six main orders:

- **Isochrysidales**: include Noelaerhabdaceae and the extinct family Prinsiaceae, based on flagella characters, biochemical characters and molecular genetics.

  Noelaerhabdaceae are characterised by placoliths, with a grill in the central area, and monocyclic proximal shields (Figure 1.1). The representatives of this order have been cultured extensively and their life-cycle is well known. Their dominant phase is diploid, non-motile, and heterococcoliths-bearing (Young *et al.*, 2003).

  Emiliania huxleyi, *Gephyrocapsa oceanica* (Figure 1.2) and *Reticulofenestra parvula* are the preeminent representatives of this order.

- **Coccosphaerales**: include the two major families Coccolithaceae and Calcidiscaceae, plus the family Pleurochrysidaceae. The grouping was made based on their structural types and supported by molecular genetic studies. Additionally the Hymenomonadaceae were added to this order, because they are closely related to the Pleurochrysidaceae.

  Coccolithaceae dominant phase of life-cycle is non-motile, with placoliths heterococcoliths (Figure 1.1). Their central area is connected by disjunct structures, which are used to define various fossil genera. Calcidiscaceae have proximal shields formed of a single layer of elements with sub-radial sutures. The connection between the proximal and distal shields is weak and they are often found separated. Pleurochrysidaceae have a rim structure which appears as a simplified version of that of the Coccolithaceae. Hymenomonadaceae are small littoral and fresh water coccolithophores. Their coccoliths are muroliths with an open central area (Figure 1.1; Young *et al.*, 2003).

  The main representatives of this order include *Coccolithus pelagicus* (Figure 1.2), *Calcidiscus leptoporus* (Figure 1.2) and *Pleurochrysis carterae*.

- **Zygodiscales**: include Helicosphaeraceae, Pontosphaeraceae and the extinct family Zygodiscaceae. These families have evolutionary connections.

  Helicosphaeraceae are ellipsoidal with a prominent flagellar opening and the coccoliths are arranged spirally around the coccosphere. Pontosphaeraceae are sub-spherical, non-motile. The coccoliths are muroliths (Figure 1.1); their central area shows a variable number of perforations (Young *et al.*, 2003).

  In this order we can distinguish *Helicosphaera carteri* (Figure 1.2), *Pontosphaera discopora* (Figure 1.2) and *Scyphosphaera apsteinii*.

- **Syracosphaerales**: include the families Syracosphaeraceae, Rhabdosphaeraceae and Calciosoleniaceae, plus the genus of uncertain placement *Coronosphaera*. The grouping was made based on the coccolith structure; typically the members of this order show more than one coccolith type (polymorphism). Coccoliths are complex, consisting of three components: a rim, a radial lath cycle and an axial structure in their centre.

  Syracosphaeraceae are motile, with elaborate coccospheres, often showing dithecatism (i.e. development of distinct inner and outer layers of coccoliths). Rhabdosphaeraceae are typically spine-bearing and non-spine-bearing coccoliths with similar shields. Calciosoleniaceae are motile, with elongate, fusiform coccospheres (Young *et al.*, 2003).

  The only species which have been cultured from this very diverse order are *Syracosphaera pulchra* (Figure 1.2), *Algirosphaera robusta* and *Coronosphaera mediterranea* (Figure 1.2; Young *et al.*, 2003). Other representatives of this order include *Acanthoica quattrospina* (Figure 1.2), *Calciopappus caudatus*, *Michaelsarsia elegans*, *Ophiaster hydroideus*, *Calciosolenia brasiliensis*, and *Rhabdosphaera clavigera* (Figure 1.2).

- **Heterococcolith and genera of uncertain placement** do not show clear affinities to any other order. They include Alisphaeraceae, Umbrellosphaeraceae, Papposphaeraceae families and narrow rimmed placoliths genera.

  Alisphaeraceae have asymmetrical coccoliths with edge directed toward flagellar opening extended into a protrusion and the entire coccolith is formed of rim units. Umbrellosphaeraceae have coccoliths which consist of a funnel-shape distal part on a flat base. The funnel elements are continuous with the

basal plate elements of the central area. Papposphaeraceae are small (4-6 μm) and lightly-calcified, and their haptonema (i.e. cylinder-shaped organelle attached near the flagella and unique to the group) is usually longer than the cell. They are usually recorded at high-latitudes (Young *et al.*, 2003).

Alisphaera gaudii (Figure 1.2), *Canistrolithus* sp., *Polycrater galapagensis*, *Umbrellosphaera irregularis* (Figure 1.2), *Tetralithoides quadrilaminata*, *Papposphaera lepida*, *Picarola margalefii* (Figure 1.2) are some representatives of these groups.

- **Nannoliths** include Braarudosphaeraceae and Ceratolithaceae families, plus nannoliths of uncertain placement. The structures characterising this order are about the same size of coccoliths, but they lack definite coccolith affinities. Braarudosphaeraceae are not grown in culture, they contain visible chloroplasts and, at present day, they occur in shelf environments, usually under lower salinity conditions. Ceratolithaceae are characterised by horseshoe-shaped nannoliths (Figure 1.1; Young *et al.*, 2003). *Braarudosphaera bigelowii* (Figure 1.2) and *Ceratolithus cristatus* (Figure 1.2) are two representatives of these families.

  Other representatives of this order include *Florisphaera profunda* (Figure 1.2), *Gladiolithus flabellatus* and *Ericiolus* sp.

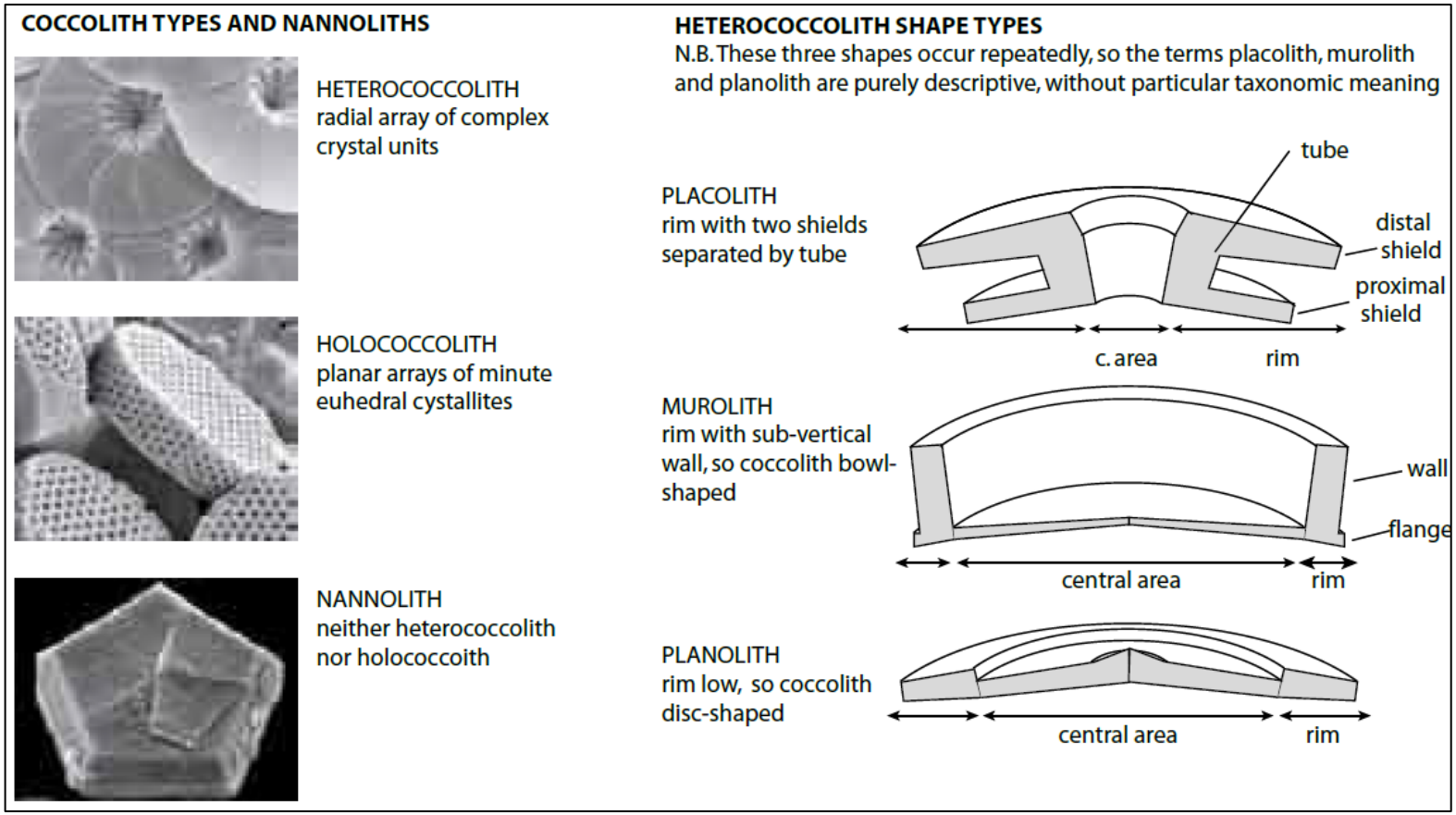


Figure 1.1 From Young *et al.* (2003); "Key terms of the coccolith morphology."

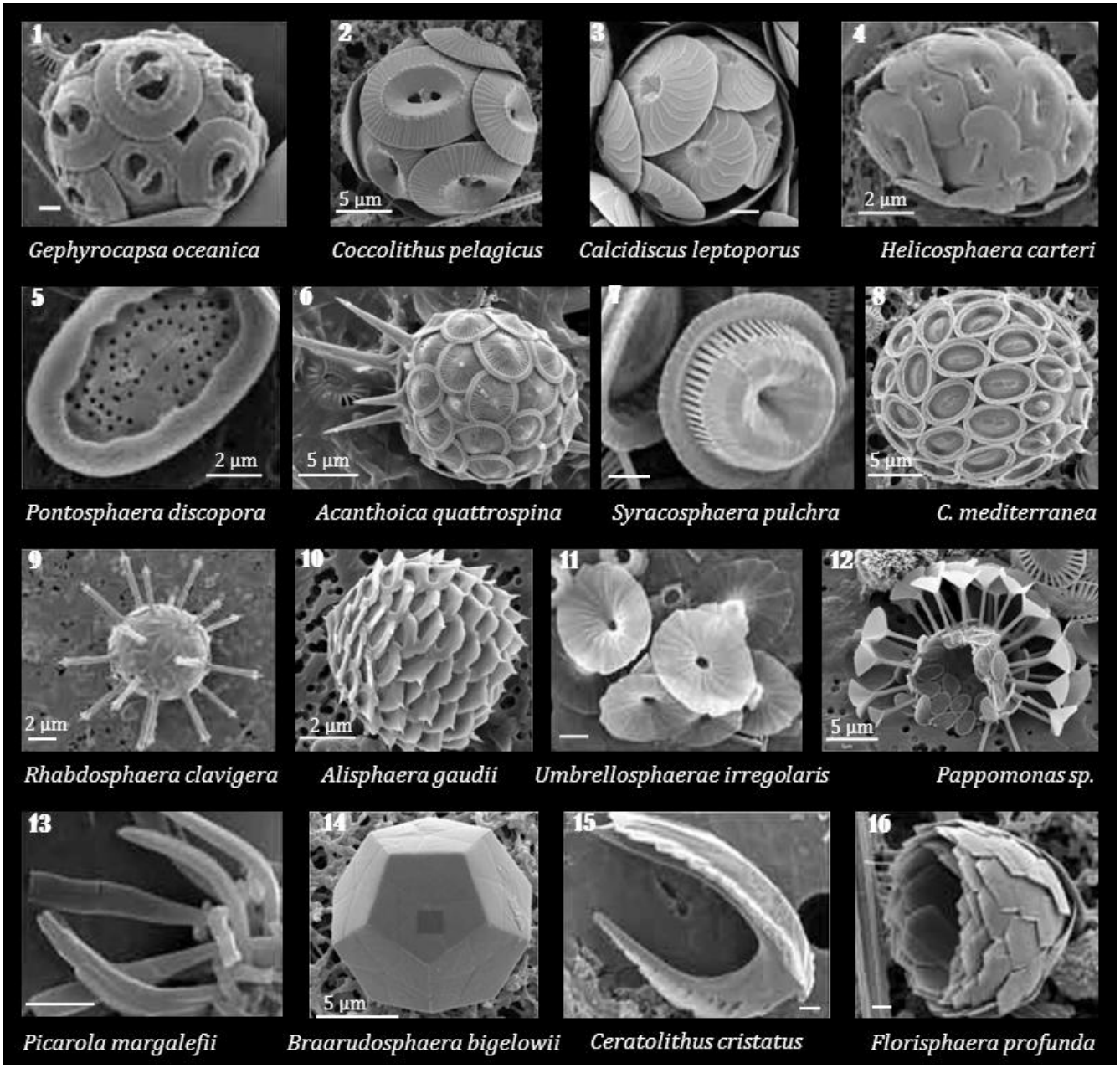


Figure 1.2 Some coccolithophore species. All the photos are taken from Young *et al.*, (2003). The scale is 1 µm if not differently indicated.

One of the most important and well-studied coccolithophore is *Emiliania huxleyi*. Calcification by this species is thought to exert a significant influence on net drawdown of atmospheric CO_2 by creating a net export of carbon to the seabed (Robertson *et al.*, 1994; Riebesell and Tortell, 2011; Westbroek *et al.*, 1993) during and following bloom events, because its coccoliths sink towards the bottom of the water column taking large amounts of organic carbon with them (i.e., ballast effect), where a significant proportion becomes lost to the carbon cycle for millennia (Coxall et al. 2005; Riebesell et al. 2009). In the short term the calcification process results in reduced sea surface alkalinity (Bates et al., 1996), which in turn reduces the surface ocean ability to drawdown CO_2 (Arrigo et al., 1999) and opposes the direct drawdown of CO_2 by photosynthesis. *E. huxleyi* is, in modern oceans, the most abundant coccolithophore (Hagino *et al.*, 2011; Okada and Mclntyre, 1979), and it is frequently the most numerous species in phytoplankton cell counts from surface water samples (Tyrrell and Merico, 2004). It is also the most widespread species, being found from the tropics to high latitude regions and from mid-ocean to inshore water, with the exception of the polar waters. Under appropriate circumstances, *E. huxleyi* regularly forms blooms which extend over thousands of square kilometres and may persist for many months (Holligan *et al.*, 1993; Tyrrell and Merico, 2004; Hagino *et al.*, 2011). These blooms become visible to satellites such as the Moderate Resolution Imaging Spectroradiometer (MODIS) when they collapse due to the mass shedding of calcium carbonate coccoliths following large scale cell death (Holligan *et al.*, 1993).

E. huxleyi contains high concentrations of dimethylsulphoniopropionate (DMSP), a precursor of DMS, which is a key compound in the global sulphur cycle and its oxidation products influence atmospheric acidity. It has also been proposed that DMS, by providing precipitation nuclei, may influence cloud formation and the

Earth's temperature (Charlson et al., 1987; Malin, 1997). *E. huxleyi* is a model organism with a fully sequenced genome (Read *et al.*, 2013) and a considerable body of physiological information (e.g. Grossi *et al.*, 2000; Conte *et al.*, 1998; Richier *et al.*, 2011; Beaufort *et al.*, 2011; Brownlee and Taylor, 2002; Poulton *et al.*, 2014; Cook *et al.*, 2011; Tyrrell and Schneider, 2008; Müller *et al.*, 2008; Stolte *et al.*, 2000; Fernandez, E. Boyd, P. Holligan, P. Harbour, 1993; Balch and Utgoff, 2009; Findlay *et al.*, 2011; Rost and Riebesell, 2004; Bell and Pond, 1996; Barcelos e Ramos *et al.*, 2010; Paasche, 2002; Rees *et al.*, 2002; Langer *et al.*, 2009; Houdan *et al.*, 2005). Moreover, different strains of *E. huxleyi* show morphological and genetic variability (Iglesias-Rodriguez *et al.*, 2006; Schroeder *et al.*, 2005). For the above reasons *E. huxleyi* has been chosen for my research project with the overall aim of assessing the genetic composition of its populations at different locations, characterised by particular bio-geographical conditions and in relation to bloom and non-bloom conditions in order to better understand the mechanisms underlying its ecosystem dynamics in the present oceans. A more detailed description of *E. huxleyi* and its life cycle is explained in 1.2.

1.1.4 Assessing phytoplankton biodiversity

Many molecular markers have been used to characterise biodiversity within species. DNA markers are useful in phylogenetic analysis, functional gene analysis and applied research, such as paternity testing and food traceability. In order to detect polymorphisms in nuclear DNA, thus assess the genetic diversity of a population, the most frequently used markers are microsatellites (Marsjan and Oldenbroek, 2007), which target non-codifying regions (Sunnucks, 2000). Microsatellites comprise duplet or triplet nucleotide repetitions and are subjected to

extremely high mutation frequencies (Jarne and Lagoda, 1996). Although the breeding structure of populations, population bottlenecks and the biogeographical history of a species are expected to affect all markers in similar ways, the functional gene variability can be different for each gene as a result of history, migration and drift. For this reason the variation within regions which might express certain advantageous traits and adaptable features might not be detected by their correlation with random markers in non-codifying regions (Van Tienderen *et al.*, 2002).

Markers which can directly target specific genes or gene families are used to amplify non-random regions, (Van Tienderen et al., 2002). Different taxonomic levels can be investigated selecting an appropriate genetic marker. At general taxonomic rank, RNA genes (small subunit, SSU, or large subunit, LSU) are well known and SSU in particular has a huge dataset of sequences (NCBI, EMBL or ARB databases), which covers virtually every living taxonomic group known to biology. Furthermore SSU is known to be universally present with the same function in all organisms, except viruses, and to contain variable regions of conservation, which enables design of primers or probes. SSU and LSU are also used in phylogenetic analyses. Furthermore, next-generation sequencing approaches, such as Illumina sequencing, enables millions of sequencing reactions to be run in parallel, at acceptable costs in a short time (Mardis, 2008). These techniques are widely utilised to assess prokaryotic and eukaryotic biodiversity within environmental samples. Whilst the V4 and V9 regions of the small subunit (SSU) rDNA are the best targets for assessing environmental diversity of microbial eukaryotes (Stoeck *et al.*, 2010), 16S rRNA is the most utilised for microbial prokaryotes (Medlin and Kooistra, 2010). Metagenomics, metaproteomics, metatranscriptomics, and proteogenomics are becoming

fundamental approaches to characterize the microbial diversity and their interactions with all the environmental factors (Rastogi and Sani, 2011). Furthermore, non-molecular techniques, such as flow cytometry combined with microscopy can be used to distinguish and quantify species in environmental samples (Medlin and Kooistra, 2010).

1.2 *Emiliania huxleyi*

Emiliania huxleyi belongs to the order Isochrysidales, family Nöelaerhabdaceae of coccolithophores (phylum Haptophyta; Figure 1.3). Paleontological studies have revealed that coccolithophore floras in the geological past were often dominated by a few cosmopolitan taxa (Hine and Weaver, 1998; Young *et al.*, 1999; Rickaby *et al.*, 2002; Moolna and Rickaby, 2012). The maximum species diversity was reached in Late Cretaceous and in some places, e.g. the white cliffs of Dover in England, extraordinary production of chalk are still visible (Westbroek *et al.*, 1989). Nöelaerhabdaceae, have dominated coccolithophore communities numerically for more than 20 million years. Related species, *Gephyrocapsa spp* and *Reticulofenestra spp*, are genetically very close but vary significantly in the morphology of the coccoliths (Beaufort *et al.*, 2011). *E. huxleyi* and *Gephyrocapsa oceanica* are genetically identical for the SSU rDNA and RuBisCO rbcL sequences (Medlin *et al.*, 1996; Fujiwara *et al.*, 2001), but *G. oceanica* has a longer fossil record than *E. huxleyi* (Hine and Weaver, 1998; Hagino *et al.*, 2011). For this reason it is thought that *E. huxleyi* diverged from *G. oceanica* circa 290 kYa (ka; Raffi *et al.*, 2006). Furthermore, both *E. huxleyi* and *G. oceanica* produce particular kinds of

long-chain alkenones and they have in common certain fine structural features (Paasche, 2002; Eglinton *et al.*, 2001; Fujiwara *et al.*, 2001).

1.2.1 E. huxleyi blooms

E. huxleyi blooms are seasonally predictable in certain shelf areas including the North Sea (Holligan *et al.*, 1983) and the Western English Channel (Groom and Holligan, 1987), along continental shelves, e.g. Patagonian Shelf (Painter *et al.*, 2010), and in open ocean, e.g. Iceland Basin (Holligan, 1992; Garcia-Soto and Fernandez, 1995). These blooms are formed by fragmented gene pools with evidence of adaptation of local populations to their environment (Beaufort *et al.*, 2011; Martínez *et al.*, 2012).

Formation of *E. huxleyi* blooms is often linked to warm, stratified conditions where silicic acid and microzooplankton concentrations are low and irradiance levels are high (Holligan *et al.*, 1983; Poulton *et al.*, 2014) with low chlorophyll-a concentrations (Tyrrell and Merico, 2004). In the western English Channel *E. huxleyi* blooms occur at the end of the positive net heat flux (NHF), an indicator of ecosystem dynamics which is involved in the regulation of the yearly community structure (Smyth *et al.*, 2014). Other factors, such as iron availability, may also be important in further regulating bloom formation in cold, nutrient-rich waters (Poulton *et al.*, 2014). The end of a bloom event is associated with the switch to negative NHF, which appears to be an important cut-off for coccolithophore species that require thermal stratification to thrive. Furthermore, it has been proven that viral infection, by *E. huxleyi* large lytic virus *Eh*Vs, plays an important role in terminating the bloom (Wilson *et al.*, 2002).

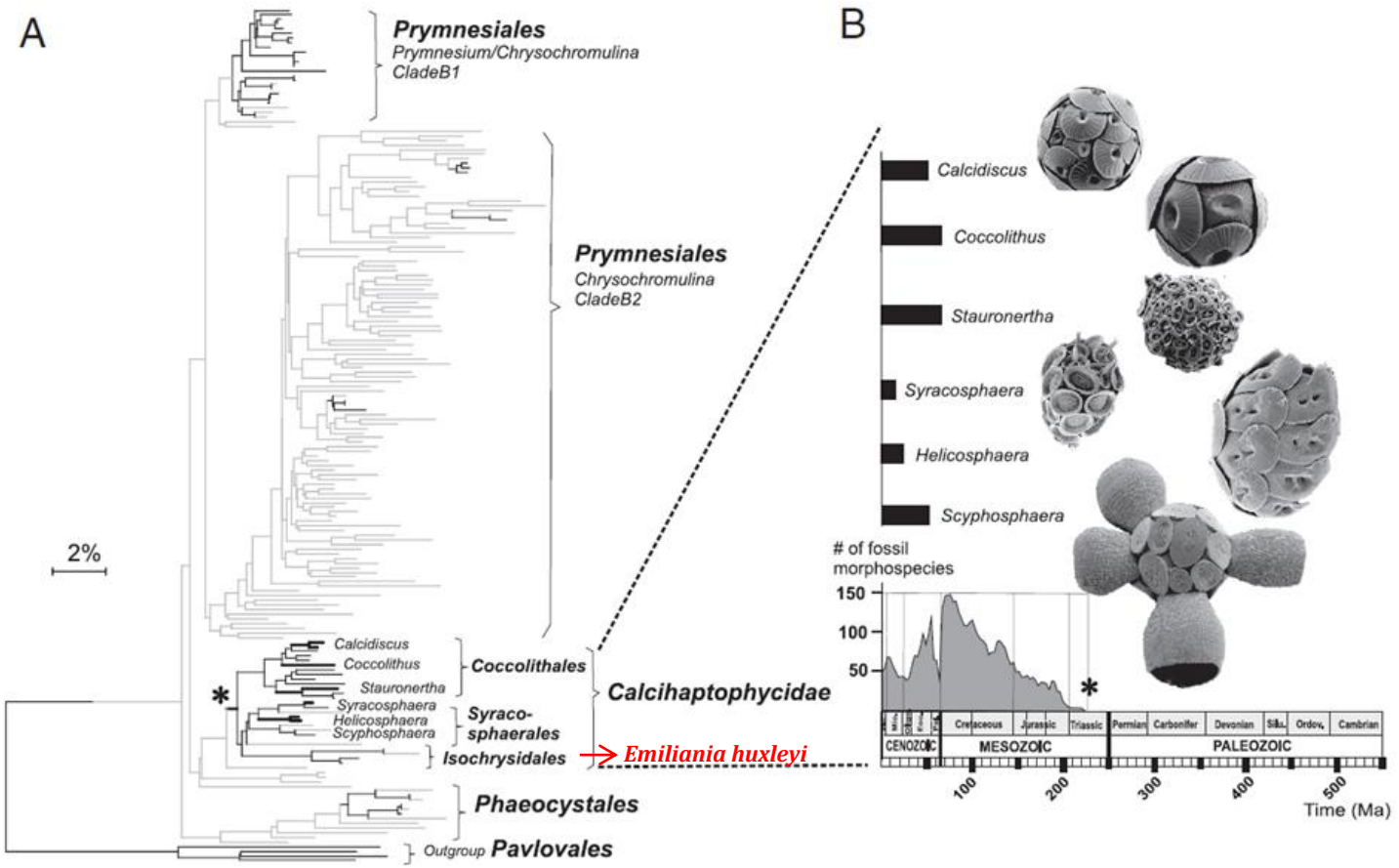


Figure 1.3 From Liu *et al.*, 2009; "Phylogenetic assessment of the haptophyte environmental diversity: A) LSU rDNA tree (5% divergence cutoff), B) Focus on the stratigraphic ranges (black rectangles) of key genera within the calcifying haptophytes. The coccolithophore fossil record (Lower Right) represents number of fossil morphospecies along time in millions of years. Black clover symbols indicate the origin of haptophyte calcification circa 220 Ma."

1.2.2 Life cycle

E. huxleyi and the Haptophytes generally, have a haplodiplontic life cycle (Figure 1.4). In *E. huxleyi*, the different life cycle phases bear either coccoliths (diploid phase) or organic scales (haploid phase) (Billard, 1994). *E. huxleyi* is characterised by three types of cell. These are the non-motile coccolith-bearing cell (C-cell), the non-motile naked cell (N-cell) and the flagellated haploid cell. The C-cell is the most studied form, it has been recorded forming blooms, and its DNA content is twice that of the haploid cell. The haploid has two flagella which allow the cell to swim (Paasche, 2002; Billard and Inouye, 2004). The N-cells apparently arise in culture after extended periods of cultivation (Green *et al.*, 1996; Houdan *et al.*, 2005).

E. huxleyi can alternate a diploid phase and a haploid phase, via meiosis and syngamy, and in this way genetic recombination occurs, maintaining high diversity, however it is still not understood what is the ecological role played by *E. huxleyi* switch between sexual and asexual reproduction, and vice versa. In 2008, a study by Frada and his colleagues (2008) demonstrated that the haploid phase of *E. huxleyi* is "invisible" to *E. huxleyi* virus, *Eh*Vs, like the Cheshire Cat in the famous novel "Alice's Adventures in Wonderland" by Lewis Carroll. While the "Red Queen Hypothesis" (Jaenike, 1978) explains the sexual reproduction as a co-evolved defensive mechanism of a host against its parasites, because it allows the organisms to survive across time becoming more genetically diverse, and subsequently more resistant to viral attack, the "Cheshire Cat Strategy" proposed for *E. huxleyi* shows that the asexual phase is used to escape death (Leung et al., 2012; Frada et al., 2008). The strategy, in fact, is efficient without variating the dominant genotypes over multiannual mesocosm experiments (Martínez et al., 2007; Frada et al., 2008), which means that *E.*

huxleyi and its pathogen co-evolved in sympatry (Leung et al., 2012), and that *E. huxleyi* energetic costs are reduced.

If nutrients and light are abundant *E. huxleyi* will divide about once a day, allowing very rapid population growth (Paasche, 2002). Müller et al (2008) showed that the cells enter the cell division cycle after reaching a critical diameter in the G1 phase, independently from the degree of calcification. The same trend was also observed in naked cells. Muller et al also demonstrated that calcification in *E. huxleyi* is confined to the G1 phase and that the cells suppress this process during the S phase, when the very sensitive process of DNA replication takes place (Müller *et al.*, 2008).

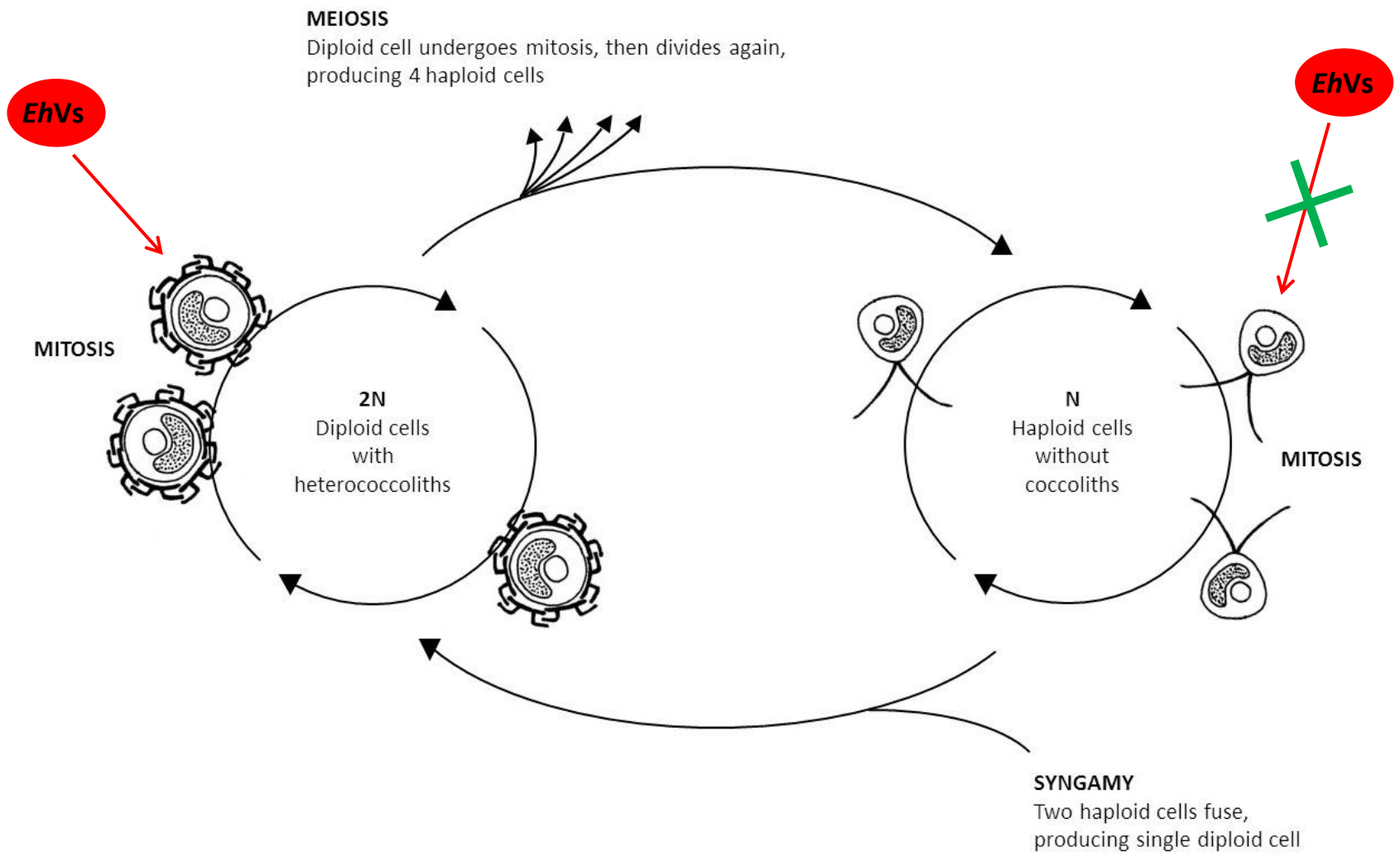


Figure 1.4 *Emiliania huxleyi* haplo-diplontic life cycle (modified from Young & Henriksen, 2003).

1.2.3 Coccolith formation

Many theories have been proposed to explain the production of coccoliths. These include protection from grazing, buoyancy regulation, light modification, providing protons for conversion of HCO_3^- to CO_2 for photosynthesis, and prevention of osmotically induced volume changes (Kleypas *et al.*, 2006). Furthermore, *E. huxleyi* sinking rate is of interest for the ecology of this species, because the movement through the water column may be essential for survival, by increasing the likelihood that it encounters patches or water layers with elevated nutrient concentrations (Paasche, 2002). Since the C-cell form does not have flagella, the motility can be increased by the presence of coccoliths. Effectively the calcite (density, 2.7 gcm^{-3}) is much heavier than organic cell matter (mean density, c. 1.05 gcm^{-3}) or seawater (density, l.027 gcm^{-3}; Young, 1994; Paasche, 2002). Stokes' law should be particularly valid for *E. huxleyi*, in view of its small size and spherical shape (Young, 1994).

Coccolith formation by *E. huxleyi* occurs in an intracellular vesicle, the coccolith vesicle, which seems to arise from a Golgi cisternae. The vesicle is closely apposed to the nuclear membrane, it apparently communicates with a membranous structure termed the 'reticular body' and it probably produces no more than one coccolith at a time (Paasche, 2002). Inside the vesicle there are coccolith constituents including an acidic, water-soluble, calcium binding polysaccharide (Corstjens *et al.*, 1998; Schroeder *et al.*, 2005). A thin organic plate (called base-plate) is used as a substrate for crystal nucleation and the accumulation of calcium and carbonate (Westbroek, P. *et al.*, 1989; Sekino and Shiraiwa, 1996). The coccolith vesicle is not present during nuclear division and is reconstituted after mitosis (Müller *et al.*, 2008; van Emburg, 1989).

1.2.4 E. huxleyi morphotypes

E. huxleyi coccoliths are heterococcoliths, comprising two elliptical shields (distal and proximal) connected by a central area. The distal shield is formed by T-shaped crystal elements (Figure 1.5). The coccoliths are composed of $CaCO_3$ and trace amounts of S, Sr and Mg (Cros *et al.*, 2013), which form complex crystal units departing strongly from simple crystal shapes (Young *et al.*, 1999; Paasche, 2002) and the coccolith shape varies according to different morphotypes. There are, in fact, six well established forms of *E. huxleyi*.

Morphotype A is characterised by medium-sized coccoliths (3-4 μm), robust distal shield elements and curved central area elements (Young et al. 2003; Figure 1.6). The A morphotype is the most common in culture collections (Iglesias-Rodriguez 2006). *E. huxleyi* B form is characterised by larger coccoliths (3.5-5 μm) delicate distal shield elements and irregular morphology in the central area (Young *et al.*, 2003). This morphotype is distinctly less calcified (Young *et al.*, 2014) and it is found rarely. Morphotype C (also called O type) shows small coccoliths (2.5-3.5 μm), delicate distal shield elements and an open central area, sometimes covered by a thin plate. There are two additional forms of *E. huxleyi*: type B/C, and type R. The B/C form shows morphology intermediate between B and C morphotypes, with medium-sized coccoliths (3-4 μm) and delicate distal shield elements (Young *et al.*, 2003) and often dominates assemblages in the southern hemisphere (Cubillos *et al.*, 2007; Holligan *et al.*, 2010; Poulton *et al.*, 2011), and subpolar waters (Hagino *et al.*, 2005). The R phenotype is an overcalcified type A, showing heavily calcified shield elements (Young *et al.*, 2003). Furthermore, a sixth variety of *E. huxleyi* has been described, the so called var. *corona* (Okada and McIntyre, 1977), which is characterised by a discontinuous elevated crown around the central area (Young *et al.*, 2003).

The volume of calcium carbonate in individual coccoliths can vary by more than an order of magnitude, with obvious implications for carbonate flux modelling (Young, 1994).

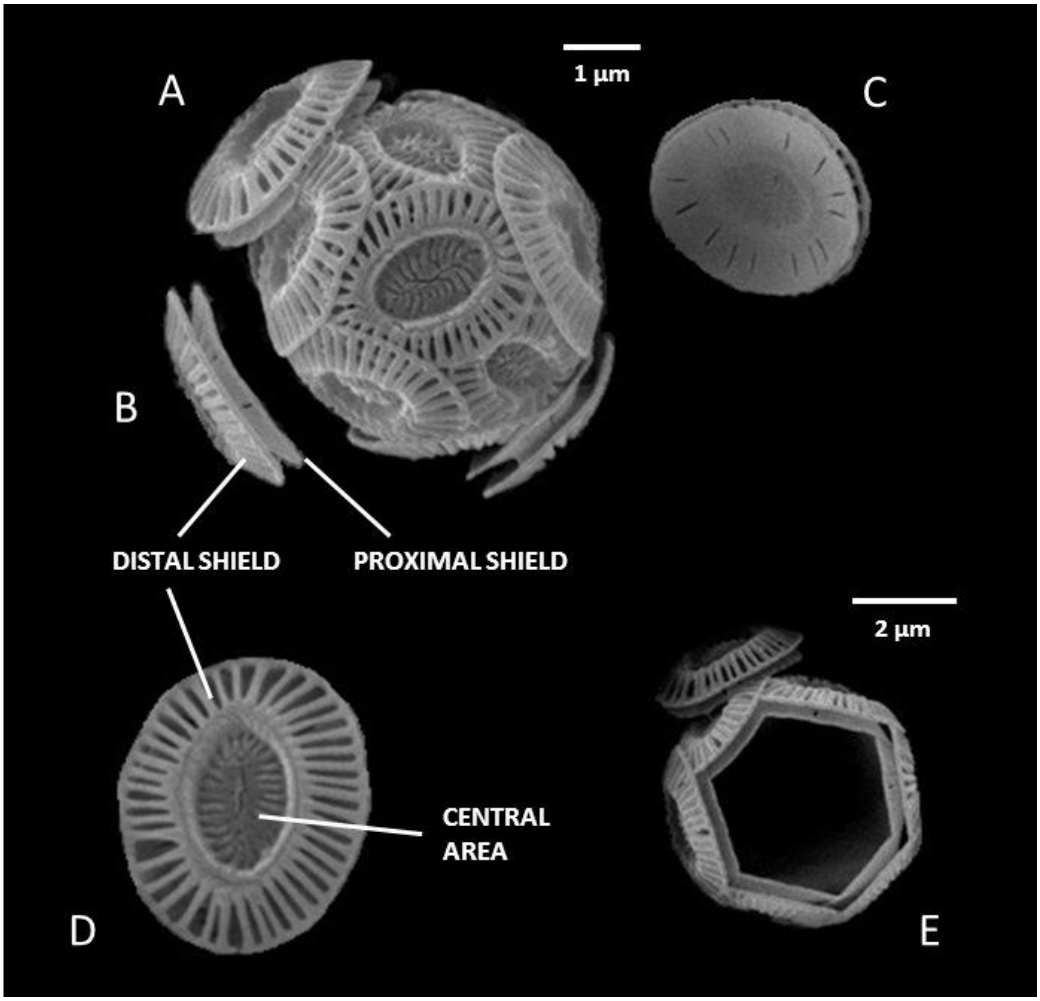


Figure 1.5 *E. huxleyi* coccolith structures. In detail: A) Scanning electron micrograph (SEM) of an intact coccosphere with two detaching-coccoliths. B) One coccolith consists of two overlapping elliptical shields equal in size, the distal shield (upper one) and the proximal shield (underside one), connected by a central area. C) Back side of the proximal shield. D) Coccolith distal shield and central area. Their elements are used to classify different morphotypes of *E. huxleyi*. E) The coccolith structure is well organised around the cell.

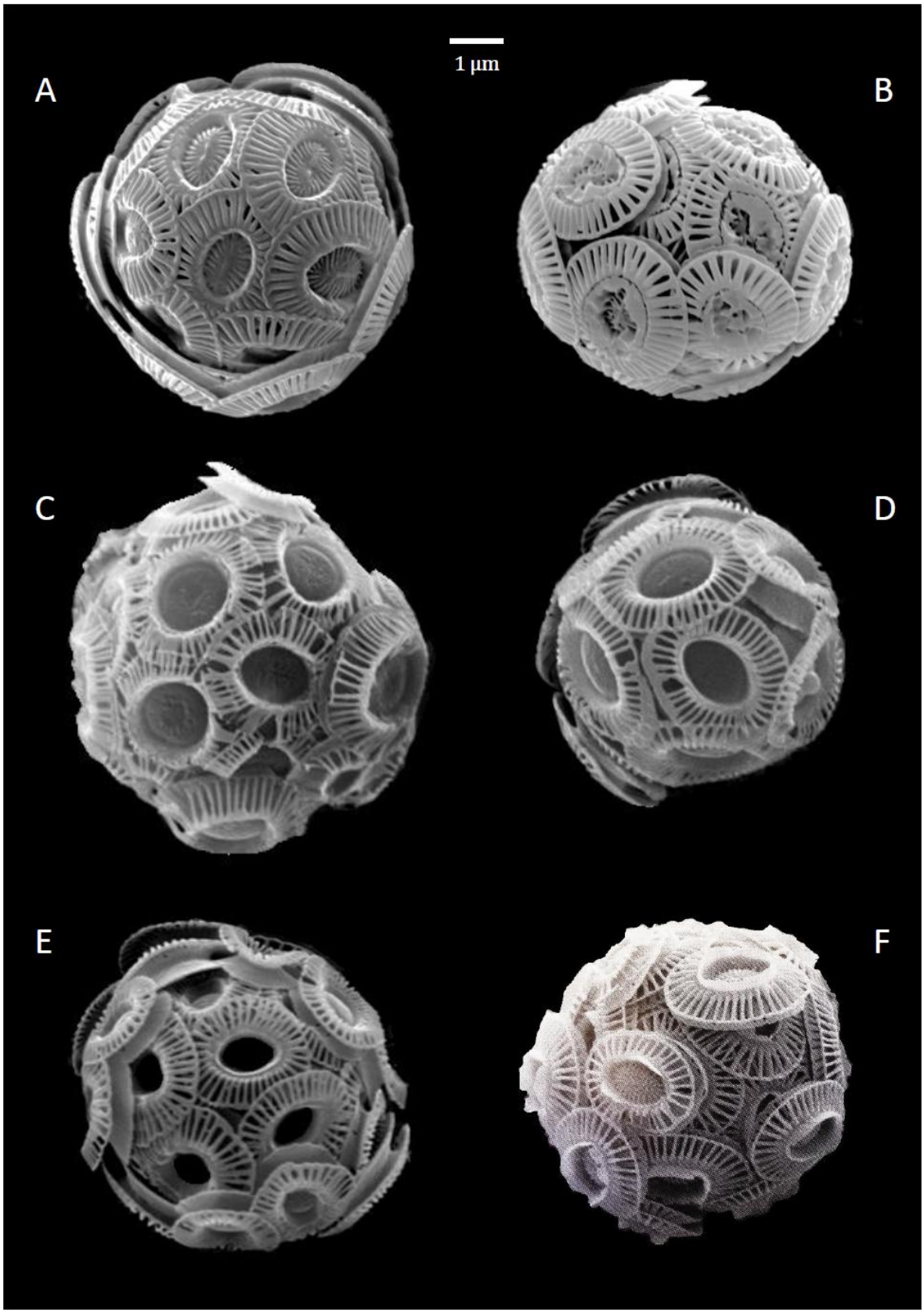


Figure 1.6 SEM images of *E. huxleyi* morphotypes: A) A form, B) R form, C) B form, D) B/C form, E) C form (or O) and F) var. *corona*, Young *et al.*, 2003.

1.2.5 Pigments

E. huxleyi pigment composition has been analysed by high-performance liquid chromatography (HPLC) analysis in a number of studies and shown to comprise chlorophyll-a, the carotenoid 19'hexanoyloxyfucoxanthin (yellow-orange) typical for Primnesiophyceae (Wright and Jeffrey, 1987), 19'-butanoyloxyfucoxanthin together with fucoxanthin, diadinoxanthin (yellow), diatoxanthin (pale orange), β-β carotene, the chlorophylls c2 and c3 and c1 (in minor quantity) and a phytol-substituted chlorophyll c (Haxo, 1985; Wright and Jeffrey, 1987; Fookes and Jeffrey, 1989; Nelson and Wakeham, 1989; Kraay *et al.*, 1992; Garrido *et al.*, 1995; Garrido and Zapata, 1998). Cook *et al*. (2011) analysed the photosynthetic pigment composition and the *tuf*A gene of different strains of *E. huxleyi* types A and B/C from the Southern Ocean. The carotenoid composition was found to be morphology-specific and so was the allelic composition of *tuf*A. Based on their findings they proposed to classify the type B/C as *E. huxleyi* var. *aurorae* (Cook *et al.*, 2011).

1.2.6 Genetic variability

In 2013 the sequencing of the first haptophyte reference genome, from the *E. huxleyi* strain CCMP1516, was completed, revealing extensive genome variability. This study revealed a pan genome likely supported by a complement of repetitive sequences: the *E. huxleyi* genome is comprises of genes distributed variably between strains additionally to common core genes. An important branch of the eukaryotic tree of life is represented by haptophytes, and *E. huxleyi* genome will be extremely useful in trying to better understand evolutionary, cellular and physiological processes (Read *et al.*, 2013).

Before *E. huxleyi* genome sequencing was completed, the variability within *E. huxleyi* sp. had been described in several studies that used both genetic and biochemical approaches. Polyclonal antibodies have been used to differentiate between morphotypes (van Bleijswijk *et al.*, 1991; Young and Westbroek, 1991; Medlin *et al.*, 1996), but they were not effective if the cells were naked (Campbell *et al.*, 1989).

Alkenone/alkenoate biomarkers have been used as a proxy of paleo-sea surface temperature (SST), to analyse *E. huxleyi* batch cultures originated from coastal and oceanic regions and maintained under controlled temperature conditions. The analysis confirmed that biogeographical variations observed in the alkenone vs. temperature relationship in natural waters is representative of differences in genetic makeup and physiological status of the local alkenone-synthesizing populations (Conte *et al.*, 1998). Furthermore, different pigment composition has been found to be strain specific (Stolte *et al.*, 2000).

In 2005 Schroeder and colleagues discovered a genetic marker that correlates with the separation of *E. huxleyi* morphotypes (Figure 1.7) and revealed that the two main morphotypes are composed of a number of distinct genotypes. This marker lies in the untranslated region of the GPA gene, so called for its high content in glutamic acid, proline and alanine and which encodes for a protein implicated in regulating coccolith structure. The marker, circa 100 bp, has been named coccolith morphology motif (CMM) and it has been used to analyse strains obtained from various locations. Four different motifs were well characterised and they described four groups in which the isolates were divided: CMM I, II, III and IV. The motif could be involved in binding the sequence that modules the initiation machinery which controls the overall morphology of the coccoliths or it could be involved in repressing and

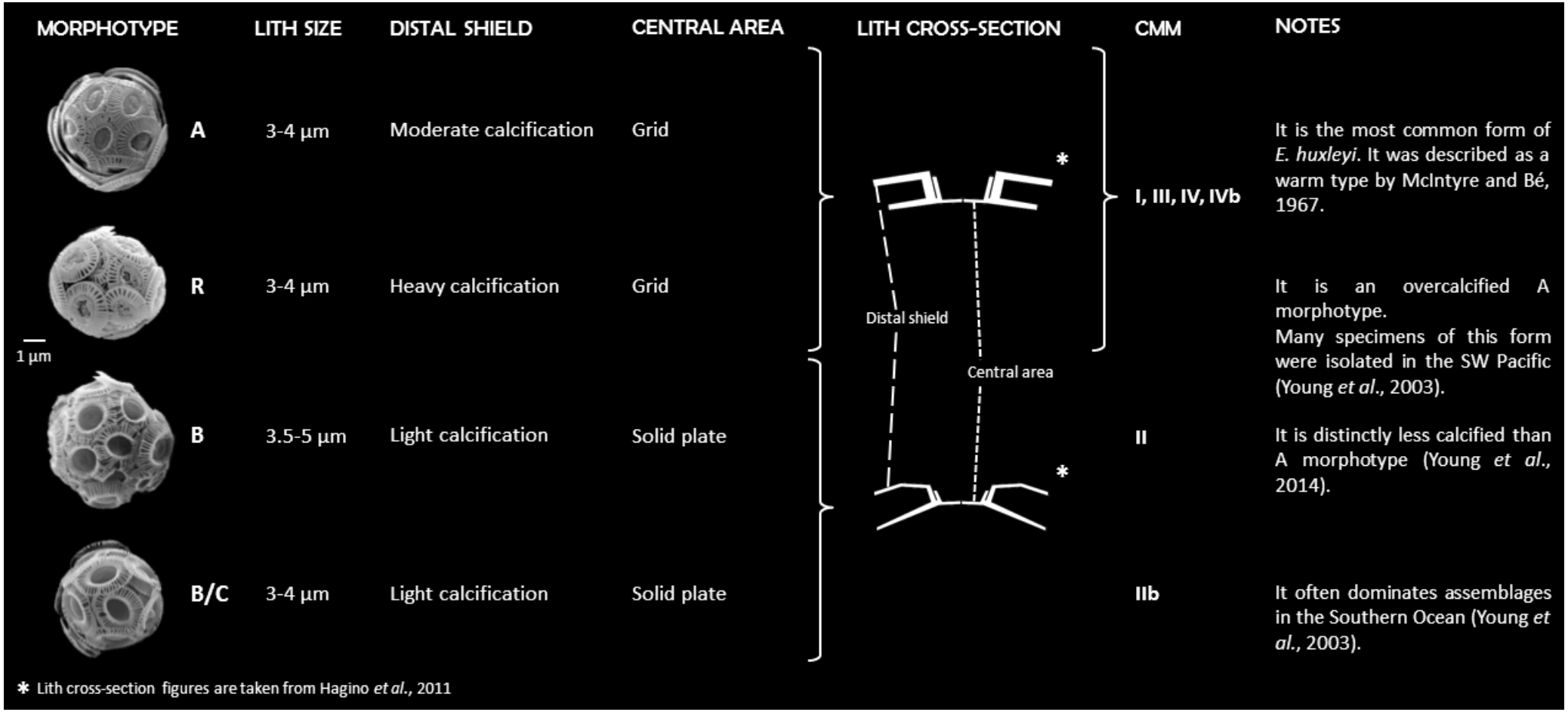


Figure 1.7 Summary of *E. huxleyi* different morphotypes and the related Coccolith Morphology Motives (CMMs).

activating proteins which control the development of the coccoliths within the coccolith vesicle (Schroeder *et al.*, 2005).

Moreover, high intraspecific genetic variability within *E. huxleyi* sp. was assessed analysing the microsatellite length variants (alleles) on 85 *E. huxleyi* clonal isolates representative of different ocean basins (Iglesias-Rodriguez *et al.*, 2006). Microsatellites are multiple repeats of simple oligonucleotides which have high mutation rates, and they are present in both coding and non-coding regions of all prokaryote or eukaryote genomes (Schlötterer and Tautz, 1992; Iglesias-Rodriguez *et al.*, 2006).

1.3 Thesis outline

Populations are dynamic and many factors, such as nutrient limitation, physical and biological interactions, migration processes and environmental factors can influence their structures. Montes-Hugo *et al.* (2009) showed a decadal variation of phytoplankton biomass and environmental factors along the Western Antarctic Peninsula associated with a rapid regional climate change. Also studies carried out by Feng *et al.* (2009), using a natural North Atlantic bloom phytoplankton community in a shipboard continuous culture incubation, revealed "highest coccolithophore abundance in samples treated at high CO_2 levels and highest chrysophyte abundance at high temperature". Further studies by Edwards and Richardson (2004) found that the level of response to climate change differs throughout the community and the seasonal cycle.

The stability debate on how the loss of biodiversity can influence ecosystem dynamics is still open, as our knowledge of the marine phytoplankton biodiversity is

still limited. Key questions relate to the extent to which phytoplankton may be able to display particular survival mechanisms (acclimation) or whether particular species or genotypes will be selected to dominate the oceans of the future (adaptation). Establishing the present condition of the marine environment and how the organisms are responding to the fast changing climate is fundamental to understanding how biodiversity will be affected.

In spite of Earth's climate changes, the combination of genetic diversity and high functional redundancy has helped to maintain the continuity of oxygenic photosynthesis. The composition of phytoplankton communities is correlated with oceanic circulation and mesoscale physical processes, which influence the fluxes of essential nutrients (Falkowski *et al.*, 1998).

E. huxleyi is one of the major pelagic producers of $CaCO_3$ in the modern ocean, and its response to water chemistry changes at the surface is of particular relevance for ocean biogeochemical cycles and climate feedback systems (O'Dea *et al.*, 2014).

Knowing that *E. huxleyi* shows genetic variability which is related to different morphotypes (Schroeder *et al.*, 2005) and that complete coccoliths of a given size can contain widely varying amounts of calcite, as a result of variation in the degree of primary calcification (Young and Westbroek, 1991; Young, 1994a), it becomes of great relevance, and it is the focus of this thesis, to assess the genetic composition, the population structure and dynamics of this key species. What is the extant genetic assemblage of *E. huxleyi* natural populations? Are some of its genotypes more suitable to certain ocean carbonate parameters? Can studies conducted on *E. huxleyi* strains maintained for years in laboratory collections be representative of the wild community?

Genetic diversity within species is at the heart of biodiversity (Frankham *et al.*, 2002), and in the same way coexisting species evolve to use different resources, reducing the effects of interspecific competition, competition among members of a single species drives ecological diversification within natural populations (Svanback and Bolnick, 2007).

In order to make an initial assessment of the diversity within the standing stock of *E. huxleyi* held within established culture collections, I developed a quick method to characterise different *E. huxleyi* genotypes. This technique is based on qPCR and it uses dual labelled probes, which can be combined together to run multiplexing amplification. The probes are designed on the *E. huxleyi* coccolith morphology motif (CMM). This method, called probe assay, was tested on *E. huxleyi* clonal isolates obtained from a cruise around the UK and in the North Sea (see Appendix I for the details on the expedition) and on samples collected from various oceans and held in culture collections for many years. Additionally, in order to test the robustness of the method, the same samples were analysed by microsatellite profiling (by Dr Stacy Krueger-Hadfield) and many of them were additionally sequenced. The results of this analysis and the description of the probe assay are illustrated in Chapter 2.

The majority of the studies found in literature regarding *E. huxleyi* are focused on bloom events, aimed to comprehend the underlying causes of these episodes; however there is only limited information about *E. huxleyi* genetic composition outside the bloom context. I analysed a six year time series of *E. huxleyi* samples collected at the L4 station of the Western Channel Observatory to understand on a temporal scale how the genetic composition of *E. huxleyi* varies when bloom events are not occurring and to assess *E. huxleyi* genetic standing stock. My working

hypothesis was that the *E. huxleyi* population is resilient from year to year, although the relative abundance of specific genotypes may change throughout the year and only specific genotypes are selected to form blooms (Chapter 3).

Furthermore, I analysed the genetic composition of recently collected *E. huxleyi* samples from the Northern and Southern Hemispheres (Appendix I). Information obtained from this study will allow us to address the more general hypothesis that despite rapid changing climate and the subsequent ocean acidification problem, certain taxa, like the abundant and wide spread *E. huxleyi*, may survive though different genotypes within these taxa will be selected (Chapter 4).

CHAPTER 2

Genotyping an *Emiliania huxleyi* (Prymnesiophyceae) bloom event in the North Sea reveals evidence of asexual reproduction

Stacy A. Krueger-Hadfield[1,2§], Cecilia Balestreri[1,3*], Joanna Schroeder[1*], Andrea Highfield[1], Pierre Helaouët[4], Jack Allum[5], Roy Moate[6], Kai T. Lohbeck[7,8], Peter I. Miller[9], Ulf Riebesell[8], Thorsten B. H. Reusch[7], Ros E. M. Rickaby[3], Jeremy Young[10], Gustaaf Hallegraeff[11], Colin Brownlee[1] and Declan C. Schroeder[1§]*

[1] Marine Biological Association, Citadel Hill Laboratory, Plymouth, PL1 2PB, UK

[2] College of Charleston, Grice Marine Laboratory, 205 Fort Johnson Road, Charleston, SC 29412, USA[2]

[3] Department of Earth Sciences, Oxford University, South Parks Road, Oxford OX1 3AN, UK

[4] Sir Alister Hardy Foundation for Ocean Science, Citadel Hill Laboratory, Plymouth, PL1 2PB, UK

[5] School of Biological Sciences, Faculty of Science, University of Plymouth, Plymouth PL4 8AA, UK

[6] Plymouth Electron Microscope Centre, Faculty of Science, University of Plymouth, Plymouth PL4 8AA, UK

[7] Evolutionary Ecology of Marine Fishes, GEOMAR Helmholtz-Centre for Ocean Research Kiel, Düsternbrooker Weg 20, 24105 Kiel, Germany,

[8] Biological Oceanography, GEOMAR Helmholtz-Centre for Ocean Research Kiel, Düsternbrooker Weg 20, 24105 Kiel, Germany

[9] Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, PL1 3DH

[10] Department of Earth Sciences, University College London, WC1E 6BT

[11] School of Plant Science, University of Tasmania, Private Bag 55, Hobart, Tasmania 7001, Australia

* Shared first author

[§] Author for correspondence

Authors' contribution

The manuscript was written by Stacy Krueger-Hadfield and Cecilia Balestreri, together with Declan Schroeder and Joanna Schroeder. The majority of the samples analysed were collected by Cecilia Balestreri during the D366 cruise (Appendix I). Cecilia Balestreri was also responsible for the clones' isolation and for the DNA extraction of these samples. The Southern Ocean strains of *Emiliania huxleyi* were provided by Gustaaf Hallegraeff. The development of the probe assay technique was started by Andrea Highfield and carried on by Cecilia Balestreri, who also optimised it and used it to analyse all the samples. Stacy Krueger-Hadfield ran the microsatellite analysis and Joanna Schroeder was responsible for the statistical analysis. Pierre Helaouët carried out the global SST analysis. Jack Allum provided most of the scanning electron microscope (SEM) photographs of the D366 samples and Roy Moate supervised the work at the SEM facility. Project design, funds awarded to, writing and revision of the manuscript, by Declan Schroeder. Kai Lohbeck, Ulf Riebesell & Thorsten Reusch provided the genetic drift data; Peter Miller provided the bloom satellite analysis; Ros Rickaby, Jeremy Young and Colin Brownlee contributed to the mentoring of Cecilia and the revision and editing of the final draft before submission.

2.1 Abstract

Due to the unprecedented rate at which our climate is changing, the ultimate consequence for many species is likely to be either extinction or migration to an alternate habitat. Certain species might, however, evolve at a rate that could make them resilient to the effects of a rapidly changing environment. This scenario is most likely to apply to species that have large population sizes and rapid generation times, such that the genetic variation required for adaptive evolution can be readily supplied. *Emiliania huxleyi* (Lohm.) Hay and Mohler (Prymnesiophyceae) is likely to be such a species as it is the most conspicuous extant calcareous phytoplankton species in our oceans with generation times of 1 day^{-1}. Here we report on a validated set of microsatellites, in conjunction with the coccolithophore morphology motif genetic marker, to genotype 93 clonal isolates collected from across the world. Of these, 52 came from a single bloom event in the North Sea collected on the D366 United Kingdom Ocean Acidification cruise in June-July 2011. There were 26 multilocus genotypes (MLGs) encountered only once in the North Sea bloom and 8 MLGs encountered twice or up to six times. Each of these repeated MLGs exhibited P_{sex} values of less than 0.05 indicating each repeated MLG was the product of asexual reproduction and not separate meiotic events. In addition, we show that the two most polymorphic microsatellite loci, EHMS37 and P01E05, are reporting on regions likely undergoing rapid genetic drift during asexual reproduction. Despite the small sample size, there were many more repeated genotypes than previously reported for other bloom-forming phytoplankton species, including a previously genotyped *E. huxleyi* bloom event. This study challenges the current assumption that sexual

reproduction predominates during bloom events. Whilst genetic diversity is high amongst extant populations of *E. huxleyi*, the root cause for this diversity and ultimate fate of these populations still requires further examination. Nonetheless, we show that certain CMM genotypes are found everywhere; while others appear to have a regional bias.

2.2 Introduction

The coccolithophore *Emiliania huxleyi* (Lohm.) Hay and Mohler (Prymnesiophyceae), is thought to be the main calcite producer on Earth (Westbroek *et al.*, 1993), it is widespread in all but extreme polar oceans and it regularly forms extensive "white water" blooms in high latitude coastal and shelf ecosystems. While the process of calcification results in decreased alkalinity of surface waters, potentially reducing the drawdown of CO_2 from the atmosphere, coccolithophores, such as *E. huxleyi*, are thought to contribute to reductions in atmospheric CO_2 by creating a net export of carbon to the seabed (Robertson *et al.*, 1994; Riebesell and Tortell, 2011).

Current estimates are that as much as 27% of the anthropogenic CO_2 produced from burning of fossil fuels released between 1959 -2011 has been absorbed by the oceans (Le Quéré *et al.*, 2013). As CO_2 reacts with seawater, it generates dramatic changes in carbonate chemistry, including decreases carbonate ions and pH (ocean acidification) and an increase in bicarbonate ions. The consequences of this overall process are commonly referred to as ocean acidification. Moreover, ongoing atmospheric warming is expected to cause significant changes to the ocean climate by the end of this century (the average temperature of the upper layers of the ocean having increased by 0.6°C over the past 100 years, IPCC, 2007). The oceans are, therefore, experiencing unprecedented levels of change, raising concerns about the impacts on key biological species such as *E. huxleyi.* The nature of such impacts will have important biological, ecological, biogeochemical and societal implications (Turley *et al.*, 2010). Langer *et al.* (2009) found that different clonal *E.*

huxleyi isolates vary in their phenotypic traits such as growth and calcification rate, suggesting a potential role for selection on standing genetic variation in shaping future populations. This mechanism was demonstrated by Lohbeck *et al.* (2012) who identified pH-driven selection on 6 clonal isolates from an *E. huxleyi* bloom near Bergen, Norway. Functional diversity within this set of clones allowed selective sorting over only 500 generations of exponential growth. These findings raise questions about the pace and relevance of such clonal sorting under natural conditions. Unfortunately, very little is known about the population biology of this key phytoplankton species and hence, forecasting how future populations will respond is difficult.

Future *E. huxleyi* populations could have a very different set of phenotypes when compared with present-day populations. This shift in phenotypic traits would have profound implications on ecosystem function and biogeochemical cycles. However, before we can address the effects of a rapidly changing climate on *E. huxleyi*, we must understand the very basic properties of its genetic diversity and ecological interactions. Martínez *et al.* (2007, 2012) described a genetically rich, but stable *E. huxleyi* population using the coccolithophore morphology motif (CMM) in the North Atlantic. The CMM lies within the 3' untranslated mRNA region of the coccolith polysaccaride associated protein GPA, which is implicated in controlling coccolith structure (Schroeder *et al.*, 2005). In addition, Iglesias-Rodriguez *et al.* (2006) and Hinz (2010) found high levels of intraspecific microsatellite genetic diversity in different *E. huxleyi* bloom events. In contrast to the CMM, microsatellites appear to be highly polymorphic markers that can resolve neutral genetic diversity within populations. The authors concluded that this is most likely driven by high

rates of sexual reproduction. However, for species with large population sizes and rapid generation times, sex is not the sole driver for high genetic diversity. Indeed, in species exhibiting large dispersal potential and geographic ranges, very high levels of genetic diversity are expected (i.e., molecular hyperdiversity, Cutter *et al.*, 2013). In the natural environment *Saccharomyces* yeasts only reproduce sexually one in every 1000 to 3000 effective generations (Tsai *et al.*, 2008). The mycorrhizal fungi (phylum Glomeromycota) are among the oldest and most successful symbionts of land plants and show no evidence of sexual reproduction (VanKuren *et al.*, 2013). Indeed, a combination of intra-individual polymorphism and effective population sizes in the Glomeromycota contribute to its evolutionary longevity.

The 10 polymorphic microsatellite markers used in Iglesias-Rodríguez *et al.* (2006) and Hinz (2010) were developed without the benefit of genome sequence information for this species (Read *et al.*, 2013). In this study, we revisited 10 polymorphic microsatellite markers developed by Iglesias-Rodriguez *et al.* (2002, 2006), thoroughly tested and critically evaluated them in order to begin characterizing genetic diversity in an *Emiliania huxleyi* bloom event sampled during the D366 Sea Surface Consortium UK Ocean Acidification cruise (www.surfaceoa.org.uk; see Appendix I). The estimated genetic diversity, as defined by both the CMM and microsatellite markers, was used to critically revise the predominant mode of reproduction during an *E. huxleyi* bloom. Moreover, clonal diversity in the North Sea bloom event is compared to a biogeographic phytoplankton data set and the adaptive potential of future *E. huxleyi* populations facing a changing ocean is discussed.

2.3 Materials and methods

2.3.1 Validation of microsatellite primers

(I) Ten polymorphic microsatellite sequence primer pairs (AJ487304-17; AJ494737-42, Table 2.1) were blasted (blastn) against the CCMP1516 genome (Read *et al.*, 2013) in order to verify the amplification of a single site within the genome.

(II) PCR conditions used are as those described in Iglesias-Rodríguez *et al*. (2002, 2006), using the following modified PCR mix: 20 µL final volume, 2 µL of at least 10 ng DNA template, 1x reaction buffer, 1.5 mmol L^{-1} $MgCl_2$, 0.25 mmol L^{-1} deoxyribonucleotide triphosphate, 250 mmol L^{-1} each of unlabeled forward and reverse primers and 1 U of taq polymerase (GoTaq Flexi, Promega). In addition, the loci which produced repeatable PCR results and for which single-locus genetic determinism was verified were tested with an annealing temperature of 54 ^{o}C in order to facilitate the multiplexing of loci in the future. Initial PCR amplification trials were visualized using 1.8% agarose gels with a 50 bp ladder (New England Biolabs, MA, USA). Each reliable locus produced the same results as when tested with the original annealing temperature. Therefore, all subsequent reactions were run at 54 ^{o}C, though for the purposes of this study, all reactions were done in simplex.

(III) In order to investigate the stability of alleles at each locus, strain no. 62 used in Lohbeck *et al.* (2012b, 2013) was genotyped at the start of the experiment and after 1300 generations of exponential growth under a set of different CO_2 conditions (i.e. mapping any changes between June 2010 to November 2012). A second strain, CCMP1516 (Read *et al.*, 2013), was also used spanning multiple

generations, varying culture conditions under alternating exponential and stationary growth conditions that resulted in loss of coccolith production.

Table 2.1 Characteristics of the 10 microsatellite markers isolated in *Emiliania huxleyi* by Iglesias-Rodríguez *et al.* (2002, 2006). N_{Bio}, total number of distinct alleles observed over the biogeographic data set and N_{NS}, total number of distinct alleles observed over the North Sea Bloom data set.

Locus	Acc. No.	Fluorescent Dye	Profile*	BLAST	Amplification proportion	A-range (bp)	NBio	NNS
EHMS37	AJ494737 AJ494738	PET	one	1	0.93	194-340	37	12
P01E05	AJ494739 AJ494740	6-FAM	one	1	0.96	106-190	28	10
P02F11	AJ487316 AJ487317	NED	one	0	0.98	98-192	21	8
P02E09	AJ494741 AJ494742	PET	one	1	0.99	82-172	10	7
P02B12	AJ487310 AJ487310	NED	one	0	1	204-224	11	4
P02E11	AJ487312 AJ487313	VIC	multiple	1	-	-	-	-
P02E10	AJ487314 AJ487315	6-FAM	multiple	5	-	-	-	-
EHMS15	AJ487304 AJ487305	VIC	multiple	2	-	-	-	-
P01F08	AJ487306 AJ487307	-	none.	0	-	-	-	-
P02A08	AJ487308 AJ487309	-	none	0	-	-	-	-

*: number of loci amplified

2.3.2 Microsatellite amplification

For optimization purposes, all successful PCR products were transferred to an ABI 3130 xL genetic analyzer (Applied Biosystems, Foster City, CA, USA) equipped with a 36 cm capillary array. The PCR mix was updated to include a fluorescently labeled forward primer: 150 mmol L^{-1} of the labeled forward primer, 100 mmol L^{-1} of the unlabeled forward primer and 250 mmol L^{-1} of the unlabeled reverse primer, where all other mix components remained unchanged. Two µL of each PCR product was added to 10 µL of loading buffer containing 0.3 µL of size standard (GeneScan – 500 Liz, Applied Biosystems, Foster City, CA, USA) plus 9.7 µL of Hi-Di formamide (Applied Biosystems, Foster City, CA, USA). The loading mix was denatured at 92°C for 3 minutes. A positive and negative control was electrophoresed with each set of samples run on the sequencer.

After optimization, a subset of known genotypes was transferred to SourceBioScience Nottingham for fragment analysis on a 3730xL DNA analyser run on a 50 cm capillary array. For all clonal isolates, 7 µL of each PCR product was sent to SourceBioScience, including positive and negative controls for each sequencer run. All genotypes were scored manually using GENEMAPPER ver. 4 (Applied Biosystems, Foster City, CA, USA).

2.3.3 UK Ocean Acidification Research Cruise

The R/V Discovery, cruise number 366, circumnavigated the British Isles in June/July 2011 as part of the UK Ocean Acidification research programme (www.surfaceoa.org.uk; see Appendix I). Samples used in this study were collected

mainly in the North Sea (5 stations, Figure 2.1) and also in the Western coast of Scotland, Bay of Biscay and Western English Channel (Table 2.2).

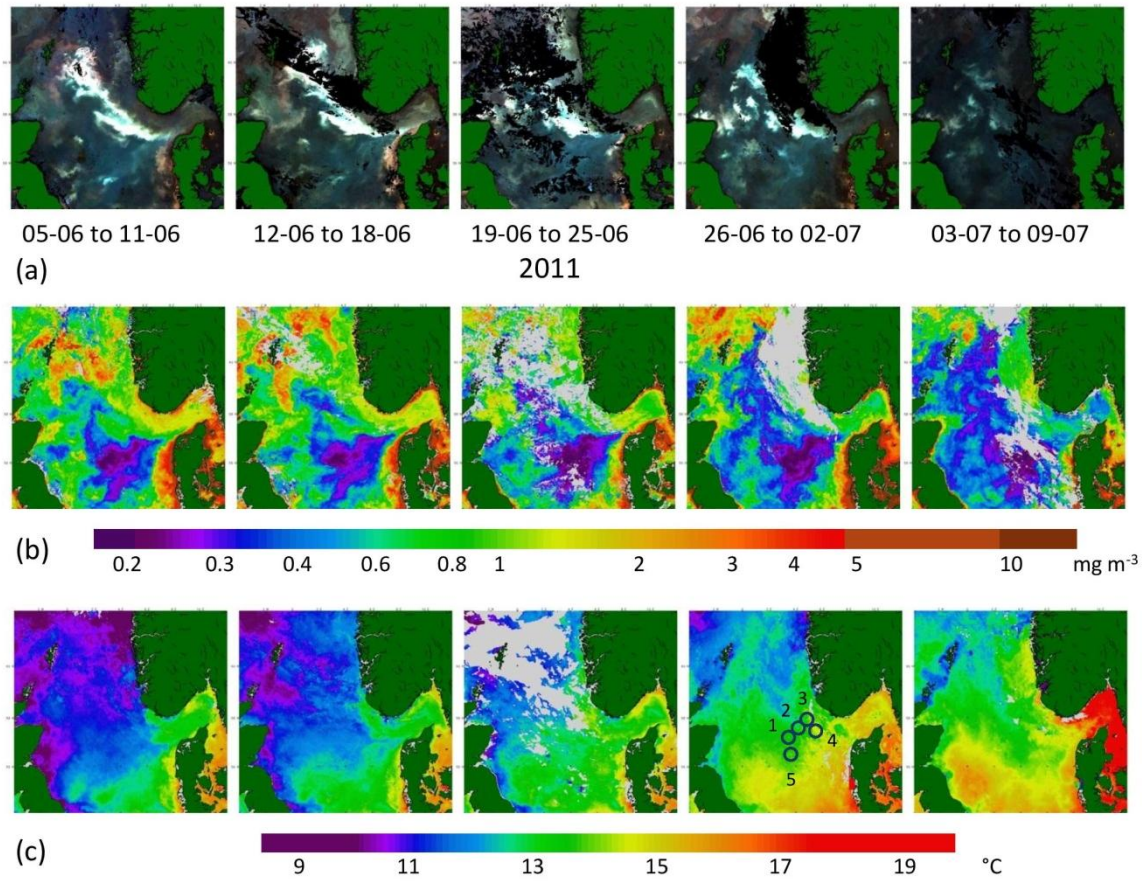


Figure 2.1 Earth observation 7-day composite data showing *Emiliania huxleyi* bloom development before, during and after cruise: (a) Enhanced ocean colour from Aqua-MODIS, showing coccoliths as bright patches and persistent cloud in black. (b) Chlorophyll-a concentration from Aqua-MODIS, with cloud in light grey. (c) Sea-surface temperature from AVHRR, where numbered circles indicate cruise stations listed in Table 2.2.

Table 2.2 *Emiliania huxleyi* isolates used in this study

Strain number	Strain ID	MLG	P_{sex} value	Station	Year	Location	Latitude	Longitude	Morphotype	CMM sequencing	CMM probe assay	EHMS37		P01E05		P02F11		P02E09		P02B12		SST	SST cluster	Culture collection
1	D366 106-2	MLG 1	3.01×10^{-16}	1	2011	North Sea	57.2	3.48	A	I (■)	I I	210	210	142	142	98	102	102	102	208	208	10.544	medium	Plymouth
2	D366 17-1	–	–	5	2011	North Sea	56.5	3.65	A	I (■)	I I	210	210	142	142	98	102	102	102	208	208	10.544	medium	Plymouth
3	D366 26-1	–	–	5	2011	North Sea	56.5	3.65	A	–	I I	210	210	142	142	98	102	102	102	208	208	10.544	medium	Plymouth
4	D366 124-1	MLG 2	7.54×10^{-05}	4	2011	North Sea	57.45	5.53	A	I (■)	I I	210	210	142	142	102	102	102	102	208	208	10.639	medium	Plymouth
5	D366 40-1	–	–	5	2011	North Sea	56.5	3.65	A	I (■)	I I	210	210	142	142	102	102	102	102	208	208	10.544	medium	Plymouth
6	D366 32-5	–	–	5	2011	North Sea	56.5	3.65	A	I (□)	I I	210	210	142	142	102	102	102	102	208	208	10.544	medium	Plymouth
7	D366 124-2	–	–	4	2011	North Sea	57.45	5.53	A	–	I I	210	210	142	142	102	102	102	102	208	208	10.639	medium	Plymouth
8	D366 124-5	–	–	4	2011	North Sea	57.45	5.53	A	–	I I	210	210	142	142	102	102	102	102	208	208	10.639	medium	Plymouth
9	D366 35-1	–	–	5	2011	North Sea	56.5	3.65	A	–	I I	210	210	142	142	102	102	102	102	208	208	10.544	medium	Plymouth
10	D366 25-3	MLG3	–	5	2011	North Sea	56.5	3.65	A	I (■)	I I	210	210	142	142	102	102	102	102	212	212	10.544	medium	Plymouth
11	D366 112-1	MLG4	2.05×10^{-07}	2	2011	North Sea	57.56	4.2	A	I (■)	I I	214	214	142	142	102	102	98	102	208	208	10.571	medium	Plymouth
12	D366 30-1	–	–	5	2011	North Sea	56.5	3.65	A	I (■)	I I	214	214	142	142	102	102	98	102	208	208	10.544	medium	Plymouth
13	D366 40-5	–	–	5	2011	North Sea	56.5	3.65	A	–	I I	214	214	142	142	102	102	98	102	208	208	10.544	medium	Plymouth
14	D366 33-3	–	–	5	2011	North Sea	56.5	3.65	A	–	I I	214	214	142	142	102	102	98	102	208	208	10.544	medium	Plymouth
15	D366 112-3	–	–	2	2011	North Sea	57.56	4.2	A	–	I I	214	214	142	142	102	102	98	102	208	208	10.571	medium	Plymouth
16	D366 112-2	MLG5	–	2	2011	North Sea	57.56	4.2	A	–	I I	214	214	132	142	102	106	98	102	208	208	10.571	medium	Plymouth
17	D366 21-5	MLG6	–	5	2011	North Sea	56.5	3.65	A	–	I I	206	218	132	142	102	106	102	102	208	208	10.544	medium	Plymouth
18	D366 22-4	MLG7	–	5	2011	North Sea	56.5	3.65	A	–	I I	196	206	132	142	102	106	102	102	208	208	10.544	medium	Plymouth
19	D366 124-3	MLG8	2.74×10^{-05}	4	2011	North Sea	57.45	5.53	A	I (□)	I I	202	210	142	142	102	102	102	102	208	208	10.639	medium	Plymouth
20	D366 124-4	–	–	4	2011	North Sea	57.45	5.53	A	–	I I	202	210	142	142	102	102	102	102	208	208	10.639	medium	Plymouth
21	D366 26-3	MLG9	0.035	5	2011	North Sea	56.5	3.65	A	I (□)	I I	206	206	142	142	102	102	102	102	208	208	10.544	medium	Plymouth
22	D366 26-4	–	–	5	2011	North Sea	56.5	3.65	A	–	I I	206	206	142	142	102	102	102	102	208	208	10.544	medium	Plymouth
23	D366 26-5	MLG10	–	5	2011	North Sea	56.5	3.65	A	I (□)	I I	206	206	142	142	102	130	102	102	208	208	10.544	medium	Plymouth
24	D366 33-1	MLG11	5.16×10^{-06}	5	2011	North Sea	56.5	3.65	A	I (■)	I I	206	206	152	152	118	134	102	102	208	208	10.544	medium	Plymouth
25	D366 33-2	–	–	5	2011	North Sea	56.5	3.65	A	–	I I	206	206	152	152	118	134	102	102	208	208	10.544	medium	Plymouth
26	D366 17-3	MLG12	–	5	2011	North Sea	56.5	3.65	A	–	I I	206	206	152	152	102	102	102	102	208	212	10.544	medium	Plymouth
27	D366 35-2	MLG13	4.15×10^{-08}	5	2011	North Sea	56.5	3.65	A	I (■)	I I	206	206	132	132	102	102	102	102	208	212	10.544	medium	Plymouth
28	D366 35-3	–	–	5	2011	North Sea	56.5	3.65	A	–	I I	206	206	132	132	102	102	102	102	208	212	10.544	medium	Plymouth
29	D366 36-4	–	–	5	2011	North Sea	56.5	3.65	A	–	I I	206	206	132	132	102	102	102	102	208	212	10.544	medium	Plymouth
30	D366 20-4	MLG14	–	5	2011	North Sea	56.5	3.65	A	I (■)	I I	206	206	132	132	102	102	102	102	208	208	10.544	medium	Plymouth
31	D366 21-3	MLG15	–	5	2011	North Sea	56.5	3.65	A	I (■)	I I	206	214	148	148	102	102	102	102	208	208	10.544	medium	Plymouth
32	D366 89-5	MLG16	–	5	2011	North Sea	56.5	3.65	A	–	I I	210	254	132	132	102	102	102	102	208	208	10.544	medium	Plymouth
33	D366 26-2	MLG17	1.61×10^{-05}	5	2011	North Sea	56.5	3.65	A	I (■)	I I	210	218	126	142	102	102	102	102	208	208	10.544	medium	Plymouth
34	D366 30-2	–	–	5	2011	North Sea	56.5	3.65	A	–	I I	210	218	126	142	102	102	102	102	208	208	10.544	medium	Plymouth
35	D366 24-1	MLG18	–	5	2011	North Sea	56.5	3.65	A	–	I I	202	210	130	130	102	102	106	106	208	208	10.544	medium	Plymouth
36	D366 37-4	MLG19	–	5	2011	North Sea	56.5	3.65	A	–	I I	200	206	132	142	102	102	106	106	208	212	10.544	medium	Plymouth
37	D366 120-1	MLG20	–	3	2011	North Sea	57.91	4.85	A	–	I I	206	210	132	152	102	102	102	102	208	208	10.571	medium	Plymouth
38	D366 40-3	MLG21	–	5	2011	North Sea	56.5	3.65	A	I (■)	I I	206	218	126	142	102	102	102	102	208	208	10.544	medium	Plymouth
39	PVDCH1	MLG1-Geo	–	–	2011	Pacific Ocean. Chile	−30.25	−71.7	R	-	I I	276	290	146	146	102	102	102	102	204	208	15.645	high	Roscoff
40	PVDCH8	MLG2-Geo	–	–	2011	Pacific Ocean. Chile	−30.25	−71.7	R	-	I I	282	294	150	150	102	102	102	102	204	208	15.645	high	Roscoff
41	PVDCH6	MLG3-Geo	3.36×10^{-08}	–	2011	Pacific Ocean. Chile	−30.25	−71.7	R	–	I I	282	282	156	156	102	102	102	102	208	208	15.645	high	Roscoff
42	PVDCH112	–	–	–	2011	Pacific Ocean. Chile	−30.16	−71.56	R	–	I I	282	282	156	156	102	102	102	102	208	208	15.645	high	Roscoff
43	PVDCH47	–	–	–	2011	Pacific Ocean. Chile	−30.16	−71.56	R	–	I I	282	282	156	156	102	102	102	102	208	208	15.645	high	Roscoff
44	PVDCH140	MLG4-Geo	–	–	2011	Pacific Ocean. Chile	−30.16	−71.56	R	–	I I	276	284	160	160	102	102	88	102	208	208	15.645	high	Roscoff
45	PVDCH148	MLG5-Geo	–	–	2011	Pacific Ocean. Chile	−30.16	−71.56	R	–	I I	268	276	146	146	102	130	98	154	208	208	15.645	high	Roscoff
46	UIO262	MLG6-Geo	–	–	2010	Oslo Fjord	59.25	10.71	A	–	I I	206	214	150	156	102	130	102	102	208	208	12	medium	University of Oslo
47	UIO269	MLG7-Geo	–	–	2010	Oslo Fjord	59.25	10.71	A	–	I I	206	214	146	156	102	130	102	102	208	208	12	medium	University of Oslo

Strain number	Strain ID	MLG	P_{sex} value	Station	Year	Location	Latitude	Longitude	Morphotype	CMM		EHMS37		P01E05		P02F11		P02E09		P02B12		SST	SST cluster	Culture collection
										sequencing	probe assay													
48	D366 106-4	MLG22	–	1	2011	North Sea	57.2	3.48	A	–	I	IV	208	218	126	142	102	102	102	102	208	208	10.544	medium	Plymouth
49	D366 106-5	MLG23	–	1	2011	North Sea	57.2	3.48	A	–	I	IV	210	212	132	132	102	102	102	116	208	208	10.544	medium	Plymouth
50	D366 120-2	MLG24	–	3	2011	North Sea	57.91	4.85	A	–	I	IVb	206	210	150	150	102	102	102	102	208	208	10.571	medium	Plymouth
51	D366 31-3	MLG25	–	5	2011	North Sea	56.5	3.65	A	–	I	IVb	206	210	132	142	102	102	102	102	208	208	10.544	medium	Plymouth
52	D366 34-1	MLG26	–	5	2011	North Sea	56.5	3.65	A	–	I	IV	210	210	132	142	102	102	82	102	208	208	10.544	medium	Plymouth
53	D366 36-5	MLG27	–	5	2011	North Sea	56.5	3.65	A	–	I	IVb	210	210	132	132	102	102	102	102	208	208	10.544	medium	Plymouth
54	D366 34-3	MLG28	–	5	2011	North Sea	56.5	3.65	A	–	I	IV	210	210	132	132	102	106	102	102	208	208	10.544	medium	Plymouth
55	D366 19-2	MLG29	–	5	2011	North Sea	56.5	3.65	A	–	I	IVb	196	210	142	142	102	106	102	102	208	208	10.544	medium	Plymouth
56	D366 36-2	MLG30	–	5	2011	North Sea	56.5	3.65	A	–	I	IVb	210	210	132	156	102	106	102	102	208	208	10.544	medium	Plymouth
57	D366 21-4	MLG31	–	5	2011	North Sea	56.5	3.65	A	–	I	IVb	202	206	132	142	102	106	102	102	208	208	10.544	medium	Plymouth
58	D366 31-2	MLG32	–	5	2011	North Sea	56.5	3.65	A	–	I	IV	210	210	160	160	102	102	102	102	208	208	10.544	medium	Plymouth
59	D366 48-3	MLG8-Geo	–	–	2011	Western coast of Scotland	56.78	−7.4	A	–	I	IVb	196	206	142	142	102	126	102	102	208	208	11.384	medium	Plymouth
60	D366 80-1	MLG9-Geo	–	–	2011	Bay of Biscay	45.7	−7.16	A	–	I	IV	214	214	132	146	102	102	102	102	208	212	15.523	high	Plymouth
61	D366 97-5	MLG10-Geo	–	–	2011	Western English Channel	50.08	−4.61	A	–	I	IV	206	226	132	142	102	102	102	102	208	212	13.007	medium	Plymouth
62	D366 48-5	MLG11-Geo	–	–	2011	Western coast of Scotland	45.7	−7.16	A	–	III	IVb	196	202	150	150	102	102	102	102	208	208	11.384	medium	Plymouth
63	D366 126-2	MLG33	–	4	2011	North Sea	57.45	5.53	A	IV (▲)	IV	IV	202	202	132	142	102	102	102	102	212	212	10.639	medium	Plymouth
64	D366 30-4	MLG34	–	5	2011	North Sea	56.5	3.65	A	IV (▶)	IV	IV	196	206	132	132	102	134	82	102	208	208	10.544	medium	Plymouth
65	D366 91-2	MLG35	–	5	2011	North Sea	56.5	3.65	A	–	IV	IVb	196	210	132	142	102	106	102	102	208	208	10.544	medium	Plymouth
66	D366 48-2	MLG12-Geo	–	–	2011	Western coast of Scotland	56.78	−7.4	A	IV (◀)	IV	IV	196	264	138	138	98	134	102	102	208	208	11.384	medium	Plymouth
67	D366 80-3	MLG13-Geo	–	–	2011	Bay of Biscay	45.7	−7.16	A	–	IV	IVb	276	276	146	146	102	102	102	102	208	208	15.523	high	Plymouth
68	D366 80-4	MLG14-Geo	3.24×10^{-07}	–	2011	Bay of Biscay	45.7	−7.16	A	–	IV	IVb	276	276	146	160	102	102	102	102	208	208	15.523	high	Plymouth
69	D366 80-5	–	–	–	2011	Bay of Biscay	45.7	−7.16	A	–	IV	IVb	276	276	146	160	102	102	102	102	208	208	15.523	high	Plymouth
70	D366 71-1	MLG15-Geo	4.42×10^{-05}	–	2011	Bay of Biscay	46.2	−7.21	A	IV (▼)	IV	IV	202	202	146	146	102	118	102	102	212	212	15.149	high	Plymouth
71	D366 71-4	–	–	–	2011	Bay of Biscay	46.2	−7.21	A	–	IV	IV	202	202	146	146	102	118	102	102	212	212	15.149	high	Plymouth
72	D366 98-1	MLG34	–	–	2011	Western English Channel	50.08	−4.61	A	IV (▲)	IV	IV	196	206	132	132	102	134	82	102	208	208	13.007	medium	Plymouth
73	D366 J31	MLG16-Geo	–	–	2011	Irish Sea	52.46	−5.9	A	–	IV	IV	196	230	150	150	98	102	82	102	208	208	12.268	medium	Roscoff
74	D366 J7	MLG17-Geo	–	–	2011	Irish Sea	52.46	−5.9	A	–	IV	IV	238	290	132	150	102	134	98	102	208	208	12.268	medium	Roscoff
75	PVDCH250	MLG18-Geo	–	–	2011	Pacific Ocean. Chile	−34.1	−79	R	–	IV	IV	206	214	130	130	102	102	82	82	208	220	16.429	high	Roscoff
76	PVDCH280	MLG19-Geo	–	–	2011	Pacific Ocean. Chile	−34.1	−79	R	–	IV	IV	202	260	142	160	102	102	102	102	208	208	16.429	high	Roscoff
77	PVDCH288	MLG20-Geo	–	–	2011	Pacific Ocean. Chile	−34.1	−79	R	–	IV	IV	202	260	150	150	102	102	98	102	208	208	16.429	high	Roscoff
78	BOUM6	MLG21-Geo	–	–	2008	Mediterranean Sea. Spain	39.1	5.35	A	–	IV	IV	202	260	150	156	102	102	102	102	208	208	19.395	high	Roscoff
79	BG10-6	MLG22-Geo	–	–	2007	Irish Sea	49.5	−10.5	A	IV (▲)	IV	IV	206	214	106	118	98	98	102	102	208	208	13.417	medium	Roscoff
80	EHSO_50.28	MLG23-Geo	–	–	2007	Southern Ocean	−49.58	149.25	A	IV (△)	IVb	IVb	210	210	142	142	102	120	102	154	208	208	9.133	medium	UTAS
81	EHSO_5.25Q	MLG24-Geo	–	–	2006	Southern Ocean	−49.58	149.25	A	–	IV	IV	210	210	132	156	102	102	102	154	208	208	9.133	medium	UTAS
82	EHSO_50.14	MLG25-Geo	–	–	2006	Southern Ocean	−49.58	149.25	A	IV (▼)	IV	IV	210	210	132	156	102	102	102	154	212	214	9.133	medium	UTAS
83	EHSO_50.25	MLG26-Geo	–	–	2006	Southern Ocean	−49.58	149.25	A	IV (▼)	IV	IV	202	206	132	156	102	102	102	154	212	212	9.133	medium	UTAS
84	EHSO_50.3	MLG27-Geo	–	–	2006	Southern Ocean	−49.58	149.25	A	–	IV	IV	210	210	132	170	102	120	102	154	208	212	9.133	medium	UTAS
85	EHBi_21	MLG28-Geo	–	–	2006	Bicheno. East Tasmania	−41.11	148.16	A	–	IVb	IVb	272	284	132	156	102	102	82	102	208	212	15.523	high	UTAS
86	CH25_90	MLG29-Geo	–	–	1990	North Sea	57.43	1.22	B	II	II	II	282	282	132	138	102	134	102	102	204	204	–	–	Plymouth
87	NG26	MLG30-Geo	–	–	2011	Tsushima Strait. Japan	32.42	128.67	C	II	IIb	IIb	206	210	142	142	102	130	98	102	208	208	21.257	high	Roscoff
88	EHSO_65.06	MLG31-Geo	–	–	2007	Southern Ocean	−54.11	146	B / C	II (●)	IIb	IIb	202	206	132	142	102	102	102	102	208	208	4.177	low	UTAS
89	EHSO_8.15	MLG32-Geo	–	–	2007	Southern Ocean	−53.55	145.55	B / C	II (●)	IIb	IIb	206	206	132	142	102	102	102	102	204	208	4.960	low	UTAS
90	EHSO_8.15Q	MLG33-Geo	–	–	2007	Southern Ocean	−53.55	145.55	B / C	-	IIb	IIb	206	206	132	132	102	102	102	170	204	208	4.960	low	UTAS
91	EHSO_65.17	MLG34-Geo	–	–	2007	Southern Ocean	−54.11	146.16	B / C	II (◯)	IIb	IIb	206	214	132	132	102	102	102	170	212	212	4.177	low	UTAS

2.3.4 Satellite Imagery

Ocean colour data from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor on the Aqua satellite were acquired from NASA OceanColor Website and processed to version R2013.0 using the PML Generic Earth Observation Processing System (GEOPS) (Shutler *et al.*, 2005). Chlorophyll-a concentration was estimated using the OC3M algorithm, and a 7-day median composite calculated from the cloud-free pixels to gain a synoptic view. The enhanced colour view is obtained from 7-day median composites of remote sensing reflectance at 547nm, 488nm and 443nm, combined as the red, green and blue channels respectively of an RGB image; hence this enhances the green-blue section of the visible spectrum. These images are useful for distinguishing different types of plankton or sediment: pure water looks blue; plankton blooms appear green or brown-red for more dense blooms; suspended sediment appears whitish/yellow; and *E. huxleyi* blooms appear brighter turquoise.

Sea-surface temperature (SST) data were generated from Advanced Very High Resolution Radiometer (AVHRR) data on NOAA satellites, acquired by NEODAAS-Dundee, and processed using the Panorama system (Miller *et al.*, 1997). The NOAA non-linear SST (NLSST) algorithm was applied, and again the 7-day median composite used to reduce the effect of clouds.

2.3.5. E. huxleyi clonal isolates

Culture strains used in this study are listed in Table 2.2. The D366 samples were screened and sorted using a flow cytometer (FACSORT, BD Biosciences, San Jose, CA, USA) and cell counts were assessed using a flow cytometer (Accuri C6, BD

Biosciences, San Jose, CA, USA) at the following thresholds: FSC 2000 and FL3 800. A dilution factor was calculated in order to obtain a starting concentration of approximately 1000 cells/mL. Each sample was subjected to a dilution-to-extinction regime in order to isolate individual cells and obtain clonal uni-algal cultures (Appendix II). All the cultures, including those additional geographically diverse strains resourced from various culture collection repositories (Table 2.2), were maintained in f/2 -Si medium (Guillard, 1975) in a constant temperature room at 15 °C and irradiated by a photon flux of 40-55 μmol $m^{-2}s^{-1}$ on a 16:8 hours LD cycle. The Qiagen DNeasy Blood and Tissue protocol (QIAGEN, Valencia, CA, USA) was used to extract DNA from each isolate.

2.3.6 Scanning Electron Microscopy

All of the samples were filtered using a 0.45 μm cellulose nitrate membrane filter, mounted onto metallic stubs using adhesive tape and coated in a thin layer of gold (Au) using an Au sputter coater. These were visualized using a JEOL 5600 Low Vacuum Scanning Electron Microscope. Scanning electron micrographs were captured at magnifications ranging between x8,000 - x20,000, and electron beam damage was minimized by operating the microscope at 15 kV. A total of 152 micrographs were captured, 62 from the environmental samples and 90 from the clonal isolates. All coccoliths were measured mainly at x20,000 magnification using ImageJ v1.38 software (rsb.info.nih.gov/ij/). Morphometrics included in analysis were distal shield length and width, central area length and width, average element length and width, and coccosphere diameter. To reduce bias and maintain a randomized sampling method during examination the surface area of the stubs was

divided into nine squares. For each sample, six squares were randomly allocated using a random number generator, and examined for coccospheres with coccoliths lying flat on the substrate.

2.3.7 CMM amplification and sequencing

Amplification of the coccolith morphology motif (CMM, Schroeder *et al.*, 2005) as achieved using a set of nested primers qCBP_F (5'-AGTCTCTCGACGCTGCCTC-3') and qCBP_R (5'-TGGCCTAGCACCAGTCTTTGG-3') corresponding to position 1203-1221 and 1283-1303, respectively, for the GPA mRNA of strain L (AF012542). The template DNA was added to 12.5 µL of QuantiTect Multiplex PCR NoROX kit master mix (Qiagen) and 1 µL for each probe (2 pmol), final volume of 25 µL for each reaction. PCR products were incubated with ExoSAP-IT (USB corporation) before being sequenced using the ABI Big Dye terminator cycle sequencing ready reaction kit version 3.1 (Applied Biosystems) at Geneservice, Cambridge, UK.

2.3.8 CMM probe design and multiplex assay

Dual labelled probes (Table 2.3, Figure 2.2) were designed based on multiple sequence alignments from reference CMM sequences (Schroeder *et al.*, 2005) and sequences generated from section 2.7. The probes were designed to be specific to a particular CMM group I to IV. Based on the sequence variation, two different probes were designed for CMM II and IV. The probes were divided into two multi-probe sets according to their fluorescent dyes and melting temperatures to allow for multiplexing (Table 2.3). The multiplex probe assay was carried out using a Corbette

Rotor-Gene™ 6000 (QIAGEN, Valencia, CA, USA). The PCR proceeded with an initial denaturation at 95 °C for 15 minutes, followed by 40 cycles of a two-step PCR: 94 °C for 60 seconds and 68 °C for 60 seconds for the first probe-set (probes I, II and III) and 94 °C for 60 seconds and 64 °C for 90 seconds for the second probe-set (probes IIb, IV and IVb). The fluorescence was acquired at the end of each annealing/extension step on the green, yellow and crimson channels.

Table 2.3 *Emiliania huxleyi* dual labelled probes for the CMM probe assay.

Multiplex	Probe	CMM	Sequence (5' -> 3')	Tm (ºC)	Dye (5')	Quencher (3')	Channel	Excitation/ Detection
1	Probe I	I	CCTGACG GGTGGTG GGCGGCG	68	6-FAM	BHQ1	Green	470 nm/ 510 nm
	Probe II	II	CGGCGAT TTTTATG CGCCCAC CA		ATTO680	BBQ650	Crimson	680 nm/ 712 nm
	Probe III	III	GATCGAG AGGCCTG ACGGGTG G		CY5	BBQ650	Red	625 nm/ 660 nm
2	Probe IIb	II	CGGCGAT TTTATGC GCCCACC A	64	HEX	BHQ1	Yellow	530 nm/ 555 nm
	Probe IV	IV	GGCGGCG ATTTTTA TGCCCGC CCCA		ATTO680	BBQ650	Crimson	680 nm/ 712 nm
	Probe IVb	IV	GGGGCGG CAATTTT ATGCCCG CCCCA		6-FAM	BHQ1	Green	470 nm/ 510 nm

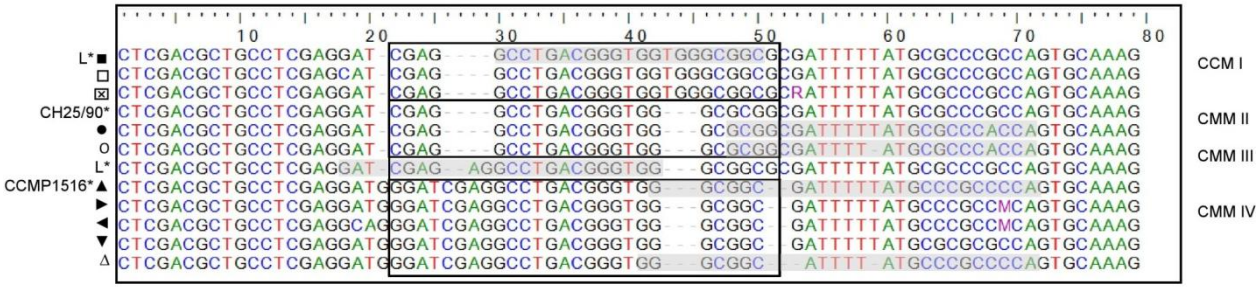


Figure 2.2 Alignment of CMM sequences produced in this study to reference CMMs (Schroeder et al. 2005). The CMM region is boxed. The dash line indicates the split between two subgroups of CMMs based on variation outside the CMM genotype. The bases shaded in grey show the positions of the probes (Table 2.3).

2.3.9 Microsatellite multilocus genotype analyses

For each of the following analyses, the biogeographic (MLG-Geo) and North Sea (MLG) bloom clonal isolates (Table 2.2) were treated separately.

Prior to analyses, the number of repeated identical multilocus microsatellite genotypes (MLG) was computed using the Mutlilocus Matches option in GENALEX, ver. 6.5 (Peakall and Smouse, 2006, 2012). This option automates detection of repeated genotypes within a dataset. The genotypic richness (R) was calculated as:

$$R = \frac{G-1}{N-1}$$

where G is the number of distinct multilocus genotypes and N is the total number of studied individuals (Dorken and Eckert, 2001). This modification of Ellstrand and Roose's (1987) index of clonal diversity was proposed by Dorken and Eckert (2001) such that the smallest possible value in a mono-clonal bloom is always 0, independently of sample size, and the maximum value is still 1, when all the different samples analysed correspond to distinct clonal lineages.

Repeated MLGs may occur due to repeated sampling of the same genet which are produced through asexual reproduction (i.e., sampling many clones of the same genotype) or two distinct sexual events wherein the resulting cells share the exact same alleles at all loci. In order to estimate whether putative genets shared the same MLG, GENCLONE 2.0 was used (Arnaud-Haond and Belkhir, 2006). For each repeated MLG, P_{sex}, which is the probability for a given multilocus genotype to be observed in N samples as a consequence of two different sexual reproductive events, was calculated. For $P_{sex} > 0.05$, duplicated multilocus genotypes were considered as different genets having arisen from two independent sexual recombination events). If $P_{sex} < 0.05$, the duplicated multilocus genotypes were considered clones of the same genet (i.e., products of asexual reproduction).

2.3.10 Null alleles and linkage disequilibria

The frequency of null alleles was estimated using a maximum likelihood estimator in the software ML-NULLFREQ (Kalinowski and Taper, 2006). Linkage disequilibrium was tested for using GENEPOP, ver. 4.1 (Rousset, 2008). In addition to physical linkage on a chromosome, disequilibria may be due to a lack of recombination caused by clonal propagation or selfing (mating system) or to differences in allele frequencies among populations (spatial genetic structure). Significance testing was done using 1,000 permutations and Bonferroni correction (Sokal and Rohlf, 1995).

2.3.11 Sampling effort

Variation in allelic richness depends, essentially, on population size—large samples are expected to have more alleles, especially rare ones, than small samples. Rarefaction (in-silico) analyses involve subsampling each sample without replacement at a range of depths. By considering these subsamples taken from each sample, samples originally of different sizes can be compared and unbiased estimates of allelic richness computed (Kalinowski, 2005). Using rarefaction, as implemented in the program HP-RARE, ver. 1.0 (Kalinowski, 2005), the mean number of alleles (i.e., the number of alleles averaged over the total number of loci used) expected with a sample size of 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 and 75 were computed. In addition, the accumulation of different genotypes sampled in the North Sea bloom was calculated for CMM and the microsatellites separately using the FASTGROUPII web-based calculator (Yu *et al.*, 2006).

2.3.12 Genetic distance

Bruvo *et al.*'s (2004) approach was used to calculate a genetic distance matrix from the alleles observed at the five microsatellite markers. The genetic distance between two 'individuals', at a single microsatellite marker, reflects the probability that the alleles of one individual mutated to the other. Probabilities are calculated using a model which assumes that slipped-strand mis-pairing is the main cause of changes in microsatellite length, resulting in single-step mutations. Notably, the Bruvo *et al.* (2004) calculation is independent of the microsatellite mutation rate, which in this study, and the majority of other studies, is unknown. A genetic distance

matrix (comparing all samples) was computed for each microsatellite marker and the average of these matrices used in the analyses described. The Polysat package (Clark and Jasieniuk, 2011) was used with R version 3.0.0 to perform the computations.

The genetic distance matrix was then analyzed using a permutational multivariate analysis of variance implemented in the R community ecology package `Vegan' (ver. 2.0-7, Oksanen *et al.*, 2012). Termed ADONIS in the software package, the function partitions the variation observed in the distance matrix into sums of square distance matrices, characterising variation attributable to specified sources. This method is a robust alternative to parametric MANOVA (multivariate analysis of variance) and to ordination methods for describing how variation is attributed to different uncontrolled covariates. ADONIS is also an alternative to AMOVA (nested analysis of variance, Excoffier *et al.*, 1992) for genetic data when there are some samples with limited numbers of individuals. Significance is assessed using *F*-statistics on sequential sums of squares from permutations of the raw data. In this study, permutational multivariate analysis of variance (ADONIS) was used to partition distance matrices among the following sources of variation in Sea Surface Temperature (SST), Northern vs. Southern Hemisphere and Locality. These tests were considered across all samples (i.e. the full genetic distance matrix) and within samples of specific CMM genotypes (i.e. submatrices of samples extracted from the full genetic distance matrix according to CMM genotype).

2.3.13 Global SSTs determination

Gridded (1° x 1°) Sea Surface Temperature (SST) data originated from the Hadley Centre (www.metoffice.gov.uk/hadobs/hadisst/). For those samples that fell outside the Hadley Centre SST coverage, i.e. the extreme coastal, their nearest SST values in a latitudinal direction were used instead. Similarly, *in situ* SST data were used for the Oslo Fjord strains. The matrices have been calculated by averaging SST values for the sampling effort (from January 2006 to December 2011). The samples were then clustered using a hierarchical clustering algorithm (termed hclust) implemented in R (version 3.0.0). The algorithm starts with each sample as a cluster in itself and merges clusters together sequentially using Ward's minimum variance criterion (Ward, 1963). The sequential merging was continued until all samples were contained in a single cluster and the subsequent tree describing how the clusters merged was `cut' to yield three clusters. These clusters formed the low, medium and high SST groups.

2.4 Results

2.4.1 Genetic inheritance, polymorphism and stability of the microsatellite markers

Loci P01F08 and P02A08 did not produce any PCR products after repeated attempts and alteration of PCR conditions (Table 2.1). These two markers were, therefore, the first to be eliminated from the suite of loci. In addition, there were no hits against the CCMP1516 genome for either of these two primer pairs (Table 2.1). Of

the remaining eight markers that produced products, P02E11, P02E10 and EHMS15 resulted in multi-allelic (i.e., more than two, the maximum number of alleles possible for a diploid) profiles. There were at least three distinct peaks corresponding to at least three different alleles (Supplementary Figure 2.1). Altering PCR conditions resulted in different allelic peaks rendering these loci unrepeatable. Moreover, P02E10 and EHMS15 primer pairs were found five and two times, respectively, in the CCMP1516 genome (Table 2.1). The multiple hits suggested these primer pairs may have amplified more than one region in the genome which corresponded to the multi-peaked profiles observed. As they were not repeatable and did not follow single-locus genetic determinism, they were rejected from further analyses.

EHMS37, P01E05, P02F11, P02E09 and P02B12 produced consistent results at their original annealing temperatures as well as the modified PCR program with an annealing temperature of 54 °C. For each of these polymorphic markers, single-locus Mendelian inheritance was assumed as only one (i.e., homozygous) or two peaks (i.e., heterozygous) were observed for each of the clonal isolates tested. For the 15 samples (5 replicates, 3 different CO_2 conditions) from Lohbeck et al. (2012) extracted at the start of the CO_2 selection experiment in 2010, there were no differences between replicates and treatments. Further, in the same replicate selection lines extracted after 1300 generations of exponential growth, there was no change in the alleles present at each locus (Table 2.4). However, CCMP1516 showed variation in allele number and size for both EHMS37 and P01E05; the two most polymorphic loci (section 3.5). When comparing the genome sequence (Read *et al.*, 2013) and previously characterized microsatellite data for this strain (Mackinder *et al.*, 2011b) to our PCR amplicons, variation extended to the locus P02E09. The loss of

the 137 PO1E05 allele in strain CCMP1516 genotyped in this study coincided with the loss of calcification, i.e. failure to produce a coccolithosphere. Unfortunately, Mackinder *et al.* (2011) did not look at this allele (Table 2.4). Moreover, CCMP1516 can no longer produce haploid flagellate life-forms (P. von Dassow, *personal communication*), therefore these genetic modifications were not due to sexual recombination.

2.4.2 D366 E. huxleyi cultures

The techniques used to isolate clonal uni-algal *E. huxleyi* strains from the D366 cruise, selected only for calcified (diploid) forms that were cultured. We successfully produced 104 isolates from single cells, 88 (85%) remained viable (data not shown). Of these, 65 D366 isolates were successfully genotyped (Table 2.2), 52 of which originated from the North Sea bloom event (Figure 2.1). *E. huxleyi* morphotype A was the only morphotype to be identified (Figure 2.3). The mean coccosphere diameter was 5.4 μm (range 3.9-7.5 μm). Coccolith dimensions (Figure 2.4) were consistent with the classic morphotype A phenotype. The mean coccolith distal shield length was 3.2 μm ranging between 2.1-4.4 μm (Figure 2.4a), and the mean distal shield width was 2.6 μm ranging from 1.5 to 4 μm (Figure 2.4b). The mean central area length was 1.6 μm (range 1.2-2.5 μm), and the mean central area width was 1.1 μm (range 0.7-1.7 μm). The mean average element length was 0.63 μm (range 0.25-0.95 μm), and the mean average element width was 0.12 μm (range 0.09-0.16 μm). All consistent with the classic morphotype A phenotype (Young *et al.,* 2003).

Table 2.4 Microsatellite stability over multiple generations.

Sample	Year	Generations	EHMS37		P01E05		P02F11		P02E09		P02B12		Source
Lohbeck*	2010	0	208	214	124	148	102	104	102	104	208	208	this study
	2012	1300	208	214	124	148	102	104	102	104	208	208	this study
CCMP1516	2007	-	341		158		no hit		100		no hit		Read *et al.* (2013)^
	2010	-	339	339	ND		119	193	96	102	212	216	Mackinder *et al.* (2011a)
	2010“	-	339	339	ND		119	193	96	102	212	216	Mackinder *et al.* (2011a)
	2010	ND	338	340	137	153	120	192	100	106	212	216	this study
	2011	ND	340	340	137	153	120	192	100	106	212	216	this study
	2012“	ND	338	340	153	153	120	192	100	106	212	216	this study

*: Lohbeck et al. (2013)
": independent loss of coccolithsphere production
^: from the genome
ND: not determined

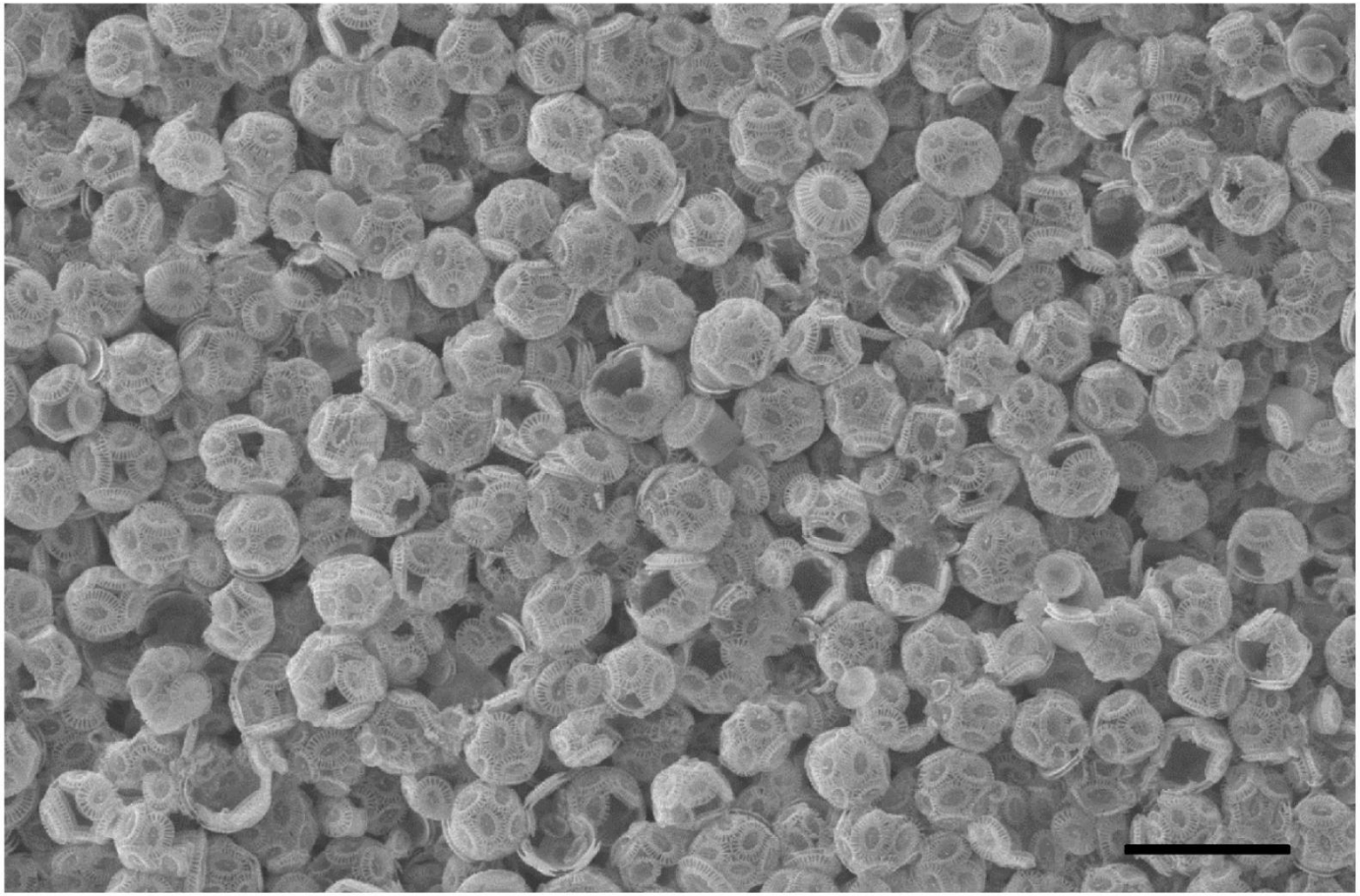


Figure 2.3 Scanning electron micrograph of a mixed *Emiliania huxleyi* culture prior to single cell isolation originating from D366 station 5 in the North Sea. Bar = 5 μm.

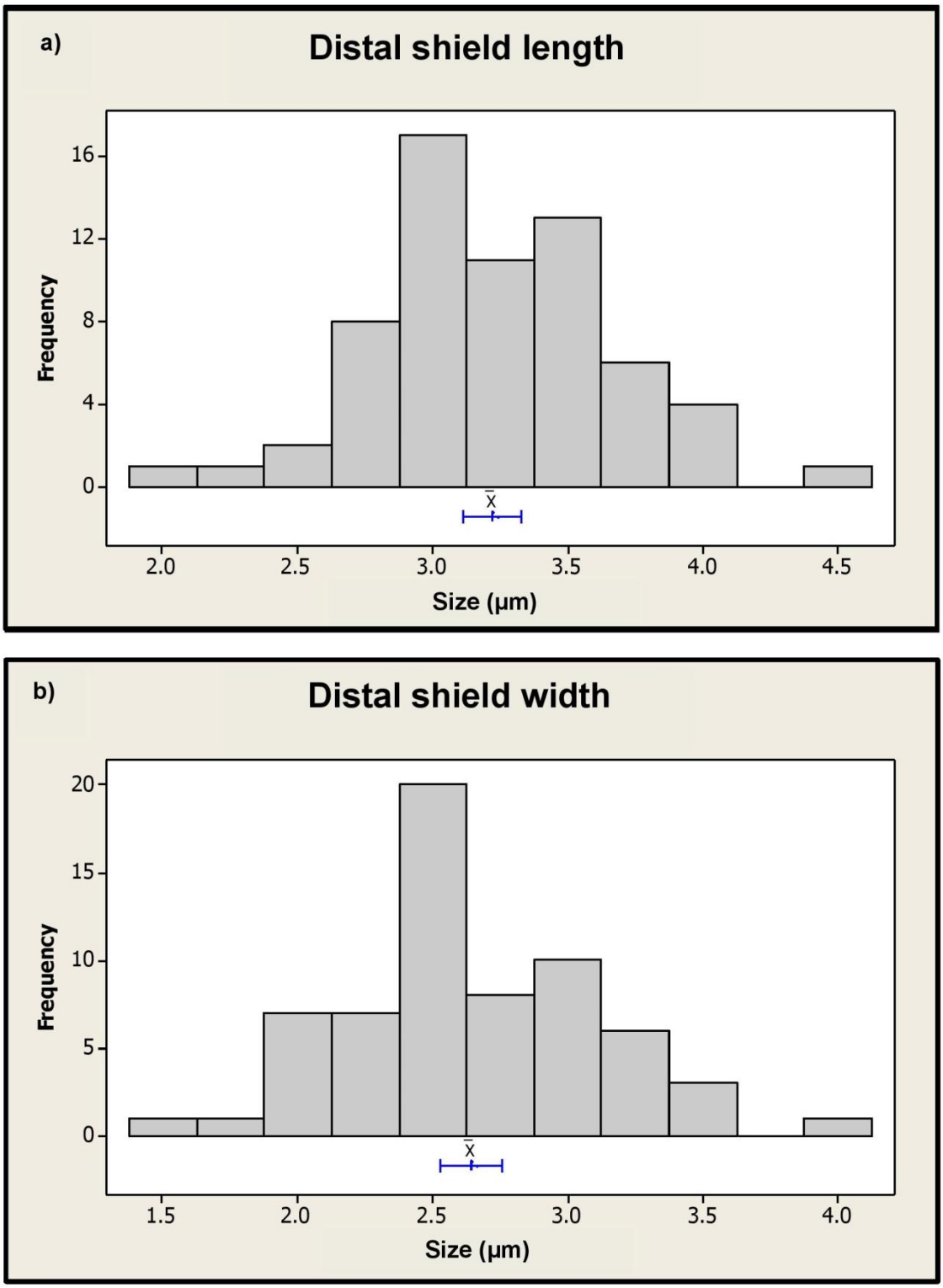


Figure 2.4 Frequency distribution histograms of all the measurements taken for distal shield length (a) and width (b): 95% t-confidence for mean is shown.

2.4.3 Biogeographic E. huxleyi cultures

A select group of 26 *E. huxleyi* strains were chosen based mainly on origin and date of isolation. Our aim was to include strains from diverse geographic locations, from both northern and southern hemispheres and disparate climatic environments. In addition, we wanted to restrict the age of the cultures to lessen the influence of genetic drift from the point of isolation. Our final data set comprised strains not more than 5 years older than D366 strains, with the only exception being strain CH25/90 (Table 2.2) as the most recent and only one of two reference strains for morphotype B (CMM II) still in culture (Schroeder *et al.*, 2005). The majority (84%) of all the biogeographic samples, including the D366 cultures, were isolated in 2011. Twenty isolates originate from the Southern hemisphere, while 6 isolates were isolated from the Mediterranean Sea, Oslo fjord, Irish Sea and Tsushima Strait, Japan (Table 2.2). The SST experienced by these strains ranged from 4.1 to 21.2 ºC (Figure 2.5). All strains could be clustered into three SST groups, namely low, <5 ºC, medium >5 & <14.3 ºC, and high >14.3 ºC (Table 2.2). The North Sea SSTs as observed by AVHRR (Figure 2.1c) are consistent with the SST clustering ranges that were based on Hadley Centre temperatures (Table 2.2).

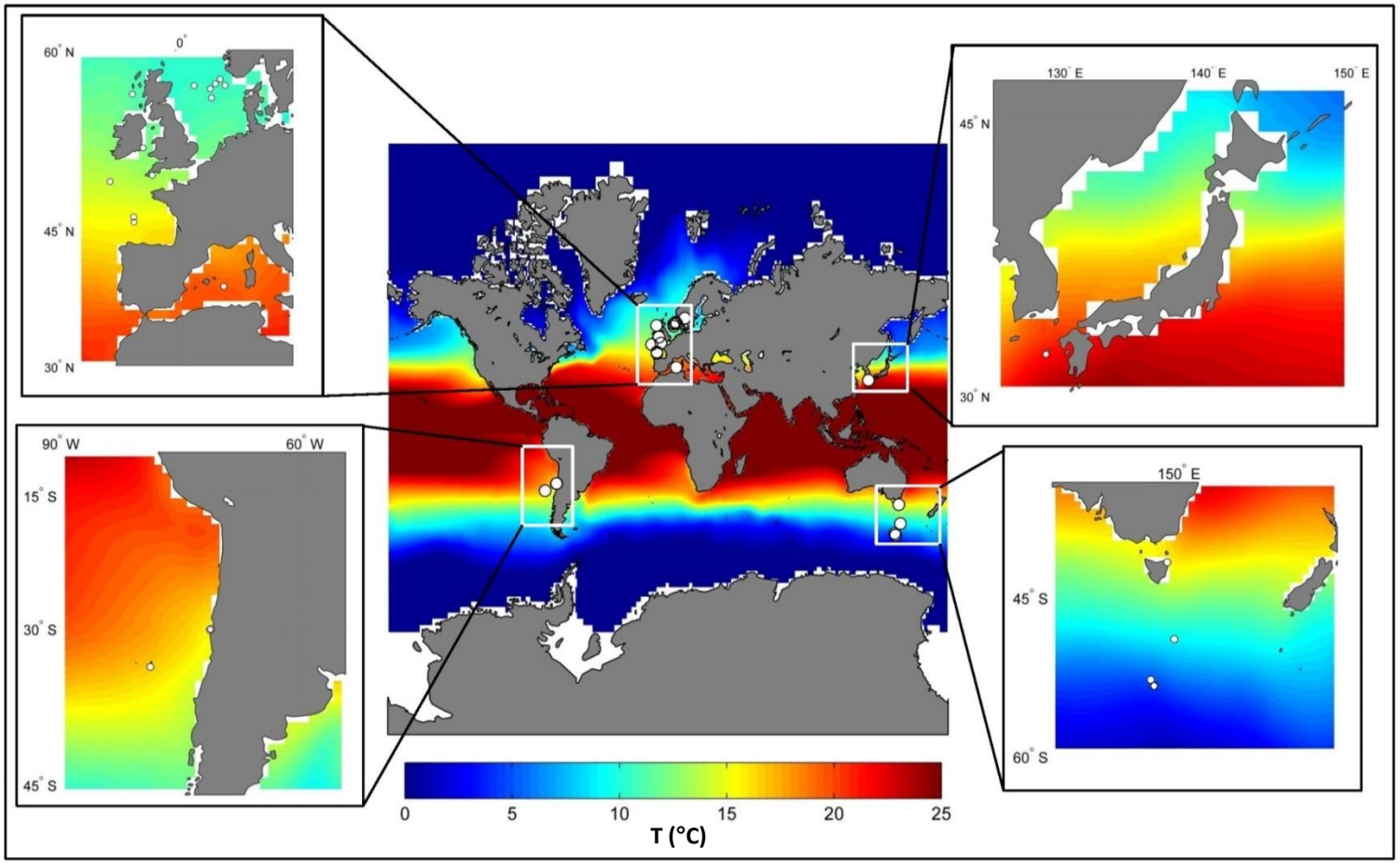


Figure 2.5 Average Sea Surface Temperature (SST) values for the sampling effort from January 2006 to December 2011 for the world's oceans. The four regions that include Europe, Japan, Chile and Australia that represent all our dataset are shown in greater detail. Key: temperature colour index from blue to red, 0ºC to 25ºC, respectively.

2.4.4 CMM genotyping

Isolates in our reduced D366 dataset could be divided into three main CMM groups, namely homozygous for CMM I, homozygous for CMM IV and heterozygous for CMM I/IV & III/IV (Table 2.2). It is, however, important to note that two of the 13 isolates that did not make the final reduced D366 dataset, produced complex MLGs and CMM profiles; all indicative of the presence of multiple genotypes in the same sample (data not shown). For technical reasons, these and the remainder 11 strains were not included in later analyses.

The CMM identity was mainly determined by applying the multiplex CMM probe assays (Supplementary Figures 2.2 and 2.3), with sequencing of CMM amplicons from a few isolates to validate the probe assay results (Table 2.2). Note that multiple CMM probes were designed to account for the additional sequence variation outside the designated CMM region (Figure 2.2). When this was taken into account for two of the main affected CMMs, namely CMM II and IV, both sets of probes improved the sensitivity of the assay.

Of the North Sea D366 clonal isolates, 38 were homozygous for CMM I, 3 were homozygous for the CMM IV and 11 were heterozygous for CMM I/IV (Table 2.2). Therefore, CMM I was the most numerically abundant genotype. CMM I in a homozygous state was also found in other geographic strains, seven were of Chilean and two of Norwegian origins (Table 2.1). Similarly, CMM IVs were distributed widely geographically, while CMM I/IVs where restricted to the Northern hemisphere.

No CMM IIs were detected in our D366 dataset. The five B/C and C morphotypes from the Southern Ocean and Tsushima Strait, respectively, were

however shown only to have the CMM II genotype (Table 2.2). There are 91 samples in this data set and of these 6 are homozygous CMM II (including the homozygous CMM II morphotype B Ch25/90 reference strain - Schroeder *et al.*, 2005). Furthermore, exactly these 6 samples are characterized by a morphotype other than type A (morphotype R being a Southern Ocean over-calcified variant of A). The probability that these non-morphotype A samples are the only CMM II genotypes by chance is $1/({}^{91}C_6) = 1.5e\text{-}09$. The number ${}^{91}C_6 = 666563898$ is the total number of ways 6 samples can be selected from 91, it suggests the observed result is highly unlikely to have occurred by random chance.

2.4.5 Microsatellite genotyping

There were significantly greater amplification rates in this study (Table 2.2) compared to Iglesias-Rodríguez *et al.* (2006; $t = 5.18$, $df = 5$, $p = 0.004$), but no difference between this study and Hinz (2010; $t = 0.75$, $df = 4$, $p = 0.493$). However, the amplification rate at locus P02B12 in Hinz (2010) was only 66%, whereas in this study it was 100%.

One hundred and eight alleles were characterized across the five microsatellite loci. The number of alleles ranged from two to ten in the North Sea bloom, whereas there were five to 17 alleles encountered on a global scale (Table 2.2). Each of the loci corresponded to a stepwise mutation model. EHMS37 was the most polymorphic locus whereas P02B12 was the least polymorphic locus. Allele frequencies are available upon request.

Of the 52 clonal isolates genotyped in the North Sea bloom, 26 MLGs were

only encountered once, five MLGs were encountered twice, two MLGs were encountered three times, one MLG was encountered five times and, finally, one MLG was encountered six times. The genotypic richness, *R*, in the North Sea was 0.667, the smallest value reported during a phytoplankton bloom. Moreover, each duplicated MLG was characterized by P_{sex} values much smaller than 0.05 (Table 2.2). In other words, it was extremely unlikely that they were the product of two independent meiotic events. All repeated microsatellite MLGs also shared the same CMM allele. Consequently, all repeated MLGs were considered descendants of the same genotype. In addition, there was also a repeated microsatellite MLG encountered three times in a bloom sampled off the coast of Chile in 2011 (Table 2.2). This repeated MLG exhibited P_{sex} values much smaller than 0.05 (Table 2.2) and, as above, was considered descendants of the same genotype.

There was no evidence of linkage disequilibrium in the North Sea bloom (i.e., all p-values were > 0.05 before Bonferroni correction). There was evidence of null alleles at each locus except P02F11 in the North Sea bloom. The null allele frequencies varied from 0.194 at EHMS37 to 0.258 at P01E05. However, as demonstrated by Krueger-Hadfield *et al.* (2011, 2013), null allele frequencies calculated in diploid stages of haploid-diploid life cycles could be biased due to violation of some of the assumptions underlying maximum likelihood estimators. Therefore, null alleles may be present in our diploid strains (i.e., a diploid strain may have been scored as homozygous at locus EHMS37, but was in fact a heterozygote for the allele amplified and for an allele that was not amplified due to, for example, a possible mutation in the primer binding site). However, the frequency estimates are likely upwardly biased and the actual numerical value should be treated with caution

as we are unsure of certain parameters of the *E. huxleyi* life cycle (i.e., mating system), which could bias the maximum likelihood estimator.

2.4.6 Sampling effort

There was a difference between CMM and the microsatellites in that the rarefaction curve for CMM genotypes reached a plateau whereas the microsatellites did not (Supplementary Figure 2.4). Although the microsatellite rarefaction curve did not plateau, at the point at which sampling was ceased, the gradient of curve was not as steep as that observed in other studies (e.g., Hinz, 2010). That said, a slight increase did occur between 50 and 75 genes sampled (Supplementary Figure 2.4).

2.4.7 Population genetic structure at different spatial scales

Using the ADONIS method to attribute variation in microsatellite Bruvo genetic distances (Figure 2.6) to variation in SST, Northern vs Southern hemispheres and locality yielded weak correlations: between 8 and 31% of the variation in the distance matrix was explained by these variables (Table 2.5). In addition, the morphotypes did not cluster together on the basis of microsatellite genetic distance, notably the four B/C morphotypes from the cooler Australian waters were dispersed between other morphotypes (Figure 2.7). Within CMM genotypes, locality explained the most variation out of the three covariates.

Table 2.5 ADONIS output with three different clustering variables: SST, Northern vs. Southern hemispheres (North vs. South) and Locality. Each model is fitted to all samples, CMM type I/I samples only, CMM type II/II samples only, CMM type I/IV samples only, and CMM type IV/IV samples only.

Clustering variables	**Samples (N)**	**R^2 "**	**DF^**
SST	All (71)	12.9	2
	I-I (28)	19.8	1
	II-II (5*)	39.0	1
	I-IV (15)	9.0	1
	IV-IV (23)	15.8	1
North vs. South	All (71)	8.8	1
	I-I (28)	19.8	1
	II-II(5*)	39.0	1
	I-IV (15)	NA	NA
	IV-IV (23)	12.8	1
Locality	All (71)	31.1	9
	I-I (28)	25.6	2
	II-II(5*)	39.0	1
	I-IV (15)	33.6	3
	IV-IV (23)	51.3	7

N: sample size

*: small sample size

": R^2 indicates the proportion (%) of variability accounted for by the clustering variable

^: DF is the number of free parameters in the model

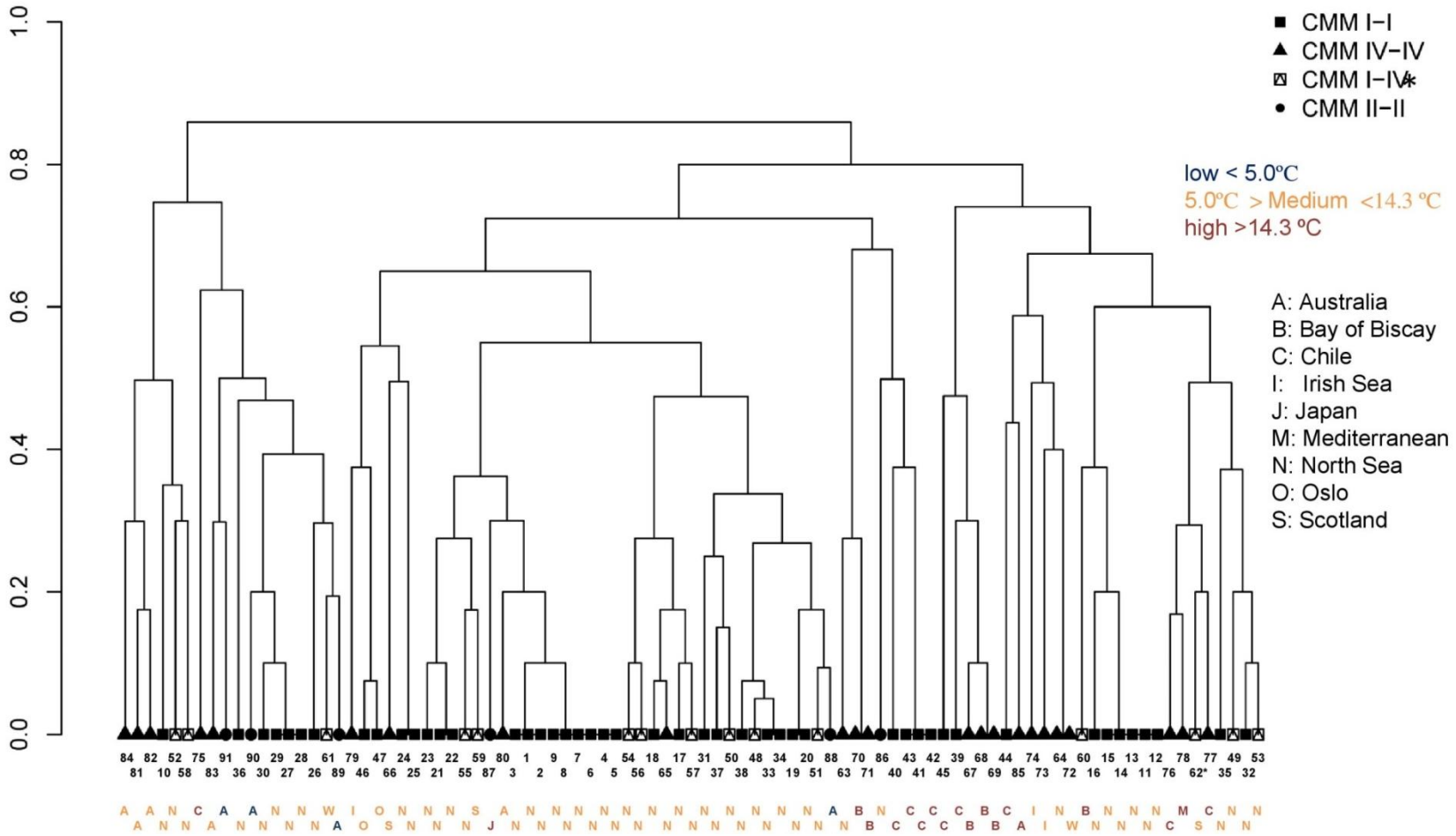
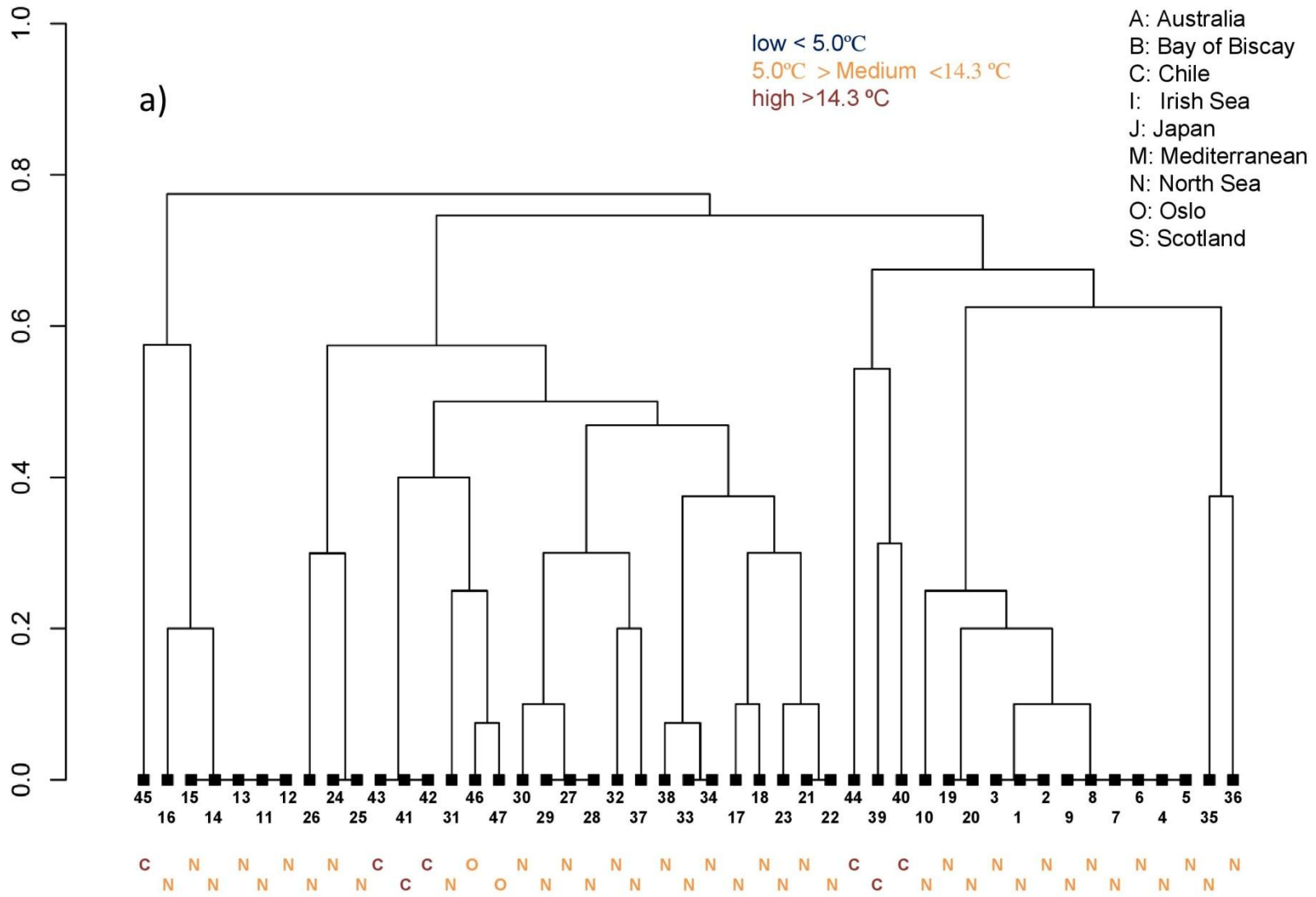
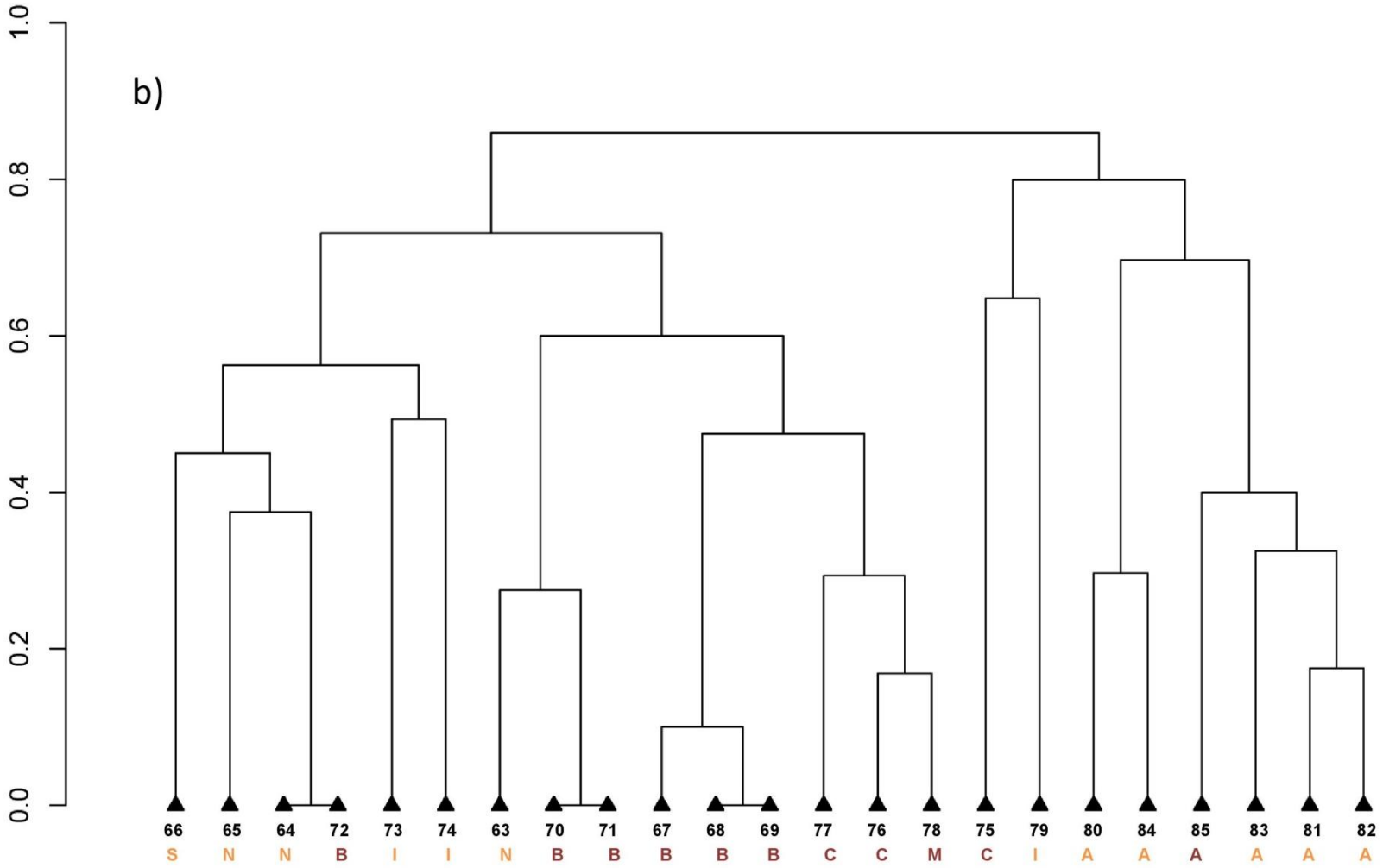


Figure 2.6 Multi-dimensional scaling plots constructed using Bruvo et al.'s (2004) genetic distance creating a 2-dimensional representation of the dissimilarity matrix used for the permutational multivariate analysis of variance (ADONIS {VEGAN} community ecology package in R) for all the samples.


a)
low < 5.0ºC
5.0ºC > Medium <14.3 ºC
high >14.3 ºC
A: Australia
B: Bay of Biscay
C: Chile
I: Irish Sea
J: Japan
M: Mediterranean
N: North Sea
O: Oslo
S: Scotland
1.0
0.8
0.6
0.4
0.2
0.0
45 15 13 12 24 43 42 46 30 27 32 38 34 18 21 44 40 19 3 2 8 6 5 36
16 14 11 26 25 41 31 47 29 28 37 33 17 23 22 39 10 20 1 9 7 4 35
C N N N N C C O N N N N N N N C C N N N N N N N
N N N N N C N O N N N N N N N C N N N N N N N


b)
1.0
0.8
0.6
0.4
0.2
0.0
66 65 64 72 73 74 63 70 71 67 68 69 77 76 78 75 79 80 84 85 83 81 82
S N N B I I N B B B B B C C M C I A A A A A A

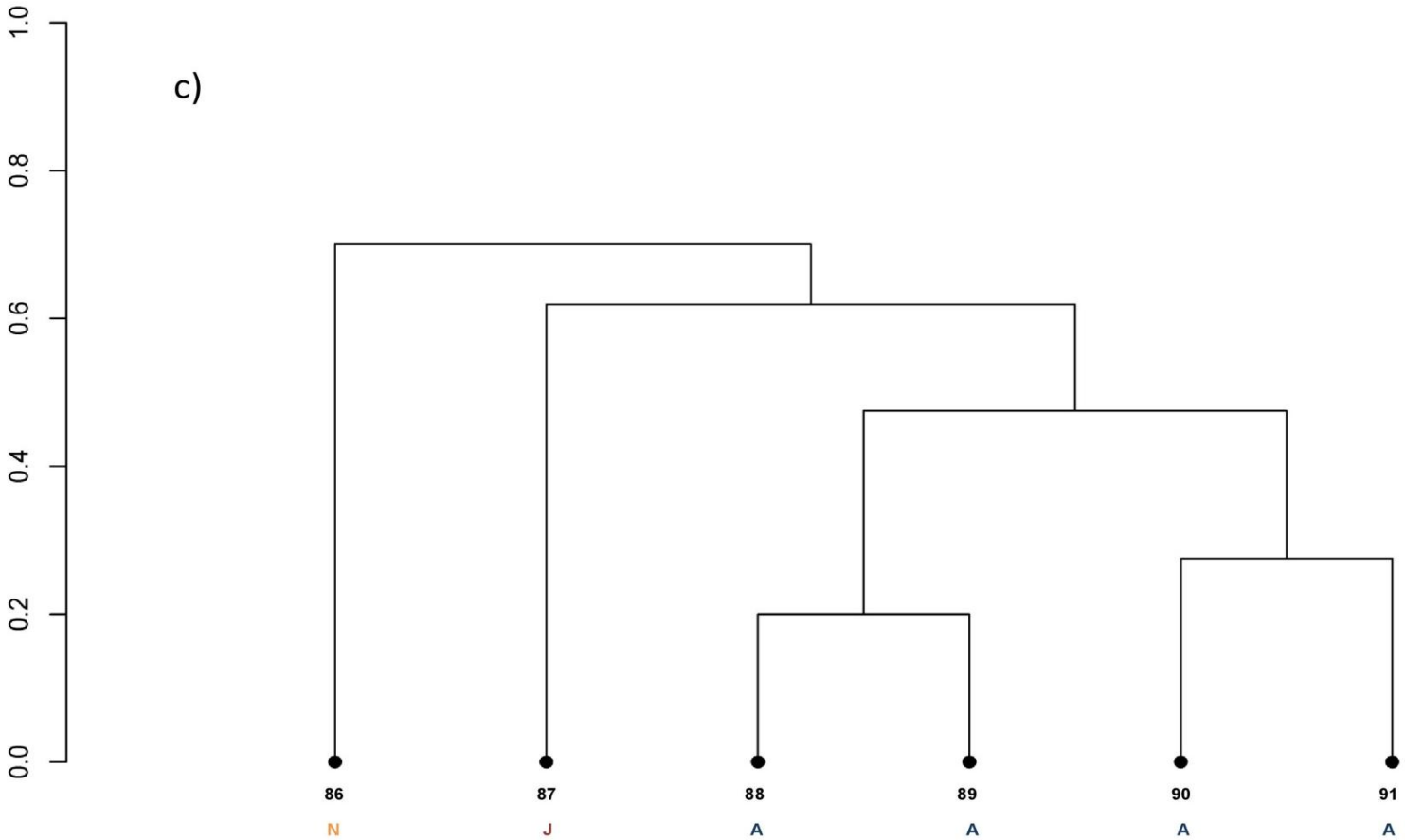


Figure 2.7 Multi-dimensional scaling plots constructed using Bruvo et al.'s (2004) genetic distance creating a 2-dimensional representation of the dissimilarity matrix used for the permutational multivariate analysis of variance (ADONIS {VEGAN} community ecology package in R) in the biogeographic group: a) CMM I, b) CMM IV & c) CMM II.

2.5 Discussion

The use of a validated set of microsatellites and the CMM functional genetic marker demonstrated clear evidence of asexual reproduction prevailing during a single *E. huxleyi* bloom event in the North Sea in 2011. Eight genotypes were encountered between two to six times across the sampling dates and locations of the bloom event. Despite the small sample size, there were many more repeated genotypes than previously reported for other bloom-forming phytoplankton species, including a previously genotyped *E. huxleyi* bloom event. This study challenges the assumption that sex drives genetic diversity within and between *E. huxleyi* populations. Whilst genetic diversity is high amongst extant populations of *E. huxleyi*, the root cause for this diversity still requires further examination in order to be able to predict the impacts of unprecedented levels of climate change are having on key biological species such as *E. huxleyi*.

2.5.1 Asexual dominance in the D366 North Sea Bloom

For population genetics, the key benefit of microsatellites is the high inter-individual variation, which makes it possible to study both intra- and inter-population genetic diversity. The evolutionary dynamics, biological function, genomic distribution and practicality of microsatellites have been summarized in a wide variety of reviews (see Schlötterer, 1998; Selkoe & Toonen, 2006). As a down-side, mutation rates may be so high that appreciable genotypic changes may occur during an observational period (e.g., Tesson *et al.*, 2013). However, whether these are real mutations or mis-scoring (discussed again below) would need more careful analysis.

Microsatellite mutation rates vary, but the typical range is thought to be 10^{-2} to 10^{-6} mutations per locus per generation (Li *et al.*, 2002). Hinz (2010) estimated the number of mutations per microsatellite locus per generation in *E. huxleyi* to be between $7x10^{-3}$ to 142 over a 15 year culture period. Assuming this calculation is meaningful for certain strains, 1 mutation per 1000 generations is expected statistically within each lineage. As each of these mutations would be selectively neutral, the probability of fixation would be negligible and would be dependent upon the size of the asexual population. In other words, even if occasional mutations occurred in uni-algal cultures, it would not be possible to detect - as seen for the Lohbeck *et al.*, (2012b) strain that did not show any changes based on microsatellite genotyping during 1300 asexual generations. However, we investigated a second strain (CCMP1516) that originates from the warmer tropical Pacific environment and has been in culture since 1991 (Schroeder *et al.*, 2005). In contrast, the strain used in Lohbeck *et al.* (2012b) originates from Bergen (relative cooler environment) and was maintained in culture for a lot less time (i.e., since 2009) and under continuous exponential growth. Our data suggests the change in selective pressure incurred due to culturing in artificial laboratory conditions over a 20 year time period has had a compounding effect on fitness. While adaptation to high pCO_2 conditions had little effect on Lohbeck strains ability to calcify (i.e. cells never lost their ability to produce coccoliths), we predict that the same would not be true for CCMP1516. We predict that it would have behaved very differently as it often loses its ability to calcify under current pCO_2 scenarios. Replicate cultures of CCMP1516 have to be kept to ensure that the calcified form of CCMP1516 is not lost for good.

Mis-scoring of alleles was certainly a problem for CCMP1516 (Table 2.4). The variations observed in the EHMS37 and P02F11 are likely as a result of noise, user interpretation and between sequencer shifts associated with the stutter peaks surrounding the "dominant" microsatellite peak (expanded upon again later). By contrast, the variations observed in P01E05 and PO2E09 are more intriguing. What is the source of this variation? Could the P01E05 loci be informative about the state of calcification? We know that the allele size 137 for PO1E05 was likely present in the genome sequence dataset (Read *et al.*, 2013) but was omitted from the final genome due to the complexities of assembly, i.e. assembly of genomes of diploid organisms eliminates subtle variation and reports mainly on a single consensus chromosomal copy. However, the disappearance of this allele in the 2012 non-calcifying strain (Table 2.4) raises important questions regarding the role of this genomic region in the calcification process. What is certain, however, is that some genomic regions within *E. huxleyi* are subject to greater genetic drift or rearrangements within an asexually maintained state. Until we determine the source and the nature of these variations and understand the effect and extent of the changes on the fitness of a diversity of strains, estimation of microsatellite mutation rates per locus for *E. huxleyi* would be futile. This in turn raises questions of the usefulness of these particular microsatellites in *E. huxleyi* population genetics.

Microsatellites have previously been used to explore genetic diversity and population structure in several bloom-forming phytoplankton (e.g., diatoms: Rynearson & Armbrust, 2000, 2004, 2005; Evans *et al.*, 2005; dinoflagellates: Alpermann *et al.*, 2009; Erdner *et al.*, 2011; Casabianca *et al.*, 2012; coccolithophores: Iglesias-Rodríguez *et al.*, 2006). High levels of intraspecific genetic variability have

been reported in all phytoplankton groups, but often these results are discussed as somewhat of a paradox. A bloom event should be dominated by asexual reproduction, as asexual reproduction is likely the only mode by which such large biomass can be generated over short time periods. Yet, the paradigm of sexual reproduction being the source of exceptional genetic diversity during bloom periods has pervaded the microbial literature. For *E. huxleyi,* we have seen that sexual recombination was not the cause of the microsatellite variation observed in CCMP1516. This has been documented in other asexually reproducing organisms, such as fungi. Sexual recombination was thought to only occur between two fungal strains of opposite mating types; however, Lin *et al.* (2005) demonstrated recombination in isogenic mating types. We have no evidence that recombination between diploid *E. huxleyi* cells are the source for the genetic variation observed, but this merely highlights the many possibilities that could explain high levels of genetic variation within species. Due to the high levels of genetic diversity and linkage equilibrium observed in our study, genetic drift had occurred, but was unlikely to have contributed to genetic diversity directly during the D366 North Sea bloom. Indeed, rare recombination events can erase any signatures of clonality, such as heterozygote excess and linkage disequilibrium (Halkett *et al.*, 2005). Yet, the fact that many genotypes were re-sampled indicates that asexual reproduction was driving the bloom formation.

This is one of the only studies which calculated P_{sex} values in order to demonstrate the origin of the repeated MLGs (sexual or asexual events). In contrast, Iglesias-Rodríguez *et al.* (2006) and Hinz (2010) reported few, if any, repeated MLGs in two previous studies on *E. huxleyi* blooms, but this is likely due to several features

of these studies which do not arise directly from the biology of this coccolithophore. First, the sample size used to calculate genetic diversity from a sampling location or time point (Iglesias-Rodríguez *et al.*, 2006) or a particular mesocosm or time point (Hinz 2010), was small and therefore, repeated genotypes may not be detected due to chance or isolation techniques. Second, Iglesias-Rodríguez *et al.* (2006) included several loci which have been shown in this study to be multi-allelic and are, therefore, not suitable for genotypic diversity estimates. Further, only seven out of the 85 isolates tested amplified at all ten loci. It is unclear from Iglesias-Rodríguez *et al.* (2006) what the genotypes were for the validated five loci used in this study and whether these genotypes were in fact different. Third, in Iglesias-Rodríguez *et al.* (2006), the authors used two microsatellites, P01E05 (potentially mutating after long periods of time in culture) and EHMS15 (multi-allelic), in isolation to describe the geographic distribution of genotypes and potential reductions in gene flow. However, if one uses restricted data sets to perform these calculations, such as between Northern and Southern hemisphere strains, spurious results will be encountered. For example, we demonstrated that SST, Northern vs. Southern hemisphere and Locality does not explain the overall clustering of the strains based on CMM or microsatellite profiling.

Iglesias-Rodriguez *et al.* (2006) also estimated the number of genotypes in the environment to be, at the minimum, 2.4×10^{20}. Yet, the computational method of calculating this value depends on locus independence. There were no calculations of linkage disequilibrium, but if one assumes the loci are independent and in linkage equilibrium based on the results of the current study, this would not be a major violation. However, the method likely overestimates the number of different

genotypes. If there were four alleles at a locus, then in Iglesias-Rodríguez *et al.*'s (2006) method, there would be six different heterozygous combinations plus the four possible homozygous states. This would then be multiplied by the next figure at the next locus and so on. The computational method used does not take into account the manner in which certain alleles are encountered or that some combinations are never found. Capture-recapture statistics is a preferred method to estimate the number of lineages within a bloom in a conservative manner.

One issue with studies, such as this in coccolithophores (also see Cook *et al.*, 2013) or in diatoms, as in Rynearson and Armbrust (2005), is the sample size of clonal isolates from a given "site". For macroalgae, it is necessary to sample at least 30 diploids and haploids (for those which have haploid-diploid life cycles) from a population (Krueger-Hadfield *et al.*, 2011). However, due to the difficulty of single cell extractions in some phytoplankton and the large scale of their distribution and bloom events, more than 30 samples of at least the diploid phase are likely to be necessary. For example, the daily sample size of clonal isolates from Rynearson and Armbrust (2005) varied from 20 to 76 with values of D ranging from 0.87 to 1.0. Plotting the N versus R resulted in a significant negative slope (r^2= 0.456, b = -0.001, p < 0.023), indicating that increasing the sample size of clonal isolates increases the chances of re-encountering a MLG.

Yet, even values in Rynearson and Armbrust (2005) with apparently sufficient sample size to detect repeated MLGs, there were still more unique MLGs encountered than in the North Sea *E. huxleyi* bloom studied here. This might be expected due to the nature of diatom blooms. Diatoms continue dividing until they reach a critical size when sexual reproduction is triggered (Chepurnov *et al.*, 2005). However,

Rynearson and Armbrust (2005) did not find any sexual stages during the sampling of a *Ditylum brightwellii* bloom event in Puget Sound. Therefore, the high genotypic diversity in the diatom bloom may have been due to past sexual events, but also resting stages of *D. brightwellii*. Resting stages can act as inocula for blooms and provide an additional diversifying effect.

2.5.2 A place for CMM

Ascribing a genetic basis to a particular coccolithophore morphotype has been attempted in several studies which were able to show some genetic differentiation among the strains tested (gpa/CMM: Schroeder *et al.*, 2005, tufA: Cook *et al.*, 2011, cox1b and atp4 Hagino *et al.*, 2011). There are four main morphotypes: Type A [*E. huxleyi* var *huxleyi*] has varying levels of calcification, global distribution and is the most prevalent in bloom events (Hagino *et al.*, 2011; Cook *et al.*, 2011, 2013). The other three, namely C [*E. huxleyi* var *kleijniae* Young & Westbroek ex Medlin & Green] (Young *et al.*, 2003), B [*E. huxleyi* var *pujosae* (Verbeck) Young & Westbroek ex Medlin & Green] and B/C [*Emiliania huxleyi* var *aurorae* Cook & Hallegraeff] are found in the most northern and southern latitudes (van Bleijswijk *et al.*, 1991; Young *et al.*, 2003; Cook *et al.*, 2013). Two other morphotypes, R (Young *et al.*, 2003; Cook *et al.*, 2011) & O (Hagino *et al.*, 2011) have been reported in the southern and northern latitudes, respectively. Schroeder *et al.* (2005) used the CMM to reinforce the partitioning of the A & B morphotypes. In addition, morphotype A has a combination of CMM I, CMM III or CMM IV alleles, while morphotype B was only found associated with CMM II. The present study has expanded on this finding by showing that the morphotype R is likely an over-calcified form of A, and more

surprisingly linking morphotypes C and B/Cs to B. While the latter share a similar biogeography, their cell sizes span the smallest (C – 2.5 μm) to the largest (B – 7 μm) for this species.

CMM I was the numerically dominant allele in the form of homozygous CMM I and heterozygous CMM I/IV. However, CMM IV was the second most abundant genotype and the most widely distributed. This was partially supported by the ADONIS variation test (i.e. locality being the greatest influence on the genetic variation for homozygous CMM IV), but also by the discovery of a CMM IV repeated MLG in the North Sea and the Western English Channel (see Table 2.2, MLG 34).

CMM II, on the other hand, was not detected in the North Sea locality. One of the original B morphotype strains, CH25/90, originated from the North Sea (van Bleijswijk *et al.*, 1994) at a location not too dissimilar from the D366 North Sea sampling sites. In addition, Martínez *et al.* (2012) reported the presence of CMM II in the North Sea in 1999. The absence of morphotype B or CMM II in our D366 culture collection raises important questions as to whether the well-documented increase in SSTs over the past decade could have negatively affected the natural habitat for this morphotype. We know that CMM IIs, including B/C and C's, predominantly or even exclusively occupy the more northern and southern latitudes. It is conceivable to predict that in the case of the North Atlantic the morphotype B's could have moved further north to cooler environments. Helaouët *et al.* (2011) showed a similar northward movement for the copepod, *Calanus*, over the past decade. Higher spatial and temporal resolution is required before we can conclude that climate change could also have attributed to the range restriction of morphotype B. Taken together, morphotype A appears to be more resilient and thus dominates at a regional and

global scale while morphotype B is more sensitive and thus likely to be more specific to the niche it occupies.

Assessing the CMMs to clonal samples obtained from environmental cultures raise some issues regarding the reliability of the results when making ecological inferences about *E. huxleyi* wild populations. Every sample taken from the environment and transferred into an artificial habitat is subjected to selection, genetic mutation and subsequent mutation accumulation, and recombination (Lakeman *et al.*, 2009). These artificial lab-induced processes alter the final interpretation of the frequencies of alleles in field populations., In order to obtain results which mirror the 'real' composition of a community, the strains should be isolated as soon as possible from the environmental material it was collected from and they should be maintained at similar conditions to that from which they were sampled (Lakeman *et al.*, 2009). Nevertheless, certain genotypes or strains more suited to these artificial lab environments and the interpretation of lab based results should always be validated with *in situ* based observations.

The true biological function of the calcium binding protein, GPA, which CMM is thought to influence (Schroeder *et al.*, 2005), remains to be resolved. Recent studies have shown that GPA is most likely not directly involved in the production of coccoliths in *E. huxleyi* (Mackinder *et al.*, 2011; Rokitta *et al.*, 2011) but there is evidence to suggest GPA binds Ca^{2+} (Corstjens *et al.*, 1998). The link between CMM and morphotypes observed in this study is clear (i.e., one in one and a half billion chance of all six CMM II's being randomly associated with morphotypes other than the dominant A morphotype). Interestingly, the plastid gene *tufA* (Cook *et al.* 2011) supports the division of *E. huxleyi* into two main subgroups or varieties (Cook *et al.*

2013), while the mitochondrial (mtDNA) *cox1b-ATP4* genes (Hagino *et al.*, 2011) found that no genetic distinction could be made. The most parsimonious explanation for this apparent discrepancy is that the chromosomal (CMM) and plastid (*tufA*) alleles are under different selection pressure, possibly as a function of their individual attributes to fitness, while the mtDNA genes provide an insight into the ancestral history of this species through their maternal line. Such discrepancies between mtDNA and chromosomal phylogenies are well documented in animal systems. For example, apparent discrepancies exist between the distributions of the lineages of mtDNA and of the two major Y-chromosome lineages in mice (Boissinot and Boursot, 1997). Some subspecies share the same mtDNA lineage but have different chromosome lineages or vice versa (Boissinot and Boursot, 1997). Partitioning *E. huxleyi* into different CMM subgroups certainly has its place in population genetics as it appears to be more informative than when using microsatellites in isolation.

2.5.3 Implications for future research in microalgal population genetics

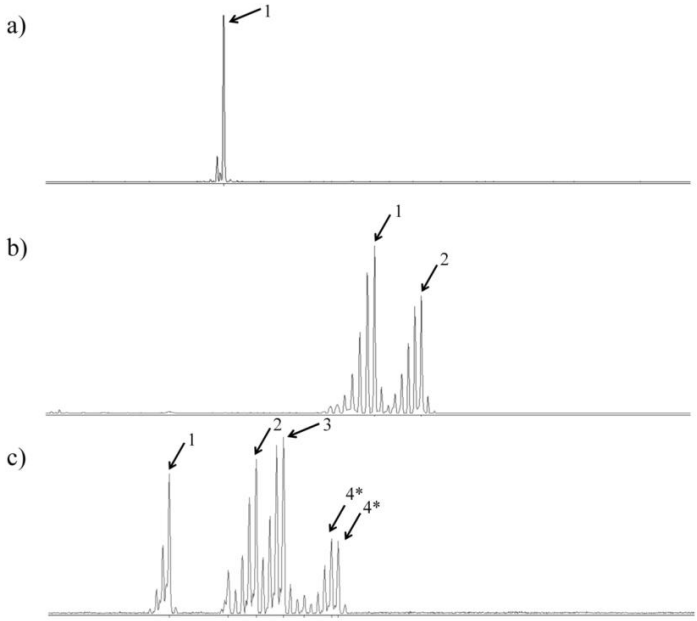
The bloom population in *E. huxleyi* appears to be relatively stable over consecutive blooms in a similar location, as also documented in *Ditylum brightwellii* (Rynearson and Armbrust, 2005). Martínez *et al.* (2007) demonstrated a stable inter-annual population using CMM genotypes using environmental DNA. However, this has been a limiting step as microsatellites necessitate clonal cultures or individuals. Preliminary data suggest certain allelic combinations are found in different years in the North Atlantic (*unpublished data*). Yet, this raises a critical point. As microalgae inhabit such a stochastic environment that changes rapidly, how should genotypes be scored? As gradations of allele frequencies, or distinctive diagnostic genotypes?

Schuller *et al.* (2012) demonstrated genetic difference in *Saccharomyces cerevisiae* were due to fine-scale allelic changes rather than diagnostic genotypes (i.e., very different allele sizes). The authors cautioned that though microsatellites are useful for population-level analyses, sub-strain level discrimination may occur due to their relatively high mutation rates. In this study, there was noise around the dominant allele of several base pairs, suggesting these alleles were recent mutations from the dominant (i.e., 100 and 104 alleles surrounding the 102 allele in P02F11, Table 2.4). Therefore, it might be necessary to treat microalgae in a similar manner to yeast. Does this represent something biological or is it simply noise? Are other bloom events in other basins dominated by the same or different alleles? Applying the techniques used in this study will enable us to respond to these questions and in so doing begin to describe the genetic structure of *E. huxleyi* in more detail. This is a critical step for further exploring host-viral dynamics (e.g., Martínez *et al.*, 2007), the occurrence of meta-population dynamics (Rynearson *et al.*, 2009), associated levels of genetic diversity (Walser and Haag, 2012) and understanding how this species will respond to climatic change or ocean acidification. High standing genetic variation and the fact that bloom events do not appear to cause a genetic bottleneck indicate that phytoplankton populations have the potential to adapt fast enough to keep pace with ongoing climate change. *E. huxleyi* is a relatively new species, having only appeared less than 300,000 years ago (Raffi *et al.*, 2006). Therefore, it will be interesting to explore the population genetics of this species in more detail in order to determine how this species has and is evolving.

2.6 Acknowledgements

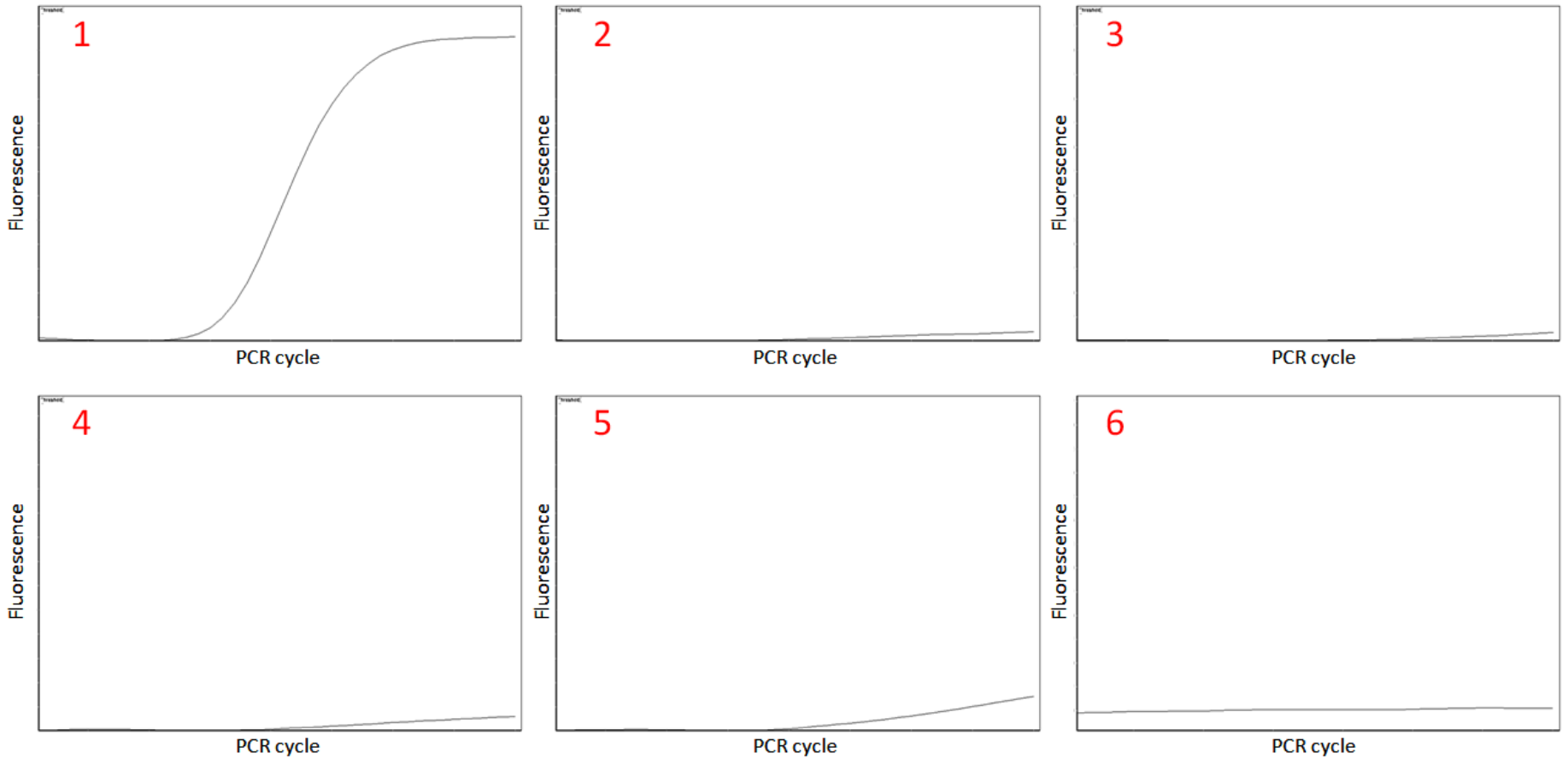
Special thanks go to Sue Cook, Bente Edvardsen, Ian Probert and Kyoko Hagino for either supplying us with DNA or live cultures for the biogeographic comparison. Thanks also go to Stephen Cotterell, Matt Hall and Gideon Mordecai for the technical advice and assistance; Mairi Knight for use of the capillary sequencer at Plymouth University. This project has been supported by Interreg IV Marinexus project (Ref: 1956/4073) and the UK Ocean Acidification programme.

2.7 Supplement

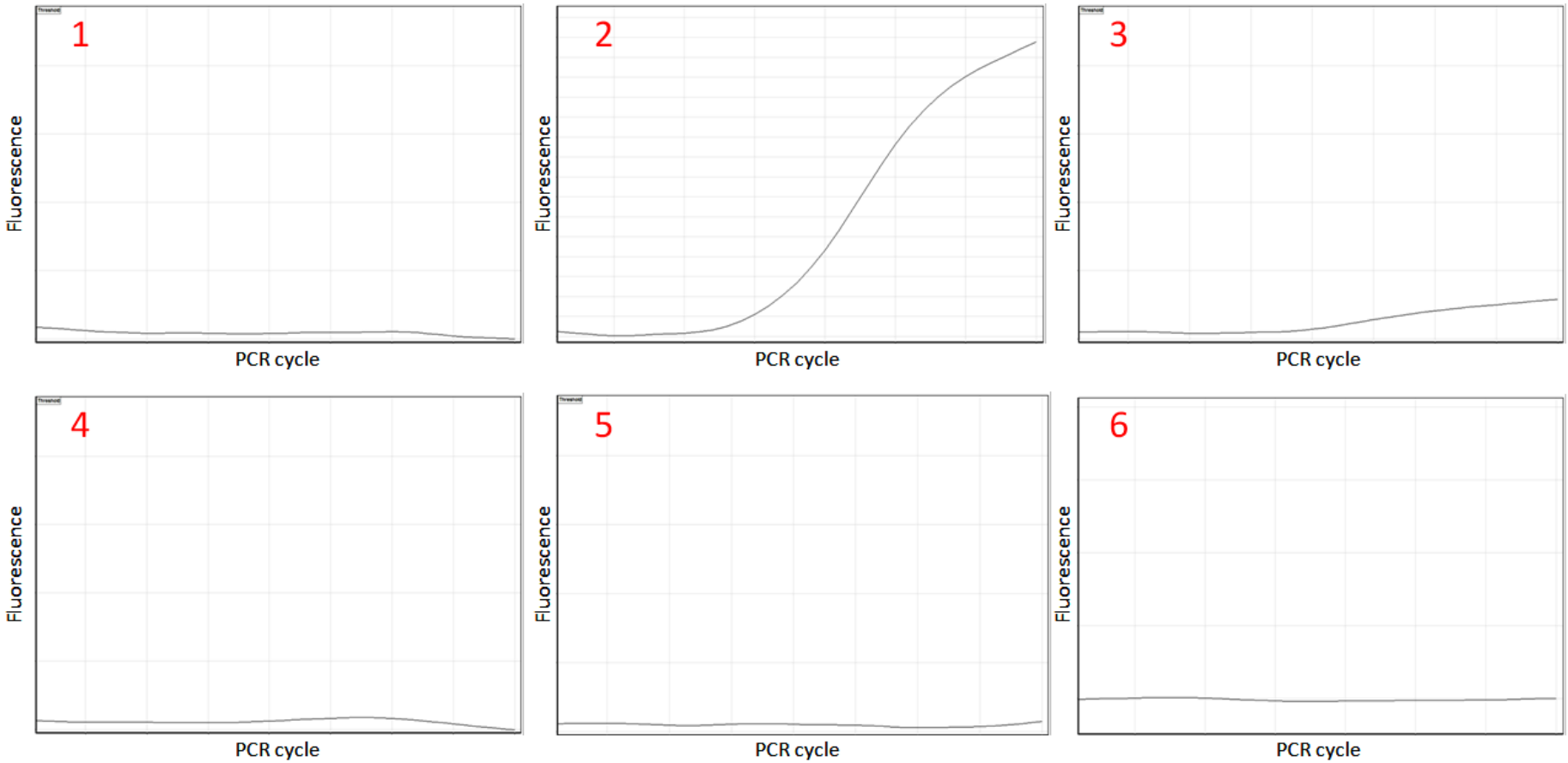


Supplementary Figure 2.1 The traces from two assumed Mendelian markers with a single BLAST hit against the *Emiliania huxleyi* CCMP 1516 genome sequence (Read et al. 2013) and one non-Mendelian marker. Each arrow indicates a putative microsatellite allele. a) Locus P02F11 did not exhibit stutter peaks occasionally associated with microsatellite trace profiles, but produced reliable single peaks. This trace demonstrates a homozygous strain at P02F11 (i.e., allele 1). b) Locus P01E05 exhibited stutter bands. However, it was possible to score the largest and furthest right peak as the locus profile was reproducible within and among strains (i.e., the same number of stutter peaks with the largest peak. This trace demonstrates a heterozygous strain at P01E05 (i.e., allele 1 and allele 2). c) Locus P02E10 produced multiple peaks for the tested strains which were not reproducible between different PCRs of the same strain. This locus was found five times in the CCMP 1516 genome. Moreover, it was impossible to score the alleles as exhibited in this trace where there were four possible alleles. The fourth allele (indicated with an asterix) does not follow the typical stutter motif pattern as the largest peak is not the furthest right. This problem was also found in different PCRs where the peak heights within the stutter motif were altered in which the largest peak could be the first, second or third. The remaining loci, EHMS15 and P02E11, exhibited similar types of patterns as demonstrated at P02E10.

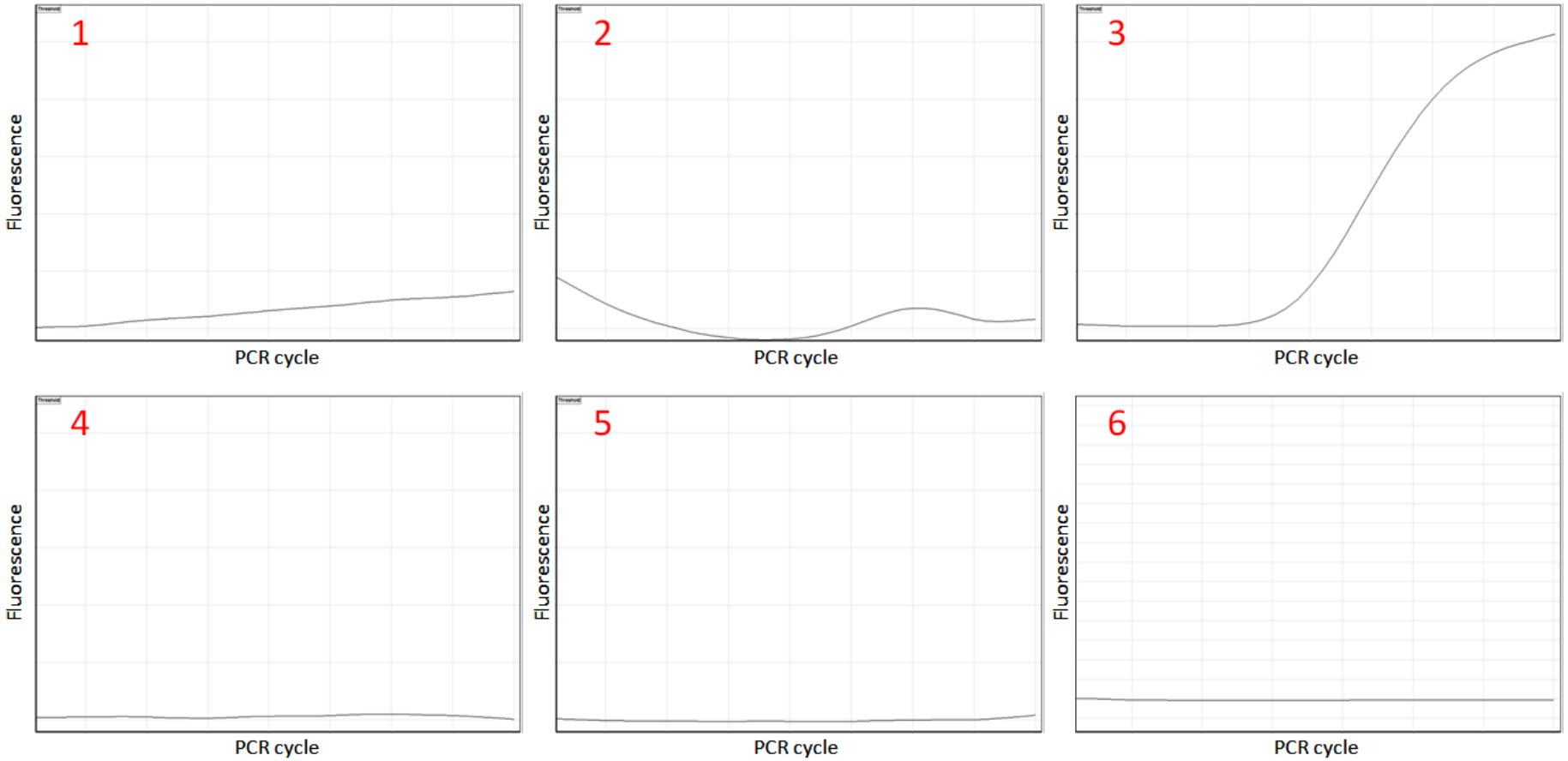
a)


1
Fluorescence
PCR cycle
2
Fluorescence
PCR cycle
3
Fluorescence
PCR cycle
4
Fluorescence
PCR cycle
5
Fluorescence
PCR cycle
6
Fluorescence
PCR cycle

b)

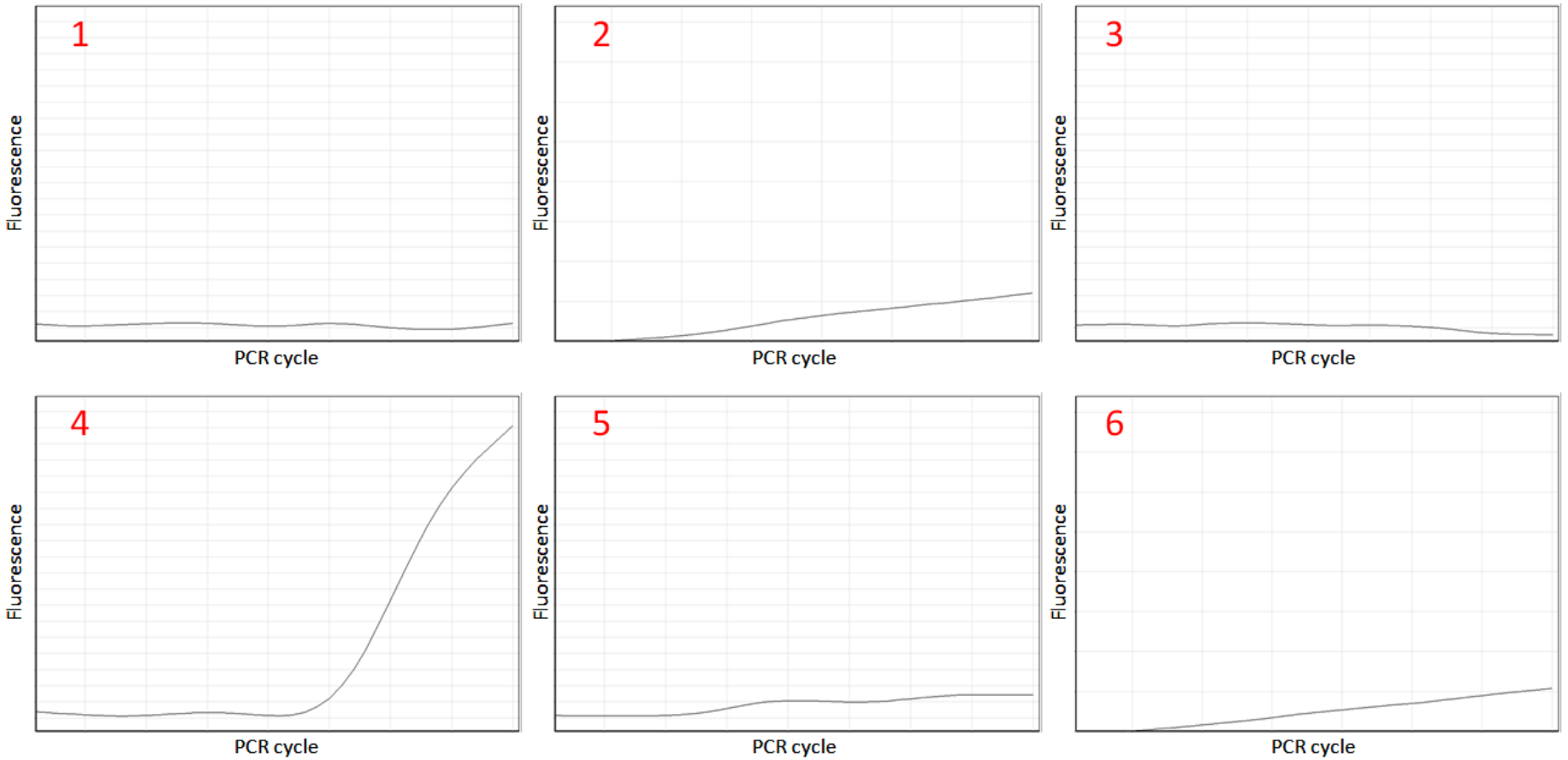


c)

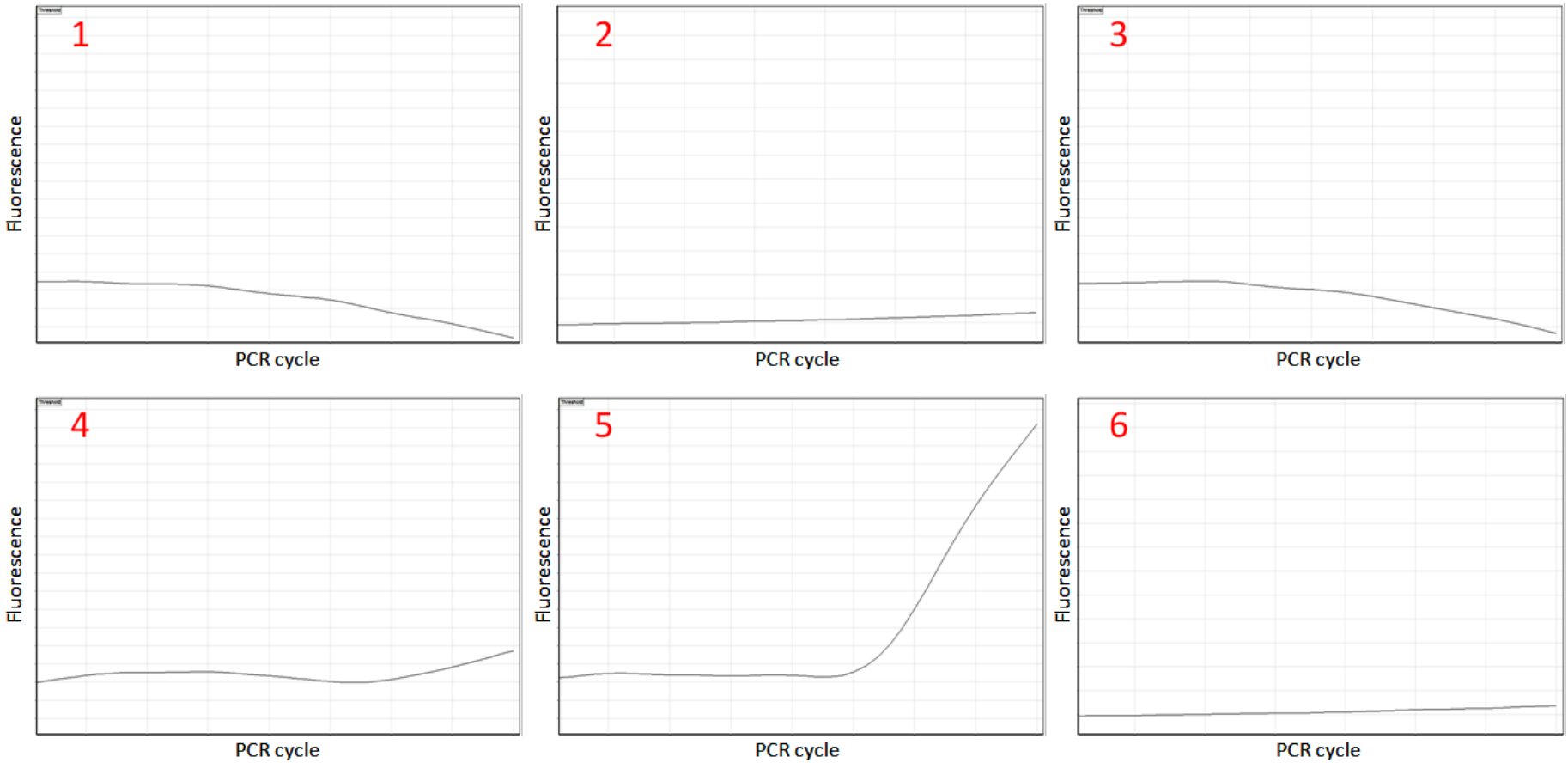


Supplementary Figure 2.2 The following panels show the probes-specificity for their respective sequences for Multiplex 1: a) probe I (6-FAM –green channel), b) probe II (ATTO680 – crimson channel) and c) probe III (CY5- red channel) (Table 2.3). Red numbers 1-6) CMM I, II, III, IV, IVb and IIb template DNA, respectively.

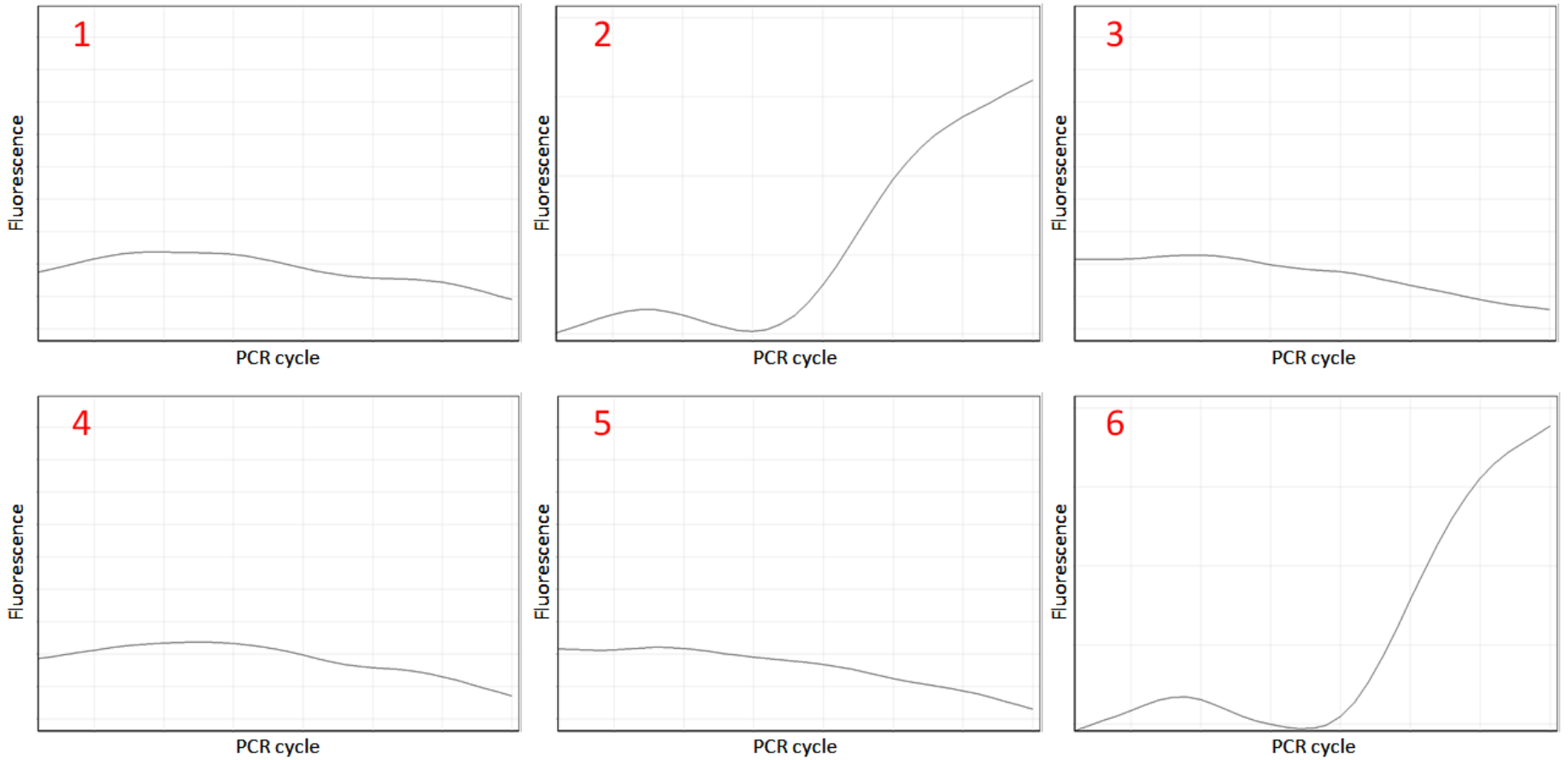
a)


1
Fluorescence
PCR cycle
2
Fluorescence
PCR cycle
3
Fluorescence
PCR cycle
4
Fluorescence
PCR cycle
5
Fluorescence
PCR cycle
6
Fluorescence
PCR cycle

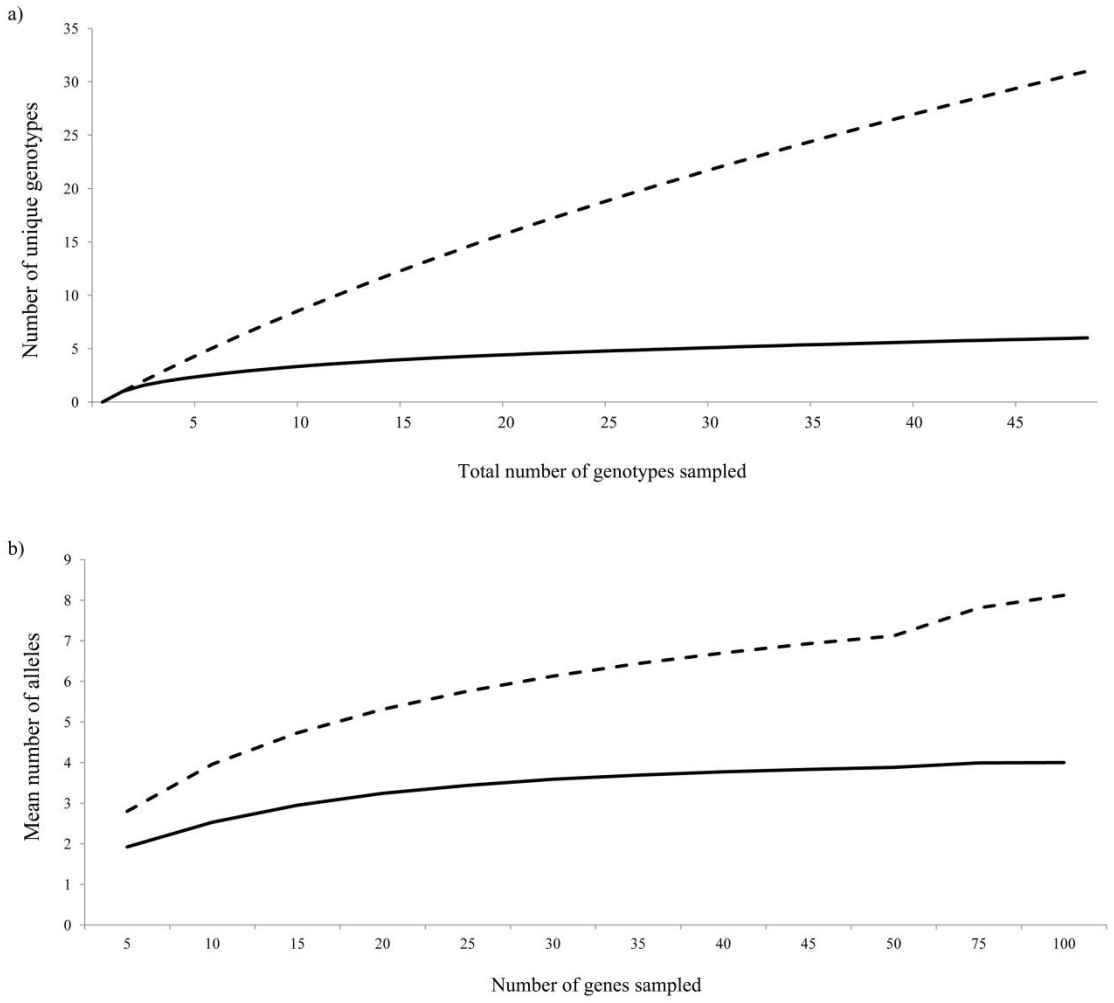
b)


1
Fluorescence
PCR cycle
2
Fluorescence
PCR cycle
3
Fluorescence
PCR cycle
4
Fluorescence
PCR cycle
5
Fluorescence
PCR cycle
6
Fluorescence
PCR cycle

c)



Supplementary Figure 2.3 The following panels show the probes-specificity for their respective sequences for Multiplex 2: a) probe IV (ATTO680 – crimson channel), probe IVb (6-FAM –green channel) and c) probe IIb (HEX–yellow channel) (Table 2.3). Red numbers 1-6) CMM I, II, III, IV, IVb and IIb template DNA, respectively.



Supplementary Figure 2.4 a) Genotype rarefaction curve for CMM genotype (solid line) and microsatellites (dashed line) generated using FASTGROUPII (Yu et al., 2006). The curve reaches a plateau for CMM genotypes whereas, there appeared to be microsatellite MLGs which were not sampled. b) The variation in the mean number of alleles (averaged over 5 microsatellite loci) observed with different numbers of genes sampled using HP-RARE (Kalinowski 2005) indicated that after sampling at least 75 genes (or at least 38 diploid isolates), the chances of encountering a rare allele decreased.

CHAPTER 3

Natural selection in the Western English Channel: an increase in regional sea surface temperature is implicated in the differential survival of *Emiliania huxleyi* bloom forming morphotypes

*Cecilia Balestreri[1,2], Anthony J. Richardson[3,4], Rosalind E. M. Rickaby[2], Tim J. Smyth[5], Colin Brownlee[1] and Declan C. Schroeder[1]**

[1] The Marine Biological Association of the United Kingdom, The Laboratory, Citadel Hill, Plymouth PL1 2PB, U.K.

[2] Department of Earth Sciences, Oxford University, South Parks Road, Oxford OX1 3AN, UK

[3] Ocean and Atmosphere Flagship, CSIRO Marine and Atmospheric Research, Ecosciences Precinct, Dutton Park, Queensland 4102, Australia

[4] Centre for Applications in Natural Resource Mathematics (CARM), School of Mathematics and Physics, University of Queensland, St Lucia, Qld 4072, Australia

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

* Corresponding author: dsch@mba.ac.uk

Authors' contribution

Cecilia Balestreri wrote the manuscript, which was edited by Declan Schroeder and Anthony Richardson. The samples were processed and the DNA was extracted from each sample by Cecilia Balestreri, who also carried out the probe assay analysis. Tim Smyth provided the net heat flux (NHF) data. Project design, revision and submission of the manuscript were provided by Declan Schroeder. Ros Rickaby and Colin Brownlee contributed to the mentoring of Cecilia and in the revision and editing of the final draft before submission.

3.1 Abstract

The coccolithophore *Emiliania huxleyi* regularly forms extensive white-water blooms that extend over thousands of square kilometres and may persist for many months in high latitude coastal and shelf ecosystems. Coccolithophores take advantage of the seawater carbonate chemistry to create insoluble carbonate structures known as coccoliths. A variety of *E. huxleyi* genotypes corresponding to either of two main coccolith morphological types, morphotype A or B, have been reported. By using a multiplex probe assay, we monitored the *in situ E. huxleyi* intraspecific genetic diversity at the L4 long-term monitoring site within the Western Channel Observatory over a period of 6 years. We report here that only 2 out of the 6 known CMM genotypes, CMM I & CMM IV, persist and dominate throughout the year; in both bloom and non-blooming periods. Both genotypes correspond to the classic morphotype A phenotype, which have in the recent past been reported to dominate North-eastern Atlantic, Western English Channel (WEC) and North Sea *E. huxleyi* bloom events. One specific genotype linked to morphotype B, that previously formed blooms in the WEC and North Sea, was only found during the winter non-blooming months in the WEC in our 6-year dataset. This recent disappearance of morphotype B forming blooms coincides with the well-documented rise since the 1980s of sea surface temperatures in this region.

3.2 Introduction

About half the anthropogenic CO_2 produced from burning of fossil fuel released between 1800 and 1994 has been absorbed by the oceans (Sabine *et al.*, 2004). As CO_2 reacts with seawater, it causes a decrease in pH (ocean acidification) and carbonate ions and an increase in bicarbonate ions (Raven *et al.*, 2005). Moreover, compounding stressors include an increase in global sea surface temperatures by 0.6 ± 0.2 °C over the 20th century (Rayner *et al.*, 2003). The complex and rapid changes in ocean physical and chemical properties raise concerns about the impact on key biological species over such short timescales. The nature of their responses will have important biological, ecological, biogeochemical and societal implications (Turley *et al.*, 2010).

The coccolithophore *Emiliania huxleyi* (Prymnesiophyceae) is one of the main calcium carbonate producers on Earth (Westbroek *et al.*, 1993) and is present in all but extreme polar oceans. In the right conditions, it regularly forms blooms which extend over thousands of square kilometres and may persist for several months (Holligan *et al.*, 1993). Coccolithophores contribute in the reduction of atmospheric CO_2, as sinking coccoliths, especially at the final stages of the bloom, transport large amounts of associated organic carbon to the deep sea enhancing the oceanic sink for carbon (Coxall *et al.*, 2005; Riebesell *et al.*, 2009; Riebesell and Tortell, 2011; Robertson *et al.*, 1994). Moreover, anthropogenic CO_2 has the potential to reduce calcification rates in various morphotypes of *E. huxleyi* (Riebesell *et al.*, 2000; Müller *et al.*, 2015).

E. huxleyi is characterised by two major morphotype groups: type A and B. Morphotype B is distinctly less calcified (Young *et al.*, 2014); and has a higher calcite carbon to organic carbon ratio than morphotype A (van Bleijswijk *et al.*, 1994). Variation occurs within each group as evidenced by both morphology (Young and Westbroek, 1991; Young *et al.*, 2014) and genetics (Schroeder *et al.*, 2005; Krueger-Hadfield *et al.*, 2014). A dual labelled probe assay technique allowed the genetic characterisation of numerous *E. huxleyi* samples for the coccolith morphology motif (CMM), discovered by Schroeder and his colleague (2005), and determined that CMM I and CMM IV, associated with morphotype A, are the most prevalent in both the Northern and Southern hemispheres, while CMM II, associated with morphotype B, is more rare (Krueger-Hadfield *et al.*, 2014). This latter observation contradicts an older study which described the dominance of morphotype B in the North Sea (van Bleijswijk *et al.*, 1991).

Krueger-Hadfield *et al.* (2014) also genotyped a bloom event that occurred in late June 2011 in the North Sea; revealing that the CMM I and CMM IV dominated the bloom and that asexual reproduction drove its formation. Moreover, in two separate mesocosm studies in Raunefjorden in western Norway (Marine Biological Field Station), conducted in June 2000 and June 2003, Martínez *et al.*, (2007) monitored the progression of independent coccolithophore bloom events and found that CMM I dominated each bloom. A subsequent study conducted at the same station in June 2008 by Sorensen *et al.* (2009) showed that *E. huxleyi* genetic composition during the bloom was still dominated by CMM I. Further studies of a bloom event in the North Sea in June 1999 (Martínez *et al.*, 2012) revealed the presence of five distinct *E. huxleyi* genotypes by DGGE analysis. Although a high genetic variability was detected

in the water column, the bloom chlorophyll maximum was characterised only by CMM IV (Martínez *et al.*, 2012). Highfield et al. (2014) showed that an *E. huxleyi* bloom event in late summer 2006, was dominated by CMM IV. The CMM IV genotype appears to dominate in a wide range of environmental niches, from cold to warm and often changeable environments (Krueger-Hadfield *et al.*, 2014).

All the aforementioned studies represent snapshots of diversity in a restricted time frame and the source of the diversity that underpins bloom development remains an open question. For example, does an extensive genetic standing stock exist from which a subset of 'fit' genotypes are selected that go on to form a bloom? Time-series data offer a great resource to assess the continuous genetic composition of a species over time. The Western English Channel (WCO, 50°15'N 4°13'E; www.westernchannelobservatory.org.uk; see Appendix I) has been studied intensively for more than 100 years (Southward *et al.*, 2005) and this abundance of data provides a robust context with which to explore temporal microbiological complexity in the calcifying haptophytes.

Several molecular studies on the L4 station of the WCO have successfully analysed the prokaryotic community structure, detecting strong seasonal patterns (Caporaso *et al.*, 2012; Gilbert *et al.*, 2009, 2012). To date, no study has described the molecular eukaryotic diversity at this station. Nonetheless, a substantial inter-annual variability in the composition of the phytoplankton community has been routinely observed at L4 through the use of traditional microscopy techniques. During the summer months, thermal stratification and increased irradiance coincided with increased coccolithophorid abundance, with *E. huxleyi* accounting for ca. 90% of the species composition (Widdicombe *et al.*, 2010). The seasonal patterns observed at the

WCO are consistent with phytoplankton succession during the year defined by turbulence and nutrients (Margalef, 1978). A high abundance of diatoms that flourish in well-mixed and high-nutrient waters, occurs during the spring followed by coccolithophores and finally by dinoflagellates, which flourish in well-stratified and low-nutrient waters, at the end of the summer/beginning of autumn (Margalef, 1978; Balch, 2004).

The switch in species dominance can also be correlated with net heat flux (NHF) at the L4 station (Smyth *et al.*, 2014). The NHF was calculated using modelled meteorological conditions and measured sea-surface temperature. The NHF takes into account five different parameters: air temperature, dew point, wind speed, cloud fraction and atmospheric pressure, and it was used as a critical indicator of ecosystem dynamics. A tight correlation has been shown between phytoplankton abundance and diversity patterns and the switches between negative and positive NHF over an annual cycle (Smyth *et al.*, 2014). The switch to positive NHF at L4 coincided with the initial dominance of phyto-flagellates and diatom species, while the switch to negative NHF toward the autumn proved to be an important cut-off for coccolithophores and dinoflagellates. Therefore, the *E. huxleyi* population at the WCO requires well-stratified waters, low-nutrient concentration, moderate density and day length approaching 16h for bloom development. Independent studies at other locations, i.e. during the summer of 1992 in the Equatorial Pacific (Balch and Kilpatrick, 1996), in the Bering Sea during 1990 (Napp and Hunt, 2001) and in the Gulf of Maine during June 2000 (Balch *et al.*, 2004), confirm these requirements. Here we used NHF to constrain the periods when *E. huxleyi* is predicted to bloom; thereby allowing us to determine the composition, by CMM profiling, of the *E. huxleyi*

community in monthly samples collected at the L4 station of the WCO, over a 6 year period through bloom and non-bloom months. We show for the first time genotypes that co-occur in non-bloom populations in the Western English Channel. However, only selected genotypes go on to develop into bloom populations.

3.3 Materials and methods

3.3.1 Sample collection and oceanographic data

Surface seawater samples were collected from the L4 station (www.westernchannelobservatory.org.uk; see Appendix I) from January 2008 to June 2013. Five litre volumes of seawater were filtered through a 0.22 µm Sterivex cartridge (Millipore) and stored at -80 °C (Gilbert *et al.*, 2009). Corresponding NHF data was collated as described by Smyth *et al.* (2014). All the cell count, the chlorophyll-a and the temperature data relative to the L4 station were found online at www.westernchannelobservatory.org.uk.

3.3.2 DNA extraction and genetic analysis

Sterivex filters were selected for DNA extraction as described in Neufeld *et al.*, (2007). The only modification to the protocol was that the Sterivexes were first broken and the filter-part removed so as to improve DNA yield efficiencies. To confirm the quality of the DNA, PCR amplification of the V9 eukaryote SSU rRNA gene region (Dunthorn *et al.*, 2012) was obtained using 1 µl of the environmental DNA added to 5 µl 5X Colourless GoTaq Flexi Buffer (Promega), 1.5 µl MgCl2 Solution 25 mM (Promega), 2.5 µl PCR Nucleotide Mix 10mM each (Promega), 1 µl Evagreen Dye

20X (Biotium), 0.1 µl GoTaq DNA Polymerase (5u/µl - Promega) up to a final volume of 25 µl for each reaction. The PCR proceeded with an initial denaturation at 94 °C for 3 minutes, followed by 40 cycles of a three step PCR: 94 °C for 45 seconds, 50 °C for 60 seconds and 72 °C for 90 seconds. The cycle was repeated 35 times.

Amplification of the CMM was carried out using the multiplex probe assay technique as described in Krueger-Hadfield *et al.*, (2014) to detect the presence of certain genotypes in each L4 sample. Each CMM tested was cloned into plasmidic vectors (pCB1-6) using TA Cloning® Kit, with pCR™2.1 Vector and One Shot® TOP10 Chemically Competent E. coli (Invitrogen, Life technologies, Paisley - UK). These were used as a positive control during each qPCR run (Corbett Life Science – Rotor Gene 6000). The positive controls were also used as an indicator of the amplification quality and quantity after setting fluorescence thresholds against standard curves (Supplementary Figure 3.1) and applying the 'dynamic tube correction' and the 'noise slope correction' according to the manufacturer's recommendations (Rotor-Gene 6000, Operator Manual, 2006). Additionally, they were used to normalize the data within and between qPCR runs. Cycle threshold (CT) values were compared (when possible) for the samples and a fold-difference value (FD) was calculated to quantify the presence of each specific CMM (Supplementary Figure 3.3). Calculation of the standard deviation for each plasmid confirmed the robustness of the technique between different PCRs (Supplementary Figure 3.2).

3.3.3 Estimation of genotypic diversity

Genotype diversity during sampling was calculated by simply using the number of coccolithophore genotypes (CMMs) recorded at each sampling event.

To assess whether particular genotypes were found preferentially during predicted bloom or non-bloom conditions, we performed a number of randomization tests. Given the number of times a particular genotype was observed at each time point during the 6 year study, we calculated the likelihood of finding a particular genotype the number of times it was actually observed during the predicted *E. huxleyi* bloom period. Specifically, these tests were based on the total (T) number of sampling times (nT=51), the number of samples collected during the predicted *E. huxleyi* bloom (B) period (nB=8); and for each genotype the total number of times it was found (nT,i, where i is the genotype) and the number of times it was sampled during predicted bloom conditions (nB,i). The predicted bloom times were set to be 8 days as predicted by Smyth *et al* 2014. We thus randomly selected nT,i (say 7 for i=CMM IVb) numbers from nT (51) random numbers without replacement – with the sequence of numbers representing a series of random sampling times for genotype i. We ran 1 million iterations for each genotype. We then summed the number of times that genotype i was found during the predicted bloom times (nhatB,i). For each random realization, this sum varied from no occurrences of the genotype in the predicted bloom times (0) to the genotype being found during all possible bloom times (8). The randomization tests thus represent the likelihood of obtaining nB,i or more occurrences of the genotype during the predicted bloom time given the distribution of nhatB,i.

3.4 Results

PCR amplification of the eukaryotic V9 SSU rRNA region was achieved for all 51 samples collected (data not shown), while *E. huxleyi* CMM amplification occurred

in 46 out of the 51. No CMM amplification occurred in three consecutive months from April to June in 2012 (Figure 3.1). The other two negative results occurred at a similar time of the year, i.e. July 2008 and May 2013.

Using the NHF criteria as defined in Smyth et al 2014, we were able to positively assign the sampling dates collected for the months of July and August to fall within the predicted *E. huxleyi* bloom period (Figure 3.1). Temperature data relative to the samples support the trend described by the NHF, while chlorophyll-a data and cell count data show some anomalies (Figure 3.1). Two CMMs, namely CMM I and CMM IV, were most prevalent over the six year period and both were present in 7 out of the 8 bloom months (Figure 3.1). For those 46 samples, where *E. huxleyi* was detected, only two months during the spring of 2008 showed an absence of CMM I. Similarly, CMM IV was not amplified for only 5 months, three of which were in 2008. CMM II amplification occurred less frequently and was absent in 2012 and 2013. Moreover, CMM II was not detected in the predicted bloom months, while the months of April, May and June gave consistent positive amplification for CMM II in 2008, 2009 and 2011. Year 2010 was unusual as CMM II appeared later on in the year, i.e. September to December 2010; however, samples were not available from March to August of this year. In addition, the occurrence of CMM II extended into early January 2011. CMM IVb was occasionally amplified, and was found in 7 out of the 46 samples and in 2 out of the 8 bloom months. Two CMMs, namely CMM III and CMM IIb, were not detected in this time series.

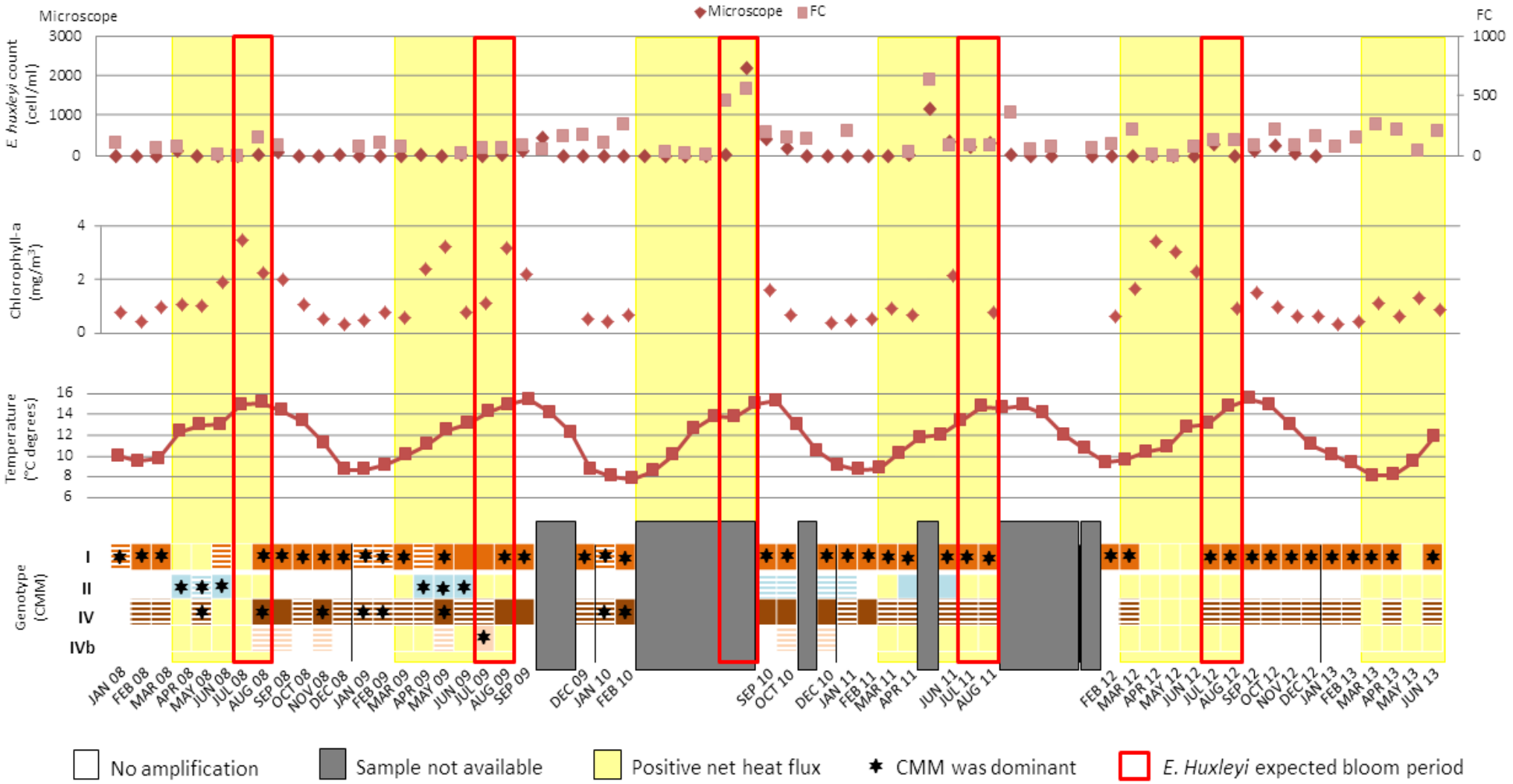

Microscope
FC
E. huxleyi count (cell/ml)
3000
2000
1000
0
1000
500
0
Chlorophyll-a (mg/m³)
4
2
0
Temperature (°C degrees)
16
14
12
10
8
6
Genotype (CMM)
I
II
IV
IVb
JAN 08
FEB 08
MAR 08
APR 08
MAY 08
JUN 08
JUL 08
AUG 08
SEP 08
OCT 08
NOV 08
DEC 08
JAN 09
FEB 09
MAR 09
APR 09
MAY 09
JUN 09
JUL 09
AUG 09
SEP 09
DEC 09
JAN 10
FEB 10
SEP 10
OCT 10
DEC 10
JAN 11
FEB 11
MAR 11
APR 11
JUN 11
JUL 11
AUG 11
FEB 12
MAR 12
APR 12
MAY 12
JUN 12
JUL 12
AUG 12
SEP 12
OCT 12
NOV 12
DEC 12
JAN 13
FEB 13
MAR 13
APR 13
MAY 13
JUN 13
No amplification
Sample not available
Positive net heat flux
CMM was dominant
E. Huxleyi expected bloom period

Figure 3.1 (previous page). Presence/absence of specific CMMs and positive net heat flux (NHF). 51 samples selected over a six year period at the L4 station in relation to the positive NHF for each year (highlighted background) are compared with the coccolithophores cell count, the chlorophyll-a (chl-a) content and the temperature (T) of the surface seawater over the same time. The presence of a coloured square, on the genotype plot, indicates a great amplification of a specific genotype, related to the corresponding CMM. A hatched square indicates the detection of a certain CMM, although the signal was too weak for an accurate quantification. The star on a square indicates the dominance of that CMM. Some months of this time series are missing, because there were no samples available.

CMM I was not only the most prevalent but also dominated the population in the presence of the other CMMs. When the amplification of two or more CMMs in the same run was quantifiable, CMM I resulted dominant in 9 samples out of 15 (FD=7.7 for September 2008; FD=5.2 for August 2009; FD=14.2 for September 2009; FD=60.5 for September 2010; FD=5.1 for October 2010; FD=11 for December 2010; FD=5.9 for February 2011; FD=23.3 for April 2011; FD=5.4 for June 2011). Only in two cases, June and July 2009, was CMM I not the dominant CMM. During these periods, CMMs II and IVb were dominant respectively (FD=7 for CMM II and FD=104.5 for CMM IVb). In four cases we were not able to establish a clear dominance of a single CMM (August and November 2008, May 2009 and February 2010; Supplementary figure 3.3).

There was no significant difference between the coccolithophore genotypic diversity during *E. huxleyi* bloom (2.00±0.33) and non-bloom (2.05±0.15) conditions. Based on the randomization tests, the probability that CMM I was found 7 or more times during the predicted bloom times (nB, CMM I=7) given that it was found in 44 samples (nT, CMM I=44) is p=0.700 (Figure 3.2a). Given how common this genotype

is, there is a 70% chance of it being found 7 or more times during the predicted bloom times. Similarly for CMM IV (Figure 3.2c), the probability of this genotype being found 7 or more times during the predicted bloom times (nB, CMM IV=7) given that it was found in 41 samples (nT, CMM IV=41) is p=0.503, suggesting that the prevalence and dominance of both CMM I and CMM IV throughout the year is a good predictor for which genotype(s) will go on to bloom in the WEC.

For CMM II (Figure 3.2b), the probability of it never being found during the predicted bloom times (nB, CMM II=0) given that it was found in 12 samples (nT, CMM II=12) is simply the value at zero of p=0.097. This is marginally significant ($p<0.1$), implying that this genotype is rarely found during bloom times as experienced in the WEC. For CMM IVb (Figure 3.2d), the probability of this genotype being found 2 or more times during the predicted bloom times (nB, CMM IVb=2) given that it was found in 7 samples (nT, CMM IVb=7) was 0.300, suggesting that this genotype could potentially bloom but is certainly not the dominant genotype.

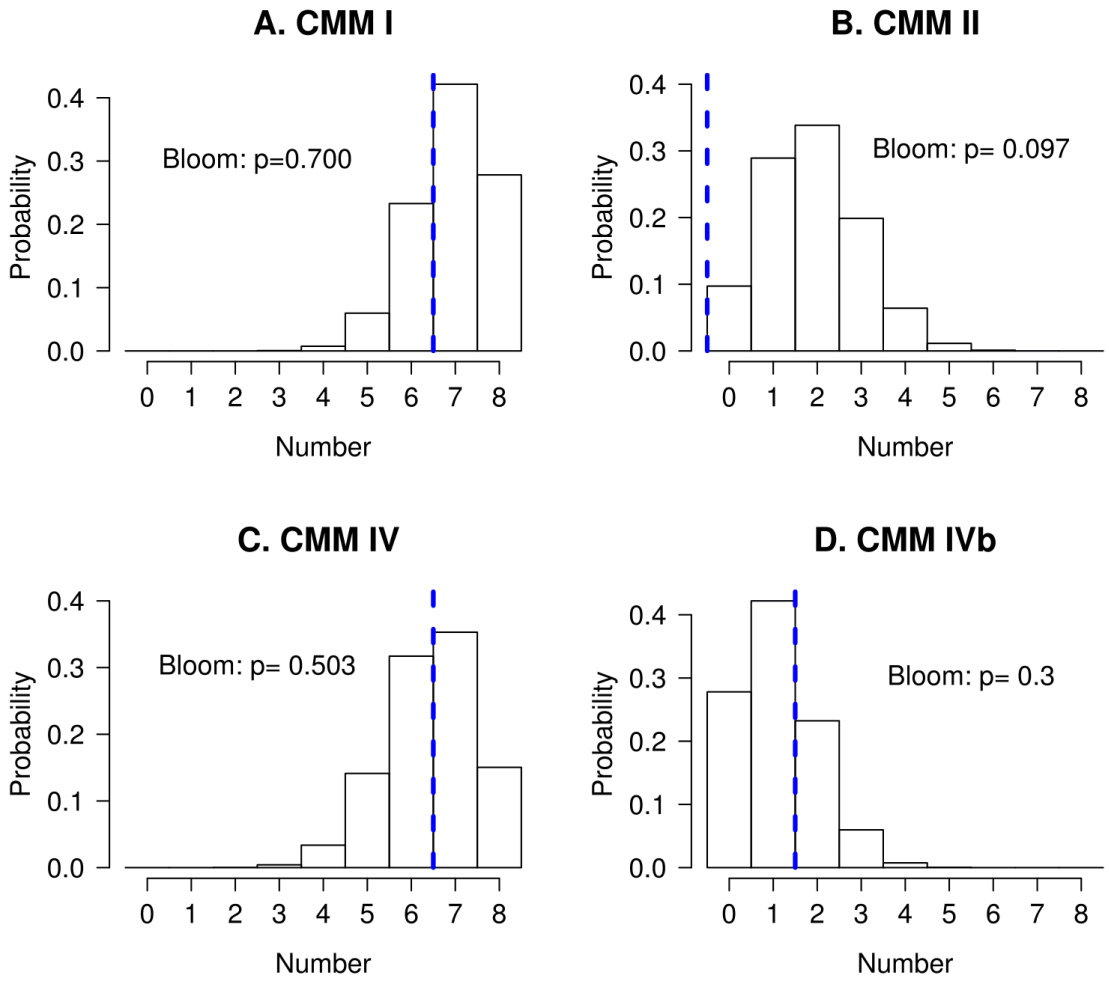


Figure 3.2 Histogram of results from randomization tests (n=1,000,000) showing the probability of each genotype being found a certain number of times during the predicted bloom period. Dashed blue line shows the observed number of times each genotype was found during the bloom. P gives the probability of the observed number of occurrences in the bloom.

3.5 Discussion

Phytoplankton succession in the WEC follows the Margalef's seasonal pattern where the spring bloom starts with a positive NHF as defined by an increase in light availability, an increase in sea surface temperature (SST) and an inhibition of vertical mixing (stratification), which cause the phytoplankton and the nutrients to be maintained at the surface (Margalef, 1978; Mann and Lazier, 2005; Miller, 2004; Widdicombe *et al.*, 2010). *E. huxleyi* in this period is found in concurrence with small phyto-flagellates (between 2 and 4 mm in size) and diatoms; however it is the diatoms which regularly dominate the spring period in the WEC. Between March and May diatoms such as *Chaetoceros* and *Thalassiosira,* are known to be the dominant species (Widdicombe *et al.*, 2010). Stratification lasts until September, with elevated chlorophyll in June, July and August, as showed in Figure 3.1, with the exception of summer 2010 and two anomaly peaks in spring 2009 and spring 2012. The temperature profile follows the NHF predicted period, with the positive NHF starting when the water gets warmer (Figure 3.1). *E. huxleyi* cell count generally does not fit the expected bloom period (Figure 3.1), but we have to take into account that the measurements used to produce the plot were taken at the sea surface, and therefore they may not be representative of the real depth at which the bloom was occurring. Breakdown in stratification and erosion of the thermocline, with consequent replenishment of nitrate to the surface layers, induces the switch to a negative NHF and determines the cut-off point for the *E. huxleyi* bloom (Smyth *et al.*, 2014).

The screening of environmental samples for a short genetic marker, such as the probes designed based on the CMM, could present some disadvantages, as it

would be possible that the marker is not unique for a certain species within a whole mixed community. When, at first, Schroeder and his colleagues (2005) discovered the correlation between the CMM sequence and different *E. huxleyi* morphotypes, they tested the oligonucleotide primers, which produced amplicons only for *E. huxleyi* and not for other Haptophyte species, such as *Gephyrocapsa oceanica, Coccolithus pelagicus*, and *Phaeocystis globosa*. *G. oceanica* and *E. huxleyi* have been proven belonging to the same species-complex (Bendif et al., 2014), therefore they are genetically very close. Nevertheless, the probes tested by Schroeder et al. (2005) and used in this study appear to be *E. huxleyi*-specific, although *E. huxleyi* genomic sequence is the only Haptophyte species fully characterised to date.

E. huxleyi morphotype A (specifically CMM I and CMM IV) clearly dominates the blooms in the WEC. This is consistent with the blooms documented in 2006 (Highfield *et al.*, 2014) and is typical of the blooms in the North Sea (Martínez *et al.*, 2012; Krueger-Hadfield *et al.*, 2014). When the NHF switches to become positive we found that the CMM II genotypes appear in the first two/three months: April, May and June of years 2008, 2009 and 2011. CMM II is positively correlated with the presence of *E. huxleyi* morphotype B (Schroeder *et al.*, 2005; Krueger-Hadfield *et al.*, 2014) and this morphotype was first isolated in a bloom event in the WEC in July 1975 (Young and Westbroek, 1991). In the 1990's, it was also observed in the North-East Atlantic region and in the North Sea (van Bleijswijk *et al.*, 1991), where it was responsible for a bloom event in the summer of 1990. In 1993 it was only found rarely during another bloom event in the same area (Van der Wal *et al.*, 1995). Therefore, morphotype B has been observed to form blooms both in the WEC and North Sea prior to the 2000's. Recently, we could not find the morphotype B in the North Sea

during the bloom event in the summer of 2011 (Krueger-Hadfield *et al.*, 2014) and this study showed no bloom event in the WEC dominated by B morphotype between 2008 and 2013. It has been well documented that the North-East Atlantic and UK coast SST have been rising at a similar rate to land air temperature. Since the 1980s the rate of rise has been about 0.2-0.6 ºC per decade (Rayner *et al.*, 2003; Figure 3.3). In the last 40 years, the warming SST in the Northern Hemisphere seems related with the disappearance of B morphotype bloom events in the Western English Channel and in the North Sea (Figure 3.3). Could morphotype B prefer cooler summer SSTs? If so, this would be consistent with our observation of its more recent presence in the cooler April to May months. An alternate explanation could be that since morphotype B is less calcified and larger than morphotype A (Young et al., 2014), it could be more susceptible to grazing by zooplankton (Fileman *et al.*, 2002). However, zooplankton abundance is low in early spring (Miller, 2004).

If we consider the annual trend of genotypes of *E. huxleyi* in the WEC, we found that CMM I and IV were generally present throughout the year, often being present outside the periods of seasonal stratification of the water column. In fact, the winter/early spring and autumn are characterised by well-mixed water columns (Smyth *et al.*, 2010). Therefore there is no requirement for a stable water column to permit *E. huxleyi* growth in this region. Balch *et al.* (2000) showed that in the Arabian Sea, the highest coccolithophorid abundance was during the highly productive inter-monsoon period (Balch *et al.*, 2000, 2001), suggesting an association between high numbers of coccolithophores and mixed overlying waters of high productivity (Balch, 2004). However, the majority of studies observing *E. huxleyi* natural blooms (and the associated measured water conditions) suggest that they tend to occur in highly

stratified water where the mixed layer depth is usually ~10-20 m, and is almost always ≤ 30 m (Nanninga and Tyrrell, 1996; Tyrrell and Taylor, 1996).

Morphotype A has been shown to dominate bloom events in the northern hemisphere, i.e. in the North-East Atlantic Ocean in June 1991 (Holligan *et al.*, 1993), in the WEC in June 1992 (Garcia-Soto and Fernandez, 1995), in the North Sea in July 1993 (Van der Wal *et al.*, 1995) and June 2011 (Krueger-Hadfield *et al.*, 2014), and in the Norwegian mesocosm experiment in 1992 (Young, 1994). The studies on coccolithophore bloom events in the southern hemisphere (Brown and Yoder, 1994; Brown and Podestá, 1997; Gayoso, 1995; Painter *et al.*, 2010) have not focused on the morphological or genotypic composition of the blooms so preventing any insight as to whether morphotype B is able to form bloom events in this hemisphere. Nevertheless, the B/C morphotype, related to the CMM IIb (Krueger-Hadfield *et al.*, 2014), is found in high abundance during the bloom event in December 2008 along the Patagonian Shelf and it has been reported that this morphotype is specifically a Southern Ocean ecotype (Poulton *et al.*, 2011). Indeed, this *E. huxleyi* form was found in many studies in the southern hemisphere (Cubillos *et al.*, 2007; Holligan *et al.*, 2010; Cook *et al.*, 2011; Müller *et al.*, 2015), and subpolar waters (Hagino *et al.*, 2005). Moreover, we could not detect the B/C morphotype (CMM IIb) in the WEC. This adds credence to the regional bias of certain CMM genotypes and supports the reporting of "ecotypes" for *E. huxleyi* (Müller *et al.*, 2015).

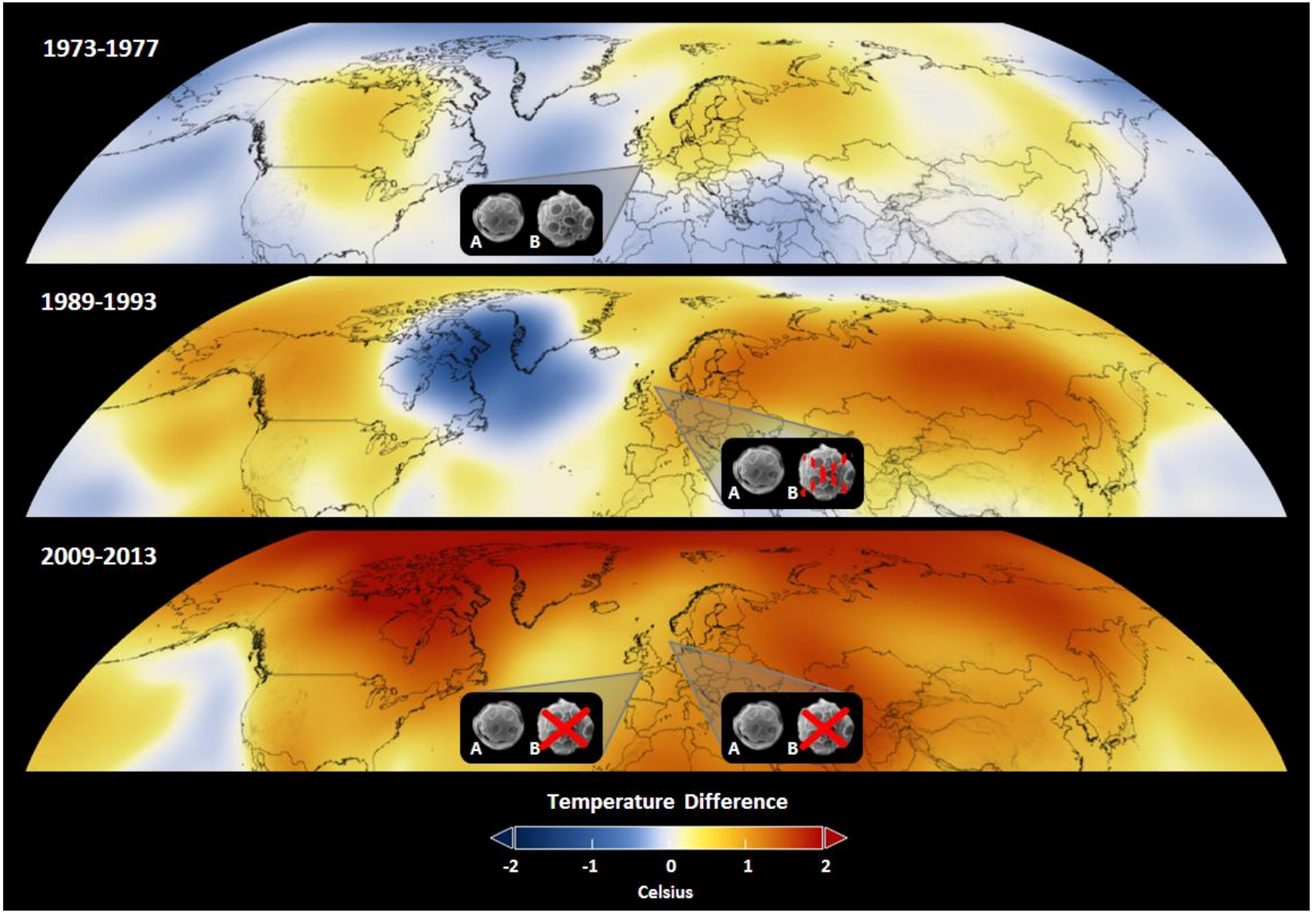
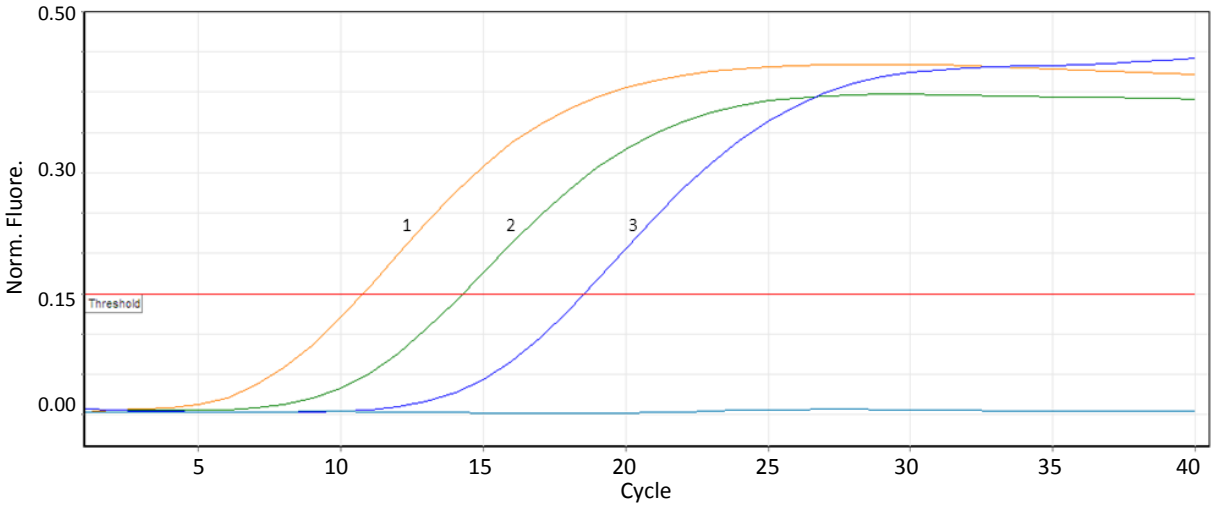

1973-1977
A
B
1989-1993
A
B
2009-2013
A
B
A
B
Temperature Difference
-2
-1
0
1
2
Celsius

Figure 3.3 (previous page). Presence of *E. huxleyi* morphotypes A and B in the last 40 years, as observed during bloom events in the Western English Channel and North Sea. The letters A and B next to *E. huxleyi* photos refer to the morphotype. The hatched red cross represents a presence followed by an absence of *E. huxleyi* in a certain area. Color-coded maps in Robinson projection display a progression of 5 year rolling average of changing global surface temperature anomalies. Three decade period, from 1951 to 1980, functioned as a baseline for the analysis of the temperature produced at NASA's GISS Surface Temperature Analysis (GISTEMP) using current data files from NOAA GHCN v3 (meteorological stations), ERSST v4 (ocean areas), and SCAR (Antarctic stations) combined as described by Hansen *et al.*, 2010. The software calculated the difference between the surface temperature in a given month and the average temperature for the same place from the baseline. Higher than normal temperatures are shown in red and lower than normal temperatures are shown in blue. Credit for the Global Temperature Anomalies maps: NASA/Goddard Space Flight Center Scientific Visualization Studio. Data provided by Robert B. Schmunk (NASA/GSFC GISS). Visualizations by Lori Perkins (NASA/GSFC). Visuals review by lead scientist and head of the Goddard Institute for Space Studies (GISS) Gavin A. Schmidt (NASA/GSFC GISS). For more information visit http://svs.gsfc.nasa.gov/goto?4252.

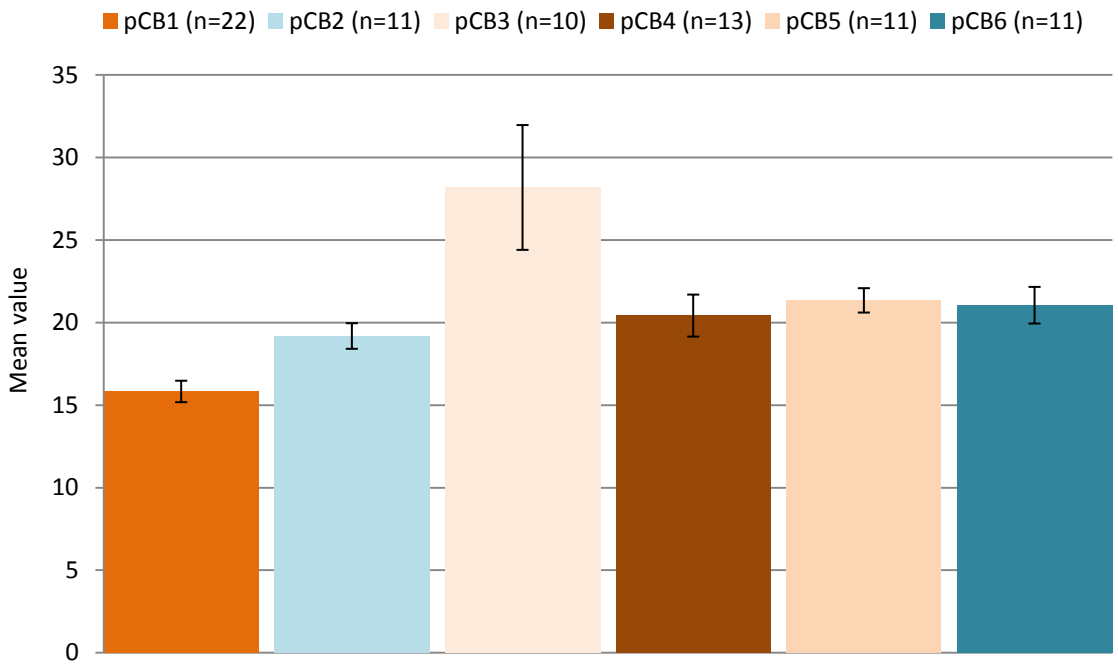
The CMM IVb genotype shows a different temporal abundance pattern to CMM IV despite their close sequence similarities (Krueger-Hadfield *et al.*, 2014). The significance of this remains unclear, especially since CMM IVb was seen to co-occur with CMM IV in the North Sea bloom of 2011 (Krueger-Hadfield *et al.*, 2014).

Finally, we have to take in account the possibility of homozygous and heterozygous CMM co-dominance. In the study of an *E. huxleyi* bloom event from 2011 in the North Sea, both homozygous clones for CMM I and CMM IV and heterozygous clones for CMM I/IV and CMM I/IVb were shown to co-dominate the bloom (Krueger-Hadfield *et al.*, 2014). In our study we could only assess which CMM allele was the most dominant in environmental samples and further analysis would be necessary to determine contingent heterozygosity.

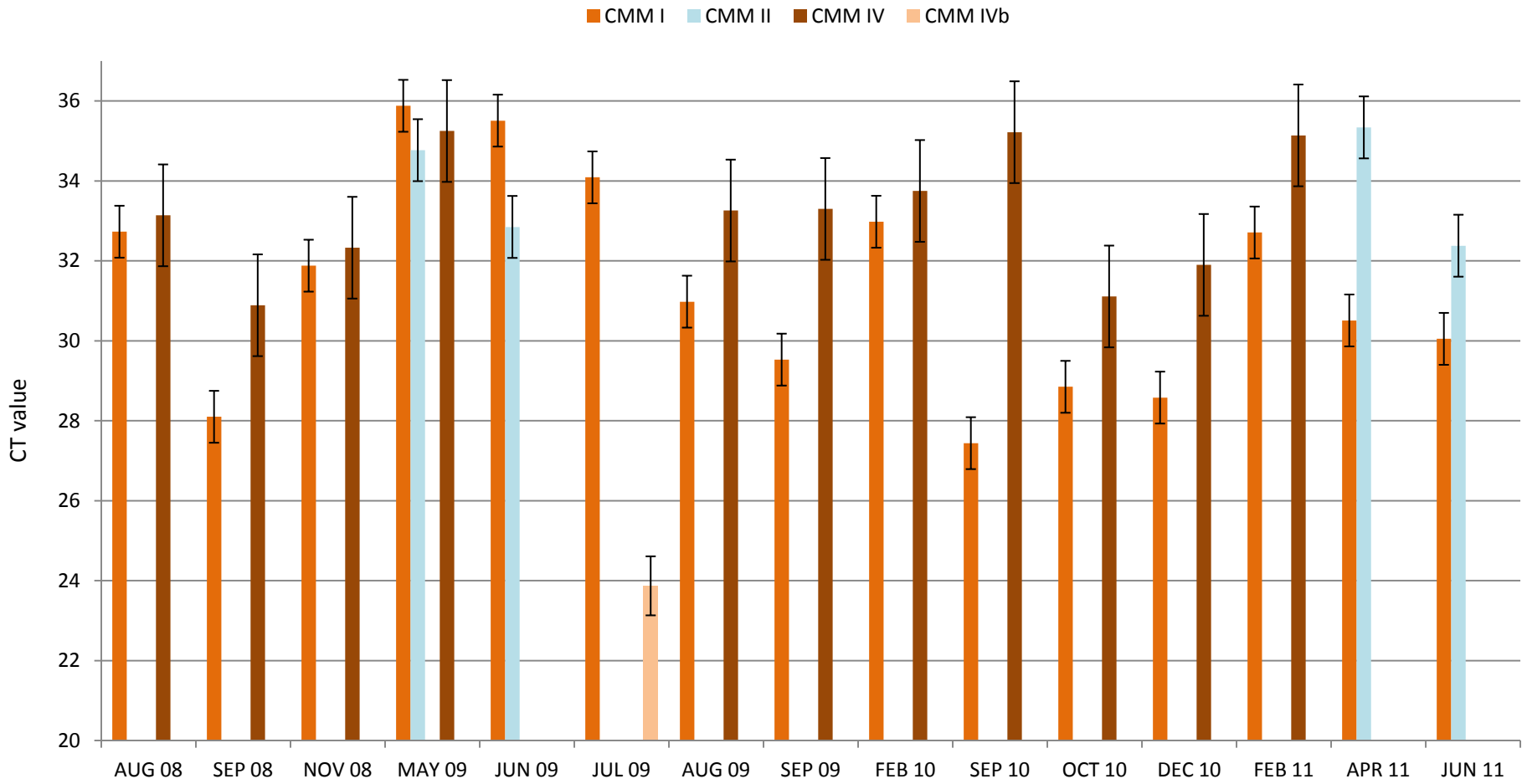
3.6 Supplement



Supplementary Figure 3.1 Different plasmidic concentration during a multiplexing run: after applying the slope correction visualisation mode and setting the fluorescence threshold at 0.15 it is observable a decreasing cycle threshold (CT) value mirroring a decreasing concentration of plasmidic material (specifically plasmid pCB2, corresponding to CMM II sequence). In detail: 1) Yellow curve, plasmid concentration 5 ng/µl, CT= 10.77; 2) Green curve, plasmid concentration 0.5 ng/µl, CT= 14.27, FD= 3.5; 3) Blue curve, plasmid concentration 0.05 ng/µl, CT= 18.56, FD= 4.29. The bottom line (light blue) represents the negative control.



Supplementary figure 3.2 Standard deviation for each CMM over multiple PCR runs. The small value of the standard deviation, with the exception of the CMM III, indicates robustness of the probe assay over multiple experiments. Represented on the plot there are the mean values for the CMM CT amplification calculated on the total number (n) of probe assays in which each CMM was recorded.



Supplementary figure 3.3 CT comparison of multiple CMMs detected in a same month. This bar plot shows the months in which multiple CMMs were recorded and their signal was strong enough for quantification. There is a correlation between the difference in CT values (CT) and the fold-difference (FD) of DNA and this determined which CMM dominated in a specific moment. Furthermore, the standard deviation bars are showed and only in four cases (August and November 2008, May 2009 and February 2010) we could not assess the dominance of a unique CMM.

CHAPTER 4

Emiliania huxleyi genotypes occupy distinct biogeographic provinces

*Cecilia Balestreri[1,3], Sally Thorpe[2], Rosalind E. M. Rickaby[3], William Balch[4], Colin Brownlee[1], and Declan C. Schroeder[1]**

[1]Marine Biological Association, Citadel Hill Laboratory, Plymouth, PL1 2PB, UK

[2]British Antarctic Survey, Madingley Road, High Cross, Cambridge, CB3 0ET, UK

[3]Department of Earth Sciences, Oxford University, South Parks Road, Oxford OX1 3AN, UK

[4]Bigelow Lab Ocean Sci, East Boothbay, ME USA.

*Corresponding author: dsch@mba.ac.uk

Authors' contribution

Cecilia Balestreri wrote the manuscript and Declan Schroeder edited it. Cecilia Balestreri also collected all the environmental samples (cruises JR271 and RR1202, see Appendix I) and she isolated the clones used in this study. Furthermore she ran the probe assay technique in order to assess the genetic composition of the collected material. Sally Thorpe produced the maps of dynamic topography and sea surface temperature for the regions taken into account in this study. William Balch organized the RR1202 cruise (Appendix I) and provided *in situ* sea surface temperature data for the samples collected in the Southern Indian Ocean. Ros Rickaby and Colin Brownlee contributed to the mentoring of Cecilia and in the revision and editing of the final draft before submission.

4.1 Abstract

Emiliania huxleyi is a coccolithophore that has a global distribution and under appropriate conditions form extensive blooms in both northern and southern latitudes. A few studies have suggested that certain *E. huxleyi* genotypes are generalists, with others having restricted biogeographical boundaries. Here we report on a biogeographic survey to determine the extant diversity of *E. huxleyi* in both the Northern and Southern hemispheres. We profiled DNA extracted from both *in situ* environmental water samples and were possible from single-celled isolated clones established from the same bodies of water. We show that *E. huxleyi* morphotype A as defined by the coccolith morphology motif (CMM) I dominates the temperate regions (10 -18 ^{o}C), while morphotypes B & B/C as defined by CMM II & IIb, respectively, dominate the polar regions (<10 ^{o}C). There was however a notable exception in the Greenland Sea, likely due to the extensive intrusion of temperate waters into this otherwise polar environment. These results confirm the genetic variability in the global *E. huxleyi* community and how the intraspecific variability is defined by geophysical features. Furthermore, our finding demonstrate that we should be cautious in describing the genetic composition of biogeographic populations based on cultures maintained in laboratory collections alone, as the cultured isolates might not be representative of real extant stocks.

4.2 Introduction

Over the last two decades, biodiversity and ecosystem function has emerged as one of the most controversial research areas in ecology (Loreau *et al.*, 2002). Species definitions further complicate matters especially in the microbial world where sequence barcodes are routinely being used to define and re-define species (e.g. Bendif *et al.*, 2013). Furthermore, genetic variation is driven by breeding, the environment and random mutations. If this variation occurs in genes, the variation might express certain advantageous traits and adaptable features (Van Tienderen *et al.*, 2002). It is predicted that a rapidly changing climate will directly influence species diversity and distribution (Thomas *et al.*, 2004).

A study by Liu et al. (2009) demonstrated that haptophytes dominate the chlorophyll-a normalized phytoplankton standing stock in modern oceans. They propose that ecological and evolutionary success of this group may have significantly impacted the oceanic carbon pump (Liu *et al.*, 2009). The marine coccolithophore *Emiliania huxleyi* is one of the most abundant calcareous phytoplankton species, it is found at all the latitudes but the extreme polar regions and under appropriate conditions it can form extensive blooms observable through satellite images (Berge, 1962; Brown & Yoder, 1994; Holligan et al., 1993; Holligan et al., 1983).

Over the past decade, many gene markers have been used to characterise the biodiversity within this species (Bendif *et al.*, 2014).

We have recently used a multiplex probe assay, using the CMM as a short molecular marker, to monitor the *in situ E. huxleyi* intraspecific genetic diversity at the L4 long-term monitoring site within the Western Channel Observatory over a

period of 6 years. We report that only 2 out of the 6 known CMM genotypes, CMM I & CMM IV, persist and dominate throughout the year; in both bloom and non-blooming periods (Balestreri *et al.,* unpublished). Both genotypes correspond to the classic morphotype A phenotype, which have in the recent past been reported to dominate North-eastern Atlantic, Western English Channel (WEC) and North Sea *E. huxleyi* bloom events. One specific genotype linked to morphotype B, that previously formed blooms in the WEC and North Sea, was only found during the winter non-blooming months in the WEC in our 6-year dataset. This study highlights the importance of characterizing the standing stock for any given region as this is fundamental to determining which *E. huxleyi* population will be selected for in future ocean scenarios.

We have also previously suggested that the two main morphotypes, A and B (which includes B/C) of *E. huxleyi* is unequally distributed in the modern oceans (Krueger-Hadfield *et al.,* 2014; Balestreri *et al*., unpublished). Here we wanted to further test this hypothesis by screening the DNA collected from environmental samples and clonal cultures isolated from numerous scientific cruises across different geo-thermal boundaries. We sampled across the geo-thermal boundary in the Southern Indian Ocean created by the Antarctic Circumpolar Current (ACC). A similar geo-thermal boundary was sampled in the Arctic Ocean, where the Gulf Stream enters the northernmost part of the Atlantic Ocean, passing south-east the Iceland, crossing the Norwegian Sea toward the Barents Sea. We show that both the Arctic and Antarctic polar waters have selected similar ecotypes, while temperate waters are dominated by two genotypes within morphotype A.

4.3 Materials and methods

4.3.1 Environmental samples

A total of 49 environmental samples (Table 4.2) were collected from CTD casts during two cruises as part of the UK Ocean Acidification research program (Table 4.1). These cruises were aimed to investigate the biology of surface ocean communities and biogeochemistry in areas known as CO_2 'sinks and sources' (www.surfaceoa.org.uk, see the cruise track in Appendix I). Additionally, we collected 26 samples during the second transect of the Great Southern Coccolithophore Belt (Table 4.1), a cruise which was aimed to examine several aspects of the coccolithophore biology and the impact of short-term ocean acidification on coccolithophore growth and calcite dissolution (www.bco-dmo.org/project/473206, see the cruise track in Appendix I).

Table 4.1 Cruise list

VESSEL	CRUISE NAME	PERIOD
RRS Discovery	D366	June/July 2011
RV Roger Revelle	RR1202	February/March 2012
RRS James Clark Ross	JR271	June/July 2012

Table 4.2 List of the environmental samples

Station number	CTD cast	Original collection area	Latitude (DD.dddddd°)	Longitude (DD.dddddd°)	Year
1	JR271 CTD 42	*Greenland Sea*	78,9871	7,9797	2012
2	JR271 CTD 29	*Greenland Sea*	78,7181	-0,0001	2012
3	JR271 CTD 41	*Greenland Sea*	78,4218	2,7657	2012
4	JR271 BIO 4	*Greenland Sea*	78,3526	-4,168	2012
5	JR271 CTD 44	*Greenland Sea*	77,9291	9,1365	2012
6	JR271 CTD 40	*Greenland Sea*	77,8465	-1,2959	2012
7	JR271 CTD 45	*Greenland Sea*	76,2619	12,5416	2012
8	JR271 CTD 27	*Greenland Sea*	76,1753	-2,5495	2012
9	JR271 CTD 48	*Eastern Barents Sea*	74,09	25,9993	2012
10	JR271 BIO 5	*Eastern Barents Sea*	72,8916	26,0017	2012
11	JR271 CTD 55	*Norwegian Sea*	71,7607	13,3949	2012
12	JR271 CTD 56	*Norwegian Sea*	71,7475	8,4428	2012
13	JR271 CTD 57	*Norwegian Sea*	71,7519	3,8717	2012
14	JR271 CTD 58	*Norwegian Sea*	71,7453	-1,2672	2012
15	JR271 CTD 59	*Norwegian Sea*	71,7517	-5,8638	2012
16	JR271 CTD 62	*Norwegian Sea*	70,5083	-10,1	2012
17	JR271 CTD 64	*North of Iceland*	67,8343	-12,1742	2012
18	JR271 CTD 65	*North of Iceland*	67,8304	-16,4218	2012
19	JR271 CTD 67	*North of Iceland*	67,8315	-20,0642	2012
20	JR271 CTD 68	*North of Iceland*	67,2642	-24,0415	2012
21	JR271 CTD 17	*South West of Iceland*	60,5942	-18,8565	2012
22	JR271 BIO 2	*South West of Iceland*	60,5942	-18,8565	2012
23	JR271 CTD 12	*South West of Iceland*	60,0014	-18,6702	2012
24	JR271 CTD 8	*North and North-West of Scotland*	60,1342	-6,7121	2012
25	JR271 CTD 10	*North and North-West of Scotland*	59,971	-11,9751	2012
26	D366 CTD 15	*North and North-West of Scotland*	59,9423	-1,7858	2011
27	D366 CTD 16	*North and North-West of Scotland*	60,0002	-2,6645	2011
28	D366 CTD 17	*North and North-West of Scotland*	59,4247	-7,7973	2011
29	D366 CTD 18	*North and North-West of Scotland*	57,4512	-11,1675	2011
30	D366 BIO 1	*North and North-West of Scotland*	56,7943	-7,4055	2011
31	JR271 CTD 6	*North Sea*	58,7397	-0,8615	2012
32	D366 CTD 11	*North Sea*	57,9245	4,865	2011
33	D366 CTD 10	*North Sea*	57,6938	4,4183	2011
34	D366 CTD 9	*North Sea*	57,5778	4,2017	2011
35	D366 CTD 13	*North Sea*	57,4595	5,5472	2011
36	D366 CTD 8	*North Sea*	57,4492	3,9582	2011
37	D366 CTD 14	*North Sea*	57,2415	4,0308	2011
38	D366 CTD 7	*North Sea*	57,201	3,4893	2011
39	D366 CTD 12	*North Sea*	57	4,9978	2011
40	D366 BIO 5	*North Sea*	56,504	3,6555	2011
41	JR271 BIO 1	*North Sea*	56,2666	2,6333	2012

42	D366 CTD 6	*North Sea*	54,372	5,1592	2011
43	D366 CTD 5	*North Sea*	53,656	4,1977	2011
44	D366 BIO 4	*North Sea*	52,994	2,4997	2011
45	D366 CTD 1	*South West of Ireland*	51,258	-11,3348	2011
46	D366 CTD 4	*Western English Channel*	50,0867	-4,6178	2011
47	D366 CTD 2	*Western English Channel*	50,029	-4,3795	2011
48	D366 CTD 3	*Western English Channel*	48,8237	-5,1492	2011
49	D366 BIO 3	*Bay of Biscay*	46,2013	-7,2215	2011
50	RR1202 CTD 1	*Southern Indian Ocean: above the ACC*	-35.507	37.455183	2012
51	RR1202 CTD 5	*Southern Indian Ocean: above the ACC*	-37.053	39.517	2012
52	RR1202 CTD 7	*Southern Indian Ocean: above the ACC*	-37.920117	40.440033	2012
53	RR1202 BIO 1	*Southern Indian Ocean: above the ACC*	-38.314983	40.958083	2012
54	RR1202 CTD 13	*Southern Indian Ocean: above the ACC*	-40.355133	43.496483	2012
55	RR1202 CTD 22	*Southern Indian Ocean: above the ACC*	-44.11375	48.178283	2012
56	RR1202 CTD 26	*Southern Indian Ocean: above the ACC*	-45.707933	50.25725	2012
57	RR1202 CTD 29	*Southern Indian Ocean: above the ACC*	-46.059317	52.646483	2012
58	RR1202 CTD 35	*Southern Indian Ocean: above the ACC*	-46.743333	57.483917	2012
59	RR1202 BIO 2	*Southern Indian Ocean: above the ACC*	-46.821367	58.312133	2012
60	RR1202 CTD 38	*Southern Indian Ocean: above the ACC*	-47.047067	59.922217	2012
61	RR1202 CTD 47	*Southern Indian Ocean: above the ACC*	-47.8981	67.37615	2012
62	RR1202 CTD 49	*Southern Indian Ocean: above the ACC*	-48.027267	69.053467	2012
63	RR1202 CTD 57	*Southern Indian Ocean: below the ACC*	-51.411017	72.602367	2012
64	RR1202 CTD 59	*Southern Indian Ocean: below the ACC*	-52.458117	73.265967	2012
65	RR1202 CTD 87	*Southern Indian Ocean: below the ACC*	-54.2439	88.139383	2012
66	RR1202 CTD 89	*Southern Indian Ocean: below the ACC*	-52.768717	90.217383	2012
67	RR1202 CTD 77	*Southern Indian Ocean: below the ACC*	-58.749433	80.8856	2012
68	RR1202 CTD 102	*Southern Indian Ocean: above the ACC*	-43.135083	102.190017	2012
69	RR1202 CTD 100	*Southern Indian Ocean: above the ACC*	-44.616767	100.498517	2012
70	RR1202 CTD 96	*Southern Indian Ocean: above the ACC*	-47.588783	96.93655	2012
71	RR1202 CTD 111	*Southern Indian Ocean: above the ACC*	-39.475433	108.934817	2012
72	RR1202 CTD 112	*Southern Indian Ocean: above the ACC*	-40.260267	109.62705	2012
73	RR1202 CTD 119	*Southern Indian Ocean: above the ACC*	-42.081683	113.399883	2012
74	RR1202 CTD 115	*Southern Indian Ocean: above the ACC*	-42.666783	111.760783	2012
75	RR1202 CTD 117	*Southern Indian Ocean: above the ACC*	-43.998867	112.9806	2012

4.3.2 Sea Surface Temperature, Mesoscale Circulation and Geo-thermal boundaries

Daily maps of absolute dynamic topography and sea surface temperature were used to examine the mesoscale circulation of all the regions in the six months prior to sampling at the stations. Images for Figures 4.1 & 4.2 were selected from the 3 month period prior to sampling at intervals of 2 weeks. The absolute dynamic topography fields were calculated by Aviso at 1/4 degree horizontal resolution from all remotely-sensed altimetry mission data available at a given time referenced to a 20 year mean. High resolution (1/20 degree) sea surface temperature data were produced from the Operational Sea surface Temperature and Ice Analysis (OSTIA) system using both in situ and satellite data (Donlon *et al.*, 2012). Based on these data the samples could be assigned to two main geo-thermal groups: cold polar seawater (<10 °C) and temperate seawater (10- 18 °C).

4.3.3 Clonal isolates

Seawater samples were transferred in vessel containing f/2 medium (Guillard, 1975) and maintained on board in incubation at 15 °C and 8 °C, for cruise D366, and cruises RR1202 & JR271 respectively. They were irradiated by a photon flux of 40-55 µmol $m^{-2}s^{-1}$ on a 16:8 hours LD cycle. Clonal isolates (Table 4.3) were established as described in Krueger-Hadfield et al. (2014). The environmental samples were screened and sorted using a FACSORT flow cytometer. Cell count was assessed using a Accuri C6 flow cytometer (threshold FSC 2000 and FL3 800) and dilution factor was calculated to obtain a starting concentration of about 1000 cell/ml. Subsequently each sample was subjected to a dilution-to-extinction regime in

order to isolate individual cells and thus obtain clonal unialgal cultures (see Appendix II). All the cultures were maintained in f/2 medium (Guillard, 1975) in a constant temperature room, at 15 °C and irradiated by a photon flux of 40-55 µmol m-2s-1 on a 16:8 hours LD cycle.

Table 4.3 List of the clonal isolates

Station number	Clone	Morphotype	CMM detected	
			Allele 1	Allele 2
1	JR271 30-1	A	I	IV
	JR271 30-2	A	I	IV
	JR271 30-3	A	I	IV
	JR271 30-4	A	I	IV
	JR271 30-5	A	I	IV
2	JR271 26-1	A	II	IV
	JR271 26-2	A	II	IV
	JR271 26-3	A	II	IV
	JR271 26-4	A	II	IV
	JR271 26-5	A	II	IV
3	JR271 29-1	A	I	I
	JR271 29-2	A	I	I
	JR271 29-3	A	I	I
	JR271 29-4	A	I	I
	JR271 29-5	A	I	I
5	JR271 31-1	A	I	I
	JR271 31-2	A	I	I
	JR271 31-3	A	I	I
	JR271 31-4	A	I	I
	JR271 31-5	A	I	I
6	JR271 28-3	A	I	I
	JR271 28-4	A	I	I

	JR271 28-5	A	I	I
	JR271 28B-1	A	I	I
	JR271 28B-2	A	I	I
	JR271 28B-3	A	I	I
	JR271 28B-4	A	I	I
	JR271 28B-5	A	I	I
	JR271 28C-1	A	I	I
	JR271 28C-2	A	I	I
	JR271 28C-4	A	I	I
	JR271 28C-5	A	I	I
7	JR271 59-2	A	IV	IV
	JR271 59-4	A	IV	IV
	JR271 59-5	A	IV	IV
9	JR271 61-1	A	I	I
	JR271 61-2	A	I	I
	JR271 61-3	A	I	I
	JR271 61-4	A	I	I
	JR271 61-5	A	I	I
10	JR271 50-2	A	I	I
	JR271 50-3	A	I	I
	JR271 50-4	A	I	I
	JR271 50-5	A	I	I
	JR271 51-1	A	I	I
	JR271 51-2	A	I	I
	JR271 51-3	A	I	I
	JR271 51-4	A	I	I
	JR271 51-5	A	I	I
	JR271 52-1	A	I	I
	JR271 52-2	A	I	I
	JR271 52-5	A	I	I
	JR271 52B-1	A	I	I
	JR271 52B-2	A	I	I

	JR271 52B-3	A	I	I
	JR271 52B-4	A	I	I
	JR271 52B-5	A	I	I
	JR271 55B-1	A	I	I
	JR271 55B-2	A	I	I
	JR271 55B-3	A	I	I
	JR271 55B-4	A	I	I
	JR271 55B-5	A	I	I
	JR271 55B-6	A	I	I
	JR271 55B-7	A	I	I
	JR271 55B-8	A	I	I
	JR271 55B-9	A	I	I
	JR271 55B-10	A	I	I
	JR271 56-1	A	I	I
	JR271 56-2	A	I	I
	JR271 56-3	A	I	I
	JR271 56-4	A	I	I
	JR271 56-5	A	I	I
	JR271 58-1	A	I	I
	JR271 58-2	A	I	I
	JR271 58-3	A	I	I
	JR271 58-4	A	I	I
	JR271 58-5	A	I	I
12	JR271 63-1	A	IV	IV
	JR271 63-2	A	IV	IV
	JR271 63-3	A	IV	IV
	JR271 63-4	A	IV	IV
	JR271 63-5	A	IV	IV
	JR271 63B-1	A	I	I
	JR271 63B-2	A	I	I
	JR271 63B-4	A	I	I
17	JR271 68-1	B/C	IIb	IIb

	JR271 68-2	B/C	IIb	IIb
	JR271 68-3	B/C	IIb	IIb
	JR271 68-4	B/C	IIb	IIb
	JR271 68-5	B/C	IIb	IIb
18	JR271 69-1	A	IVb	IVb
	JR271 69-2	A	IVb	IVb
	JR271 69-3	A	IV	IVb
	JR271 69-4	A	IV	IVb
	JR271 69-5	A	IVb	IVb
21	JR271 23-2	A	I	I
	JR271 23-3	A	I	I
	JR271 23-4	A	I	I
	JR271 23-5	A	I	I
22	JR271 10-1	A	I	I
	JR271 10-2	A	I	I
	JR271 10-3	A	I	I
	JR271 10-4	A	I	I
	JR271 10-5	A	I	I
	JR271 12-1	A	I	I
	JR271 12-2	A	I	I
	JR271 12-3	A	I	I
	JR271 12-4	A	I	I
	JR271 12-5	A	I	I
	JR271 12B-1	A	IV	IV
	JR271 12B-2	A	IV	IV
	JR271 12B-3	A	IV	IV
	JR271 12B-4	A	IV	IV
	JR271 12B-5	A	IV	IV
	JR271 13-1	A	IV	IVb
	JR271 13-2	A	I	I
	JR271 13-3	A	IV	IVb
	JR271 13-4	A	IV	IVb

	JR271 13-5	A	I	I
	JR271 15-2	A	I	I
	JR271 15-3	A	I	I
	JR271 15-4	A	I	I
	JR271 15-5	A	I	I
	JR271 15B-1	A	IV	IV
	JR271 15B-2	A	IV	IV
	JR271 15B-3	A	IV	IV
	JR271 15B-4	A	IV	IV
	JR271 15B-5	A	IV	IV
	JR271 16-1	A	I	I
	JR271 16-2	A	I	I
	JR271 16-3	A	I	I
	JR271 16-4	A	I	I
	JR271 16-5	A	I	I
	JR271 17-2	A	I	I
	JR271 17-3	A	I	I
	JR271 17-4	A	I	I
	JR271 17-5	A	III	III
	JR271 18-1	A	I	I
	JR271 18-2	A	I	I
	JR271 18-3	A	I	I
	JR271 18-4	A	I	I
	JR271 18-5	A	I	I
30	D366 48-2	A	IV	IV
	D366 48-3	A	I	IVb
	D366 48-5	A	III	IVb
31	JR271 19-1	A	I	I
	JR271 19-2	A	I	I
	JR271 19-3	A	IV	IV
	JR271 19-4	A	I	I
32	D366 120-1	A	I	I

	D366 120-2	A	I	IVb
34	D366 112-1	A	I	I
	D366 112-2	A	I	I
	D366 112-3	A	I	I
35	D366 124-1	A	I	I
	D366 124-2	A	I	I
	D366 124-3	A	I	I
	D366 124-4	A	I	I
	D366 124-5	A	I	I
	D366 126-2	A	IV	IV
38	D366 106-2	A	I	I
	D366 106-4	A	I	IV
	D366 106-5	A	I	IV
40	D366 17-1	A	I	I
	D366 17-3	A	I	I
	D366 19-2	A	I	IVb
	D366 20-4	A	I	I
	D366 21-3	A	I	I
	D366 21-4	A	I	IVb
	D366 21-5	A	I	I
	D366 22-4	A	I	I
	D366 24-1	A	I	I
	D366 25-3	A	I	I
	D366 26-1	A	I	I
	D366 26-2	A	I	I
	D366 26-3	A	I	I
	D366 26-4	A	I	I
	D366 26-5	A	I	I
	D366 30-1	A	I	I
	D366 30-2	A	I	I
	D366 30-4	A	IV	IV
	D366 31-2	A	I	IV

	D366 31-3	A	I	IVb
	D366 32-5	A	I	I
	D366 33-1	A	I	I
	D366 33-2	A	I	I
	D366 33-3	A	I	I
	D366 34-1	A	I	IV
	D366 34-3	A	I	IV
	D366 35-1	A	I	I
	D366 35-2	A	I	I
	D366 35-3	A	I	I
	D366 36-2	A	I	IVb
	D366 36-4	A	I	I
	D366 36-5	A	I	IVb
	D366 37-4	A	I	I
	D366 40-1	A	I	I
	D366 40-3	A	I	I
	D366 40-5	A	I	I
	D366 89-5	A	I	I
	D366 91-2	A	IV	IVb
46	D366 97-5	A	I	IV
	D366 98-1	A	IV	IV
49	D366 71-1	A	IV	IV
	D366 71-4	A	IV	IV
50	RR1202 5-1	A	IV	IV
	RR1202 5-2	A	IV	IV
	RR1202 5-3	A	IV	IV
	RR1202 6-1	A	I	I
	RR1202 6-2	A	I	I
	RR1202 7-1	A	IV	IV
	RR1202 7-2	A	IV	IV
	RR1202 7-3	A	IV	IV
	RR1202 42-1	A	IV	IV

	RR1202 42-2	A	IV	IV
	RR1202 42-3	A	IV	IV
	RR1202 42-4	A	IV	IV
	RR1202 42-5	A	IV	IV
51	RR1202 22-1	A	I	I
	RR1202 22-2	A	I	I
52	RR1202 23-1	A	IV	IV
	RR1202 23-2	A	IV	IV
	RR1202 23-3	A	IV	IV
53	RR1202 1-1	A	IV	IV
	RR1202 1-3	A	IV	IV
	RR1202 1-4	A	IV	IV
	RR1202 1-6	A	IV	IV
	RR1202 1-8	A	IV	IV
	RR1202 19-1	A	IV	IV
	RR1202 19-2	A	IV	IV
	RR1202 19-3	A	IV	IV
	RR1202 19-4	A	IV	IV
	RR1202 19-5	A	IV	IV
	RR1202 21-1	A	IV	IV
	RR1202 21-2	A	IV	IV
	RR1202 21-3	A	IV	IV
	RR1202 21-4	A	IV	IV
	RR1202 21-5	A	IV	IV
	RR1202 40-1	A	IV	IV
	RR1202 40-2	A	IV	IV
	RR1202 40-3	A	IV	IV
	RR1202 40-4	A	IV	IV
	RR1202 40-5	A	IV	IV
	RR1202 63-1	A	IV	IV
	RR1202 63-3	A	IV	IV
	RR1202 63-4	A	IV	IV

	RR1202 63-5	A	IV	IV
	RR1202 63B-1	A	IV	IV
	RR1202 69-1	A	I	I
	RR1202 69-2	A	I	I
	RR1202 69-3	A	I	I
	RR1202 69-4	A	IV	IV
	RR1202 69-5	A	IV	IV
	RR1202 69B-1	A	IV	IV
	RR1202 69B-2	A	IV	IV
	RR1202 69B-3	A	IV	IV
	RR1202 69B-4	A	IV	IV
	RR1202 69B-5	A	IV	IV
64	RR1202 36-1	A	I	I
	RR1202 36-2	A	I	IV
	RR1202 36-3	A	I	IV
	RR1202 36-4	A	IV	IV
	RR1202 36-5	A	I	IV
68	RR1202 30-1	A	IV	IV
	RR1202 30-2	A	IV	IV
	RR1202 30-3	A	IV	IV
	RR1202 30-4	A	IV	IV
	RR1202 30-5	A	IV	IV
72	RR1202 14-1	A	III	III
	RR1202 14-2	A	III	III
	RR1202 14-3	A	III	III
	RR1202 14-4	A	III	III
	RR1202 14-5	A	III	III
73	RR1202 52-1	A	IV	IV
	RR1202 52-2	A	IV	IV
	RR1202 52-3	A	IV	IV
	RR1202 52-5	A	IV	IV
	RR1202 52B-1	A	IV	IV

	RR1202 52B-2	A	IV	IV
	RR1202 52B-3	A	IV	IV
	RR1202 52B-4	A	IV	IV
	RR1202 52B-5	A	IV	IV
	RR1202 55-1	A	IV	IV
	RR1202 55-2	A	IV	IV
	RR1202 55-3	A	IV	IV
	RR1202 55-4	A	IV	IV
	RR1202 55-5	A	IV	IV
75	RR1202 16-1	A	IV	IV
	RR1202 16-2	A	IV	IV
	RR1202 16-3	A	IV	IV
	RR1202 16-4	A	IV	IV
	RR1202 16-5	A	IV	IV

4.3.4 Scanning Electron Microscopy (SEM)

All the clonal cultures were filtered using a 0.45 µm cellulose nitrate membrane filter, mounted onto metallic stubs using adhesive tape and coated in a thin layer of gold (Au) using an Au sputter coater. These were visualized using a JEOL 5600 Low Vacuum Scanning Electron Microscope. Scanning electron micrographs were captured at magnifications ranging between x7.500 - x20,000, and electron beam damage was minimized by operating the microscope at 15 kV.

4.3.5 DNA extraction

From each CTD cast up to 1 L quantities of seawater was collected into Nalgene bottles and subsequently filtered using a vacuum pump and a filter rig through 0.45 µm polycarbonate membrane filters (PALL Corporation, Michigan, USA). Each filter was rinsed into a petri dish with 2ml of phosphate buffered saline (PBS) solution and the final solution was collected into an eppendorf tube. The solution from the eppendorf tubes was used to carry out DNA extraction *in situ* for each environmental sample, using Qiagen DNeasy Blood and Tissue kit protocol (QIAGEN, Valencia, CA, USA), and subsequently in our laboratory on each isolated strain and the clonality was confirmed by GPA gene as described in Krueger-Hadfield et al (2014).

4.3.6 Probe assay

Amplification of the coccolith morphology motif was carried out using multiplex probe assay technique previously described in Krueger-Hadfield *et al.*, 2014 and Balestreri *et al.,* unpublished.

4.4 Results

4.4.1 Culture and DNA collection

From the starting material collected on board of each cruise we were able to maintain in the laboratory in UK a total of 75 environmental cultures (Appendix II), of which 25 belonging to polar regions and 50 belonging to temperate regions. Out of the 25 'polar' samples, 5 were collected in the Southern Indian Ocean, below the Antarctic Circumpolar Current (ACC), while the remaining 20 'polar' samples were collected in the Greenland Sea, Eastern Barents Sea, the Norwegian Sea and North of Iceland (Figure 4.1 and 4.2). 21 'temperate' samples were collected in the Southern Indian Ocean, above the ACC and the remaining 29 'temperate' samples were collected in the ocean South-West of Iceland, North and North-West of Scotland, in the North Sea, in the Western English Channel, South West of Ireland, and in the Bay of Biscay (Figure 4.1 and 4.2).

Subsequently, we succeeded in isolating 292 clones, of which 100 clones from polar regions: 95 of them belonging to the Northern Hemisphere (45 of them from the Eastern Barents Sea, and 35 of them from the Greenland Sea) and 5 of them belonging to the Southern Hemisphere. The remaining 192 clones were obtained from temperate regions and 109 of them belonged to the Northern Hemisphere (56 of them from the North Sea, and 47 of them from South-West of Iceland), while 83 of them belonged to the Southern Hemisphere. DNA was successfully extracted for all the environmental and clonal isolate samples.

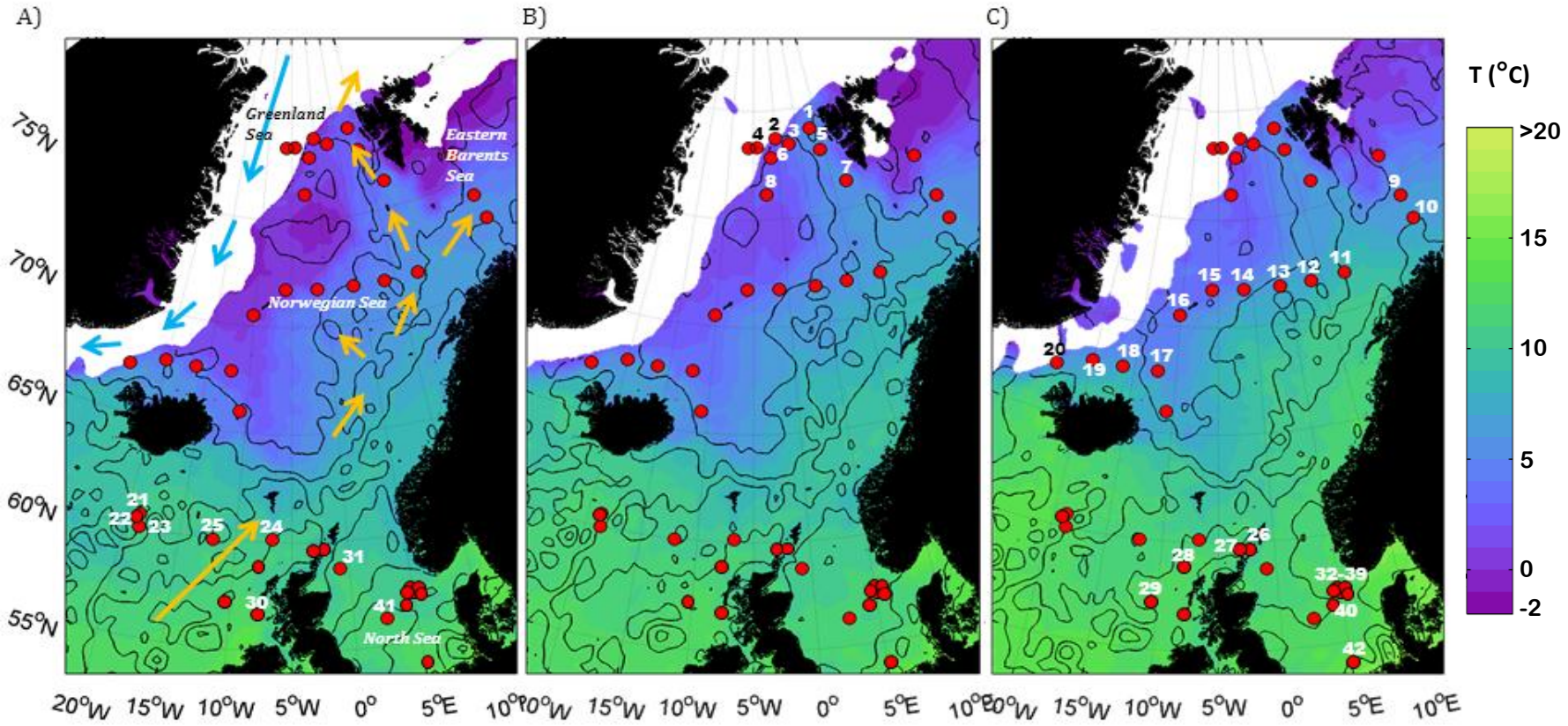


Figure 4.1 Composite images of the environmental sample locations for the Northern Hemisphere, obtained combining water flows and sea surface temperature (SST in °C). The black lines are dynamic height and the closed contours show eddies. The main current flows are reported by arrows: blue arrows for cold water coming from the North and orange arrows for the warmer currents coming from the Atlantic Ocean. Each red dot represents a sample and the number above the dot indicates the closest time of the year at which the sample was collected; in detail: A) refers to 1/06/2012, B) refers to 15/06/2012, and C) refers to 29/06/2012. Note that samples 26-30, 32-40 and 42 were collected in June/July of 2011and these stations on the maps are indicative of a similar temperature pattern.

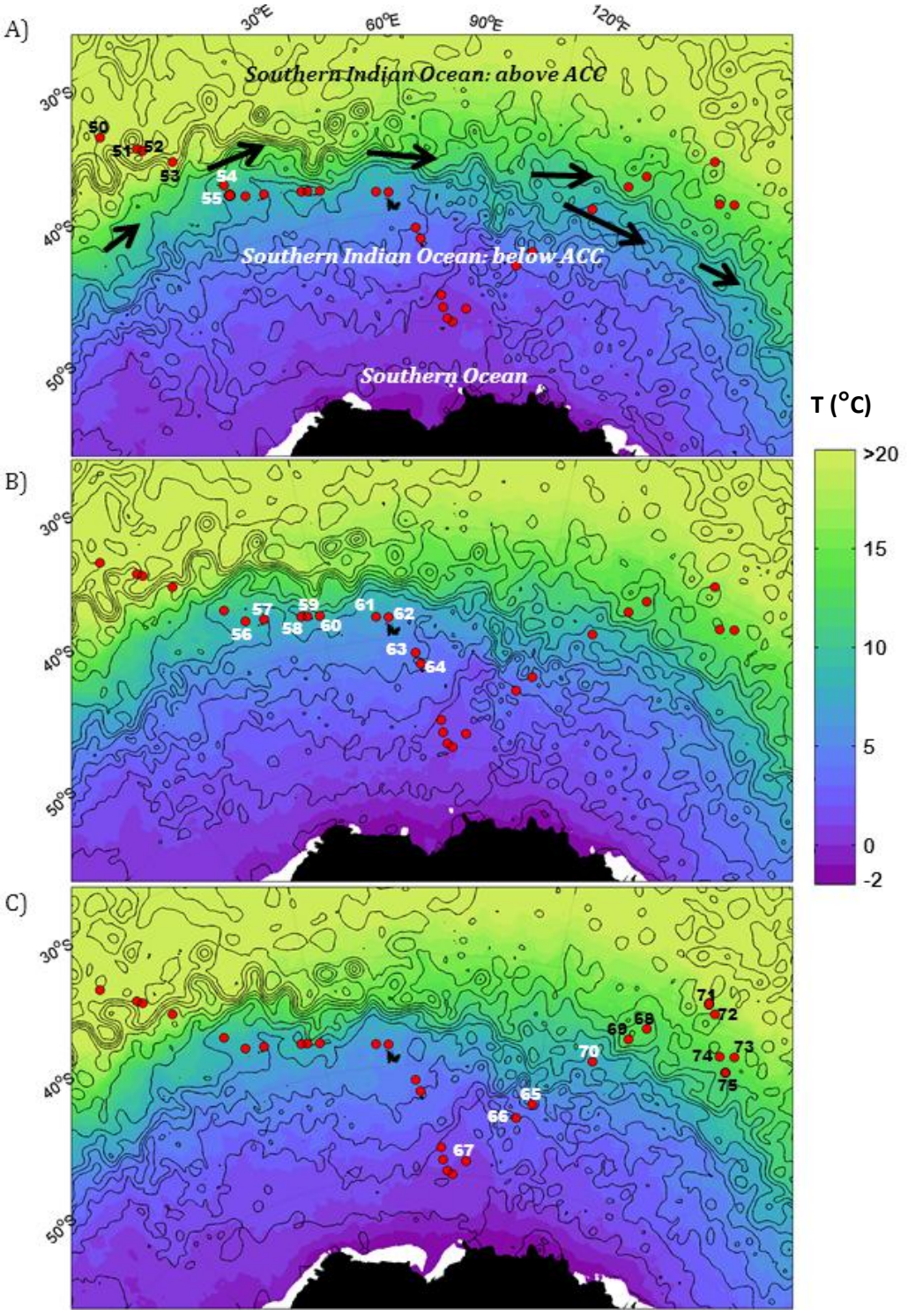


Figure 4.2 Composite images of the environmental sample locations for the Southern Hemisphere, obtained combining water flows and sea surface temperature (SST in °C). The black lines are dynamic height and the closed contours show eddies. The main current flow, the Antarctic Circumpolar Current (ACC) is reported by black arrows. Each red dot represents a sample and the number above the dot indicates the closest time of the year at which the sample was collected; in detail: A) refers to 15/02/2012, B) refers to 29/02/2012, and C) refers to 14/03/2012.

4.4.2 Genetic diversity of E. huxleyi in the Northern Hemisphere

A total of 49 environmental samples were screened for the CMM region. The samples collected in the Northern Hemisphere were dominated by CMM I, which was detected in 13 out of 20 in the polar region and 24 out of 29 in the temperate region (Figure 4.3). CMM II was found in 2 out of 20 in the polar region and 5 out of 29 in the temperate region. CMM IIb was detected 6 times in the polar region and 1 time only in the temperate region (Figure 4.3). CMM III, IV and IVb were amplified in 1, 13 and 3 samples and in 2, 2 and 2 in the temperate and polar region respectively. Seven samples out of 22 collected during the D366 cruise were sampled in a bloom area in the North Sea.

The CMM screening of the clonal cultures isolated from the temperate region showed that 74 resulted homozygous for CMM I, 16 of them were homozygous for CMM IV and 1 sample was homozygous for CMM III. The other strains studied were heterozygous: 6 strains had CMM I-IV, 7 strains had CMM I-IVb, 4 strains had CMM IV-IVb, 1 strain had CMM III-IVb (Figure 4.4).

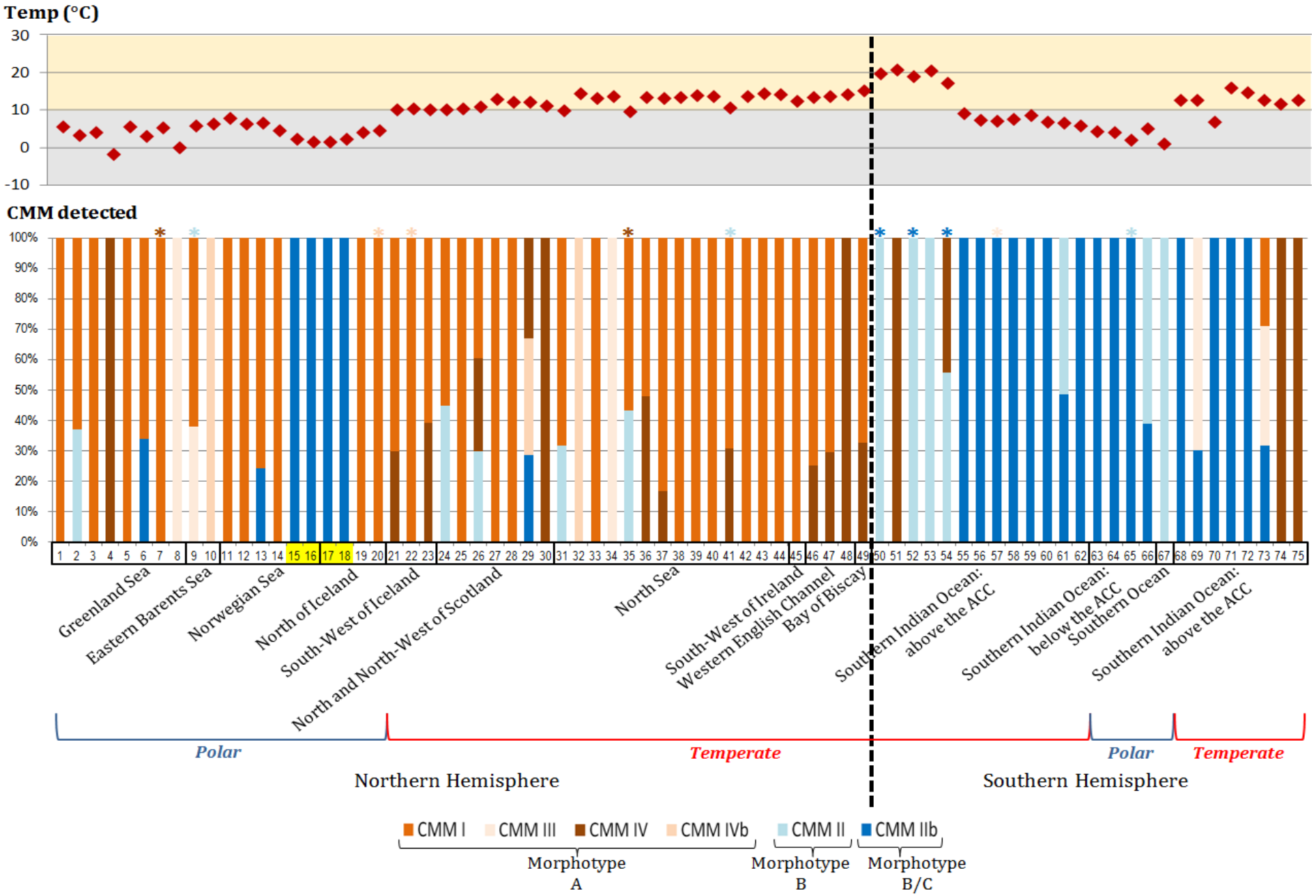

Temp (°C)
30
20
10
0
-10
CMM detected
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75
Greenland Sea
Eastern Barents Sea
Norwegian Sea
North of Iceland
South-West of Iceland
North and North-West of Scotland
North Sea
South-West of Ireland
Western English Channel
Bay of Biscay
Southern Indian Ocean: above the ACC
Southern Indian Ocean: below the ACC
Southern Ocean
Southern Indian Ocean: above the ACC
Polar
Temperate
Polar
Temperate
Northern Hemisphere
Southern Hemisphere
CMM I
CMM III
CMM IV
CMM IVb
CMM II
CMM IIb
Morphotype A
Morphotype B
Morphotype B/C

Figure 4.3 (previous page) Bar-plot of all the environmental samples and the respective sea water temperature (T). The dotted line indicates the Northern/Southern hemisphere division. Each geographical province is highlighted by a black rectangle grouping the data-point on the x axis. The bars show the different CMM composition for each station and the stars on top of the bars indicate a non-quantifiable, but detected, amplification. The CMM are grouped according to the relative morphotypes. The temperature plot shows the T *in situ* for each environmental sample.

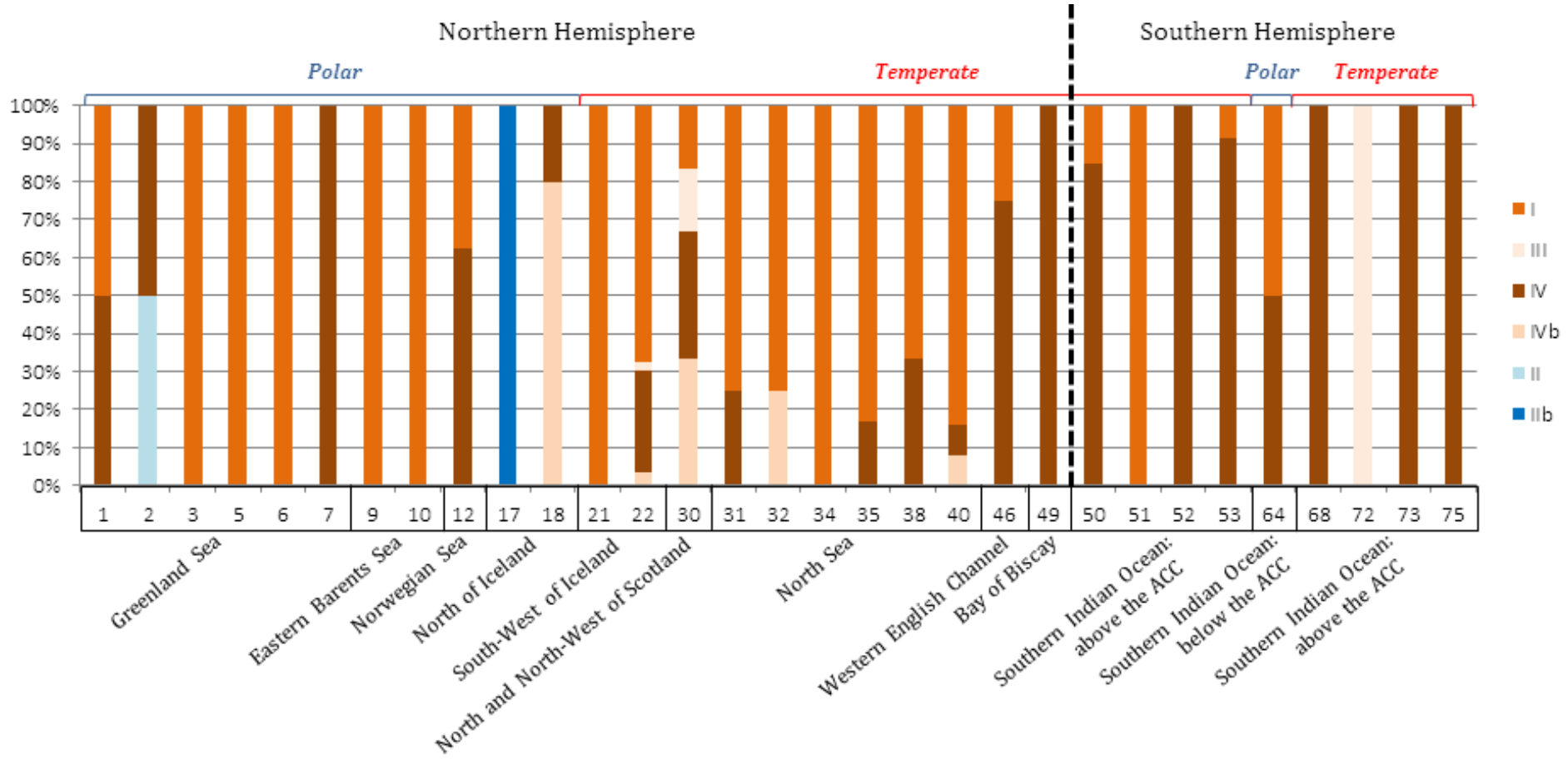


Figure 4.4 Bar-plot of the allele frequency detected in the clones. The dotted line indicates the Northern/Southern hemisphere division. Each geographical province is highlighted by a black rectangle grouping the data-point on the x axis. The bars show the allelic composition for each station.

4.4.3 Genetic diversity of E. huxleyi in the Southern Hemisphere

A total of 26 environmental samples from the Southern Hemisphere were screened for the CMM region and the majority of them were dominated by CMM IIb, both in polar and temperate regions: 4 out of 5 and 17 out of 21 samples respectively (Figure 4.3). CMM II was found in 3 out of 5 samples in the polar region and in 5 over 21 in the temperate region. The CMM I was amplified only in 1 samples from the temperate region and CMM III & IV were found 3 and 4 times exclusively in the temperate region (Figure 4.3).

The CMM composition changed when we analysed the 87 clonal isolates obtained from the original environmental cultures. Almost all the clones resulted homozygous, with the exception of 3 clones from the polar region having the allelic composition CMM I-IV. The CMM I-I was found in 1 sample from the polar region and in 7 samples in the temperate region. CMM IV-IV was detected in 1 sample in the polar region and 71 times in the temperate region, accounting as the main allelic combination in this region. Additionally, 5 samples from the temperate region showed CMM III-III composition (Figure 4.4).

4.5 Discussion

The amplification of the CMM region was not successful for more than a third of the samples collected at temperature <10 °C and it occurred significantly more often in samples collected in the Southern Ocean below 50 °S. This is probably explained by the fact that *E. huxleyi* is worldwide spread with the exception of the

extreme polar regions (Holligan *et al.*, 1993; Tyrrell and Merico, 2004; Hagino *et al.*, 2011), which are characterised by low sea temperatures (Antarctic Convergence <6 °C, Deacon, 1982; Arctic Ocean <4°C, Steele, Ermold, & Zhang, 2008). Nevertheless, in the Arctic Sea we could amplify samples collected at 79 °N and only 5 samples were not amplified for the CMM sequence and the majority of them belonged to the lower temperature group (<10 °C), supporting the idea that a low seawater temperature increases the probability of not finding *E. huxleyi* and subsequently the PCR failure for the GPA gene. We have to take into account that individual strains of *E. huxleyi* have different growth rate tolerances to temperature, in the range between 2 °C and 27 °C (Brand, 1981), therefore *E. huxleyi* specific growth rate may vary greatly, making extrapolation of strain-specific data to hypothetical scenarios such as models difficult (Fielding, 2013).

The screening of environmental samples revealed the dominance (>50% of the alleles detected) of CMM I in 38 out of the 75 site In fact, in 75% of the cases CMM I dominated the Northern Hemisphere sites, consistent with the results previously described by Krueger-Hadfield *et al.* (2014). On only one occasion CMM I could be found (<30% of the CMM alleles) in the oceans below the equatorial line, but nonetheless in a temperate environment. CMM I is associated with morphotype A (Schroeder *et al.*, 2005) and our findings confirm many oceanic observations (J. Young, personal communication) that morphotype A is the most wide spread and abundant in the modern oceans. More specifically, morphotype A and B are unequally distributed in the Northeast Atlantic region (van Bleijswijk et al., 1991) as well as in the South-west and South-east Atlantic (van Bleijswijk et al., 1994) and the morphotype B is found rarely compared to the morphotype A (Krueger-Hadfield *et*

al., 2014). In line with this, we amplified CMM II, which is related to the genuine B form, only 15 out of the 75 sites (20%) and it only dominated (>50% of the alleles detected) at 6 out of the 75 sites (8%) all exclusively to the Southern Hemisphere. We previously reported the presence of CMM II in the Western English Channel (WEC) in the non-blooming winter months (Balestreri *et al.,* unpublished), hypothesizing that its absence in the summer blooming months could be explained by the increase SSTs observed in recent years in the North Atlantic (Rayner *et al.*, 2003). In this study we note that CMM II could still be found (between 30 -45%) in more northerly latitudes in the summer blooming period months. This suggests that CMM II can still persist in this warming environment and the conditions have not deteriorated sufficiently to see its extinction. This further suggests that something else is at play in the WEC that has seen its disappearance in the summer months. Nonetheless, CMM II is in decline in the Northern Hemisphere, were it once formed blooms as the recorded by van Bleijswijk *et al.* (1991). Our study also opens a new possibility that CMM II might in fact be a Southern Hemisphere morphotype or ecotype that through ocean circulation or anthropogenic forcing resulted in incursions into the Northern Hemisphere regions.

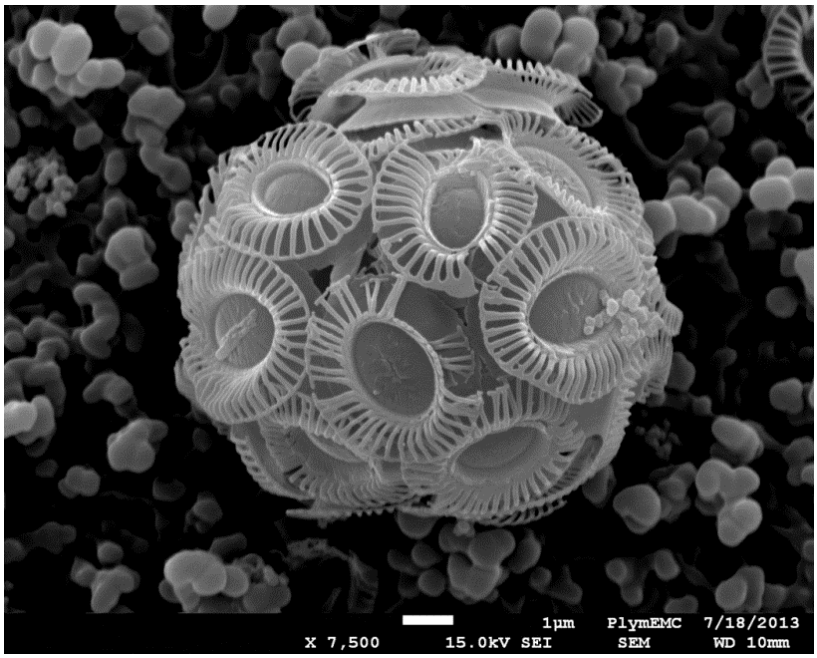
Over 205 total clones obtained from the samples collected in the Northern Hemisphere, 141 (69%) amplified for CMM I, as we would have expected after the CMM analysis of the environmental samples. In addition, the isolation of clones from samples collected in the Southern Indian Ocean produced 8 out 87 (9%) CMM I clones which confirms the presence of this genotype in the Southern Hemisphere in a smaller proportion. Furthermore, 7 of these southern CMMs I were isolated from a seawater temperature >10 °C and from locations influenced by the Agulhas Return

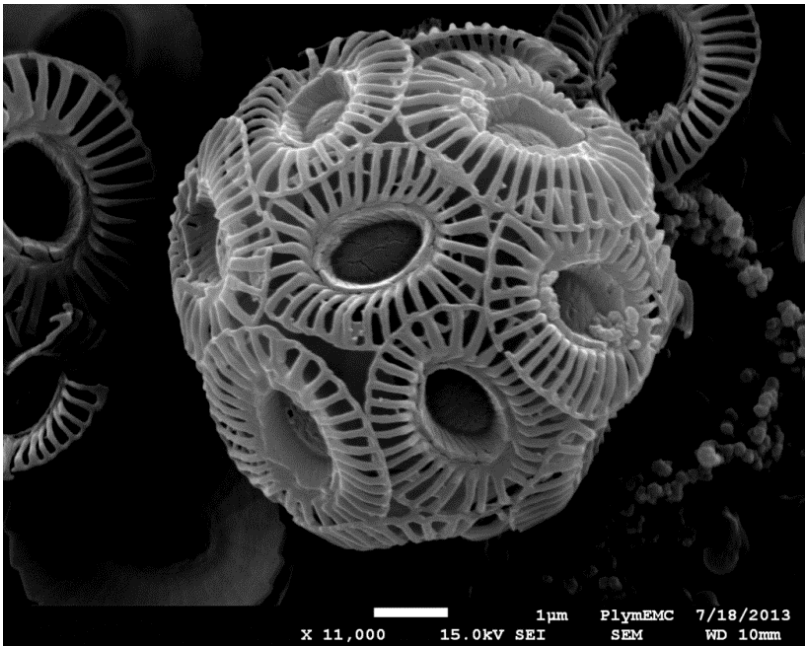
Current (Lutjeharms and Ansorge, 2001). This current shows enhanced primary productivity, with chlorophyll-a concentrations significantly higher than the surrounding South Indian Ocean (Mann and Lazier, 2005). CMM I was also present as one of the alleles in many of the heterozygous strains (21 clones isolated from the Northern Hemisphere). As previously observed by Krueger-Hadfield *et al.* (2014), based on clonal culture collections alone CMM I appears to be more resilient and thus it can dominate temperate regions at a regional and global scale.

CMM II was found 15 times in our environmental samples, 7 times in the Northern Hemisphere and 8 times in the Southern Hemisphere. Moreover, we found to be the dominant (>50% alleles detected) in 5 of the 13 (38%) temperate region influenced by the Agulhas Return Current. This is in direct contrast to the absence of the CMM II allele in any of the clonal isolates collected from either the temperate or polar regions. Our inability to bring the CMM II or morphotype B strains into culture suggests that this genotype has different nutrient and/or temperatures to that of morphotype A. Furthermore, we know that morphotype B is distinctly less calcified than morphotype A (Young et al. 2014) and maybe this factor makes this morphotype more susceptible to damage during the isolation process. This finding nonetheless highlights once again the biases associated with culture collections and that we would have been wholly incorrect to infer from our culture collection that CMM II or morphotype B in absent in the Southern Indian Ocean. In fact the most southerly sample site was exclusively CMM II.

Although CMM II (morphotype B) was not prevalent in any of the Northern Hemisphere environmental samples, the related genotype CMM IIb (correlated to the B/C morphotype) was detected on 5 occasions (4 of which were exclusively CMM IIb)

in samples collected in the Norwegian Sea and North of Iceland, an area where the cold water of the East Greenland Current separates from the main branch and turns southeast, contributing to the East Icelandic Current (Aagaard and Coachman, 1968; Delworth *et al.*, 1997; Woodgate *et al.*, 1999). These samples were therefore represented true polar waters. We confirmed the presence of *E. huxleyi* B/C morphotype through scanning electron microscopy (SEM; Figure 4.5). Five clones isolated from one environmental sample collected in the sea-area northern of Iceland were found to be B/C phenotype and their CMM analysis confirmed that their genotype was CMM IIb, supporting the correspondence between the morphotype and this specific genetic sequence.





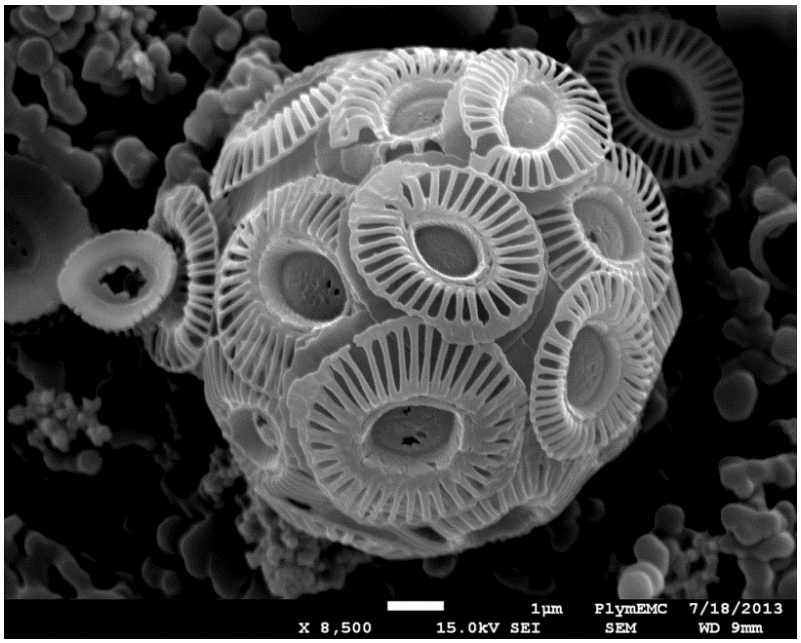


Figure 4.5 Electron Scanning Microscope (SEM) photos of *E. huxleyi* B/C morphotype found in clonal isolates North of Iceland.

In the Southern Hemisphere we found CMM IIb more frequently: in 21 samples out of 26 (81%). At 14 sites, CMM IIb was the exclusive genotype; 11 found in the temperate region, 3 in the polar region with one exception showing evidence of the presence of CMM II.

In the polar regions 1 environmental samples were dominated by CMM IV and its related genotype, CMM IVb, was also found to dominate on 1 occasion. In the temperate region we found CMM IV on 17 occasions and CMM IVb on 2 occasions. We know that these genotypes are found in the North Sea and Western English Channel (Krueger-Hadfield *et al.* 2014, Balestreri et al. unpublished) and that CMM IV can dominate bloom events in the WEC (Highfield *et al.*, 2014). The genotype IV appears to adapt to a wide range of environmental niches, from cold to warm and often changeable environments (Krueger-Hadfield *et al.*, 2014). However, as the Arctic polar region especially the Greenland Sea was heavy impacted by the temperate waters, the CMM IIb should be considered as the true Arctic genotype while CMMs I, III, IV & IVb are relics or active intrusions from temperate waters.

Out of 292 isolated clones, 96 (33%) amplified for the CMM IV, which means that this allele in present circa one in three of our isolates in the culture collection. Conversely only 3 strains showed CMM IVb, and they were isolated from cold seawater collected North of Iceland. The majority of the CMM IV clones were isolated from environmental material originated in the Southern Hemisphere in polar regions and in areas influenced by the Agulhas Return Current in the region South-East of South Africa and by Leeuwin Current in the region South-West of Western Australia. Only one clone was isolated from the North of Heard Island and McDonald Islands, in a polar region. The diversity of environments in which we found CMM IV and IVb

supports the hypothesis that these genotypes are able to adapt to a variety of niches. Moreover, CMM IV and IVb alleles were found in all the heterozygous strains, meaning that these genotypes mate with a higher frequency. The main heterozygous combinations detected in our study were CMM I-IV, I-IVb and IV-IVb and these findings are in line with Krueger-Hadfield *et al.* (2014) study, who showed co-existence of these alleles in presence of a bloom event dominated by homozygous CMM I, occurred in the North Sea in the summer of 2012.

CMM III was amplified on only 6 occasions in the environmental samples, 2 times in polar regions and 4 times in temperate regions. It was found in the Greenland Sea, Eastern Barents Sea and also in the North Sea, and in three locations South-West of Western Australia. This CMM is a rare genotype, despite its ability to occupy a multiplicity of conditions.

In summary, the most dominant genotype was CMM I (51% of the positive amplified samples), followed by CMM IIb (37%), which are related to *E. huxleyi* A and B/C morphotypes, respectively. Furthermore, we showed that the former dominated (75 %) in the Northern Hemisphere, while the latter dominated the Southern regions (81%). To our knowledge, this is the first study that reports on such a clear provincial delineation. Other studies have reported that the B/C morphotype of *E. huxleyi* is found in the Patagonian Shelf, in the Southern Ocean and in sub-polar Antarctic waters (Poulton *et al.*, 2011; Cook *et al.*, 2011; Cubillos *et al.*, 2007; Hagino *et al.*, 2005). We now can confirm the only other location that it can be found is in the true polar waters north of Iceland. Morphotype A (CMM I) is a temperate strain, while Morphotype B/C (CMM IIb) is a polar strain. In addition, morphotype B (CMM II) also appears to prefer the Southern Hemisphere and polar waters. However, this

assessment does not find an exact correspondence in the genetic composition that we observed in the clonal isolates. Although the dominant genotype among the 292 clones was again CMM I (51% of the total samples), the second most frequently was CMM IV, being present in more than 32% of the cases. Given this, we recommend that culture collections should not be used in isolation for the genetic profiling of *E. huxleyi* populations. The genetic variability that characterizes environmental material is biased toward a smaller number of genotypes grown in artificial conditions and it is not necessary representative of the real population structure of *E. huxleyi*.

We have to take into account, especially when making ecological statements based on lab based genetic studies, that every organism could not be obtained by laboratory culturing and isolation as a result of the artificial selection pressure placed on them. In most of the cases, artificial growth selection deeply impacts on the true genetic assessment of the original population, which can lead to misinterpretation of the real picture. In a laboratory environment the real ecological dynamics between organisms are not respected: competition for a resource, e.g. nutrients, might be compromised or overturned by the organism whose physiology better adapt to the artificial conditions (Service and Rose, 1985). The culture establishment is moreover altered by irradiance, salinity, nutrients and light which often differ from the natural environment (Lakeman *et al.*, 2009). Furthermore, the maintenance of a culture for long periods can lead to physiological and phenotypical changes in an organism, e.g. strain of calcifying *E. huxleyi* kept in a manipulated environment for a long time can lose their ability to calcify (Green *et al.*, 1996; Houdan *et al.*, 2005).

E. huxleyi occupies a varied range of habitat and therefore as a "species" is likely resilient to the changing ocean conditions. This is consistent with the numerous

studies that demonstrate the varied responses of *E. huxleyi* strains challenged with future climatic exposures (Iglesias-Rodriguez *et al.*, 2008; Riebesell *et al.*, 2008; Müller *et al.*, 2015; Rost and Riebesell, 2004; Meyer and Riebesell, 2015). That said we observed a clear separation of dominant genotypes both in habitat and between Northern and Southern hemispheres. Additionally, we observed a greater heterogeneity and heterozygosity in the Northern Hemisphere and in proximity of the continental shelves. In open ocean environments, far from the continents and at oligotrophic conditions, such as the Southern Indian Ocean, only a couple of genotypes prevailed in the environmental samples. Future experiments need to address the differential responses of strains/genotypes isolated from their respective habitats. Our current knowledge is largely limited to a CMM I, morphotype A response scenario.

CHAPTER 5

General discussion

"*Considering that the more prevalent coccolithophore species appear to be vulnerable to ocean acidification, a local or global shift in the species composition or a replacement by other photoautotrophic organisms may occur and could affect higher trophic levels and ocean biogeochemical cycling*" (Meyer and Riebesell, 2015).

The question of how the *Emiliania huxleyi* species-complex successfully colonized very diverse surface ocean habitats while its close relatives remained more ecologically restricted has broader implications for understanding controls on phytoplankton adaptation to new and changing habitats (Bendif *et al.*, 2015; Read *et al.*, 2013). Not only *E. huxleyi* is numerically the most abundant coccolithophore species, but it is found in several different habitats (Müller *et al.*, 2015). The work carried out in my PhD project, demonstrated and confirmed a high genetic variability associated with different morphotypes (Chapter 2 - Krueger-Hadfield *et al.*, 2014). Furthermore, it showed a dominance of certain CMM genotypes to specific geographical regions both in time (Chapter 3) and space (Chapter 4).

One of the main effects of anthropogenic CO_2 emissions that challenges calcifiers in current oceans is the acidification of seawater. The lowest pH values are found in upwelling areas, such as Equatorial Pacific and Arabian Sea, where the surface moves to deeper layers as the rich deep water with lower pH value is brought to the surface. Conversely, in areas where the biological production and export are elevated, the dissolved CO_2 is converted into organic carbon by phytoplankton and exported in deeper oceans resulting in higher pH values in the surface waters (Raven *et al.*, 2005). The Southern Ocean is known as one of the main sinks for CO_2, but

studies by Le Quéré *et al.* (2007) found that this sink has weakened between 1981 and 2004 by 0.08 petagrams of carbon per year per decade due alterations in the Southern Ocean winds. Moreover, low levels of iron and irradiance, especially during the winter, make the Southern Ocean not as productive as may be expected, although it is generally considered a nutrient rich ocean (Harper *et al.*, 2011). In the temperate and sub-polar regions of the North Atlantic and North Pacific, there are long-term records that the pH varies on a seasonal cycle (Bellerby *et al.*, 2005; Wootton *et al.*, 2008; Olafsson *et al.*, 2009) because of changes in physical ocean properties, such as temperature and salinity, but also because of the biological processes (Findlay, 2010). Nevertheless, in the Arctic Ocean the alkalinity level is already lower than many other oceans and overall polar pH levels are changing twice as fast as tropical ones; pre-industrial pH 8.2 dropped to pH 8.1, indicating increased acidity (Harper *et al.*, 2011).

Another main effect of increasing CO_2 emissions is the rise in ocean temperature (Feng *et al.*, 2009), and its negative effects on marine organisms have been shown in numerous studies. Global surface temperature has increased 0.2 °C per decade in the past 30 years, especially in the Western Equatorial Pacific and Indian Oceans, which are approximately as warm now as at the Holocene maximum (Hansen *et al.*, 2006). Moreover, the upper layer of the ocean retains anomalous heat, which is originated from changing solar irradiance (White *et al.*, 1997). In the region of the Weddell Sea from 35°S to 65°S, the Southern Ocean has warmed by 0.17 °C in the upper 1000 metres since the 1950s (Gille, 2002) and a predominantly sea-ice-free Arctic is predicted for the end of this century (Johannessen et al., 2004).

Morphotype A of *E. huxleyi* appears to dominate the wild populations in the Northern Hemisphere, while the morphotypes B and B/C are more frequently found in

environmental samples collected in the Southern Hemisphere. We know that the different strains of *E. huxleyi* respond to acidification in diverse ways and that they might differ genetically (Langer *et al.*, 2009); it remains unclear which mechanism underlies the selection of a morphotype over the other in a specific region.

Physiological experiments need to be conducted in order to delineate *E. huxleyi* strain-specific response to altered pH conditions. A newly submitted study by Rickaby *et al.* (unpublished) on recently isolated, genetically diverse, strains of *E. huxleyi* collected in the Northern and Southern Hemispheres shows that the physiological properties of each strain is related to the natural environmental concentration of carbonate ion at the site of isolation. The projection of the study anticipates the future outcompetition of the morphotypes B/C and R by more rapidly photosynthesising, and lightly calcified strains of morphotype A but with their rate of calcification highly dependent on the surface ocean saturation state (Rickaby *et al*., unpublished).

The work carried out for this thesis showed that *E. huxleyi* blooms are characterised by one or two dominant genotypes, even if natural variability is found within its genetic standing stock. This matches with our finding regarding asexual reproduction driving a bloom event in the North Sea in 2011. Furthermore, this study brought light on *E. huxleyi* population dynamics in the Western English Channel over a longer time period. Four genotypes were detected: two of them found constantly over the years and able to produce blooming events (CMM I & IV), and the remaining two (CMM II & IVb), which were sporadically found during the cold seasons, and unable to form blooms, even if, in the past, bloom events dominated by *E. huxleyi*

morphotype B, accounting for CMM II, were recorded in the same region taken in exam during this study (van Bleijswijk *et al.*, 1991; Van der Wal *et al.*, 1995). The reasons underlying the selection of certain genotypes to thrive during the blooms are still unknown. We can hypothesize that zooplankton grazing, changes in nutrient availability, variation in salinity due to abundant precipitation and the rise in seawater temperature might be involved in the selection and not exclude each other.

E. huxleyi is a species-complex (Bendif *et al.*, 2015), composed of numerous genotypes, and given that the short and long-term survival of species depends in part on the options that their genetic diversity gives them for adapting to change (Hatton *et al.*, 2011), we can speculate that *E. huxleyi* sp. is likely to survive future ocean conditions. Nevertheless, there are different lines of evidence that suggest *E. huxleyi* is increasingly expanding its range into the polar oceans (Winter *et al.*, 2014). It has been predicted that one of the likely results of climate change is an alteration in the distribution and the abundance of species (Thomas *et al.*, 2004). A global meta-analysis, conducted on more than 1700 species, documented significant range shifts averaging 6.1 km per decade towards the poles (or metres per decade upward) for 279 species, and significant mean advancement of spring events by 2.3 days per decade (Parmesan *et al.*, 2003). This poses the question of whether both the two main morphotypes of *E. huxleyi* are globally shifting or if morphotype B specifically is moving toward the Polar Regions. *E. huxleyi* B and B/C morphotypes show a lower degree of calcification and these forms are prevalent in the Southern Hemisphere. Morphotype A is regularly found in the Northern Hemisphere and we found it in only one sample collected in a temperate region of the Southern Hemisphere. However, there is a large part of the Southern Hemisphere still unexplored for the different *E.*

huxleyi morphotypes and genotypes. Expeditions in the Southern Atlantic and Pacific Oceans would fill the gap and bring light on *E. huxleyi* community composition in these areas.

The so called "supergroups" redefined the organisation of the eukaryotic organisms at a high taxonomic level, and this was possible through the use of genomic data and the rise of phylogenomics (Liu *et al.*, 2010). Numerous genetic markers have been used to characterise *E. huxleyi* genetic delineation from its closest relative *G. oceanica*: the nuclear 18S rDNA and 28S rDNA, the plastidial 16S rDNA, *rbcL*, *tufA*, and *petA* and the mitochondrial *cox1*, *cox2*, *cox3*, *rpl16*, and *dam* (Bendif *et al.*, 2014; Liu *et al.*, 2010). *E. huxleyi* shows extensive genome variability and as much as 25% variability in gene content might occur between strains of the *E. huxleyi* species-complex (Bendif *et al.*, 2015; Read *et al.*, 2013). It is still unclear what are the phylogenetic links between *E. huxleyi* species and the lineages to which they belong and how many coccolithophore existed in the past that are not found today (de Vargas and Probert, 2004). We have to take in account that many evolutionary processes can affect the genetic diversity of natural populations, e.g. the Wahlund effect, the gene flow via migration, natural selection, inbreeding, spontaneous mutation, and random genetic drift (Hartl and Clark, 2007). Analysis of mutations of coccolithophores using population genetics, phylogenetic, and genomic approaches may help to answer some of these questions (de Vargas and Probert, 2004). Moreover, next generation sequencing (NGS) technologies have improved in the last few years: we can now rely on reads of sufficient length for accurate annotation and assembly of whole operons and the future studies will be aimed to experimental test the gene function (Temperton and Giovannoni, 2012). NGS technology has made high-

resolution biodiversity assessments of environmental samples possible (Hadziavdic *et al.*, 2014), using transcriptome or genomic data obtained through NGS to search for new genetic markers, which can develop a strategy to analyse the gene function.

Our analysis of environmental samples and laboratory clones showed a bias toward certain genotypes and morphotypes. This is reasonable, as it is shown in many studies, e.g. on bacteria (Russell, 2003), phytoplankton (Lewis and Hamm, 1986), invertebrates (Regoli and Principato, 1995), and mammals (Crabbe *et al.*, 1999), how difficult is to translate experimental findings to the environmental situations.

All these observations lead to the need of being cautious in making general statements about the physiology, genetic composition and behaviour of the organism exclusively using laboratory material, especially the ones with a fast turnover, even if substantial changes in community composition in less than a few hours are not expected (Fuhrman *et al.*, 2015).

The disappearance of *E. huxleyi* morphotype B from temperate waters and its movement toward northernmost latitudes could affect the community dynamics in those areas: B morphotype is, in fact, less calcified than A morphotype and grazers' diet might be affected and subsequently bigger predators' diet. It is still unknown if there are substantial differences in the physiology of different *E. huxleyi* genotypes, e.g. DMSP production, which has been shown to be metabolised by the grazer *Oxyrrhis marina* (Wolfe *et al.*, 1994). *O. marina* has also been proven to feed preferentially on *E. huxleyi* infected by its virus, *Eh*V-86 (Evans and Wilson, 2008), and although a correlation between different *E. huxleyi* viruses and different *E. huxleyi* morphotypes

and/or genotypes has not been demonstrated. in the study by Schroeder and his colleagues (2003) a diverse virus succession was found in the presence of a *E. huxleyi* bloom in a mesocosm experiment, suggesting that despite a genetic variation in the virus population, only few genotypes were able to infect and terminate the bloom. In the paper they leave an open question: whether or not the virus genetic-specificity is linked to a host genetic-specificity and the question is still without answer.

In the research work conducted for this thesis, *E. huxleyi* morphotype A shows a greater genetic variability than morphotypes B or B/C. Furthermore, heterozygosity was found in numerous clones characterised by A morphotype (CMM I-IV, I-IVb, IV-IVb, II-IV, III-IVb), but it was never found in presence of B/C morphotype. Might this mean that morphotype A more frequently reproduce sexually? Perhaps the Red Queen Hypothesis is relevant for morphotype A, and the Cheshire Cat Strategy is prevalent for morphotype B. Moreover, morphotype B is genetically characterised by homozygous CMM II (Krueger-Hadfield *et al.*, 2014), however the heterozygous CMM II-IV shows morphotype A characteristics. Does this mean that allele IV is dominant?

Since the calcite production is strictly correlated with cell-specific calcification rate and irradiance (Poulton *et al.*, 2014) a shift toward colder regions of *E. huxleyi* B and B/C morphotypes could have implications on the biogeochemistry of those waters, especially if these morphotypes are outcompeted by the stronger A morphotype in the formation of large blooms. In these areas the iron availability could also play an important role for the bloom formation. *E. huxleyi* species survival is not in danger in the light of the changing climate, nevertheless, the shift in genetic

composition may lead to shift in the food web dynamics and subsequently in how carbon and nutrients are metabolised in the ocean.

In order to answer some of the key questions that this study raises, e.g. what is the physiological response of different genotypes to the changing climate, what are the trophic effects due to the movement of certain morphotypes toward higher latitudes, whether and how the carbon pump will be affected by a possible disappearance of the morphotype B, and how different genotypes interact in the wild population,, we need in the first place more sampling efforts, especially in areas not yet studied for the genetic composition of *E. huxleyi* standing stock, and more genetic markers. The scientific community needs to address its laboratory studies on freshly collected samples and newly isolated clones and physiological variability should be determined *in situ*. Furthermore, it would be extremely valuable to acquire elucidation on the specific genetic adaptations, whether it exists, of particular strains studies relative to other coccolithophore species. Finally, we should look at studies of other phytoplanktonic species to determine whether the same principles of genetic diversity and distributions also apply to other organisms.

APPENDICES

APPENDIX I: Cruises & L4 Station

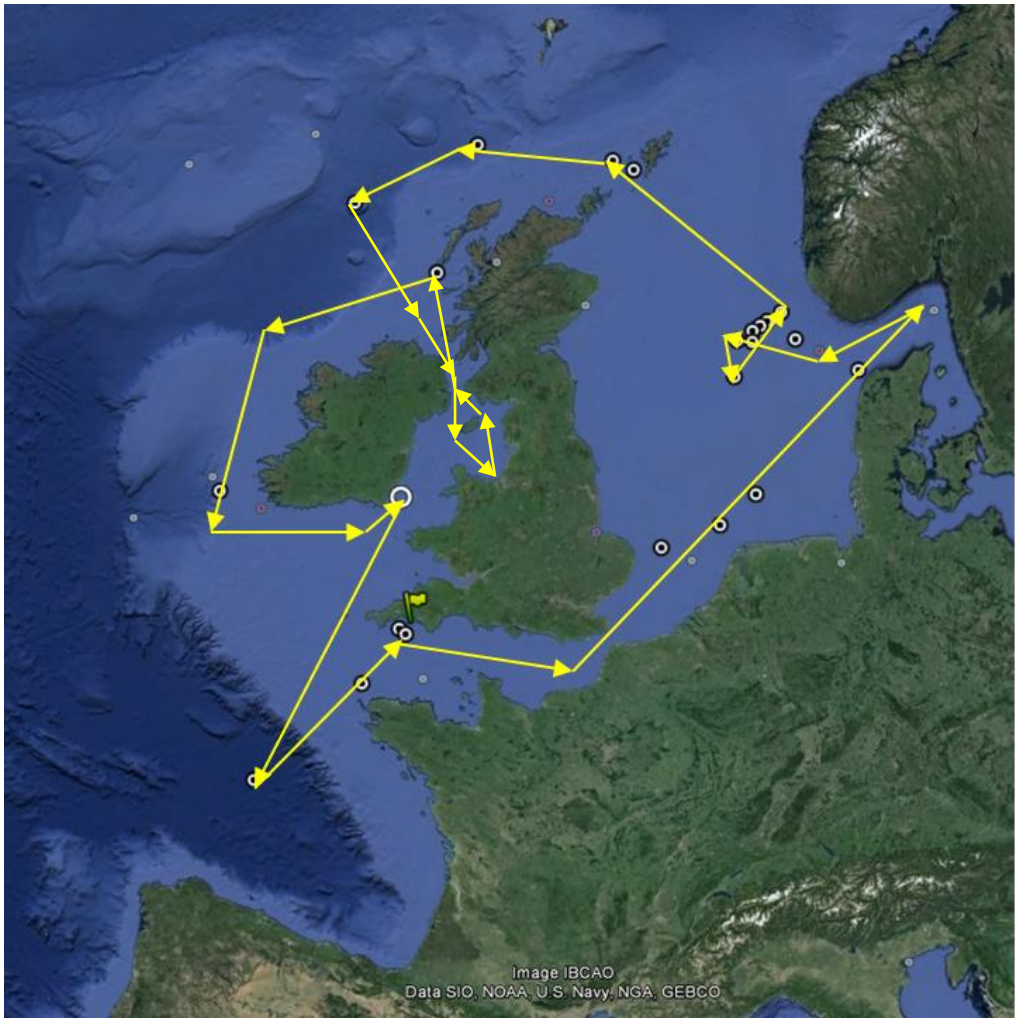
In order to assess the present genetic composition of *E. huxleyi* populations and interpret the new data in the light of its population structure in a changing climate we used different sampling approaches.

Cruises

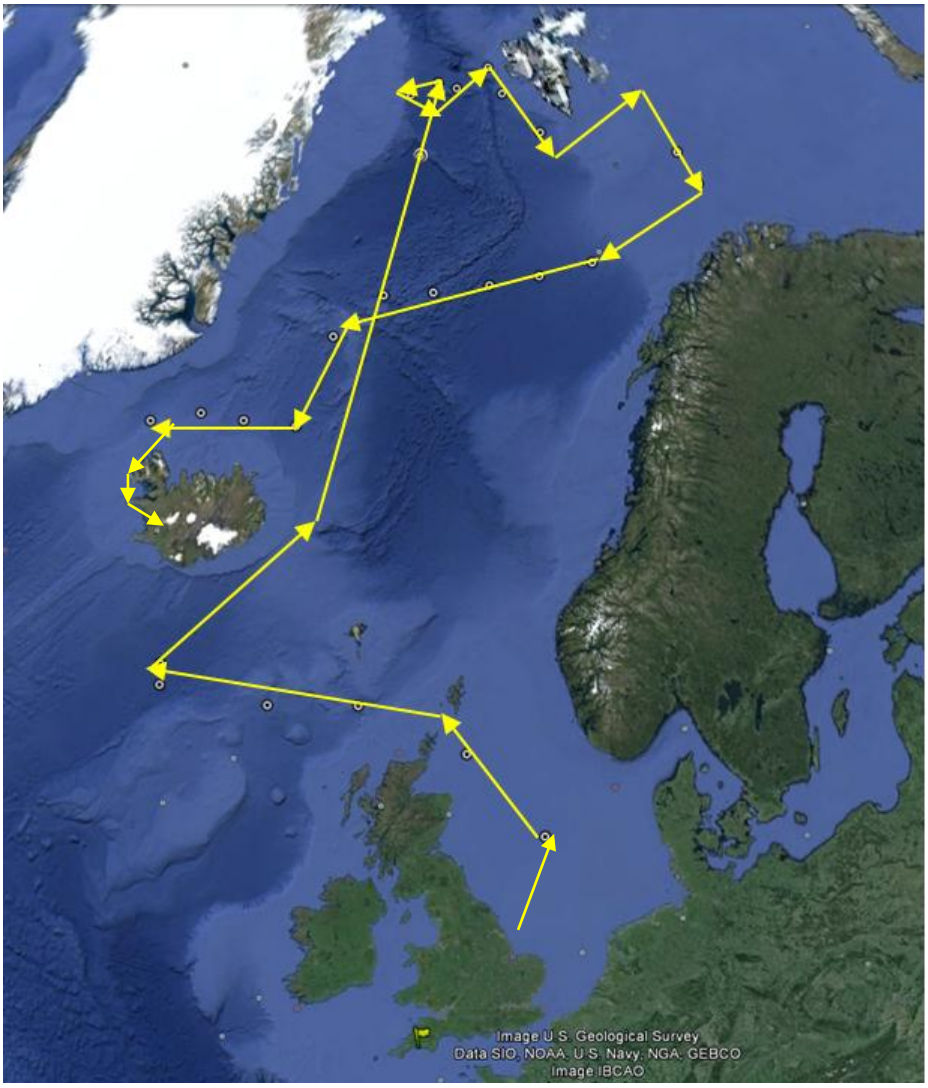
One fundamental resource for this research has been provided by the field work that I carried out in the Northern and in the Southern Hemispheres on board three scientific cruises. These campaigns allowed me to sample seawater from different geographical regions around the world, wide-ranging in latitude and longitude.

Two of these cruises were part of the UK Ocean Acidification (UKOA) research program, funded by the Natural Environment Research Council (NERC). The aim of these expeditions was to investigate the biology of surface ocean communities and biogeochemistry in areas known as CO_2 'sinks and sources' (www.surfaceoa.org.uk). Additionally, samples were collected along the second transect of the Great Southern Coccolithophore Belt, a campaign which aimed to examine several aspects of coccolithophore biology and the impact of short-term ocean acidification on coccolithophore growth and calcite dissolution (www.bco-dmo.org/project/473206). My participation in this cruise was funded by an UK Ocean Acidification Added Value Grant from NERC.

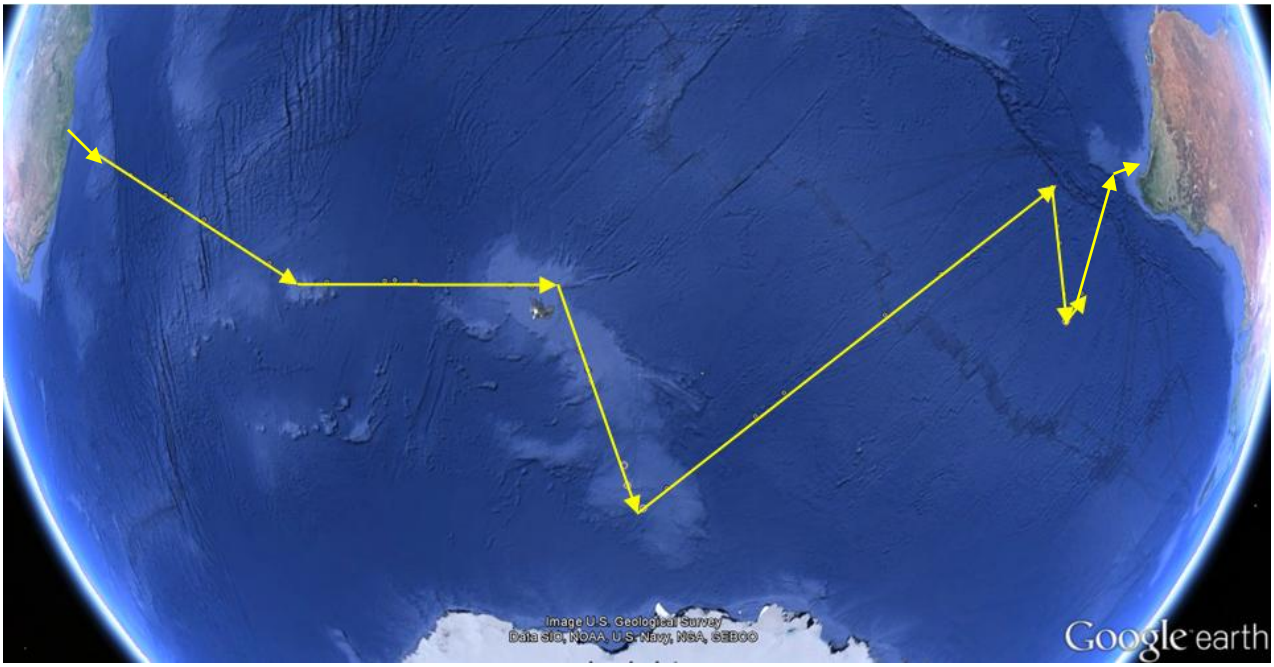
♦ **Cruise D366** (6th June 2011, Liverpool - 10th July 2011, Liverpool) on board the RRS Discovery (UK).



♦ **Cruise JR271** (1st June 2012, Immingham-UK - 3rd July 2012, Reykjavik) on board the RRS James Clark Ross (UK).



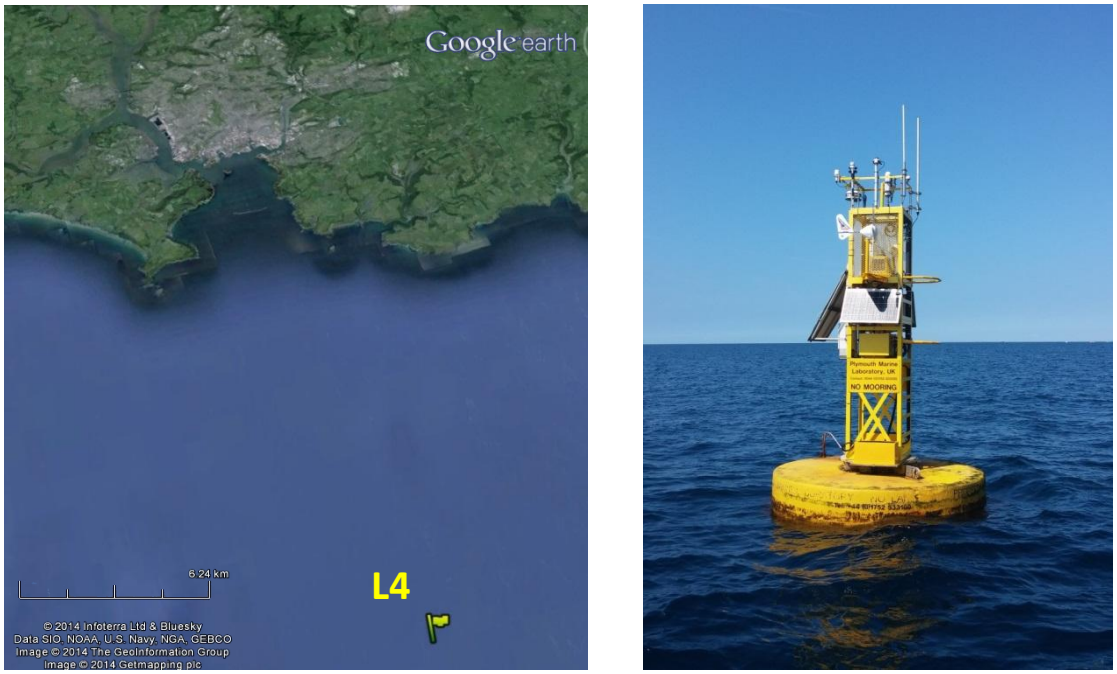
♦ **Cruise RR1202** (18th February 2012, Durban-South Africa - 23rd March 2012, Fremantle-Australia) on board the R/V Roger Revelle (USA).



DNA extraction from the environmental samples was carried out on board each cruise and the DNA samples were stored frozen. These samples collected *in-situ* provided an important biological resource, representing snapshots of *E. huxleyi* community structure at particular locations at present day CO_2 concentrations.

Western Channel Observatory (WCO-Plymouth)

WCO (Appendix Figure I) provided another important data resource. The WCO has some of the longest time-series in the world for zooplankton and phytoplankton, with data collected over a century (www.westernchannelobservatory.org.uk). This great collection supplied a large amount of environmental samples and it gave me the opportunity to study the *E. huxleyi* population composition over a 6 year time at the L4 station.



Appendix Figure I. L4 station of the Western Channel Observatory

APPENDIX II: Culture collection

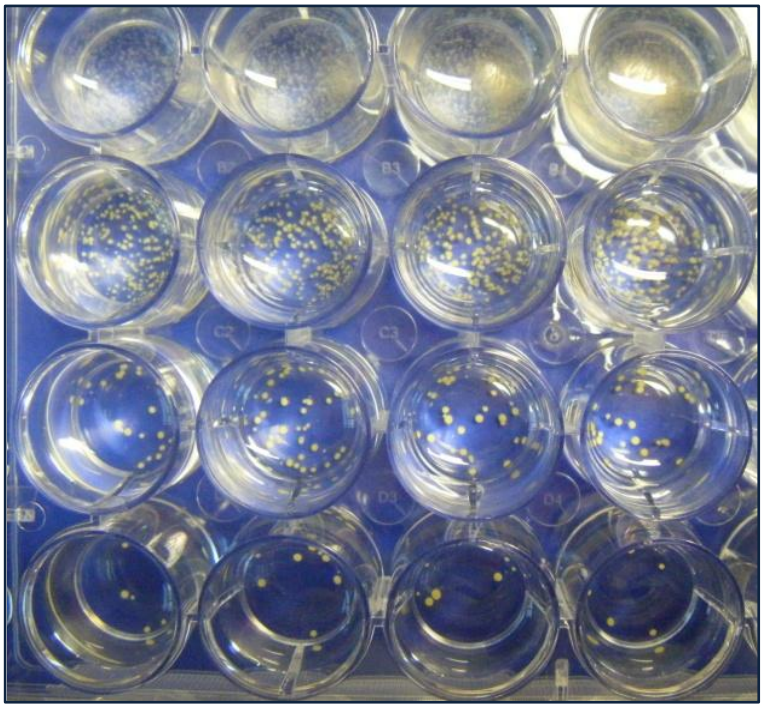
An algal culture collection was established (Appendix Figure II) from the water samples collected during the cruise work. Many culture vessels containing f/2 – Si (Guillard, 1975) culture medium were inoculated with the environmental water samples, maintained in an on board incubator and subsequently transported to the laboratory at the MBA in Plymouth (UK). Uni-algal cultures were obtained by screening the environmental water samples using a FACSORT flow cytometer (FACSORT, BD Biosciences, San Jose, CA, USA), which allowed for the selection of distinct features and characteristics associated with coccolithophores, diatoms and picoeukaryotes. The cells isolated in this way were transferred into culture vessels containing f/2 nutrient media and were incubated at 15°C at a light intensity between 40 and 55 µmol/m^2s with a 16:8 light:dark cycle. Dilution to extinction was carried

out (see the details in 'Box method') in order to obtain clonal uni-algal colonies and subsequently clonal cultures.

Appendix Figure II. Established culture collection of 292 clonal cultures.

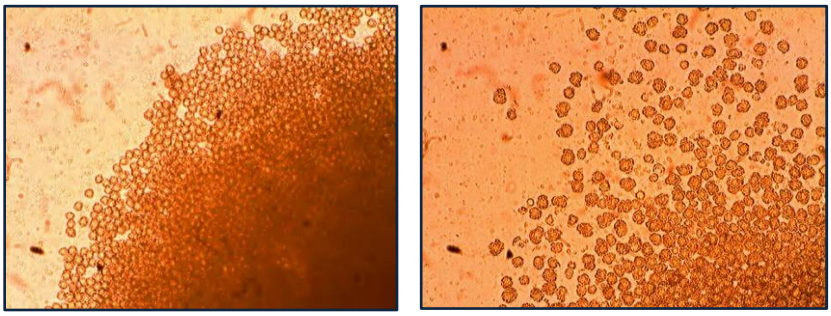
BOX METHOD

Dilution to extinction technique



Appendix Figure III. Plate with *E. huxleyi* colonies growing in each well at different concentrations. The top row contained thousands of colonies per well, the bottom row contained between 1 and 10 colonies per well.

The cell-concentration in each culture was calculated by flow cytometry. Subsequently each sample was transferred in the first row of a plate and consecutively diluted into a new well in order to obtain a low concentration of cells in the last row of the plate (Appendix Figure III). *E. huxleyi* calcifying cells are not motile and they sink through the water column. According to this principle, we allowed the cells to grow and to form colonies on the bottom of each well. Finally we picked up five colonies for each sample, trying to isolate colonies which showed visible differences under the light microscope (Appendix Figure IV), and we transferred the so-acquired cells into new vessels filled with f/2 nutrient media.



Appendix Figure IV. *E. huxleyi* colonies grown-up at the bottom of plate-wells. Although a first observation was made by optical microscopy and the resolution scale was not sufficiently detailed to investigate the coccolith morphologies, a difference in the cell size between the colonies was observable and it has been used in order to pick up different strains of *E. huxleyi* (microscope enlargement: 200X).

APPENDIX III: Conferences attended and publications

Relevant conferences

February **2015**, ASLO Meeting, Granada, Spain. *'Microbial diversity at the Western Channel Observatory as assessed by NGS'.*

September **2014**, The Biennial Challenger Conference for Marine Science, Plymouth University, UK. *'The dark side of the bloom'*.

September **2012**, The Third Symposium on The Ocean in a High-CO_2 World, Monterey, CA, U.S.A. *'Genetic and physiological plasticity within extant phytoplankton assemblages'.*

Publications derived from this study

Published

Genotyping an *Emiliania huxleyi* (prymnesiophyceae) bloom event in the North Sea reveals evidence of asexual reproduction
S. A. Krueger-Hadfield*****, C. Balestreri*****, J. Schroeder*****, A. Highfield, P. Helaouët, J. Allum, R. Moate, K. T. Lohbeck, P. I. Miller, U. Riebesell, T. B. H. Reusch, R. E. M. Rickaby, J. Young, G. Hallegraeff, C. Brownlee, and D. C. Schroeder Biogeosciences, 11, 5215-5234, **2014**. Impact Factor: 3.753

(* shared first authors, denotes equal contribution)

Submitted (ISME Journal)

Natural Selection in the Western English Channel: an increase in regional sea surface temperature is implicated in the differential survival of *Emiliania huxleyi* bloom forming morphotypes
Cecilia Balestreri, Anthony J. Richardson, Rosalind E. M. Rickaby, Tim J. Smyth, Colin Brownlee and Declan C. Schroeder

Completed (for future submission to ISME Journal)

***Emiliania huxleyi* genotypes occupy distinct biogeographic provinces**
Cecilia Balestreri, Sally Thorpe, Rosalind E. M. Rickaby, William Balch, Colin Brownlee, and Declan C. Schroeder

REFERENCES

Aagaard K, Coachman L. (1968). The East Greenland Current north of Denmark Strait: part I. *Arctic* **21**: 181–200.

Allison S, Martiny J. (2008). Resistance, resilience, and redundancy in microbial communities. *Proc Natl Acad Sci U S A* **105**: 11512–11519.

Alpermann TJ, Beszteri B, John U, Tillmann U, Cembella AD. (2009). Implications of life-history transitions on the population genetic structure of the toxigenic marine dinoflagellate Alexandrium tamarense. *Mol Ecol* **18**: 2122–2133.

Anderson DT. (1994). Barnacles: structure, function, development and evolution. London: Chapman & Hall. p.357

Arnaud-Haond S, Belkhir K. (2006). GENECLONE: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. *Mol Ecol Notes* **7**: 15–17.

Arrigo KR, Robinson DH, Worthen DL, Dunbar RB, DiTullio GR, VanWoert M, *et al.* (1999). Phytoplankton community structure and the drawdown of nutrients and CO_2 in the Southern Ocean. *Science (80-)* **283**: 365–367.

Balch W, Utgoff P. (2009). Potential interactions among ocean acidification, coccolithophores, and the optical properties of seawater. *Oceanography* **22**: 146–159.

Balch WM. (2004). Using new techniques for re-evaluating the physiological ecology of coccolithophores. In: *Coccolithophores. From molecular processes to global impact*. pp 1–23.

Balch WM, Drapeau DT, Bowler BC, Booth ES, Goes JI, Ashe a., *et al.* (2004). A multi-year record of hydrographic and bio-optical properties in the Gulf of Maine: I. Spatial and temporal variability. *Prog Oceanogr* **63**: 57–98.

Balch WM, Drapeau DT, Fritz JJ. (2000). Monsoonal forcing of calcification in the Arabian Sea. *Deep Res Part II Top Stud Oceanogr* **47**: 1301–1337.

Balch WM, Drapeau DT, Fritz JJ, Bowler BC, Nolan J. (2001). Optical backscattering in the Arabian Sea - Continuous underway measurements of particulate inorganic and organic carbon. *Deep Res Part I Oceanogr Res Pap* **48**: 2423–2452.

Balch WM, Kilpatrick K. (1996). Calcification rates in the equatorial Pacific along 140°W. *Deep Res Part II* **43**: 971–993.

Balestreri C, Richardson AJ, Rickaby REM, Smyth TJ, Brownlee C, Schroeder DC (submitted for publication - 2015). Natural selection in the Western English Channel: an increase in regional sea surface temperature is implicated in the differential survival of *Emiliania huxleyi* bloom forming morphotypes.

Barcelos e Ramos J, Müller MN, Riebesell U. (2010). Short-term response of the coccolithophore *Emiliania huxleyi* to an abrupt change in seawater carbon dioxide

concentrations. *Biogeosciences* **7**: 177–186.

Barnes RD. (1963). Invertebrate zoology. Philadelphia: Saunders. p.632

Baross J a., Hoffman SE. (1985). Submarine hydrothermal vents and associated gradient environments as sites for the origin and evolution of life. *Orig Life Evol Biosph* **15**: 327–345.

Bates NR, Michaels AF, Knap AH. (1996). Alkalinity changes in the Sargasso Sea: geochemical evidence of calcification? *Mar Chem* **51**: 347–358.

Bäuerlein E, Behrens P, Epple M, Pickett-Heaps J, Mann S, Pompe W. (2009). Handbook of biomineralization: biomimetic and bioinspired chemistry.

Beardall J, Raven JA. (2013). Calcification and ocean acidification: new insights from the coccolithophore *Emiliania huxleyi*. *New Phytol* **199**: 1–3.

Beaufort L, Probert I, de Garidel-Thoron T, Bendif EM, Ruiz-Pino D, Metzl N, *et al.* (2011). Sensitivity of coccolithophores to carbonate chemistry and ocean acidification. *Nature* **476**: 80–3.

Bell M V, Pond D. (1996). Lipid composition during growth of motile and coccolith forms of *Emiliania huxleyi*. *Phytochemistry* **41**: 465–471.

Bellerby R, Olsen A, Furevik T, Anderson L. (2005). Response of the surface ocean CO_2 system in the Nordic Seas and North Atlantic to climate change. In: *Climate Variability in the Nordic Seas: Geophysical Monograph Series*. pp 189–198.

Bendif E, Al-Kandari M, Probert I, Schroeder DC. (2013). Microalgae systematics: can we still fit a square peg in a round hole when new evidence says it is a triangle? *Mar Biol* **1**: 15–17.

Bendif EM, Probert I, Carmichael M, Romac S, Hagino K, de Vargas C. (2014). Genetic delineation between and within the widespread coccolithophore morpho-species *Emiliania huxleyi* and *Gephyrocapsa oceanica* (Haptophyta) Mock T (ed). *J Phycol* **50**: 140–148.

Bendif EM, Probert I, Young JR, von Dassow P. (2015). Morphological and phylogenetic characterization of new *Gephyrocapsa* isolates suggests introgressive hybridization in the *Emiliania*/*Gephyrocapsa* complex (Haptophyta). *Protist* **166**: 323-336.

Berge G. (1962). Discoloration of the sea due to *Coccolithus huxleyi* 'bloom'. *Sarsia* **6**: 27–40.

Berner R a. (1990). Atmospheric carbon dioxide levels over phanerozoic time. *Science (80-)* **249**: 1382–6.

Billard C. (1994). Life cycles. In: *The Haptophyte Algae* (J.C. Green and B.S.C. Leadbeater, eds). Oxford: Clarendon Press. p.464

Billard C, Inouye I. (2004). What is new in coccolithophore biology? In: *Coccolithophores. From Molecular Processes to Global Impact*. pp 1–29.

Boissinot S, Boursot P. (1997). Discordant phylogeographic patterns between the Y chromosome and mitochondrial DNA in the house mouse: selection on the Y chromosome? *Genetics* **146**: 1019–34.

Bouchet P, Rocroi J-P, Frýda J, Hausdorf B, Ponder W, Valdés Á, *et al.* (2005). Classification and nomenclator of gastropod families. *Malacol Int J Malacol* **47**: 1–397.

Bourne GC. (1900). The Anthozoa. In: *A Treatise on Zoology* (Lankester, E. Ray Sir, ed). London: Adam and Charles Black, pp 1–84.

Brand LE. (1981). Genetic variability in reproduction rates in marine phytoplankton populations. *Evolution (N Y)* **35**: 1117–1127.

Brown C, Podestá G. (1997). Remote sensing of coccolithophore blooms in the western South Atlantic Ocean. *Remote Sens Environ* **4257**: 83–91.

Brown C, Yoder J. (1994). Coccolithophorid blooms in the global ocean. *J Geophys Res* **99**: 7467–7482.

Brownlee C, Taylor AR. (2002). Algal calcification and silification. In: *Encyclopedia of Life Sciences*. London: Nature Publishing Group. pp 1–6.

Bruvo R, Michiels NK, D'Souza TG, Schulenburg H. (2004). A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Mol Ecol* **13**: 2101–2106.

Butchart SHM, Walpole M, Collen B, van Strien A, Scharlemann JPW, Almond REA, *et al.* (2010). Global biodiversity: indicators of recent declines. *Science (80-)* **328**: 1164–1168.

Campbell L, Shapiro LP, Haugen EM, Morris L. (1989). Immunochemical approaches to the identification of the ultraplankton: assets and limitations. *Coast Estuar Stud Vol* **35**: 39–56.

Caporaso JG, Paszkiewicz K, Field D, Knight R, Gilbert JA. (2012). The Western English Channel contains a persistent microbial seed bank. *ISME J* **6**: 1089–93.

Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings C, Venail P, *et al.* (2012). Biodiversity loss and its impact on humanity. *Nature* **486**: 59–67.

Casabianca S, Penna A, Pecchioli E, Jordi A, Basterretxea G, Vernesi C. (2012). Population genetic structure and connectivity of the harmful dinoflagellate *Alexandrium minutum* in the Mediterranean Sea. *Proc Biol Sci* **279**: 129–38.

Charlson RJ, Lovelock JE, Andreae MO, Warren SG. (1987). Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature* **326**: 655 – 661.

Chepurnov VA, Mann DG, Sabbe K, Vannerum K, Casteleyn G, Verleyen E, *et al.* (2005). Sexual reproduction, mating system, chloroplast dynamics and abrupt cell size reduction in *Pseudo-nitzschia pungens* from the North Sea (Bacillariophyta). *Eur J Phycol* **40**: 379–395.

Clark L V., Jasieniuk M. (2011). POLYSAT: an R package for polyploid microsatellite

analysis. *Mol Ecol Resour* **11**: 562–566.

Conte MH, Thompson A, Lesley D, Harris RP. (1998). Genetic and physiological influences on the alkenone/alkenoate versus growth temperature relationship in *Emiliania huxleyi* and *Gephyrocapsa oceanica*. *Geochim Cosmochim Acta* **62**: 51–68.

Cook SS, Jones RC, Vaillancourt RE, Hallegraeff GM. (2013). Genetic differentiation among Australian and Southern Ocean populations of the ubiquitous coccolithophore *Emiliania huxleyi* (Haptophyta). *Phycologia* **52**: 368–374.

Cook SS, Whittock L, Wright SW, Hallegraeff GM. (2011). Photosynthetic pigment and genetic differences between two Southern Ocean morphotypes of *Emiliania huxleyi* (Haptophyta)1. *J Phycol* **47**: 615–626.

Corstjens PL, van der Kooij A, Linschooten C, Brouwers G-J, Westbroek P, Jong EWDV. (1998). Gpa, a calcium-binding protein in the coccolithophorid *Emiliania huxleyi* (Prymnesiophyceae). *J Phycol* **34**: 622–630.

Cotner JB, Biddanda B a. (2002). Small players, large role: microbial influence on biogeochemical processes in pelagic aquatic ecosystems. *Ecosystems* **5**: 105–121.

Coxall HK, Wilson PA, Pälike H, Lear CH, Backman J. (2005). Rapid stepwise onset of Antarctic glaciation and deeper calcite compensation in the Pacific Ocean. *Nature* **433**: 53–57.

Crabbe JC, Wahlsten D, Dudek BC. (1999). Genetics of mouse behavior: interactions with laboratory environment. *Science (80-)* **284**: 1670–1672.

Cros L, Fortuño JM, Estrada M. (2013). Elemental composition of coccoliths: Mg/Ca relationships. *Sci Mar* **77**: 63–67.

Cubasch U, Meehl GA, Boer GJ, Stouffer RJ, Dix M, Noda A, *et al.* (2001). Projections of future climate change. In: Kim J-W, Stone J (eds). *IPCC Third Assessment Report; Climate Change 2001- Working Group I: The Scientific Basis*. pp 526–582.

Cubillos J, Wright S, Nash G, de Salas M, Griffiths B, Tilbrook B, *et al.* (2007). Calcification morphotypes of the coccolithophorid *Emiliania huxleyi* in the Southern Ocean: changes in 2001 to 2006 compared to historical data. *Mar Ecol Prog Ser* **348**: 47–54.

Cutter AD, Jovelin R, Dey A. (2013). Molecular hyperdiversity and evolution in very large populations. *Mol Ecol* **22**: 2074–2095.

Deacon G. (1982). Physical and biological zonation in the Southern Ocean. *Deep Sea Res* **29**: 1–15.

Delworth TL, Manabe S, Stouffer RJ. (1997). Multidecadal climate variability in the Greenland Sea and surrounding regions: a coupled model simulation. *Geophys Res Lett* **24**: 257–260.

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

Doney SC, Fabry VJ, Feely R a, Kleypas J a. (2009). Ocean acidification: the other CO_2 problem. *Ann Rev Mar Sci* **1**: 169–192.

Donlon CJ, Martin M, Stark J, Roberts-Jones J, Fiedler E, Wimmer W. (2012). The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. *Remote Sens Environ* **116**: 140–158.

Dorken ME, Eckert CG. (2001). Severely reduced sexual reproduction in northern populations of a clonal plant, *Decodon verticillatus* (Lythraceae). *J Ecol* **89**: 339–350.

Dunthorn M, Klier J, Bunge J, Stoeck T. (2012). Comparing the hyper-variable V4 and V9 regions of the small subunit rDNA for assessment of ciliate environmental diversity. *J Eukaryot Microbiol* **59**: 185–187.

Edwards M, Richardson AJ. (2004). Impact of climate change on marine pelagic phenology and trophic mismatch. *Nature* **430**: 881–4.

Eglinton TI, Conte MH, Eglinton G, Hayes JM. (2001). Proceedings of a workshop on alkenone-based paleoceanographic indicators. *Geochemistry Geophys Geosystems* **2**.

Ellstrand NC, Roose ML. (1987). Patterns of genotypic diversity in clonal plant species. *Am J Bot* **74**: 123–131.

Erdner DL, Richlen M, McCauley LAR, Anderson DM. (2011). Diversity and dynamics of a widespread bloom of the toxic dinoflagellate *Alexandrium fundyense*. *PLoS One* **6**: e22965.

Evans C, Wilson WH. (2008). Preferential grazing of *Oxyrrhis marina* on virus infected *Emiliania huxleyi*. *Limnol Oceanogr* **53**: 2035–2040.

Evans KM, Kühn SF, Hayes PK. (2005). High levels of genetic diversity and low levels of genetic differentiation in North Sea *Pseudo-nitzschia pungens* (Bacillariophyceae) populations. *J Phycol* **41**: 506–514.

Excoffier L, Smouse PE, Quattro JM. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479–91.

Fabry VJ, Seibel BA, Feely RA, Orr JC. (2008). Impacts of ocean acidification on marine fauna and ecosystem processes. *ICES J Mar Sci* **65**: 414–432.

Falkowski PG, Barber RT, Smetacek V. (1998). Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science (80-)* **281**: 200–206.

Feng Y, Hare C, Leblanc K, Rose J, Zhang Y, DiTullio G, *et al.* (2009). Effects of increased pCO_2 and temperature on the North Atlantic spring bloom. The phytoplankton community and biogeochemical response. *Mar Ecol Prog Ser* **388**: 13–25.

Fernandez, E. Boyd, P. Holligan, P. Harbour D. (1993). Production of organic and inorganic carbon within a large-scale coccolithophore bloom in the northeast Atlantic Ocean. *Mar Ecol Prog Ser* **97**: 271–285.

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science (80-)* **281**: 237–240.

Fielding SR. (2013). *Emiliania huxleyi* specific growth rate dependence on temperature. *Limnol Oceanogr* **58**: 663–666.

Fileman ES, Cummings DG, Llewellyn CA. (2002). Microplankton community structure and the impact of microzooplankton grazing during an *Emiliania huxleyi* bloom, off the Devon coast. *J Mar Biol Assoc UK* **82**: 359–368.

Findlay H. (2010). Ocean acidification in the Arctic : a summary from current observations. Report of The Catlin Arctic Survey 2010.

Findlay HS, Calosi P, Crawfurd KJ. (2011). Determinants of the PIC:POC response in the coccolithophore *Emiliania huxleyi* under future ocean acidification scenarios. *Limnol Oceanogr* **56**: 1168–1178.

Fookes CJR, Jeffrey SW. (1989). The structure of chlorophyll c3, a novel marine photosynthetic pigment. *J Chem Soc Chem Commun* 1827.

Frada M, Probert I, Allen MJ, Wilson WH, de Vargas C. (2008). The 'Cheshire Cat' escape strategy of the coccolithophore *Emiliania huxleyi* in response to viral infection. *Proc Natl Acad Sci* **105**: 15944–15949.

Frankham R, Briscoe DA, Ballou JD. (2002). Introduction to conservation genetics. Cambridge: Cambridge University Press. p.617

Free A, Barton NH. (2007). Do evolution and ecology need the Gaia hypothesis? *Trends Ecol Evol* **22**: 611–619.

Fuhrman J a., Cram J a., Needham DM. (2015). Marine microbial community dynamics and their ecological interpretation. *Nat Rev Microbiol* 1–14.

Fujiwara S, Tsuzuki M, Kawachi M, Minaka N, Inouye I. (2001). Molecular Phylogeny of the Haptophyta based on the rbcL gene and sequence variation in the spacer region of the RUBISCO operon. *J Phycol* **37**: 121–129.

Garcia-Soto C, Fernandez E. (1995). Evolution and structure of a shelf coccolithophore bloom in the Western English Channel. *J Plankton Res* **17**: 2011–2036.

Garrido JL, Zapata M. (1998). Detection of new pigments from *Emiliania huxleyi* (Prymnesiophyceae) by high herformance liquid chromatography, liquid chromatography-mass spectrometry, visible spectroscopy, and fast atom bombardment mass spectrometry. *J Phycol* **34**: 70–78.

Garrido JL, Zapata M, Muñiz S. (1995). Spectral characterization of new chlorophyll c pigments isolated from *Emiliania Huxleyi* (Prymnesiophyceae) by High-Performance Liquid Chromatography. *J Phycol* **31**: 761–768.

Gayoso A. (1995). Bloom of *Emiliania huxleyi* (Prymnesiophyceae) in the western South Atlantic Ocean. *J Plankton Res* **17**: 1717–1722.

Gilbert JA, Field D, Swift P, Newbold L, Oliver A, Smyth T, *et al.* (2009). The seasonal structure of microbial communities in the Western English Channel. *Environ Microbiol* **11**: 3132–9.

Gilbert JA, Steele JA, Caporaso JG, Steinbrück L, Reeder J, Temperton B, *et al.* (2012). Defining seasonal marine microbial community dynamics. *ISME J* **6**: 298–308.

Gille ST. (2002). Warming of the Southern Ocean since the 1950s. *Science (80-)* **295**: 1275–1277.

Green JC, Course PA, Tarran GA. (1996). The life-cycle of *Emiliania huxleyi*: a brief review and a study of relative ploidy levels analysed by flow cytometry. *J Mar Syst* **9**: 33–44.

Groom SB, Holligan PM. (1987). Remote sensing of coccolithophore blooms. *Adv Sp Res* **7**: 73–78.

Grosell M. (2014). Calcium carbonate contributors. In: *International Innovation*. pp 90-92

Grossi V, Raphel D, Aubert C, Rontani JF. (2000). The effect of growth temperature on the long-chain alkenes composition in the marine coccolithophorid *Emiliania huxleyi*. *Phytochemistry* **54**: 393–9.

Guillard RRL. (1975). Culture of phytoplankton for feeding marine invertebrates. In: *Culture of Marine Invertebrate Animals*. pp 29–60.

Guinotte JM, Fabry VJ. (2008). Ocean acidification and its potential effects on marine ecosystems. *Ann N Y Acad Sci* **1134**: 320–42.

Hadziavdic K, Lekang K, Lanzen A, Jonassen I, Thompson EM, Troedsson C. (2014). Characterization of the 18s rRNA gene for designing universal eukaryote specific primers. *PLoS One* **9**.

Hagino K, Bendif EM, Young JR, Kogame K, Probert I, Takano Y, *et al.* (2011). New evidence for morphological and genetic variation in the cosmopolitan coccolithophore *Emiliania huxleyi* (Prymnesiophyceae) from the cox1b-ATP4 GENES. *J Phycol* **47**: 1164–1176.

Hagino K, Okada H, Matsuoka H. (2005). Coccolithophore assemblages and morphotypes of *Emiliania huxleyi* in the boundary zone between the cold Oyashio and warm Kuroshio currents off the coast of Japan. *Mar Micropaleontol* **55**: 19–47.

Halkett F, Simon J, Balloux F. (2005). Tackling the population genetics of clonal and partially clonal organisms. *Trends Ecol Evol* **20**: 194–201.

Hansen J, Ruedy R, Sato M, Lo K. (2010). Global surface temperature change. *Rev Geophys* **48**: RG4004.

Hansen J, Sato M, Ruedy R, Lo K, Lea DW, Medina-Elizade M. (2006). Global temperature change. *Proc Natl Acad Sci* **103**: 14288–14293.

Harper P, Joy R, Kanowski P, Mackay R, McKenzie N, Ward T. (2011). Chapter 7:

Antarctic Environment. In: *Independent report to the Australian Government Minister for Sustainability, Environment, Water, Population and Communities.* Canberra. pp. 466-565

Hartl DL, Clark AG. (2007). Principles of Population Genetics. Sunderland: Sinauer Associates, Inc. Publishers. p.488

Hatton TJ, Cork S, Harper P, Joy R, Kanowski P, Mackay R, *et al.* (2011). Chapter 8: Biodiversity. In: *Independent report to the Australian Government Minister for Sustainability, Environment, Water, Population and Communities .* Canberra. pp. 573-689

Haxo FT. (1985). Photosynthetic action spectrum of the coccolithophorid, *Emiliania huxleyi* (haptophyceae): 19'hexanoyloxyfucoxanthin as antenna pigment. *J Phycol* **21**: 282–287.

Head M. (1996). Chapter 30. Modern dinoflagellate cysts and their biological affinities. In: *Palynology: Principles and Applications*. American Association of Stratigraphic Palynologists Foundation, pp 1197–1248.

Helaouët P, Beaugrand G, Reid PC. (2011). Macrophysiology of *Calanus finmarchicus* in the North Atlantic Ocean. *Prog Oceanogr* **91**: 217–228.

Highfield A, Evans C, Walne A, Miller PI, Schroeder DC. (2014). How many Coccolithovirus genotypes does it take to terminate an *Emiliania huxleyi* bloom? *Virology* **466-467**: 138–145.

Hine N, Weaver PPE. (1998). Quaternary. In: *Calcareous Nannofossil Biostratigraphy* (Bown PR, ed). Springer - British Micropalaeontological Society Publications Series. pp 266–283.

Hinz DJ. (2010). *Emiliania huxleyi* and climate change: a genetic and biogeographic investigation of bloom dynamics for a key phytoplankton species in the global carbon cycle. PhD Thesis. University of Southampton.

Holligan PM. (1992). Do marine phytoplankton influence global climate? In: *Primary Productivity and Biogeochemical Cycles in the Sea* (Falkowski PG, Woodhead AD, Vivirito K, eds). Springer US, pp 487–501.

Holligan PM, Charalampopoulou a., Hutson R. (2010). Seasonal distributions of the coccolithophore, *Emiliania huxleyi*, and of particulate inorganic carbon in surface waters of the Scotia Sea. *J Mar Syst* **82**: 195–205.

Holligan PM, Fernández E, Aiken J, Balch WM, Boyd P, Burkill PH, *et al.* (1993). A biogeochemical study of the coccolithophore, *Emiliania huxleyi*, in the North Atlantic. *Global Biogeochem Cycles* **7**: 879–900.

Holligan PM, Viollier M, Harbour DS, Camus P, Champagne-Philippe M. (1983). Satellite and ship studies of coccolithophore production along a continental shelf edge. *Nature* **304**: 339–342.

Houdan A, Probert I, Van Lenning K, Lefebvre S. (2005). Comparison of photosynthetic responses in diploid and haploid life-cycle phases of *Emiliania huxleyi*

(Prymnesiophyceae). *Mar Ecol Prog Ser* **292**: 139–146.

Iglesias-Rodriguez DM, Schofield OM, Batley J, Medlin LK, Hayes PK. (2006). Intraspecific genetic diversity in the marine coccolithophore *Emiliania huxleyi* (Prymnesiophyceae): the use of microsatellite analysis in marine phytoplankton population studies. *J Phycol* **42**: 526–536.

Iglesias-Rodriguez MD, Halloran PR, Rickaby REM, Hall IR, Colmenero-Hidalgo E, Gittins JR, *et al.* (2008). Phytoplankton Calcification in a High-CO_2 World. *Science (80-)* **320**.

Iglesias-Rodríguez MD, Probert I, Batley J. (2006). Microsatellite cross-amplification in coccolithophores: application in population diversity studies. *Hereditas* **143**: 99–102.

Iglesias-Rodriguez MD, Saez AG, Groben R, Edwards KJ, Batley J, Medlin LK, *et al.* (2002). Polymorphic microsatellite loci in global populations of the marine coccolithophorid *Emiliania huxleyi*. *Mol Ecol Notes* **2**: 495–497.

Jaenike J. (1978). A hypothesis to account for the maintenance of sex within populations. *Evol Theory* **3**: 191–194.

Jarne P, Lagoda P. (1996). Microsatellites, from molecules to populations and back. *Trends Ecol Evol* **11**.

Johannessen OM, Bengtsson L, Miles MW, Kuzmina SI, Semenov VA, Alekseev G V., *et al.* (2004). Arctic climate change: observed and modelled temperature and sea-ice variability. *Tellus* **56A**: 328–341.

Kalinowski ST. (2005). HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes* **5**: 187–189.

Kalinowski ST, Taper ML. (2006). Maximum likelihood estimation of the frequency of null alleles at microsatellite loci. *Conserv Genet* **7**: 991–995.

Kennedy WJ, Taylor JD, Hall A. (1969). Environmental and biological controls on bivalve shell mineralogy. *Biol Rev* **44**: 499–530.

Kennett JP, Srinivasan MS. (1983). Neogene planktonic foraminifera: a phylogenetic atlas. Stroudsburg: Hutchinson Ross Publishing Company. pp.265

Kleypas JA, Feely RA, Fabry VJ, Langdon C, Sabine CL, Robbins LL. (2006). Impacts of ocean acidification on coral reefs and other marine calcifiers : a guide for future research, report of a workshop held 18-20 April 2005, St Petersburg, FL, sponsored by NSF, NOAA, and the U.S. Geological Survey.

Kraay GW, Zapata M, Veldhuis MJW. (1992). Separations of chlorophylls c1, c2, and c3 of marine phytoplankton by reversed-phase-c18-high-performance liquid chromatography. *J Phycol* **28**: 708–712.

Kroeker KJ, Kordas RL, Crim R, Hendriks IE, Ramajo L, Singh GS, *et al.* (2013). Impacts of ocean acidification on marine organisms: quantifying sensitivities and interaction with warming. *Glob Chang Biol* **19**: 1884–1896.

Kroeker KJ, Kordas RL, Crim RN, Singh GG. (2010). Meta-analysis reveals negative yet variable effects of ocean acidification on marine organisms. *Ecol Lett* **13**: 1419–1434.

Krueger-Hadfield SA, Balestreri C, Schroeder J, Highfield A, Helaouët P, Allum J, *et al.* (2014). Genotyping an *Emiliania huxleyi* (prymnesiophyceae) bloom event in the North Sea reveals evidence of asexual reproduction. *Biogeosciences* **11**: 5215–5234.

Krueger-Hadfield SA, Collén J, Daguin-Thiébaut C, Valero M. (2011). Genetic population structure and mating system in *Chondrus crispus* (Rhodophyta). *J Phycol* **47**: 440–450.

Krueger-Hadfield SA, Roze D, Mauger S, Valero M. (2013). Intergametophytic selfing and microgeographic genetic structure shape populations of the intertidal red seaweed *Chondrus crispus*. *Mol Ecol* **22**: 3242–3260.

Kwok R. (2009). Fish are crucial in oceanic carbon cycle. *Nature News*. Published online 15 January 2009

Lakeman MB, von Dassow P, Cattolico RA. (2009). The strain concept in phytoplankton ecology. *Harmful Algae* **8**: 746–758.

Langer G, Nehrke G, Probert I, Ly J, Ziveri P. (2009). Strain-specific responses of *Emiliania huxleyi* to changing seawater carbonate chemistry. *Biogeosciences* **6**: 2637–2646.

Leung TLF, King KC, Wolinska J. (2012). Escape from the Red Queen: an overlooked scenario in coevolutionary studies. *Oikos* **121**: 641–645.

Lewis MA, Hamm BG. (1986). Environmental modification of the photosynthetic response of lake plankton to surfactants and significance to a laboratory-field comparison. *Water Res* **20**: 1575–1582.

Li Y-C, Korol AB, Fahima T, Beiles A, Nevo E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol* **11**: 2453–2465.

Lin X, Hull CM, Heitman J. (2005). Sexual reproduction between partners of the same mating type in *Cryptococcus neoformans*. *Nature* **434**: 1017–1021.

Liu H, Aris-Brosou S, Probert I, De Vargas C. (2010). A time line of the environmental genetics of the haptophytes. *Mol Biol Evol* **27**: 161–176.

Liu H, Probert I, Uitz J, Claustre H, Aris-Brosou S, Frada M, *et al.* (2009). Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proc Natl Acad Sci* **106**: 12803–12808.

Lohbeck KT, Riebesell U, Collins S, Reusch TBH. (2013). Functional genetic divergence in high CO_2 adapted *Emiliania huxleyi* populations. *Evolution (N Y)* **67**: 1892–1900.

Lohbeck KT, Riebesell U, Reusch TBH. (2012a). Adaptive evolution of a key phytoplankton species to ocean acidification. *Nat Geosci* **5**: 346–351.

Lohbeck KT, Riebesell U, Reusch TBH. (2012b). Adaptive evolution of a key

phytoplankton species to ocean acidification. *Nat Geosci* **5**: 346–351.

Loreau M, Downing A, Emmerson M, Gonzalez A, Hughes J, Inchausti P, *et al.* (2002). A new look at the relationship between diversity and stability. In: *Biodiversity and ecosystem functioning: synthesis and perspectives* (Loreau M, Naeem S, Inchausti P, eds). New York: Oxford University Press Inc. pp 79–91.

Lutjeharms J, Ansorge I. (2001). The Agulhas return current. *J Mar Syst* 115–138.

Mackinder L, Bach L, Schulz K, Wheeler G, Schroeder D, Riebesell U, *et al.* (2011a). The molecular basis of inorganic carbon uptake mechanisms in the coccolithophore *Emiliania huxleyi*. *Eur J Phycol* **46**: 142–143.

Mackinder L, Wheeler G, Schroeder D, von Dassow P, Riebesell U, Brownlee C. (2011b). Expression of biomineralization-related ion transport genes in *Emiliania huxleyi*. *Environ Microbiol* **13**: 3250–3265.

Malin G. (1997). Sulphur, climate and the microbial maze. *Nature* **387**: 857–859.

Mann K, Lazier J. (2005). Dynamics of marine ecosystems: biological-physical interactions in the oceans. Third edition. Hoboken: Blackwell Publishing. pp.512

Mardis ER. (2008). Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* **9**: 387–402.

Margalef R. (1978). Life-forms of phytoplankton as survival alternatives in an unstable environment. *Oceanol acta* **1**: 493–509.

Marsjan PA, Oldenbroek JK. (2007). Molecular markers – a tool for exploring genetic diversity. In: *The State of the World's Animal Genetic Resources for Food and Agriculture*. pp 359–379.

Martínez JM, Schroeder DC, Larsen A, Bratbak G, Wilson WH. (2007). Molecular dynamics of *Emiliania huxleyi* and cooccurring viruses during two separate mesocosm studies. *Appl Environ Microbiol* **73**: 554–62.

Martínez JM, Schroeder DC, Wilson WH. (2012). Dynamics and genotypic composition of *Emiliania huxleyi* and their co-occurring viruses during a coccolithophore bloom in the North Sea. *FEMS Microbiol Ecol* **81**: 315–23.

McCann KS. (2000). The diversity-stability debate. *Nature* **405**: 228–33.

Medlin L, Barker G, Campbell L, Green J, Hayes P, Marie D, *et al.* (1996). Genetic characterisation of *Emiliania huxleyi* (Haptophyta). *J Mar Syst* **9**: 13–31.

Medlin LK, Kooistra WHCF. (2010). Methods to estimate the diversity in the marine photosynthetic protist community with illustrations from case studies: a review. *Diversity* **2**:973-1014.

Meyer J, Riebesell U. (2015). Reviews and Syntheses: Responses of coccolithophores to ocean acidification: a meta-analysis. *Biogeosciences* **12**: 1671–1682.

Mieszkowska N, Leaper R, Moore P, Kendall M, Burrows M, Lear D, *et al.* (2006). Marine biodiversity and climate change: assessing and predicting the influence of

climatic change using intertidal rocky shore biota. Scottish Natural Heritage Commissioned Report No.202 (ROAME No.F01AA402).

Miller CB. (2004). Biological Oceanography. Hoboken: Wiley-Blackwell. pp.504

Miller P, Groom S, McManus A, Selley J, Mironnet N. (1997). Panorama: a semi-automated AVHRR and CZCS system for observation of coastal and ocean processes. In: *RSS97: Observations and Interactions, Proceedings of the Remote Sensing Society, Reading, September 1997*. pp 539–544.

Montes-Hugo M, Doney SC, Ducklow HW, Fraser W, Martinson D, Stammerjohn SE, *et al.* (2009). Recent changes in phytoplankton communities associated with rapid regional climate change along the Western Antarctic Peninsula. *Science (80-)* **323**: 1470–1473.

Moolna A, Rickaby REM. (2012). Interaction of the coccolithophore *Gephyrocapsa oceanica* with its carbon environment: response to a recreated high-CO_2 geological past. *Geobiology* **10**: 72–81.

Müller M, Trull T, Hallegraeff G. (2015). Differing responses of three Southern Ocean *Emiliania huxleyi* ecotypes to changing seawater carbonate chemistry. *Mar Ecol Prog Ser* **531**: 81–90.

Müller MN, Antia AN, Laroche J. (2008). Influence of cell cycle phase on calcification in the coccolithophore *Emiliania huxleyi*. *Limnol Oceanogr* **53**: 506–512.

Nanninga HJ, Tyrrell T. (1996). Importance of light for the formation of algal blooms by *Emiliania huxleyi*. *Mar Ecol Prog Ser* **136**: 195–203.

Napp JM, Hunt GL. (2001). Anomalous conditions in the south-eastern Bering Sea 1997: linkages among climate, weather, ocean, and biology. *Fish Oceanogr* **10**: 61–68.

Nelson JR, Wakeham SG. (1989). A Phytol-substituted chlorophyll c from *Emiliania huxleyi* (Prymnesiophyceae). *J Phycol* **25**: 761–766.

Neufeld JD, Schäfer H, Cox MJ, Boden R, McDonald IR, Murrell JC. (2007). Stable-isotope probing implicates *Methylophaga* spp and novel Gammaproteobacteria in marine methanol and methylamine metabolism. *ISME J* **1**: 480–91.

O'Dea SA, Gibbs SJ, Bown PR, Young JR, Poulton AJ, Newsam C, *et al.* (2014). Coccolithophore calcification response to past ocean acidification and climate change. *Nat Commun* **5**: 5363.

Okada H, McIntyre A. (1977). Modern coccolithophores of the Pacific and North Atlantic Oceans. *Micropaleontology* **23**: 1–55.

Okada H, Mclntyre A. (1979). Seasonal distribution of modern coccolithophores in the Western North Atlantic Ocean. *Mar Biol* **54**: 319–328.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, *et al.* (2012). vegan: community ecology package. R package version 2.0-5.

Olafsson J, Olfsdottir S, Benoit-Cattin A, Danielsen M, Arnarson T, Takahashi T. (2009).

Rate of Iceland Sea acidification from time series measurements. *Biogeosciences Discuss* **6**: 5251–5270.

Paasche E. (2002). A review of the coccolithophorid *Emiliania huxleyi* (Prymnesiophyceae), with particular reference to growth, coccolith formation, and calcification-photosynthesis interactions. *Phycologia* **40**: 503–529.

Painter SC, Poulton AJ, Allen JT, Pidcock R, Balch WM. (2010). The COPAS'08 expedition to the Patagonian Shelf: Physical and environmental conditions during the 2008 coccolithophore bloom. *Cont Shelf Res* **30**: 1907–1923.

Parmesan C, Parmesan C, Yohe G, Yohe G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**: 37–42.

Peakall R, Smouse PE. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update. *Bioinformatics* **28**: 2537–9.

Peakall R, Smouse PE. (2006). GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* **6**: 288–295.

Pedrós-Alió C. (2006). Marine microbial diversity: can it be determined? *Trends Microbiol* **14**: 257–63.

Poulton A, Young J, Bates N, Balch W. (2011). Biometry of detached *Emiliania huxleyi* coccoliths along the Patagonian Shelf. *Mar Ecol Prog Ser* **443**: 1–17.

Poulton AJ, Stinchcombe MC, Achterberg EP, Bakker DCE, Dumousseaud C, Lawson HE, *et al.* (2014). Coccolithophores on the north-west European shelf: calcification rates and environmental controls. *Biogeosciences* **11**: 3919–3940.

Le Quéré C, Andres RJ, Boden T, Conway T, Houghton RA, House JI, *et al.* (2013). The global carbon budget 1959–2011. *Earth Syst Sci Data* **5**: 165–185.

Le Quéré C, Rödenbeck C, Buitenhuis ET, Conway TJ, Langenfelds R, Gomez A, *et al.* (2007). Saturation of the southern ocean CO_2 sink due to recent climate change. *Science (80-)* **316**: 1735–8.

Raffi I, Backman J, Fornaciari E, Pälike H, Rio D, Lourens L, *et al.* (2006). A review of calcareous nannofossil astrobiochronology encompassing the past 25 million years. *Quat Sci Rev* **25**: 3113–3137.

Rastogi G, Sani RK. (2011). Chapter 2: Molecular techniques to assess microbial community structure, function, and dynamics in the environment. In: *Microbes and microbial technology*. New York: Springer. pp.29-58

Raven J, Caldeira K, Elderfield H, Hoegh-Guldberg O, Liss P, Riebesell U, *et al.* (2005). Ocean acidification due to increasing atmospheric carbon dioxide. *R Soc* 60.

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander L V., Rowell DP, *et al.* (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J Geophys Res* **108**: 4407.

Read B a, Kegel J, Klute MJ, Kuo A, Lefebvre SC, Maumus F, *et al.* (2013). Pan genome

of the phytoplankton *Emiliania* underpins its global distribution. *Nature* **499**: 209–13.

Rees AP, Woodward EM., Robinson C, Cummings DG, Tarran G a, Joint I. (2002). Size-fractionated nitrogen uptake and carbon fixation during a developing coccolithophore bloom in the North Sea during June 1999. *Deep Sea Res Part II Top Stud Oceanogr* **49**: 2905–2927.

Regoli F, Principato G. (1995). Glutathione, glutathione-dependent and antioxidant enzymes in mussel, *Mytilus galloprovincialis*, exposed to metals under field and laboratory conditions: Implications for the use of biochemical biomarkers. *Aquat Toxicol* **31**: 143–164.

Richier S, Fiorini S, Kerros M-E, von Dassow P, Gattuso J-P. (2011). Response of the calcifying coccolithophore *Emiliania huxleyi* to low pH/high pCO_2: from physiology to molecular level. *Mar Biol* **158**: 551–560.

Rickaby REM, Hermoso M, Lee RBY, Rae BD, Heureux AMC, Balestreri C, *et al.* (submitted for publication - 2015). Environmental carbonate chemistry selects for phenotype of recently isolated strains of *Emiliania huxleyi*.

Rickaby REM, Schrag DP, Zondervan I, Riebesell U. (2002). Growth rate dependence of Sr incorporation during calcification of *Emiliania huxleyi*. *Global Biogeochem Cycles* **16**: 1–8.

Riebesell U, Bellerby RGJ, Engel A, Fabry VJ, Hutchins D a, Reusch TBH, *et al.* (2008). Comment on 'Phytoplankton calcification in a high-CO_2 world'. *Science* **322**: 1466; author reply 1466.

Riebesell U, Körtzinger A, Oschlies A. (2009). Sensitivities of marine carbon fluxes to ocean change. *Proc Natl Acad Sci U S A* **106**: 20602–9.

Riebesell U, Tortell PD. (2011). Effects of ocean acidification on pelagic organisms and ecosystems. In: *Ocean Acidification* (Gattuso J-P, Hansson L, eds). Oxford: Oxford University Press, pp 99–121.

Riebesell U, Zondervan I, Rost B, Tortell PD, Zeebe RE, Morel FM. (2000). Reduced calcification of marine plankton in response to increased atmospheric CO_2. *Nature* **407**: 364–367.

Robertson JE, Robinson C, Turner DR, Holligan P, Watson a. J, Boyd P, *et al.* (1994). The impact of a coccolithophore bloom on oceanic carbon uptake in the northeast Atlantic during summer 1991. *Deep Sea Res Part I Oceanogr Res Pap* **41**: 297–314.

Rokitta SD, de Nooijer LJ, Trimborn S, de Vargas C, Rost B, John U. (2011). Transcriptome analyses reveal differential gene expression patterns between the life-cycle stages of *Emiliania huxleyi* (Haptophyta) and reflect specialization to different ecological niches. *J Phycol* **47**: 829–838.

Rost B, Riebesell U. (2004). Coccolithophores and the biological pump: responses to environmental changes. *Coccolithophores* 99–125.

Rost B, Zondervan I, Wolf-Gladrow D. (2008). Sensitivity of phytoplankton to future changes in ocean carbonate chemistry: current knowledge, contradictions and

research directions. *Mar Ecol Prog Ser* **373**: 227–237.

Rousset F. (2008). genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour* **8**: 103–106.

Russell a. D. (2003). Biocide use and antibiotic resistance: the relevance of laboratory findings to clinical and environmental situations. *Lancet Infect Dis* **3**: 794–803.

Rynearson TA, Armbrust EV. (2000). DNA fingerprinting reveals extensive genetic diversity in a field population of the centric diatom *Ditylum brightwellii*. *Limnol Oceanogr* **45**: 1329–1340.

Rynearson TA, Armbrust EV. (2004). Genetic differentiation among populations of the planktonic marine diatom *Ditylum brightwellii* (Bacillariophyceae). *J Phycol* **40**: 34–43.

Rynearson TA, Armbrust EV. (2005). Maintenance of clonal diversity during a spring bloom of the centric diatom *Ditylum brightwellii*. *Mol Ecol* **14**: 1631–40.

Rynearson TA, Lin EO, Armbrust EV. (2009). Metapopulation structure in the planktonic diatom *Ditylum brightwellii* (Bacillariophyceae). *Protist* **160**: 111–121.

Sabine CL, Feely R a, Gruber N, Key RM, Lee K, Bullister JL, *et al.* (2004). The oceanic sink for anthropogenic CO_2. *Science* **305**: 367–71.

Sagan C. (1961). On the Origin and Planetary Distribution of Life. *Radiat Res* **15**: 174–192.

Schlötterer C. (1998). Genome evolution: Are microsatellites really simple sequences? *Curr Biol* **8**: R132–R134.

Schlötterer C, Tautz D. (1992). Slippage synthesis of simple sequence DNA. *Nucleic Acids Res* **20**: 211–215.

Schroeder DC, Biggi GF, Hall M, Davy J, Martinez JM, Richardson AJ, *et al.* (2005). A genetic marker to separate *Emiliania huxleyi* (Prymnesiophyceae) morphotypes. *J Phycol* **41**: 874–879.

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.

Schuller D, Cardoso F, Sousa S, Gomes P, Gomes AC, Santos MAS, *et al.* (2012). Genetic diversity and population structure of *Saccharomyces cerevisiae* strains isolated from different grape varieties and winemaking regions. *PLoS One* **7**: e32507.

Sekino K, Shiraiwa Y. (1996). Evidence for the involvement of mitochondrial respiration in calcification in a marine coccolithophorid, *Emiliania huxleyi*. *Plant Cell Physiol* **37**: 1030–1033.

Selkoe KA, Toonen RJ. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol Lett* **9**: 615–29.

Service PM, Rose MR. (1985). Genetic covariation among life-history components : the effect of novel environments. *Evolution (N Y)* **39**: 943–945.

Shutler JD, Smyth TJ, Land PE, Groom SB. (2005). A near-real time automatic MODIS data processing system. *Int J Remote Sens* **26**: 1049–1055.

Smyth TJ, Allen I, Atkinson A, Bruun JT, Harmer R a, Pingree RD, *et al.* (2014). Ocean Net Heat Flux influences seasonal to interannual patterns of plankton abundance. *PLoS One* **9**: e98709.

Smyth TJ, Fishwick JR, AL-Moosawi L, Cummings DG, Harris C, Kitidis V, *et al.* (2010). A broad spatio-temporal view of the Western English Channel observatory. *J Plankton Res* **32**: 585–601.

Sokal RR, Rohlf FJ. (1995). Biometry: the principles and practice of statistics in biological research. 3rd ed. New York: W. H. Freeman and Co. pp.880

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

Southward AJ, Langmead O, Hardman-Mountford NJ, Aiken J, Boalch GT, Dando PR, *et al.* (2005). Long-term oceanographic and ecological research in the Western English Channel. *Adv Mar Biol* **47**: 1–105.

Steele M, Ermold W, Zhang J. (2008). Arctic Ocean surface warming trends over the past 100 years. *Geophys Res Lett* **35**: L02614.

Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner H-W, *et al.* (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol Ecol* **19**: 21–31.

Stolte W, Kraay GW, Noordeloos A a M, Riegman R. (2000). Genetic and physiological variation in pigment composition of *Emiliania*. *Strain* **539**: 529–539.

Sunnucks P. (2000). Efficient genetic markers for population biology. *Trends Ecol Evol* **15**: 199–203.

Svanback R, Bolnick DI. (2007). Intraspecific competition drives increased resource use diversity within a natural population. *Proc R Soc B Biol Sci* **274**: 839–844.

Temperton B, Giovannoni SJ. (2012). Metagenomics: microbial diversity through a scratched lens. *Curr Opin Microbiol* **15**: 605–12.

Tesson SVM, Legrand C, van Oosterhout C, Montresor M, Kooistra WHCF, Procaccini G. (2013). Mendelian inheritance pattern and high mutation rates of microsatellite alleles in the diatom *Pseudo-nitzschia multistriata*. *Protist* **164**: 89–100.

Thomas CD, Thomas CD, Cameron A, Cameron A, Green RE, Green RE, *et al.* (2004). Extinction risk from climate change. *Nature* **427**: 145–8.

Van Tienderen PH, De Haan A a., Dan Der Linden CG, Vosman B. (2002). Biodiversity assessment using markers for ecologically important traits. *Trends Ecol Evol* **17**: 577–582.

Tsai IJ, Bensasson D, Burt A, Koufopanou V. (2008). Population genomics of the wild yeast *Saccharomyces paradoxus*: quantifying the life cycle. *Proc Natl Acad Sci* **105**:

4957–4962.

Turley C, Brownlee C, Findlay H, Mangi S, Ridgwell A, Schmidt D, *et al.* (2010). Ocean acidification. In: MCCIP Annual Report Card 2010-11, MCCIP Science Review. www.mccip.org.uk/arc.

Tyrrell T, Merico A. (2004). *Emiliania huxleyi*: bloom observations and the conditions that induce them. *Coccolithophores*.

Tyrrell T, Schneider B. (2008). Coccolithophores and calcite saturation state in the Baltic and Black Seas. *Biogeosciences, Eur Geosci Union*.

Tyrrell T, Taylor A. (1996). A modelling study of *Emiliania huxleyi* in the NE Atlantic. *J Mar Syst* 83–112.

VanKuren NW, den Bakker HC, Morton JB, Pawlowska TE. (2013). Ribosomal RNA gene diversity, effective population size, and evolutionary longevity in asexual Glomeromycota. *Evolution (N Y)* **67**: 207–224.

Vitousek PM, Mooney HA, Lubchenco J, Melillo JM. (1997). Human domination of Earth's ecosystems. *Science (80-)* **277**: 494–499.

van Bleijswijk J, van der Wal P, Kempers R, Veldhuis M, Young JR, Muyzer G, *et al.* (1991). Distribution of two types of *Emiliania huxleyi* (Prymnesiophyceae) in the northeast Atlantic region as determined by immunofluorescence and coccolith morphology. *J Phycol* **27**: 566–570.

van Bleijswijk JD, Kempers RS, Veldhuis MJ, Westbroek P. (1994). Cell and growth characteristics of types A and B of *Emiliania huxleyi* (Prymnesiophyceae) as determined by flow cytometry and chemical analyses. *J Phycol* **30**: 230–241.

van der Wal P, Kempers R, Veldhuis M. (1995). Production and downward flux of organic matter and calcite in a North Sea bloom of the coccolithophore *Emiliania huxleyi*. *Mar Ecol Prog Ser* **126**: 247–265.

van Emburg P. (1989). Coccolith formation in *Emiliania huxleyi*. PhD Thesis. Leiden University.

Walser B, Haag CR. (2012). Strong intraspecific variation in genetic diversity and genetic differentiation in *Daphnia magna*: the effects of population turnover and population size. *Mol Ecol* **21**: 851–61.

Ward JH. (1963). Hierarchical Grouping to Optimize an Objective Function. *J Am Stat Assoc* **58**: 236–244.

Westbroek P, Brown CW, Bleijswijk J Van, Brownlee C, Brummer GJ, Conte M, *et al.* (1993). A model system approach to biological climate forcing. The example of *Emiliania huxleyi*. *Glob Planet Change* **8**: 27–46.

Westbroek P, Young JR, Linschooten K. (1989). Coccolith production (biomineralization) in the marine alga *Emiliania huxleyi*. *J Protozool* **36**: 368–373.

White WB, Lean J, Cayan DR, Dettinger MD. (1997). Response of global upper ocean

temperature to changing solar irradiance. *J Geophys Res* **102**: 3255.

Widdicombe CE, Eloire D, Harbour D, Harris RP, Somerfield PJ. (2010). Long-term phytoplankton community dynamics in the Western English Channel. *J Plankton Res* **32**: 643–655.

Wilson WH, Tarran G a., Schroeder D, Cox M, Oke J, Malin G. (2002). Isolation of viruses responsible for the demise of an *Emiliania huxleyi* bloom in the English Channel. *J Mar Biol Assoc UK* **82**: S002531540200560X.

Winter A, Henderiks J, Beaufort L, Rickaby REM, Brown CW. (2014). Poleward expansion of the coccolithophore *Emiliania huxleyi*. *J Plankton Res* **36**: 316–325.

Wolfe G V., Sherr EB, Sherr BF. (1994). Release and consumption of DMSP from *Emiliania huxleyi* during grazing by *Oxyrrhis marina*. *Mar Ecol Prog Ser* **111**: 111–119.

Woodgate RA, Fahrbach E, Rohardt G. (1999). Structure and transports of the East Greenland Current at 75°N from moored current meters. *J Geophys Res* **104**: 18059–18072.

Wootton T, Pfister C, Forester J. (2008). Dynamic patterns and ecological impacts of declining ocean pH in a high-resolution multi-year dataset. *Proc Natl Acad Sci* **105**: 18848–18853.

Worm B, Barbier EB, Beaumont N, Duffy JE, Folke C, Halpern BS, *et al.* (2006). Impacts of biodiversity loss on ocean ecosystem services. *Science (80-)* **314**: 787–790.

Wright S, Jeffrey S. (1987). Fucoxanthin pigment markers of marine phytoplankton analysed by HPLC and HPTLC . *Mar Ecol Prog Ser* **38**: 259–266.

Young J. (1994a). Variation in *Emiliania huxleyi* coccolith morphology in samples from the Norwegian EHUX experiment, 1992. *Sarsia* 37–41.

Young J, Davis S, Bown P, Mann S. (1999). Coccolith ultrastructure and biomineralisation. *J Struct Biol* **126**: 195–215.

Young J, Geisen M, Cros L, Kleijne A, Sprengel C, Probert I, *et al.* (2003). A guide to extant coccolithophore taxonomy. *J Nannoplankt Res*.

Young J, Westbroek P. (1991). Genotypic variation in the coccolithophorid species *Emiliania huxleyi*. *Mar Micropaleontol* **18**: 5–23.

Young JR. (1994b). Functions of coccoliths. *Coccolithophores* 63–82.

Young JR, Henriksen K. (2003). Biomineralization within vesicles: the calcite of coccoliths. *Rev Mineral Geochemistry* **54**: 189–215.

Young JR, Poulton a. J, Tyrrell T. (2014). Morphology of *Emiliania huxleyi* coccoliths on the northwestern European shelf – is there an influence of carbonate chemistry? *Biogeosciences* **11**: 4771–4782.

Yu Y, Breitbart M, McNairnie P, Rohwer F. (2006). FastGroupII: a web-based bioinformatics platform for analyses of large 16S rDNA libraries. *BMC Bioinformatics* **7**: 57.